5-2015

Binding Interactions of (R)- and (S)-hydroxypropyl-CoM Dehydrogenases and the Zinc Knuckle Proteins Air1 and Air2

Jeremy W. Bakelar
Utah State University

BINDING INTERACTIONS OF (*R*)- AND (*S*)-HYDROXYPROPYL-COM

DEHYDROGENASES AND THE ZINC KNUCKLE PROTEINS AIR1 AND AIR2

by

Jeremy W. Bakelar

A dissertation submitted in partial fulfillment

of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Biochemistry

Approved:

_____	_____
Dr. Sean J. Johnson	Dr. Lance C. Seefeldt
Major Professor	Committee Member
_____	_____
Dr. Joan M. Hevel	Dr. Gregory J. Podgorski
Committee Member	Committee Member
_____	_____
Dr. Alvan C. Hengge	Dr. Mark R. McLellan
Committee Member	Vice President for Research and
	Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2015

ABSTRACT

Binding Interactions of (*R*)- and (*S*)-hydroxypropyl-CoM Dehydrogenases

and the Zinc Knuckle Proteins Air1 and Air2

by

Jeremy W. Bakelar, Doctor of Philosophy

Utah State University, 2015

Major Professor: Dr. Sean J. Johnson
Department: Chemistry and Biochemistry

This work is focused on understanding protein function by describing how

paralogous proteins with overlapping and distinct functions interact with their substrates

and with other proteins. Two model systems are the subject of this research: (1) the

stereospecific dehydrogenases *R*- and *S*-HPCDH, and (2) the zinc knuckle proteins Air1

and Air2.

R- and *S*-HPCDH are homologous enzymes that are central to the metabolism of

propylene and epoxide in the soil bacterium *Xanthobacter autotrophicus*. The bacterium

produces *R*- and *S*-HPCDH simultaneously to facilitate transformation of *R*- and *S*-

enantiomers of epoxypropane to a common achiral product 2-ketopropyl-CoM (2-KPC).

Both *R*- and *S*-HPCDH are highly stereospecific for their respective substrates as each

enzyme displays less than 0.5% activity with the opposite substrate isomer. Presented

here are substrate-bound x-ray crystal structures of *S*-HPCDH. Comparisons to the

previously reported product-bound structure of *R*-HPCDH reveal structural differences that provide each enzyme with a distinct substrate binding pocket. These structures demonstrate how chiral discrimination by *R*- and *S*-HPCDH results from alternative binding of the distal end of substrates within each substrate binding pocket, providing a structural basis for stereospecificity displayed by *R*- and *S*-HPCDH.

Air1 and Air2 are homologous eukaryotic proteins that individually function within a trimeric protein complex called TRAMP. In the nucleus, TRAMP participates in RNA surveillance, processing, and turnover by stimulating the 3'-5' exonucleolytic degradation of targeted RNAs by the nuclear exosome. Previous studies have indicated that within TRAMP Air1 and Air2 provide crucial protein-protein interactions that link the individual subunits of the complex. However, the mechanistic details of these protein-protein interactions are poorly understood. The work in this dissertation has characterized a previously unknown binding interface between Air2 and another TRAMP component, the helicase Mtr4. This interaction may explain how helicase activity is modulated in TRAMP. In addition to TRAMP protein interactions, preliminary studies have identified a small region of Air1 that is required for modulating the activity of a protein that is not found in TRAMP, the methyltransferase Hmt1. Collectively, these studies provide important characterization of Air1 and Air2 protein-binding interactions, and establish a foundation for future research efforts aimed at exploring Air protein function.

(149 pages)

PUBLIC ABSTRACT

Binding Interactions of (*R*)- and (*S*)-hydroxypropyl-CoM Dehydrogenases

and the Zinc Knuckle Proteins Air1 and Air2

by

Jeremy W. Bakelar, Doctor of Philosophy

Utah State University, 2015

Major Professor: Dr. Sean J. Johnson
Department: Chemistry and Biochemistry

A thorough understanding of protein function requires knowledge of how proteins interact with their substrates and with other proteins. The work entailed in this dissertation describes the binding interactions of proteins from two different model systems: (1) the dehydrogenase enzymes *R*- and *S*-HPCDH and (2) the zinc knuckle proteins Air1 and Air2.

R- and *S*-HPCDH are highly similar enzymes (42% identical) that function in a unique metabolic pathway found in the soil bacterium *Xanthobacter autotrophicus*. The bacterium produces *R*- and *S*-HPCDH simultaneously to facilitate the transformation of two different forms of the organic molecule epoxypropane to a common product that can be further metabolized and used as a source of energy for the microbe. *R*- and *S*-HPCDH are highly specific for either the right-handed or left-handed form (*R*- or *S*- forms) of their substrate molecules, respectively. Presented here are x-ray crystal structures

(structural models) of *S*-HPCDH. Comparisons to the previously reported structure of *R*-HPCDH reveal structural differences that provide each enzyme with a distinct preference for binding and processing either the *R*- or *S*- form of their substrate molecules, demonstrating a structural basis for substrate preference by *R*- and *S*-HPCDH.

Air1 and Air2 are highly similar (45% identical) eukaryotic proteins that individually function within an essential three-protein complex called TRAMP. In the nucleus, TRAMP functions in RNA surveillance which is used to monitor different types of RNA molecules found in the nucleus and stimulate the degradation of any RNAs that need to be further processed or eliminated. Previous studies have indicated that Air1 and Air2 are involved in mediating crucial protein-protein interactions that link together the individual protein subunits of TRAMP. The work in this dissertation characterizes a previously unknown binding interface between Air2 and another TRAMP protein-component, the helicase Mtr4. Importantly, this interaction may explain how the functional activity of Mtr4 is modulated upon formation of TRAMP, a critical TRAMP functionality. In addition to protein interactions within TRAMP, this work has also identified a small region of Air1 that binds and regulates the activity of a protein that is not part of TRAMP, the methyltransferase Hmt1. Collectively, these studies reveal important and previously unknown binding interactions of the multifaceted proteins Air1 and Air2, and provide a foundation for future research efforts aimed at understanding their functions.

To my family Nichole, Kyler, and Carsen I dedicate this dissertation

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION I: DEHYDROGENASES OF BACTERIAL EPOXIDE

CARBOXYLATION

During the last century, the increase of human dependence on industrial chemical processes has resulted in the release of a myriad of potentially detrimental carbon based compounds into the environment. These environmental contaminants include commonly recognized greenhouse gases like CO and CO_2, and also include many short-chain unsaturated hydrocarbons such as ethylene, propylene, butylene, and styrene; all of which are produced on a massive scale worldwide. The epoxides formed from alkenes and their halogenated counterparts are especially reactive molecules that may have toxic, mutagenic, and carcinogenic effects on biological organisms [1, 2]. High concentrations of these hazardous molecules can often be detected in soils and ground water samples in areas near industrial sites [3], posing a great risk to the local biota and to human health. Interestingly, several bacterial species have been identified that are capable of detoxifying and converting short chain (C_2-C_6) alkenes and epoxides into usable non-reactive central metabolites. One such species is *Xanthobacter autotrophicus* Py2, which is capable of growth using propylene or its corresponding epoxide, epoxypropane, as its sole source of carbon and energy [4]. Bacteria containing such metabolic pathways for converting alkenes and epoxides into non-harmful molecules are generally believed to play an essential role in the re-mineralization of this carbon in the global carbon cycle [3].

It is of considerable interest to understand the biological mechanisms used by microbes to eliminate various toxic hydrocarbons, as both the whole organisms and the

enzymes comprising their novel metabolic pathways have potential bioremediation and biotechnological applications. The focus of research presented in this dissertation is to gain further insight into the metabolic pathway of epoxide degradation used by the bacterium *X. autotrophicus* Py2. Of particular interest are two stereospecific dehydrogenases which allow the bacterium to metabolize both *R*- and *S*- enantiomers of epoxypropane.

Sources of propylene and epoxypropane

Propylene and its oxidized derivative epoxypropane are generated from both anthropogenic and biogenic sources. Propylene is a three carbon alkene that is produced primarily as a byproduct of petroleum refining wherein the process of steam cracking large hydrocarbon feed-stocks is used to produce other smaller hydrocarbons such as ethylene. In 2011 it was estimated that the global capacity of industrial propylene production was nearly 70 million metric tons and expected to increase at a rate of 5% annually [5]. In the United States, nearly two thirds of the propylene that is produced is used for the manufacturing of plastics in the form of polypropylene, and 17% is used directly for the production of epoxypropane (also known as propylene oxide) [6]. Epoxypropane is a highly reactive molecule that is used as a versatile chemical intermediate for the production of many other compounds. The reactivity of epoxypropane stems from its strained three-membered ring and propensity to undergo nucleophilic attack by a number of different compounds including organic and inorganic acids and bases, alcohols, and amines. The major industrial products derived from epoxypropane include polyglycol ethers used in polyurethane foams, and propylene

glycols which are used in a number of applications including polyester resins, chemical solvents, pharmaceuticals, foods, antifreeze, and many others. By volume propylene and epoxypropane are amongst the top 50 chemicals produced worldwide [7]. As the production of these potentially toxic compounds continues to increase, the volume of propylene and epoxypropane that are emitted into local environments also increases. For example, since 1983 the annual US emissions of propylene and epoxypropane have been estimated to have increased from around 440 thousand tons to over 700 thousand tons [8, 9].

In addition to industrial production, propylene and epoxypropane are also produced, albeit at a much smaller scale, from biological sources. Propylene for example is one of many small alkenes that are generated and excreted by plant vegetation and fungi. Propylene is also formed as a product of the combustion of organic materials, i.e. burning of biomass and fossil fuels [9]. Epoxypropane and other aliphatic epoxides can be produced biologically in some eukaryotic pathways and by several alkene oxidizing bacetria. The epoxides generated *in vivo* are generally short lived intermediate molecules that are produced by the initial step in biological pathways that are used for alkene detoxification, or in metabolic pathways that allow some bacteria to utilize alkenes and epoxides as carbon and energy sources. In each type of pathway, epoxides are produced by the oxidation of alkenes by alkene monooxygenase enzymes. Alkene monooxygenases (AMOs) typically exhibit broad substrate specificities and convert various alkenes into epoxides by incorporating O_2 across the olefin bond of the alkene, as illustrated for the substrate propylene and product epoxypropane in Figure 1-1. Notably, most AMOs are somewhat stereoselective producing both *R*- and *S*- enantiomers of aliphatic epoxides.
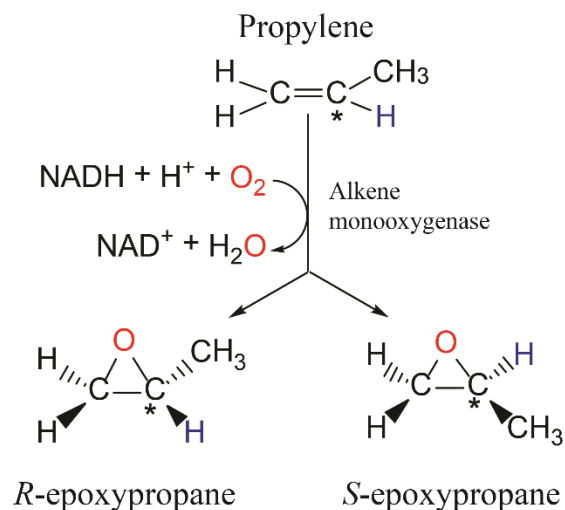
Figure 1-1. Oxidation of propylene catalyzed by alkene monooxygenase.

The monooxygenases used in alkene detoxification pathways include the heme containing family of cytochrome P450 enzymes found throughout eukaryotes and prokaryotes [10], and other non-heme containing monooxygenases found exclusively in prokaryotes [11-13]. AMOs used to generate aliphatic epoxides such as epoxypropane for carbon and energy sources have been isolated from two bacterial species, *X. autotrophicus* Py2 and *R. rhodochrous* B-276 [14, 15]. In these bacteria, alkene metabolism is initiated by the epoxidation of alkenes (i.e. propylene to epoxypropane) by a non-heme di-iron type AMO.

Biological Reactivity of Epoxides

Epoxides are likely to have adverse effects on biological systems. The strong electrophilic nature of epoxides allows them to readily form covalent adducts with a number of biological macromolecules, including DNA, RNA, and proteins [2]. In DNA

and RNA, the reactive nucleophiles are nitrogen atoms of purine and pyrimidine bases. In

proteins, the reactive nucleophiles include nitrogen atoms of the imidazole ring of

histidine, and sulfur atoms found in the side chains of cysteine and methionine residues.

In each case, the biological nucleophile attacks and opens the epoxide oxirane ring,

creating a conjugated product with the nucleophile covalently bound. (Scheme 1-1).

Scheme 1-1.
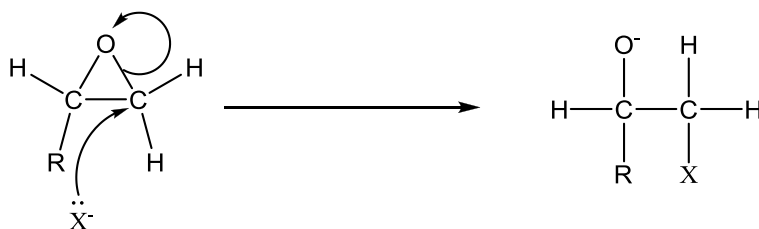
Such modifications of protein and nucleic acids have profound effects on cellular

functions. Therefore, many aliphatic epoxides, including epoxypropane, are characterized

as having toxic, mutagenic, and carcinogenic effects on living organisms [1, 2]. For

organisms that contain alkene degradation pathways, the initial conversion of an alkene

to epoxide results in a more reactive and potentially harmful compound that must be

quickly transformed into a non-harmful molecule.

Biogenic remediation of epoxides

To circumvent the toxic effects of epoxide reactivity, microbial systems use

nucleophiles other than nucleic acids and protein to react with epoxides and render them

inactive as electrophiles. In these reactions, epoxides are either converted into a less

detrimental compound which can then be excreted by the microbe (detoxification

pathway), or converted into an organic metabolite that can be used as a source of carbon

and energy via productive metabolism (metabolic pathway). The differences between the

pathways lay in the different nucleophiles that are used to react with the epoxide and the

fate of the product that is formed. In the case of detoxification pathways, many types of

bacteria contain detoxification enzymes such as glutathione S-transferases (GSTs) and

epoxide hydrolayses which use glutathione or water, respectively, as nucleophiles to

attack and open the epoxide ring [16-18] (Figure 1-2A, B). The less reactive product can

then be excreted by the bacterium or used for other biological purposes. In the case of

productive metabolism, several pathways have been characterized in which the first step

involves the nucleophilic addition of a biological molecule such as glutathione or water,

as described above for detoxification pathways. Another strategy for metabolizing

epoxides is found in styrene utilizing bacteria where metabolism of styrene proceeds by

isomerization of styrene oxide to the corresponding aldehyde phenylacetaldehyde [19]

(Figure 1-2C). Another less common strategy is used by several aerobic bacteria that

have been isolated with propylene and epoxypropane as the only source of carbon and

energy. In these pathways a thiol of the atypical cofactor, 2-mercaptoethanesulfonate

(coenzyme M; CoM), is used as the reactive nucleophile which attacks and opens the

epoxide ring, generating a coenzyme M conjugate (Figure 1-2D). The coenzyme M

conjugate is then further metabolized in a series of reactions including a final

carboxylation step producing a molecule of acetoacetate which can then be transformed

into acetyl-CoA for energy production. This unique pathway, which has become known

as the epoxide carboxylation pathway, was initially discovered and characterized within
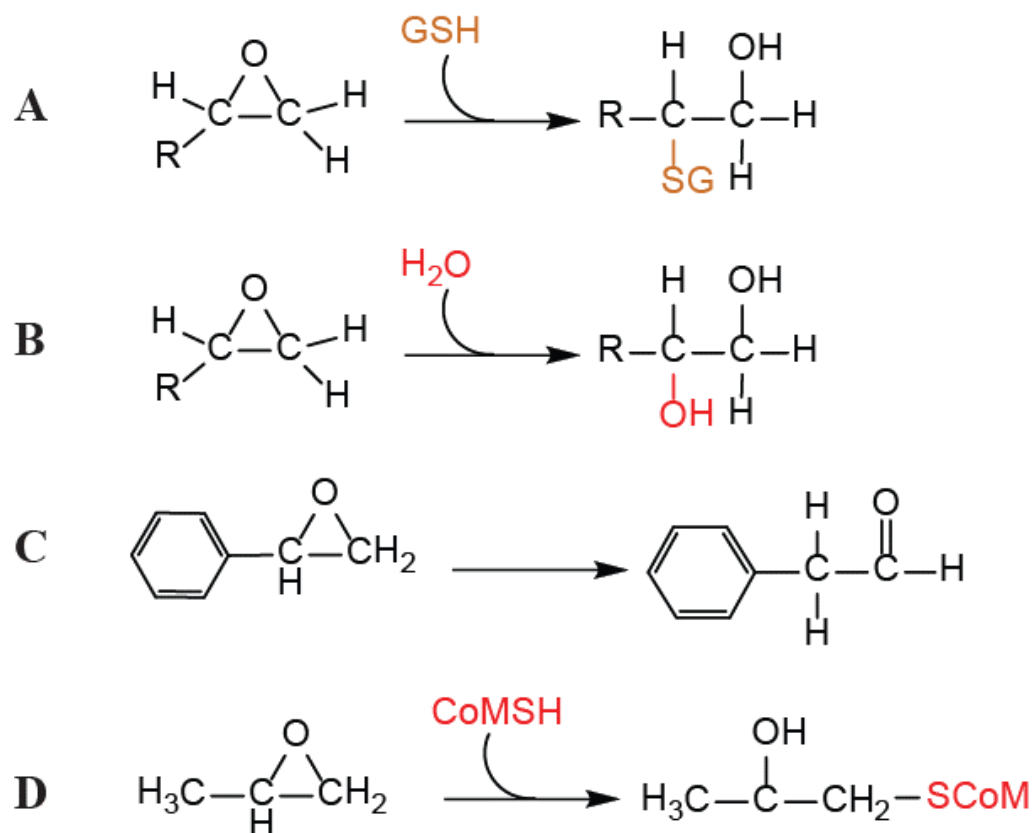
Figure 1-2. Biological strategies of epoxide metabolism. (A) Glutathione transferase, (B) epoxide hydrolase, (C) Styrene oxide isomerase, (D) epoxyalkane:CoM transferase.

the soil bacterium *X. autotrophicus* [20, 21] strain Py2, and later extended to *R. rhodochrous* strain B276 [22]. Since then, homologues of each of the four enzymes of the pathway have been found in other alkene metabolizing bacteria including additional strains of *Xanthobacter* and *R.rhodochrous, Mycobacterium, Pseudomonas,* and *Alcaligenes*. Overall, the pathway is a very effective way to convert a hazardous 3-carbon compound into a relatively inert 4-carbon metabolite. Throughout the last decade, there has been significant interest in understanding the various enzymatic steps of bacterial epoxide carboxylation, as each of the enzymes in the pathway and the various bacterial

species that contain them have potential bioremediation and biotechnological applications.

The epoxide carboxylation pathway of *X. autotrophicus* Py2

As mentioned previously, the first step in metabolizing short-chain alkenes like propylene is the oxidative conversion of the alkene into its corresponding epoxide by AMOs. In *X. autotrophicus* Py2 the AMO used is a multicomponent NADH-dependent enzyme that is highly stereoselective, as demonstrated by propylene oxidation in which the AMO produces a racemic mixture of 95% *R*-epoxypropane and 5% *S*-epoxypropane [15, 21]. In the presence of CO_2, both enantiomers of epoxypropane can be effectively metabolized by the bacterium, as each is transformed into a molecule of acetoacetate using a three-step four-enzyme pathway called the epoxide carboxylation pathway [21]. The pathway converts *R*-and *S*- enantiomers of a variety of short-chain aliphatic epoxides into a β-keto acid which can then be converted into two molecules of acetyl-CoA (Figure 1-3). Although aliphatic epoxides of varying chain lengths (C_2-C_6) can be effectively metabolized via the epoxide carboxylation pathway [3], the most extensively characterized substrate is epoxypropane [23-25]. Therefore, the following descriptions of the pathway are in the context of the conversion of epoxypropane to the β-keto acid acetoacetate. Experimental evidence has indicated that the reactions catalyzed at each of the three steps are fully reversible. The pathway requires NADPH, NAD^+, and fixes a molecule of CO_2.

A multitude of biochemical and structural studies have revealed unique functions and many mechanistic details of the enzymes at each of the three steps of the epoxide carboxylation pathway. The first step involves the opening of the epoxide ring and
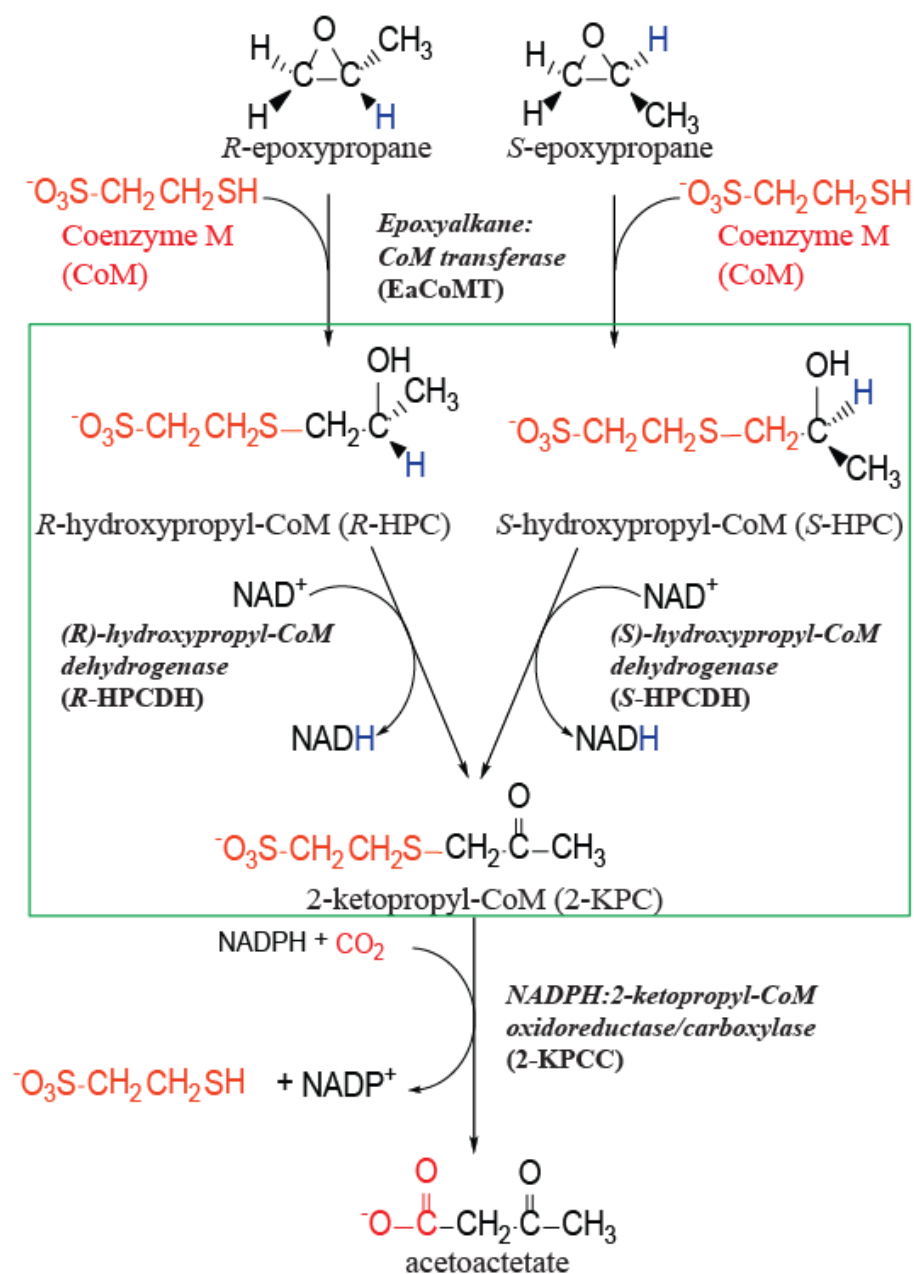
Figure 1-3. Three steps of the epoxide carboxylation pathway in *X. autotrophicus* Py2. Enzymes catalyzing reactions are written in italics. Green box indicates the second step which is catalyzed by the two enzymes *R*-hydroxypropyl-CoM dehydrogenase (*R*-HPCDH) and *S*-hydroxypropyl-CoM dehydrogenase (*S*-HPCDH).

nucleophilic addition of the cofactor CoM. This reaction is catalyzed by the enzyme

epoxyalkane coenzyme M transferase (EaCoMT) [20, 23, 26]. Sequence analysis

indicates that EaCoMT belongs to the Zn-containing alkyl transferase family of enzymes

which catalyze nucleophilic substitution reactions using activated thiols as nucleophiles.

Like other Zn-containing alkyl transferases, EaCoM activates its reactant thiol group

(thiol of CoM) by coordinating it to a Zinc ion. The metal ion coordination lowers the

pK_a of the CoM thiol by 1.7 pH units (9.1 to 7.4), facilitating nucleophilic attack by the

now deprotonated thiolate, and subsequent covalent addition of CoM to the aliphatic

epoxide (Figure 1-4) [27, 28]. EaCoMT is unique in the family of Zn-containing alkyl

transferases because it catalyzes a nucleophilic addition reaction rather than a

nucleophilic substitution reaction. The utility of CoM in the pathway is evident in the

remaining two steps, as each of the following three enzymes in the pathway use the

negatively charged sulfonate moiety of CoM as a convenient *molecular handle* to
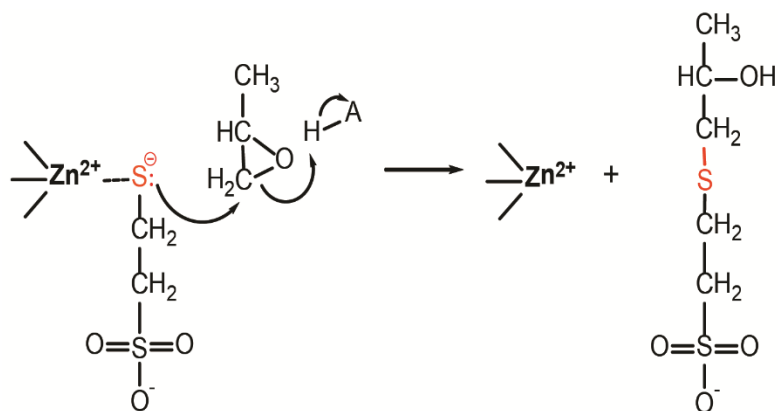
Figure 1-4. Covalent addition of CoM to epoxypropane by
epoxyalkane coenzyme M transferase (EaCoMT). A Zinc ion
activates the CoM thiol (colored red) for nucleophilic attack
and subsequent addition of CoM to epoxypropane, producing
R- and S-hydroxypropyl CoM (R- and S- HPC).

properly orient substrates for catalysis [29-33].

