NORTHWESTERN UNIVERSITY

Structural and Functional Analyses of the Core Components of the Mammalian Sin3A Corepressor Complex

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Interdepartmental Biological Sciences Program

By

Yuan He

EVANSTON, ILLINOIS

June 2008

ABSTRACT

Structural and Functional Analyses of the Core Components of the Mammalian Sin3A Corepressor Complex

Yuan He

Eukaryotic gene transcription is a coordinated process involving the participation of a large number of proteins. Sequence-specific DNA-binding transcription factors function as activators or repressors of transcription and frequently recruit transcriptional coregulators that lack specific DNA-binding activity. These multi-protein complexes with intrinsic chromatin-modifying and chromatin-remodeling activities have emerged as key regulators of eukaryotic gene transcription. The Sin3 complex was one of the first corepressor complexes to be identified and is found in organisms ranging from yeast to human. In mammals, it is thought to be composed of at least ten polypeptides besides Sin3A/B. The Sin3 proteins appear to serve a dual role as molecular adapters bridging the enzymatic components of the complex with promoter-bound repressors and also as molecular scaffolds for the assembly of the complexes. To better understand how these multi-subunit complexes are recruited to specific loci on the genome and how the core components of the complexes are built into functional assemblies, I have conducted structure-function analysis toward the mammalian Sin3A corepressor complex using solution-state NMR spectroscopy as the primary tool.

My results suggested that the PAH1 and PAH2 domains in the mammalian Sin3 proteins are structurally independent and ruled out any heterotypic association between the paralogous mSin3A and mSin3B proteins via interactions involving the mSin3A PAH2 domain. Preliminary results from structure determination of the mSin3A PAH3–SAP30 SID complex indicated that the PAH3 domain forms a four-helix bundle similar to other PAH domains, but

with helix α3 much shorter than in the PAH1 and PAH2 domains. In addition, the SAP30 SID domain adopts a structure comprising three helices, which is a departure from the structural motif used by proteins targeting the mSin3A PAH1 and PAH2 domains. Furthermore, I discovered and determined the structure of a novel zinc finger motif in SAP30, which appears to function as a nucleic acid binding domain and implies a previously unknown function for the Sin3 corepressor complex. Finally, a bottom-up approach has been taken to incrementally reconstitute and characterize a sub-complex comprising subunits specific to the Sin3A/B corepressor complexes, and a portion of the co-expressed proteins remained in a soluble form. These studies have furthered our understanding of the mechanisms of recruitment and complex assembly for Sin3 while providing paradigmatic basis for characterization of other HDAC-associated corepressor complexes.

Acknowledgements

The work included here could not have been accomplished without the help from many people. The Sin3 work was done in collaboration with the Ayer lab at Huntsman Cancer Institute in Salt Lake City, Utah, and the ubiquitin work was done in collaboration with the Hicke lab at Northwestern University, Evanston, Illinois. Herein I thank both past and present members of these labs for their contributions.

I thank both past and present members of the Radhakrishnan lab. In particular, Rebecca Imhoff and Ivana Vucojicic for their contributions in mSin3A PAH3, SAP30, RBP1, and polycistronic co-expression construct generation and protein purification. I also thank Anirban Sahu for his contributions in SAP30 ZnF mutants construct generation and protein purification. In addition, I thank Richard Kang, Hisae Matsuura for their initial development of the ubiquitin work and James Zhang for his aid in the structure determination of the ubiquitin work.

I thank Svetoslava Stamenova, Michael French, and Zachary Kramer of the Hicke lab for their contributions to the ubiquitin work. In particular, Svetoslava discovered the Sla1 SH3-3 domain as an ubiquitin binder.

I thank Mohan Kaadige of the Ayer lab for conducting histone modification pull-down assay to identify the potential binder for SAP30 ZnF.

I thank the NMR managers Ben Ramirez and Yongbo Zhang for their assistance with NMR data collection and discussion involving data analysis.

I also thank my thesis committee, Amy Rosenzweig, Curt Horvath, Alfonso Mondragón, and Hamid Band for their essential guidance in my research over the years.

Finally, I thank my mentor Ishwar Radhakrishnan for the research opportunity and teaching me not only how to perform research but also how to think in a scientific way. I greatly appreciate the years I have spent in his lab and his always keeping the office door open.

As always, I thank my parents and my wife Hong for their ongoing support and encouragement.

Table of Contents

Page

List of Figures

List of Tables

CHAPTER ONE

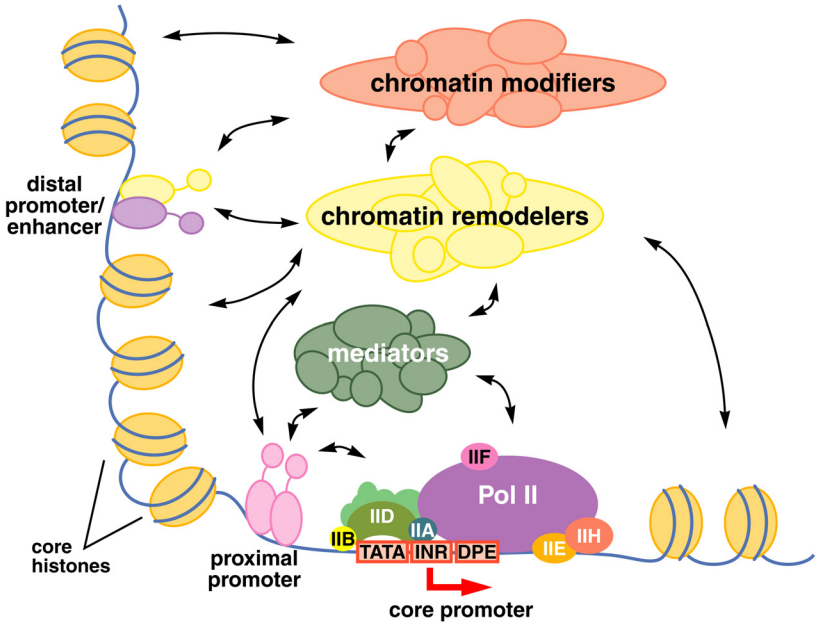
Introduction

1.1 Introduction

Since the structure of DNA, the depository of hereditary information in the cell was discovered in 1953, remarkable progress has been accomplished in deciphering the mechanisms in which the four-letter code directs the formation of life. According to the "central dogma" of molecular biology, transcription occurs as the first step for a living cell to read out the genetic information from a gene. As expected, incorrect spatial and/or temporal patterns of gene transcription lead to the pathogenesis of many human diseases (Stillman, 1998).

Although the mechanism of eukaryotic gene transcription is similar in concept to those found in prokaryotes, the regulatory network in eukaryotes is far more intricate, not only with a larger number of genes to coordinate but also requiring much more protein factors to participate (Elgin and Workman, 2000). In eukaryotes, RNA polymerase II (pol II) along with general transcription factors function as the basal machinery for the synthesis of mRNA precursors coding for all the proteins. Like its counterparts in bacteria, pol II is a multi-subunit enzyme, which is responsible for transcribing the gene, but without the capability for recognizing promoters on its own (Roeder, 1991). A series of general transcription factors including IIA, IIB and IID (TFIIA, TFIIB and TFIID) are assembled onto the basal promoters in a defined manner to form the preinitiation complex, which not only serves as a bridge for the pol II binding but also ensures accurate initiation and triggers further elongation (Figure 1.1). Herein I am focusing on eukaryotic gene transcription unless otherwise specified.

Figure 1.1 **Eukaryotic gene transcription regulation.** Eukaryotic transcription is a coordinated process involving the participation of a large number of proteins. The basal transcription machinery includes the RNA polymerase, a multi-subunit enzyme that is responsible for transcribing the gene and the general transcription factors that ensure accurate initiation of transcription.



Sequence-specific DNA binding transcription factors function as activators or repressors of transcription by simultaneously recognizing both the promoter sequences and other coregulators through their DNA-binding domains and transactivation or transrepression domains, respectively (Buratowski, 1995). These multi-subunit coregulators with intrinsic chromatin-modifying and/or chromatin-remodeling activities lack specific DNA-binding activity, yet have emerged as key regulators of eukaryotic gene transcription. The mediator complex serves as a bridge between DNA-bound factors and components of the basal transcriptional machinery, all of which appear to work together as a holoenzyme (Koleske and Young, 1995). Chromatin remodelers can enhance nucleosome mobility and dynamics by repositioning nucleosomes through an ATP-dependent process. Chromatin modifiers post-translationally modify the protein components of chromatin by acetylation, methylation, phosphorylation, ubiquitination, and so on (Kingston et al., 1996). The role of the chromatin remodeling and modifying activities is to perturb or restore chromatin structure so as to make the template more or less permissive to transcription. Although different promoters such as those containing a TATA box, initiator (INR) element, and downstream promoter element (DPE) with preinitiation complex associated are important for basal transcription, additional upstream regulatory elements including diverse promoters and enhancers are required for further enhancing transcriptional activity (Figure 1.1).

Eukaryotic gene transcription is subject to multiple levels of regulation, one of which has been mentioned above as the involvement of different types of transcription factors. In addition, both the compact packing of the nucleosome core particle and higher-order structure of chromatin have also been suggested in affecting the transcriptional level (Kornberg and Lorch, 1991). In fact, it is becoming increasingly apparent that chromatin has a very dynamic local

environment according to various genetic and biochemical studies, thereby adding another level of regulation to gene transcription.

1.2 Post-translational modification of histones in eukaryotic gene transcription regulation

Histones are relatively small, yet highly conserved proteins with masses between 10 and 21 kDa, and constitute the major protein component of chromatin. There are five families of histones including H2A, H2B, H3, and H4, also known as the core histones in addition to the linker H1 and H5 histones that play a distinct role in chromatin condensation compared to the others. Histone octamer composed of two copies of each core histone is wrapped around by 146 bp of DNA for less than two turns to form the basic unit of chromatin – nucleosome. The core histone proteins in chromatin are frequently subject to numerous reversible modifications at multiple sites, mostly located at their flexible N- and C-terminal tails protruding out of the nucleosome. Distinct combinations of those covalent modifications, including acetylation, methylation, ubiquitination, and phosphorylation, are thought to define a 'histone code', that in turn leads to different transcriptional responses (Jenuwein and Allis, 2001). Furthermore, these codes are essential in maintaining different dynamic states of higher-order chromatin structure, thereby playing important roles in gene replication, DNA repair and genome integrity maintenance (Bhaumik et al., 2007). I summarize below various aspects of two better-characterized histone modifications – acetylation and methylation –including the mechanisms by which enzymes define the steady state levels of these chromatin modifications on the one hand, and how these modifications are recognized by a variety of chromatin binding modules on the other.

Acetylation

The most extensively studied modification of histones is acetylation, whose global level is finely tuned by the opposing catalytic actions of two distinct types of enzyme, histone acetyltransferases (HATs) or histone deacetylases (HDACs). HATs catalyze the addition of an acetyl group from acetyl coenzyme A (acetyl-CoA) to the ε-amine of lysine, neutralizing its positive charge and potentially keeping chromatin under an 'open' or transcriptionally-permissive environment, whereas HDACs hydrolyze the acetyl group off the chromatin substrates thereby switching the chromatin into a 'closed' or repressive architecture for transcription downregulation. Interestingly, these enzymes are commonly found in multi-subunit transcriptional coregulator complexes. For example, various HATs have been found in a number of transcription coactivators, such as Gcn5/PCAF (Herrera et al., 1997), CBP/p300 (Korzus et al., 1998) and p160/SRC-1 (Spencer et al., 1997), while transcription corepressors including mSin3A (Hassig et al., 1997; Laherty et al., 1997), NCoR/SMRT (Guenther et al., 2001) and Mi-2/NURD (Xue et al., 1998) possess intrinsic HDAC activity.

Despite its role in disrupting higher-order chromatin structure (Wolffe and Hayes, 1999), acetylation can also be recognized by diverse transcriptional coregulator complexes harboring a protein-protein interaction module called bromodomains (Zeng and Zhou, 2002). Bromodomains were the first structurally characterized histone-binding module and comprise an antiparallel left-handed four-helix bundle with a hydrophobic cavity on one end of the protein serving as the acetyl-lysine (AcK) binding site (Dhalluin et al., 1999). The interface between the protein and the ligand is dominated by hydrophobic interactions and evolutionarily reminiscent of the recognition mode between acetyl-CoA and histone acetyltransferases. A number of proteins containing multiple bromodomains have been since characterized and they appear to

recognize multiple acetylation marks within both histone and non-histone proteins (Jacobson et al., 2000; VanDemark et al., 2007).

Methylation

The majority of histone methylation has been shown to occur on histones H3 and H4. Unlike lysine residue as the exclusive target for HAT, both lysine and arginine residues can be methylated and there can also be multiple levels of methylation toward the same residue, such as mono- or dimethylation for arginine and mono-, di- or trimethylation for lysine (Bhaumik et al., 2007). Another distinct feature of histone methylation is the great diversity of histone methyltransferases (HMTs) and the high specificity of each enzyme. Histone methylation has become one of the most active fields in chromatin research during the past decade partially because of the discovery of methylated arginine deiminase, methylated lysine amine oxidase and hydroxylase, all of which are also referred to as histone demethylases (Klose and Zhang, 2007; Shi, 2007). Like HDAC's activity in removing acetyl group from chromatin substrates, these demethylases are responsible for actively antagonizing HMTs in the overall turnover of the modification. Likewise in the case of methyltransferase, these demethylases also have their unique substrate specificities.

Five different protein folds have been identified thus far for specifically targeting histone methylation, including chromodomain, tudor domain, MBT domain, PHD-finger module, and WD40 repeats (Taverna et al., 2007). These histone-binding modules can be further divided into two general families, the Royal superfamily folds (chromodomain, tudor, MBT) sharing an SH3-like β-barrel structure in addition to the PHD-finger family comprising a Cys_4-His-Cys_3 type zinc-finger. Different methylation binding modules come in different flavors. Whereas some

readers can differentiate state-specific methylation, there are also examples for sequence-context specificity (Taverna et al., 2007).

1.3 HDAC-associated transcriptional corepressor complexes

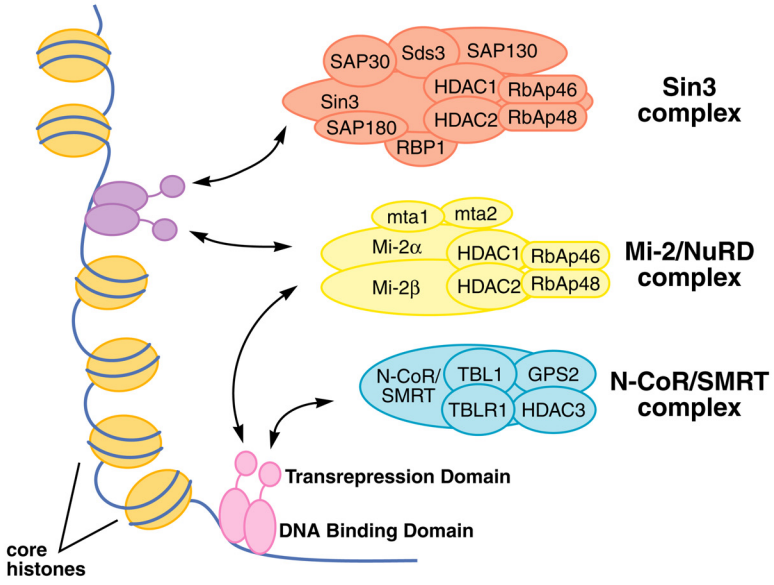
Histone deacetylation of acetylated lysine residues is correlated with restricted accessibility of underlying DNA sequences and thereby the decrease of transcriptional activity. These enzymes involved in maintaining the steady-state levels of cellular acetylation are often found in various multi-protein transcriptional corepressor complexes. How such a small group of complexes is recruited to specific loci of the genome by either DNA-binding factors or proteins harboring chromatin-binding modules is of great interest.

The mammalian HDACs have been grouped into three classes based on their homology to the ones found in yeast. Class I HDACs (1, 2, 3, 8, 11) and class II HDACs (4, 5, 6, 7, 9, 10) are similar to Rpd3p and Hda1p in yeast, respectively, whereas class III HDACs are unrelated at the sequence level to Class I and Class II HDACs and are homologous to Sir2p in yeast (Thiagalingam et al., 2003).

The cellular role of Class I HDAC3 has been revealed following the discovery of nuclear receptor corepressor (N-CoR)/silencing mediator of retinoic and thyroid receptors (SMRT)-HDAC3 corepressor complex (Karagianni and Wong, 2007). Ubiquitously expressed HDAC3 can not only be found in the nucleus where HDAC1 and HDAC2 are exclusively located, but also exists in the cytoplasm and the plasma membrane. It shares approximately 50% identity with both HDAC1 and HDAC2, and is capable of catalyzing the deacetylation of diverse histone substrates. Both N-CoR and SMRT were discovered as the scaffold proteins connecting the unliganded nuclear receptors such as thyroid hormone (TR) and retinoic acid receptors (RAR)

with the catalytic moiety – HDAC3, thereby coupling transcriptional repression to histone deacetylation (Figure 1.2). Biochemical purification studies have revealed additional components of the complex including the WD-40 repeats containing protein transducin β-like 1 (TBL1), TBL1-related protein (TBLR1), and the G-protein pathway suppressors 2 (GPS2) (Karagianni and Wong, 2007). While mSin3A/B and other HDACs have also been reported to associate with this corepressor, a core complex formed by N-CoR/SMRT, TBL1/TBLR1, GPS2 and HDAC3 is thought to be a functional entity.

Figure 1.2 **Coregulator complexes in transcriptional repression.** Histone deacetylases are commonly found in various transcriptional corepressor complexes that do not possess intrinsic DNA-binding activity. They are recruited to specific loci of the genome by either DNA-binding factors or proteins harboring chromatin-binding modules. The Sin3 and Mi-2/NuRD complexes share a common sub-complex comprising HDAC1, HDAC2, RbAp46 and RbAp48, while gene repression exerted by the N-CoR/SMRT complex relies on the action of another closely related enzyme – HDAC3.



The Mi-2/nucleosome remodeling and deacetylase (NuRD) complex is a unique example of HDAC-associated corepressor complex which not only have enzymatic activity of HDACs but also contain chromatin remodeling activity as its name implies (Denslow and Wade, 2007). The two helicase-like ATPases Mi-2α and Mi-2β belong to a subclass of the SWI/SNF family of remodeling complexes and are capable of sliding nucleosomes along the DNA. They function not only as a molecular scaffold for the assembly of the entire complex, but also bring in another chromatin-related enzymatic function – histone deacetylation (Figure 1.2). It is currently unclear how these two independent enzymatic activities are integrated for its typical function in transcriptional repression. Like the N-CoR/SMRT complex mentioned above, several different forms of Mi-2/NuRD complex have also been reported based on different complex purification strategies, but all of them share a common sub-complex comprising HDAC1, HDAC2, retinoblastoma protein (Rb)-associated proteins RbAp46 and RbAp48 (Jepsen and Rosenfeld, 2002). In fact, this particular sub-complex also exists in the Sin3 corepressor complex that will be discussed below. The precise function of the other two components of the Mi-2/NuRD complex – mta1 and mta2 still remain unclear.

1.4 The mammalian Sin3 corepressor complex

The Sin3 complex is a somewhat better-characterized HDAC-associated corepressor complex that also harbors the HDAC1/HDAC2/RbAp46/RbAp48 sub-complex in mammalian cells (Silverstein and Ekwall, 2005). Unlike the Mi-2/NuRD corepressor complex, Sin3 has been found in a wide variety of species including plant, animal and yeast as well. Besides the scaffolding proteins Sin3A/B and the enzymatic sub-complex, at least another six Sin3-associated polypeptides exist in this mega-dalton mammalian complex, including SAP25,

SAP30, SAP45/Sds3, SAP130, SAP180/BCAA and RBP1 (Figure 1.2). The Sin3 protein (~150 kDa) appears to serve a dual role as molecular adapter bridging the enzymatic components of the complex with DNA-bound repressors and also as molecular scaffold for the assembly of the complex. Furthermore, the Sin3 complex has been implicated in collaboration with a number of other chromatin-directed activities including SWI/SNF family of remodeling complexes such as Brg1 and hBrm (Sif et al., 2001), glycosaminyltransferases such as OGT (Yang et al., 2002), and HMTs such as ESET (Yang et al., 2003).

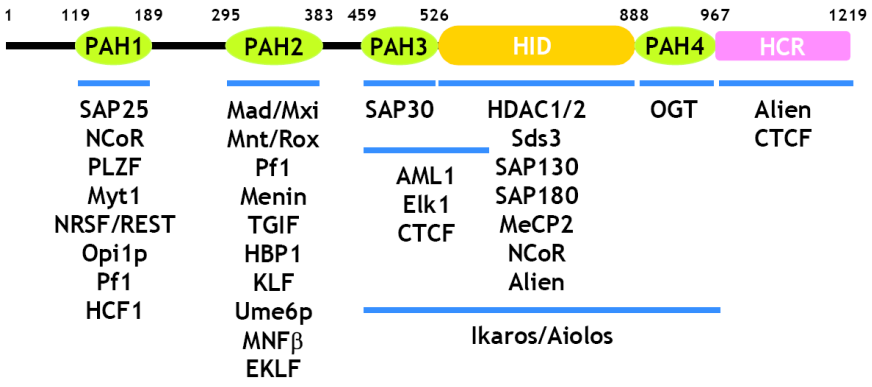
Recruitment of the mammalian Sin3A corepressor complex

The Mad family of proteins were the first series of transcription factors revealed in direct involvements in Sin3-dependent transcriptional repression (McArthur et al., 1998; Schreiber-Agus and DePinho, 1998). Since then, a plethora of gene-specific transcription factors have been described in recruiting the Sin3 corepressor complex such as p53 (Murphy et al., 1999), NRSF/REST (Huang et al., 1999; Naruse et al., 1999), AML1 (Lutterbach et al., 2000), TGIF (Wotton et al., 2001), Krüppel like factors (KLFs) (Zhang et al., 2001), HBP1 (Swanson et al., 2004), and Myt1 (Romm et al., 2005). The essential role of Sin3 in neoplastic and normal development is further emphasized by both mutational analysis and conditional knockout experiments of Sin3 (Cowley et al., 2005; Dannenberg et al., 2005).

The evolutionarily conserved Sin3 proteins harbor six individual domains including four imperfect repeats of paired amphipathic helices (PAH) motifs, a HDAC interaction domain (HID), and a C-terminal highly conserved region (HCR) (Figure 1.3) (Silverstein and Ekwall, 2005). In the emerging view of the domain structure, the two PAH domains in the N-terminus of the mSin3A protein appear to function mainly in the recruitment of diverse promoter-bound

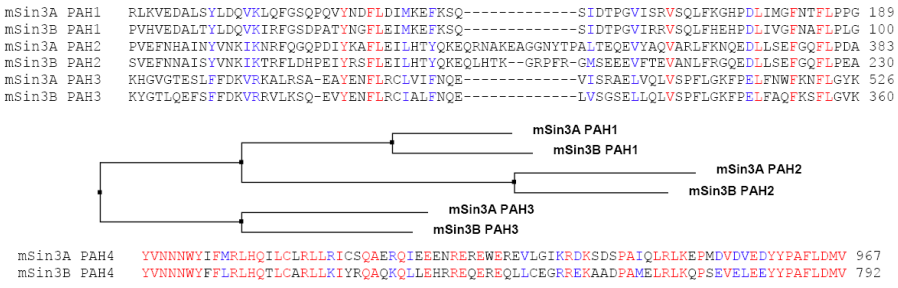
transcription factors, whereas the portion traversing both PAH3 and HID domains is responsible for complex assembly. It is currently unclear what specific role the C-terminal region of Sin3 plays.

Figure 1.3 **Domain structure of the mammalian Sin3A corepressor complex.** The corresponding boundaries of various domains in mSin3A are depicted by the residue numbers on the top. Regions of the protein required for either recruitment of protein factors or interaction with the core components of the complex are marked as blue lines.



Sequence analysis indicates that PAH1 and PAH2 domains are closely related to each other, but PAH3 departs from them based on the low degree of sequence similarity (Figure 1.4). Although PAH4 was originally predicted as a paired amphipathic helix domain as the other three, it is unlikely to share the same fold due to poor sequence conservation. At the structural level, most of the studies have been focused on the PAH1 and PAH2 domains. Both domains have been shown to adopt a left-handed four-helix bundle structure according to a number of structural analyses toward the Sin3 interaction domain (SID)-PAH complexes including Mad1 SID-mSin3A PAH2 (Brubaker et al., 2000), HBP1 SID-mSin3A PAH2 (Swanson et al., 2004), NRSF SID-mSin3B PAH1 (Nomura et al., 2005), and SAP25 SID-mSin3A PAH1 (Sahu et al., 2008). These studies also revealed that two distinct types of sequence motifs within various transcription factors bind to the same hydrophobic cleft on the corresponding PAH domains through their non-polar side of the amphipathic helix. Moreover, the Mad1 SID (type I) adopts a reversed orientation relative to the HBP1 SID in the interactions with mSin3A PAH2 (type II), and a similar scheme is shared by the SIDs of NRSF and SAP25 in their PAH1 interactions. However, there are a number of interactors possessing neither type of sequence motif yet still have the capabilities to bind to the one of the PAH domains, which indicates that additional recognition modes likely exist. Alternatively, some of these interactions might be mediated by proteins that harbor known motifs.