The second step in the pathway, and the focus of Chapter 2 of this dissertation, involves two dehydrogenase enzymes, (R)-hydroxypropyl-coenzyme M dehydrogenase (R-HPCDH) and (S)-hydroxypropyl-coenzyme M dehydrogenase (S-HPCDH). These are homologous enzymes (42% sequence identity) that are highly stereospecific for either the R- or S-enantiomers of HPC. In this step, both R- and S- HPCDH are used *in concert* to convert each enantiomer of HPC into the same achiral product 2-ketopropyl CoM (2-KPC). A remarkable feature of R- and S-HPCDH is their ability to discriminate between R- and S- enantiomers of HPC, as each enzyme exhibits only 0.5%-1% activity when using the opposite HPC isomer as a substrate [21]. R- and S-HPCDH are members of the *classical* short-chain dehydrogenase reductase (SDR) superfamily of enzymes. Enzymes in this family are NAD(P)(H) dependent enzymes that share a common catalytic tetrad (Tyr–Lys–Ser–Asn) and catalytic mechanism, as shown for the conversion of R-HPC to 2-KPC in Figure 1-5. In the mechanism shown, the role of the catalytic tetrad residues and the chemistry around the chiral carbon (C2 in R-HPC) is analogous to the general mechanism described for other SDR enzymes [34, 35]. Specifically, the tetrad serine assists in positioning the hydroxyl group of the substrate near the catalytic tyrosine. The tetrad lysine has a dual role in coordinating NAD^+ and in lowering the pK_a of the catalytic tyrosine hydroxyl group through a proton relay that connects the tyrosine, through hydrogen bonding, to the bulk solvent. The proton relay path is generated through side chains of three of the catalytic tetrad residues (Tyr-Lys-Asn), hydroxyl group of the nicotinamide ribose, and water molecules that lead away from the active site to bulk solvent. In its deprotonated form, the tyrosine acts as a general base abstracting a proton

from the C2 hydroxyl group of the substrate. A hydride is simultaneously transferred

from the C2 carbon to NAD$^+$, forming NADH and the ketone 2-ketopropyl coenzyme M

(2-KPC) as products.
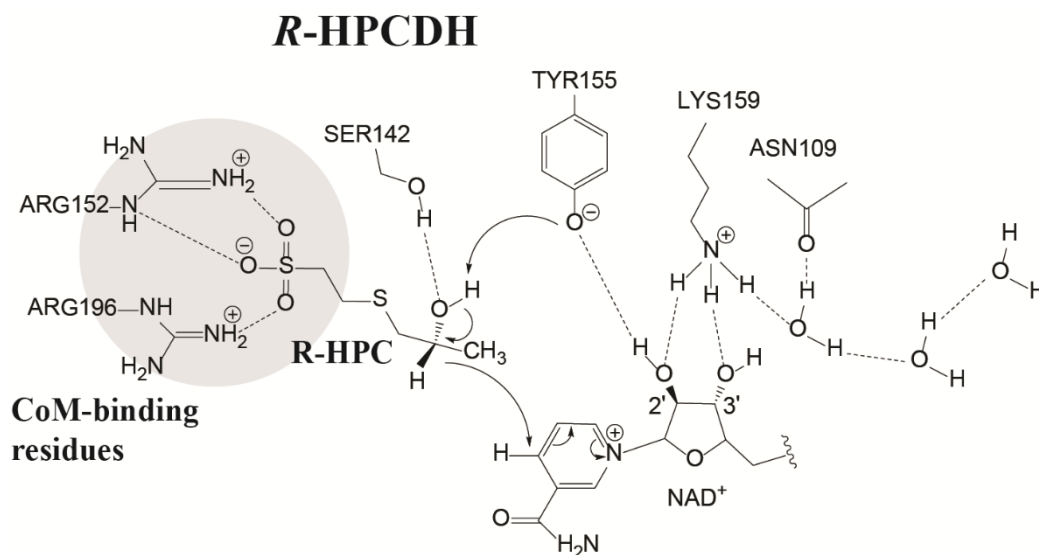
Figure 1-5. Oxidation of *R*-HPC by *R*-hydroxypropyl-coenzyme M dehydrogenase (*R*-HPCDH). Binding of the sulfonate moiety of CoM is indicated by a grey circle. Catalytic tetrad residues (Ser-Tyr-Lys-Asn) are shown. Tyr155 acts as a general base and a hydride is transferred from *R*-HPC to NAD$^+$. A proton relay connects catalytic residues to bulk solvent through water mediated hydrogen bonding. Image was modified from Sliwa, *et al.* 2010.

carboxylation of 2-KPC and the regeneration of CoM by the enzyme NADPH:2-

ketopropyl-CoM carboxylase/oxidoreductase (2-KPCC). Sequence analysis has indicated

that 2-KPCC belongs to the FAD containing NADPH:disulfide oxidoreductase (DSOR)

family of enzymes [36]. All enzymes in this family employ a general mechanistic

strategy in which a reduced form of flavin adenine dinucleotide (FADH) is used to reduce

a conserved cysteine disulfide bond in the active site, which activates the cysteine

residues for catalysis [37]. The cysteine residues proximal and distal, with respect to

FAD, are termed the flavin and interchange thiols, respectively. Typical DSOR enzymes

use the activated thiols to catalyze the two electron reduction of a substrate disulfide bond

(Figure 1-6), the interchange thiol attacks one of the sulfur atoms of a substrate disulfide

bond, resulting in disulfide bond cleavage and a mixed disulfide formed between the

interchange thiol and the substrate disulfide. The catalytic cysteine pair is then re-

oxidized leading to the reduction and subsequent release of the second substrate thiol.
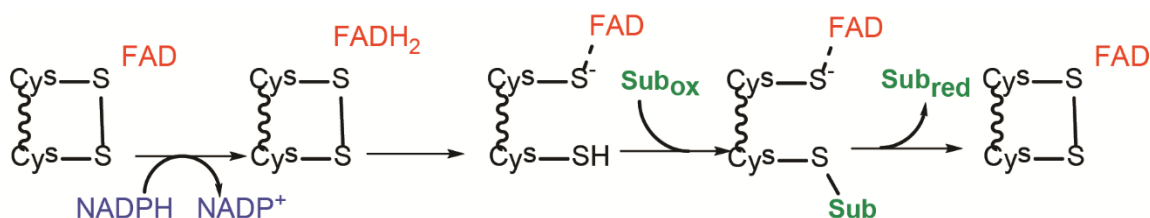
Figure 1-6. Enzymatic steps catalyzed by the DSOR family of enzymes. A substrate with an oxidized disulfide bond is attacked by the interchange thiol. Oxidation of the catalytic cysteine pair reduces the covalently bound substrate and the reduced substrate is released.

2-KPCC is unique amongst DSOR enzymes in at least two respects. First, it

catalyzes the reduction of a thioether rather than a disulfide bond. Second, it is the only

known carboxylase within the DSOR family. Mechanistic studies [38] have provided

evidence for a mechanism where the interchange thiol attacks the thioether bond of 2-

KPC resulting in a mixed disulfide between CoM and the interchange thiol, and the

formation of an enolacetone anion. The enolacetone anion then undergoes carboxylation

producing acetoacetate. Re-oxidation of the catalytic cysteine pair releases CoM, which

can then be reused in the in the first step of the epoxide carboxylation pathway.

The initial mechanistic studies of 2-KPCC left many unanswered questions about

the unique features of its mechanism, including how the substrate is bound, and how the

highly reactive enolacetone anion is stabilized and carboxylated. A variety of X-ray crystal structures have provided significant insights to these questions [32, 39, 40]. The different crystal structures of 2-KPCC include; a substrate-free form (apo-enzyme), enzyme bound to the substrate 2-KPC, enzyme with a mixed disulfide of CoM, and a CO_2 bound form. The 2-KPC bound structure revealed that the enzyme uses two positively charged amino acids to bind the sulfonate moiety of CoM. Upon substrate binding a conformational change occurs that creates a hydrophobic pocket around the substrate 2-KPC which appears to be accessible only by a small hydrophobic channel, presumably for CO_2. A hydrogen bonding network was also identified in the 2-KPC bound structure which has been proposed to stabilize the enolacetone anion. The hydrogen bonding network consists of an ordered water molecule that is hydrogen bonded to the carbonyl oxygen of 2-KPC and two histidine residues (H84 and H137). It is presumed that hydrogen bonding by the His-oriented water molecule serves to both stabilize the enolacetone anion and orient its methylene group for attack on CO_2.

The novelty of the epoxide carboxylation pathway is exhibited at each of the three steps. (1) The usage of CoM, (2) the use of two stereospecific dehydrogenases, and (3) the fixation of CO_2. The discovery of CoM as the 4-carbon carrier in this pathway and the fixation of CO_2 were surprising and unprecedented. Prior to its discovery in this bacterial pathway, CoM the smallest known biological cofactor, had only been identified to function in methanogenisis in archaebacterial [41-43]. A Carboxylation step consuming CO_2 had also never been demonstrated in any metabolic pathway used for small hydrocarbon metabolism. The simultaneous usage of two homologous dehydrogenases in the same pathway is also a very rare occurrence throughout biology.

As mentioned above, extensive biochemical and structural studies have revealed novel functions and mechanisms of the enzymes at each of the pathway steps. However, the mechanisms that govern stereospecificity of the dehydrogenases *R*- and *S*-HPCDH is a feature that unlike other novel aspects of the pathway, has not been fully characterized. In addition, because *R*- and *S*-HPCDH are homologous enzymes catalyzing the same reaction but with opposite stereochemistry, they are a great model for studying mechanisms of stereospecificity in enzyme catalysis, and more specifically, the mechanisms controlling specificity within the SDR superfamily of enzymes.

SDR superfamily of enzymes

R-and *S*-HPCDH belong to the *classical* short-chain dehydrogenase/reductase (SDR) super family of enzymes. The SDR family represents one of the largest and oldest enzyme families known with over 60,000 members annotated in sequence databases [44]. The characteristics that define this protein family are chain lengths of approximately 250 amino acids, a large N-terminal domain (Rossman fold) containing a common sequence motif (TGxxxGxG) for binding NAD(P)(H), a conserved catalytic tetrad (serine, tyrosine, lysine, asparagine), and a small C-terminal domain that typically functions in substrate binding . SDR enzymes are arguably the most widely distributed group of enzymes. Members have been identified throughout the three domains of life and represent multiple enzyme classes including isomerases, transferases, lyases, and oxidoreductases. Although these enzymes have conserved residues for binding NAD(P)(H) and catalysis, they have very broad substrate specificities including substrates such as xenobiotics, steroids, sugars, and aliphatic alcohols [34, 44, 45].

Another interesting feature of the SDR superfamily is that despite relatively low sequence homology (15-30% sequence identity), and diverse substrate preferences, the 3-dimensional structures of SDR enzymes (over 400 in the PDB) are clearly homologous with nearly super-imposable tertiary and quaternary structures. The predominant structural feature of SDR proteins is a large N-terminal domain composed of a common α/β-folding pattern characterized by a central β-sheet typical of a Rossmann-fold with three α-helices on each side (Figure 1-7). It is generally understood that members of the SDR superfamily utilize this common N-terminal Rossman fold domain as a structural scaffold in which minor modifications to a smaller C-terminal extension (the C-terminal domain) confer large differences in substrate preference.

The Rossman fold dinucleotide binding motif is the most common fold found in the Protein Data Bank (PDB), and the various enzymes of the SDR superfamily highlight the versatility of this fold in being able to accommodate diverse substrates for various types of chemical transformations. Because of this versatility, many SDR enzymes have been targets of enzyme engineering for biocatalysis [46-50]. Of particular interest in multiple industries is the engineering of stereospecific and stereoselective dehydrogenases that can be used to produce certain alcohol isomers or clarify racemic mixtures. The potential commercial value of dehydrogenases within the SDR superfamily for stereospecific and stereoselective transformations has stimulated significant interest in elucidating the
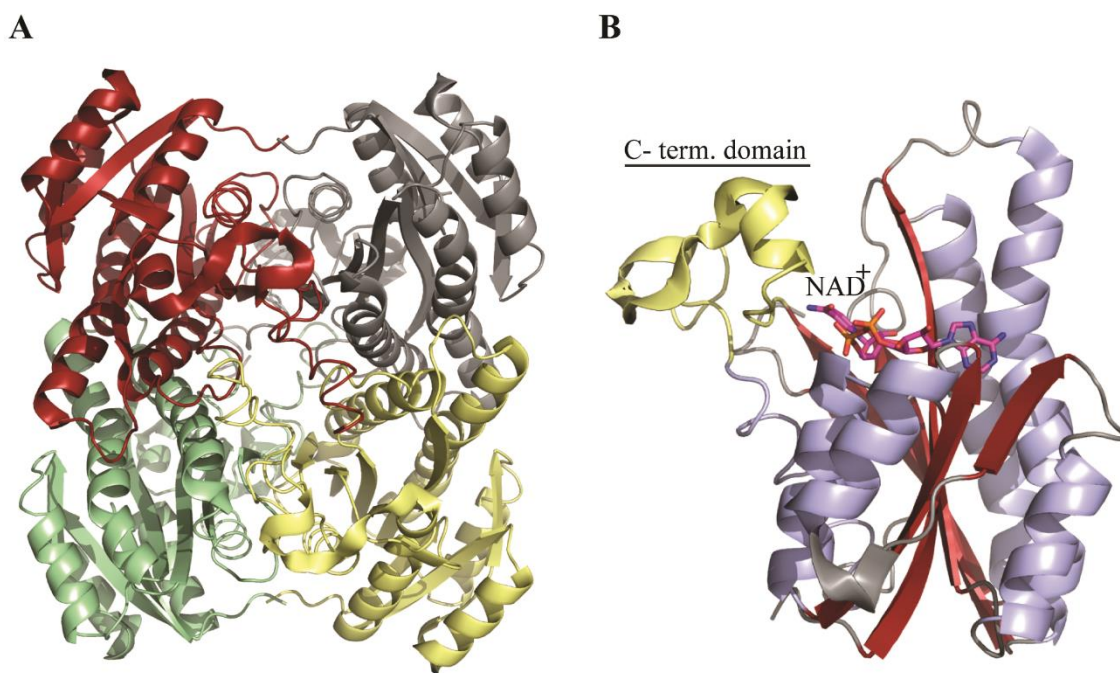
Figure 1-7. Ribbon diagram of classical SDR enzyme (3α/20β HSD; PDB: 2HSD) (A) Tetramer. (B) Monomer; the Rossman fold is depicted with β-strands in red and α-helicies in light blue. The C-terminal domain is colored yellow and NAD^+ is shown as sticks.

catalytic and structural features that govern there mechanisms [49, 51, 52]. Furthermore, once biochemical and molecular details of a particular enzyme is known, enzymatic properties such as specificity and selectivity can be modulated through directed evolution or mutagenesis to fine-tune the enzyme for various functions.

Within the SDR superfamily there are numerous stereospecific enzymes. However, few stereospecific enzymes have been found to function simultaneously with a homologous partner-enzyme. Accordingly, *R*- and *S*-HPCDH represent a unique and ideal model system to examine stereospecificity within the SDR superfamily because they are homologous proteins performing essentially identical chemistry in the same

pathway, yet both enzymes are highly specific for their respective substrates (*R*-HPC or *S*-HPC).

Mechanistic studies of *R*-and *S*-HPCDH

Both *R* and *S*-HPCDH have been recombinantly expressed, purified, and characterized biochemically. Mutagenesis studies have revealed that like the enzyme 2-KPCC (the following step in the pathway), Both *R*- and *S*-HPCDH possess two positively charged amino acids that are involved in binding the negatively charged sulfonate tail (sulfonate moiety of CoM) of the substrate. These residues have been identified as R152 and R196 in *R*-HPCDH, and R211 and K214 in *S*-HPCDH [53, 54]. Site directed mutagenesis of these residues in either enzyme dramatically alters their ability to oxidize natural substrates, whereas the same mutations do not alter the oxidation of similar substrates lacking the sulfonate moiety [33, 54]. These studies suggest that in each enzyme the sulfonate binding amino acids are important contributors for aligning the natural substrates properly for catalysis. From these studies, it was also hypothesized that alternative positioning of the substrate binding residues within each enzyme could be a contributing factor to their opposite stereospecificity.

Inhibition and kinetic analyses have suggested differences in how each enzyme controls stereospecificity. Inhibition studies showed that the substrate *S*-HPC is a competitive inhibitor of *R*-HPC oxidation in *R*-HPCDH, with a K_{ic} nearly identical to the *Km* of the natural substrate [54]. This observation indicated that *R*-HPCDH has a similar preference for binding either substrate (*R*-HPC or *S*-HPC). Similar studies performed with the *S*-HPCDH enzyme showed that the *R*-HPC substrate is not a competitive

inhibitor of S-HPC oxidation [33], indicating that unlike R-HPCDH, the S-enzyme has a much greater preference for binding its natural substrate S-HPC over R-HPC.

To help clarify differences observed by inhibition studies, a side-by-side kinetic analysis of R- and S-HPCDH for oxidizing each substrate isomer (R-HPC and S-HPC) was performed [33]. The catalytic efficiencies (k_{cat}/K_m) of each enzyme for its "natural" substrate isomer are in the range of 1 to 8 x 10^5. Whereas the catalytic efficiencies of each enzyme using the opposite substrate isomers are each nearly three-orders of magnitude lower. Thus, as previously described, both enzymes are highly effective at discriminating between different HPC stereoisomers. Interestingly, this kinetic study revealed that stereospecificity in the two enzymes is governed by different kinetic mechanisms. In the R-enzyme, the difference in catalytic efficiencies between the two substrates is governed largely by a difference in the values of k_{cat} (turnover rate), which is 402 times lower when using S-HPC as substrate compared to the R-HPC substrate. Alternatively, in the S-enzyme the difference in catalytic efficiencies between the two substrates is predominated by a large difference in the value of K_m for each substrate (K_m is 209 times higher for R-HPC compared to S-HPC). This increase in K_m may be indicative of an inherent inability of the R-HPC substrate to effectively bind within the catalytic site of the S-enzyme.

In addition to biochemical characterization, R-HPCDH has been characterized structurally [30]. A 1.8 Å crystal structure of R-HPCDH bound to NAD^+ and the product 2-KPC showed that the enzyme is structurally homologous to other members of the SDR superfamily. Specifically, like many other SDR enzymes its crystal structure is a tetramer with each monomer composed of a 2-domain organization (Figure 1-8). The large N-

A

B
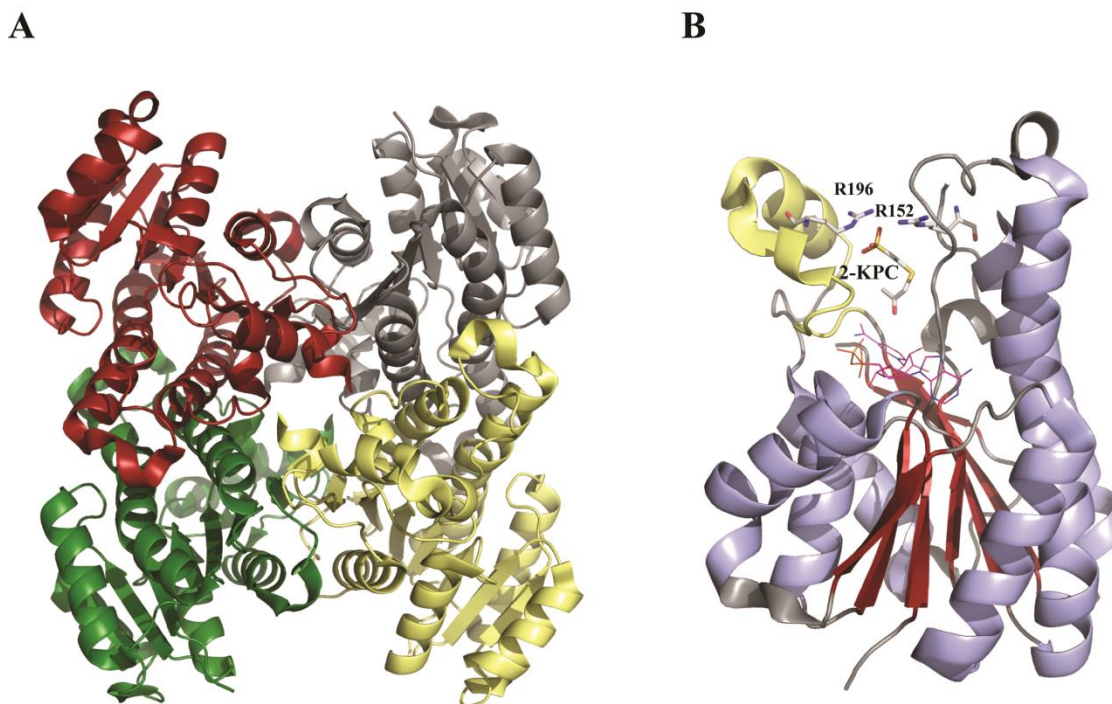
Figure 1-8. Ribbon diagram of *R*-HPCDH (PDB: 2CFC). (A) Tetramer. (B) Monomer; Rossman fold is depicted with β-strands in red and α-helicies in light blue. The C-terminal domain is colored yellow and NAD$^+$ is shown as lines. The product 2-KPC is shown with the sulfonate moiety of CoM bound by R196 and R152.

terminal domain is in the form of a Rossman fold composed of seven parallel β-strands

flanked on each side by three α helices. Unambiguous electron density of residues bound

to 2-KPC confirmed that R152 and R196 are used to bind directly to the sulfonate moiety

of CoM. Unfortunately, the positioning of the reactive carbonyl group of 2-KPC at the

catalytic site could not be observed in the crystal structure. Instead, the structure exhibits

the 2-KPC carbonyl group more than seven angstroms from the catalytic tyrosine,

presumably representing a conformation of 2-KPC exiting the active site [30]. In an

attempt to identify structural differences between *R*- and *S*-HPCDH, The *R*-HPCDH

structure was used to construct a homology model of *S*-HPCDH [33]. As expected, the

homology model predicted nearly opposite positioning of *S*-HPCDH sulfonate binding

residues compared to the sulfonate binding residues in *R*-HPCDH.

Taken together, results from biochemical and structural studies have been used to

formulate a model for stereospecificity in *R*- and *S*-HPCDH. According to this "mirror

image" model, opposite stereospecificity of *R*- and *S*-HPCDH is obtained via opposite

positioning of sulfonate binding residues in each enzyme, which results in reverse

orientations of the chemical groups attached to the C2 carbon (chiral carbon) of each

substrate relative to a common catalytic site [30, 33, 54]. The different kinetic values

described above for each enzyme can be rationalized by this model in the following way:

In the *R*-enzyme both substrates can bind at the catalytic site with similar affinity, but a

960-fold increase in catalytic efficiency using *R*-HPC over *S*-HPC is achieved because

the *R*-HPC reactive group is in the proper stereo-orientation for catalysis. When the *S*-

HPC substrate binds to the *R*-enzyme its C2 hydrogen is oriented in the opposite direction

pointing away from NAD$^+$. This opposite orientation inhibits direct hydride transfer from

C2 of the substrate to NAD$^+$, and results in a 402-fold decrease in the value of k_{cat} for

oxidation of *S*-HPC compared to *R*-HPC. In the *S*-enzyme, the *R*-HPC substrate cannot

effectively bind at the catalytic site because its C2 methyl group is not accommodated

due to steric clashes with amino acid side chains or NAD$^+$, resulting in a 209-fold

increase in the K_m value for oxidizing the *R*-HPC substrate compared to the *S*-HPC

substrate. Although this model is plausible, it does not describe the actual structural

differences that may account for the observed kinetic differences between the enzymes,

and it cannot be convincingly validated without a comparative structural analysis

between the two enzymes and their substrate binding interactions at the catalytic site.

Concluding Remarks

The mechanism that governs stereospecificity of *R*- and *S*-HPCDH is one of the last outstanding questions regarding the remarkable pathway of bacterial epoxide carboxylation. The work described in Chapter 2 of this dissertation provides the first side-by-side structural comparison between *R*- and *S*-HPCDH, and the first structural analysis of how an HPC substrate is aligned within the active site of one of these enzymes (*S*-HPCDH). The results from these studies reveal unique substrate binding pockets displayed by each enzyme, a structural role of methionine residues within each active site, and a structural basis for stereospecificity displayed by *R*- and *S*-HPCDH. Furthermore, the structural mechanisms used by this pair of dehydrogenases to discriminate between two highly similar substrates may provide valuable insight for the rational design of other enzymes within the SDR superfamily.

REFERENCES

[1]   C.E. Voogd, J.J. van der Stel, J.J. Jacobs, Mutation Res. 89 (1981) 269-282.

[2]   D.R. Wade, S.C. Airy, J.E. Sinsheimer, Mutation Res. 58 (1978) 217-223.

[3]   S. Hartmans, J.A. de Bont, W. Harder, FEMS Microbiol. Rev. 5 (1989) 235-264.

[4]   J.R. Allen, S.A. Ensign, J. Bacteriology 178 (1996) 1469-1472.

[5]   I.C. Consulting, IHS Chemical Economics Handbook, Hargrove Pub. New Haven CT, 2011.

[6]   C. Research, Market Study: Propylene, 2011.

[7]   D.C. Company, Usage of Propylene and Propylene Oxide 2014.

[8]   Middleton, Stockwell, Carter, Atmospheric Environment 24A (1990) 1107-1133.

[9]   Altshuller, Atmospheric Environment 17 (1983) 2131-2165.

[10]   F.P. Guengerich, J. Biol. Chem. 266 (1991) 10019-10022.

[11]   A.B. Hooper, T. Vannelli, D.J. Bergmann, D.M. Arciero, Antonie Van Leeuwenhoek, Atmosphere 71 (1997) 59-67.

[12]   J.D. Lipscomb, Annual Review of Microbiology 48 (1994) 371-399.

[13]   K. McClay, B.G. Fox, R.J. Steffan, Applied and Environmental Microbiology 66 (2000) 1877-1882.

[14]   A. Miura, H. Dalton, Bioscience Biotechnology and Biochemistry 59 (1995) 853-859.

[15]   F.J. Small, S.A. Ensign, J. Biol. Chem. 272 (1997) 24913-24920.

[16]   R.N. Armstrong, Chemical Research in Toxicology 10 (1997) 2-18.

[17]   R.N. Armstrong, C.S. Cassidy, Drug Metabolism Review 32 (2000) 327-338.

[18]   A.E. Salinas, M.G. Wong, Current Medicinal Chemistry 6 (1999) 279-309.

[19]   S. Hartmans, J.P. Smits, M.J. Vanderwerf, F. Volkering, J.A.M. Debont, Applied and Environmental Microbiology 55 (1989) 2850-2855.

[20]   J.R. Allen, D.D. Clark, J.G. Krum, S.A. Ensign, Proc. Natl Acad Sci USA 96 (1999) 8432-8437.

[21]   J.R. Allen, S.A. Ensign, Biochemistry 38 (1999) 247-256.

[22]   J.G. Krum, S.A. Ensign, J. Bacteriology 182 (2000) 2629-2634.

[23]   J.R. Allen, S.A. Ensign, J. Bacteriology 180 (1998) 2072-2078.

[24]   S.A. Ensign, F.J. Small, J.R. Allen, M.K. Sluis, Archives of Microbiology 169 (1998) 179-187.

[25]   F.J. Small, S.A. Ensign, J. Bacteriology 177 (1995) 6170-6175.

[26]   J.R. Allen, S.A. Ensign, J. Bacteriology 179 (1997) 3110-3115.

[27] J.M. Boyd, S.A. Ensign, Biochemistry 44 (2005) 13151-13162.

[28] J.G. Krum, H. Ellsworth, R.R. Sargeant, G. Rich, S.A. Ensign, Biochemistry 41 (2002) 5005-5014.

[29] J.W. Bakelar, D.A. Sliwa, S.J. Johnson, Archives of Biochemistry and Biophysics 533 (2013) 62-68.

[30] A.M. Krishnakumar, B.P. Nocek, D.D. Clark, S.A. Ensign, J.W. Peters, Biochemistry 45 (2006) 8831-8840.

[31] A.M. Krishnakumar, D. Sliwa, J.A. Endrizzi, E.S. Boyd, S.A. Ensign, J.W. Peters, Microbiology and Molecular Biology Reviews : MMBR 72 (2008) 445-456.

[32] B. Nocek, S.B. Jang, M.S. Jeong, D.D. Clark, S.A. Ensign, J.W. Peters, Biochemistry 41 (2002) 12907-12913.

[33] D.A. Sliwa, A.M. Krishnakumar, J.W. Peters, S.A. Ensign, Biochemistry 49 (2010) 3487-3498.

[34] C. Filling, K.D. Berndt, J. Benach, S. Knapp, T. Prozorovski, E. Nordling, R. Ladenstein, H. Jornvall, U. Oppermann, J. Biol. Chem. 277 (2002) 25677-25684.

[35] H. Jornvall, B. Persson, M. Krook, S. Atrian, R. Gonzalez-Duarte, J. Jeffery, D. Ghosh, Biochemistry 34 (1995) 6003-6013.

[36] S.A. Ensign, J.R. Allen, Annual Review of Biochemistry 72 (2003) 55-76.

[37] M.A. Kofoed, D.A. Wampler, A.S. Pandey, J.W. Peters, S.A. Ensign, J. Biol. Chem. 193 (2011) 4904-4913.

[38] D.D. Clark, J.R. Allen, S.A. Ensign, Biochemistry 39 (2000) 1294-1304.

[39] A.S. Pandey, D.W. Mulder, S.A. Ensign, J.W. Peters, FEBS Letters 585 (2011) 459-464.

[40]   A.S. Pandey, B. Nocek, D.D. Clark, S.A. Ensign, J.W. Peters, Biochemistry 45 (2006) 113-120.

[41]   J.R. Allen, D.D. Clark, J.G. Krum, S.A. Ensign, Proc. Natl Acad Sci USA 96 (1999) 8432-8437.

[42]   R.K. Thauer, Microbiology 144 (Pt 9) (1998) 2377-2406.

[43]   R.S. Wolfe, Annual Review of Microbiology 45 (1991) 1-35.

[44]   Y. Kallberg, U. Oppermann, B. Persson, FEBS J. 277 (2010) 2375-2386.

[45]   B. Persson, Y. Kallberg, J.E. Bray, E. Bruford, S.L. Dellaporta, A.D. Favia, R.G. Duarte, H. Jornvall, K.L. Kavanagh, N. Kedishvili, M. Kisiela, E. Maserk, R. Mindnich, S. Orchard, T.M. Penning, J.M. Thornton, J. Adamski, U. Oppermann, Chemico-Biological Interactions 178 (2009) 94-98.

[46]   S. Banta, M. Boston, A. Jarnagin, S. Anderson, Metabolic Engineering 4 (2002) 273-284.

[47]   S. Banta, B.A. Swanson, S. Wu, A. Jarnagin, S. Anderson, Biochemistry 41 (2002) 6226-6236.

[48]   T.W. Johannes, R.D. Woodyer, H.M. Zhao, Applied and Environmental Microbiology 71 (2005) 5728-5734.

[49]   H.J. Moon, M.K. Tiwari, R. Singh, Y.C. Kang, J.K. Lee, Applied and Environmental Microbiology 78 (2012) 3079-3086.

[50]   R. Woodyer, T. Johannes, W. van der Donk, H.M. Zhao, Abstracts of Papers of the American Chemical Society 229 (2005) U229-U229.

[51]   R. Montalvo-Gonzalez, D. Chavez, G. Aguirre, M. Parra-Hake, R. Somanathan, Synthetic Communications 39 (2009) 2737-2746.

[52]   V. Sotolongo, D.V. Johnson, D. Wahnon, I.W. Wainer, Chirality 11 (1999) 39-45.

[53]   D.D. Clark, J.M. Boyd, S.A. Ensign, Biochemistry 43 (2004) 6763-6771.

[54]   D.D. Clark, S.A. Ensign, Biochemistry 41 (2002) 2727-2740.

CHAPTER 2

CRYSTAL STRUCTURES OF *S*-HPCDH REVEAL DETERMINANTS OF

STEREOSPECIFICITY FOR *R*-AND *S*-HYDROXYPROPYL-COENZYME M

DEHYDROGENASES[1,2]

ABSTRACT

(*R*)- and (*S*)-hydroxypropyl-coenzyme M dehydrogenases (*R*- and *S*-HPCDH) are stereospecific enzymes that are central to the metabolism of propylene and epoxide in *Xanthobacter autotrophicus*. The bacterium produces *R*- and *S*-HPCDH simultaneously to facilitate transformation of *R*- and *S*-enantiomers of epoxypropane to a common achiral product 2-ketopropyl-CoM (2-KPC). Both *R*- and *S*-HPCDH are highly specific for their respective substrates as each enzyme displays less than 0.5% activity with the opposite substrate isomer. In order to elucidate the structural basis for stereospecificity displayed by *R*- and *S*-HPCDH we have determined substrate bound crystal structures of *S*-HPCDH to 1.6 Å resolution. Comparisons to the previously reported product-bound structure of *R*-HPCDH reveal that although the placement of catalytic residues within the active site of each enzyme is nearly identical, structural differences in the surrounding area provide each enzyme with a distinct substrate binding pocket. These structures demonstrate how chiral discrimination by *R*- and *S*-HPCDH results from alternative binding of the distal end of substrates within each substrate binding pocket.

INTRODUCTION

Propylene metabolism in the soil bacterium *Xanthobacter autotrophicus* strain Py2 involves the conversion of chiral epoxides to acetoactetate [1-4]. The initial steps in the pathway involve the epoxidation of propylene followed by nucleophilic addition of the small cofactor coenzyme M to the *R*-and *S*-enantiomers of epoxypropane, resulting in a racemic mixture of *R*- and *S*-hydroxypropyl CoM (*R*- and *S*-HPC, respectively). *R*- and *S*-HPC are then converted into a common achiral product 2-ketopropyl-CoM (2-KPC) using two highly stereospecific dehydrogenases, (*R*)-hydroxypropyl-coenzyme M dehydrogenase (*R*-HPCDH) and (*S*)-hydroxypropyl-coenzyme M dehydrogenase (*S*-HPCDH) (Figure 2-1). A remarkable feature of *R*- and *S*-HPCDH (40 % sequence identity) is the extremely high substrate stereospecificity exhibited by these enzymes (each enzyme displays less than 0.5% activity with the opposite substrate isomer [2]), although each enzyme catalyzes the same chemical reaction.
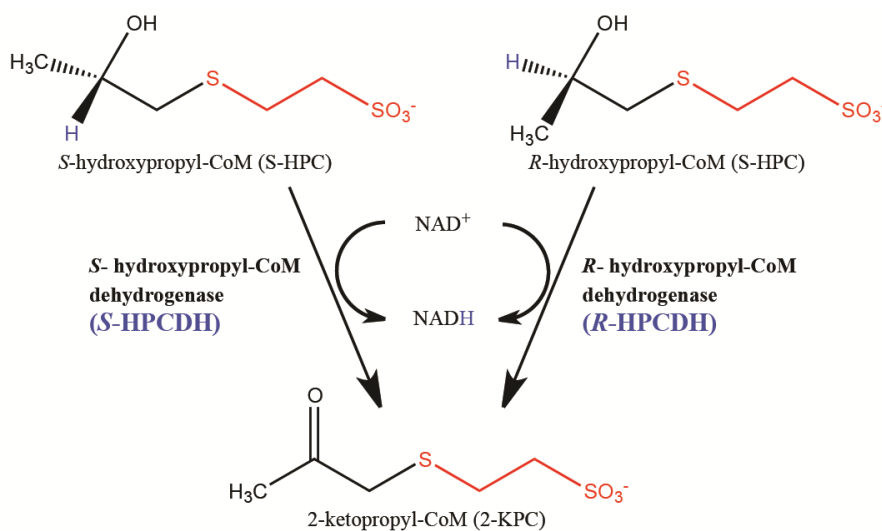
Figure 2-1. *R*-HPCDH and *S*-HPCDH convert *R*-and *S*-hydroxypropyl CoM (*R*- and *S*-HPCDH) to a common product 2-ketopropyl-CoM (2-KPC).

R- and *S*-HPCDH have been characterized as members of the "classical" short-chain dehydrogenase/reductase (SDR) superfamily of enzymes. SDR enzymes are found throughout eukaryotic and prokaryotic species and constitute one of the largest enzyme families known with over 60,000 members annotated in sequence databases [5]. SDR enzymes are categorized in several enzyme classes including lyases, isomerases, and oxidoreductases, and they act on a broad range of substrates including steroids, sugars, xenobiotics, and aliphatic alcohols [5-7]. Despite the relatively low sequence identity between SDR enzymes (15-30% sequence identity), structural studies have revealed a common scaffold that includes N- and C- terminal domains [8]. The N-terminal domain is a highly conserved Rossman-fold structure that contains a catalytic triad/tetrad (Tyr-Lys-Ser-Asn in the case of *R*- and *S*-HPCDH) and a GxxxGxG motif involved in NAD(P)(H) binding . The C-terminal domain is more variable, both in terms of sequence and structure, but generally contains a substrate binding loop region that is involved in substrate binding and specificity [8-11]. Although numerous SDR superfamily enzymes exhibit stereospecificity, few stereospecific enzymes have been found to function in concert with a homologous partner-enzyme. *R*- and *S*-HPCDH represent an ideal model system to examine stereospecificity because they perform identical chemistry in the same pathway, yet both enzymes are highly specific for their respective substrates (*R*-HPC or *S*-HPC).

Several biochemical studies, and a product bound structure of *R*-HPCDH has led to a proposed mechanism for *R*- and *S*-HPCDH [12-16]. In this mechanism, the role of the catalytic tetrad residues (Tyr-Lys-Ser-Asn) and the chemistry around the chiral carbon (C2) is analogous to the general mechanism described for other SDR enzymes [6,

17]. Specifically, the catalytic tetrad serine assists in positioning the hydroxyl group of the substrate near the catalytic tyrosine. The tyrosine residue in its deprotonated form acts as a general base for proton abstraction from the substrate hydroxyl group, and a hydride is transferred from the C2 carbon of the substrate to NAD^+. Lysine has a dual role in coordinating NAD(H) and in lowering the pK_a of the catalytic tyrosine hydroxyl group through a proton relay involving water molecules and the catalytic tetrad asparagine. The product bound structure of *R*-HPCDH reveals a pair of positively charged amino acids (R152, R196) that coordinate the sulfonate tail of 2-KPC [16]. Kinetic analysis of *S*-HPCDH mutants suggest that positively charged amino acids (R211 and K214) may also contribute to sulfonate binding in the *S*-enzyme [15]. It has been proposed that in each enzyme the sulfonate binding residues are used to properly orient the reactive groups of the substrate at the catalytic site [13, 15, 16]. Although a crystal structure of *R*-HPCDH is available, there are no structures of *S*-HPCDH. Furthermore, no substrate bound structures of *R*- or *S*-HPCDH are currently available. Here we report high-resolution binary (NAD^+ bound) and ternary (*S*-HPC/NADH bound) complex crystal structures of *S*-HPCDH. These structures provide insight into the structural architecture of the active site and substrate binding path of *S*-HPCDH. In addition, the structures allow for the first structural comparison between *S*-HPCDH and *R*-HPCDH, clarifying structural mechanisms of stereospecificity employed by these enzymes.

MATERIALS AND METHODS

Expression, purification, and crystallization

Three sequence variants of *S*-HPCDH exist on a single megaplasmid in *Xanthobacter autotrophicus* strain Py2 [18]. The most well behaved *S*-HPCDH variant *in vitro* (previously designated as *S*-HPCDH3) was expressed and purified in *E. coli* as previously described [15]. Protein was concentrated to 13 mg/ml in the presence of 25 % glycerol and stored at -80 °C. Protein was thawed on ice and crystallized by sitting drop vapor diffusion at 4 °C at a 1:1 protein:well drop ratio in a well solution consisting of 0.1 M Bis-Tris, pH 6.5, 0.35M ammonium acetate and 27 % polyethylene glycol 3350. Crystals of *S*-HPCDH bound to NAD^+ (binary complex) were grown in the presence of 50 μM NAD^+ and 50 μM *S*-HPC. The crystal was transferred directly into a cryoprotectant composed of the mother liquor, 50 μM *S*-HPC and 10 % glycerol, and flash-cooled in liquid nitrogen. *S*-HPC/NADH bound crystals (ternary complex) were grown in the presence of 50 μM NADH and 50 μM *S*-HPC. The crystal was transferred directly into a cryoprotectant composed of the mother liquor without ammonium acetate, 140 μM *S*-HPC and 10 % glycerol, and flash-cooled in liquid nitrogen.

Data collection

Diffraction data were generated and collected using a home-source x-ray generator (Rigaku RU-200 and MicroMax-007HF) and detector (Rigaku R-AXIS IV++). Data were processed using the HKL2000 program suite [19]. The *S*-HPCDH crystals belong to space group $P2_12_12$ with unit cell dimensions of a= 116 Å, b= 128 Å, and c= 58 Å. Data statistics are summarized in Table 2-1.

Structure determination, model building and refinement

The binary complex S-HPCDH structure was solved by molecular replacement using the *R*-HPCDH structure ([16], PDB code, 2cfc) as a molecular replacement search model. The ternary complex structure was solved by molecular replacement using the refined binary complex as a search model. Molecular replacement was performed using Phaser [20] from the Phenix program suite [21].

Manual building of the *S*-HPCDH structures was done using the program COOT [22]. Electron density maps were high in quality, permitting unambiguous identification and positioning of the majority of amino acids in each monomer of asymmetric unit. Discontinuous electron density prohibited modeling of the first two N-terminal amino acids in each monomer, and a loop region beginning at position 200 and extending to positions 204-207, depending on the monomer. The one exception is one monomer in the ternary complex structure, in which the entire loop region is observed. Although *S*-HPC was included in the crystallization and cryo solutions for the NAD$^+$ bound structure, no electron density was observed for *S*-HPC. Instead, an acetate molecule (originating from the crystallization buffer) was observed at the active site, essentially mimicking the position of the *S*-HPC reactive center. Refinement of the *S*-HPCDH structures were performed using phenix.refine [23] from the Phenix program suite [21]. Geometry statistics were calculated using MolProbity [24,25]. The refined coordinates and structure factors were deposited in the protein data bank under accession ID 4GH5 (binary complex) and 4ITU (ternary complex). Figures were made using Pymol [26].

Table 2-1. Data collection and refinement statistics for *S*-HPCDH structures. Values in parentheses correspond to those in the outer resolution shell.

S-HPCDH structure	Binary complex	Ternary complex
Data collection		
Wavelength (Å)	1.5418	1.5418
Resolution range (Å)	35.0 - 1.60	35.0 - 1.60
Outer shell (Å)	1.66 – 1.60	1.66 – 1.60
No. of reflections		
unique	113,574	108,385
total	708,549	1,179,096
Average redundancy	6.2 (2.8)	10.9 (7.9)
Mean $I/\sigma(I)$	25.8 (1.9)	37.0 (3.3)
Completeness (%)	98.3 (86.0)	93.7 (63.9)
R_{sym} (%)[a]	6.3 (38.9)	5.7 (37.4)
Space group	$P2_12_12$	$P2_12_12$
# of protein molecules/ asym. unit	4	4
Unit cell dimensions		
a, b, c (Å)	116.1, 127.9, 58.5	116.0, 128.4, 58.4
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Refinement		
R_{work}/R_{free} (%)[b]	17.1/19.9	16.4/19.5
Atoms in the structure	8,114	8,098
protein	6,895	6,977
waters	1027	931
ligands	192	220
Average B factor (Å^2)	13.8	14.3
protein	12.5	13.2
water	22.5	23.4
NAD$^+$/NADH	11.0	11.0
S-HPC	---	22.4
rmsd bond (Å)/angle (°)	0.008/1.104	0.007/1.114
Protein geometry [c]		
Ramachandran outliers (%)	0	0
Ramachandran favored (%)	98.9	98.1
Rotamer outliers (%)	0	0.9
PDB ID	4GH5	4ITU

[a] $R_{sym} = (\Sigma|(I - \langle I \rangle)|)/(\Sigma I)$, where $\langle I \rangle$ is the average intensity of multiple measurements.

[b] $R_{work} = (\Sigma|F_{obs} - F_{calc}|)/(\Sigma|F_{obs}|)$ and is calculated using all data; R_{free} is the R-factor based on 5% of the data excluded from refinement.

[c] Ramachandran statistics were calculated using the MolProbity server [24,25].

RESULTS

Overall structure of *S*-HPCDH

The binary and ternary crystal structure of *S*-HPCDH both adopt nearly identical protein conformations (RMSD = 0.091 Å over 904 residues). The quaternary structure of *S*-HPCDH is a homo-tetramer (Figure 2-2), similar to the previously reported structure of *R*-HPCDH [16] and other SDR proteins [8]. The tetramer is assembled such that each subunit interacts with the other three subunits. As shown in Figure 2 (for clarity, subunit A is described, but analogous interactions are observed for each subunit in the tetramer), major interactions between subunit A and the other subunits include a four helix bundle between the long helices α*D* and α*E* of subunits A and C (Figure 2-2, box 1) and several hydrophobic interactions between the α*F* helices of subunits A and B (Figure 2-2, box 2). An additional interaction is also observed at the N-termini of each subunit in which several water molecules are coordinated by R4 and Q237 from each subunit (Figure 2-2, box 3). A similar N-terminal interaction has not been observed in other SDR family structures including *R*-HPCDH, and to our knowledge is a unique feature of *S*-HPCDH. The tetramer is further stabilized by a crossover interaction between diagonally spaced subunits (subunits A and D in Figure 2-2, box 4). This interaction involves adjacent C-terminal carboxylate groups of I258, which coordinate two water molecules. A similar interaction between C-terminal carboxylate groups was also observed in the *R*-HPCDH structure [16].
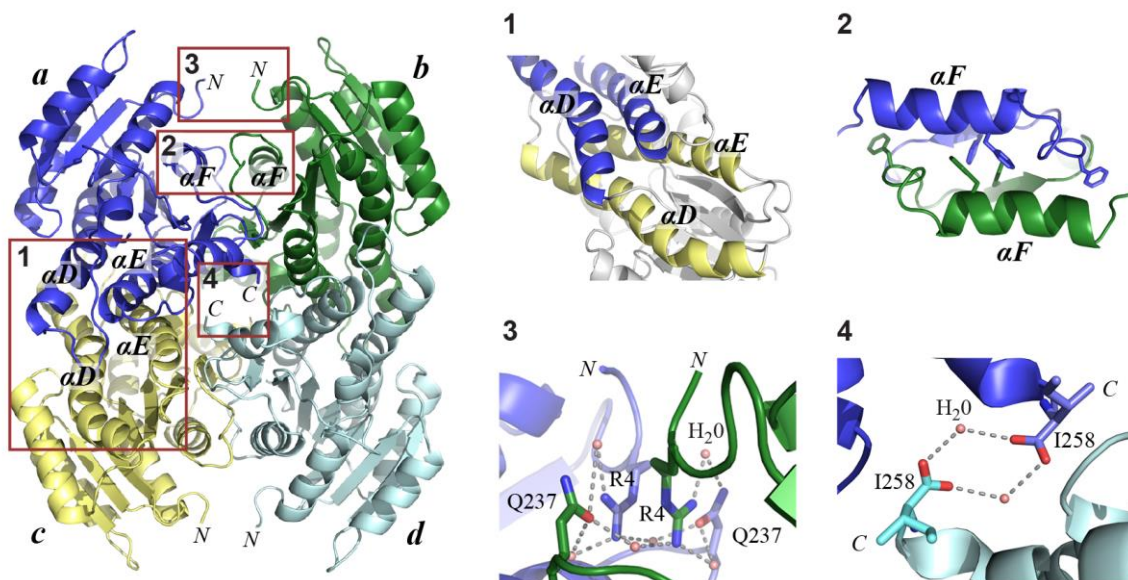
Figure 2-2. Tetrameric structure of S-HPCDH. Each subunit of S-HPCDH (labeled *a-d*) interacts directly with each of the other three subunits. Direct interactions between subunit *a* and the other three subunits are indicated with boxes labeled 1-4.

Each monomer of *S*-HPCDH has a bipartite domain structure with the active site located at the bottom of a cleft formed between the N- and C- terminal domains (Figure 2-3). The large N-terminal domain comprises the core of the structure and is in the form of a NAD(H) binding Rossman-fold motif common to SDR enzymes [8]. The dinucleotide binding Rossman-fold found in all SDR enzymes is composed of a central twisted seven-stranded parallel β-sheet (β*A*-β*G*) that is flanked on each side by three α-helices (α*A*-α*F*). In *S*-HPCDH, α*D* and α*E* are located on the same side of the Rossman fold and are about twice the length of the other α-helices, resulting in an overall N-terminal domain with two α-helices that are much longer on one side (Figure 2-3A). Two loops connect α*D* and α*E* to their respective β-strands (α*D* and α*E* to β*D* and β*E*, respectively), and these long loops constitute one side of the active site cleft.
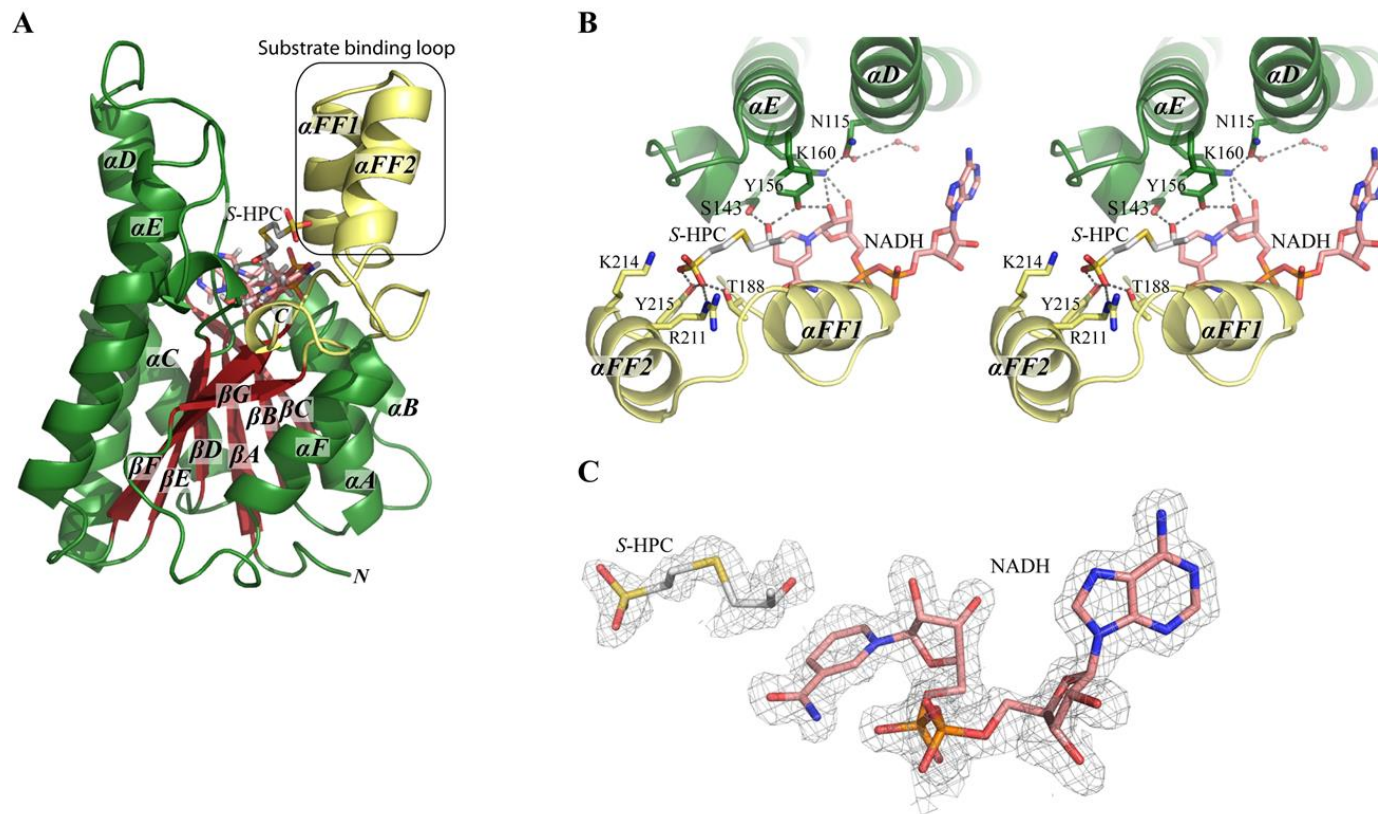
Figure 2-3. Monomeric structure of *S*-HPCDH. (A) Cartoon representation of one of the monomers of *S*-HPCDH ternary complex. The N- terminal Rossman fold domain is depicted with β strands colored red, loops colored green, and α-helices colored green. The C-terminal domain is colored yellow. NADH (pink) and *S*-HPC (gray) molecules are shown as sticks. The region spanning helices αFF1 and αFF2 is commonly referred to among SDR enzymes as the substrate binding loop. (B) Catalytic site of *S*-HPCDH ternary complex. *S*-HPC is bound with its reactive OH group coordinated between S143 and Y156. The sulfonate tail of *S*-HPC is coordinated by T188, R211 and Y215. Water molecules that contribute to a proposed proton relay are shown as red spheres. (C) A 2Fo-Fc omit map surrounding *S*-HPC and NADH in the ternary structure, contoured at 1 σ.

The last several C-terminal residues (following β*G*) and an extensive loop

inserted between β*F* and α*F* fold into a small C-terminal domain. The C-terminal domain

contains two additional α-helices (*αFF1, αFF2*) that protrude upward forming two sides

of a lobe and the other side of the active site cleft. This lobe is referred to as the substrate

binding loop, following the nomenclature employed for other SDR enzymes [9-11]. In

the SDR superfamily, the substrate binding loop exhibits a high degree of sequence

variation between proteins and often contains residues important for substrate binding

[11]. In the *S*-HPCDH structure, R211 and K214 (which have been implicated in

substrate binding) are both located on α*FF2* (Figure 2-3B). The residues connecting

α*FF1* to α*FF2* (amino acids 200-204) are observed in one monomer of the ternary

complex, but appear to be somewhat flexible as indicated by high b-factors and

discontinuous electron density for the rest of the monomers in the ternary structure and

all of those in the binary structure.

Active site and substrate binding

The active site of *S*-HPCDH is located in a cleft between the N- and C-terminal

domains (Figure 2-3). In both the binary and ternary structures, NAD^+/NADH bind *S*-

HPCDH in an extended conformation along an axis parallel to a plane created by the C-

termini of the central β-sheet. Specific protein interactions are consistent with NAD(H)

binding observed in *R*-HPCDH and other SDR enzymes [7, 16]. In contrast with the *R*-

HPCDH structure, strong electron density is observed for all copies of the entire

NAD^+/NADH molecule, including the nicotinamide ring (Figure 2-3C), with average b-

factors of 11.0 $Å^2$ for each complex. Since the crystallization experiments for both the

binary and ternary complexes were performed aerobically, it is possible that NADH was oxidized to NAD^+, which could be difficult to distinguish at 1.6 Å resolution. If NAD^+ were present at the active site of the ternary complex (which was co-crystallized with *S*-HPC and NADH), then we would expect a reaction to proceed converting *S*-HPC to 2-KPC. However, *S*-HPC is clearly observed within the active site of the ternary structure, suggesting that NADH rather than NAD^+ is bound.

In the ternary complex, *S*-HPC binds in a pocket that extends from above the nicotinamide ring of NADH toward α*FF*2 of the substrate binding loop (Figure 2-3B). Each of the four *S*-HPC molecules is well ordered along its entire length (Figure 2-3C) with average b-factors of 22.4 $Å^2$. The sulfonate tail of *S*-HPC is coordinated by T188, R211 and Y215 (Figure 2-3B). K214 is also located near the sulfonate tail, but electron density for the terminal side chain atoms is weak and no direct interaction is observed. One side of the *S*-HPC binding pocket is entirely composed of residues located on α*FF1* (M194, G195, L198, and L199), and the other side of the pocket is composed of hydrophobic residues (V144, A145 and I150) provided by the N-terminal domain. On the reactive end of the *S*-HPC molecule, the OH group is coordinated by direct interactions with the S143 and Y156 side chains. This orientation places the chiral center of *S*-HPC directly above the nicotinamide ring of NADH (Figure 2-3B,C).

The observed orientation of *S*-HPC, NADH and the catalytic tetrad (Y156, K160, S143, N115) in the ternary complex is consistent with the catalytic mechanism described for other SDR enzymes [6, 17]. Specifically, the general acid/base Y156 is positioned between S143 and K160, with its catalytic hydroxyl forming hydrogen bonds with the 2' hydroxyl of the ribose ring of NADH and the OH group of *S*-HPC (Figure 2-3B). The ε

amino group of K160 forms a bidentate hydrogen bond with the 2' and 3' hydroxyl groups of the nicotinamide ribose. A water molecule is also observed to be coordinated between N115 and K160. This water molecule is further coordinated to a string of other waters, suggesting a proton relay chain connecting the bulk solvent to the active site tyrosine using N115, K160 and NAD(H) as has been proposed for other SDR family enzymes [6, 27, 28]. Notably, no conformational differences are observed between binary and ternary structures for the catalytic tetrad side chains or the residues surrounding the *S*-HPC binding site, with the exception of R211.This observation suggests that the substrate binding pocket is fully configured for *S*-HPC binding once NAD(H) is bound, and may be consistent with an ordered binding mechanism for *S*-HPCDH.

Structural comparison of *S*-HPCDH and *R*-HPCDH

The overall structure of *S*-HPCDH is highly homologous to the *R*-HPCDH structure and other SDR family enzymes. The arrangement of the catalytic tetrad is superimposable between the *R*- and *S*- enzymes (RMSD=0.24 Å). However, superposition of the *R*-HPCDH and *S*-HPCDH structures reveals three regions within each monomer that display different structural conformations: a small shift in the C-terminal end of αB and two larger conformational rearrangements in the substrate binding region above the active site (Figure 2-4A, B). The shift observed in αB does not appear to impact the structure of the active site or the substrate binding region, and likely has no functional consequence. The larger rearrangements directly impact the substrate binding region and are each centered around a three amino acid insertion/deletion located on a loop connecting βD and αD, and on a loop connecting αFF1 and αFF2.
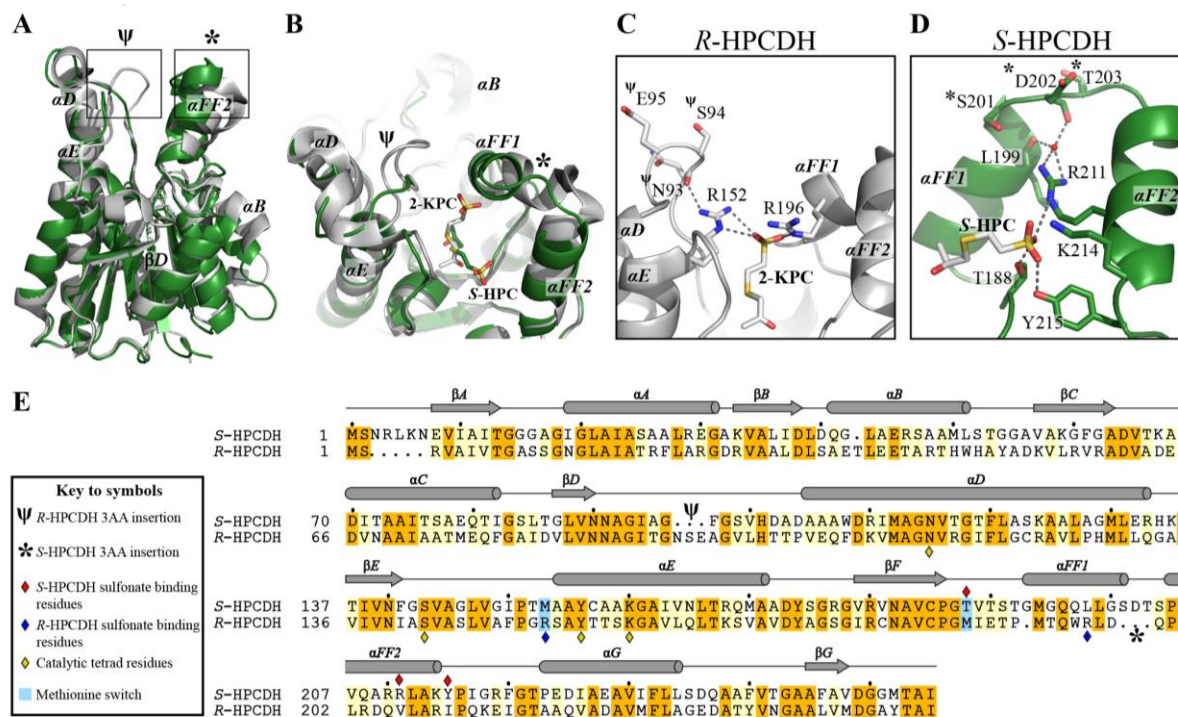
Figure 2-4. Structural differences between *R*-HPCDH and *S*-HPCDH. The largest differences center around 3-residue insertions within each enzyme. Insertions are indicated by Ψ in *R*-HPCDH and by * in *S*-HPCDH. (A) Superposition of *R*-HPCDH (gray) and *S*-HPCDH (green). Sites of major structural differences are indicated with boxes. (B) Top view (as compared to view in panel (A)) of superimposed *R*- and *S*-HPCDH structures. Bound molecules (*S*-HPC in *S*-HPCDH and 2-KPC in *R*-HPCDH) are shown as sticks. Carbon atoms are colored green for *S*-HPC and gray for 2-KPC (C) View of the 3-residue insertion in *R*-HPCDH (N93, S94, E95), and binding of the product 2-KPC at the active site. The backbone carbonyl oxygen atom for N93 is also shown. (D) View of the 3-residue insertion in *S*-HPCDH (S201,D202,T203) and binding of the substrate *S*-HPC at the active site. The backbone carbonyl oxygen atoms for L199 and D202 are also shown. (E) Sequence alignment of *R*- and *S*-HPCDH. Identical residues are shaded orange; similar residues are shaded yellow. The positions of the 3-residue insertions are indicated as Ψ in *R*-HPCDH and * in *S*-HPCDH. Sulfonate binding residues and catalytic tetrad residues are indicated with diamonds. Residues that contribute to the methionine switch are shaded blue.