Figure 1.4 **Sequence analysis of the PAH domains in the mammalian Sin3 proteins.** Multiple sequence alignment of the mSin3A/B PAH domains based on CLUSTAL W. Their relatedness are shown in the evolutionary tree except for the PAH4 domains which are distinctly related to the others.

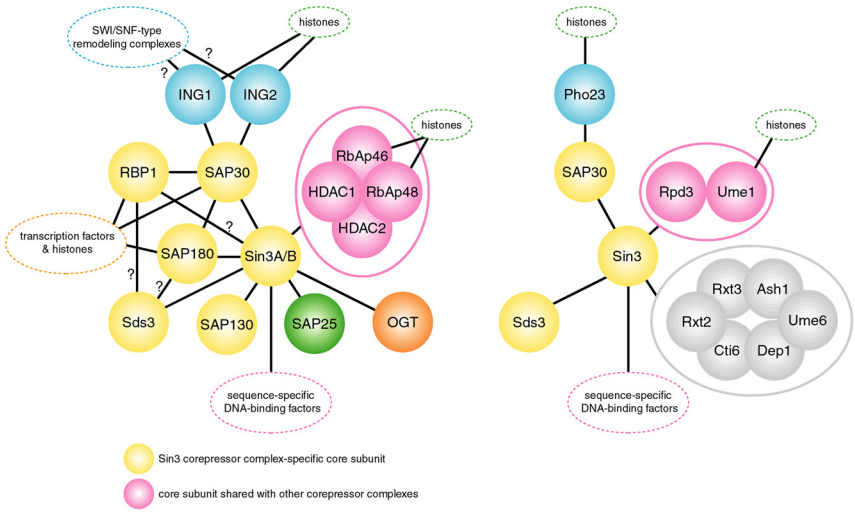


Assembly of the mammalian Sin3A corepressor complex

A number of biochemical and mass spectrometric studies have implicated SAP30, Sds3/SAP45, SAP130, SAP180/BCAA and RBP1 as the key constituents of the mammalian Sin3 corepressor complexes (Figure 1.5), and the PAH3 and HID domains are thought to harbor all the necessary determinants for interactions with these proteins (Figure 1.3) (Silverstein and Ekwall, 2005).

As one of the best-studied interactors of the PAH3 domain, SAP30 seems to play a particularly important role in stabilizing the core complex (Laherty et al., 1998; Zhang et al., 1998). Not only is it consistently detected in the corepressor complex ranging from yeast to human, its deletion leads to a similar phenotype as those of *sin3* and *rpd3* in yeast (Zhang et al., 1998; Sun and Hampsey, 1999; Bernstein et al., 2000). However, about it is unclear how this rather small protein (220 aa) is able to simultaneously engage so many targets besides Sin3 including RBP1 (Lai et al., 2001), SAP180 (Binda et al., 2006), the ING family proteins (Kuzmichev et al., 2002; Shi et al., 2006) and other transcription factors such as the NCoR, CIR, and yin yang 1 (YY1), (Laherty et al., 1998; Hsieh et al., 1999; Huang et al., 2003) (Figure 1.5).

Figure 1.5 **Interaction networks of the Sin3 corepressor complexes' components in mouse (left) and yeast (right).** Orthologues of various components are shown using identical colors and similar positions. Subunits identified in only a subset of the complex are colored differently.



As another constitutive component of the mammalian Sin3 corepressor complex, Sds3/SAP45 has been implicated in associating with the HID region of Sin3 together with two other key components of the complex – SAP130 and SAP180/BCAA (Alland et al., 2002; David et al., 2003; Fleischer et al., 2003; Carrozza et al., 2005a). In fact, the Sin3 corepressor complex has also been shown to target breast cancer metastasis suppressor 1 (BRMS1) and p40, both of which are Sds3/SAP45's homologues and can associate with RBP1 specifically (Nikolaev et al., 2004). Moreover, RBP1 and SAP180/BCAA share extensive similarities not only at the sequence level but also in their binding partners such as SAP30 (Binda et al., 2006). Accordingly, the Sin3 corepressor complex is cooperatively constituted by all these core components within a quite complicated interaction network (Figure 1.5). However, little is known about how the core corepressor complex is assembled and how the HDAC sub-complex is engaged.

Saccharomyces cerevisiae & pombe Rpd3 complex (Rpd3L & Rpd3S complexes)

The Sin3 corepressor complex plays an important role in the promoter-dependent transcription repression. But recent genetic and biochemical studies conducted in budding and fission yeast have identified a smaller HDAC-associated Sin3 complex (Rpd3S) harboring unique subunits that is absent in the larger complex (Rpd3L) (Table 1.1) (Carrozza et al., 2005b; Keogh et al., 2005; Nicolas et al., 2007). This newly described complex turns out to function as a repressor for aberrant transcription initiation in intragenic regions. Its current model suggests that a specific histone mark (H3K36Me2/3) directed by the elongating RNA pol II-associated HMT (Set2) serves a recruiting signal for the Sin3 complex (Rpd3S), which in turn restore a 'closed' chromatin structure via diminishing the high level of acetylation associated with

transcribing pol II. Interestingly, the two unique components of Rpd3S in *Saccharomyces cerevisiae*, Eaf3 and Rco1, are both necessary for interactions with methylated H3K36 and it appears that a combination of Eaf3 chromodomain and Rco1 PHD-finger domain are sufficient for chromatin targeting (Li et al., 2007).

Table 1.1 **Distinct HDAC-associated Sin3 corepressor complexes serving various putative roles.** The canonical Sin3 corepressor complexes recruited by diverse promoter-bound transcription factors function in suppression of aberrant transcription in centromeric regions, whereas a compositionally distinct format of the complex appears to be involved in suppressing intragenic transcription initiation.

Complex I *S. pombe*	Rpd3L *S. cerevisiae*	Mammalian
Pst1	Sin3	Sin3A/B
Clr6	Rpd3	HDAC1/2
Png2	Pho2	ING1b/2‡
Sds3	Sds3	Sds3
Prw1	Ume1	RbAp46/48
-	SAP30	SAP30
Cti6*,Rxt2*,Rxt3*	Cti6,Rxt2,Rxt3	-
-	Dep1	-
-	-	RBP1/SAP180,SAP130
-	-	OGT‡,SAP25‡

‡: not a constitutive or 'core' component
*: exists in *S. pombe* but not detected in Complex I

Complex II *S. pombe*	Rpd3S *S. cerevisiae*	Mammalian$
Pst2	Sin3	Sin3A/B
Clr6	Rpd3	HDAC1/2
Cph1/2	Rco1	Pf1
Alp13	Eaf3	MRG15
Prw1	Ume1	RbAp46/48

$: not proteomically-characterized

CHAPTER TWO

Solution NMR Studies of Apo-mSin3A and mSin3B Reveal that the PAH1 and PAH2 Domains are Structurally Independent

2.1 Introduction

The recruitment of chromatin-modifying and/or chromatin-remodeling enzymes by sequence-specific DNA-binding factors is a common, albeit important, step in the pathway leading to the activation or repression of target gene transcription in eukaryotes (Ptashne and Gann, 2002). An impressive array of chromatin-modifying and chromatin-remodeling activities has been identified over the past decade but these enzymes are frequently found in large multi-protein coregulator complexes with the additional subunits presumably lending specificity to the process (Elgin and Workman, 2000). The Sin3 corepressor complex comprising at least ten subunits is one of only a handful of major corepressor complexes identified thus far in mammalian cells that exerts its negative effects on gene transcription through the action of histone deacetylases (HDACs; (Hassig et al., 1997; Laherty et al., 1997; Laherty et al., 1998; Zhang et al., 1998; Alland et al., 2002; Fleischer et al., 2003)). The Sin3 corepressor has a dual role functioning not only as a molecular scaffold for complex assembly but also as a molecular adapter bridging HDACs with an astonishingly large and diverse group of DNA-binding transcription factors and chromatin-binding proteins (Ayer, 1999; Knoepfler and Eisenman, 1999; Silverstein and Ekwall, 2005). The interactions are generally mediated through one or more of six discrete regions conserved from yeast to human including four imperfect copies of the PAH domain. The second copy of the PAH domain (PAH2) is the site for interactions with numerous factors including members of the Mad family (McArthur et al., 1998; Schreiber-Agus and DePinho, 1998), Mnt/Rox (Hurlin et al., 1997), Pf1 (Yochum and Ayer, 2001), HBP1 (Swanson et al., 2004), the KLF family (Zhang et al., 2001), TGIF (Wotton et al., 2001), Menin (Kim et al., 2003), MNF-β (Yang et al., 2000), and EKLF (Chen and Bieker, 2004).

Our previous structural studies revealed that the mammalian Sin3A (mSin3A) PAH2 domain interacted through a hydrophobic cleft with diverse targets by binding to chain-reversible sequence motifs, thereby defining at least two distinct classes of PAH2 interactors (Brubaker et al., 2000; Swanson et al., 2004). The apo-mSin3A PAH2 domain exhibits conformational heterogeneity, with one of the conformers existing in a partially unfolded state, and homodimerizes with modest affinity suggesting plausible mechanisms for binding to a broad range of targets through new surfaces on the one hand and for regulating Sin3 function via occlusion of the hydrophobic cleft on the other (Zhang et al., 2006). The PAH2 domain of the paralogous mSin3B protein engages targets in a manner similar to mSin3A PAH2 (Spronk et al., 2000; van Ingen et al., 2004), but the apo-protein is monomeric and is fully folded, although a small population of a minor conformer has been reported (van Ingen et al., 2006). The properties of these domains in the context of the full-length or much larger fragments of the Sin3 protein are unexplored and the assumption of the autonomous nature of the individual domains at the structural level has not been thoroughly tested. The results of my studies described below have been recently published (He and Radhakrishnan, 2008).

2.2 Materials and Methods

Expression and purification of mSin3A PAH2, mSin3A PAH1/PAH2, mSin3B PAH2 and mSin3B PAH1/PAH2

The coding sequence of mSin3A PAH1/PAH2 corresponding residues 118−385 was amplified by PCR and inserted into the pMCSG7 expression vector (Stols et al., 2002). All cloned gene segments were confirmed by DNA sequencing. *E. coli* BL21(DE3) cells (Novagen, Madison, WI) containing the vector were grown at 37 °C in M9 minimal media. The growth

temperature was shifted to 20 °C when the $OD_{600\ nm}$ reached approximately 0.7. Expression of the His_6-tagged protein was induced using 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and the cells were harvested 16 h thereafter. Cell pellets were suspended in 20 mM sodium phosphate buffer (pH 7.5) containing 0.3 M sodium chloride, 2 mM Tris (2-carboxy-ethyl) phosphine hydrochloride (TCEP), 1 mM phenyl-methylsulfonyl (PMSF), 1 μM leupeptin, 1 mM pepstatin, and 0.1% Triton X-100, lysed via sonication and centrifuged. The supernatant was incubated with the His-Select Nickel resin (Sigma, St. Louis, MO) for 30 minutes. Bound proteins were cleaved from the resin by incubating with tobacco etch virus (TEV) protease for 4 h at 22 °C followed by overnight incubation at 4 °C. The mixture was centrifuged and the desired protein in the supernatant was purified to homogeneity via reversed-phase HPLC using a C18 column (Vydac, Hesperia, CA) and a linear gradient of 0.1% trifluoroacetic acid (TFA) and 0.1% TFA in 80% acetonitrile and lyophilized. An mSin3A PAH1/PAH2 sample uniformly labeled with ^{15}N isotope was produced as described above, except that cells were grown in M9 minimal media containing ^{15}N-ammonium sulfate (Spectra Stable Isotopes, Columbia, MD).

The coding sequences of mSin3A PAH2, mSin3B PAH2 and mSin3B PAH1/PAH2 domains corresponding to residues 295−385, 145−252 and 29−252, respectively, were amplified by PCR and inserted into the pMCSG10 expression vector (Stols et al., 2002). All cloned gene segments were confirmed by DNA sequencing. *E. coli* BL21(DE3) cells (Novagen, Madison, WI) containing the vector were grown at 37 °C in M9 minimal media. The growth temperature was shifted to 20 °C when the $OD_{600\ nm}$ reached approximately 0.7. Expression of the GST-tagged protein was induced using 1 mM IPTG, and the cells were harvested 16 h thereafter. Cell pellets were suspended in PBS buffer (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 8.5), 5 mM EDTA, 2 mM DTT, 1 mM PMSF, 1 μM leupeptin, 1 mM pepstatin,

and 0.1% Triton X-100, lysed via sonication and centrifuged. The supernatant was incubated with glutathione sepharose resin (GE Healthcare) for 30 minutes. Bound proteins were cleaved from the resin by incubating with TEV protease for 4 h at 22 °C followed by overnight incubation at 4 °C. The mixture was centrifuged and the target protein in the supernatant was purified to homogeneity via reversed-phase HPLC using a C18 column and a linear gradient of 0.1% TFA and 0.1% TFA in 80% acetonitrile and lyophilized. Samples of mSin3A PAH2, mSin3B PAH2 and mSin3B PAH1/PAH2 uniformly labeled with ^{15}N isotope were produced as described above, except that cells were grown in M9 minimal media containing ^{15}N-ammonium sulfate.

NMR spectroscopy

NMR samples were prepared by dissolving the dry, lyophilized protein powder in 20 mM sodium phosphate buffer (pH 6.0) containing 0.2% (w/v) NaN_3 and 2 mM DTT-d_{10}. Protein concentrations were determined spectrophotometrically (Gill and von Hippel, 1989). Identical sample concentrations were employed for recording NMR spectra of the recombinant mSin3A/B PAH2 and PAH1/PAH2 proteins. NMR data were acquired on a Varian Inova 600 MHz spectrometer at 25 ºC. NMR data processing and analysis were performed using an in-house modified version of Felix 98.0 (Felix NMR, San Diego, CA).

2.3 Results and Discussion

The weak tendency of the apo-mSin3A PAH2 domain to dimerize (K_d = 280 μM) prompted us to ask whether this might reflect an innate ability to interact with related PAH domains. Biologically, this could be a mechanism to thwart low-affinity interactions with the

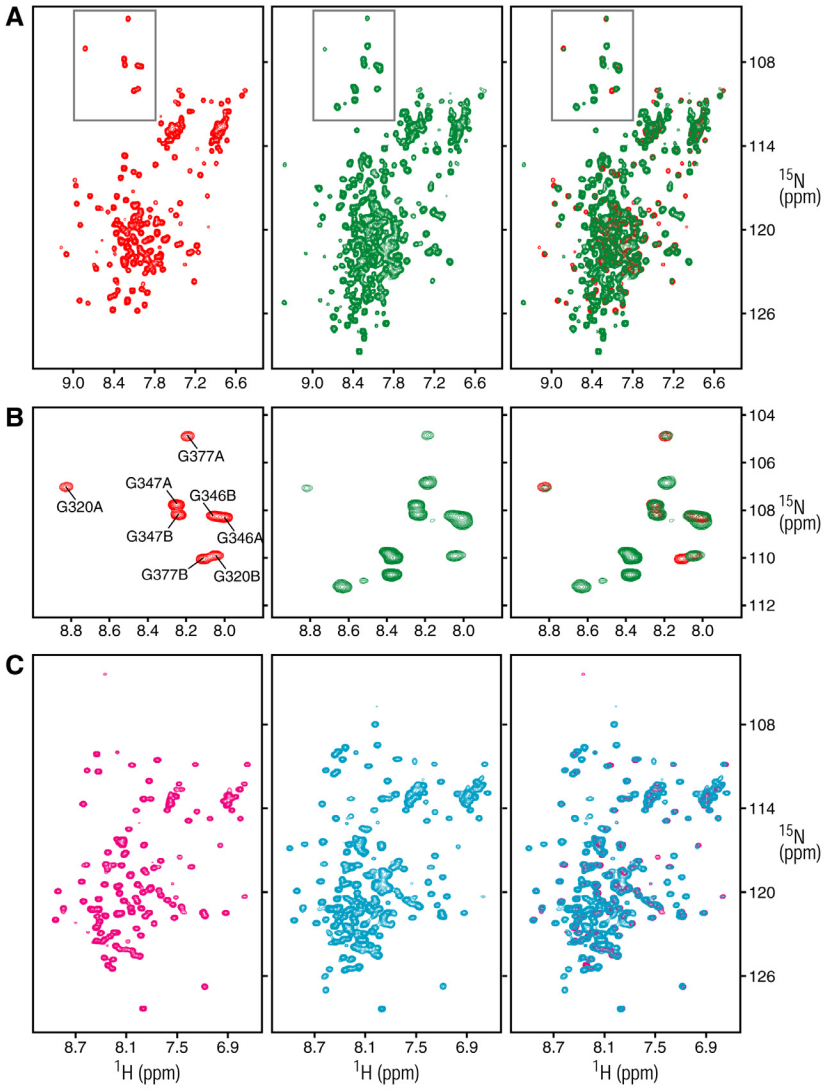
ubiquitous Φ-X-X-Φ-Φ sequence motifs found in transcription factors (Plevin et al., 2005). Potential candidates for mSin3A PAH2 interactions include the mSin3B PAH2 domain with which the domain shares 65% sequence identity and 81% similarity, the mSin3A PAH1 domain (33% identity, 63% similarity) and mSin3B PAH1 (36% identity, 56% similarity). The PAH3 and PAH4 domains were not considered as potential interactors as they are either constitutively associated with core components of the corepressor complex (e.g. SAP30 in the case of PAH3; (Laherty et al., 1998; Zhang et al., 1998)) or have diverged substantially at the sequence level (e.g. PAH4) and are considered unrelated to the other PAH domains.

For the initial studies, we titrated purified mSin3A PAH1 with ^{15}N-labeled mSin3A PAH2. We also expressed and purified mSin3B PAH2 and titrated with ^{15}N-mSin3A PAH2. At sub-millimolar concentrations, neither mSin3A PAH1 nor mSin3B PAH2 produced any changes in the ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra of mSin3A PAH2 (data not shown). Since weak *inter*-molecular association could translate to a stronger *intra*-molecular interaction when the domains are tethered in *cis*, we generated constructs spanning the PAH1 and PAH2 domains (designated PAH1/PAH2) of mSin3A. Expressing and purifying sufficient quantities of mSin3A PAH1/PAH2 was a challenge as the two domains are separated by a long linker region (ca. 110 residues) that rendered the protein readily susceptible to degradation by bacterial proteases. By optimizing the expression tags, inducing protein expression at low temperatures, and taking precautions to limit protease action during the purification stages we were able to generate sufficient amounts of pure protein for NMR studies. Before dissolving the dry, lyophilized powder in NMR buffer, the integrity of the sample was confirmed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). NMR spectra were recorded immediately after sample preparation. The ^{1}H-^{15}N HSQC spectrum of this construct recorded at

0.35 mM concentration is characterized by a broad range of peak intensities and relatively modest amide proton chemical shift dispersion. The vast majority of (mostly intense) resonances fall between 7.5 and 8.5 ppm typical of natively unfolded regions mostly likely from the segment linking the two domains (Figure 1A). The spectrum also shares similarities with the spectrum for the apo-PAH2 domain recorded at the same concentration under identical conditions (Zhang et al., 2006). Indeed, the conformational heterogeneity of apo-PAH2, which is characterized by the appearance of more than one set of correlations, is also evident in the apo-mSin3A PAH1/PAH2 spectrum (Figure 1A). This is readily apparent in the upfield region of the ^{15}N spectrum populated by glycine resonances (Figure 1B). Seven out of eight correlations belonging to the four glycine residues in the PAH2 domain exhibit identical ^{1}H and ^{15}N shifts in the two constructs. The chemical shifts of the correlations belonging to Gly377 in the B-conformation differ slightly, although its origin is unknown and it might arise from subtle differences in solution conditions and/or sample concentrations. An overwhelming majority of the remaining correlations in the apo-mSin3A PAH2 spectrum overlaps with those in the apo-mSin3A PAH1/PAH2 spectrum (Figure 1A, *right panel*). Some of the weaker correlations in the apo-mSin3A PAH2 spectrum (mainly those belonging to the B-conformation, whose population increases with concentration; (Zhang et al., 2006)) exhibit diminished intensity in the apo-mSin3A PAH1/PAH2 spectrum, but this could be attributed in part to the substantial increase in overall molecular weight of the dimeric forms (~60 kDa for apo-mSin3A PAH1/PAH2 versus ~20 kDa for apo-mSin3A PAH2). Furthermore, the PAH2 correlations in the PAH1/PAH2 spectrum exhibit similar concentration-dependent changes in chemical shifts noted previously in the apo-mSin3A PAH2 spectrum ((Zhang et al., 2006); data not shown). As in the case of the apo-PAH2, these changes likely reflect homo-dimerization of apo-PAH1/PAH2 mediated by

PAH2 with the dimer exhibiting fast dissociation kinetics on the NMR timescale. Since our model for an intramolecular interaction between the PAH1 and PAH2 domains involves the same PAH2 surface that is used for homo-dimerization and interaction with other targets, a key prediction of this model would be the loss of conformational heterogeneity and homo-dimerization ability of the mSin3A PAH2 domain in the PAH1/PAH2 construct. The conservation of these mSin3A PAH2 properties in the apo-mSin3A PAH1/PAH2 construct thus strongly argues against this model.

Figure 2.1 **NMR spectra of the N-terminal PAH domains of mSin3A and mSin3B reveal the absence of intramolecular interactions between the domains.** (A) ^{1}H-^{15}N correlated spectra of the mSin3A PAH2 (red), mSin3A PAH1/PAH2 (green), and an overlay of the two spectra (right) recorded at 25 °C in 20 mM sodium phosphate buffer (pH 6.0). Protein concentrations were 0.35 mM and identical NMR data acquisition, processing, display, and contouring threshold parameters were used. (B) Expanded plots corresponding to the glycine region (gray boxes) of the spectra shown in panel A. The assignments for the glycine residues in the two conformers (designated A and B) are shown (Zhang et al., 2006). (C) ^{1}H-^{15}N correlated spectra of mSin3B PAH2 (magenta), mSin3B PAH1/PAH2 (cyan), and an overlay of the two spectra (right) recorded at 25 °C in 20 mM sodium phosphate buffer (pH 6.0). Protein concentrations were 0.24 mM and identical NMR data acquisition, processing, display, and contouring threshold parameters were used.


A
B
C
G377A
G347A
G346B
G320A
G347B
G346A
G377B
G320B
^{15}N (ppm)
^{1}H (ppm)

In parallel with the studies described above, we also sought to explore possible interaction between the mSin3B PAH1 and PAH2 domains. A construct spanning the PAH1 and PAH2 domains of mSin3B was generated. The PAH domains of mSin3B, unlike its mSin3A counterpart, are separated by a much shorter linker region (ca. 50 residues). The integrity of the purified sample prior to NMR studies was again confirmed by SDS-PAGE. The quality of the ^{1}H-^{15}N HSQC spectrum of mSin3B PAH1/PAH2 is superior to that of a comparable construct of mSin3A (Figure 1C). Intense correlations are seen around 8 ppm in the proton spectrum, presumably from the natively unfolded linker segment. In contrast, the mSin3B ^{1}H-^{15}N HSQC spectrum is characterized by better dispersion with the hallmarks of a folded domain (van Ingen et al., 2006). Most of the correlations in the mSin3B PAH2 spectrum overlap well with those in the mSin3B PAH1/PAH2 spectrum (Figure 1C). Additionally, no severe line broadening effects are witnessed for PAH2 resonances in the PAH1/PAH2 spectrum, consistent with the absence of stable, intramolecular interactions involving the PAH domains. We note that two isoforms of mSin3B comprising the PAH1 and PAH2 domains have been described (Koipally et al., 1999). Our studies suggest that these isoforms are likely to antagonize the functions of the full-length forms by effectively competing for recruitment by cognate targets. Finally, titrations of unlabeled mSin3B PAH1/PAH2 with ^{15}N-labeled mSin3A PAH2 at sub-millimolar concentrations produced no changes in the NMR spectrum (data not shown), ruling out heterotypic interactions between these segments.

In summary, we have tested whether the PAH2 domains of mSin3 proteins serve as interaction sites for related PAH domains. Our studies confirm the absence of such inter-domain interactions at the intermolecular and intramolecular levels and suggest that the PAH1 and PAH2 domains of mSin3A and mSin3B are structurally independent. The N-terminal segments

spanning the PAH1 and PAH2 domains within the respective proteins are thus likely to be broadly available for heterotypic protein-protein interactions. The weak propensity of the apo-mSin3A PAH2 domain to homo-dimerize and exist in two different conformations is preserved in the mSin3A PAH1/PAH2 construct. The properties of the apo-mSin3A PAH2 domain thus distinguish this domain from the other PAH1 and PAH2 domains.