Repositioning of the loop between β*D* and α*D* results from a loss of three amino acids in

S-HPCDH (Figure 4E). In the *R*-HPCDH structure, these three residues (N93, S94, and

E95) flank the entrance of the substrate binding pocket, and the main chain carbonyl

oxygen from one of the residues (N93) forms a hydrogen bond to the amide nitrogen of

R152 (Figure 4C). R152 is one of two arginine residues used by the *R*-enzyme to bind

the sulfonate tail of endogenous substrates and the product 2-KPC [13, 16]. The deletion

of these three residues in *S*-HPCDH collapses and reconfigures this substrate binding

region such that the binding pocket found in *R*-HPCDH is lost.

Conversely, three amino acids (residues S201, D202, and T203) are inserted in

the loop between α*FF1* and α*FF2* in *S*-HPCDH (Figure 4D, E). Although the loop

appears to be mobile in the crystal (weak electron density prevented complete modeling

of the loop in 3 of 4 monomers in the ternary structure), the net effect is an opening up of

the space between α*FF1* and α*FF2*, as compared to *R*-HPCDH (Figure 4A, B). The

increased separation between α*FF1* and α*FF2*, combined with amino acid substitutions

in the surrounding region, facilitates formation of a new substrate binding pocket. Main

chain carbonyl oxygen atoms (L199 and D202) from the loop between α*FF1* and α*FF2*

help orient R211 for *S*-HPC sulfonate binding through water mediated hydrogen bonds in

a manner analogous to the loop stabilization of R196 observed in the *R*-HPCDH substrate

binding pocket.

While the three residue insertions/deletions play the most dramatic role in

reconfiguring and repositioning the substrate binding pockets of *R*- and *S*-HPCDH, other

sequence differences throughout the region also contribute. Most notably is a

"methionine switch" where M153 in *S*-HPCDH replaces the sulfonate binding residue
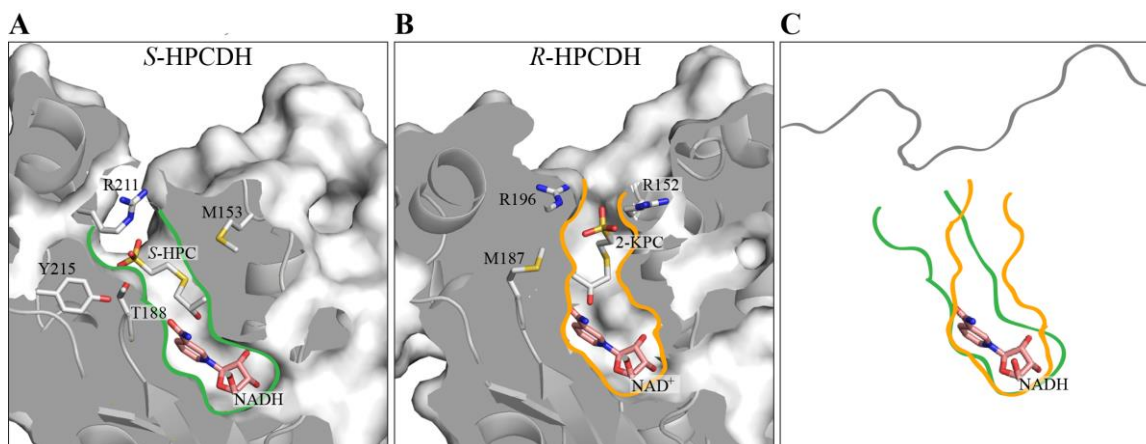
Figure 2-5. Distinct substrate binding pockets are observed in *R*-HPCDH- and *S*-HPCDH. (A) Sliced view of *S*-HPCDH bound to the substrate *S*-HPC. The shape of the substrate binding pocket is outlined in green. (B) Sliced view of *R*-HPCDH bound to the product 2-KPC. The shape of the substrate binding pocket in *R*-HPCDH is outlined in orange. (C) Superimposed outlines of the two substrate binding pockets highlights the distinct shape and directionality of the distal ends of each substrate binding pocket

R152 of *R*-HPCDH. Conversely, M187 in *R*-HPCDH replaces the sulfonate binding

residue T188 of *S*-HPCDH. In each case, the methionine residue creates a steric block

that closes off the path to the sulfonate binding site observed in the other enzyme (Figure

2-5). The net effect of all these amino acid substitutions and structural rearrangements is

the formation of two distinct substrate binding pockets that are rotated by approximately

60° with respect to each other (Figure 2-5). In each case, the pocket contains residues that

can bind the sulfonate moiety at the distal end of the substrate, which differentially

orients the reactive end of the substrate toward the structurally conserved catalytic site.

DISCUSSION

R-and *S*-HPCDH represent a rare set of stereospecific enzymes that both function

in a single metabolic pathway to produce the same product from different enantiomeric

substrates. This provides an ideal system for understanding mechanisms of stereospecificity. Previous biochemical studies have shown that stereochemical discrimination is governed by different kinetic parameters in each enzyme [15]. However, a structural understanding of this system has been limited because only the *R*-HPCDH structure was available [15, 16].

In order to more completely understand the structural mechanisms used to achieve stereospecificity we have now determined the crystal structure of *S*-HPCDH. It was previously proposed that specificity of *R*- and *S*-HPCDH is controlled in large part by alternative locations of sulfonate binding residues in each enzyme [13, 15, 16]. In the current *S*-HPCDH structure we confirmed that there was alternative placement of substrate sulfonate binding residues when compared to sulfonate binding residues in the *R*-HPCDH structure. Not foreseen, however, was how sequence variations within the active site cleft of *R*- and *S*-HPCDH provide each enzyme with distinct substrate binding pockets. Significantly, the unique arrangement of these stereospecific pockets doesn't appear to arise through a simple inversion of the active site. Rather, each pocket adopts unique shape and binding characteristics at the site of sulfonate binding (Figure 2-5C).

To visualize how the specific shape and orientation of the substrate binding pocket in *S*-HPCDH affects specificity, we modeled *R*-HPC into the *S*-HPC binding pocket of *S*-HPCDH (Figure 2-6). Effective catalysis for either substrate (*S*-HPC or *R*-HPC) requires that the C2 (chiral carbon) hydroxyl group is positioned for proton abstraction between S143 and Y156, and the C2 hydrogen is pointed towards the b-face of the nicotinamide ring of NAD^+ making it available for hydride abstraction [17]. In the
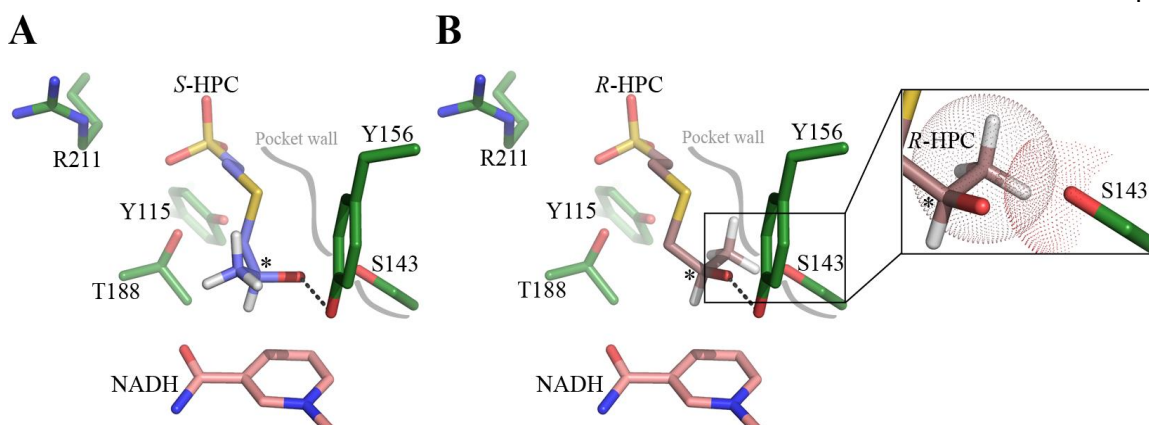
Figure 2-6. *S*-HPC and a model of *R*-HPC within the active site of the *S*-HPCDH ternary complex. Both substrates are shown with their C2 hydroxyl group and C2 hydrogen atoms oriented for effective catalysis. The C2 carbon is indicated with an asterisk (*). Residues within the active site are displayed with carbon atoms colored green. Carbon atoms of NADH are colored pink. The surface of the substrate binding pocket is shown as a gray line. (A) The C2 methyl group of *S*-HPC is spatially accommodated within the active site. (B) When *R*-HPC is properly aligned for catalysis, the C2 methyl group is directed toward the pocket wall and clashes directly with S143. Insert depicts the potential clash of the C2 methyl group of *R*-HPC with S143 of *S*-HPCDH. The molecular surface of the methyl group carbon atom (C1) of *R*-HPC and the side chain oxygen atom of S143 are shown as dot representation.

ternary complex, the C2 methyl group of *S*-HPC is positioned within an open region of the substrate binding pocket where it is free of steric hindrance (Figure 2-6A). In contrast, when *R*-HPC is modeled into the catalytic site in the same fashion, the C2 methyl group of *R*-HPC is oriented directly toward the wall of the substrate binding pocket, resulting in significant steric clashes with the side chain of the catalytic serine (S143) (Figure 2-6B). This observation is consistent with the 290-fold increase in K_m observed when *R*-HPC is used as a substrate for *S*-HPCDH [15], and provides a structural basis for stereospecificity displayed by *S*-HPCDH.

In summary, several strategies are employed by the *R*- and *S*- HPCDH system to achieve stereospecificity. The enzymes utilize a common catalytic site but reorient the

substrate binding path as it moves away from the active site. Reorientation of the binding path is achieved by a three residue insertion/deletion and alternately expands or collapses the binding pocket and assists in coordinating the distal end of the substrate. Additional mutations that include a methionine switch further modulate access to and the shape of the substrate binding path. These strategies may provide a basis for rational design of stereospecificity in other systems.

REFERENCES

[1]   J.R. Allen, S.A. Ensign, J. Biol. Chem. 272 (1997) 32121-32128.

[2]   J.R. Allen, D.D. Clark, J.G. Krum, S.A. Ensign, Proc. Natl Acad Sci USA 96 (1999) 8432-8437.

[3]   J.R. Allen, S.A. Ensign, Biochemistry 38 (1999) 247-256.

[4]   S.A. Ensign, Biochemistry 40 (2001) 5845-5853.

[5]   Y. Kallberg, U. Oppermann, B. Persson, Febs J 277 (2010) 2375-2386.

[6]   C. Filling, K.D. Berndt, J. Benach, S. Knapp, T. Prozorovski, E. Nordling, R. Ladenstein, H. Jornvall, U. Oppermann, J. Biol. Chem. 277 (2002) 25677-25684.

[7]   B. Persson, Y. Kallberg, J.E. Bray, E. Bruford, S.L. Dellaporta, A.D. Favia, R.G. Duarte, H. Jornvall, K.L. Kavanagh, N. Kedishvili, M. Kisiela, E. Maserk, R. Mindnich, S. Orchard, T.M. Penning, J.M. Thornton, J. Adamski, U. Oppermann, Chemico-Biological Interactions 178 (2009) 94-98.

[8]   K. Kavanagh, H. Jornvall, B. Persson, U. Oppermann, Cell Mol. Life Sci. 65 (2008) 3895-3906.

[9]   C. Grimm, E. Maser, E. Mobus, G. Klebe, K. Reuter, R. Ficner, J. of Biol. Chem. 275 (2000) 41333-41339.

[10]   R. Yamazawa, Y. Nakajima, K. Mushiake, T. Yoshimoto, K. Ito, J Biochemistry 149 (2011) 701-712.

[11]   N. Tanaka, T. Nonaka, K.T. Nakamura, A. Hara, Current Organic Chemistry 5 (2001) 89-111.

[12]   D.D. Clark, S.A. Ensign, Biochemistry 41 (2002) 2727-2740.

[13]   D.D. Clark, J.M. Boyd, S.A. Ensign, Biochemistry 43 (2004) 6763-6771.

[14]   A.M. Krishnakumar, D. Sliwa, J.A. Endrizzi, E.S. Boyd, S.A. Ensign, J.W. Peters, Microbiol Molecular Biology R 72 (2008) 445-+.

[15]   D.A. Sliwa, A.M. Krishnakumar, J.W. Peters, S.A. Ensign, Biochemistry 49 (2010) 3487-3498.

[16]   A.M. Krishnakumar, B.P. Nocek, D.D. Clark, S.A. Ensign, J.W. Peters, Biochemistry 45 (2006) 8831-8840.

[17]   H. Jornvall, B. Persson, M. Krook, S. Atrian, R. Gonzalezduarte, J. Jeffery, D. Ghosh, Biochemistry 34 (1995) 6003-6013.

[18]   J.G. Krum, S.A. Ensign, J. of Bacteriology 183 (2001) 2172-2177.

[19]   Z. Otwinowski, W. Minor, Methods Enzymology 276 (1997) 307-326.

[20]   A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, J. of Applied Crystallography 40 (2007) 658-674.

[21]   P.D. Adams, P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W.

Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, Acta Crystallographica 66 (2010) 213-221.

[22]   P. Emsley, K. Cowtan, Acta Crystallographica 60 (2004) 2126-2132.

[23]   P.V. Afonine, R.W. Grosse-Kunstleve, & P.D. Adams., CCP4 Newsl. 42, contribution 8. (2005).

[24]   V.B. Chen, W.B. Arendall, 3rd, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, Acta Crystallograph. 66 (2010) 12-21.

[25]   I.W. Davis, A. Leaver-Fay, V.B. Chen, J.N. Block, G.J. Kapral, X. Wang, L.W. Murray, W.B. Arendall, 3rd, J. Snoeyink, J.S. Richardson, D.C. Richardson, Nucl. Acids Res. 35 (2007) W375-383.

[26]   The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.

[27]   U. Oppermann, C. Filling, M. Hult, N. Shafqat, X.Q. Wu, M. Lindh, J. Shafqat, E. Nordling, Y. Kallberg, B. Persson, H. Jornvall, Chemico-Biological Interact 143 (2003) 247-253.

[28]   A. Koumanov, J. Benach, S. Atrian, R. Gonzalez-Duarte, A. Karshikoff, R. Ladenstein, Proteins 51 (2003) 289-298.

CHAPTER 3

INTRODUCTION II: AIR PROTEINS AND RNA REGULATION

In eukaryotes, RNA levels and RNA quality are tightly regulated

Eukaryotic cells produce a multitude of different types of RNA. In addition to protein-coding mRNA, the transcriptome is also comprised of many non-coding RNAs including transfer RNAs (tRNA), ribosomal RNAs (rRNA), small nuclear and nucleolar RNAs (snRNA and snoRNA), cryptic unstable transcripts (CUT's), and intergenic sequences (introns) [1-3]. Most of these non-coding RNAs have direct regulatory roles in gene expression and translation. After transcription most if not all RNAs are subject to a variety of maturation events to become fully functional RNAs. These posttranscriptional modifications often first involve folding of the RNAs into intricate three dimensional structures, followed by a variety of processing steps such as internal cleavage, end trimming, and covalent modifications. The cellular levels and quality of RNAs produced in the cell must be constantly monitored and controlled in order to ensure proper cell function and survival. In humans, defects in RNA processing and regulation have been linked to many disease states including neurodegenerative diseases and cancer [4-6], highlighting the importance of understanding the biological mechanisms controlling RNA maturation and stability.

In the nucleus of eukaryotic cells RNA biogenesis and stability is regulated by RNA quality control and surveillance pathways. These pathways are comprised of a variety of molecular machines including ribonucleases, RNA binding proteins, and many others, which ensure that only properly processed RNAs are stabilized and all other non-

functional RNAs – including aberrantly processed RNAs and end-trimmings – are degraded. In yeast, the nuclear TRAMP complex is required for RNA surveillance and 3' maturation of many RNAs [7-10]. TRAMP is a three-protein complex composed of an RNA binding protein (Air1 or Air2), a RNA helicase (Mtr4) and a RNA poly(A) polymerase (Trf4 or Trf5). Together these proteins identify particular RNAs, and stimulate their degradation via the 3'-5' exonuclolytic activities of the RNA exosome, the primary RNA degrader in the nucleus.

Intricate to TRAMP function are the RNA binding proteins Air1 and Air2 (collectively referred to as Air proteins). The Air proteins are the identifiers of a myriad of RNA species that are targeted for decay, and they have also been shown mediate crucial protein-protein interaction within TRAMP [11-13]. In addition to their function within TRAMP, the Air proteins have also been implicated in regulating the activity of other proteins that are involved in nuclear mRNA transport [14]. In this role, the Air proteins may serve a function in RNA quality control by regulating the quality of mRNA transcripts that are exported to the cytoplasm. Although it is known that the Air proteins are involved in these various processes, details of how they interact with other proteins and RNA is lacking. The research presented in the last portion of this dissertation (Chapters 4-6) aims to further the understanding of Air protein functionality, with particular focus on the different protein-protein binding interactions that they employ for various functions.

The RNA Exosome is required for RNA degradation

In eukaryotes, a multi-protein complex called the RNA exosome serves the primary role of 3'-5' RNA degradation in the nucleus and cytoplasm. Essential functions performed by the exosome include 3'-5' end-processing of nuclear RNA precursors such as rRNAs, sn(o)RNAs, and tRNAs, turnover of both non-coding RNA in the nucleus and protein-coding mRNA in the cytoplasm, and RNA surveillance - where many types of aberrant non-functional RNAs are identified and degraded [8, 9, 15-21]. Crystal structures of the exosome from eukaryotes and archaea reveal a conserved structural architecture throughout species. The exosome is composed of a nine-subunit core (EXO-9). Six of these subunits are RNase PH domain containing proteins that form a hexameric barrel-like structure with a central channel. The other three subunits contain S1/KH RNA binding domains that assemble into a trimer that caps one side of the barrel-like hexamer, forming an entry pore to the central channel [22-24]. This EXO-9 core is conserved in eukaryotes and archaea. However, in contrast to the archaeal exosome, in eukaryotes the RNase PH domain proteins that comprise EXO-9 have apparently lost activity during evolution [23, 24], and nucleolytic activity is instead attained by association with additional proteins. In *S. cerevisieae*, the EXO-9 core associates with the endo- and 3'-5' exo-ribonuclease Rrp44 (Dis3) to form the catalytically active EXO-10 complex, which functions in both the nucleus and cytoplasm [25]. Structural analyses have shown that the Rrp44 nuclease associates near the bottom of the exosome core on the opposite side from the trimeric cap. In this position, unstructured ssRNAs can thread through the narrow entry pore (10-12 Å across) and into the central channel of the core to meet the exonucleolytic site of Rrp44 [25, 26]. In the nucleus, EXO-10 also associates

with the non-essential 3'-5'exoribonuclease Rrp6 to form the EXO-11 complex [27].

Rrp6 binds to the exosome trimeric cap and is required for particular degradation

functions, such as processing of 5.8S rRNA and distinct processing events of sn(o)RNAs

[28-30].

The RNA exosome requires cofactors

The functional exosome requires additional proteins, collectively called exosome

co-factors, to stimulate its activity and substrate specificity *in vivo* [31]. The requirement

of protein cofactors is not well understood, but likely involves specific recognition of

RNA substrates and providing an unstructured 3' end suitable for exonucleolytic

degradation. In *S. cerevisiae* the major exosome-activating cofactors include two related

DExH-box helicases Mtr4 and Ski2 and their associated proteins. Ski2 is found only in

the cytoplasm where it associates with two putative RNA binding proteins Ski3 and Ski8

to form the physiological multimeric SKI complex [23, 32]. In the cytoplasm the SKI

complex is required for mRNA surveillance and mRNA turnover by EXO-10. The

DExH-box helicase Mtr4 is strictly nuclear, and unlike Ski2, Mtr4 has some independent

functions such as the processing of 5.8S rRNA [27, 33-35]. Mtr4 is also found in a

physiological protein complex called TRAMP, which in addition to Mtr4 contains a RNA

binding protein (Air1 or Air2) and a RNA poly(A) polymerase (Trf4 or Trf5). In budding

yeast, TRAMP is essential and required for nuclear RNA surveillance and processing by

EXO-10 and EXO-11.

The TRAMP complex is required for nuclear RNA surveillance and processing

The **Tr**f4/5-**A**ir1/2-**M**tr4-**P**olyadenylation complex (TRAMP) of *S. cerevisiae* was discovered independently by three different groups in 2005 [9, 31, 36]. These initial studies indicated that TRAMP is composed of the DExH box RNA helicase Mtr4, a poly(A) polymerase Trf4, and one of two paralogous zinc knuckle proteins Air1 or Air2 (45% sequence identity). Later studies indicated that another TRAMP complex is formed by a Trf4 paralog Trf5 (48% sequence identity), which only associates with Air1. Thus, two types of TRAMP complexes are thought to exist. TRAMP4 contains Mtr4, Trf4, and either Air1 or Air2, and TRAMP5 contains Mtr4, Trf5, and only Air1. The duplicated Air and Trf genes likely arose during a whole genome duplication event in *S. cerevisiae* about 100 million years ago [37, 38], and therefore other eukaryotes have only one ortholog of each. In humans, the Air and Trf orthologs are named ZCCHC7 and PAPD5, respectfully. Furthermore, ZCCHC7 and PAPD5 have been shown to form a trimeric complex with human Mtr4 (hMtr4) [39], suggesting conservation of the TRAMP complex between fungi and animals.

TRAMP directly stimulates the nuclear exosome for processing and degradation of many types of RNA including tRNAs, sn(o)RNAs, rRNAs, pre-mRNAs, CUTs (cryptic unstable transcripts), and a number of intergenic transcripts [7-10]. In the current model of TRAMP function (Figure 3-1), Air1/2 identifies particular RNA substrates, Trf4/5 adds a short (~4-5 nt) poly(A) tail to the 3' end of the RNA, and Mtr4 unwinds any RNA secondary structure that may exist [9, 31, 36, 40, 41]. This unwinding and polyadenylation by TRAMP is presumed to provide an unstructured single stranded 3' end of the RNA, facilitating its degradation by the exosome [41].
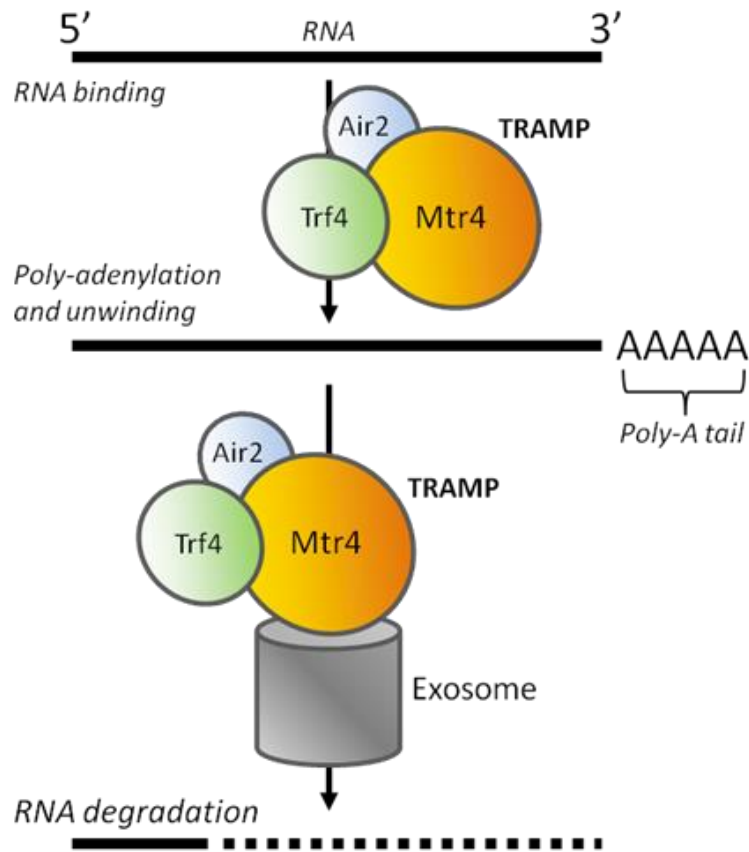
Figure 3-1. Tramp mediated RNA degradation by TRAMP4 (Trf4, Air2, and Mtr4). The TRAMP complex identifies RNAs, adds a short 3' poly(A) tail (4-5 nt), and stimulates their degradation by the nuclear exosome.

Interestingly, the function of polyadenylation by TRAMP is in contrast to the traditionally described function of polyadenylation by canonical poly(A) polymerases (PAPs). In eukaryotes, polyadenylation by PAP stabilizes pre-mRNAs, and requires recognition of specific sequence elements on the pre-mRNA transcript [42]. On the other hand, TRAMP polyadenylation stimulates RNA degradation, and TRAMP substrates have no identifiable similarities in sequence, structure, or function. [40, 41, 43]. After the initial discovery of TRAMP, a central question was how polyadenylation by PAP stabilizes RNA, while polyadenylation by TRAMP targets RNA for destruction? Recent

reports have revealed the likely answer to this question. The key difference governing the fate of polyadenylated RNA is likely attributed to a marked difference in the length of the 3' poly(A) tail. The oligo(A) tails appended by PAP are typically several dozen to hundreds of nucleotides long [44]. Recent data has indicated that the length of oligo(A) tails appended by TRAMP peaks at 4-5 nucleotides [8]. During pre-mRNA maturation, addition of at least 12 adenosines are required for binding of the canonical poly(A) binding protein Pab1, which serves in part to stabilize nuclear mRNA [44]. TRAMP oligo(A) tails are apparently too short for Pab1 binding, and as a result are left unprotected and subject to degradation by the exosome [8].

The non-canonical Poly(A) polymerases Trf4 and Trf5

Trf4 was originally identified in a synthetic lethal screen with DNA topoisomerase I (Top1) mutants, and hence was given the name **T**opoisomerase I **r**elated **f**unction **4** protein. Further genetic studies identified the Trf4 paralog Trf5 (48% sequence identity) [45]. Neither Trf4 nor Trf5 are individually essential. However, a Trf4Δ Trf5Δ double mutant is synthetically lethal [45, 46]. Homologues of Trf4 and Trf5 have been reported throughout eukaryotes including Cid14 in *Sz. Pombe*, and PAPD5 in humans [39, 47, 48].

Trf4 and Trf5 are members of the Cid1-family of non-canonical poly(A) polymerases. Members of the Cid1-family share common features with the canonical PAP, such as a homologous catalytic domain (NTP transferase domain) and central domain (PAP-associated domain) [47]. However, unlike the canonical PAP, most Cid1-family members lack a recognizable RNA recognition motif (RRM) used to identify and

bind RNA substrates. To compensate for lack of an RRM, many Cid1 family members are presumed to employ additional RNA-binding proteins to target their substrates for polyadenylation [47], such as Air1 and Air2 homologues in yeast and humans. Accordingly, the presence of either Air1 or Air2 is required for the polyadenylation activity of Trf4 *in vitro*, and Air1Δ Air2Δ cells exhibit loss of polyadenylation of non-coding RNAs *in vivo* [9, 31, 36].

The zinc knuckle proteins Air1 and Air2

In the TRAMP complex, Air1 and Air2 are involved in protein-protein and protein-RNA binding interactions. The various binding interactions of Air1 and Air2 and the implications of their binding interactions are the focus of this dissertation. The Air proteins were originally identified in a two-hybrid screen as proteins that could interact with the yeast arginine methyl transferase Hmt1, and were thus given the name **A**rginine methyltransferase **i**nteracting **r**ing finger proteins [14]. Air1 and Air2 are paralogs that each contain five adjacent CCHC-type (C-X_2-C-X_4-H-X_4-C) zinc knuckle motifs (also referred to as ring fingers) located between extended N- and C-terminal sequences (Figure 3-2). The Air proteins contain 45% sequence identity with the majority of identity conserved in the region comprising the zinc knuckles (68% sequence identity). Zinc knuckle motifs (ZnKs) are cysteine rich sequences that structurally fold into reverse turns to coordinate a zinc metal ion using three cysteine residues and one histidine. ZnKs were initially characterized in retroviral nucleocapsid proteins where they are used for binding ssRNA [49]. Many ZnK proteins have also been found throughout eukaryotes from yeast

to humans, although their functions in eukaryotic systems have not been well

characterized.
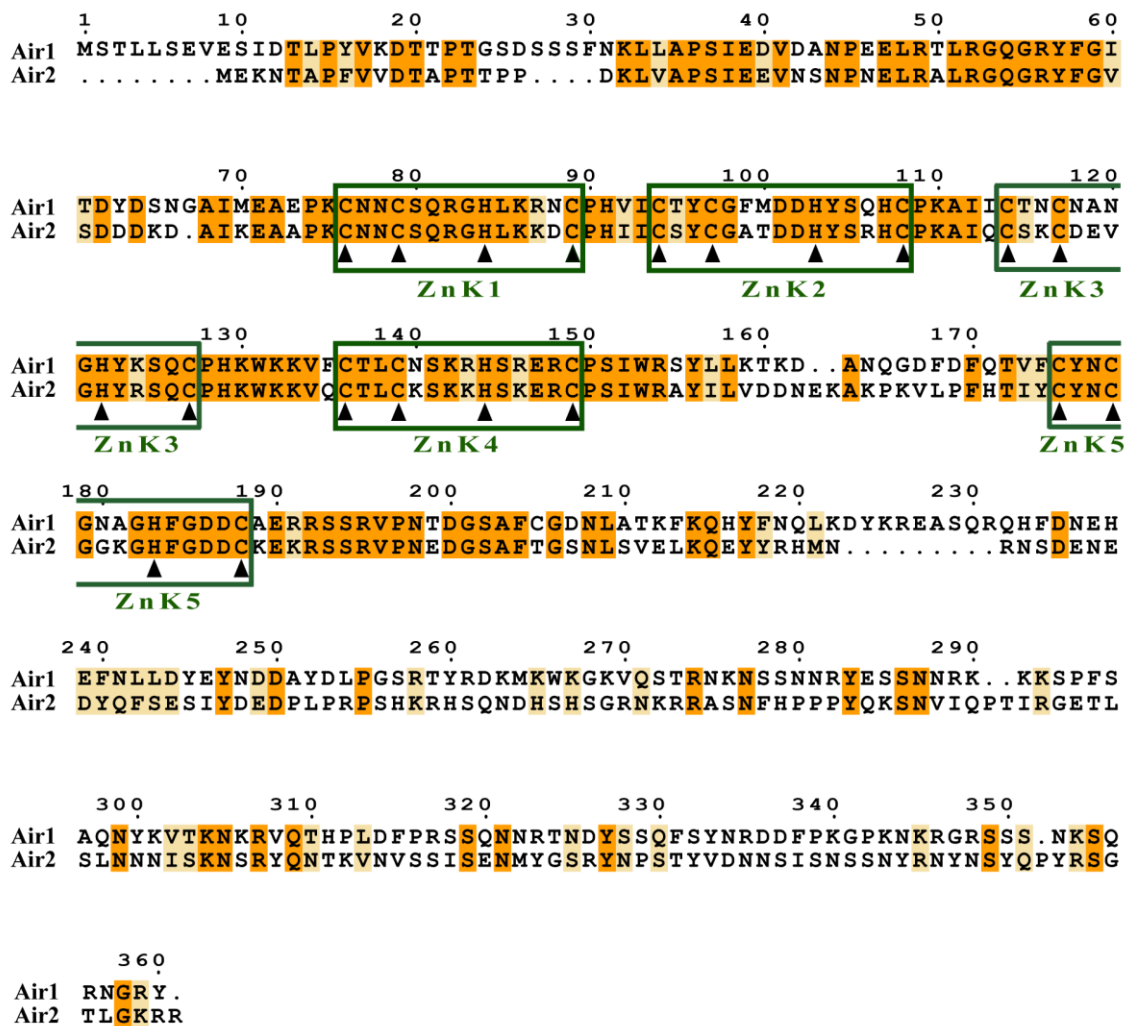
Figure 3-2. Air1 and Air2 sequence alignment. Identical sequences are colored orange. Semi-conserved sequences are colored light yellow. Zinc knuckles (ZnKs) are indicated by green boxes and ZnK residues are indicated by black triangles

The Air proteins provide RNA binding in TRAMP

Since the initial discovery of TRAMP, the zinc knuckles of Air1 and Air2 have been presumed to function as the RNA binding component of the complex. This idea was first made evident in mutagenesis experiments which showed that polyadenylation of an aberrant form of a structured RNA (tRNA[Ala]) by a Trf4-Air2 complex was significantly impaired upon deletion of the three most N-terminal ZnKs of Air2 [12]. This result correlated with previous experiments (described above) that indicated Air1 and Air2 are required for Trf4 polyadenylation of RNAs, including polyadenylation of the structured precursor-RNA hypomodified tRNA[iMet] [9, 13, 31, 36]. Notably, hypomodified tRNA[iMet] is the most well characterized RNA substrate of TRAMP. Recent reports have further clarified the involvement of Air2 ZnKs in binding aberrant forms of structured RNAs. In 2012, Holub *et al.* used NMR studies to reveal that binding of hypomodified tRNA[iMet] by Air2 involved the second, third, and fourth ZnKs [13]. In that same report, fluorescence anisotropy experiments showed that an Air2 truncation mutant containing all five ZnKs could bind directly to an unstructured 12 nucleotide poly(A) oligo with a K_d in the low micromolar range, suggesting that Air2 ZnKs also bind unstructured RNAs.

Air1 and Air2 provide substrate specificity in TRAMP

In *S. cerevisiae*, TRAMP4 and TRAMP5 have been shown to have different sub-nuclear localizations. For example, GFP-fusions of Trf4 and Air2 show that TRAMP4 can be found throughout the nucleus, whereas GFP-fusions of Trf5 and Air1 are primarily enriched within the nucleolus [11, 50, 51], which is the site of rRNA maturation. Likewise, Trf5 has been shown to preferentially stimulate degradation of abarrent rRNA

precursors [52]. This compartmentalization of TRAMP5 could be explained by distinct

substrate specificities between TRAMP complexes provided by either the Trf or Air

components. Because the Air proteins have been shown to be the likely source of

TRAMP RNA binding, they are also the likely source governing TRAMP RNA

specificity [11, 13]. Indeed, a report by Schmidt *et al*. identified that Air1 and Air2

govern TRAMP substrate specificity for vast number of different RNA types [53]. In that

study a combination of genetic and deep sequencing techniques revealed that both of the

Air proteins have overlapping function in directly stimulating the polyadenylation and

degradation of most types of identified TRAMP substrates, including many rRNAs,

sn(o)RNAs, CUTs, and mRNAs. In addition, the analysis also identified some differences

in specificity between Air1 and Air2. For example, Air2 is independently involved in

stimulating polyadenylation and degradation of several snoRNAs and spliceosomal

RNAs, as well as regulating levels of some mRNA transcripts which code for enzymes

involved in iron and carbon metabolism. Air1 was also shown to be independently

involved in regulating levels of certain mRNAs which code for proteins involved in copy

number control and regulation of the 2μ plasmid found in most yeast strains. Moreover,

the study by Schmidt *et al.* made clear that the Air proteins are key for providing

substrate specificity in TRAMP and the RNA exosome. However, the question of how

characteristic differences between Air1 and Air2 might control differences in specificity

between different TRAMP complexes is still unknown.

Air1 and Air2 mediate protein-protein interactions in TRAMP

A central question of TRAMP function has been how the three protein subunits associate with one another to form a functional complex. Initial studies of TRAMP indicated that Air2 and Trf4 form a tight complex *in vitro*. However, identification of the binding interactions between the two proteins remained elusive until five years later. In 2010, a crystal structure of a minimal Air2-Trf4 heterodimer revealed the binding interface between Air2 and Trf4 [12]. The crystal structure exhibits only the catalytic and central domains of Trf4 and the last two zinc knuckles of Air2 (Figure 3-3). In the model, the last zinc knuckle of Air2 (ZnK5) and sequences within the linker between ZnK4 and ZnK5 (L4) form direct interactions with the central domain of Trf4. Additional studies have also shown that recombinant Air2 with mutations in L4 or ZnK5 failed to co-purify and activate Trf4 [11, 13], indicating that the binding regions observed in the crystal structure represent the minimal binding interface between Air2 and Trf4. In addition, ZnK5 is the most conserved zinc knuckle amongst Air2 homologues, and the L4 binding residues constitute a conserved motif IWRxYxL, found in all Air2 homologues [11, 13]. The Air2 binding surface of Trf4 is also conserved in all Trf4 homologues. These observations suggest that the Air2-Trf4 binding interface observed in the crystal structure is phylogenetically conserved across species.

Identifying how Mtr4 associates with the Air2-Trf4 has proven to be more difficult, and therefore Mtr4 protein-protein interactions within TRAMP remain poorly characterized. Part of the complexity of identifying protein-binding interactions with Mtr4 is due to the instability of recombinant Trf and Air proteins that contain extended sequences beyond those observed in the Air2-Trf4 crystal structure. However, a recent

A

B
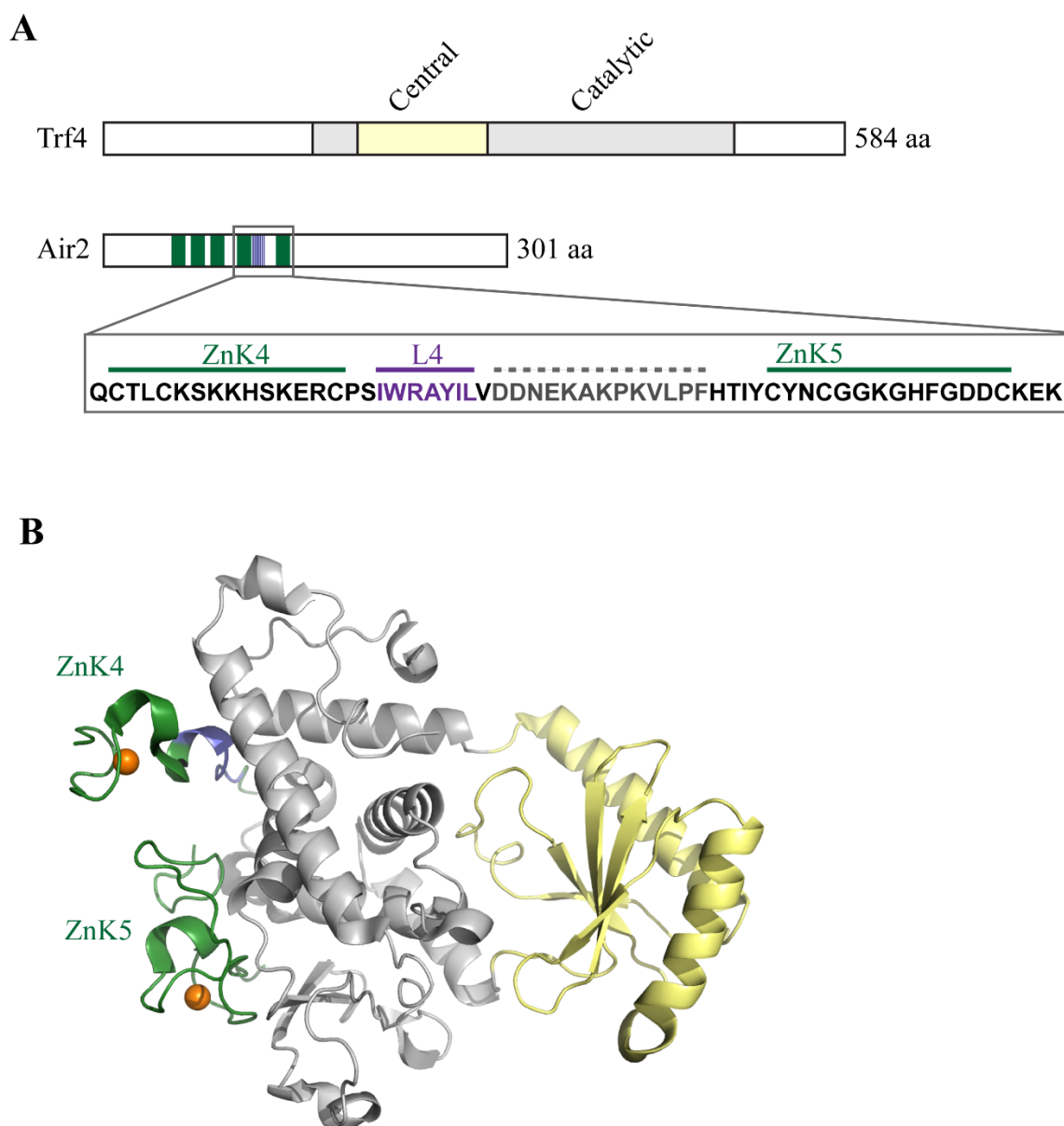
Figure 3-3. Crystal structure of a fragment of Air2 with the core domains of Trf4 (PDB id: 3NYB). **(A)** Domain organization of Trf4 and Air2. For Trf4 (1-584 aa), the catalytic domain (161-189, 316-481 aa) is colored grey. The central domain is colored light yellow. For Air2 (1-301 aa), Zinc knuckles are indicated by green boxes, Trf4 binding motif IWRxYxL in linker 4 (L4) is indicated by purple lines. Insert box shows sequences of Air2 ZnK4-ZnK5 with a dashed line indicating residues not observed in the crystal structure. **(B)** Cartoon ribbon diagram of the minimum Air2-Trf4 heterodimer. Protein domains are colored as in panel A. In Air2, zinc ions are shown as orange spheres and Trf4 binding residues in L4 are colored purple.

publication indicated that the N-terminus of Air2 is required to co-immunoprecipitate Mtr4 with an Air2-Trf4 heterodimer, suggesting that Air2 may also facilitate binding to Mtr4. In Chapter 5 of this dissertation I identify and characterize binding between a small peptide fragment of Air2 and Mtr4, revealing a binding interaction used by Air2 to bridge the three protein subunits of TRAMP.

Air proteins are regulators of the methyltransfease Hmt1 and mRNA export.

As mentioned previously, the Air proteins were initially identified as proteins that interact with the yeast arginine methyltarnsferase Hmt1 [14]. In that study, Air1 and Air2 were shown to directly bind Hmt1 in yeast two- hybrid assays in which the RGG-box domain of the mRNA transport protein Npl3 was used as bait. The RGG-box domain of Npl3 is both a site used for binding mRNAs that are shuttled to the cytoplasm, and a site of methylation by Hmt1. In budding yeast, nuclear export of Npl3, and thus export of its mRNA cargo, requires methylation by Hmt1. Furthermore, Air1 was shown to inhibit Hmt1 methylation of Npl3 *in vitro* and *in vivo*, suggesting that binding of Air1 to Hmt1 inhibits methylation, and thus inhibits mRNA export. To verify this model *in vivo*, Npl3 and mRNA localization was analyzed in *S. cerevisiae* strains containing single gene deletions of Air1 and Air2, and Air1Δ Air2Δ double mutant strains. The single gene deletions had no observable effect on cell viability or Npl3/RNA transport. However, double mutant strains had a severe growth defect, suggesting some functional overlap of Air1 and Air2. The double mutant strain also exhibited nuclear retention of Npl3 and accumulation of Poly(A) mRNAs in the nucleus. Collectively these experiments

indicated that indeed the Air proteins can regulate methylation activity of Hmt1 and influence nuclear mRNA transport.

Modification of proteins by arginine methylation serves many critical functions. For example, in eukaryotes arginine methylation is involved in RNA processing, transcriptional regulation, signal transduction, and DNA repair [54]. Currently, Air1 is the only known protein that has been demonstrated to regulate the activity of arginine methylation in *S. cerevisiae*. Understanding the mechanistic details of Hmt1 regulation would provide valuable insight to how arginine methylation is regulated endogenously. In Chapter 6 of this dissertation I describe my own research efforts aimed at understanding the interaction between the Air proteins and Hmt1, and also outline future directions for that research.

The DExH-box helicase Mtr4

The RNA helicase Mtr4 was first identified in a yeast genetic screen for identifying mutants that accumulate Poly(A) RNA in the nucleus. At that time, it was presumed that the accumulated poly(A) RNAs were mRNAs, and therefore the name **m**RNA **tr**ansport **4** was given [55]. Mtr4 was the first characterized exosome co-factor, and current knowledge indicates that Mtr4 and its homologues (including other Ski2-like helicases) are required for all known exosome functions *in vivo*. Mtr4 homologues and associated protein complexes have been identified across species from fungi to humans. In *S. cerevisiae* Mtr4 can function independent of other proteins to stimulate the exosome, such as 5.8s rRNA processing [27, 33-35]. However, the majority of characterized functions of Mtr4 are in the context of the TRAMP complex. Like other

Ski2-like helicases, Mtr4 has 3'-5' unwinding activity and RNA dependent ATPase

activity. It is generally presumed that the helicase function of Mtr4 is used to prepare

RNA substrates for exosome degradation by unwinding RNA secondary structure and

displacing bound proteins from ribonucleoprotein complexes.

Crystal structures of Mtr4, including apo and RNA-bound forms have provided

insight into Mtr4s domain architecture and RNA binding path (Figure 3-4) [34, 56].

Similar to other Ski2-like helicases, Mtr4 has a four-domain helicase core that is

composed of two canonical RecA-like folds (domains 1 and 2), a winged helix domain

(domain 3), and a helical bundle domain, also referred to as the ratchet helix domain

(domain 4). A large 256 amino acid insertion (arch domain) is also observed to be

inserted into the winged helix domain. The arch domain is a prominent feature that rises

above the helicase core and terminates in a globular structure called the fist (also called

the KOW domain) [34, 56]. The arch domain is unique to Mtr4 and other Ski2-like RNA

helicases. Notably, in Mtr4 the arch domain is not required for helicase activity or

binding to Air2-Trf4, but is required for 5.8S rRNA processing *in vivo* and binding

tRNA$_i^{Met}$ *in vitro* [34, 56]. In all DExH-box helicases, the RecA1 and RecA2 domains

contain conserved sequence motifs for binding and hydrolysis of ATP, a requirement for

helicase function [57]. In the structures of Mtr4, the RecA domains pack against the

winged helix and ratchet helix domains, forming a central channel for ssRNA. [34, 56].

To visualize RNA duplex unwinding in Mtr4, the apo-Mtr4 model was used to

construct a model of nucleic acid duplex unwinding according to the DNA bound

structure of the Ski2-like DNA helicase Hel308 (Figure 3-4-B) [34]. The duplex

unwinding model shows that a conserved β-hairpin loop in the second RecA domain
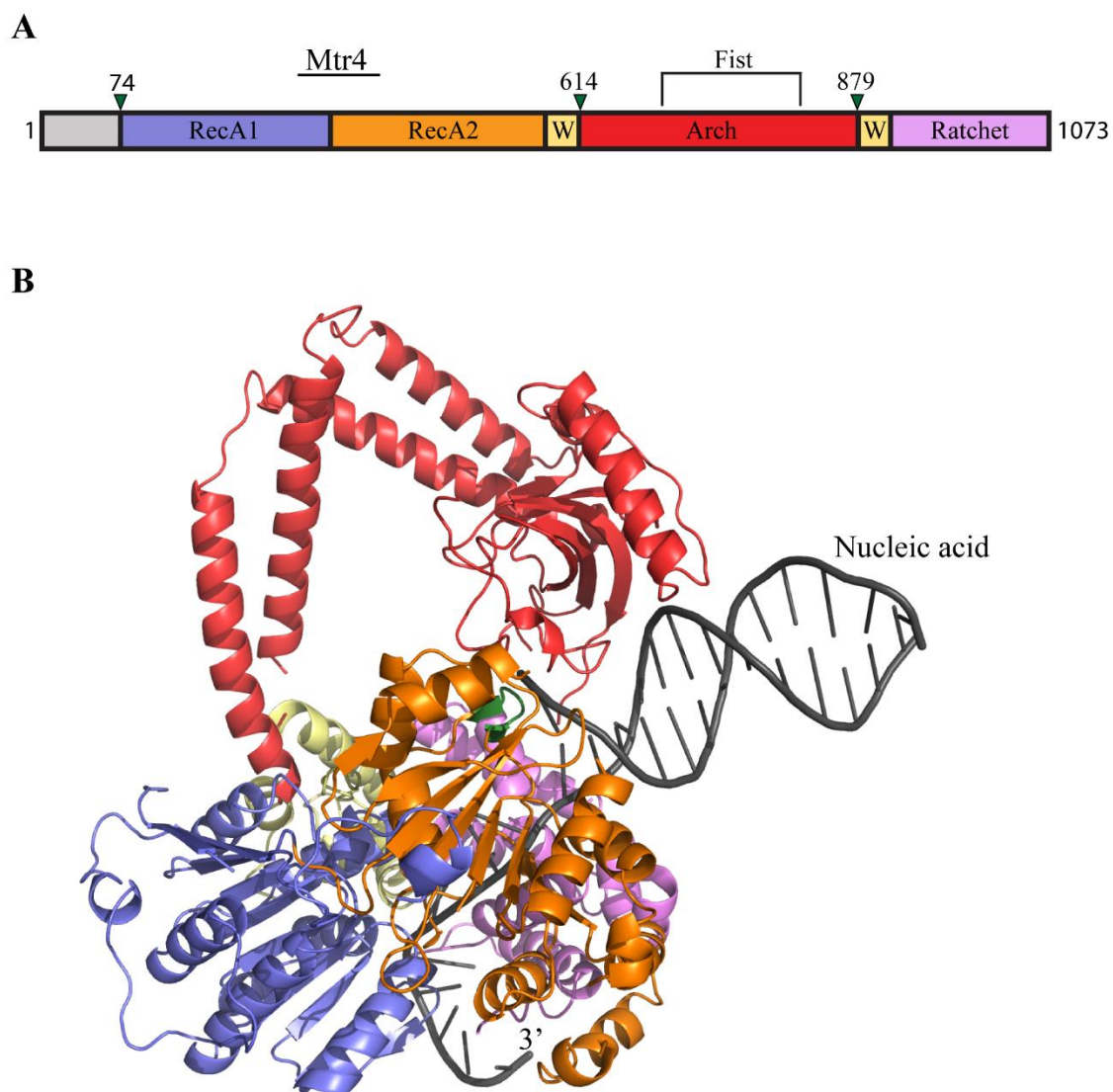
Figure 3-4. Crystal structure of Mtr4 (PDB id: 3L9O) **(A)** Domain organization of Mtr4. Domain names are shown. The winged helix domain is abbreviated as W. **(B)** Cartoon ribbon diagram of Mtr4 (residues 74-1073) from *S. cerevisiae*. Domains are colored as in panel A. To visualize binding of double-stranded nucleic acid, the apo-form structure was modeled with partially unwound nucleic acid from the Hel308 structure (Jackson *et al*., 2010). The β-hairpin loop (colored green) extends from RecA2 and facilitates strand separation.

facilitates strand separation as the RNA enters into the central channel. As the nucleic

acid passes through the channel it contacts each of the RecA domains and the ratchet

helix domains before it exits the helicase at the base. Furthermore, the surface of the

RNA exit site appears to be structured to accommodate the dimensions of the EXO9 core,

and thus it has been proposed that Mtr4 feeds RNA directly into the exosome [34].

In TRAMP, Mtr4 modulates Poly(A) tail length

As previously mentioned, TRAMP primarily adds only five adenosines to the

3'end of its RNA targets *in vivo* [8]. Several studies have identified that this poly(A) tail

length restriction is due to a secondary functional activity of Mtr4. Mtr4 was first

implicated in regulating Poly(A) tail length in studies by the Tollervey lab, which showed

Mtr4 depletion *in vivo* results in hyperadenylation of TRAMP substrates [52]. More

recently, *in vitro* studies by the Jankowsky lab revealed that in the context of TRAMP,

Mtr4 stimulates polyadenylation by Trf4 until approximately four adenosines are added

to the RNA [58]. After the critical number of four adenosines is added, polyadenylation is

markedly suppressed, and when Mtr4 is removed, poly(A) restriction is abolished,

suggesting that Mtr4 modulates poly(A) tail length. Mutations that prevent helicase

unwinding activity further indicated that modulation of polymerase activity was not due

to the unwinding activity of Mtr4. Other studies by the Jankowsky lab have showed that

isolated Mtr4 and TRAMP both have a distinct preference for unwinding RNA substrates

with a 4-6 nt poly(A) tail [59]. Together, these results indicate that Mtr4 possesses an

inherent poly(A) sensing mechanism independent of helicase function, which allows the

helicase to identify the number of 3'adenosines and stimulate Poly(A) addition until the preferred number of terminal adenosines is acquired.

Modulating the activity of another protein by identifying particular nucleic acid features is unprecedented amongst RNA helicases. Central to this unique function is Mtr4s ability to identify and discriminate between substrates that have a poly(A) tail and those that do not. A recent publication by our group [60] revealed that Mtr4 uses specific sequences located on the ratchet helix of domain 4 to identify RNA poly(A) tails. Specifically, R1030 (conserved in Mtr4 homologues) is used to detect 3' poly(A) sequences and promote unwinding of polyadenylated RNA substrates *in vitro*. Mutation of R1030 resulted in decreased unwinding rates of polyadenylated RNA to levels comparable to substrates without a poly(A) tail. These findings correlate with the RNA-bound crystal structure which showed direct interactions between R1030 and a RNA adenine base.

Mtr4 helicase activity is stimulated by TRAMP

Another interesting feature of Mtr4 function is that its RNA unwinding rate is enhanced by nine-fold in the context of TRAMP [59]. Because Mtr4 has some functions independent of TRAMP, this regulation of Mtr4 activity could be used to help avoid the energy cost of unwinding substrates that have not first been identified by TRAMP. Intriguingly, the increase in unwinding activity upon TRAMP formation is observed regardless of poly(A) tail length, suggesting that formation of TRAMP independently facilities modulation of both Trf4 polyadenylation and Mtr4 helicase activity. However, it is not clear how Mtr4s interaction with Air2-Trf4 stimulates helicase unwinding rates.

One possibility is that binding of Air2-Trf4 causes a structural rearrangement of Mtr4 that improves RNA substrate binding, ATP binding, or ATPase activity. Such protein-induced conformational changes have been demonstrated to enhance helicase activity for a variety of DExH-box helicases [61-64]. Comparatively, *in vitro* studies show that TRAMP has an elevated ATP binding affinity compared to Mtr4 alone, and importantly, this increased affinity for ATP is not due to the adenylation activity of Trf4 (which also requires ATP). This observation raises the possibility that binding of Mtr4 to Air2-Trf4 increases Mtr4s ATP binding affinity, which could facilitate the enhanced helicase activity observed in TRAMP. A better understanding of helicase activity modulation in TRAMP requires additional knowledge of how the Air2-Trf4 heterodimer interacts with Mtr4. Accordingly, in Chapter 5 of this dissertation, I a direct binding interaction is identified between Air2 and the ATP binding domains of Mtr4 (RecA domains).

Concluding remarks

The central goal of research presented in the following chapters of this dissertation is to extend the functional and mechanistic characterization of Air1 and Air2 in the context of their various binding interactions and physiological implications therein. Air1 and Air2 are functionally multi-faceted proteins capable of associating with various protein-partners and RNA species to regulate the cellular levels many RNA types. In Chapter 5, a previously unknown binding interface is identified between Air2 and the TRAMP complex component Mtr4. That work successfully narrows down an Air2-Mtr4 binding interface that includes the first 29 amino acids of Air2 and the RecA domains of Mtr4. This interaction likely explains a key functionality of helicase regulation in

TRAMP. Chapter 6 describes preliminary studies and directions for future efforts aimed at characterizing the binding interactions of Air proteins to Mtr4, and the methyltransferase protein Hmt1. Collectively, these efforts provide important characterization of Air1 and Air2 protein binding interactions with different protein binding-partners, and provide a foundation for future research aimed at understanding Air protein function.

REFERENCES

[1]   P.P. Amaral, M.E. Dinger, T.R. Mercer, J.S. Mattick, Science 319 (2008) 1787-1789.

[2]   C.A. Brosnan, O. Voinnet, Current Opinion in Cell Biology 21 (2009) 416-425.

[3]   J.E. Wilusz, H. Sunwoo, D.L. Spector, Genes & Development 23 (2009) 1494-1504.

[4]   D. Astuti, M.R. Morris, W.N. Cooper, R.H. Staals, N.C. Wake, G.A. Fews, H. Gill, D. Gentle, S. Shuib, C.J. Ricketts, T. Cole, A.J. van Essen, R.A. van Lingen, G. Neri, J.M. Opitz, P. Rump, I. Stolte-Dijkstra, F. Muller, G.J. Pruijn, F. Latif, E.R. Maher, Nature Genetics 44 (2012) 277-284.

[5]   A.G. Bassuk, Y.Z. Chen, S.D. Batish, N. Nagan, P. Opal, P.F. Chance, C.L. Bennett, Neurogenetics 8 (2007) 45-49.

[6]   A.V. Philips, T.A. Cooper, Cellular and Molecular Life Sciences 57 (2000) 235-249.

[7]   S. San Paolo, S. Vanacova, L. Schenk, T. Scherrer, D. Blank, W. Keller, A.P. Gerber, PLoS Genetics 5 (2009) e1000555.

[8]   W. Wlotzka, G. Kudla, S. Granneman, D. Tollervey, EMBO J. 30 (2011) 1790-1803.

[9]   F. Wyers, M. Rougemaille, G. Badis, J.C. Rousselle, M.E. Dufour, J. Boulay, B. Regnault, F. Devaux, A. Namane, B. Seraphin, D. Libri, A. Jacquier, Cell 121 (2005) 725-737.

[10]   K.Y. Kong, H.M. Tang, K. Pan, Z. Huang, T.H. Lee, A.G. Hinnebusch, D.Y. Jin, C.M. Wong, Nucl. Acids Res. 42 (2014) 643-660.