CHAPTER THREE

Structural Basis for the mSin3A PAH3 Domain Recognition by the SAP30 SID domain

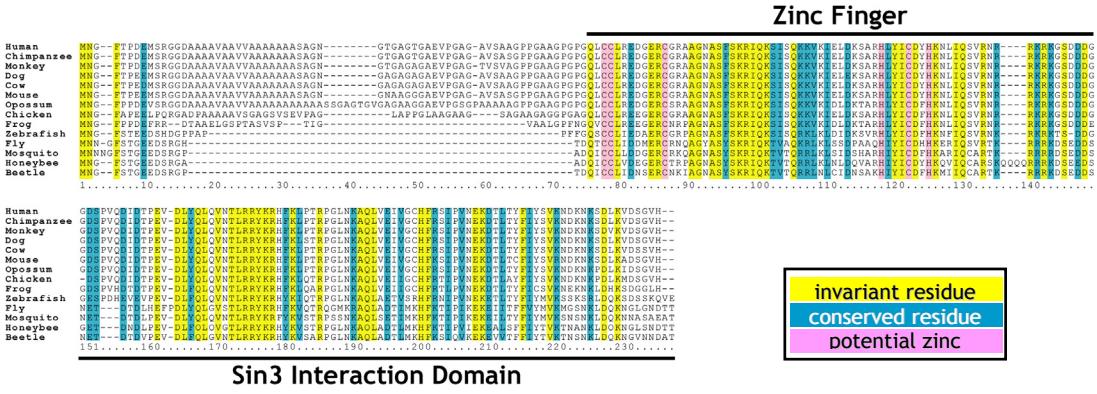
3.1 Introduction

As one of the better-defined interactors of the PAH3 domain and a presumed protein-protein interaction hub within the Sin3 corepressor complex (Figure 1.3, 1.5), SAP30 appears to play an important role in complex assembly (Laherty et al., 1998; Zhang et al., 1998). Not only is it consistently detected in the corepressor complex ranging from yeast to human, its deletion leads to a similar phenotype as those of *sin3* and *rpd3* in yeast (Zhang et al., 1998; Sun and Hampsey, 1999; Bernstein et al., 2000). However, little is know about how this rather small protein (220 aa) is able to simultaneously engage so many targets besides Sin3 including RBP1 (Lai et al., 2001), SAP180 (Binda et al., 2006), the ING family proteins (Kuzmichev et al., 2002; Shi et al., 2006) and other transcription factors such as the NCoR, CIR, and yin yang 1 (YY1), (Laherty et al., 1998; Hsieh et al., 1999; Huang et al., 2003).

SAP30 is predicted to harbor a large amount of intrinsically unstructured regions according to multiple secondary structure prediction tools. The mSin3A PAH3 domain is both necessary and sufficient for interactions with its C-terminal 91-residue segment, which is highly conserved throughout a broad range of species including yeast and human (Figure 3.1, (Laherty et al., 1998)). Since the PAH3 domain is not closely related to the PAH1 and PAH2 domains of mSin3A (Figure 1.4, PAH3 only shares 25% and 16% sequence identity with PAH1 and PAH2), it is not clear whether it adopts the canonical left-handed four-helix bundle structure of the PAH1 and PAH2 domains. Indeed, in mSin3A PAH3, a proline residue (Pro507) should severely shorten the length of the third helix and the polar characteristic of the residues within the first helix is quite different with those found in the other PAH domains, which are involved in targeting the corresponding SIDs (Figure A2.1). Therefore, the interaction mode utilized by SAP30 SID and mSin3A PAH3 is likely to be considerably distinct (Brubaker et al., 2000;

Spronk et al., 2000; Swanson et al., 2004; van Ingen et al., 2004; Nomura et al., 2005; Sahu et al., 2008). As a first step towards clarifying the mechanism of assembly of the Sin3 corepressor complex, I have undertaken structural studies of the mSin3A PAH3-SAP30 SID interaction.

Figure 3.1 **Sequence analysis of the 30 kDa Sin3 associated polypeptide within metazoans.** Multiple sequence alignment based on CLUSTAL W suggests that it contains two discrete and highly conserved domains in addition to a poorly conserved N-terminus. This central region contains a zinc finger motif (see Chapter 4).



3.2 Materials and Methods

Protein expression and purification

The coding sequence of SAP30 ΔN (residues 64-220), SAP30 SID (residues 130-220), and mSin3A PAH3 (residues 456-528) was amplified by PCR and inserted into the pMCSG23 and pMCSG7 expression vector (Stols et al., 2002) respectively. All cloned gene segments were confirmed by DNA sequencing. *E. coli* BL21(DE3) cells (Novagen, Madison, WI) co-transformed by the vectors were grown at 37 °C in LB broth (EMD Chemicals Inc.). The growth temperature was shifted to 20 °C when the $OD_{600\ nm}$ reached approximately 0.6. Expression of both proteins was induced using 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and the cells were harvested 16 h thereafter. Cell pellets were suspended in 50 mM Tris-HCl buffer (pH 8) containing 2 mM Tris (2-carboxy-ethyl) phosphine hydrochloride (TCEP), 1 mM phenyl-methylsulfonyl (PMSF), 1 μM leupeptin, 1 mM pepstatin, and 0.1% Triton X-100, lysed via sonication followed by DNAse treatment for 15 minutes at 4 °C and centrifuged. The supernatant was incubated with the His-Select Nickel resin (Sigma, St. Louis, MO) for 30 minutes. The bound proteins were eluted using 50 mM Tris-HCl buffer (pH 8) containing 2 mM TCEP and 300 mM Imidazole. The eluted proteins were incubated with tobacco etch virus (TEV) protease for 4 h at 22 °C followed by overnight incubation at 4 °C. The mixture's pH was adjusted from 8 to 2 using HCl and from 2 to 6 using NaOH followed by centrifugation, to selectively precipitate MBP. The desired protein in the supernatant was purified to homogeneity via reversed-phase HPLC using a C18 column (Vydac, Hesperia, CA) and a linear gradient of 0.1% trifluoroacetic acid (TFA) and 0.1% TFA in 80% acetonitrile and lyophilized. Samples uniformly labeled with ^{15}N and/or ^{13}C isotopes were produced using the same procedure except that cells were grown in M9 minimal medium containing ^{15}N-ammonium sulfate and/or ^{13}C-D-

glucose (Spectra Stable Isotopes, Columbia, MD) respectively. The identities of the proteins were confirmed by electrospray ionization–mass spectrometry (ESI-MS).

The coding sequence of ING2 SAID domain (residues 19-132) was amplified by PCR and inserted into the pMCSG7 expression vector (Stols et al., 2002). *E. coli* BL21(DE3) cells (Novagen, Madison, WI) containing the vector were grown at 37 °C in LB broth (EMD Chemicals Inc.). The growth temperature was shifted to 20 °C when the $OD_{600\ nm}$ reached approximately 0.7. Expression of the His_6-tagged protein was induced using 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and the cells were harvested 16 h thereafter. Cell pellets were suspended in 20 mM sodium phosphate buffer (pH 8.0) containing 200 mM sodium chloride, 2 mM TCEP, 1 mM phenyl-methylsulfonyl (PMSF), 1 μM leupeptin, 1 mM pepstatin, and 0.1% Triton X-100, lysed via sonication and centrifuged. The pellet was then suspended in the above buffer containing 4 M urea followed by sonication and centrifuged again. The supernatant in denatured buffer was incubated with the His-Select Nickel resin (Sigma, St. Louis, MO) for 30 minutes. Bound proteins were eluted using the equilibration buffer containing 300 mM imidazole followed by incubating with tobacco etch virus (TEV) protease for 4 h at 22 °C and overnight incubation at 4 °C. The mixture was centrifuged and the desired protein in the supernatant was purified to homogeneity via reversed-phase HPLC using a C18 column (Vydac, Hesperia, CA) and a linear gradient of 0.1% trifluoroacetic acid (TFA) and 0.1% TFA in 80% acetonitrile and lyophilized.

The coding sequence of RBP1 R2 domain (residues 1167-1230) was amplified by PCR and inserted into the pMCSG7 expression vector (Stols et al., 2002). *E. coli* Codon Plus cells (Novagen, Madison, WI) containing the vector were grown at 37 °C in LB broth (EMD Chemicals Inc.). The growth temperature was shifted to 20 °C when the $OD_{600\ nm}$ reached

approximately 0.7. Expression of the His_6-tagged protein was induced using 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and the cells were harvested 16 h thereafter. Cell pellets were suspended in 50 mM Tris-HCl buffer (pH 8.0) containing 200 mM sodium chloride, 2 mM TCEP, 1 mM phenyl-methylsulfonyl (PMSF), 1 μM leupeptin, 1 mM pepstatin, and 0.1% Triton X-100, lysed via sonication and centrifuged. The supernatant was incubated with the His-Select Nickel resin (Sigma, St. Louis, MO) for 30 minutes. Bound proteins were eluted using the equilibration buffer containing 300 mM imidazole followed by incubating with tobacco etch virus (TEV) protease for 4 h at 22 °C and overnight incubation at 4 °C. The mixture was centrifuged and the desired protein in the supernatant was purified to homogeneity via reversed-phase HPLC using a C18 column (Vydac, Hesperia, CA) and a linear gradient of 0.1% trifluoroacetic acid (TFA) and 0.1% TFA in 80% acetonitrile and lyophilized.

NMR samples and spectroscopy

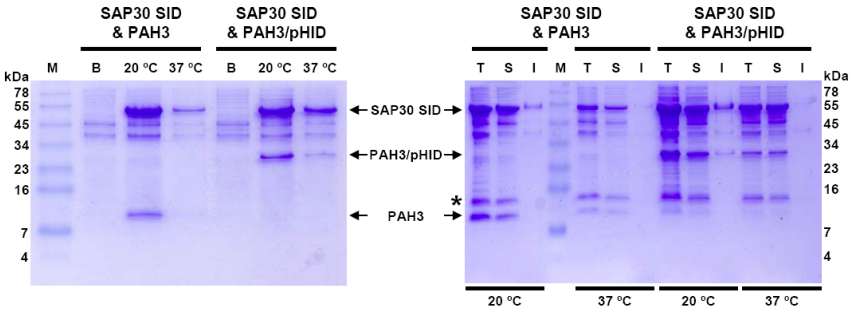
All NMR samples were prepared in 20 mM sodium phosphate buffer (pH 6.5), 2 mM DTT-d_{10}, and 0.2% (w/v) NaN_3. Complexes of ^{13}C,^{15}N-labeled protein with unlabeled one were generated by dissolving the ^{13}C,^{15}N-labeled protein and a equivalent amount of the unlabeled protein in 20 mM sodium phosphate buffer (pH 6), 2 mM DTT, 8 M urea, followed by dialysis against 20 mM sodium phosphate buffer (pH 6.5), 2 mM DTT using Spectra/por DispoDialyzer with 1,000 Da cutoff (Spectrum Laboratories Inc., Rancho Dominguez, CA) and buffer exchange into 20 mM sodium phosphate buffer (pH 6.5), 2 mM DTT-d_{10}, and 0.2% (w/v) NaN_3 using Centricon with 3,000 Da cutoff (YM-3, Millipore Corporation). Protein concentrations were measured spectrophotometrically (Gill and von Hippel, 1989). NMR data were acquired on a Varian Inova 600 MHz spectrometer at 35 ºC. NMR data processing and analysis were

performed using an in-house modified version of Felix 98.0 (Accelrys) and NMRView (Radhakrishnan et al., 1999; Johnson, 2004). Backbone ^{1}H, ^{15}N, and ^{13}C resonances for the both mSin3A PAH3 and SAP30 SID were assigned by analyzing three-dimensional (3D) HNCACB, CBCA(CO)NH, C(CO)NH-TOCSY, HNCO, (Grzesiek and Bax, 1993; Ferentz and Wagner, 2000).

3.3 Results and Discussion

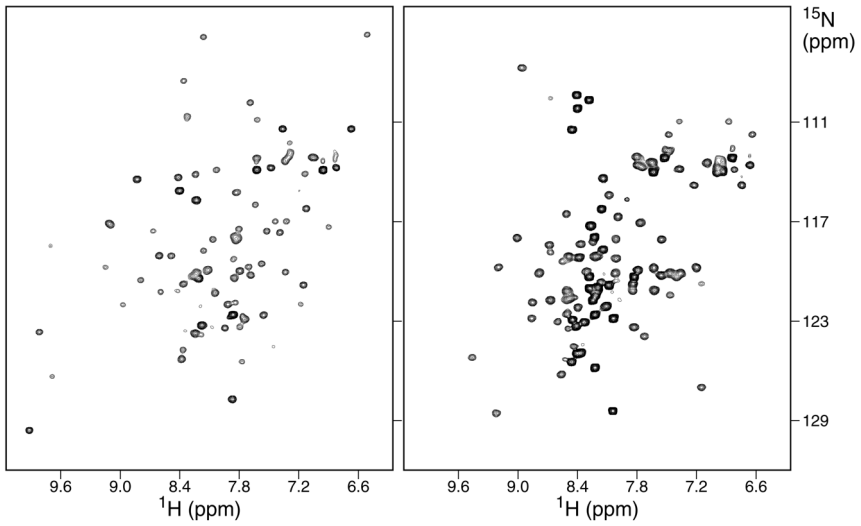
Multiple attempts were made to express various mSin3A constructs spanning the PAH3 domain in isolation, but all of them failed to express in *E. coli* possibly because of their susceptibility towards intracellular proteases. Surprisingly, co-expression with MBP-tagged SAP30 SID led to high yield production of both proteins at comparable amounts, indicating that the direct physical interaction between SAP30 SID and PAH3 improved the stability of PAH3 domain during expression (Figure 3.2). In fact, the interaction between the two polypeptides was of such high affinity that the complex formed during co-expression could be disrupted only using denaturants or organic solvents. Similar results have been obtained in the course of studies of other transcription factor complexes (Demarest et al., 2002).

Figure 3.2 **SDS-PAGE analyses of the co-expression (*left*) and solubility tests (*right*) of two His_6-tagged mSin3A PAH3 constructs with MBP-tagged SAP30SID (conducted by Rebecca Imhoff).** Protein levels from expression at different temperatures were checked by SDS-PAGE followed by Coomassie Blue staining. Arrowheads depict protein bands of interest. "*" indicates lysozyme used in cell lysis. Abbreviations: B: before induction; T: total cell lysate; S: soluble supernatant; I: inclusion-body pellet; M: molecular weight marker.



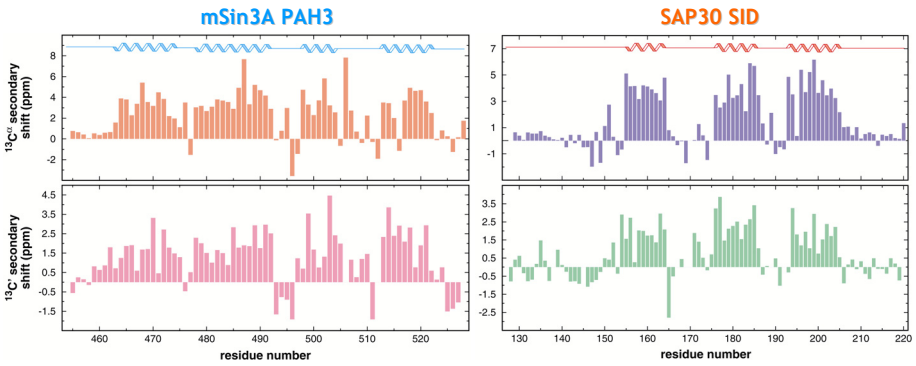
Attempts to study the apo-forms of the mSin3A PAH3 and SAP30 SID were precluded because of solubility issues close to physiological pH and also due to the poor quality of the NMR spectra, which suggested an aggregated protein. I then generated two 1:1 complex samples with either mSin3A PAH3 or SAP30 SID ^{15}N-labeled. The quality of ^{1}H-^{15}N HSQC spectra for both complexes increased dramatically with reasonable amide proton dispersion and a single-set of resonances, indicating that both proteins were properly folded and conformationally pure in the presence of their interacting partners (Figure 3.3).

Figure 3.3 **^{1}H-^{15}N HSQC spectra of mSin3A PAH3 in the presence of an equimolar amount of unlabeled SAP30 SID (*left*) and that of SAP30 SID in the presence of one equivalent amount of unlabeled mSin3A PAH3 at 35 °C.** The NMR buffer used here contained 20 mM sodium phosphate (pH 6.5), 5 mM DTT, and 0.2% sodium azide.



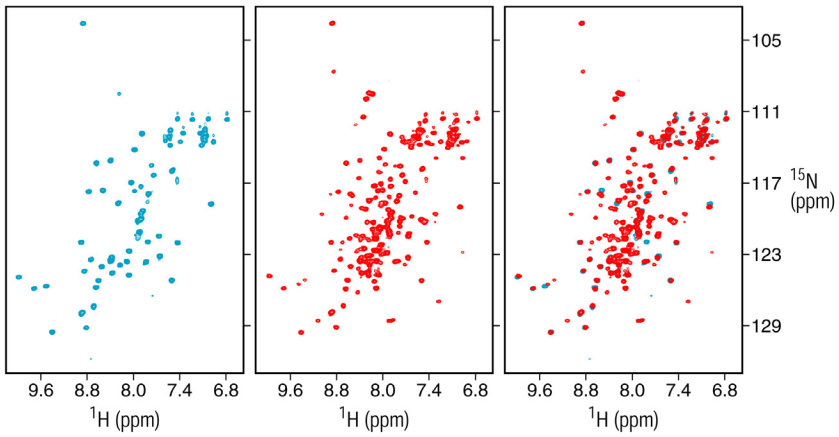
Two 1:1 complex samples with either mSin3A PAH3 or SAP30 SID $^{13}C,^{15}N$-labeled were generated and the backbone resonances of both proteins within the corresponding complexes were assigned using triple-resonance experiments. The $^{13}C^{\alpha}$ and $^{13}C'$ secondary chemical shift analyses of backbone nuclei in the respective proteins indicated that the PAH3 and SAP30 SID domains are largely helical (Figure 3.4). Like other PAH domains the PAH3 domain forms a four-helix bundle, but with helix $\alpha 3$ much shorter than in the PAH1 and PAH2 domains (Brubaker et al., 2000; Spronk et al., 2000; Swanson et al., 2004; van Ingen et al., 2004; Nomura et al., 2005; Sahu et al., 2008). Intriguingly, the SAP30 SID adopts a structure comprising three helices, which is a departure from the structural motif used by proteins targeting the mSin3A PAH1 and PAH2 domains. In addition, comparable large number of inter-molecular NOEs were detected in both 3D $^{15}N/^{13}C$-filtered, $^{15}N/^{13}C$-edited NOESY spectra conducted using the two complexes, which further confirmed that a different interaction mode might be present in this complex and it should be feasible to obtain the complex structure using the similar approach as those used in the structure determination of other SID-PAH complexes (data not shown). It is at present unclear how the three SAP30 SID helices are arranged and how they engage the PAH domain.

Figure 3.4 **The secondary chemical shifts of backbone nuclei in both mSin3A PAH3 (*left*) and SAP30 SID (*right*) according to assignments of C^{α} and carbonyl carbon indicates that the PAH3 and SID domains are largely helical.**



SAP30 sequence analysis revealed relatively high levels of conservation in the linker region between the zinc-finger and SID domains of SAP30 (Figure 3.1). In order to investigate the possibility of an interaction between the SAP30 ZnF and SAP30 SID in the context of the full-length protein, I made two other constructs – SAP30 ZnF (residues 64-131) harboring only the zinc-finger motif and SAP30 ΔN (residues 64-220) spanning both conserved domains. A 1:1 complex of ^{15}N-labeled SAP30 ΔN with mSin3A PAH3 was made using the same procedures as the complex generation of SAP30 SID with PAH3. The ^{1}H-^{15}N HSQC spectra of both SAP30 ZnF and the SAP30 ΔN-PAH3 complex were characterized by a single-set of resonances and reasonable amide proton chemical shift dispersion indicative of conformationally pure and properly folded proteins in the sample solution (Figure 3.5). Most of the correlations in the SAP30 ZnF spectrum overlap well with those in the SAP30 ΔN-PAH3 complex spectrum. In addition, no changes in resonance positions or line-broadening effects were noticed for SAP30 resonances within the SAP30 ΔN-PAH3 complex spectrum, consistent with the absence of any stable, intramolecular interactions involving the zinc-finger and the SID domains.

Figure 3.5 **^{1}H-^{15}N HSQC spectra of SAP30 ZnF (residues 64-131) (*left*), SAP30 ΔN (residues 64-220) complexed with mSin3A PAH3 (*middle*), and overlay of the above two (*right*) at 25 °C.** Protein concentrations were 0.2 mM in both cases. The NMR buffer contained 20 mM Tris-acetate (pH 6.5), 5 mM DTT, and 0.2% sodium azide. Identical NMR data acquisition, processing, display, and contouring threshold parameters were used.



Extended studies have been made in the characterization of interactions between SAP30 and two of its better-characterized binding partners – ING2 SAID domain and RBP1 R2 domain. Since the interaction domain of SAP30 involved in targeting those two proteins have not been definitively established, NMR titration experiments were initially carried using ^{15}N-labeled SAP30 ZnF and ^{15}N-labeled mSin3A PAH3-SAP30 SID (both proteins in the complex were ^{15}N-labeled) complex separately. Titration of ^{15}N-labeled SAP30 ZnF with stoichiometric amounts of unlabeled RBP1 R2 failed to produce any perturbation in the ^{1}H-^{15}N HSQC spectrum (data not shown). Similarly, titrations of the ^{15}N-labeled mSin3A PAH3-SAP30 SID complex with either unlabeled ING2 SAID or unlabeled RBP1 R2 also failed to produce any perturbation in the corresponding ^{1}H-^{15}N HSQC spectrum, implying that neither ING2 SAID nor RBP1 R2 interacted with the SAP30 SID-PAH3 complex (data not shown). However, the titration of ^{15}N-labeled SAP30 ZnF with ING2 SAID led to precipitation of both proteins, leading to an inconclusive result. A separate titration experiment conducted using ^{15}N-labeled SAP30 ΔN-mSin3A PAH3 complex (where only SAP30 ΔN was ^{15}N-labeled) and unlabeled ING2 SAID led to the precipitation of ING2 SAID, again leading to an inconclusive result (data not shown).

In summary, a co-expression strategy has to be taken in order to prepare both proteins from the mSin3A PAH3-SAP30 SID complex for structural analysis. However, they readily formed aggregates and a co-refolding step was necessary for reconstitute the complex. The interaction between the two components is likely to be at the sub-micromolar range according to the ability to co-purify the complex from co-expression and a surprisingly large number of inter-molecular NOEs between the two proteins. Preliminary structural analysis suggested that PAH3 domain forms a four-helix bundle, but with helix α3 much shorter than in the PAH1 and PAH2 domains. More interestingly, the SAP30 SID adopts a structure comprising three helices, which

is a departure from the structural motif used by proteins targeting the mSin3A PAH1 and PAH2 domains. The complex structure of mSin3A PAH3-SAP30 SID will not only provide the first platform for clarifying the corepressor complex assembly, but also permit potential discovery of the interaction surface for their further recruitment. Although SAP30 was proposed to serve as the interaction interface for the RBP1 R2 and the ING2 SAID domains, these interactions could not be recapitulated *in vitro* with bacterially-expressed, purified proteins/domains. These results suggest that the RBP1 and ING2 interactions with the Sin3 corepressor complex may be mediated by other protein(s) or rely on presently uncharacterized post-translational modifications.

CHAPTER FOUR

Structure-Function Analysis of a Novel Zinc Finger Motif in SAP30

4.1 Introduction

Sequence-specific DNA-binding transcription factors function as activators or repressors of transcription and frequently recruit transcriptional coregulators that lack specific DNA-binding activity. In the meanwhile, these coregulator complexes can also be recruited onto specific loci on the genome though various chromatin-binding proteins (Figure 1.1). As the first identified global transcriptional corepressor complex, Sin3 has already been shown to interact with a diverse group of chromatin-targeting protein factors to equalize its lack of intrinsic DNA-binding activity (Figure 1.3, 1.5). However, this well-known conception might need some review since several chromatin regulatory complexes have been implicated to harbor DNA-binding domain within their core components (Da et al., 2006; Tu et al., 2008).

As a protein-protein interaction hub within the Sin3 corepressor complex (Figure 1.3, 1.5), SAP30 seems to play a particularly important role in stabilizing the core complex. In the course of its structural characterization, a novel zinc-finger motif (designated as SAP30 ZnF) was identified to be located within this 220-residue protein. The initial functional studies of this newly discovered zinc-finger motif were focused on its potential role in recognizing histone codes. However, that possibility has been potentially ruled out after testing most of the known histone modifications. Recent studies indicated that it might be a nucleic acid binding domain, which could be another example for the existence of intrinsic DNA-binding activity within coregulator complex.