[11]   M.B. Fasken, S.W. Leung, A. Banerjee, M.O. Kodani, R. Chavez, E.A. Bowman, M.K. Purohit, M.E. Rubinson, E.H. Rubinson, A.H. Corbett, J. Biol. Chem. 286 (2011) 37429-37445.

[12]   S. Hamill, S.L. Wolin, K.M. Reinisch, J. Biol. Chem. 107 (2010) 15045-15050.

[13]   P. Holub, J. Lalakova, H. Cerna, J. Pasulka, M. Sarazova, K. Hrazdilova, M.S. Arce, F. Hobor, R. Stefl, S. Vanacova, Nucl. Acids Res 40 (2012) 5679-5693.

[14]   K. Inoue, T. Mizuno, K. Wada, M. Hagiwara, J. Biol. Chem. 275 (2000) 32793-32799.

[15]   T. Carneiro, C. Carvalho, J. Braga, J. Rino, L. Milligan, D. Tollervey, M. Carmo-Fonseca, Molecular and Cellular Biology 27 (2007) 4157-4165.

[16]   B. Das, S. Das, F. Sherman, Proc. Natl Acad Sci USA 103 (2006) 10871-10876.

[17]   C.A. Davis, M. Ares, Jr., Proc. Natl Acad Sci USA 103 (2006) 3262-3267.

[18]   C. Dez, J. Houseley, D. Tollervey, EMBO J. 25 (2006) 1534-1546.

[19]   P. Grzechnik, J. Kufel, Molecular Cell 32 (2008) 247-258.

[20]   P. Hilleren, T. McCarthy, M. Rosbash, R. Parker, T.H. Jensen, Nature 413 (2001) 538-542.

[21]   S. Kadaba, A. Krueger, T. Trice, A.M. Krecic, A.G. Hinnebusch, J. Anderson, Genes & Development 18 (2004) 1227-1240.

[22]   K. Buttner, K. Wenig, K.P. Hopfner, Molecular Cell 20 (2005) 461-471.

[23]   J. Houseley, J. LaCava, D. Tollervey, Nature Reviews 7 (2006) 529-539.

[24]   Q. Liu, J.C. Greimann, C.D. Lima, Cell 127 (2006) 1223-1237.

[25]   F. Bonneau, J. Basquin, J. Ebert, E. Lorentzen, E. Conti, Cell 139 (2009) 547-559.

[26]   D.L. Makino, M. Baumgartner, E. Conti, Nature 495 70-75.

[27]   C. Allmang, E. Petfalski, A. Podtelejnikov, M. Mann, D. Tollervey, P. Mitchell, Genes & Development 13 (1999) 2148-2158.

[28]   R.K. Gudipati, Z. Xu, A. Lebreton, B. Seraphin, L.M. Steinmetz, A. Jacquier, D. Libri, Molecular Cell 48 (2012) 409-421.

[29]   A. Lardenois, Y. Liu, T. Walther, F. Chalmel, B. Evrard, M. Granovskaia, A. Chu, R.W. Davis, L.M. Steinmetz, M. Primig, Proc. Natl Acad Sci USA 108 1058-1063.

[30]   E. Thomson, D. Tollervey, Molecular and Cellular Biology 30 976-984.

[31]   J. LaCava, J. Houseley, C. Saveanu, E. Petfalski, E. Thompson, A. Jacquier, D. Tollervey, Cell 121 (2005) 713-724.

[32]   J.T. Brown, X. Bai, A.W. Johnson, RNA 6 (2000) 449-457.

[33]   J. de la Cruz, D. Kressler, D. Tollervey, P. Linder, EMBO J. 17 (1998) 1128-1140.

[34]   R.N. Jackson, A.A. Klauer, B.J. Hintze, H. Robinson, A. van Hoof, S.J. Johnson, EMBO J. 29 (2010) 2205-2216.

[35]   S. Kadaba, X. Wang, J.T. Anderson, RNA 12 (2006) 508-521.

[36] S. Vanacova, J. Wolf, G. Martin, D. Blank, S. Dettwiler, A. Friedlein, H. Langen, G. Keith, W. Keller, PLoS Biology 3 (2005) e189.

[37] M. Kellis, B.W. Birren, E.S. Lander, Nature 428 (2004) 617-624.

[38] C. Seoighe, K.H. Wolfe, Proc. Natl Acad Sci USA 95 (1998) 4447-4452.

[39] M. Lubas, M.S. Christensen, M.S. Kristiansen, M. Domanski, L.G. Falkenby, S. Lykke-Andersen, J.S. Andersen, A. Dziembowski, T.H. Jensen, Molecular Cell 43 624-637.

[40] J.T. Anderson, X. Wang, Critical Reviews in Biochemistry and Molecular Biology 44 (2009) 16-24.

[41] J. Houseley, D. Tollervey, Biochimica et Biophysica Acta 1779 (2008) 239-246.

[42] W. Keller, Cell 81 (1995) 829-832.

[43] X. Wang, H. Jia, E. Jankowsky, J.T. Anderson, RNA 14 (2008) 107-116.

[44] A.B. Sachs, R.W. Davis, R.D. Kornberg, Molecular and Cellular Biology 7 (1987) 3268-3276.

[45] B.U. Sadoff, S. Heath-Pagliuso, I.B. Castano, Y. Zhu, F.S. Kieff, M.F. Christman, Genetics 141 (1995) 465-479.

[46] Z. Wang, I.B. Castano, A. De Las Penas, C. Adams, M.F. Christman, Science 289 (2000) 774-779.

[47] A.L. Stevenson, C.J. Norbury, Yeast 23 (2006) 991-1000.

[48] T.Z. Win, S. Draper, R.L. Read, J. Pearce, C.J. Norbury, S.W. Wang, Molecular and Cellular Biology 26 (2006) 1710-1721.

[49] V. D'Souza, M.F. Summers, Nature Reviews Microbiology 3 (2005) 643-655.

[50]   W.K. Huh, J.V. Falvo, L.C. Gerke, A.S. Carroll, R.W. Howson, J.S. Weissman, E.K. O'Shea, Nature 425 (2003) 686-691.

[51]   C. Walowsky, D.J. Fitzhugh, I.B. Castano, J.Y. Ju, N.A. Levin, M.F. Christman, J. Biol. Chem. 274 (1999) 7302-7308.

[52]   J. Houseley, D. Tollervey, EMBO Reports 7 (2006) 205-211.

[53]   K. Schmidt, Z. Xu, D.H. Mathews, J.S. Butler, RNA 18 1934-1945.

[54]   M.T. Bedford, S. Richard, Molecular Cell 18 (2005) 263-272.

[55]   T. Kadowaki, S. Chen, M. Hitomi, E. Jacobs, C. Kumagai, S. Liang, R. Schneiter, D. Singleton, J. Wisniewska, A.M. Tartakoff, J. of Cell Biology 126 (1994) 649-659.

[56]   J.R. Weir, F. Bonneau, J. Hentschel, E. Conti, Proc. Natl Acad Sci USA 107 (2010) 12139-12144.

[57]   M.R. Singleton, M.S. Dillingham, D.B. Wigley, Annual Review of Biochemistry 76 (2007) 23-50.

[58]   H. Jia, X. Wang, F. Liu, U.P. Guenther, S. Srinivasan, J.T. Anderson, E. Jankowsky, Cell 145 (2012) 890-901.

[59]   H. Jia, X. Wang, J.T. Anderson, E. Jankowsky, Proc. Natl Acad Sci USA 109 (2012) 7292-7297.

[60]   L.L. Taylor, R.N. Jackson, M. Rexhepaj, A.K. King, L.K. Lott, A. van Hoof, S.J. Johnson, Nucl. Acids Res 42 (2014) 13861-13872.

[61]   N.L. Korneeva, E.A. First, C.A. Benoit, R.E. Rhoads, J. Biol. Chem. 280 (2005) 1872-1881.

[62]   J.P. Staley, J.L. Woolford, Jr., Current Opinion in Cell Biology 21 (2009) 109-118.

[63]   N. Tanaka, A. Aronova, B. Schwer, Genes & Dev. 21 (2007) 2312-2325.

[64]   T. Tanaka, T. Mizukoshi, K. Sasaki, D. Kohda, H. Masai, J. of Biolog. Chem. 282 (2007) 19917-19927.

CHAPTER 4

A COMPREHENSIVE DESCRIPTION OF METHODS USED TO CHARACTERIZE

AIR PROTEINS

ABSTRACT

Many methodologies have been developed in order to conduct structural and biochemical studies of Air-proteins and other TRAMP components. These methods include construction of expression constructs, purification strategies, protein binding studies, and crystallization. Methods were also developed for conducting binding studies and initiating crystallization trials. This chapter describes detailed methods used to characterize TRAMP proteins, with a particular emphasis on Air1 and Air2. The expression constructs and purification strategies mentioned are those that have produced the greatest amount of protein expression and purification yields.

INTRODUCTION

All three protein subunits of TRAMP including both of the Air proteins (Air1 and Air2), Trf4, and Mtr4 has been studied as part of this dissertation work. Many obstacles were overcome to develop methods for protein expression and purification of isolated proteins and protein co-complexes. Assays for characterizing protein binding interactions were also developed, and several crystallization studies for TRAMP proteins have been initiated.

One of the major obstacles encountered during this research was the difficulty of expression and purification of recombinant TRAMP proteins. At the beginning of this

dissertation work the only protein subunit of TRAMP that had been successfully expressed in *E.coli* was Mtr4. My efforts started with developing a protocol to express and purify isolated Trf4. Initial protein expression trials were conducted using various expression and growth parameters in both *E.coli* and yeast. To overcome difficulties in expression and solubility, a protein refolding protocol was developed that could successfully isolate a soluble form of truncated Trf4.

A soluble full-length form of Trf4 was later obtained by co-expressing Trf4 with another TRAMP component, the zinc knuckle protein Air2. Both proteins were co-expressed on a single plasmid (pET-Duet1) which resulted in soluble Trf4-Air2 complexes that could be purified. In addition to full length proteins, different truncation mutants of Trf4 and Air2 were also cloned into the same co-expression plasmid. These constructs also produced soluble Trf4-Air2 complexes that were purified and used for many crystallization screening trials. Unfortunately, none of those trials have produced protein crystals. However, those expression constructs and purifications methods will likely be of value for future functional and structural studies.

The greatest amount of success in expression and purification of TRAMP complex proteins has been with the zinc knuckle proteins Air1 and Air2. In that work many different truncation mutants and protocols were developed to successfully purify both of the Air-proteins. Purified Air1 and Air2 have been used to characterize important binding interactions between Air-proteins and the helicase Mtr4 (Chapter 5), as well as binding interactions between Air-proteins and proteins that are not part of the TRAMP complex, such as the methyltransferase Hmt1 (Chapter 6). In addition, many crystallization trials have been initiated using Air protein truncation mutants. These

crystallization trials include co-crystallization of Air-proteins with Mtr4 and with Hmt1. In both of those cases, crystals have been obtained and are awaiting x-ray diffraction analysis.

This chapter describes in detail various methods that have been developed and used during my dissertation research on proteins from the TRAMP complex. It is anticipated that this chapter will serve as a useful guide for future researchers studying proteins within the TRAMP complex. Particular attention is paid to methods used for expression and purification of the Air proteins.

EXPRESSION AND PURIFICATION OF RECOMBINANT PROTEINS

Expression and purification of Trf4

Expression analysis of isolated Trf4 was initiated by inserting the full-length Trf4 gene sequence from *S.cerevisiae* into a pET 151-D-topo *E. coli* expression vector (Invitrogen). Primer design and insertion of Trf4 was performed following the pET 151-D-topo manual. All expression tests using this full length Trf4 construct resulted in no detectable expression. To improve expression of Trf4, PCR primers were designed to amplify a truncation mutant of Trf4 encoding residues 113-516 (Trf4$^{113-516}$). These residues encompass the region of Trf4 exhibited in the Trf4-Air2 crystal structure (Trf4 residues 159-481) which includes the Trf4 catalytic and central domains [1]. Trf4$^{113-516}$ also contains residues specified as a minimal interaction region with Mtr4 by yeast two hybrid analysis [2, 3]. The PCR primers used for amplification were ordered from a commercial vendor (Integrated DNA Technologies), and the full-length Trf4 construct was used as a template for the PCR reaction. The purified PCR product was then inserted

into an empty pET 151-D-topo vector using restriction cloning. The resulting construct

was expressed in BL21(DE3) codon+RIL *E.coli* cells (Agilent Technologies) using

IPTG induction and auto-induction protocols [4, 5]. Cell lysis was performed by re-

suspending cells in lysis buffer (10% glycerol, 500mM NaCl, 2mM β-ME, 10mM

imidazole) followed by sonication, and centrifugation. SDS-PAGE analysis indicated that

the truncated form of Trf4 was highly expressed, but the protein was only observed in the

insoluble fraction.

Denaturation and refolding of Trf4$^{113-516}$

A protein denaturing-refolding protocol has been used to acquire soluble Trf4$^{113-516}$. Approximately 20g of cells were re-suspended in lysis buffer, lysed by sonication,

and centrifuged. The insoluble fraction (cell-pellet from centrifugation) was then re-

suspended in lysis buffer containing 6M urea. The 6M Urea solution was centrifuged at

20,000 rpm and the soluble fraction was dialyzed step-wise to 250mM urea. The

following day, protein was refolded on a Ni-affinity column by adding the 250mM urea

solution to nickel resin followed by washing and eluting with solutions containing no

urea. Refolded Trf4$^{113-516}$ was then further purified using heparin-affinity

chromatography. SDS-PAGE analysis revealed that this purification strategy yielded

recombinant Trf4$^{113-516}$ protein that was greater than 80% pure. This refolding procedure

yielded approximately 3mg of Trf4$^{113-516}$ from a 2 L growth. To date, this is the only

known protocol for purification of isolated Trf4.

Co-expression and co-purification of Trf4-Air2

Expression construct design

For expression and purification of a Trf4-Air2 heterodimer, a pET-Duet1 expression plasmid containing the full length genes of Trf4 and Air2 was obtained from a collaborator (Dr. Echard Jankowsky at Case Western Reserve University). This construct contains a full-length Air2 gene with an N-terminal 6x-histidine tag in cloning site 1, and the full-length Trf4 gene in cloning site 2. Using that construct as a template, a second pETduet1 Trf4-Air2 expression construct was made. This construct is analogous to the former except the Air2 gene has been truncated to express only amino acids 1-223. The truncated Air2 gene was created by site directed mutagenesis following a modified version of the QuickChange[TM] Site-Directed Mutagenesis protocol (Agilent Technologies). The only modification to the QuickChange[TM] protocol was that the recommended polymerase and buffer was substituted with PfuUltra[TM] polymerase and PfuUltra Buffer (Stratagene). The Air2 mutagenesis replaced the codon encoding Air2 Ser224 with STOP224. The mutagenesis site was identified using sequence analysis and secondary structural prediction to identify a suitable site approximately downstream of the Air2 zinc knuckle 5 motif. The Air2 protein encoded by this construct contains the N-terminus and all five Zinc knuckles.

Expression and purification

Co-expression of Trf4 with Air2 alleviates some of the intractable problems (such as no expression, low solubility, and protein aggregation) that I have encountered upon expression and purification of isolated proteins. For example, no expression is detectable

in other constructs that contain isolated full-length versions of Air1, Air2, or Trf4. In addition, all other constructs that contain isolated truncation mutants express proteins that generally have low solubility and aggregate during gel filtration. In contrast, each of the Trf4-Air2 pETduet1 constructs mentioned above express Trf4-Air2 heterodimers that are nearly completely soluble and do not form aggregates during gel filtration.

An identical procedure has been used to express and purify each of the Trf4-Air2 heterodimers. Protein expression was induced in BL21(DE3) codon+RIL *E.coli* cells (Agilent Technologies) using an auto induction protocol [5]. Cells were grown in 2.5 L baffled Erlenmeyer flasks with a culture volume of 500 mL autoinduction media. Appropriate antibiotics were added to the growth media for selectable resistance (ampicillin and chloramphenicol). Because the zinc knuckles motifs of Air2 require zinc ions to fold properly, $ZnSO_4$ (250uM) was also added to growth media prior to inoculation with *E.coli* cells. Cell cultures were incubated at 37° C with shaking at 300 rpm for five hours. After the initial incubation, cultures were moved to a room temperature shaker (220 rpm) for an additional 24 hours. Cells were harvested by centrifugation and stored at -80° C.

Table 4-1. Buffer solutions for Air protein preparations

Buffer component	Lysis Buffer	Buffer A	Buffer B	Sizing Buffer
Glycerol	10.00%	10.00%	10.00%	10.00%
HEPES pH 7.5	50 mM	50 mM	50 mM	50 mM
β-ME	2 mM	2 mM	2 mM	2 mM
NaCl	500 mM	150 mM	1 M	100 mM
Volume used	1 L mL	1 L	500 mL	1L

Cell lysis was performed by first re-suspending cell pellets in Lysis Buffer (Table 4-1) containing 10mM imidazole and protease inhibitors (1 µg/mL aprotinin, 1.4 µg/mL pepstatin, and 1 µg/mL leupeptin) and lysozyme (0.2mg/mL). Approximately 3 mL Lysis Buffer was added for every 1 mg of cells. Cells were lysed by sonication followed by centrifugation at 20,000 rpm for 30 minutes at 4° C. The clarified lysate was then added to 5 mL of Qiagen or GE brand nickel-affinity resin for 1 hour at 4° C. After incubation the Ni resin was washed with 500 mL lysis buffer, followed by 200 mL Buffer B, and then 100 mL Buffer A (Table 4-1) protein was eluted with 50 mL Buffer A containing 500 mM imidazole. Ni-column elutions were loaded directly onto a 5 mL heparin column pre-equilibrated with Buffer A. The majority of contaminating proteins remaining after the Ni-column step flowed through the heparin column. A 100 mL gradient from 0 to 100% Buffer B was performed using AKTA prime chromatography system. Protein eluted from the heparin column at approximately 40% Buffer B (500mM NaCl). SDS-PAGE analysis showed that the Trf4-Air2 complex was purified to greater than 80% purity. However, two major contaminating bands (the majority of contaminants) were observed that run approximately 5 and 10 kDa below the Air2 band. These bands were later identified as Air2 degradation products (verified by western blot analysis). Notably, these two degradation bands are observed in every preparation of Air1 and Air2.

The majority of Air2 degradation can be removed using a Phenyl HP hydrophobic interaction column. Therefore, heparin elutions were dialyzed against HIC Buffer A (Buffer A in which the salt component is 1 M AmSO4) and loaded onto a phenyl column. The majority of all Air2 degradation products flow through the phenyl column. A 100 mL gradient of 0-100% HIC Buffer B (Buffer B in which the salt component is 100mM

AmSO$_4$) was then performed. After Phenyl purification the elutions were added to a Superdex 200 26/60 gel-filtration column pre-equilibrated with Buffer A without ZnSO$_4$. Each of the Trf4-Air2 heterodimers elute from the gel filtration column as a single peak according to its molecular weight. The final purified proteins are more than 97% pure and can be concentrated (in a spin concentrator) to at least 15mg/mL. Polyadenylation activity of purified recombinant proteins produced by this protocol has been verified by our collaborators in the Jankowsky lab. Furthermore, I have performed this preparation many times and tested over 2000 crystallization conditions. No crystals were obtained from these trials. The crystal structure of Air2-Trf4 reported by Hamill *et al.* [1] suggests that further truncations of Trf4 and Air2 may be required to obtain crystals.

Although Trf4-Air2 heterodimers can be successfully purified, two major problems exist with the aforementioned protocol. The first major concern is that although these co-expression constructs enable protein expression, the amount of expression is still not optimal. Second, the purification procedure requires several steps over multiple days. Together, these issues prohibit large scale production of purified proteins. The most efficient protein preps that I conducted with these constructs and protocol yielded approximately 1 mg of purified Trf4-Air2 from 1L culture.

Expression and purification of codon optimized Air1 and Air2

Construct design

To improve expression and purification yield of isolated Air1 and Air2 proteins, codon optimized genes (optimized for *E,coli* expression) of Air1 and Air2 were purchased from a commercial vendor (Genscript). Both of the codon optimized genes that

were purchased contain coding sequences for the first amino acid through zinc knuckle

five (N-ZnK5). Specifically, the codon optimized gene sequences were each designed to

terminate according to the last residue of Air2 that is observed in the Air2 (ZnK4-5)-Trf4

crystal structure. In that crystal structure the last Air2 residue is aa198, which resides 31

residues down-stream of ZnK5. In the Air1 sequence, the homologous amino acid is

aa209. Accordingly, the codon optimized genes contain the sequences coding for amino

acids 1-209 of Air1 (Air1 N-ZnK5) and amino acids 1-198 of Air2 (Air2 N-ZnK5). In

addition to coding sequences, the codon optimized genes were designed to also contain

sequences at the N-terminus (preceding the coding sequence) for a TEV cleavable 6x-

histidine tag and a FLAG-tag. The N-terminus of each codon optimized gene is shown in

(Figure 4-1B). The 6x-histidine tag was added to aid purification and the FLAG-tag was

added to facilitate pull-down studies using anti-FLAG resin. Each codon optimized gene

arrived from the vendor in a shuttle vector. Therefore, appropriate restriction enzyme

sites were designed to flank the N- and C-termini in order to cut and paste the sequences

directly into the expression vector pET-Duet1 (cloning site1). In addition, a BamHI site

was also designed into the sequence downstream of the FLAG-tag so that genes without

any tags could also be inserted directly into pET-Duet1 (cloning site1). As a caution to

future researchers, it should be noted that although the FLAG-tagged Air2 sequence can

be inserted directly into pET-Duet1, the BamHI sight is out of frame. This was due to a

mistake in the original sequence design. However, an in-frame Air2 (N-ZnK5) gene

sequence was PCR amplified using the codon optimized gene as a template and cloned

into empty pET-Duet1. To date, Both FLAG-tagged and non-FLAG tagged versions of

codon optimized Air1 and Air2 have been cloned into cloning site 1 of pET-Duet1.
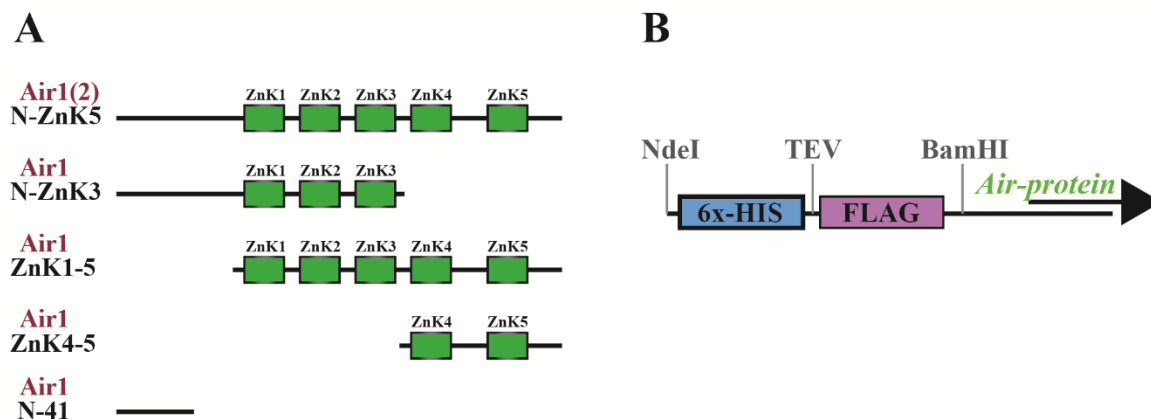
Figure 4-1. Air1 and Air2 codon optimized expression constructs. (A) Overview of Air1 and Air2 codon optimized gene sequence. Each construct shown has been made with and without a FLAG-tag. (B) The N-terminus of codon optimized gene sequences.

To characterize regions of Air1 that are important for function (such as protein-protein interactions), a variety of truncation mutants have been produced by PCR using the codon optimized Air1 gene as a template, followed by restriction cloning into pET duet1 (Figure 4-1A). Design of these truncation mutants focused on different ZnKs of Air1 and the N-terminus. Each of the codon optimized Air1 genes shown in Figure 4-1 can be purified to > 90% purity.

General cautions about Air-protein preparation

Before I describe expression and purification of Air1 proteins, some words of general caution should be given about working with the Air1 or Air2. First, although protein expression is greatly improved by codon optimized genes, about 50-70% of expressed protein is insoluble. This solubility issue has been observed upon expression of all Air1 and Air2 constructs tested. Because of this problem, large growths of 4-6 liters are typically grown for each preparation. Second, degradation products are commonly

observed that run 5-10 kDa below Air-proteins in SDS-PAGE analysis. I have found that the best way to avoid these degradation products is to avoid growing cells beyond the recommended time (indicated below) and maintain protein solutions at 4° C throughout the entire preparation. For example, during lysis I incubate cells in an ice slurry and avoid sonicating more than 15 second intervals at a time. I also perform Ni-column batch binding and elutions in the 4° C room. The final caution is that Air proteins are generally prone to aggregation and precipitation. To help avoid this problem, at least 12% glycerol should be used in all buffers; 12% glycerol is especially required during protein concentration procedures. Protein precipitation also seems to occur while concentrating if imidazole is present in the buffer solution. Air-proteins will also precipitate if there is a sudden change in buffer solutions. For example, I have observed precipitation when buffer exchanging by directly adding a volume of exchange buffer that is more than 25% of the total protein solution volume. Buffer exchanging by dialysis also causes protein precipitation. To avoid protein precipitation during buffer exchanges, researchers should use gel filtration, affinity chromatography, or directly add a volume of exchange buffer that is no more than 20 % of the total protein solution volume. Finally, protein preparations that are prolonged to more than 72 hours typically result in protein aggregation and precipitate. Therefore, Air proteins should be purified in less than three days.

Expression and purification of Air1 proteins

  The following general method was used to express and purify different Air1 proteins expressed by codon optimized pET-Duet1 constructs. Expression constructs

were transformed into in BL21 (DE3) codon+RIL *E.coli* cells (Agilent Technologies).

Single colonies from a fresh transformation plate or a glycerol stock was used to

inoculate an overnight starter culture consisting of 30 mL LB media (in a 75 mL baffled

flask) to which $ZnSO_4$ was added to a final concentration of 250 µM. Appropriate

antibiotics were also added to all culture media (ampicillin and chloramphenicol). Starter

cultures were grown at 37° C for 12-16 hours (overnight). Approximately 5 mL of the

starter culture was then used to inoculate 500 mL of media (also containing 250 uM

$ZnSO_4$) in 2.5 L baffled flasks. The overnight culture can be used to inoculate either LB

or super-broth media (a nutrient rich media). Equal protein is produced by either growth

media, the difference between the two is cost and time. LB growths are less expensive

and require 18 hours of growth after induction, whereas super-broth growths must be

harvested 4 hours after induction. Growth beyond these time-points results in a higher

proportion of insoluble protein and protein degradation. After the overnight culture was

added, the larger cell cultures were incubated at 37° C with shaking at 300 rpm until

culture density reached an OD600 of 0.6-0.8 (1.5-2.5 hours). Protein expression was then

induced by adding 0.3 mM IPTG. Cultures were then moved to a room temperature

shaker (180-200 rpm) for 4-16 hours depending on the growth media (as indicated

above). Cells were harvested by centrifugation and stored at -80° C.

During protein purification, initial buffers including lysis buffer and buffers used

during Ni-column procedures all contained 200 µM $ZnSO_4$. This was done to help ensure

that Air2 zinc knuckles did not exchange zinc ions for nickel ions during the Ni-column

purification step. To begin protein purification, frozen cells were re-suspended in lysis

buffer (Table 1-1) containing 10 mM imidazole and protease inhibitors (1 µg/mL

aprotinin, 1.4 µg/mL pepstatin, and 1 µg/mL leupeptin) and lysozyme (0.2mg/mL).

Approximately 3 mL lysis buffer was added for every 1 mg of cells. Cells were lysed by

sonication followed by centrifugation at 20,000 rpm for 30 minutes at 4° C. The clarified

lysate was then added to 5 mL of GOLDBIO or GE brand nickel-affinity resin for 1 hour

at 4° C. After incubation the Ni-resin was washed with 700 mL lysis buffer, followed by

200 mL buffer B, and then 100 mL buffer A (Table 1-2). These thorough washing steps

are necessary to obtain the best purity results. Throughout all of my preparations of

different Air proteins, The Ni-column step was the most important step for purification,

and extensive washing during this step made a big difference in the final quantity of

purified protein obtained. Protein was eluted from the Ni-column in three elution steps.

Each elution contained 20 mL of buffer A containing 500 mM imidazole which was

incubated on the nickel resin for five minutes prior to elution.

The Ni-column elutions were then loaded directly onto a 5 mL heparin column

pre-equilibrated with buffer A (containing no $ZnSO_4$). A 100 mL gradient from 0 to

100% buffer B was performed using AKTA prime or AKTA purifier chromatography

system. The Air1 proteins typically elute from a heparin column as a very broad peak that

spans about 80 mL from 30% -70% buffer B (400-800 mM NaCl). Purified protein that is

80%-90% pure is typically found in only 10-20 mL of elutions within this broad peak.

This is a point where much protein can be lost (especially if the purification during the

Ni-column step was not efficient). A typical preparation of protein from a 6 L growth

results in 10-15 mg of purified protein after heparin. If 80-90% purity is suitable then

protein can be buffer exchanged at this point and frozen. If increased purification is

needed then gel filtration can be used.

For purification by gel filtration, Air1 protein solutions were first concentrated using a standard spin-concentrator. Protein concentration was carried out by centrifugation in the spin-concentrator at 2200 rpm for 15 minute intervals. Protein was mixed sufficiently between each spin using a 1mL pipette. The protein solution was concentrated to between 1 and 5 mL and then loaded into an injection loop of the AKTA prime/purifier system to which a Superdex 200 26/60 column was attached. The gel filtration column was pre-equilibrated with sizing buffer (Table 4-2) and ran overnight. All of the Air1 proteins (and Air2 proteins) elute off of the sizing column in multiple peaks. One peak is an Air-protein aggregate that elutes in the void volume. The other peaks also primarily contain the Air-protein and different levels of minor contaminants. The contaminants are generally proteins that are 70-80 kDa is size as well as Air protein degradation products. There is always a peak containing purified protein that elutes off the column at a volume that is consistent with an Air protein monomer. However, the size of this peak, and the relative amount of protein that it represents has been inconsistent between multiple preps of the same protein. The reason for this inconsistency is not known. The best method for increasing the amount of purified Air protein monomer on gel filtration is to use a buffer that contains at or below 100 mM NaCl. In all Air-protein preps tested, a NaCl concentration of around 50 mM NaCl drastically reduced the aggregate peak and increased the amount of protein obtained from the peak corresponding to a protein monomer. The most efficient preparations Air1 (N-ZnK5) and Air2 (N-ZnK5) from a 6 L growth after gel filtration produced only 1-2 mg of protein. This small amount is due mainly to the fact that the majority of Air protein either elutes as an aggregate or elutes with other contaminants. Again, the majority of contaminants

are often Air protein degradation products, and fractions that contained these contaminants were not pooled.

Modified purification of Air2 proteins and Air1 ZnK4-5

In addition to the above protocol, the Air2 proteins and only Air1 ZnKs 4-5 can be purified in a much more effective way. The only difference to the Air1 purification protocol is that a mono-Q column is used in place of a heparin column. Two Air2 protein constructs, Air2 N-ZnK5 and N-ZnK3, and Air1 ZnKs4-5 have been purified by this method. In each case the Air-proteins flow through the mono-Q column and all observed contaminants stick to the column. The result is protein that is purified to greater than 98% purity (indicated by SDS-PAGE). In addition, The Air1 ZnK4-5 that was purified using a Q-column eluted off of a gel filtration column as a single peak without aggregation. Furthermore 10 mg of pure protein was obtained from that prep. This is interesting because other preps of Air1 ZnK4-5 that involved a heparin column produced a large aggregate peak during gel filtration. The reason for this difference is not known, but the Q-column method seems to work. Noteworthy, other Air1 proteins stick to the Q-column without further purification. Air2 constructs have not been analyzed by gel filtration following mono-Q purification. This should be done by future researchers.

Co-expression and purification of Air2 (N-ZnK5) and Hmt1 (K13S)

For functional studies regarding Air proteins and Hmt1, undergraduate researchers (Emily Frampton and Kaleb Chatland) and I have begun constructing co-expression constructs of Air-proteins and Hmt1. Currently, we have been successful in producing one of these construct which contains codon optimized Air2 (N-ZnK5) in

cloning site 1 and Hmt1 (K13S) in cloning site 2. We have also been successful in co-purifying these proteins and growing crystals of proteins purified using that protocol.

Expression and purification

Like other described protein expression methods, the Air2-Hmt1 construct was transformed into BL21 (DE3) codon+RIL *E.coli* cells. A single colony from a fresh transformation plate was used to inoculate a 30 mL LB overnight culture supplemented with 250 µM $ZnSO_4$ and antibiotics (ampicillin and chloramphenicol). The overnight starter cultures were grown at 37° C for 16 hours (overnight). Approximately 5 mL of the starter culture was then used to inoculate four 500 mL cultures of LB media (also containing 250 µM $ZnSO_4$) in 2.5 L baffled flasks. The larger cell cultures were then incubated at 37° C with shaking at 300 rpm until culture density reached an OD600 of 0.3. Protein expression was then induced by adding 0.05 mM IPTG. Cultures were then moved to a room temperature shaker that was adjusted to shake flasks at 200 rpm for 20 hours. After 20 hours, cells were harvested and stored at -80° C.

Cells were lysed as described above for Air1 protein preparations. The one difference is that the lysis buffer used contained tris buffer pH 7.5, 50 mM NaCl, and no imidazole. Low salt was used in order to ensure complex formation. Other studies have shown that when imidazole is added, Hmt1 does not co-elute with Air2 during Ni-column purification. Instead, when imidazole is added a higher proportion of equal molar amounts of Air2 and Hmt1 are observed in the flow-through, suggesting that Air2-Hmt1 complexes may have a weaker affinity for Ni-resin compared to isolated Air2. Another interesting feature that should be pointed out is that unlike isolated Air2 preparations, co-

expression with Hmt1 results in greater than 80% soluble Air2 protein and no Air2

degradation products are observed by SDS-PAGE. After nickel column purification the

proteins are greater than 90% pure.

Nickel column elutions were pooled and concentrated to 5 mL using a spin-

concentrator. The concentrator was centrifuged at 2000 rpm at 15 minute intervals and

protein was mixed sufficiently between each spin using a 1mL pipette. After

concentrating to 5 mL the protein solution was loaded into an injection loop of the AKTA

purifier system and loaded onto a Superdex 200 26/60 column. The gel filtration column

was pre-equilibrated with sizing buffer similar to that indicated in Table 4-1. The key

differences again are tris pH 7.5 and 50mM NaCl) and run overnight. The elution profile

of gel filtration showed 6 total peaks and an aggregation peak in the void volume.

According to the chromatogram, the first two peaks (indicated as peak 1 and peak 2 in lab

notebooks) were about three times larger than the other peaks. SDS-PAGE analysis

revealed that peak 2 contained the highest amount of purified Air2 and Hmt1 that

additionally appeared to be in equal molar ratio amounts. Peak 2 eluted from the gel

filtration column at 170 mL which corresponds to a protein complex of about 160-200

kilodaltons. Fractions corresponding to peak 2 were pooled and concentrated to 20

mg/mL according to the extinction coefficient for Hmt1-Air2 (N-ZnK5) which was

predicted by the ExPASy-ProtParam online bioinformatics tool [6].

Improved expression and purification of isolated Air2 (N-ZnK5)

Effective purification of isolated Air2 can be achieved by co-expressing Air2 with

Hmt1. As mentioned in the previous section, co-expression of Air2 (N-ZnK5) with Hmt1

produces Air2 that is greater than 80% soluble and does not appear to have the

degradation products observed when Air2 is expressed alone. After co-expression with

Hmt1, stable Air2 can be isolated using buffers containing high concentrations of salt

during lysis and initial purification steps. Below, is a general description of a method

used to purify stable Air2 (N-ZnK5) after co-expression with Hmt1.

Air2 (N-ZnK5) Purification

For the following protein preparation the *E.coli* growth parameters and protein

expression was identical to that described above for Air2-Hmt1 co-expression. To begin

the lysis procedure, approximately 20 grams of cells (corresponding to 2 liters of LB

culture) were re-suspended in 60 mL Lysis Buffer (Table 4-2) containing protease

inhibitors (1 μg/mL aprotinin, 1.4 μg/mL pepstatin, and 1 μg/mL leupeptin) and

lysozyme (0.2mg/mL). The solution was incubated on ice 20 minutes, sonicated, and then

centrifuged at 18,000 rpm for 40 minutes. The soluble lysate was added to 7 mL GE

brand Nickel-resin (pre-equilibrated with Lysis Buffer) and incubated at 4° C on a

rotating table for 45 minutes. After incubation, the Ni-resin was added to a 50 mL hand-

column and washed with 700 mL Lysis Buffer (Table 4-2) followed by 100 mL Wash

Buffer 1 (Table 4-2). Notably, Wash Buffer 1 contains 1M AmSO$_4$, and washes with that

buffer can be used to help remove any excess Hmt1. The Ni-resin was finally washed

with 100 mL Buffer A. Bound protein was then isolated from the Ni-resin by incubating

the resin for 5 minutes with 15 mL Buffer A containing 300 mM imidazole followed by

elution. The final elution step was repeated three times. Different preps have shown (by

Table 4-2. Buffer solutions for Air protein preparations

Buffer component	Lysis Buffer	Wash Buffer 1	Buffer A	Buffer B	Buffer C
Glycerol	10.00%	10.00%	12.00%	12.00%	12.00%
Tris pH 7.5	50 mM	50 mM	50 mM	50 mM	50 mM
β-ME	2 mM	2 mM	2 mM	2 mM	2 mM
NaCl	1M	----	150mM	150 mM	100 mM
AmSO$_4$	----	1M	----	----	----
ZnSO$_4$	0.2M	0.2M	0.2M	0.2M	----
Volume used	1 L	200 mL	500 mL	500 mL	500 mL

SDS-PAGE) that after Ni-resin purification, Air2 is typically purified to greater than 90% purity with minor protein contaminants around 70-75 kDa in size.

Further purification of Air2 (N-ZnK5) can be obtained using an anion exchange column. The three Ni-column elutions (described above) were pooled (~45 mL total) and loaded onto a 5 mL mono-Q column pre-equilibrated with Buffer A (Table 4-2). A 100 mL gradient from 0 to 100% Buffer B (Table 4-2) was performed using AKTA purifier chromatography system. Importantly, Air2 (N-ZnK5) flows through the mono-Q column while all remaining contaminants do not, resulting in a flow-through sample that contains stable Air2 purified to greater than 98% purity. The purified solution was then buffer exchanged to Buffer C (Table 4-2) using a HiPrep 26/10 desalting column (GE Healthcare Life Sciences). Buffer exchanging by gel filtration is necessary because concentrating with imidazole present results in protein precipitation. After buffer exchanging, the protein was concentrated to approximately 8 mg/mL followed by addition of 20% glycerol and storage at -80° C. Air2 has been successfully purified

multiple times using the above protocol, and each time approximately 5-10 mg of protein

was obtained per liter of cell culture. Moreover, using this protocol provides more than

tenfold the amount of pure Air2 than can be obtained without co-expression.

CRYSTALLIZATION TRIALS

Co-crystallization trials of Air2 (N-ZnK5) and Hmt1 (K13S)

Co-expressed and co-purified Air2 (N-ZnK5) - Hmt1 (K13S) complex (described

previously) was used for sitting-drop vapor diffusion crystallization screening trials using

an Art Robbins Gryphon Crystallization Robot (Art Robbins Instruments) located in the

Johnson lab. Intelli-plate 96-well crystallization plates (Art Robbins Instruments) were

used for crystallization trials. Five different crystallizations screens each containing 96

individual crystallization solutions were used to set up five 96-well crystallization plates.

The crystallizations screens used include the MCSG suite (Microlytic) which consist of

four different screens (MCSG 1-4), and the INDEX screen (Hampton Research).

Crystallization plates were incubated at $4°$ C. After three weeks protein crystals were

observed in many crystallization conditions. Crystals from MCSG 4 at location G4 in the

96 well plate (well solution is 0.1 M Sodium Citrate:HCl pH 5.6, 10% (w/v) PEG 4000,

10% (v/v) 2-Propanol) was analyzed for diffraction quality using a home-source x-ray

generator (MicroMax-007HF) and detector (Rigaku R-Axis IV++). The cryo-solution

used for these crystals consisted of the MCSG4-G4 well-solution with 15% glycerol. All

three crystals that were analyzed exhibited x-ray diffraction that was consistent with

protein crystals. The best diffracting crystal diffracted x-rays to 7 Å. Remaining crystals

in the G4-well were later analyzed by SDS-PAGE and SYPRO Ruby protein gel stain

(SIGMA-ALDRICH). The gel analysis confirmed that these crystals contain both Air2

and Hmt1. Furthermore, another crystal from that same condition was sent to the

synchrotron source SSRL for diffraction analysis. Unfortunately that particular crystal

had poor x-ray diffraction (~20Å). However, the diffraction pattern suggested that the

crystal was indeed protein. Additionally, an x-ray excitation scan was conducted on that

crystal at SSRL which detected the presence of zinc ions, suggesting that the crystal

contained Air2 zinc knuckles. Future researchers should consider this particular crystal

hit for generating Air2-Hmt1 crystals. In addition, other crystallization hits remain

untested for diffraction quality.

Co-crystallization trials of Air2 (N-29) peptide and Mtr4

Crystallization trials have also been initiated for Air2 N-29 peptide with both full

length Mtr4 and archless Mtr4. These crystallization trials were initiated by purifying

Mtr4 proteins according to established protocols. Each of the Mtr4 proteins were

concentrated to 14-46 mg/mL in crystallization buffer (50mM Hepes pH 7.5, 5%

glycerol, 100mM NaCl, 2mM β-ME). Approximately 0.1 g of lyophilized unlabeled Air2

N-29 peptide (synthesized by Dr. Joshua Price at Brigham Young University) was also

suspended in 100 μl of the same crystallization buffer. The Air2 peptide and Mtr4

solutions were then mixed with in a molar ratio of 1.2:1.0 Air2:Mtr4. This protein

solution was incubated for 30 minutes on ice and then crystallization trials were initiated

using the same equipment and screens described above for Air2-Hmt1 crystallization

trials. The only difference is that for each Air2-Mtr4 crystallization trial, the five crystal

screens were set up at both 4° C and room temperature. After 3 months, several different

crystal hits have been identified only in the trials containing archless Mtr4 at room temperature.

Analysis of RNA binding by Air1 and Air2

To date, only one report has confirmed direct interactions of Air proteins with an RNA substrate [7]. In that report NMR and fluorescence anisotropy studies demonstrated that Air2 Zinc knuckles 2-4 were involved in binding hypomodified tRNA[iMet], and ZnK's 1-5 can bind a 15nt poly(A) RNA *in vitro*. In order to detect and further characterize RNA binding by various recombinant Air1/2 proteins, a fluorescence based assay was developed which monitors the change in intrinsic fluorescence of Air1/2 proteins as an RNA substrate is titrated into an Air1/2 protein solution. This RNA binding assay has been used to qualitatively identify binding of purified Air1 (N-ZnK5) and Air2 (N-ZnK5) to both U6 snRNA and tRNA[imet] *in vitro.* Additional purifications of these and other Air1/2-RNA complexes should be used for x-ray crystallography studies.

As an example of the assay procedure; purified Air1 (N-ZnK5) was concentrated to 2 μM in binding buffer (50 mM Hepes pH 7.5, 12% glycerol, 2 mM βME, 50 μM NaCl) and 1.9 mL was added to a 2 mL quartz cuvette. U6 snRNA and tRNA[imet] were titrated with increasing concentrations (2 -20 μM) and the change in fluorescence between 320 nm and 400 nm was monitored using a steady-state-photon counting spectrofluorometer, PC1 with Vinci software, from ISS Instruments (Antony lab, USU). As shown in Figure 4-2, a significant change in fluorescence is observed after the addition of a tenfold molar excess of RNA substrate (20 μM RNA). Similar results have also been obtained for analyzing Air2 (N-ZnK5) binding to tRNA[imet]. This assay can be

Figure 4-2. Air1 (N-ZnK5) binds tRNA^{iMet} *in vitro*. RNA binding results in a change in Air1 intrinsic fluorescence. The change in fluorescence upon addition of 20 μM tRNA^{iMet} was monitored in triplicate.

used as a qualitative assessment of Air-RNA interactions to guide potential crystallization trials or identify RNA substrates and RNA binding regions of Air proteins. Currently, Air protein concentrations below 2 μM have not been tested and future researchers are encouraged to investigate whether a significant change in intrinsic fluorescence (upon RNA binding) can be observed when using nano molar concentrations of Air1/2. If changes in fluorescence can be detected while using low enough concentrations of Air proteins, then it is likely that a quantitative binding curve could be established to further characterize Air-RNA binding interactions.

Figure 4-2. Air1 (N-ZnK5) binds tRNA^{iMet} *in vitro*. RNA binding results in a change in Air1 intrinsic fluorescence. The change in fluorescence upon addition of 20 μM tRNA^{iMet} was monitored in triplicate.

used as a qualitative assessment of Air-RNA interactions to guide potential crystallization trials or identify RNA substrates and RNA binding regions of Air proteins. Currently, Air protein concentrations below 2 μM have not been tested and future researchers are encouraged to investigate whether a significant change in intrinsic fluorescence (upon RNA binding) can be observed when using nano molar concentrations of Air1/2. If changes in fluorescence can be detected while using low enough concentrations of Air proteins, then it is likely that a quantitative binding curve could be established to further characterize Air-RNA binding interactions.

REFERENCES

[1]   S. Hamill, S.L. Wolin, K.M. Reinisch, Proc. Natl. Acad. of Sci. USA 107 (2010) 15045-15050.

[2]   J. LaCava, J. Houseley, C. Saveanu, E. Petfalski, E. Thompson, A. Jacquier, D. Tollervey, Cell 121 (2005) 713-724.

[3]   S. Vanacova, J. Wolf, G. Martin, D. Blank, S. Dettwiler, A. Friedlein, H. Langen, G. Keith, W. Keller, PLoS Biol. 3 (2005) e189.

[4]   A. Sivashanmugam, V. Murray, C. Cui, Y. Zhang, J. Wang, Q. Li, Protein Sci. 18 (2009) 936-948.

[5]   F.W. Studier, Prot. Exp. Pur. 41 (2005) 207-234.

[6]   H.C. Gasteiger E., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A., Proteomics Hrotocols Handbook, in: W. J.M. (Ed.), Protein Identification and Analysis Tools on the ExPASy Server, Humana Press, 2005, pp. 571-607.

[7]   P. Holub, J. Lalakova, H. Cerna, J. Pasulka, M. Sarazova, K. Hrazdilova, M.S. Arce, F. Hobor, R. Stefl, S. Vanacova, Nucl. Acids Res. 40 (2012) 5679-5693.

CHAPTER 5

TRAMP ASSEMBLY INVOLVES BINDING INTERACTION OF

THE N-TERMINUS OF AIR2 AND THE RECA DOMAINS OF MTR4

ABSTRACT

In *Saccharomyces cerevisiae* many non-coding RNAs are processed and degraded by RNA surveillance systems primarily involving the 3'-5' exonucleolytic activities of the nuclear RNA exosome. TRAMP complexes are exosome cofactors that identify and polyadenylate RNAs, a process that stimulates their degradation. Each TRAMP complex is composed of three proteins, a RNA binding protein (Air1 or Air2), a poly(A) polymerase (Trf4 or Trf5) , and a RNA helicase (Mtr4). Previous studies have revealed that Air2 and Trf4 form a tight complex, but how an Air2-Trf4 heterodimer associates with Mtr4, and how formation of TRAMP modulates Mtr4 helicase activity has been poorly characterized. Experiments reported here identify an important interaction site between the N-terminus of Air2 and Mtr4. Additionally we show that a homologous region of Air1 binds Mtr4, indicating common binding interfaces in different TRAMP complexes.

INTRODUCTION

To avoid the accumulation of aberrant and unneeded RNA transcripts the eukaryotic cell has evolved nuclear RNA surveillance pathways which ensure rapid turnover of many types of non-coding RNAs. In these pathways a large multimeric protein complex called the nuclear RNA exosome serves the primary role for RNA

degradation. In Yeast, the nuclear exosome is a ring-like structure composed of nine inactive core subunits (exo9-core) that associate with two catalytically active 3'-5' exo-ribonucleases Rrp44 (exo10-complex) and Rrp6 (exo11-complex) [1-3]. *In vivo*, the specificity and exonucleolytic activity of the exosome requires additional proteins that are collectively referred to as exosome cofactors. TRAMP (Trf4p/Air2p/Mtr4p polyadenylation complex) is a key exosome co-factor that stimulates the degradation of a wide variety of RNA substrates including aberrant forms of tRNAs, snoRNAs, rRNA processing intermediates, cryptic unstable transcripts (CUTs), and pre-mRNA splicing intermediates [4-10].

TRAMP is a three-protein complex composed of a RNA binding zinc-knuckle protein (Air1 or Air2), a non-canonical poly(A) polymerase (Trf4 or Trf5), and a RNA helicase (Mtr4) [10-12]. In the current model of TRAMP function, Air1/2 identifies particular RNA substrates, Trf4/5 adds a short (~4-5 nt) poly(A) tail to the 3' end of the RNA, and Mtr4 unwinds any RNA secondary structure that may exist [10-16]. This unwinding and polyadenylation by TRAMP provides an unstructured single stranded 3' end of the RNA, which is a favorable substrate for the exosome to start degrading [14]. In this way, TRAMP plays a critical role in ridding the cell of non functional and unwanted RNA transcripts. TRAMP mediated RNA surveillance is essential and found throughout eukaryotic species from yeast to humans [13, 17]. In humans, defects in RNA surveillance and processing have been linked to many disease states [18-20], highlighting the importance of understanding the molecular details of RNA surveillance and TRAMP complex function.

In the yeast *Saccharomyces cerevisiae*, each of the Air proteins (Air1 and Air2: 45% seq. id) and Trf proteins (Trf4 and Trf5: 48% seq. id) are parologues that likely arose as result of a whole genome duplication event in *S. cerevisiae* [21], and therefore other eukaryotes have only one ortholog of each. Subsequently, in *S. cerevisiae*, two general types of TRAMP complexes have been described; each one containing Mtr4 and different combinations of Air and Trf components. TRAMP4 contains Trf4 and Air2 or Air1, whereas TRAMP5 contains Trf5 and only Air1 [10-12, 22]. TRAMP4 and TRAMP5 have been shown to have distinct sub-nuclear localizations, and each complex is involved in processing specific RNA substrates. For example, Trf5-GFP and Air1-GFP fusions are primarily enriched within the nucleolus [23-25], where rRNA transcription and processing occurs. Accordingly, TRAMP5 preferentially enhances degradation of aberrant rRNA precursors. TRAMP4 is required for polyadenylation and degradation of structured RNA such as $tRNA_i^{met}$ [12, 23], and regulating levels of transcripts encoding proteins involved in carbon metabolism [26]. Despite these differences in RNA specificity, growth phenotypes indicate that there is some essential functional overlap between the different TRAMP complexes. In *S. cerevisiae*, none of the Trf or Air proteins are individually essential, as single gene deletions do not result in growth defects [27, 28]. However, a Trf4Δ Trf5Δ double mutant is inviable [29], and an Air1Δ Air2Δ double mutant has a severe growth defect [30]. An Mtr4Δ mutant is also inviable [27]. These growth phenotypes further suggest that together the TRAMP complexes (TRAMP4 and TRAMP5) and their individual subunits perform critical functions, but why assembly into a complex is necessary, and how different complexes are assembled is not clear.

Several reports have indicated some functional significance of TRAMP complex formation. A crystal structure of Air2-Trf4 and other biochemical studies have made clear that Air2 and Trf4 form a tight complex [23, 31], and this Air2-Trf4 heterodimer is required for Trf4 polymerase activity *in vitro* [10-12]. It is still unknown how Mtr4 interacts with the other TRAMP subunits. However, functional studies have revealed that when Mtr4 is bound as part of an in-tact TRAMP complex it serves as a critical regulator of the polyadenylation activity of Trf4 on a variety of RNA substrates [16]. Additionally, the Air2-Trf4 heterodimer directly stimulates the helicase unwinding activity of Mtr4 [32]. This 3-way coupling between TRAMP components is presumed to ensure that a preferred number of ~4-5 adenosines are added to the 3'end of RNA substrates, which enhances duplex unwinding by Mtr4 and provides a docking site for the exosome [16, 32]. Although these studies highlight important aspects of TRAMP complex assembly, how the Air2-Trf4 dimer associates with Mtr4, and how this association might facilitate cooperative interactions between protein subunits remains to be described. Understanding these important characteristics of TRAMP requires more detail about how the complex is assembled. In this work, we identify a direct binding interaction between Air2 and Mtr4. In addition, we have narrowed down a binding interface of the two proteins to be located within the first 29 amino acids of Air2 and the ATP binding/hydrolysis domains (RecA1 and RecA2 domains) of Mtr4.

MATERIALS AND METHODS

Recombinant protein expression and purification

The construction, expression, and purification of full length Mtr4 and Mtr4 truncation mutants have been described previously [33]. For Air1 residues 1-41, a codon optimized gene sequence containing an N-terminal FLAG-tag followed by a 6X His-tag was inserted into the co-expression vector pET-Duet1 and transformed into BL21(DE3) RIL cells. Cells were grown using superbroth media, and protein expression was induced during log-phase of growth with 0.5mM IPTG. Cells were harvested approximately four hours after induction. Protein was purified using Nickel NTA resin followed by gel-filtration.

Fluorescence anisotropy

Binding analysis of Mtr4 to Air2 was carried out using fluorescence anisotropy. An Air2 peptide comprising residues 1-29 with an N-terminal fluorescein label was obtained from Dr. Joshua Price at Brigham Young University. Binding reactions were buffered in 50 mM HEPES (pH 7.5), 100 mM sodium chloride, 0.1 mg/mL BSA, and 12% glycerol. Concentrations of the labeled Air2$^{1-29}$ peptide was held constant at 80 nM and titrated with increasing concentrations of Mtr4. Samples were incubated for three minutes after each titration, as changes in anisotropy were not observed beyond this incubation time. Anisotropy at each titration point was measured ten times and averaged. Anisotropy measurements were obtained using a steady-state-photon counting spectrofluorometer, PC1 with Vinci software, from ISS Instruments. Excitation and

emission slits were adjusted to 0.5 nm and temperature was maintained at 25°C. The

excitation wavelength was 495 nm and emission anisotropy was measured at 521 nm.

FLAG-tag pull-down studies

      FLAG-tag co-immunoprecipitation was conducted using batch method in a 2mL

vessel. Purified recombinant FLAG-tagged Air1(residues 1-41) was incubated with 100µl

of ANTI-FLAG M2 affinity gel (Sigma-Aldrich) and washed with binding buffer (50

mM HEPES pH 7.5, 100mM NaCl, 5% glycerol). Purified Mtr4 was then added and

incubated for 1 hour followed by additional washes with binding buffer. Bound proteins

were eluted with binding buffer containing 100 µg/mL FLAG peptide, and visualized by

SDS-PAGE.

RESULTS

A 29 amino acid peptide in the N-terminus of Air2 is sufficient for Mtr4 interaction

      The binding interactions that link together the three protein subunits of TRAMP

have only been described for the binding interface between Air1/2 and Trf4/5 [23, 31,

34]. However, recent reports by Holub *et al.* have indicated that the N-terminus of Air2 is

required for co-immunoprecipitation of an Air2-Trf4 heterodimer with Mtr4 [34],

suggesting that the N-terminus of Air2 is important for TRAMP assembly. To test

whether the N-terminus of Air2 can directly bind Mtr4, we used fluorescence anisotropy

to analyze binding between Air2 residues 1-29 and purified recombinant Mtr4. As shown

in Figure 5-1B, a fluorescently labeled Air2 peptide directly bound to full length Mtr4

with a K_d of 6.7 µM. This binding was specific as no binding was observed for this
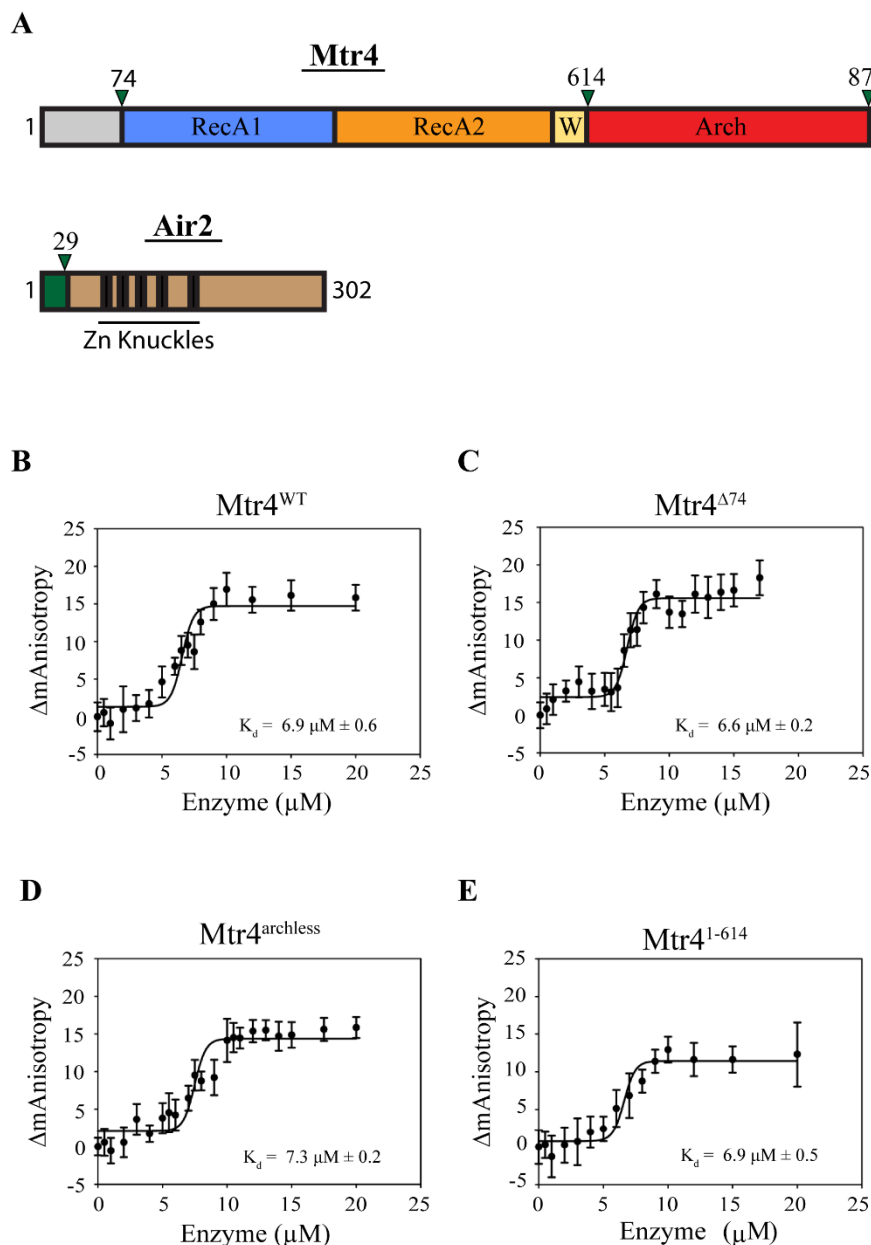
Figure 5-1. A 29 amino acid peptide in the N-terminus of Air2 interacts with the RecA domains of Mtr4. (A) Domain analysis of Mtr4 and Air2. Green triangles indicate various truncation mutants used for anisotropy analysis. B-E: Fluorescence anisotropy analysis of a Air2 peptide binding to (B) Mtr4 full length, (C) Mtr4 with the unstructured N-terminus deleted, (D) Mtr4 with the arch domain deleted, and (E) Mtr4 consisting of the RecA domains and a C-terminal deletion.

peptide to other non-TRAMP proteins (data not shown). To identify the region of Mtr4 used for binding Air2, we further probed Mtr4-Air2 peptide interactions with truncated forms of Mtr4 (Figure 5-1A). Collectively, an N-terminal truncation (Mtr4$^{\Delta74}$, Figure 5-1B), an arch deletion (Mtr4^{archless}; Figure 5-1C), and C-terminal deletion (Mtr4$^{1-614}$, Figure 5-1D), all displayed similar binding affinities. Importantly, Mtr4$^{1-614}$ contains a small portion of the winged helix domain (residues 576-614) and the RecA1 and RecA2 domains (domains 1 and 2 in Mtr4) which are involved in ATPase activity and RNA binding. From these analyses we conclude that the N-terminus of Air2 directly binds to Mtr4. Specifically, this binding interface involves the first 29 residues of Air2 and the region of Mtr4 encompassing both of the RecA domains.