4.2 Materials and Methods

Protein expression and purification

The coding sequence of SAP30 ZnF corresponding residues 64−131 was amplified by PCR and inserted into the pMCSG7 expression vector (Stols et al., 2002). All cloned gene segments were confirmed by DNA sequencing. *E. coli* BL21(DE3) cells (Novagen, Madison, WI) containing the vector were grown at 37 °C in LB broth (EMD Chemicals Inc.). The growth temperature was shifted to 20 °C when the OD_{600nm} reached approximately 0.6. Expression of the His_6-tagged protein was induced using 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and the cells were harvested 16 h thereafter. Cell pellets were suspended in 50 mM Tris-HCl buffer (pH 8) containing 0.2 M sodium chloride, 2 mM Tris (2-carboxy-ethyl) phosphine hydrochloride (TCEP), 1 mM phenyl-methylsulfonyl (PMSF), 1 μM leupeptin, 1 mM pepstatin, and 0.1% Triton X-100, lysed via sonication followed by DNAse treatment for 15 minutes at 4 °C and centrifuged. The supernatant was incubated with the His-Select Nickel resin (Sigma, St. Louis, MO) for 30 minutes. The resin was washed by 50 mM Tris-HCl buffer (pH 8) containing 0.5 M sodium chloride and bound proteins were eluted using 50 mM Tris-HCl buffer (pH 8) containing 0.2 M sodium chloride, 2 mM TCEP, and 300 mM Imidazole. The eluted proteins were incubated with tobacco etch virus (TEV) protease for 4 h at 22 °C followed by overnight incubation at 4 °C. The mixture was centrifuged and the desired protein in the supernatant was purified to homogeneity via reversed-phase HPLC using a C18 column (Vydac, Hesperia, CA) and a linear gradient of 0.1% trifluoroacetic acid (TFA) and 0.1% TFA in 80% acetonitrile and lyophilized. Samples uniformly labeled with ^{15}N and/or ^{13}C isotopes were produced using the same procedure except that cells were grown in M9 minimal medium containing ^{15}N-ammonium sulfate and/or ^{13}C-D-glucose (Spectra Stable Isotopes, Columbia, MD) respectively. The alanine mutants of the SAP30 ZnF were generated by QuikChange site-directed mutagenesis (Stratagene). The 15N-labeled samples for mutants C68A and H108A were produced using the

same procedure as the wild type, but the samples for C67A and C112A were solubilized from inclusion body pellets followed by purification under denatured conditions. The identities of the proteins were confirmed by electrospray ionization–mass spectrometry (ESI-MS).

NMR samples

NMR samples were prepared by dissolving the dry, lyophilized protein powder in 20 mM Tris acetate buffer (pH 6), 2 mM DTT-d_{10}, and 0.2% (w/v) NaN_3. Protein concentrations were determined spectrophotometrically (Gill and von Hippel, 1989). The concentrations of the ^{15}N,^{13}C-labeled sample and the ^{15}N-labeled sample for DNA titrations were 0.56 mM and 0.2 mM, respectively. An equimolar complex of ^{15}N-labeled SAP30 ZnF and DNA duplex was generated by directly combining the two.

NMR spectroscopy and structure determination

NMR data were acquired on a Varian Inova 600 MHz spectrometer at 25 ºC. NMR data processing and analysis were performed using an in-house modified version of Felix 98.0 (Accelrys) and NMRView (Radhakrishnan et al., 1999; Johnson, 2004). Backbone and side chain ^1H, ^{15}N, and ^{13}C resonances for the SAP30 ZnF were assigned by analyzing three-dimensional (3D) HNCACB, C(CO)NH-TOCSY, HNCO, H(CCO)NH-TOCSY, HCCH-TOCSY, and HCCH-COSY spectra (Grzesiek and Bax, 1993; Ferentz and Wagner, 2000). Aromatic proton resonances were assigned from 2D ^1H-^{13}C HSQC spectra and 3D HCCH-COSY and HCCH-TOCSY spectra (Otting and Wuthrich, 1990; Lohr and Ruterjans, 1996).

For structure determination, backbone ϕ and ψ torsion angle restraints were derived from an analysis of H^{α}, C^{α}, C^{β}, C' and backbone ^{15}N chemical shifts using TALOS (Cornilescu et al.,

1999). Restraints were imposed only for those residues that exhibited TALOS reliability scores of 9. NOE-based distance restraints for each component of the binary complex were derived from 3D ^{15}N-edited NOESY (mixing time, τ_m = 75 ms) spectrum recorded in H_2O, 3D aliphatic ^{13}C-edited NOESY (τ_m = 60 ms) and 3D aromatic ^{13}C-edited NOESY (τ_m = 80 ms) spectra recorded in D_2O. All NOEs were calibrated and assigned iteratively and automatically by ARIA (version 1.2) (Linge et al., 2003, 2004) and were checked manually between successive rounds of refinement. Structures were calculated using ARIA in conjunction with CNS (Brunger et al., 1998). Structures were calculated from extended conformations as starting models. A total of 80 structures were computed in the final iteration, 40 of which were refined in the presence of a shell of explicit water solvent, and the 20 structures with the lowest restraint energies, restraint violations, and RMS deviations from ideal covalent geometry were selected for structural analysis. The final structures were analyzed using PROCHECK and CNS (Laskowski et al., 1996; Brunger et al., 1998; Salerno et al., 2004). Molecular images were generated using CHIMERA (Pettersen et al., 2004).

4.3 Results and Discussion

In the course of our previous sequence analysis of SAP30, another highly conserved region was revealed sandwiched between the SID domain at the carboxyl terminus and an N-terminal low-complexity region (Figure 3.1). Interestingly, this domain can only be found in metazoans ranging from fly to human and it appears to be a novel motif because its sequence does not conform to any of the known motifs in sequence databases. An MBP-tagged construct of SAP30 spanning residues 64-131 (corresponding to the SAP30 ZnF) was expressed and purified using affinity chromatography. After TEV protease treatment to cleave off the MBP-tag

in addition to EDTA treatment to eliminate any residual metal ions, the ^{1}H-^{15}N correlated spectrum of this protein was recorded. The spectrum was characterized by poor chemical shift dispersion of amide proton cross-peaks indicative of a natively unfolded conformation adopted by the protein in solution (Figure 4.1A). However, due to the fact that six invariant cysteine and histidine residues exist in this highly conserved region across different species, we tested whether it could harbor a zinc finger motif by adding one equivalent amount of $ZnCl_2$ to the sample. Extensive chemical shift dispersion was noticed upon zinc addition, which indicated that zinc ion was essential for this domain to adopt its 3D structure and thereby confirmed the hypothesis that SAP30 harbored a zinc finger (Figure 4.1B). Moreover, a more rigorous zinc titration was conducted by recording the same type of experiments. With the addition of increasing amounts of zinc, a new set of resonances appeared while the other set corresponding to the poor amide proton dispersion disappeared implying that the zinc coordination of this protein falls into the slow exchange regime on the NMR time-scale and that it bound to zinc with high affinity. The spectrum did not change upon further addition of zinc implying that a single zinc-binding site was involved (Figure 4.2).

Figure 4.1 **^{1}H-^{15}N correlated HSQC spectra of SAP30 ZnF in the absence (A) and presence (B) of an equivalent amount of $ZnCl_2$ at 25 °C.** The backbone amide proton and nitrogen resonances are labeled with residue numbers from 63 to 131 of SAP30. Side-chain resonances are depicted by either pink horizontal lines or red dotted lines.

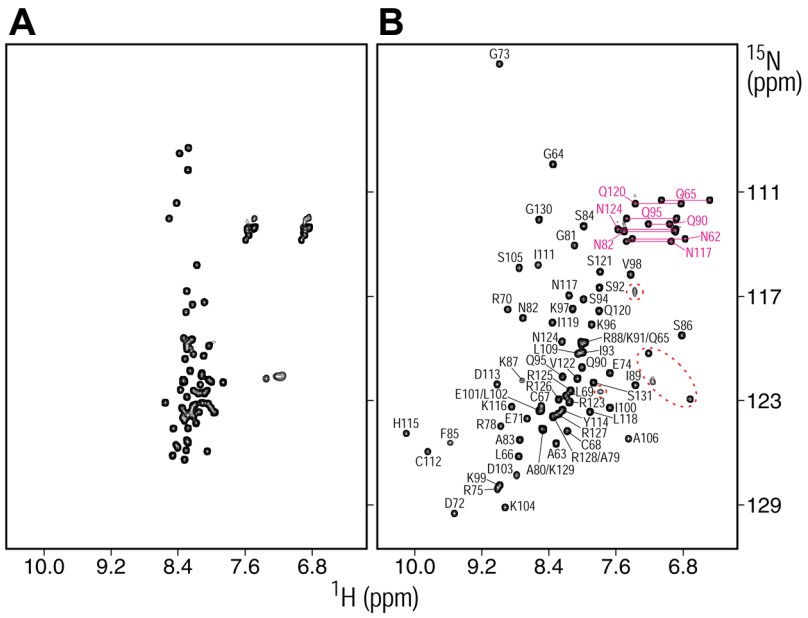
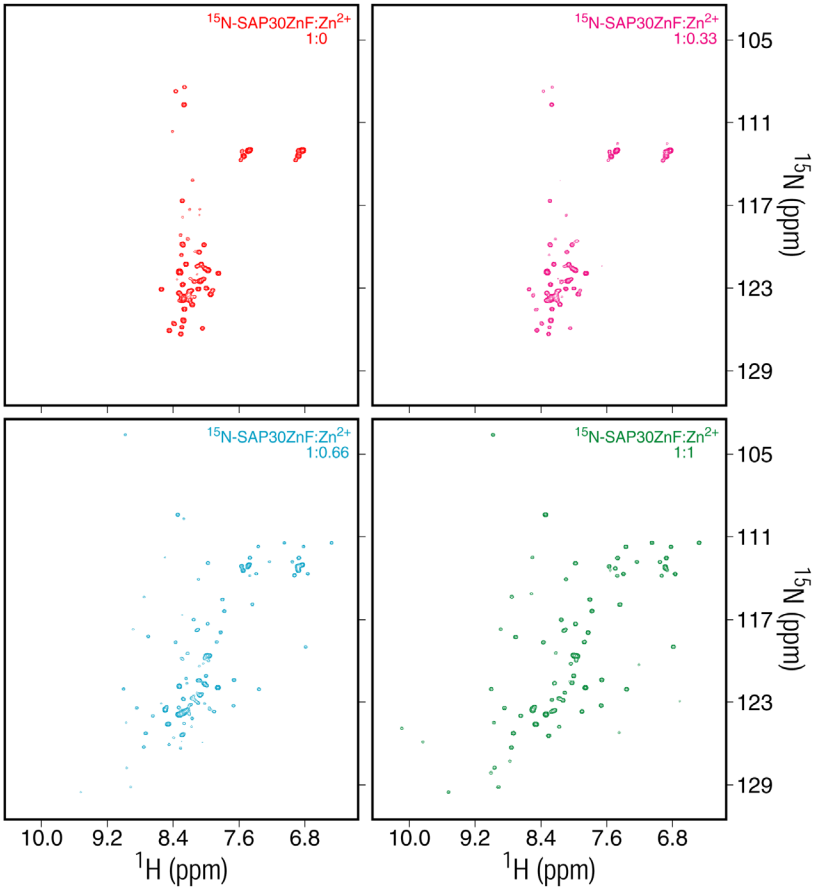
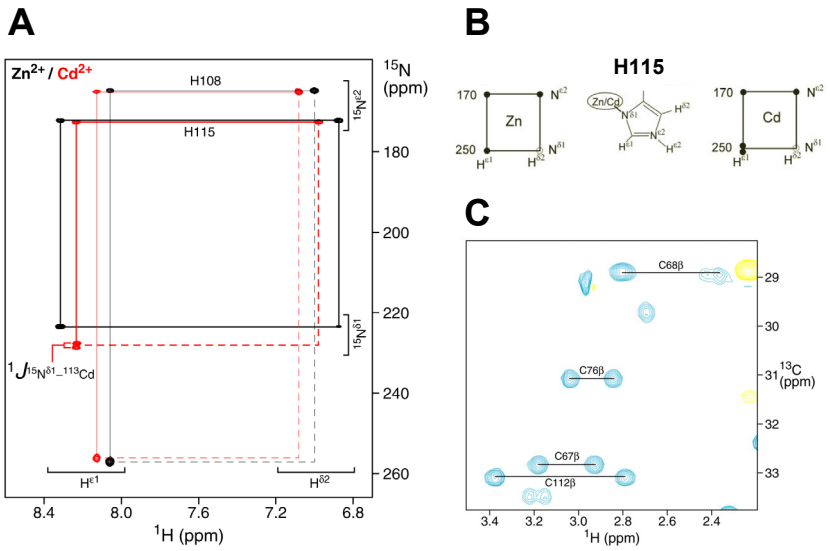


Figure 4.2 **^{1}H-^{15}N correlated HSQC spectra of SAP30 ZnF in the absence and with the addition of increasing amount of $ZnCl_2$ at 25 °C.** The molar ratios of 15N-labeled SAP30 ZnF and ZnCl2 are depicted on the upper right corner of the corresponding spectrum. Identical data acquisition, processing and display parameters were used.



Assuming a tetrahedral coordination for the single zinc ion, there appeared to be an excess of potential zinc-binding residues in the SAP30 ZnF construct (four cysteine and two histidine residues). To determine whether either of the two histidine residues' side-chains were involved in zinc coordination, we prepared ^{15}N-labeled SAP30 ZnF either in the presence of zinc or where the zinc ion was replaced by ^{113}Cd, which has a NMR active nucleus and is often used to explore the zinc coordination within a zinc-finger motif (Legge et al., 2004). The ^{113}Cd-loaded SAP30 ZnF sample shared a similar well-dispersed HSQC spectrum with the zinc-loaded form in Figure 4.1B (data not shown) implying that this zinc-finger motif also adopts a similar three-dimension fold in the presence of ^{113}Cd. Surprisingly however, there was a clear difference in the 2D heteronuclear multiple quantum coherence (HMQC) spectra of the two metal-loaded samples. The cross-peak at 235 ppm (^{15}N) and 8.2 ppm (^{1}H) corresponding to the $N^{\delta 1}$-$H^{\varepsilon 1}$ correlation of H115 is split by the ^{15}N-^{113}Cd coupling suggesting that the $N^{\delta 1}$ atom is directly involved in coordinating the metal ion, whereas the $N^{\delta 1}$/ $N^{\varepsilon 2}$ resonances of H112 are unaffected by any ^{15}N-^{113}Cd scalar coupling (Figure 4.3A). These studies not only revealed one out of two invariant histidine residues involving in the zinc coordination, but also unambiguously assigned the nitrogen atom that is responsible for the covalent bonding with zinc (Figure 4.3B).

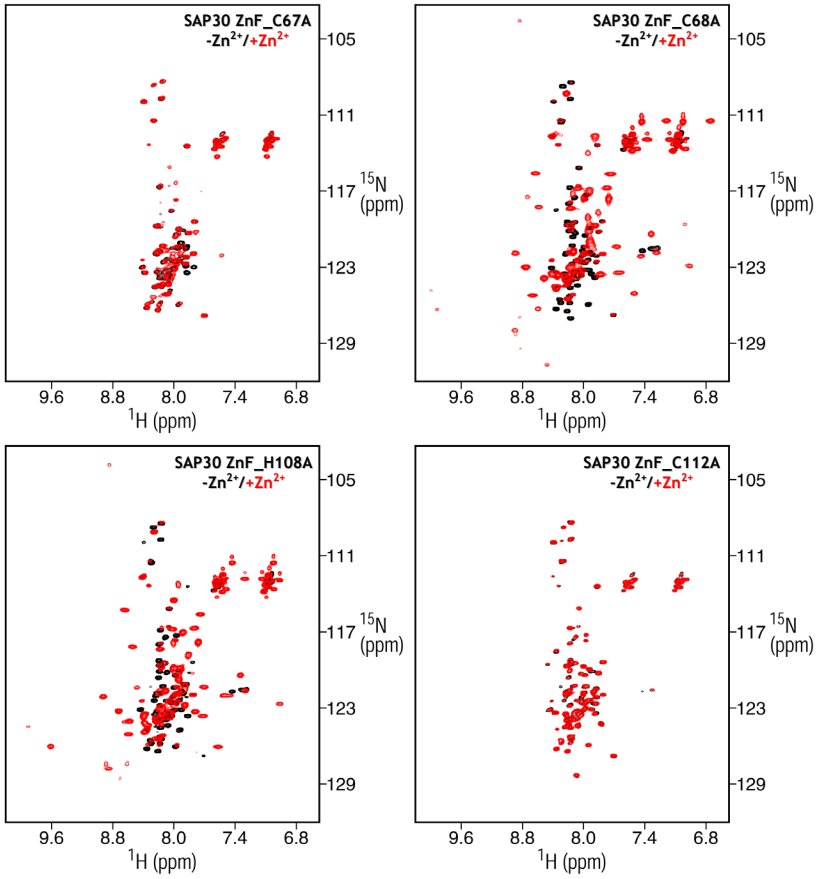
Figure 4.3 **Determination of zinc coordination pattern using NMR.** (A) ^{1}H-^{15}N correlated HMQC spectrum of SAP30 ZnF loaded with zinc (black) or ^{113}Cd (red). (B) Cartoon showing H115 side-chain ^{1}H-^{15}N through-bond correlation for explaining cross-peak splitting caused by ^{15}N-^{113}Cd coupling in (A). (C) ^{1}H-^{13}C HSQC of SAP30 ZnF indicating C^{β} correlations of the four cysteines.



In order to characterize which three out of the four cysteine residues along with H115 coordinate zinc ion, ^{1}H-^{13}C HSQC spectrum of SAP30 ZnF was recorded because the C^{β} shift of cysteine is sensitive to the ionization and redox states. Three out of four cysteines' C^{β} ($C67^{\beta}$, $C76^{\beta}$, and $C112^{\beta}$) shifted downfield as a result of zinc coordination (Figure 4.3C). Furthermore, the thiol proton resonance of C68 can be observed implying that C68 was not involved in zinc coordination. Collectively, these results indicated that C67, C76, C112 and H115 are the zinc coordinating residues.

To further confirm the NMR studies described above about determining the zinc coordinating residues, each cysteine and histidine residues were mutated into an alanine. HSQC spectrum overlays clearly showed that well-dispersed amide proton resonances were detected in zinc titrations of C68A and H108A mutants, which are similar to the wild-type and indicative of non-zinc binding residues (Figure 4.1). However, the dispersion level of two mutants (C67A and C112A) did not change much upon zinc addition, implying that they were not able to adopt three-dimensional structure even in the presence of zinc and thereby C67 and C112 are zinc-coordinating residues (Figure 4.4). Interestingly, both C68A and H108A mutant constructs were expressed in soluble form, similar as the wild-type, whereas C67A and C112A were exclusively expressed within inclusion bodies, suggesting that the ability of this zinc finger to coordinate zinc is important for staying in solution during overexpression. The C76A and H115A mutants could not be analyzed because they expressed poorly.

Figure 4.4 **SAP30 ZnF alanine mutations probed by 2D HSQC spectra.** Overlays of HSQC clearly show that C67A and C112A mutants indeed disrupt folding of the domain upon zinc addition indicated by the failure to increase amide proton dispersion upon zinc addition, whereas C68A and H108A mutants are largely unaffected.



We decided to determine the structure of this novel zinc-finger motif and hoped the structure itself might give some clues about its uncovered cellular roles. Essentially all the resonances were assigned using standard through-bond approaches and the structure was determined using a combination of tetrahedral zinc coordination geometry, chemical shift-based torsion angle and nuclear Overhauser effect (NOE)-based distance restraints (Table 4.1). Structures of reasonable quality and precision and in agreement with input experimental data were obtained (Table 4.1 and Figure 4.5). The SAP30 ZnF appears to adopt a novel structure since it does not conform to any of the known folds. It comprises two α-helices and two β-strands, which span residues 87-95, 113-120 and 69-71, 100-102, respectively. The tetrahedral geometry for zinc coordination is well-defined; a parallel calculation without incorporation of the zinc coordination restraints resulted in a similar overall structure. $\{^{1}H\}$-^{15}N heteronuclear NOE measurements indicated a relatively rigid backbone except for the regions corresponding to the extreme N- and C-terminus (Figure 4.6). However, lower than average NOEs were measured for the segment immediately C-terminal to the α2 helix.

Table 4.1 NMR Structure Determination Statistics for the SAP30 ZnF

Restraint Statistics	
NOE-based distance restraints	
Unambiguous NOE-based restraints	1046
Intraresidue	543
Sequential ($\| i - j \| = 1$)	174
Medium-range ($1 < \| i - j \| \leq 4$)	123
Intramolecular long-range ($\| i - j \| > 4$)	206
Ambiguous NOE-based restraints	234
Intramolecular NOE-based restraints	1280
Hydrogen bonding distance restraints	18
Zinc coordination restraints (distance, improper)	4, 1
Torsion angle restraints	40 (17 ϕ, 17 ψ, 6 χ^1)
Structure Quality of NMR Ensemble	
Restraint satisfaction	
Root-mean-square differences for distance restraints (Å)	0.012 ± 0.002
Root-mean-square differences for torsion angle restraints (°)	0.355 ± 0.328
Deviations from ideal covalent geometry	
Bond lengths (Å)	0.004 ± 0.000
Bond angles (°)	0.480 ± 0.023
Impropers (°)	1.330 ± 0.123
Ramachandran plot statistics (%)	
Residues in most favored regions	76.0
Residues in allowed regions	20.9
Residues in disallowed regions	3.1
Average Atomic Root-Mean-Square Deviations from the Average Structure (Å)	
All atoms	3.22
All atoms except disordered regions[a]	1.55
All atoms in secondary structural elements[b]	1.74
Backbone atoms (N, C^α, C')	
All residues	3.22
All residues excluding highly disordered regions[a]	0.77
All residues in secondary structural elements[b]	0.68

[a]: disordered regions include residues 61–65 and 123–131
[b]: secondary structural elements include residues 69–71, 87–95, 100–102, and 113–120

Figure 4.5 **The SAP30 ZnF defines a novel fold. 180° rotated views of both the C^{α} trace (*top panels*) in the ensemble and the representative structure (*bottom panels*) of the SAP30 ZnF are colored as rainbow from the N-terminus (blue) to the C-terminus (red).** Zinc ions are indicated as pink spheres.

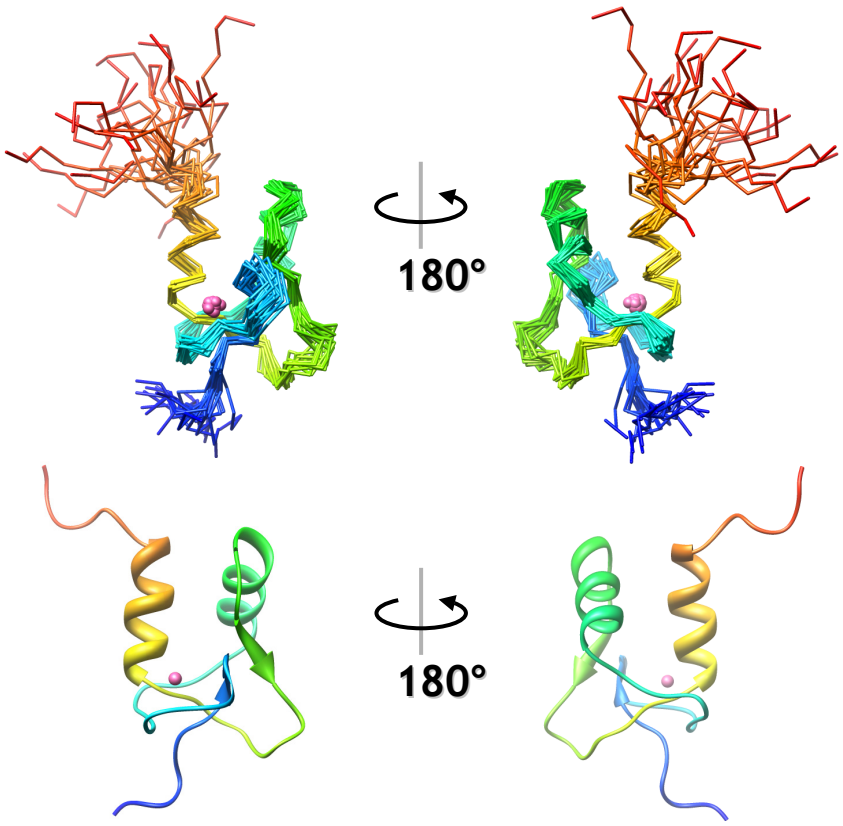
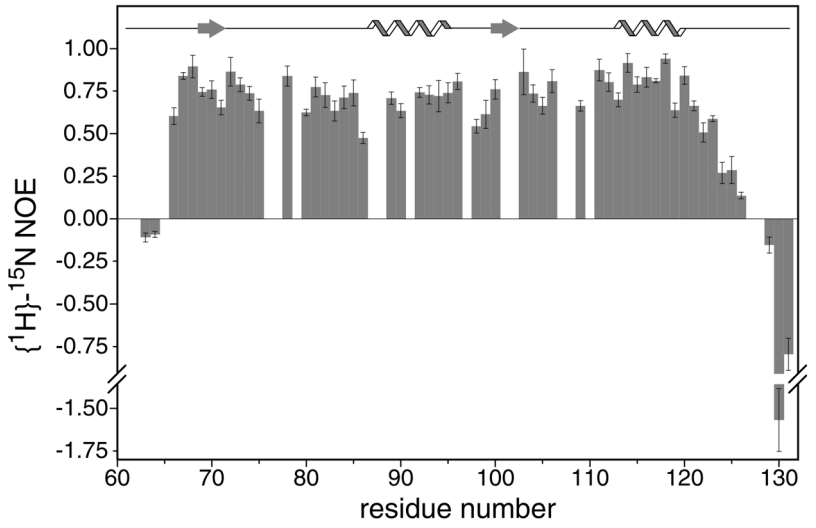


Figure 4.6 **A plot of the {^{1}H}-^{15}N heteronuclear NOE to assess the flexibility of the polypeptide backbone.** The error bars represent the standard deviations from three independent measurements.



For characterizing the functional role of this newly discovered zinc finger motif, we initially tested whether it might contribute to the SAP30 interactions with ING2 SAID domain or RBP1 R2 domain by NMR titration experiments. But none of these studies suggested any direct interaction between the constructs we used (data not shown). We then tested whether the SAP30 ZnF might have similar function as the PHD zinc-finger motif, which has been described as a chromatin-binding module, partially because the conservation analysis revealed a potential ligand-binding site with a relatively small surface area (Figure 4.7). However, no specific interaction has been detected by NMR titrations using various ligands, including acetyllysine, trimethyllysine, dimethylarginine, phosphoserine, phosphothreonine, ubiquitin and a small subset of histone peptides harboring some of the above modifications.

Although we initially excluded a DNA-binding function for the SAP30 ZnF as SAP30 is an integral component of the Sin3 corepressor complex, further investigation revealed a positively charged surface on the structure, which is highly conserved (Figure 4.7). Therefore, we titrated an approximately equimolar amount of a 20nt DNA duplex to a ^{15}N-labeled SAP30 ZnF sample. Surprisingly, the addition of DNA not only induced resonance broadening all across the motif except for its flexible N- and C-termini, presumably caused by the slower tumbling of the ZnF-DNA complex, but also lead to substantial perturbation of a subset of resonances in the HSQC spectrum, indicative of a specific interaction between the protein and DNA (Figure 4.8). Moreover, the most perturbed residues on the ZnF during DNA addition appear to nicely correspond to the ones having the lowest resonance intensity, such as residues 84-87 and 121-123 (Figure 4.8B). More rigorous DNA titrations using DNA duplexes of different lengths will be necessary for confirming this interaction. Chemical shift mapping analysis using the structure of the free ZnF revealed two discrete regions possibly involved in

binding DNA, including residues in the first helix and in the C-terminal region following the second helix (Figure 4.9). Since the mapping analysis did not give rise to a well-defined binding surface spanning the C-terminus of the ZnF, a secondary structure rearrangement remains a formal possibility for this region to recognize the DNA. In order to test this possibility, a secondary chemical shift analysis will be performed by reassigning the C^{α}, C', and H^{α} chemical shifts in the context of the DNA duplex.

Figure 4.7 **Structure of the SAP30 ZnF shown in the cartoon representations highlighting either different types of residues (*top pannels*) or various degrees of conservations (*middle and bottom pannels*).**

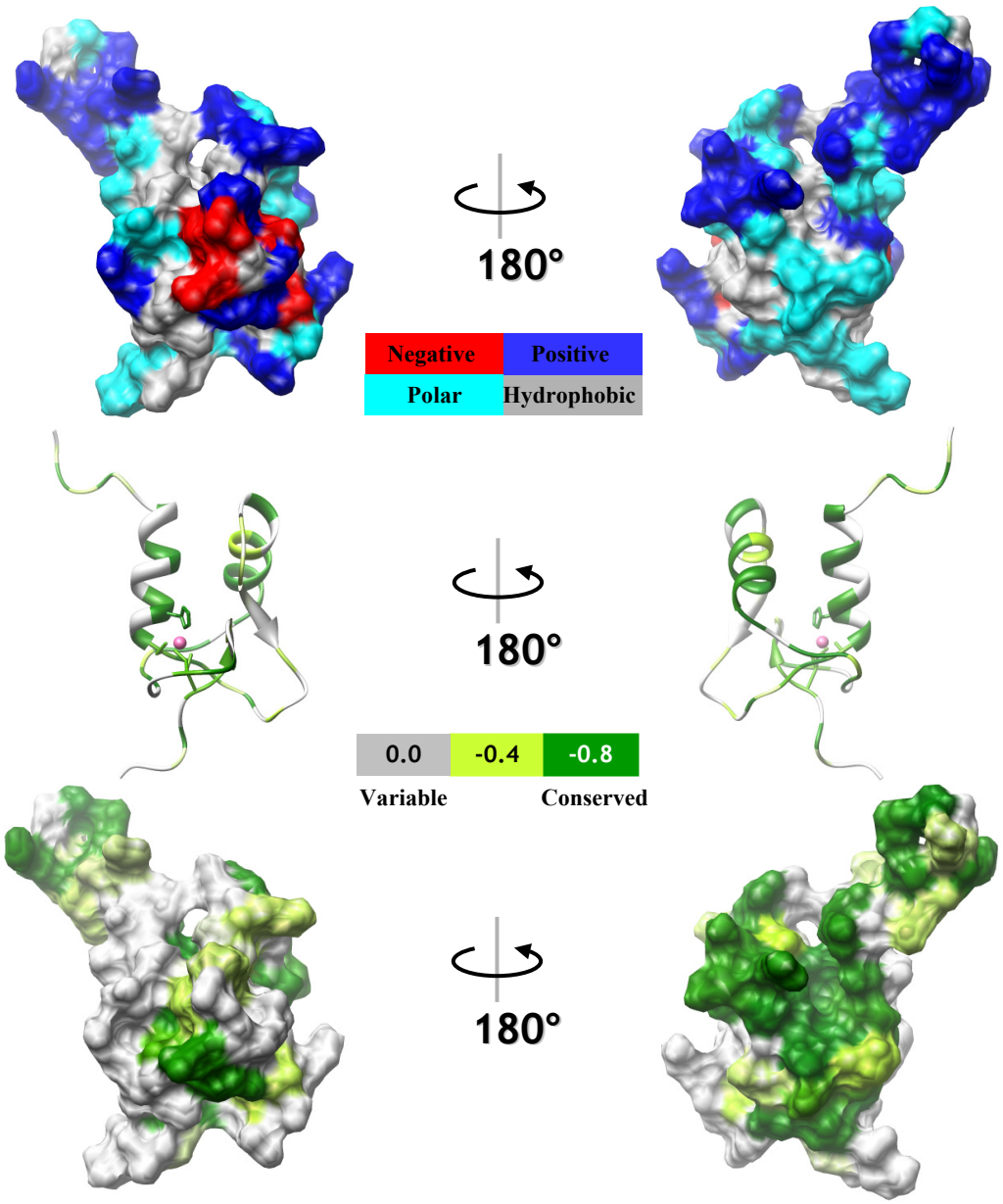


Figure 4.8 **NMR titration of 15N-labeled SAP30 ZnF with a DNA duplex.** (A) Overlay of HSQC spectra recorded in the absence (*blue*) and the presence (*red*) of a DNA duplex. For the spectrum of the apo-form, only the outside contours are depicted for easier resonances tracking. (B) Plots of both chemical shift deviation and resonance intensity ratio derived from (A) are shown against residue number. Residues underwent biggest perturbation upon addition of DNA duplex are labeled.

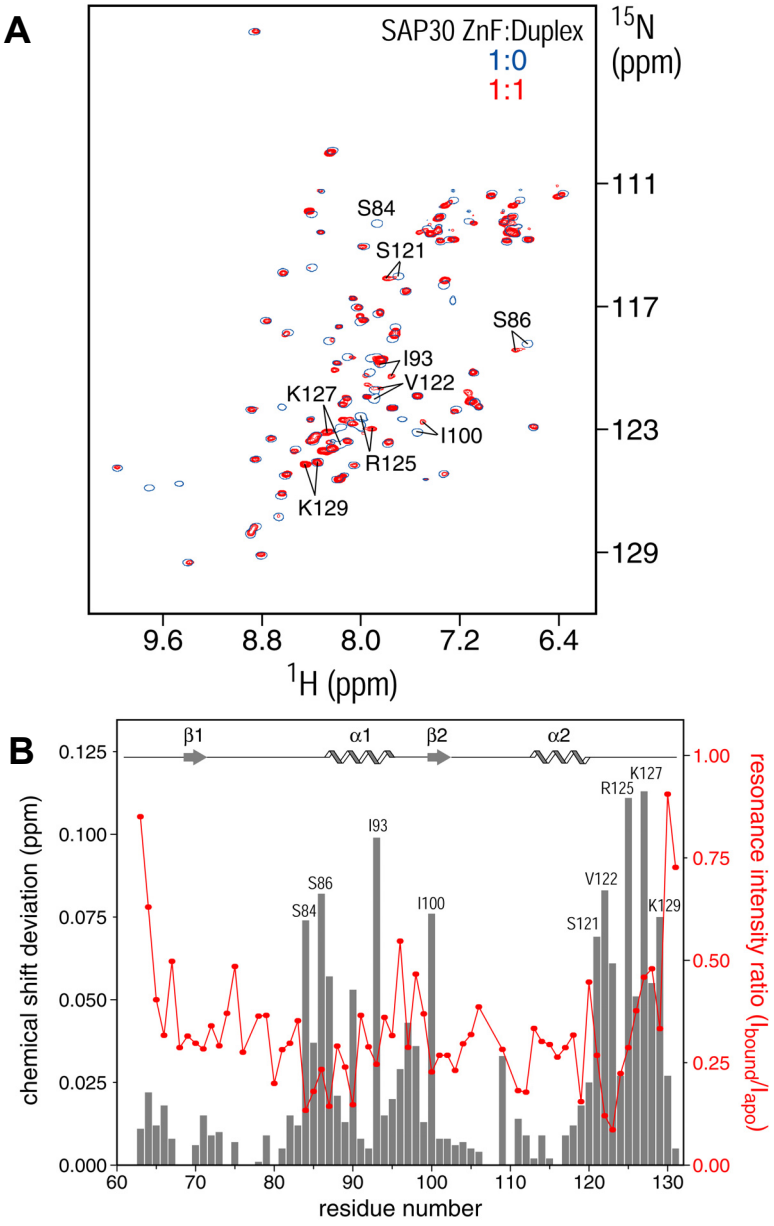
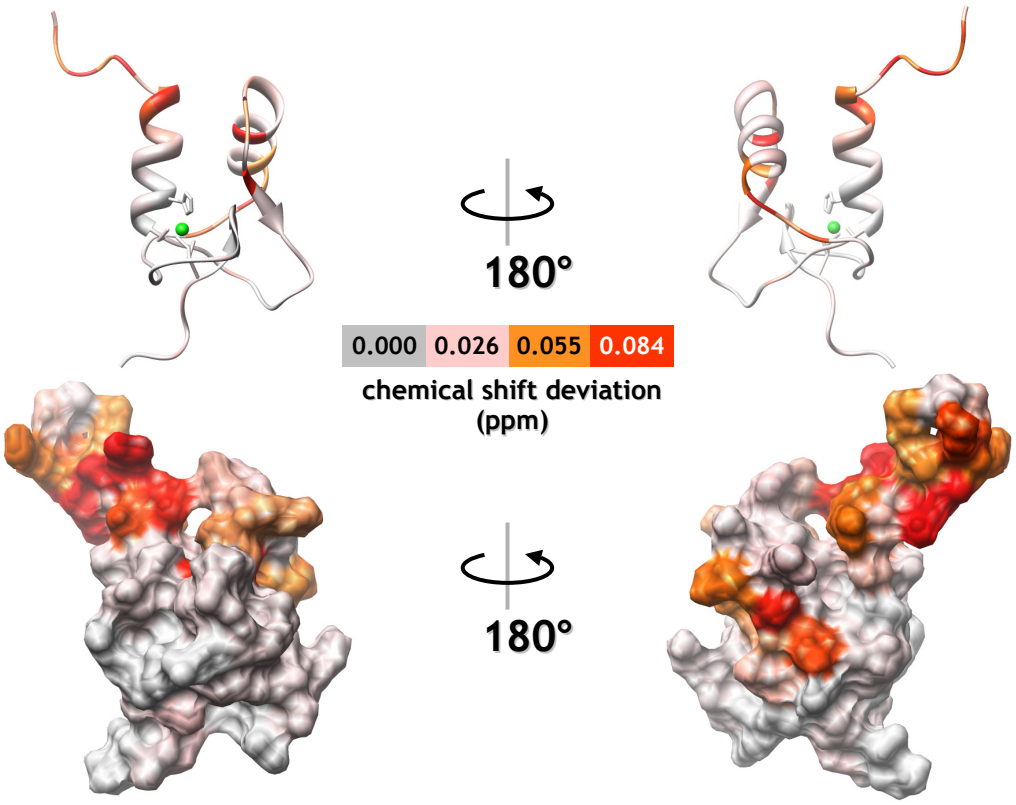


Figure 4.9 **Structure of the SAP30 ZnF shown in the cartoon representations highlighting various degrees of perturbation upon addition of DNA duplex (ribbon structure, *top pannels*; surface representation, *bottom pannels*).**



In Summary, I have conducted structural and functional analyses to a newly discovered zinc-finger motif within the evolutionarily conserved SAP30. Using a combination of mutagenesis and long-range NOEs characterization, the zinc coordination pattern was successfully defined, which was incorporated later into the structural determination of the zinc-finger motif. Structures of reasonable quality and precision and in agreement with input experimental data were obtained and they typically comprise two α-helices and two β-strands. This motif appears to adopt a novel structure since it does not conform to any of the known folds, and interestingly, highly conserved residues within this motif compose a relatively small, but contiguous surface encompassing the two helices, which is potentially its ligand-binding site. HSQC-based NMR technique was then used to characterize its biological role, which led to a series of negative results. Several protein domains that have been implicated in targeting SAP30 were firstly tested. However, there was no compelling evidence showing any direct interactions in my hand. A number of ligands containing various histone modifications were then tested by NMR titrations, yet again led to ruling out the possibility of this motif to be a histone code reader. Fortunately, a possible function for this zinc-finger motif was finally revealed by further sequence analysis and NMR titration experiments using a DNA duplex, which indicated that there might be a direct interaction between the domain and the DNA. This could be another example to overturn the conventional view of the lack of intrinsic DNA-binding activity within chromatin regulatory complexes (Da et al., 2006; Tu et al., 2008). At this moment we do not know if this novel motif is sequence-specific DNA-binding domain, and in fact, it is also possible for it to target RNA instead, both of which require further investigation and might lead to a previously unknown function for the Sin3 corepressor complex.

CHAPTER FIVE

Summary and Prospectus

5.1 Apparent promiscuity of interactions involving mSin3A PAH domains

As a key co-regulator of the eukaryotic gene transcription, the Sin3 proteins appear to serve not only as molecular adapters bridging the enzymatic components of the complex with numerous transcription regulators but also as molecular scaffolds for the assembly of distinct corepressor complexes with various intracellular roles. In spite of their lack of common motifs, the Sin3 proteins harbor four PAH domains that are highly conserved ranging from yeast to human. The N-terminal regions of the Sin3 proteins spanning the first two PAH domains have been proposed to function as targeting sites for a plethora of promoter-bound transcription factors, while the PAH3 along with HID domains appear to lay the platform for the complex assembly (Figure 1.3). During the past decade, tremendous progress has been accomplished in getting to know how the Sin3 proteins are recruited onto specific loci on the genome, yet apparent promiscuity for these diverse-function proteins still exist. For example, is there any biological implication for the mSin3A PAH2 domain to have the intrinsic ability to form homodimers? Does mSin3A PAH3 share a similar recognizing mode with PAH1 and PAH2 domains in binding their targets? Since SAP30 serves as a particularly important protein-protein interaction hub within the complex, how does the interaction between PAH3 and SAP30 contributes to the variable complex assembly. Answering these questions will not only further our understanding of the mechanisms of recruitment and complex assembly for Sin3 but also provide paradigmatic basis for characterization of other HDAC-associated corepressor complexes.

Previous structural studies revealed that the mammalian Sin3A (mSin3A) PAH2 domain interacted through a hydrophobic cleft with diverse targets by binding to chain-reversible sequence motifs, thereby defining at least two distinct classes of PAH2 interactors (Brubaker et

al., 2000; Swanson et al., 2004). The apo-mSin3A PAH2 domain exhibits conformational heterogeneity, with one of the conformers existing in a partially unfolded state, and homodimerizes with modest affinity suggesting plausible mechanisms for binding to a broad range of targets through new surfaces on the one hand and for regulating Sin3 function via occlusion of the hydrophobic cleft on the other (Zhang et al., 2006). The PAH2 domain of the paralogous mSin3B protein engages targets in a manner similar to mSin3A PAH2 (Spronk et al., 2000; van Ingen et al., 2004), but the apo-protein is monomeric and is fully folded, although a small population of a minor conformer has been reported (van Ingen et al., 2006). Here we have shown that the domains are structurally independent and the properties of the individual domains, such as the conformational heterogeneity and the ability to homo-dimerize of mSin3A PAH2, are preserved in constructs that span both PAH domains. Our results thus suggest that the N-terminal segments of the Sin3 proteins are broadly available for interactions with other proteins and that the PAH domains are organized into structurally independent modules. Our data also rule out any heterotypic association between the paralogous mSin3A and mSin3B proteins via interactions involving the mSin3A PAH2 domain.

SAP30 has been proposed to interact with several transcription factors but its interaction with Sin3 is the best characterized and is mediated through the PAH3 domain. As a first step towards clarifying the mechanism of assembly of the Sin3 corepressor complex, we are undertaking structural studies of the mSin3A PAH3-SAP30 SID interaction. The ^{1}H-^{15}N HSQC spectra of the respective proteins in the complex are indicative of folded domains, although the presence of poorly dispersed, narrow resonances in the SAP30 SID spectrum indicates natively unfolded regions. The secondary chemical shifts of backbone nuclei in the respective proteins indicates that the PAH3 and SID domains are largely helical. Like other PAH domains the

PAH3 domain forms a four-helix bundle, but with helix α3 much shorter than in the PAH1 and PAH2 domains. Intriguingly, the SAP30 SID adopts a structure comprising three helices which is a departure from the structural motif used by proteins targeting the mSin3A PAH1 and PAH2 domains. The determination of the complex structure is presently underway, and it is still unclear how the three SAP30 SID helices are arranged and how they engage the PAH domain.

In the course of our characterization of the mSin3A–SAP30 interaction, we discovered a novel zinc finger motif included within SAP30, whose sequence does not conform to any of the known motifs in sequence databases. To gain insights into its uncovered cellular role, we obtained the solution structure of this zinc-finger motif using solution NMR methods. Two close but discrete positively charged surfaces are transected by a deep hydrophobic cleft. Interestingly, the above region of the motif is also highly conserved and a NMR titration analysis conducted using a DNA duplex suggested this particular surface of the molecule indeed overlap with its positively charge region. It is highly possible that this newly discovered zinc-finger motif functions in binding nucleic acid and our findings implies a previously unknown function for the Sin3 corepressor complex.

5.2 Functional roles of the evolutionarily conserved SAP30

SAP30 is one of the best-characterized Sin3 associated polypeptide and highly conserved across a broad range of species ranging from yeast to human. But how it is able to simultaneously engage so many targets besides Sin3 including RBP1 (Lai et al., 2001), SAP180 (Binda et al., 2006), the ING family proteins (Kuzmichev et al., 2002; Shi et al., 2006) and other transcription factors such as the NCoR, CIR, and yin yang 1 (YY1), (Laherty et al., 1998; Hsieh et al., 1999; Huang et al., 2003) still remains elusive. Furthermore, our structural and functional

analyses of different domains in the mammalian SAP30 raised more questions for us to answer. Does the SAP30 ZnF also recognize RNA? What does its nucleic acid binding activity implicates? Are there specific interactions between SAP30 and RBP1, ING2? If not, what is the real biological role for such a tight interaction between mSin3A PAH3 and SAP30 SID?

Besides its well-known function in promoter-dependent transcription repression, the Sin3 corepressor complex has also been shown to participate in epigenetic signal recognition in the pericentric heterochromatin, thereby play a particularly important role in normal chromosome segregation (David et al., 2003; Silverstein et al., 2003; Xin et al., 2004). So far it is still not clear whether SAP30 plays any role in the recruitment of the corepressor complex to specific loci into the centromeres and whether the SAP30 ZnF is a sequence specific DNA binding domain, but a SELEX (Systematic Evolution of Ligands by Exponential Enrichment)-based screening will be explored in the near future to address this possibility.

Intriguingly, recent genetic and biochemical studies of a closely related protein of SAP30 – SAP30-like (SAP30L) may answer some of the above questions (Lindfors et al., 2003; Viiri et al., 2006; Korkeamaki et al., 2008). SAP30L is a 183-residue protein sharing around 70% sequence identity with its mammalian counterpart and only present in higher organisms. It contains both the zinc-finger motif and the C-terminal SID in addition to a much shorter N-terminal low-complexity region (Lindfors et al., 2003). It has been recently indicated in targeting the mSin3A into the nucleolus through a putative 8AA nucleolar localization signal, which is also present in SAP30 and overlaps with the first helix of the SAP30 SID domain (Viiri et al., 2006). Therefore, RNA binding activity remains a formal possibility for the zinc-finger motif within SAP30.

5.3 Reconstitution of the core Sin3 corepressor complex

Compared to the overwhelming studies of recruitment mechanisms of the Sin3 corepressor complexes, the investigation for how the corepressor is assembled obviously lagged behind. Our structure-functional analysis of the mSin3A PAH3-SAP30 SID could potentially serves as a starting point for dissecting the corepressor complex. We have attempted to extend this study by reconstituting multi-subunit complexes with other protein factors that has been shown previously to target this portion of the corepressor including the ING proteins and RBP1. However the failure of seeing direct interactions in my hands indicated that there might be other proteins involved in these processes or some unidentified post-translational modification may be necessary for their interactions.

Furthermore, it still remains elusive for how the HID domain of the Sin3 proteins is engaged with other key entities of the complex including SAP45, SAP130, and SAP180, and how the HDAC core complex including HDAC1, HDAC2, RbAp46/48 is recruited onto the core complex of Sin3. Our reconstitution analysis making use of a polycistronic expression strategy thus far (Appendix three) provided a promising alternative toward our final goal – elucidation of the protein-protein interaction network at the domain-level in the Sin3 corepressor complex. Indeed, this method has been successfully used to evaluate protein complexes with inadequately characterized interaction networks like in our case (Tan, 2001; Tan et al., 2005; Fitzgerald et al., 2006; Tolia and Joshua-Tor, 2006).

References

Alland L, David G, Shen-Li H, Potes J, Muhle R, Lee HC, Hou H, Jr., Chen K, DePinho RA (2002) Identification of mammalian Sds3 as an integral component of the Sin3/histone deacetylase corepressor complex. Mol Cell Biol 22:2743-2750.

Ayer DE (1999) Histone deacetylases: transcriptional repression with SINers and NuRDs. Trends Cell Biol 9:193-198.

Ayscough KR (2005) Coupling actin dynamics to the endocytic process in Saccharomyces cerevisiae. Protoplasma 226:81-88.

Beal R, Deveraux Q, Xia G, Rechsteiner M, Pickart C (1996) Surface hydrophobic residues of multiubiquitin chains essential for proteolytic targeting. Proc Natl Acad Sci U S A 93:861-866.

Bernstein BE, Tong JK, Schreiber SL (2000) Genomewide studies of histone deacetylase function in yeast. Proc Natl Acad Sci U S A 97:13708-13713.

Bhaumik SR, Smith E, Shilatifard A (2007) Covalent modifications of histones during development and disease pathogenesis. Nat Struct Mol Biol 14:1008-1016.

Binda O, Roy JS, Branton PE (2006) RBP1 family proteins exhibit SUMOylation-dependent transcriptional repression and induce cell growth inhibition reminiscent of senescence. Mol Cell Biol 26:1917-1931.

Brubaker K, Cowley SM, Huang K, Loo L, Yochum GS, Ayer DE, Eisenman RN, Radhakrishnan I (2000) Solution structure of the interacting domains of the Mad-Sin3 complex: implications for recruitment of a chromatin-modifying complex. Cell 103:655-665.

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 54 (Pt 5):905-921.

Buratowski S (1995) Mechanisms of gene activation. Science 270:1773-1774.

Carrozza MJ, Florens L, Swanson SK, Shia WJ, Anderson S, Yates J, Washburn MP, Workman JL (2005a) Stable incorporation of sequence specific repressors Ash1 and Ume6 into the Rpd3L complex. Biochim Biophys Acta 1731:77-87; discussion 75-76.

Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia WJ, Anderson S, Yates J, Washburn MP, Workman JL (2005b) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell 123:581-592.

Chen X, Bieker JJ (2004) Stage-specific repression by the EKLF transcriptional activator. Mol Cell Biol 24:10416-10424.

Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13:289-302.

Cowley SM, Iritani BM, Mendrysa SM, Xu T, Cheng PF, Yada J, Liggitt HD, Eisenman RN (2005) The mSin3A chromatin-modifying complex is essential for embryogenesis and T-cell development. Mol Cell Biol 25:6990-7004.

Da G, Lenkart J, Zhao K, Shiekhattar R, Cairns BR, Marmorstein R (2006) Structure and function of the SWIRM domain, a conserved protein module found in chromatin regulatory complexes. Proc Natl Acad Sci U S A 103:2057-2062.

Dang DT, Pevsner J, Yang VW (2000) The biology of the mammalian Kruppel-like family of transcription factors. Int J Biochem Cell Biol 32:1103-1121.

Dannenberg JH, David G, Zhong S, van der Torre J, Wong WH, Depinho RA (2005) mSin3A corepressor regulates diverse transcriptional networks governing normal and neoplastic growth and survival. Genes Dev 19:1581-1595.

David G, Turner GM, Yao Y, Protopopov A, DePinho RA (2003) mSin3-associated protein, mSds3, is essential for pericentric heterochromatin formation and chromosome segregation in mammalian cells. Genes Dev 17:2396-2405.

Demarest SJ, Martinez-Yamout M, Chung J, Chen H, Xu W, Dyson HJ, Evans RM, Wright PE (2002) Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. Nature 415:549-553.

Denslow SA, Wade PA (2007) The human Mi-2/NuRD complex and gene regulation. Oncogene 26:5433-5438.

Dhalluin C, Carlson JE, Zeng L, He C, Aggarwal AK, Zhou MM (1999) Structure and ligand of a histone acetyltransferase bromodomain. Nature 399:491-496.

Dikic I (2002) CIN85/CMS family of adaptor molecules. FEBS Lett 529:110-115.

Elgin SCR, Workman JL (2000) Chromatin structure and gene expression, Second Edition. New York, NY: Oxford University Press.

Ellenrieder V, Buck A, Harth A, Jungert K, Buchholz M, Adler G, Urrutia R, Gress TM (2004) KLF11 mediates a critical mechanism in TGF-beta signaling that is inactivated by Erk-MAPK in pancreatic cancer cells. Gastroenterology 127:607-620.

Ferentz AE, Wagner G (2000) NMR spectroscopy: a multifaceted approach to macromolecular structure. Q Rev Biophys 33:29-65.

Fitzgerald DJ, Berger P, Schaffitzel C, Yamada K, Richmond TJ, Berger I (2006) Protein complex expression by using multigene baculoviral vectors. Nat Methods 3:1021-1032.

Fleischer TC, Yun UJ, Ayer DE (2003) Identification and characterization of three new components of the mSin3A corepressor complex. Mol Cell Biol 23:3456-3467.

Gill SC, von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem 182:319-326.

Grzesiek S, Bax A (1993) Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. J Biomol NMR 3:185-204.

Guenther MG, Barak O, Lazar MA (2001) The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3. Mol Cell Biol 21:6091-6101.

Hassig CA, Fleischer TC, Billin AN, Schreiber SL, Ayer DE (1997) Histone deacetylase activity is required for full transcriptional repression by mSin3A. Cell 89:341-347.

He Y, Radhakrishnan I (2008) Solution NMR studies of apo-mSin3A and mSin3B reveal that the PAH1 and PAH2 domains are structurally independent. Protein Sci in press.

He Y, Hicke L, Radhakrishnan I (2007) Structural basis for ubiquitin recognition by SH3 domains. J Mol Biol 373:190-196.

Herrera JE, Bergel M, Yang XJ, Nakatani Y, Bustin M (1997) The histone acetyltransferase activity of human GCN5 and PCAF is stabilized by coenzymes. J Biol Chem 272:27253-27258.

Hicke L (2001) Protein regulation by monoubiquitin. Nat Rev Mol Cell Biol 2:195-201.

Hicke L, Dunn R (2003) Regulation of membrane protein transport by ubiquitin and ubiquitin-binding proteins. Annu Rev Cell Dev Biol 19:141-172.

Hsieh JJ, Zhou S, Chen L, Young DB, Hayward SD (1999) CIR, a corepressor linking the DNA binding factor CBF1 to the histone deacetylase complex. Proc Natl Acad Sci U S A 96:23-28.

Huang NE, Lin CH, Lin YS, Yu WC (2003) Modulation of YY1 activity by SAP30. Biochem Biophys Res Commun 306:267-275.

Huang Y, Myers SJ, Dingledine R (1999) Transcriptional repression by REST: recruitment of Sin3A and histone deacetylase to neuronal genes. Nat Neurosci 2:867-872.

Hurley JH, Lee S, Prag G (2006) Ubiquitin-binding domains. Biochem J 399:361-372.

Hurlin PJ, Queva C, Eisenman RN (1997) Mnt: a novel Max-interacting protein and Myc antagonist. Curr Top Microbiol Immunol 224:115-121.

Jacobson RH, Ladurner AG, King DS, Tjian R (2000) Structure and function of a human TAFII250 double bromodomain module. Science 288:1422-1425.

Jenuwein T, Allis CD (2001) Translating the histone code. Science 293:1074-1080.

Jepsen K, Rosenfeld MG (2002) Biological roles and mechanistic actions of co-repressor complexes. J Cell Sci 115:689-698.

Johnson BA (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. Methods Mol Biol 278:313-352.

Jozic D, Cardenes N, Deribe YL, Moncalian G, Hoeller D, Groemping Y, Dikic I, Rittinger K, Bravo J (2005) Cbl promotes clustering of endocytic adaptor proteins. Nat Struct Mol Biol 12:972-979.

Kaczynski J, Cook T, Urrutia R (2003) Sp1- and Kruppel-like transcription factors. Genome Biol 4:206.

Kang RS, Daniels CM, Francis SA, Shih SC, Salerno WJ, Hicke L, Radhakrishnan I (2003) Solution structure of a CUE-ubiquitin complex reveals a conserved mode of ubiquitin binding. Cell 113:621-630.

Karagianni P, Wong J (2007) HDAC3: taking the SMRT-N-CoRrect road to repression. Oncogene 26:5439-5449.

Keogh MC, Kurdistani SK, Morris SA, Ahn SH, Podolny V, Collins SR, Schuldiner M, Chin K, Punna T, Thompson NJ, Boone C, Emili A, Weissman JS, Hughes TR, Strahl BD, Grunstein M, Greenblatt JF, Buratowski S, Krogan NJ (2005) Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell 123:593-605.

Kim H, Lee JE, Cho EJ, Liu JO, Youn HD (2003) Menin, a tumor suppressor, represses JunD-mediated transcriptional activity by association with an mSin3A-histone deacetylase complex. Cancer Res 63:6135-6139.

Kingston RE, Bunker CA, Imbalzano AN (1996) Repression and activation by multiprotein complexes that alter chromatin structure. Genes Dev 10:905-920.

Klose RJ, Zhang Y (2007) Regulation of histone methylation by demethylimination and demethylation. Nat Rev Mol Cell Biol 8:307-318.

Knoepfler PS, Eisenman RN (1999) Sin meets NuRD and other tails of repression. Cell 99:447-450.

Koipally J, Renold A, Kim J, Georgopoulos K (1999) Repression by Ikaros and Aiolos is mediated through histone deacetylase complexes. EMBO J 18:3090-3100.

Koleske AJ, Young RA (1995) The RNA polymerase II holoenzyme and its implications for gene regulation. Trends Biochem Sci 20:113-116.

Korkeamaki H, Viiri K, Kukkonen MK, Maki M, Lohi O (2008) Alternative mRNA splicing of SAP30L regulates its transcriptional repression activity. FEBS Lett 582:379-384.

Kornberg RD, Lorch Y (1991) Irresistible force meets immovable object: transcription and the nucleosome. Cell 67:833-836.

Korzus E, Torchia J, Rose DW, Xu L, Kurokawa R, McInerney EM, Mullen TM, Glass CK, Rosenfeld MG (1998) Transcription factor-specific requirements for coactivators and their acetyltransferase functions. Science 279:703-707.

Kuzmichev A, Zhang Y, Erdjument-Bromage H, Tempst P, Reinberg D (2002) Role of the Sin3-histone deacetylase complex in growth regulation by the candidate tumor suppressor p33(ING1). Mol Cell Biol 22:835-848.

Laherty CD, Yang WM, Sun JM, Davie JR, Seto E, Eisenman RN (1997) Histone deacetylases associated with the mSin3 corepressor mediate Mad transcriptional repression. Cell 89:349-356.

Laherty CD, Billin AN, Lavinsky RM, Yochum GS, Bush AC, Sun JM, Mullen TM, Davie JR, Rose DW, Glass CK, Rosenfeld MG, Ayer DE, Eisenman RN (1998) SAP30, a component of the mSin3 corepressor complex involved in N-CoR-mediated repression by specific transcription factors. Mol Cell 2:33-42.

Lai A, Kennedy BK, Barbie DA, Bertos NR, Yang XJ, Theberge MC, Tsai SC, Seto E, Zhang Y, Kuzmichev A, Lane WS, Reinberg D, Harlow E, Branton PE (2001) RBP1 recruits the mSIN3-histone deacetylase complex to the pocket of retinoblastoma tumor suppressor family proteins found in limited discrete regions of the nucleus at growth arrest. Mol Cell Biol 21:2918-2932.

Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8:477-486.

Legge GB, Martinez-Yamout MA, Hambly DM, Trinh T, Lee BM, Dyson HJ, Wright PE (2004) ZZ domain of CBP: an unusual zinc finger fold in a protein interaction module. J Mol Biol 343:1081-1093.

Li B, Gogol M, Carey M, Lee D, Seidel C, Workman JL (2007) Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin. Science 316:1050-1054.

Li SS (2005) Specificity and versatility of SH3 and other proline-recognition domains: structural basis and implications for cellular signal transduction. Biochem J 390:641-653.

Lindfors K, Viiri KM, Niittynen M, Heinonen TY, Maki M, Kainulainen H (2003) TGF-beta induces the expression of SAP30L, a novel nuclear protein. BMC Genomics 4:53.

Linge JP, Habeck M, Rieping W, Nilges M (2003) ARIA: automated NOE assignment and NMR structure calculation. Bioinformatics 19:315-316.

Linge JP, Habeck M, Rieping W, Nilges M (2004) Correction of spin diffusion during iterative automated NOE assignment. J Magn Reson 167:334-342.

Lohr F, Ruterjans H (1996) Novel Pulse Sequences for the Resonance Assignment of Aromatic Side Chains in 13C-Labeled Proteins. J Magn Reson B 112:259-268.

Lutterbach B, Westendorf JJ, Linggi B, Isaac S, Seto E, Hiebert SW (2000) A mechanism of repression by acute myeloid leukemia-1, the target of multiple chromosomal translocations in acute leukemia. J Biol Chem 275:651-656.

Mahadev RK, Di Pietro SM, Olson JM, Piao HL, Payne GS, Overduin M (2007) Structure of Sla1p homology domain 1 and interaction with the NPFxD endocytic internalization motif. Embo J 26:1963-1971.

Mallory MJ, Strich R (2003) Ume1p represses meiotic gene transcription in Saccharomyces cerevisiae through interaction with the histone deacetylase Rpd3p. J Biol Chem 278:44727-44734.

Mayer BJ (2001) SH3 domains: complexity in moderation. J Cell Sci 114:1253-1263.

McArthur GA, Laherty CD, Queva C, Hurlin PJ, Loo L, James L, Grandori C, Gallant P, Shiio Y, Hokanson WC, Bush AC, Cheng PF, Lawrence QA, Pulverer B, Koskinen PJ, Foley KP, Ayer DE, Eisenman RN (1998) The Mad protein family links transcriptional repression to cell differentiation. Cold Spring Harb Symp Quant Biol 63:423-433.

Murphy M, Ahn J, Walker KK, Hoffman WH, Evans RM, Levine AJ, George DL (1999) Transcriptional repression by wild-type p53 utilizes histone deacetylases, mediated by interaction with mSin3a. Genes Dev 13:2490-2501.

Naruse Y, Aoki T, Kojima T, Mori N (1999) Neural restrictive silencer factor recruits mSin3 and histone deacetylase complex to repress neuron-specific target genes. Proc Natl Acad Sci U S A 96:13691-13696.

Nicolas E, Yamada T, Cam HP, Fitzgerald PC, Kobayashi R, Grewal SI (2007) Distinct roles of HDAC complexes in promoter silencing, antisense suppression and DNA damage protection. Nat Struct Mol Biol 14:372-380.

Nikolaev AY, Papanikolaou NA, Li M, Qin J, Gu W (2004) Identification of a novel BRMS1-homologue protein p40 as a component of the mSin3A/p33(ING1b)/HDAC1 deacetylase complex. Biochem Biophys Res Commun 323:1216-1222.

Nomura M, Uda-Tochio H, Murai K, Mori N, Nishimura Y (2005) The neural repressor NRSF/REST binds the PAH1 domain of the Sin3 corepressor by using its distinct short hydrophobic helix. J Mol Biol 354:903-915.

Otting G, Wuthrich K (1990) Heteronuclear filters in two-dimensional [1H,1H]-NMR spectroscopy: combined use with isotope labelling for studies of macromolecular conformation and intermolecular interactions. Q Rev Biophys 23:39-96.

Pang YP, Kumar GA, Zhang JS, Urrutia R (2003) Differential binding of Sin3 interacting repressor domains to the PAH2 domain of Sin3A. FEBS Lett 548:108-112.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25:1605-1612.

Plevin MJ, Mills MM, Ikura M (2005) The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. Trends Biochem Sci 30:66-69.

Ptashne M, Gann A (2002) Genes and Signals, First Edition. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Radhakrishnan I, Perez-Alvarado GC, Parker D, Dyson HJ, Montminy MR, Wright PE (1999) Structural analyses of CREB-CBP transcriptional activator-coactivator complexes by NMR spectroscopy: implications for mapping the boundaries of structural domains. J Mol Biol 287:859-865.

Roeder RG (1991) The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. Trends Biochem Sci 16:402-408.

Romm E, Nielsen JA, Kim JG, Hudson LD (2005) Myt1 family recruits histone deacetylase to regulate neural transcription. J Neurochem 93:1444-1453.

Sahu SC, Swanson KA, Kang RS, Huang K, Brubaker K, Ratcliff K, Radhakrishnan I (2008) Conserved themes in target recognition by the PAH1 and PAH2 domains of the Sin3 transcriptional corepressor. J Mol Biol 375:1444-1456.

Salerno WJ, Seaver SM, Armstrong BR, Radhakrishnan I (2004) MONSTER: inferring non-covalent interactions in macromolecular structures from atomic coordinate data. Nucleic Acids Res 32:W566-568.

Schreiber-Agus N, DePinho RA (1998) Repression by the Mad(Mxi1)-Sin3 complex. Bioessays 20:808-818.

Sharma M, Sun Z (2001) 5'TG3' interacting factor interacts with Sin3A and represses AR-mediated transcription. Mol Endocrinol 15:1918-1928.

Shi X, Hong T, Walter KL, Ewalt M, Michishita E, Hung T, Carney D, Pena P, Lan F, Kaadige MR, Lacoste N, Cayrou C, Davrazou F, Saha A, Cairns BR, Ayer DE, Kutateladze TG, Shi Y, Cote J, Chua KF, Gozani O (2006) ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature 442:96-99.

Shi Y (2007) Histone lysine demethylases: emerging roles in development, physiology and disease. Nat Rev Genet 8:829-833.

Sif S, Saurin AJ, Imbalzano AN, Kingston RE (2001) Purification and characterization of mSin3A-containing Brg1 and hBrm chromatin remodeling complexes. Genes Dev 15:603-618.

Silverstein RA, Ekwall K (2005) Sin3: a flexible regulator of global gene expression and genome stability. Curr Genet 47:1-17.

Silverstein RA, Richardson W, Levin H, Allshire R, Ekwall K (2003) A new role for the transcriptional corepressor SIN3; regulation of centromeres. Curr Biol 13:68-72.

Spencer TE, Jenster G, Burcin MM, Allis CD, Zhou J, Mizzen CA, McKenna NJ, Onate SA, Tsai SY, Tsai MJ, O'Malley BW (1997) Steroid receptor coactivator-1 is a histone acetyltransferase. Nature 389:194-198.

Spronk CA, Tessari M, Kaan AM, Jansen JF, Vermeulen M, Stunnenberg HG, Vuister GW (2000) The Mad1-Sin3B interaction involves a novel helical fold. Nat Struct Biol 7:1100-1104.

Stamenova SD, French ME, He Y, Francis SA, Kramer ZB, Hicke L (2007) Ubiquitin binds to and regulates a subset of SH3 domains. Mol Cell 25:273-284.

Stillman B (1998) Mechanisms of Transcription. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Stols L, Gu M, Dieckman L, Raffen R, Collart FR, Donnelly MI (2002) A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. Protein Exp Purif 25:8-15.

Sun ZW, Hampsey M (1999) A general requirement for the Sin3-Rpd3 histone deacetylase complex in regulating silencing in Saccharomyces cerevisiae. Genetics 152:921-932.

Swanson KA, Knoepfler PS, Huang K, Kang RS, Cowley SM, Laherty CD, Eisenman RN, Radhakrishnan I (2004) HBP1 and Mad1 repressors bind the Sin3 corepressor PAH2 domain with opposite helical orientations. Nat Struct Mol Biol 11:738-746.

Tan S (2001) A modular polycistronic expression system for overexpressing protein complexes in Escherichia coli. Protein Expr Purif 21:224-234.

Tan S, Kern RC, Selleck W (2005) The pST44 polycistronic expression system for producing protein complexes in Escherichia coli. Protein Expr Purif 40:385-395.

Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol 14:1025-1040.

Thiagalingam S, Cheng KH, Lee HJ, Mineva N, Thiagalingam A, Ponte JF (2003) Histone deacetylases: unique players in shaping the epigenetic histone code. Ann N Y Acad Sci 983:84-100.

Tolia NH, Joshua-Tor L (2006) Strategies for protein coexpression in Escherichia coli. Nat Methods 3:55-64.

Tu S, Teng YC, Yuan C, Wu YT, Chan MY, Cheng AN, Lin PH, Juan LJ, Tsai MD (2008) The ARID domain of the H3K4 demethylase RBP2 binds to a DNA CCGCCC motif. Nat Struct Mol Biol 15:419-421.

van Ingen H, Baltussen MA, Aelen J, Vuister GW (2006) Role of Structural and Dynamical Plasticity in Sin3: The Free PAH2 Domain is a Folded Module in mSin3B. J Mol Biol.

van Ingen H, Lasonder E, Jansen JF, Kaan AM, Spronk CA, Stunnenberg HG, Vuister GW (2004) Extension of the binding motif of the Sin3 interacting domain of the Mad family proteins. Biochemistry 43:46-54.

VanDemark AP, Kasten MM, Ferris E, Heroux A, Hill CP, Cairns BR (2007) Autoregulation of the rsc4 tandem bromodomain by gcn5 acetylation. Mol Cell 27:817-828.

Viiri KM, Korkeamaki H, Kukkonen MK, Nieminen LK, Lindfors K, Peterson P, Maki M, Kainulainen H, Lohi O (2006) SAP30L interacts with members of the Sin3A corepressor complex and targets Sin3A to the nucleolus. Nucleic Acids Res 34:3288-3298.

Vijay-Kumar S, Bugg CE, Cook WJ (1987) Structure of ubiquitin refined at 1.8 A resolution. J Mol Biol 194:531-544.

Wittekind M, Mapelli C, Lee V, Goldfarb V, Friedrichs MS, Meyers CA, Mueller L (1997) Solution structure of the Grb2 N-terminal SH3 domain complexed with a ten-residue peptide derived from SOS: direct refinement against NOEs, J-couplings and 1H and 13C chemical shifts. J Mol Biol 267:933-952.

Wolffe AP, Hayes JJ (1999) Chromatin disruption and modification. Nucleic Acids Res 27:711-720.

Wotton D, Knoepfler PS, Laherty CD, Eisenman RN, Massague J (2001) The Smad transcriptional corepressor TGIF recruits mSin3. Cell Growth Diff 12:457-463.

Xin H, Yoon HG, Singh PB, Wong J, Qin J (2004) Components of a pathway maintaining histone modification and heterochromatin protein 1 binding at the pericentric heterochromatin in Mammalian cells. J Biol Chem 279:9539-9546.

Xue Y, Wong J, Moreno GT, Young MK, Cote J, Wang W (1998) NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. Mol Cell 2:851-861.

Yang L, Mei Q, Zielinska-Kwiatkowska A, Matsui Y, Blackburn ML, Benedetti D, Krumm AA, Taborsky GJ, Jr., Chansky HA (2003) An ERG (ets-related gene)-associated histone methyltransferase interacts with histone deacetylases 1/2 and transcription co-repressors mSin3A/B. Biochem J 369:651-657.

Yang Q, Kong Y, Rothermel B, Garry DJ, Bassel-Duby R, Williams RS (2000) The winged-helix/forkhead protein myocyte nuclear factor beta (MNF-β) forms a co-repressor complex with mammalian sin3B. Biochem J 345 Pt 2:335-343.

Yang X, Zhang F, Kudlow JE (2002) Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein O-GlcNAcylation to transcriptional repression. Cell 110:69-80.

Yochum GS, Ayer DE (2001) Pf1, a novel PHD zinc finger protein that links the TLE corepressor to the mSin3A-histone deacetylase complex. Mol Cell Biol 21:4110-4118.

Zeng L, Zhou MM (2002) Bromodomain: an acetyl-lysine binding domain. FEBS Lett 513:124-128.

Zhang JS, Moncrieffe MC, Kaczynski J, Ellenrieder V, Prendergast FG, Urrutia R (2001) A conserved alpha-helical motif mediates the interaction of Sp1-like transcriptional repressors with the corepressor mSin3A. Mol Cell Biol 21:5041-5049.

Zhang Y, Zhang Z, Demeler B, Radhakrishnan I (2006) Coupled unfolding and dimerization by the PAH2 domain of the mammalian Sin3A corepressor. J Mol Biol 360:7-14.

Zhang Y, Sun ZW, Iratni R, Erdjument-Bromage H, Tempst P, Hampsey M, Reinberg D (1998) SAP30, a novel protein conserved between human and yeast, is a component of a histone deacetylase complex. Mol Cell 1:1021-1031.

APPENDIX ONE

Structural Basis for Ubiquitin Recognition by SH3 Domains

A1.1 Introduction

The SH3 domain is a protein-protein interaction module found in many proteins in diverse eukaryotes. The best-characterized function of these domains, which were initially described in signaling proteins, is to bind proline-enriched sequences of the type P-x-x-P (x: any residue) (Mayer, 2001). The repertoire of proteins in which these domains are found as well as that of sequence and structural motifs recognized by these domains appear to be substantially broader than initially expected (Li, 2005). The recent discovery of ubiquitin as a target of SH3 domains thus not only adds to the diversity but also brings two key players in cellular signaling into the same picture (Stamenova et al., 2007). Monoubiquitination of target proteins serves to signal internalization and sorting through the endocytic pathway and gene activation at the transcriptional level (Hicke, 2001; Hicke and Dunn, 2003). More than a dozen distinct ubiquitin-binding motifs have been described that function in the direct recognition of the ubiquitin signal (Hurley et al., 2006), but none of these motifs shares any overt similarity to SH3 domains. Similarly, ubiquitin does not contain any of the known SH3 binding helical motifs (Li, 2005). To gain insights into the mode of ubiquitin binding by SH3 domains, we determined the solution structure of the third SH3 domain of the yeast Sla1 protein (henceforth designated Sla1 SH3-3) in complex with monoubiquitin. The Sla1 protein is localized at the cell cortex and is thought to couple actin dynamics with endocytosis by functioning as an adaptor interacting with multiple components of the actin cytoskeleton and the endocytic machinery besides engaging specific signals within transmembrane proteins (Ayscough, 2005). Sla1 orthologs exist in a broad range of species and the mammalian ortholog CIN85, which retains the ubiquitin binding function, also functions as an adaptor and is specifically implicated in coordinating the internalization and

destruction of activated receptor tyrosine kinases (Dikic, 2002). The results of my studies described below have been published (He et al., 2007).

A1.2 Materials and Methods

Production of the Sla1 SH3-3 domain and ubiquitin

The *Saccharomyces cerevisiae* Sla1 SH3-3 domain corresponding to residues 350–420 was expressed as a recombinant protein and purified as described previously (Stamenova et al., 2007). The identity of the protein was confirmed by electrospray ionization mass spectrometry. Sla1 SH3-3 protein samples uniformly labeled with ^{15}N and/or ^{13}C isotopes were produced using the same procedure except that cells were grown in M9 minimal medium containing ^{15}N-ammonium sulfate and/or ^{13}C-D-glucose (Spectra Stable Isotopes, Columbia, MD) respectively. Unlabeled and uniformly ^{15}N- and $^{15}N,^{13}C$-labeled samples of *Saccharomyces cerevisiae* ubiquitin were produced as described previously (Beal et al., 1996; Kang et al., 2003).

NMR samples

All NMR samples were prepared in 20 mM sodium phosphate buffer (pH 6), 2 mM DTT-d_{10}, and 0.2% (w/v) NaN_3. An equimolar complex of $^{15}N,^{13}C$-labeled Sla1 SH3-3 and unlabeled ubiquitin was generated by titrating Sla1 SH3-3 with ubiquitin until a 1:1 ratio was attained. The progress of the titration was monitored by recording one-dimensional (1D) ^{1}H and two-dimensional (2D) ^{1}H-^{15}N correlated spectra. An equimolar $^{15}N,^{13}C$-labeled ubiquitin-unlabeled Sla1 SH3-3 complex was generated in a similar manner. Protein concentrations were measured spectrophotometrically (Gill and von Hippel, 1989) and the concentrations of the NMR samples were ~0.9 mM.

NMR spectroscopy and structure determination

NMR data were acquired on a Varian Inova 600 MHz spectrometer at 45 ºC. NMR data processing and analysis were performed using an in-house modified version of Felix 98.0 (Accelrys) and NMRView (Radhakrishnan et al., 1999; Johnson, 2004). Backbone and side chain ^{1}H, ^{15}N, and ^{13}C resonances for the Sla1 SH3-3 domain and ubiquitin were assigned by analyzing three-dimensional (3D) HNCACB, C(CO)NH-TOCSY, HNCO, H(CCO)NH-TOCSY, and HCCH-COSY spectra (Grzesiek and Bax, 1993; Ferentz and Wagner, 2000). Aromatic proton resonances were assigned from 2D ^{15}N,^{13}C-double-half-filtered TOCSY spectra and 3D HCCH-COSY and HCCH-TOCSY spectra (Otting and Wuthrich, 1990; Lohr and Ruterjans, 1996).

For structure determination, backbone ϕ and ψ torsion angle restraints were derived from an analysis of H^{α}, C^{α}, C^{β}, C' and backbone ^{15}N chemical shifts using TALOS (Cornilescu et al., 1999). Restraints were imposed only for those residues that exhibited TALOS reliability scores of 9. NOE-based distance restraints for each component of the binary complex were derived from 3D ^{15}N-edited NOESY (mixing time, τ_m = 80 ms) and 3D ^{13}C-filtered, ^{13}C-edited NOESY (τ_m = 120 ms) spectra recorded in H_2O, 3D aliphatic ^{13}C-edited NOESY (τ_m = 65 ms) and 2D ^{15}N,^{13}C-double-half-filtered NOESY (τ_m = 60 ms) spectra recorded in D_2O. Intermolecular NOEs were assigned manually and given upper bounds of 6 Å. All other NOEs were calibrated and assigned iteratively and automatically by ARIA (version 1.2) (Linge et al., 2003, 2004) and were checked manually between successive rounds of refinement. Structures were calculated using ARIA in conjunction with CNS (Brunger et al., 1998). Structures were calculated from extended conformations as starting models. A total of 80 structures were computed in the final

iteration, 40 of which were refined in the presence of a shell of explicit water solvent, and the 20 structures with the lowest restraint energies, restraint violations, and RMS deviations from ideal covalent geometry were selected for structural analysis. The final structures were analyzed using PROCHECK, MONSTER, and CNS (Laskowski et al., 1996; Brunger et al., 1998; Salerno et al., 2004). Molecular images were generated using CHIMERA (Pettersen et al., 2004).

Accession code

The RCSB PDB accession code for the coordinates and the restraint tables used for the calculation of the Sla1 SH3-3–ubiquitin complex is 2JT4.

A1.3 Results and Discussion

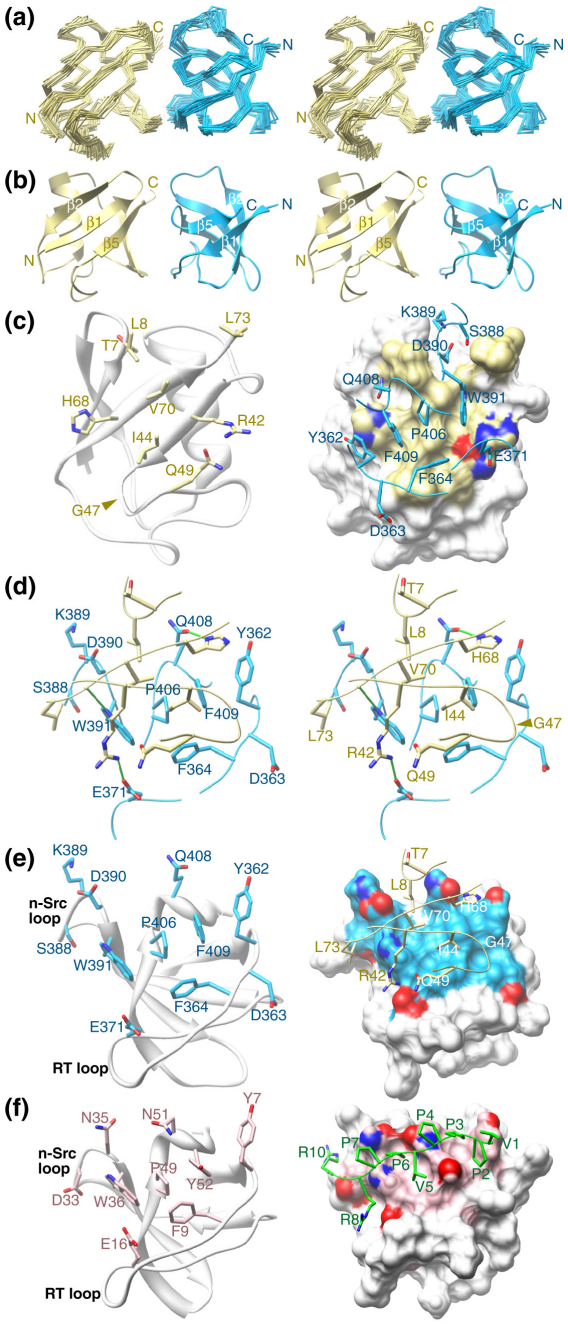
We had previously shown using solution NMR spectroscopy that like other ubiquitin-binding motifs Sla1 SH3-3 could bind ubiquitin specifically but with modest affinity (K_d ~40 μM) (Stamenova et al., 2007). To alleviate exchange-broadening effects associated with a subset of resonances at the protein-protein interface, the NMR structural studies were conducted at a slightly elevated temperature (45 °C; K_d ~400 μM). Essentially all the resonances belonging to each protein in the complex were assigned using standard through-bond approaches and structures were determined using a combination of chemical shift-based torsion angle and nuclear Overhauser effect (NOE)-based distance restraints (Table A1.1). Structures of reasonable quality and precision and in agreement with input experimental data were obtained (Table 1 & Figure 1(a)).

Table A1.1 NMR Structure Determination Statistics for the Sla1 SH3-3-Ubiquitin Complex

Restraint Statistics	
NOE-based distance restraints	
Unambiguous NOE-based restraints	2817
Intraresidue	1277
Sequential ($\| i - j \| = 1$)	445
Medium-range ($1 < \| i - j \| \leq 4$)	211
Intramolecular long-range ($\| i - j \| > 4$)	756
Intermolecular	128
Ambiguous NOE-based restraints	589
Intramolecular NOE-based restraints	1503 SH3, 1775 ubiquitin
Hydrogen bonding distance restraints	70
Torsion angle restraints	166 (83 ϕ, 83 ψ)
Structure Quality of NMR Ensemble	
Restraint satisfaction	
Root-mean-square differences for distance restraints	0.012 ± 0.001 Å
Root-mean-square differences for torsion angle restraints	0.124° ± 0.038°
Deviations from ideal covalent geometry	
Bond lengths	0.003 ± 0.000 Å
Bond angles	0.453° ± 0.011°
Impropers	1.340° ± 0.087°
Ramachandran plot statistics	
Residues in most favored regions	85.0%
Residues in allowed regions	14.4%
Residues in disallowed regions	0.6%
Average Atomic Root-Mean-Square Deviations from the Average Structure	
All atoms	2.46 Å
All atoms except disordered regions[a]	1.39 Å
All atoms in secondary structural elements[b]	1.16 Å
Backbone atoms (N, C^{α}, C')	
All residues	1.96 Å
All residues excluding disordered regions[a]	0.77 Å
All residues in secondary structural elements[b]	0.59 Å

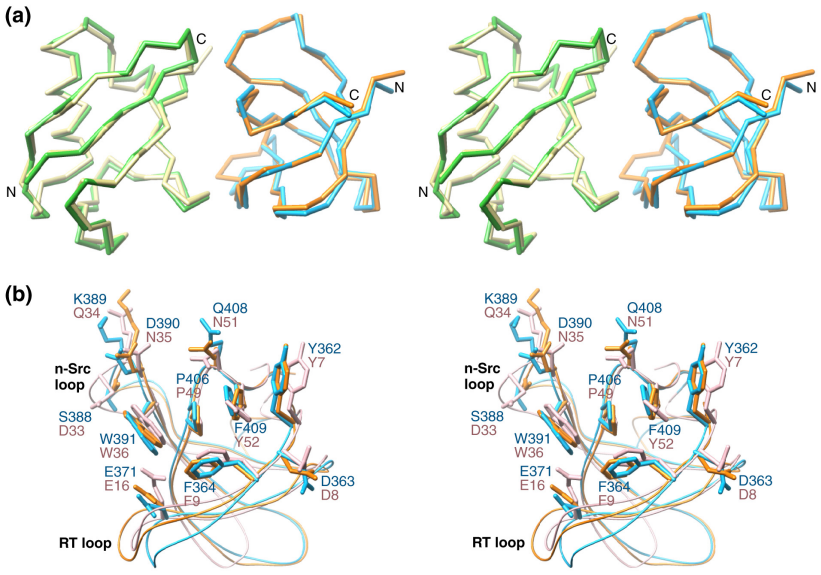
[a]: disordered regions include residues 350–354 and 414–420 of Sla1 and 74–76 of ubiquitin
[b]: secondary structural elements include residues 356–360, 379–384, 391–396, 402–406, and 410–412 of Sla1 and 2–7, 12–16, 23–34, 41–45, 48–50, and 66–71 of ubiquitin

Figure A1.1 **The Sla1 SH3-3 domain binds ubiquitin through a similar surface used for interactions with helical motifs.** (a) Stereo views of the C^{α} trace following a best-fit superposition of backbone atoms in well-ordered regions corresponding to residues 355–413 of Sla1 and 1–73 of ubiquitin in the ensemble. The Sla1 SH3-3 domain is colored blue and ubiquitin is shown in yellow. The disordered regions corresponding to residues 350–354 and 414–420 of Sla1 and 74–76 of ubiquitin have been omitted for clarity. (b) Stereo views of the representative structure from the NMR ensemble of the Sla1 SH3-3–ubiquitin complex with the same coloring scheme as in panel (a). (c) Structure of ubiquitin in the complex shown in a cartoon representation along with the side chains deemed to be interacting with Sla1 SH3-3 (*left panel*). A view of the molecular surface of ubiquitin shown along with the ubiquitin-interacting side chains and the corresponding backbone segments of Sla1 SH3-3 (*right panel*). Oxygen and nitrogen atoms are shown in red and dark blue, respectively. (d) Stereo views of the Sla1 SH3-3–ubiquitin interface depicting the side chains of the interacting residues and the corresponding backbone segments. Intermolecular hydrogen bonding interactions are depicted as solid green lines. (e) Structure of Sla1 SH3-3 in the complex shown in a cartoon representation along with the ubiquitin interacting side chains (*left panel*). A view of the molecular surface of the SH3 domain shown along with the SH3-interacting side chains and the corresponding backbone segments of ubiquitin (*right panel*). (f) Structure of the Grb2 SH3-SOS peptide complex (PDB code: 1GBR (Wittekind et al., 1997)) with the SH3 domain presented in the same orientation as Sla1 SH3-3 in panel (e). The side chains of Grb2 deemed to be interacting with the proline-rich peptide are shown in light pink; oxygen and nitrogen atoms are shown in red and dark blue, respectively. The peptide backbone and side chains are colored in green (*right*).


(a)
C
C
N
N
C
C
N
N
(b)
β2
β1
β5
C
N
β2
β1
β5
C
N
β2
β1
β5
C
N
β2
β1
β5
C
N
(c)
T7
L8
L73
H68
V70
R42
I44
Q49
G47
K389
S388
D390
Q408
W391
P406
Y362
F409
E371
F364
D363
(d)
K389
D390
Q408
Y362
S388
P406
W391
F409
F364
D363
E371
T7
L8
H68
V70
I44
G47
L73
R42
Q49
(e)
K389
D390
Q408
Y362
n-Src loop
S388
P406
W391
F409
F364
D363
E371
RT loop
T7
L8
H68
V70
L73
I44
G47
R42
Q49
(f)
N35
N51
Y7
n-Src loop
P49
Y52
D33
W36
F9
E16
RT loop
P4
P3
V1
R10
P7
P6
P2
V5
R8

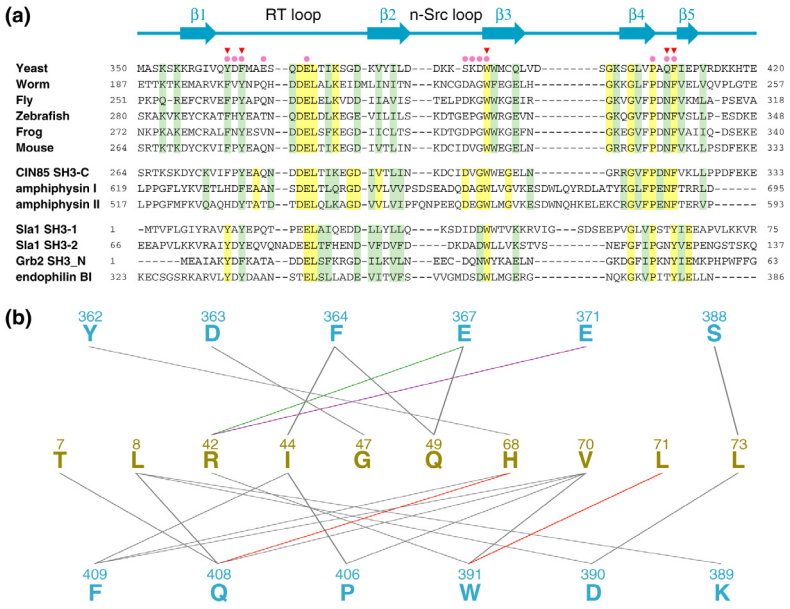
Both ubiquitin and the Sla1 SH3 domain adopt the well-characterized eponymous folds in the complex (Figure 1(b)). The five β-strands of Sla1 SH3-3 span residues 356–360, 379–384, 391–396, 402–406 and 410–412; a 3_{10}-helix separates strands β4 and β5. The polypeptide backbones of the two proteins in the complex are remarkably similar to those of the free forms with atomic root-mean-square deviations of 0.87 Å for ubiquitin (Vijay-Kumar et al., 1987) and 0.88 Å for the SH3 domain (Figure 2(a); PDB codes 1UBQ and 1Z9Z, respectively). Similarly, only small conformational rearrangements upon complex formation are noted for the SH3 side chains (Figure 2(b)). Interestingly, two molecules are found in the asymmetric unit in the crystal structure of apo Sla1 SH3-3 with the molecules participating in homotypic interactions via the ubiquitin binding surface (PDB code: 1Z9Z). However, we find no evidence for homodimerization by NMR as the Sla1 SH3-3 domain spectra are characterized by narrow resonances that broaden somewhat upon complex formation with ubiquitin due to the increased molecular size of the complex (data not shown).

Figure A1.2 **Economy in conformational rearrangements upon complex formation between Sla1 SH3-3 and ubiquitin and basis for specificity of SH3-ubiquitin interactions.** (a) Stereo views of the C^{α} traces of Sla1 SH3-3 and ubiquitin following a best-fit superposition of backbone atoms corresponding to residues 357–415 of Sla1 and 1–73 of ubiquitin in the NMR structure of the complex with the corresponding segments in the crystal structures of the apo forms. The Sla1 SH3-3 domain is colored blue (complex) and orange (apo; PDB code: 1Z9Z) whereas ubiquitin is shown in yellow (complex) and green (apo; PDB code: 1UBQ (Vijay-Kumar et al., 1987)). (b) Stereo views of the ubiquitin-binding residues of Sla1 SH3-3 in the apo (orange) and ubiquitin-bound (blue) states and comparison with the equivalent residues in the Grb2 SH3 domain (light pink) (Wittekind et al., 1997). The side chains are shown in stick representation whereas the backbones are shown as worms. Residues and important structural landmarks are annotated. Notice also the relatively modest changes in side chain conformation for Sla1 SH3-3 upon complex formation with ubiquitin.



The Sla1 SH3-3–ubiquitin complex interface extends over an area averaging 470 $Å^2$ (± 36 $Å^2$) in the NMR ensemble. Like many other ubiquitin-binding motifs (Hurley et al., 2006), the SH3 domain targets a primarily hydrophobic ubiquitin surface comprising Leu8, Ile44, Val70, and Leu73 surrounded by an array of polar residues (Figure 1(c) & 1(d)). The complementary surface on the SH3 domain is dominated by hydrophobic residues including Tyr362, Phe364, Trp391, Pro406, and Phe409 that form a relatively smooth, shallow groove (Figure 1(e)). The complex appears to be stabilized by intermolecular hydrogen bonding interactions between the side chains of His68 and Gln408, Arg42 and Glu371, and between the backbone carbonyl of Leu71 and the side chain of Trp391 and by electrostatic interactions involving Arg42 of ubiquitin and Glu367 and Glu371 of Sla1 SH3-3 (Figure 1(d) & Figure 3 (a)). We note that these non-bonding interactions were consistently detected in over 60% of the conformers in the NMR ensemble and were inferred from the structures resulting from the structure determination protocol rather than being explicitly entered as experimentally-derived restraints. These interactions are consistent with previous mutagenesis studies that implicated a subset of these residues in stabilizing the complex including Ile44 and Val70 of ubiquitin and Tyr362, Phe364, Trp391, Gln408, and Phe409 of Sla1 SH3-3 as mutations of any of these residues to alanine were found to severely diminish the SH3-ubiquitin interaction (Figure 3) (Stamenova et al., 2007).

Figure A1.3 **Sequence conservation and non-covalent interactions mediated by residues of SH3 domains with ubiquitin-binding activity.** (a) CLUSTAL W-guided multiple sequence alignment of a selection of SH3 domains. The proteins were grouped into three functional classes including Sla1 orthologs (*top*), proteins with demonstrable ubiquitin-binding activity (*middle*) and those lacking this activity (*bottom*) (Stamenova et al., 2007). Conserved and invariant residues in each group are shaded green and yellow, respectively. The solid pink circles and red inverted triangles denote Sla1 SH3-3 residues deemed to contact ubiquitin in the NMR structure and those residues that when mutated to alanine diminish monoubiquitin binding (Stamenova et al., 2007), respectively. A cartoon depicting the location of secondary structural elements in Sla1 SH3-3 is shown on top. (b) A catalog of non-covalent intermolecular interactions in the Sla1 SH3-3-ubiquitin complex detected in ≥60% of conformers in the NMR ensemble (Salerno et al., 2004). Sla1 SH3-3 and ubiquitin residues are colored in blue and yellow, respectively. The lines connect interacting residues. Line colors indicate the type of interaction (green: electrostatic; red: electrostatic; purple: salt bridge; gray: hydrophobic).



The amino-terminal Grb2 SH3 domain represents the closest homologue of Sla1 SH3-3 (37% sequence identity) for which a high-resolution structure has been described (Wittekind et al., 1997). A comparison of the structures of the Sla1 SH3-3–ubiquitin complex with that of the Grb2 SH3 domain in complex with a proline-rich sequence reveals substantial overlap between the ligand-binding surfaces of the two SH3 domains (Figure 1(f)). The vast majority of the ligand-binding residues within the two domains are either identical or conserved and this feature is shared with other SH3 domains whose structures in complex with proline-containing peptides have been described (Li, 2005). Unlike Sla1 SH3-3, the Grb2 SH3 domain does not exhibit discernible ubiquitin-binding activity (Stamenova et al., 2007), so what attribute(s) confers a particular SH3 domain with the ability to bind ubiquitin? Intriguingly, the side chain conformations of the ligand-binding residues are similar for the Grb2 and Sla1 SH3 domains (Figure 2(b)), implying that the functional differences are unlikely to originate from distinct conformational preferences. A key affinity and specificity determinant for ubiquitin-binding was previously postulated to reside with a phenylalanine residue (Phe409 of Sla1) (Stamenova et al., 2007). This residue is located at the heart of the hydrophobic interface and is shielded from the solvent in the SH3-ubiquitin complex (Figure 1(d)). The Grb2 SH3 domain harbors a tyrosine residue at the equivalent position (Figures 1(e), 1(f) and 2(b)) and mutation of Sla1 Phe409 to tyrosine was found to abrogate ubiquitin binding (Stamenova et al., 2007). The most likely explanation based on the structure of the Sla1 SH3-3-ubiquitin complex is that the introduction of a hydroxyl group results in unfavorable steric clashes between this moiety and Ile44 of ubiquitin. This is probably exacerbated by the absence of suitable hydrogen bonding donor and acceptor groups near the hydroxyl group. Predictably, a phenylalanine rather than a tyrosine is consistently detected at this position in the third SH3 domain of Sla1 orthologs including

mammalian CIN85 and in other SH3 domains possessing ubiquitin-binding activity (Figure 3(a)). Conversely, a tyrosine residue is consistently detected in SH3 domains lacking ubiquitin-binding activity. Interestingly, the phenylalanine does not preclude Sla1 SH3-3 from binding proline-rich ligands (Stamenova et al., 2007), consistent with previous structural analyses of other SH3 domains. The Sla1 SH3 domain thus provides a striking example of how disparate functions can be accommodated within a common interaction surface through subtle changes in sequence.

Although the phenylalanine residue is an important affinity and specificity determinant for ubiquitin-binding, it is likely that the full or near-full complement of non-covalent interactions described for the Sla1 SH3-3-ubiquitin complex is essential for efficient binding. Indeed, mutating Tyr64 to a phenylalanine in the Sla1 SH3-1 domain failed to produce a protein competent to bind ubiquitin (Stamenova et al., 2007). One explanation for this observation could be the lack of a Sla1 SH3-1 residue equivalent to Gln408 capable of hydrogen bonding interactions (Figure 3(b)). Similar considerations could apply to other phenylalanine-harboring SH3 domains that might fail to bind ubiquitin.

The pervasiveness of ubiquitin-binding motifs in endocytic proteins underscores the importance of this signal in endocytosis. It is intriguing that multiple endocytic proteins harbor SH3 domains with ubiquitin-binding activity while simultaneously retaining their ability to bind proline-rich ligands, suggesting a dynamic interplay between these two processes that warrants further investigation. Finally, the availability of detailed molecular pictures of all the key domains for the Sla1/CIN85 protein in complex with their physiologically relevant targets sets the stage for a comprehensive functional analyses of this evolutionarily conserved protein. Intriguingly, all four domains including the three SH3 domains and the SHD1 domain unique to

Sla1 share structural similarity but exhibit considerable diversity in the way they engage their targets, that are themselves structurally varied (Jozic et al., 2005; Mahadev et al., 2007). This could be a general theme for adaptor proteins involved in intracellular signaling.

APPENDIX TWO

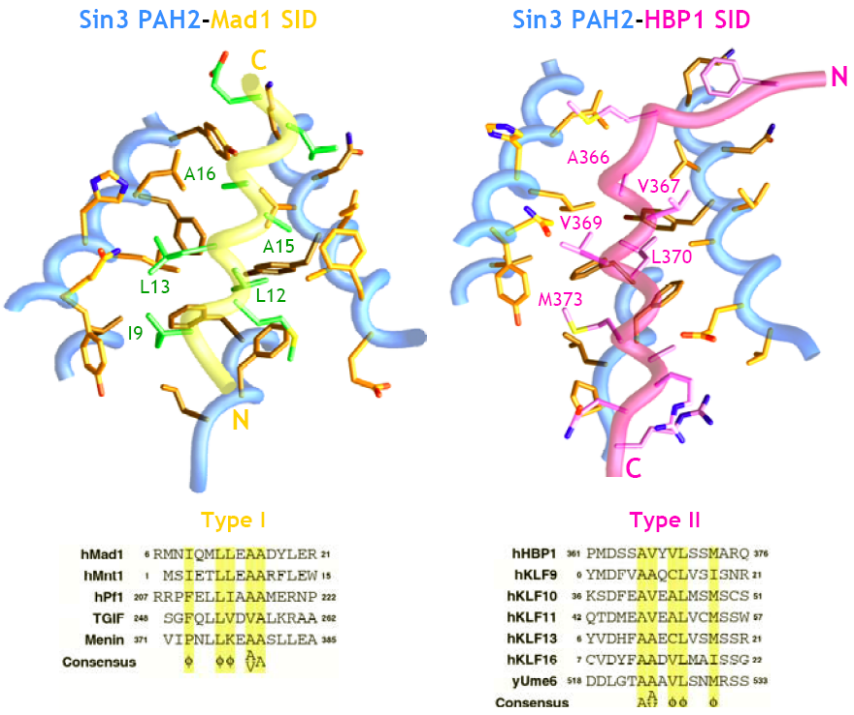
Test of the Generality of the Two Distinct Modes of mSin3A PAH2 Interaction with the SID Domains of Transcriptional Repressors TGIF and KLF11

A2.1 Introduction

The recruitment of transcriptional coregulators possessing intrinsic chromatin-modifying and/or chromatin-remodeling activities by diverse promoter-bound transcription factors is a common, albeit important, step in the eukaryotic gene transcription regulation (Ptashne and Gann, 2002). The past decade has brought us the identification of a series of chromatin-modifying and chromatin-remodeling enzymes, which are frequently found in large, multi-subunit coregulator complexes that lack intrinsic DNA-binding activity (Elgin and Workman, 2000). The Sin3 complex is one of the first corepressor complexes to be identified and is found in organisms ranging from yeast to human, whose negative effects on gene transcription is exerted through the action of histone deacetylases (HDACs; (Hassig et al., 1997; Laherty et al., 1997; Laherty et al., 1998; Zhang et al., 1998; Alland et al., 2002; Fleischer et al., 2003)). The Sin3 corepressor has a dual role functioning not only as a molecular scaffold for complex assembly but also as a molecular adapter bridging HDACs with an impressive array of DNA-binding transcription factors and chromatin-binding proteins (Ayer, 1999; Knoepfler and Eisenman, 1999; Silverstein and Ekwall, 2005). The interactions are generally mediated through one or more of six discrete conserved regions including four imperfect copies of the paired amphipathic helix (PAH) domain. The N-terminal portions of the Sin3 proteins including the first two PAH domains serve as the main 'docking' sites for its recruitment by the transcription factors, and the second copy of the PAH domain (PAH2) is the site for interactions with numerous factors including members of the Mad family (McArthur et al., 1998; Schreiber-Agus and DePinho, 1998), Mnt/Rox (Hurlin et al., 1997), Pf1 (Yochum and Ayer, 2001), HBP1 (Swanson et al., 2004), the KLF family (Zhang et al., 2001), TGIF (Wotton et al., 2001), Menin (Kim et al., 2003), MNF-β (Yang et al., 2000), and EKLF (Chen and Bieker, 2004).

At the molecular level, mSin3A PAH2 domain has been shown to adopt a left-handed four-helix bundle architecture in the Mad1 SID-mSin3A PAH2 (Brubaker et al., 2000) and HBP1 SID-mSin3A PAH2 (Swanson et al., 2004) complexes. These studies also revealed that two distinct types of sequence motifs in the SIDs of Mad1 and HBP1 bind to the same hydrophobic cleft of PAH2 domain through their non-polar side of the amphipathic helix. Moreover, the Mad1 SID (type I) adopts a reversed orientation relative to the HBP1 SID (type II) in the interactions with mSin3A PAH2 domain (Figure A2.1). Although this deduced designation combining both sequence similarities and secondary structure prediction results successfully guided the mapping of the precise location of the SIDs within a number of PAH2-associated factors, some constraints used to define the respective sequence motifs needed to be relaxed to accommodate variations when additional biochemically-characterized PAH2 interactors were identified (Swanson et al., 2004). In order to improve the accuracy of these predictions and test the generality of the classification, we have initiated the structural analyses of PAH2 interactions with one member of each type of SID (e.g. TG-interacting factor (TGIF) SID, type I; Krüppel like factor (KLF)11 SID, type II).

Figure A2.1 **Comparison of the complex structures of the Mad1 SID-mSin3A PAH2 (*left*) and HBP1 SID-mSin3A PAH2 (*right*).** Both SIDs adopting an amphipathic helical structure bind to the same hydrophobic cleft on mSin3A PAH2, but with reversed orientation relative to each other (*top*). Side-chains of residues involved in the interactions are shown in the representative structures and conserved SID residues based on structure-guided sequence analyses (*bottom*) are highlighted. (Modified from (Swanson et al., 2004))



A2.3 Materials and Methods

Production of mSin3A PAH2, TGIF SID and KLF11 SID

The mSin3A PAH2 domain corresponding to residues 295-383 was expressed as a recombinant protein and purified as described previously (Brubaker et al., 2000). PAH2 samples uniformly labeled with ^{15}N isotope was produced using the same procedure except that cells were grown in M9 minimal medium containing ^{15}N-ammonium sulfate (Spectra Stable Isotopes, Columbia, MD). The identity of the protein was confirmed by electrospray ionization–mass spectrometry (ESI-MS). A 16-residue peptide of TGIF SID (residues 248-263) and a 14-residue peptide of KLF11 SID (residues 43-56) were synthesized using automated procedures (University of Utah DNA/Peptide Facility) and purified to homogeneity using reversed-phase HPLC. The identity of the peptides were also confirmed by ESI-MS.

NMR samples and spectroscopy

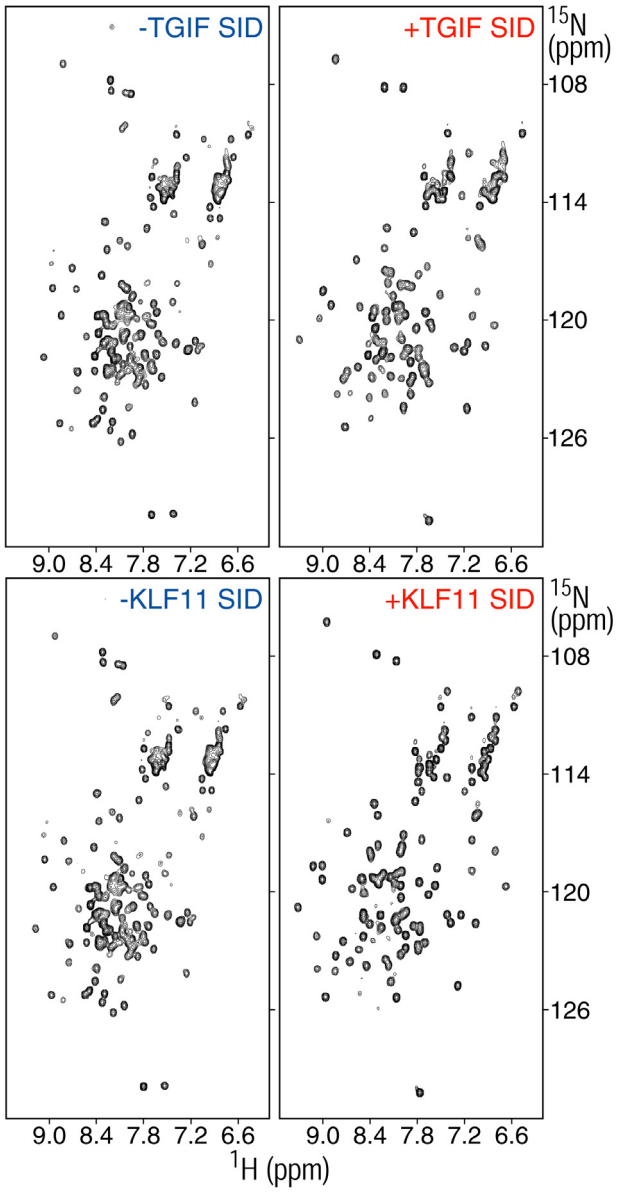
All NMR samples were prepared in 20 mM sodium phosphate buffer (pH 6), 2 mM DTT-d_{10}, and 0.2% (w/v) NaN_3. Complex of ^{15}N-labeled mSin3A PAH2 and either peptide was generated by dissolving ^{15}N-labeled PAH2 and excess of peptide in 20 mM sodium phosphate buffer (pH 6), 2 mM DTT, 8 M urea, followed by dialysis against 20 mM sodium phosphate buffer (pH 6), 2 mM DTT using Spectra/por DispoDialyzer with 1,000 Da cutoff (Spectrum Laboratories Inc., Rancho Dominguez, CA) and buffer exchange into 20 mM sodium phosphate buffer (pH 6), 2 mM DTT-d_{10}, and 0.2% (w/v) NaN_3 using Centricon with 3,000 Da cutoff (YM-3, Millipore Corporation). Protein concentrations were measured spectrophotometrically (Gill and von Hippel, 1989) and the concentrations of the NMR samples were 0.2 mM and 0.1 mM for ^{15}N-mSin3A PAH2-TGIF SID and -KLF11 SID, resepctively.

The complex formations were confirmed by recording one-dimensional (1D) ^{1}H and two-dimensional (2D) ^{1}H-^{15}N correlated spectra. NMR data were acquired on a Varian Inova 600 MHz spectrometer at 15 ºC. NMR data processing and analysis were performed using an in-house modified version of Felix 98.0 (Accelrys) (Radhakrishnan et al., 1999).

A2.3 Results and Discussion

TGIF functions as a transcriptional repressor by recruiting the Sin3 corepressor complex to the Smad complex for antagonizing transforming growth factorbeta (TGFβ)-mediated stimulation (Sharma and Sun, 2001; Wotton et al., 2001), and it has been biochemically shown to directly interact with the PAH2 domain of Sin3 through its C-terminal SID. The type I multiple sequence alignment identifies a similar conservation pattern between the Mad1 SID and TGIF SID despite of a conserved substitution from an alanine to valine at position 7 (Figure A2.1). To test whether this prediction is correct and TGIF SID is indeed a *bona fide* Sin3 interaction domain, we titrated ^{15}N-labeled mSin3A PAH2 using a purified 16-residue peptide corresponding to the TGIF SID sequence (residues 248-263). The ^{1}H-^{15}N HSQC spectrum of the apo-mSin3A PAH2 domain is characterized by at least two sets of resonances indicative of conformational heterogeneity (Figure A2.2). However, like the generation of other SID-mSin3A PAH2 complexes, the heterogeneity disappears upon the addition of the peptide and many of the PAH2 resonances shift to new positions, indicative of a specific interaction between PAH2 and the peptide with reasonable affinity based on slow exchange kinetics of the complex (Figure A2.2). The spectrum of bound PAH2 is of high quality with 79 of the expected 83 backbone correlations readily detected.

Figure A2.2 **^{1}H-^{15}N correlated HSQC spectra of mSin3A PAH2 in the absence and presence of TGIF (*top panels*) and KLF11 (*bottom panels*) SID peptides at 15 ºC in 20 mM sodium phosphate buffer, pH6.** The spectra in the top panels were recorded at 0.1 mM ^{15}N-PAH2 concentration whereas those in the bottom panels were recorded at 0.2 mM protein concentration. Identical data acquisition, processing and display parameters were used.



KLF protein family functions as transcription activators and/or repressors depending on the associated promoters. All the family members typically comprise a highly conserved C-terminal region harboring three Cys_2His_2-type zinc finger motifs for DNA-binding in addition to the N-terminal trans-activation or -repression domain, which varies between different family members (Dang et al., 2000; Kaczynski et al., 2003). KLF11 is one of the better-characterized factors in this protein family and is also involved in TGFβ-mediated signaling by Sin3-dependent repression of Smad7, which in turn positively regulates the pathway (Ellenrieder et al., 2004). The precisely mapped SID within KLF11 is also homologously present in other KLF family members (Figure A2.1). A computation-based model of KLF11 SID-mSin3A PAH2 complex structure gave rise to a contrasting result compared to our structure-based prediction (Pang et al., 2003). In order to confirm our proposed classification and improve our ability to make reliable prediction in the future, we have titrated ^{15}N-labeled mSin3A PAH2 using synthesized KLF11 SID peptide (residues 43-56). A similar change of the HSQC spectrum in terms of disappearance of heterogeneity and shift of a subset of resonances is noticed, indicative again of a specific interaction, however with lower affinity based on relatively fast exchange kinetics noticed for the complex (Figure A2.2). The spectrum is of high quality with 80 of the expected 83 backbone correlations readily detectable. Studies of detailed mode for both SID-mSin3A PAH2 interactions are presently underway.

APPENDIX THREE

Reconstitution of the Core Sin3 Corepressor Complex

A3.1 Introduction

The Sin3 complex is one of the first corepressor complexes to be identified and found in organisms ranging from yeast to human. Besides the scaffold protein Sin3A/B and the enzymatic HDAC1/HDAC2/RbAp46/RbAp48 sub-complex, another four Sin3-associated polypeptides have been established as the key constituents of the Sin3 complex in mammalian cells, including SAP30, SAP45/Sds3, SAP130, and SAP180/BCAA (Figure 1.5) (Bernstein et al., 2000; Alland et al., 2002; David et al., 2003; Fleischer et al., 2003; Mallory and Strich, 2003; Carrozza et al., 2005a). Although the PAH3 and HID domains of Sin3 are thought to harbor all the necessary determinants for interactions with the aforementioned proteins, it is currently unclear how they are assembled to form the complex and what the functional role of each SAPs is besides stabilizing the core repressor. Moreover, there is no structural information about these proteins except for the Sin3 PAH domains.

Substantial efforts have been made into pairwise co-expression of numerous HID constructs with each SAPs by previous group members due to the fact that these proteins have been shown to interact with Sin3 through the intact HID region (Figure 1.3), but progress with these studies have been impeded by challenges associated with expression, purification, and stability issues. Therefore, it is very likely that these core components of the complex are mutually dependent on each other like in the case of SAP30 SID-mSin3A PAH3 complex generation, and multiple partners may be required for proper folding of the Sin3 HID domain. Since similar challenges have been met by several studies of other multi-protein complexes, and co-expression of multiple subunits has often improved the outcome (Tan, 2001; Tan et al., 2005; Fitzgerald et al., 2006; Tolia and Joshua-Tor, 2006). However, co-expression strategies vary and

we decided to use the polycistronic expression strategy developed by Song Tan and co-workers (Tan, 2001; Tan et al., 2005).

In this strategy, fragments encoding the Sin3 interaction domains of SAP30, SAP130, and SAP180 were introduced along with that of the mSin3A PAH3/HID region in a pET-based expression vector for expression in *E.coli* host. Since the expression system was designed to express only four polypeptides, the fifth subunit Sds3/SAP45 was expressed by co-transforming the bacteria using an appropriate vector.

A3.2 Materials and Methods

Constructs design

The coding sequence of SAP30 ΔN (residues 64-200), SAP130 SID (residues 846-1047), mSin3A PAH3/HID (residues 456-830) with a N-terminal His_6-tag, and SAP180 SID (residues 957-1312) were amplified by PCR and inserted into the monocistronic expression vector pET3aTr individually (Tan, 2001). The pST39-SAP30-SAP130-PAH/HID-SAP180 co-expression vector was then constructed by subcloning the *Xba*I-*Bam*HI SAP30, *Eco*RI-*Hind*III SAP130, *Sac*I-*Kpn*I His6-mSin3A PAH3/HID, and *Bsp*EI-*Mlu*I SAP180 cassettes sequentially from the corresponding pET3aTr vectors into pST39. The internal *Hind*III site of SAP30 and the internal *Kpn*I site of PAH3/HID were removed in a translationally silent manner by QuikChange site-directed mutagenesis (Stratagene). Two stop codon 'TAATAA' was introduced into the co-expression vector right after residue number 680 of mSin3A PAH3/HID to make the truncation mutant of mSin3A. The coding sequence of full-length SAP45/Sds3 (residues 1-328) was amplified by PCR and inserted into the pMCSG23 expression vector (Stols et al., 2002). The

coding regions of all expression constructs of pET3aTr, pST39, and pMCSG23 were confirmed by DNA sequencing.

Protein expression and purification

Escherichia coli BL21(DE3) cells (Novagen, Madison, WI) containing the vector were grown at 37 °C in LB broth (EMD Chemicals Inc.). The growth temperature was either maintained at 37 °C or shifted to 20 °C depending on different expression temperatures when the $OD_{600\ nm}$ reached approximately 0.7. Expression of protein in isolation or co-expression using polycistronic expression vectors was induced using 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and the cells were harvested 4 h (37 °C) or 16 h (20 °C) thereafter.

Cell pellet of co-expression of pMCSG23-SAP45 and pST39-SAP30-SAP130-PAH/HID-SAP180 was suspended in 50 mM Tris-HCl buffer (pH 8.6) containing 0.3 M sodium chloride, 5 mM Tris (2-carboxy-ethyl) phosphine hydrochloride (TCEP), 1 mM phenyl-methylsulfonyl (PMSF), 1 μM leupeptin, 1 mM pepstatin, and 0.1% Triton X-100, lysed via lysozyme for 30 minutes and DNAse treatment for 15 minutes at 4 °C followed by centrifugation (14,000 rpm for 15 minutes at 4 °C). Pellet was further solubilized for SDS-PAGE using the buffer above containing 8 M urea.

Cell pellet of expression of pST39-SAP30-SAP130-PAH/HID-SAP180 and pST39-SAP30-SAP130-PAH/pHID-SAP180 at 20 °C was first suspended in 50 mM Tris-HCl buffer (pH 8.0) containing 150 mM sodium chloride, 5 mM TCEP, 1 mM phenyl-methylsulfonyl (PMSF), 1 μM leupeptin, 1 mM pepstatin, and 0.1% Triton X-100, lysed via sonication and centrifuged. The pellet was then suspended in 50 mM Tris-HCl buffer (pH 8.0) containing 1 M

sodium chloride, 5 mM TCEP followed by sonication, DNAse treatment for 15 minutes at 4 °C and centrifuged. The supernatant was kept for SDS-PAGE.

The coding sequence of ING2 SAID domain (residues 19-132) was amplified by PCR and inserted into the pMCSG7 expression vector (Stols et al., 2002). *E. coli* BL21(DE3) cells (Novagen, Madison, WI) containing the vector were grown at 37 °C in LB broth (EMD Chemicals Inc.). The growth temperature was shifted to 20 °C when the $OD_{600\ nm}$ reached approximately 0.7. Expression of the His_6-tagged protein was induced using 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and the cells were harvested 16 h thereafter. Cell pellets were suspended in 20 mM sodium phosphate buffer (pH 8.0) containing 200 mM sodium chloride, 2 mM TCEP, 1 mM phenyl-methylsulfonyl (PMSF), 1 μM leupeptin, 1 mM pepstatin, and 0.1% Triton X-100, lysed via sonication and centrifuged. The pellet was then suspended in the above buffer containing 4 M urea followed by sonication and centrifuged again. The supernatant in denatured buffer was incubated with the His-Select Nickel resin (Sigma, St. Louis, MO) for 30 minutes. Bound proteins were eluted using the equilibration buffer containing 300 mM imidazole followed by incubating with tobacco etch virus (TEV) protease for 4 h at 22 °C and overnight incubation at 4 °C. The mixture was centrifuged and the desired protein in the supernatant was purified to homogeneity via reversed-phase HPLC using a C18 column (Vydac, Hesperia, CA) and a linear gradient of 0.1% trifluoroacetic acid (TFA) and 0.1% TFA in 80% acetonitrile and lyophilized.

The coding sequence of RBP1 R2 domain (residues 1167-1230) was amplified by PCR and inserted into the pMCSG7 expression vector (Stols et al., 2002). *E. coli* Codon Plus cells (Novagen, Madison, WI) containing the vector were grown at 37 °C in LB broth (EMD Chemicals Inc.). The growth temperature was shifted to 20 °C when the $OD_{600\ nm}$ reached

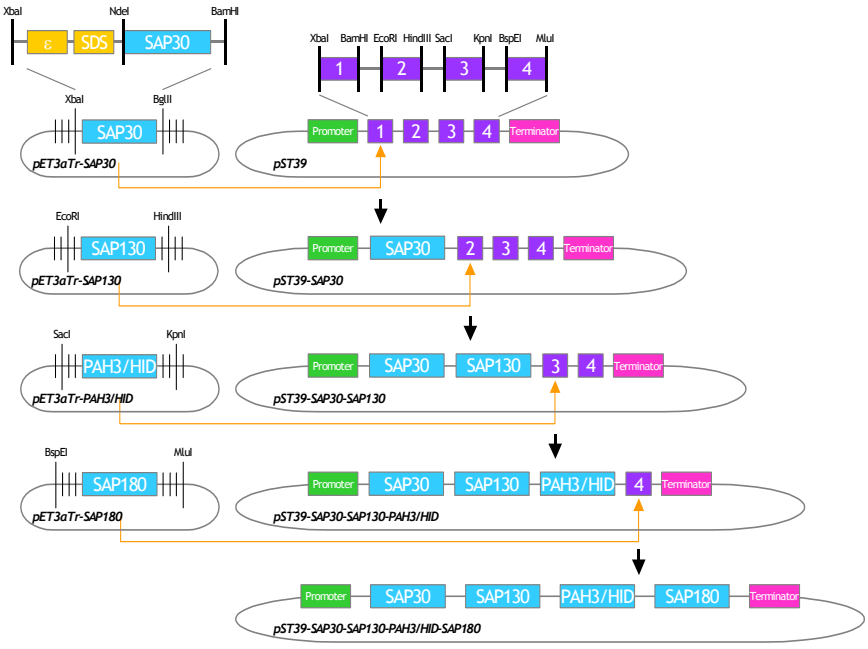
approximately 0.7. Expression of the His_6-tagged protein was induced using 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and the cells were harvested 16 h thereafter. Cell pellets were suspended in 50 mM Tris-HCl buffer (pH 8.0) containing 200 mM sodium chloride, 2 mM TCEP, 1 mM phenyl-methylsulfonyl (PMSF), 1 μM leupeptin, 1 mM pepstatin, and 0.1% Triton X-100, lysed via sonication and centrifuged. The supernatant was incubated with the His-Select Nickel resin (Sigma, St. Louis, MO) for 30 minutes. Bound proteins were eluted using the equilibration buffer containing 300 mM imidazole followed by incubating with tobacco etch virus (TEV) protease for 4 h at 22 °C and overnight incubation at 4 °C. The mixture was centrifuged and the desired protein in the supernatant was purified to homogeneity via reversed-phase HPLC using a C18 column (Vydac, Hesperia, CA) and a linear gradient of 0.1% trifluoroacetic acid (TFA) and 0.1% TFA in 80% acetonitrile and lyophilized.

A3.3 Results and Discussion

A bottom-up approach was used to incrementally reconstitute and characterize a sub-complex comprising subunits specific to the Sin3A/B corepressor complexes. Since the expression system we were using was designed to express only four polypeptides using a single vector, and we had five potential subunits to reconstitute, two separate constructs were made. In the first construct, fragments encoding SAP30 ΔN (residues 64-200), SAP130 SID (residues 846-1047), and SAP180 SID (residues 957-1312) were introduced along with that of the mSin3A PAH3/HID region (residues 456-830) with a N-terminal His_6-tag in a pET-based expression vector with the ColE1 replicon simply because these three SAPs'SID domains have been characterized and shown to target this region of mSin3A (Fleischer et al., 2003). The fifth subunit Sds3/SAP45 was then inserted into a vector carrying the CloDF13 replicon with a N-

terminal MBP-tag followed by the full-length protein. Details of cloning procedure are shown in Figure A3.1 (Rebecca Imhoff initiated the cloning procedures) and all five proteins were thereby co-expressed by co-transforming the bacteria using both expression vectors.

Figure A3.1 **A cartoon showing the cloning design for generating the polycistronic expression system used to express four genes (SAP30, SAP130 SID, His6-tagged mSin3A PAH3/HID, and BCAA/SAP180 SID) within a single expression vector.**



Expression trials of four of the five subunits conducted in isolation revealed that the proteins were expressed, albeit at different levels (Figure A3.2 (A), (B)). However, all five proteins appeared to co-express well and at comparable levels (Figure A3.2 (C)). Sds3/SAP45 appeared to express much better than others, but this may be because the protein is expressed from a different vector. The expression of SAP180 SID is difficult to evaluate as it is obscured by an *E. coli* protein. Unfortunately, with the exception of the Sds3/SAP45 protein, which is expressed with an MBP tag, none of the proteins appear to stay in solution when co-expressed during the first trial of purification. However, comparable amounts of the expressed proteins could be found in the pellet after cell lysing by sonication and centrigugation (Figure A3.2 (C)). Further optimization of the expression and purification protocols made a portion of the co-expressed proteins remaining in a soluble form. In addition, a sub-complex containing four proteins could be purified under native conditions and similar results were also shown with a truncation mutant of mSin3A spanning the PAH3 and pHID regions (residues 456-680) (Figure A3.2 (D)). Further studies of reconstituting this four-subunit complex with fragments of interest for SAP45, RBP1 and ING2 are in progress (Figure A3.2 (E), (F)).

Figure A3.2 **SDS-PAGE analysis of the expression and purification tests of Sin3 core repressor components.** Expression levels of the proteins when expressed in isolation shown in (A) and (B) at different temperatures, and in conjuction (C) were checked by SDS-PAGE followed by coomassie staining. PAH3/pHID and SAP130 migrate similarly, but can be distinguished on higher-resolution gels. Lack of co-purification for SAP180 SID conducted with mSin3A PAH3/pHID indicates that it might binds the carboxyl terminal of the HID. Arrowheads depict protein bands of interest. Abbreviations: BI: before induction; AI: after induction; S: soluble supernatant; I: inclusion-body pellet; AP: after purification.