The N-terminus of Air1 can facilitate interactions with Mtr4.

Unlike Air2, Air1 is found in both TRAMP4 (Trf4/Air2(1)/Mtr4) and TRAMP5 (Trf5/Air1/Mtr4) complexes. Unique protein interactions of Air1 and Air2 within TRAMP could be a contributing factor to various substrate specificities or functionality of different TRAMP complexes. It is generally assumed that Air1 and Air2 use the same conserved residues (IWRxYxL, and zinc knuckle 5) for binding Trf4 and Trf5, respectfully [23, 34]. Therefore, we tested if both Air1 and Air2 also use a common binding region for interacting with Mtr4. Sequence alignment of the N-termini of Air1 and Air2 indicate that Air2 residues 1-29 are homologous to the region of Air1 that include residues 1-41 (Figure 5-2). To test if Air1 residues 1-41 can bind Mtr4, we conducted pull-down experiments with purified recombinant Air1 residues 1-41 and Mtr4. Figure 5-2B reveals that FLAG-tagged Air1 residues 1-41 can effectively pull

down Mtr4 *in vitro.* These results indicate that Air1 and Air2 share a common region at

their respective N-termini for binding Mtr4.

A

B
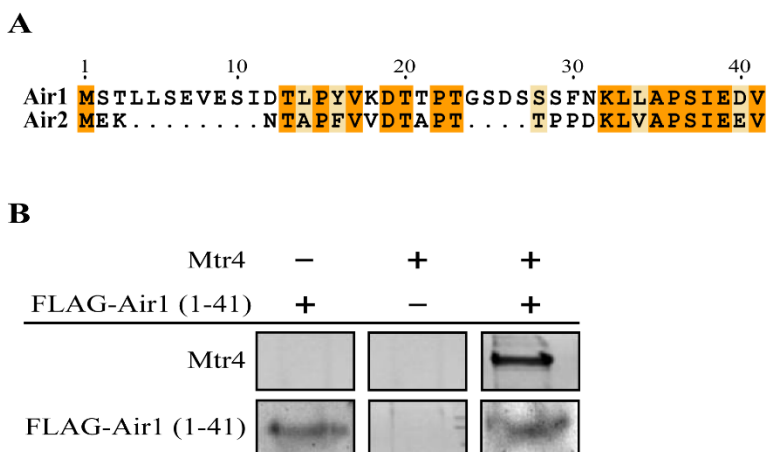
Figure 5-2. Air1 residues 1-41 bind to Mtr4 *in vitro*. (A) Sequence alignment of
Air1 and Air2 N-termini. Identical residues are highlighted orange. Semi-
conserved residues are highlighted light yellow. (B) Pulldown assay with FLAG-
Air1 residues 1-41 and recombinant Mtr4. Bound proteins were eluted with
FLAG peptide and analyzed by SDS-PAGE

DISCUSSION

In this work we have revealed that a small peptide in Air2 can directly bind to

Mtr4, and that a homologous region of Air1 can also form a stable interaction with Mtr4.

The previous crystal structure of Air2 bound to Trf4 shows that Air2 uses its fifth zinc

knuckle and a short sequence upstream of that zinc knuckle to bind Trf4. The Air2

peptide that we have identified as binding to Mtr4 is outside of the Air2-Trf4 binding

interface, and likely represents a region used to tether the Air2-Trf4 heterodimer to Mtr4.

Furthermore, our anisotropy measurements indicate that a region encompassing the two

RecA domains of Mtr4 is used to bind the Air2 peptide, as neither the N-terminal or C-

terminal domains are required for this interaction.

Binding of Air2 to the RecA domains of Mtr4 could explain how the helicase activity of Mtr4 is modulated in TRAMP. In Mtr4 and other Ski2-like helicases, the RecA1 and RecA2 domains contain conserved sequence motifs for binding and hydrolysis of ATP, a requirement for helicase function [35]. It has been shown that TRAMP formation enhances the unwinding activity of Mtr4 by increasing Mtr4s affinity for ATP and increasing rate constants for helicase unwinding [16]. This alteration of Mtr4 function by Air2-Trf4 suggests that within the TRAMP complex there is a functional coupling between Mtr4 and Air2-Trf4. It is easy to speculate that this coupling is provided via binding of Air2 to the RecA domains, which imparts a conformational change in structure that stimulates binding and hydrolysis of ATP. Such an arrangement could ensure that unwinding by Mtr4 is directed at RNA substrates that have first been identified by Air2-Trf4.

During the preparation of this manuscript Falk *et al.* also identified that the N-terminus of Air2 can bind directly to Mtr4 [36]. Specifically, they showed that a dimeric protein complex (connected by a linker) of Trf4 111-490 and Air2 1-190 interacts with Mtr4 with high affinity (K_d = 310 nM). A 49 residue deletion of the Trf4 peptide from that construct (Trf4 160-490) resulted in lower affinity binding with K_d = 6.9 μM. This measurement is very close to our measurements for K_d of the isolated Air2 peptide. Our group has also identified a 26 amino acid peptide of Trf5 (residues 98-124) directly binds Mtr4 with a modest K_d of 10 μM [37]. The sequence of that Trf5 peptide is conserved in Trf4 residues 111-140. Thus, the amino acids corresponding to the isolated Air2 and Trf5 peptides used by our group reside within the Air2-Trf4 fusion peptide used by Falk *et al.* (Trf4 111-490 and Air2 1-190) for their binding studies that showed tight binding to

Mtr4 (K_d = 310 nM). Taken together, these data suggest that a small peptide region of

Air1/2 (reported in this work) and small peptide region of Trf4/5 function together to

bind Mtr4 with high affinity.

The Falk *et al.* manuscript also presented a crystal structure of the Air2-Trf4

fusion peptide bound to Mtr4. The model includes Mtr4 residues 117-1073, Trf4 residues

121-127, and Air2 residues 6-52. The Air2 binding interactions primarily involve Air2

residues 26, 29, 35, and 44, which form electrostatic interactions with domain 4 of Mtr4

(Figure 5-3A). Importantly, unlike our solution binding studies; Falk *et al* did not report

binding interactions between Air2 residues 1-29 to any portion of the RecA domains.

Instead, the crystal structure displays a sharp bend at residue 19 which directs the N-

terminus of Air2 away from the bound Mtr4 molecule where it forms multiple binding

interactions to a symmetry related Mtr4 molecule. This secondary binding interface

involves Air2 residues 7, 8, 11, and a region of the Mtr4 Arch-domain known as the Fist

(Mtr4^{Fist}, residues 665-815).

To verify if binding observed in the crystal structure between Air2 and Mtr4^{Fist}

represents a *bona fide* interaction, we performed anisotropy measurements using our

labeled Air2 peptide to recombinantly purified Mtr4^{Fist}. Figure 5-3-B shows that we

observed no binding between the Air2 peptide and isolated Mtr4^{Fist}. Thus, our solution

binding assays and the crystal structure do not fully agree on the binding interactions

between Air2 and Mtr4. One explanation for the discrepancy is that in the crystal structure

the binding of Air2 to a symmetry related Mtr4 molecule is an artifact of crystallization,

and this crystallization-induced binding precludes native binding of Air2 residues 1-29 to

the RecA domains. Another possibility is that the interactions observed in the crystal

A

B
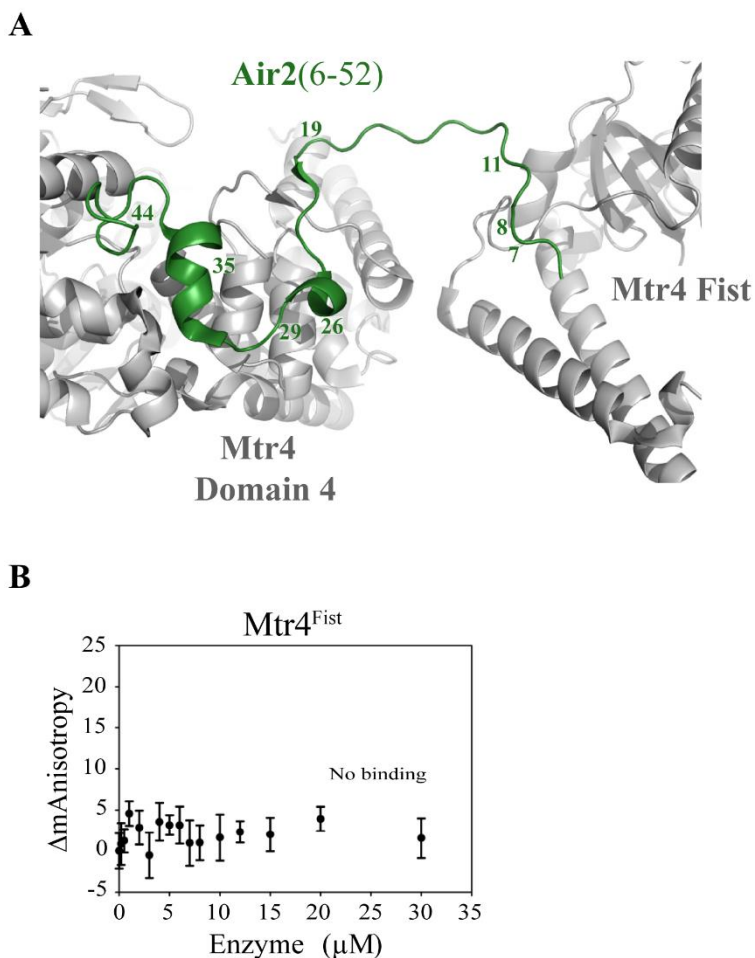
Figure 5-3. Binding interactions between the N-terminus of Air2 and Mtr4[Fist] are not observed *in vitro*. (A) Crystal structure of Mtr4 bound to Trf4 and Air2 peptides. Zoomed- in view highlights the binding interface between the Air2 peptide and two symmetry related molecules of Mtr4. Air2 peptide and Air2 residues involved in binding are colored green. Mtr4 molecules are colored grey. (B) Fluorescence anisotropy analysis of labeled Air2[1-29] peptide binding to recombinant Mtr4[Fist].

structure may be partially influenced by the Air2-Trf4 construct, which was fused together

in a non-native conformation to facilitate crystallization.

In conclusion, we have identified a region of Air2 that directly binds to Mtr4 and

likely bridges together the three proteins of TRAMP. As described above, a recent report

has also identified that the Air2 N-terminus binds to Mtr4. However, our work goes

beyond that described by Falk *et al*, as we have identified a binding interface not observed in the crystal structure and not characterized in any other previous study. Furthermore, this binding interface may explain modulation of Mtr4 reaction parameters observed upon TRAMP complex formation. In addition, we identified that the homologous region of Air1 can also bind to Mtr4 *in vitro*, suggesting that Air2 and Air1 use a common strategy to link the Trf4/5 and Mtr4 components in different TRAMP complexes.

REFERENCES

[1]   F. Bonneau, J. Basquin, J. Ebert, E. Lorentzen, E. Conti, Cell 139 (2009) 547-559.

[2]   Q. Liu, J.C. Greimann, C.D. Lima, Cell 127 (2006) 1223-1237.

[3]   H.W. Wang, J. Wang, F. Ding, K. Callahan, M.A. Bratkowski, J.S. Butler, E. Nogales, A. Ke, Proc. Natl. Acad. Sci. USA 104 (2007) 16844-16849.

[4]   T. Carneiro, C. Carvalho, J. Braga, J. Rino, L. Milligan, D. Tollervey, M. Carmo-Fonseca, Molecular and Cellular Biology 27 (2007) 4157-4165.

[5]   C.A. Davis, M. Ares, Jr., Proc. Natl. Acad. Sci. USA 103 (2006) 3262-3267.

[6]   C. Dez, J. Houseley, D. Tollervey, EMBO J. 25 (2006) 1534-1546.

[7]   P. Grzechnik, J. Kufel, Mol. Cell 32 (2008) 247-258.

[8]   S. Kadaba, A. Krueger, T. Trice, A.M. Krecic, A.G. Hinnebusch, J. Anderson, Genes & Develop. 18 (2004) 1227-1240.

[9]   K.Y. Kong, H.M. Tang, K. Pan, Z. Huang, T.H. Lee, A.G. Hinnebusch, D.Y. Jin, C.M. Wong, Nucl. Acids Res. 42 (2014) 643-660.

[10] F. Wyers, M. Rougemaille, G. Badis, J.C. Rousselle, M.E. Dufour, J. Boulay, B. Regnault, F. Devaux, A. Namane, B. Seraphin, D. Libri, A. Jacquier, Cell 121 (2005) 725-737.

[11] J. LaCava, J. Houseley, C. Saveanu, E. Petfalski, E. Thompson, A. Jacquier, D. Tollervey, Cell 121 (2005) 713-724.

[12] S. Vanacova, J. Wolf, G. Martin, D. Blank, S. Dettwiler, A. Friedlein, H. Langen, G. Keith, W. Keller, PLoS Biol. 3 (2005) e189.

[13] J.T. Anderson, X. Wang, Critical Reviews in Biochemistry and Molecular Biology 44 (2009) 16-24.

[14] J. Houseley, D. Tollervey, Biochimica et Biophysica Acta 1779 (2008) 239-246.

[15] J.M. Houseley, Z. Wang, G.J. Brock, J. Soloway, R. Artero, M. Perez-Alonso, K.M. O'Dell, D.G. Monckton, Human Mol. Genetics 14 (2005) 873-883.

[16] H. Jia, X. Wang, F. Liu, U.P. Guenther, S. Srinivasan, J.T. Anderson, E. Jankowsky, Cell 145 (2012) 890-901.

[17] J. Houseley, D. Tollervey, Cell 136 (2009) 763-776.

[18] D. Astuti, M.R. Morris, W.N. Cooper, R.H. Staals, N.C. Wake, G.A. Fews, H. Gill, D. Gentle, S. Shuib, C.J. Ricketts, T. Cole, A.J. van Essen, R.A. van Lingen, G. Neri, J.M. Opitz, P. Rump, I. Stolte-Dijkstra, F. Muller, G.J. Pruijn, F. Latif, E.R. Maher, Nature Genetics 44 277-284.

[19] A.G. Bassuk, Y.Z. Chen, S.D. Batish, N. Nagan, P. Opal, P.F. Chance, C.L. Bennett, Neurogenetics 8 (2007) 45-49.

[20] O. Bruserud, Current Pharmaceutical Biotechnology 8 (2007) 318-319.

[21] I. Wapinski, A. Pfeffer, N. Friedman, A. Regev, Nature 449 (2007) 54-61.

[22]   J. Houseley, D. Tollervey, EMBO Reports 7 (2006) 205-211.

[23]   M.B. Fasken, S.W. Leung, A. Banerjee, M.O. Kodani, R. Chavez, E.A. Bowman, M.K. Purohit, M.E. Rubinson, E.H. Rubinson, A.H. Corbett, J. of Biol. Chem. 286 (2011) 37429-37445.

[24]   W.K. Huh, J.V. Falvo, L.C. Gerke, A.S. Carroll, R.W. Howson, J.S. Weissman, E.K. O'Shea, Nature 425 (2003) 686-691.

[25]   C. Walowsky, D.J. Fitzhugh, I.B. Castano, J.Y. Ju, N.A. Levin, M.F. Christman, The Journal of Biological Chemistry 274 (1999) 7302-7308.

[26]   K. Schmidt, Z. Xu, D.H. Mathews, J.S. Butler, RNA (2012) 1934-1945.

[27]   J. de la Cruz, D. Kressler, M. Rojo, D. Tollervey, P. Linder, RNA 4 (1998) 1268-1281.

[28]   G. Giaever, C. Nislow, Genetics 197 451-465.

[29]   I.B. Castano, S. Heath-Pagliuso, B.U. Sadoff, D.J. Fitzhugh, M.F. Christman, Nucleic Acids Research 24 (1996) 2404-2410.

[30]   K. Inoue, T. Mizuno, K. Wada, M. Hagiwara, J. of Biol. Chem. 275 (2000) 32793-32799.

[31]   S. Hamill, S.L. Wolin, K.M. Reinisch, Pr Proc. Natl. Acad. Sci. USA 107 (2010) 15045-15050.

[32]   H. Jia, X. Wang, J.T. Anderson, E. Jankowsky, Proc. Natl. Acad. Sci. USA 109 (2011) 7292-7297.

[33]   R.N. Jackson, A.A. Klauer, B.J. Hintze, H. Robinson, A. van Hoof, S.J. Johnson, EMBO J. 29 (2010) 2205-2216.

[34]   P. Holub, J. Lalakova, H. Cerna, J. Pasulka, M. Sarazova, K. Hrazdilova, M.S. Arce, F. Hobor, R. Stefl, S. Vanacova, Nucl. Acids Res. 40 (2012) 5679-5693.

[35]   M.R. Singleton, M.S. Dillingham, D.B. Wigley, Annual Review of Biochemistry 76 (2007) 23-50.

[36]   S. Falk, J.R. Weir, J. Hentschel, P. Reichelt, F. Bonneau, E. Conti, Molecular Cell 55 (2014) 856-867.

[37]   J.S. Losh, A.K. King, J. Bakelar, L. Taylor, J. Loomis, J.A. Rosenzweig, S.J. Johnson, A. van Hoof, Nucl. Acids Res. 1 (2015)

CHAPTER 6

SUMMARY AND FUTURE DIRECTIONS

Summary: The dehydrogenases *R*-and *S*-HPCDH

      Throughout biology there are few examples of homologous enzymes that function within the same biological pathway to catalyze similar reactions with opposite stereospecificity. This unique characteristic of *R*- and *S*-HPCDH makes them an ideal model system to study enzymatic mechanisms of substrate recognition, specifically stereospecificity. When I began this work a substantial amount of functional and mechanistic studies had already characterized many stereospecific properties of *R*- and *S*-HPCDH. For example, kinetic studies revealed that stereospecificity is governed by alternative kinetic mechanisms in each enzyme [1]. However, the structural differences between the enzymes that enable them to have opposing stereospecificity had not been determined. Chapter 4 describes the first x-ray crystal structure of *S*-HPCDH, and the first substrate-bound crystal structure of either enzyme. Using the *S*-HPCDH crystal structure and a previously determined crystal structure of *R*-SHPCDH, I presented a structural comparison of *R*-and *S*-HPCDH that revealed previously unknown sequence and structural differences employed by each enzyme to facilitate stereospecificity.

      The main structural difference between *R*- and *S*-HPCDH is that each enzyme has a unique substrate binding pocket. When compared to each other, the substrate binding pockets share a common catalytic site, but differ in orientation as they lead away from the catalytic site to alternative substrate binding residues at the periphery of each pocket. In addition, the substrate-bound form of *S*-HPCDH provided the ability to analyze substrate

binding in these enzymes for the first time. These observations combined with previously reported kinetic data now provide a thorough model of stereospecificity for *R*- and *S*-HPCDH. In *S*-HPCDH, the difference in substrate binding pocket orientation provides a binding pocket that accommodates the *S*-substrate isoform but sterically hinders binding of the *R*-substrate isoform. Specifically, when the *R*-substrate (*R*-HPC) is modeled into the active site of *S*-HPCDH the C2 methyl group clashes directly with a catalytic serine residue. The substrate binding pocket of *R*-HPCDH appears to have a more open conformation that can accommodate both substrate isomers with equal propensity. However, as mentioned by previous models [1], binding of the *S*-substrate in *R*-HPCDH likely leads to opposite orientation of substrate reactive groups at the catalytic site, and therefore effective catalysis does not occur for the *S*-substrate in *R*-HPCDH. These structural differences correlate nicely with kinetic data that indicated stereospecificity of *S*-HPCDH is provided by a large difference in the value of K_m between different substrates isoforms, whereas stereospecificity of *R*-HPCDH is controlled by large differences in the value of k_{cat} between different substrate isomers. To our knowledge, the analysis given in Chapter 4 represents the first side-by-side structural comparison of two SDR-enzymes that function together to act upon two alcohols in the same metabolic pathway. Furthermore, this information was essential to clarify the mechanisms of stereospecificity used by *R*-and *S*-HPCDH in the unique pathway of bacterial epoxide carboxylation.

Summary: Air proteins and RNA regulation

The conserved zinc knuckle proteins Air1 and Air2 (collectively referred to as Air proteins) are required for RNA processing and RNA surveillance in eukaryotes. Each Air protein functions as part of a trimeric protein complex called TRAMP (TRAMP4 and TRAMP5) [2-5], which serves to activate 3'-end degradation of targeted RNA substrates by the nuclear exosome. Previous studies have indicated that within TRAMP4 and TRAMP5 the Air proteins play a central role in both RNA binding and forming protein-protein interactions [6-13] (Figure 6-1). However, the binding interactions between Air proteins and various binding partners (protein and RNA) have remained poorly characterized.

The research presented in this dissertation has extended the knowledge of Air protein binding interactions, and provided a foundation for future research aimed at exploring Air protein function. The principal finding of this work regarding the Air proteins is the characterization of a previously unknown binding interface between Air2 and another TRAMP component, the helicase Mtr4 (Chapter 5). The binding region of the two proteins was narrowed down to the first 29 residues of Air2 and the RecA1 and RecA2 domains of Mtr4. Because the RecA domains of Mtr4 are used for binding and hydrolysis of ATP, binding of Air2 to this region could explain why both ATP binding affinity and helicase activity of Mtr4 are enhanced upon formation of the TRAMP complex [14, 15]. Notably, binding of accessory proteins has been shown to increase ATP affinity and helicase activity in many other helicases [16-20]. Molecular detail of the identified Air2-Mtr4 interaction and how such an interaction modulates Mtr4 activity remains to be determined through future biochemical and structural studies.
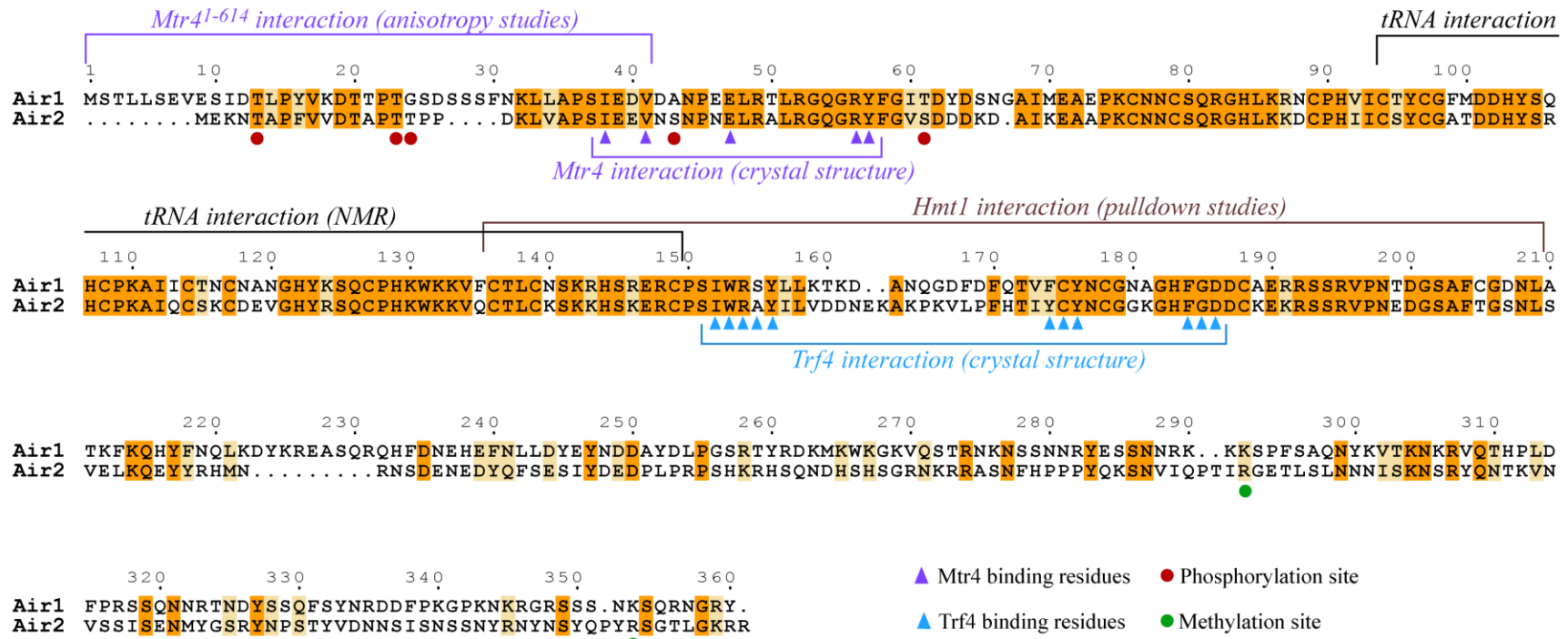
Figure 6-1. Characterized protein-protein and RNA binding interfaces of Air proteins. Air1 and Air2 sequences are aligned with conserved residues colored orange and semi-conserved residues colored light yellow. The indicated interactions are those characterized for Air2 binding to Mtr4 [6], tRNA [7], Trf4 [8], and Hmt1 (unpublished data). Colored triangles indicate specific residues of Air2 involved in protein binding. Colored circles indicate Air2 residues that are sites of posttranslational modifications [9-11].

In addition to binding Mtr4, the Air proteins also bind to different types of RNA and to other proteins such as the methyltransferase Hmt1 [13, 21] (Figure 6-1). These additional binding interactions of Air1 and Air2 are also not well characterized, and several outstanding questions remain about Air protein function in these processes.detailed description of future directions including existing preliminary data aimed at answering remaining questions regarding Air protein binding interactions is given below.

Future directions: Air proteins and protein binding interactions

Does Air2 modulate Mtr4 activity?

Fluorescence anisotropy studies have shown that Air2 residues 1-29 (Air2$^{1-29}$) can directly bind to the RecA domains of Mtr4 (Chapter 5), and it was postulated that binding of Air2 to this region may enhance ATP affinity and/or helicase activity. More compelling evidence is required to identify if binding of Air2$^{1-29}$ to Mtr4 results in modulation Mtr4 reaction parameters. This possibility can be easily tested in the Johnson lab using routine assays to analyze Mtr4 ATP binding [15], ATP hydrolysis [22], and helicase unwinding [22], all in the presence of unlabeled Air2$^{1-29}$ peptide. The Johnson lab currently has approximately 10 mg of unlabeled Air2 1-29 peptide (synthesized by Dr. Joshua Price at Brigham Young University) and other materials that are needed to conduct these experiments.

How does the N-terminus of Air2 specifically interact with Mtr4 RecA domains?

Characterization of the molecular interactions involved in forming the binding interface between Air2[1-29] and the Mtr4 RecA domains requires identifying the specific amino acids involved in binding. As mentioned in Chapter 5, Falk *et al*. recently reported a crystal structure and biochemical data that also identified that the N-terminus of Air2 could associate with Mtr4 [6]. However, in that study none of the first 29 amino acids of Air2 were reported to bind to the RecA domains of Mtr4. Instead, the crystal structure showed that Air2 residues 7, 8, and 11 were directly bound to the Arch domain of a symmetry related Mtr4 molecule. Using fluorescence anisotropy, I was able to demonstrate that this interaction between Air2 and the Mtr4 Arch was likely not a genuine interaction but rather an artifact of crystallization (Chapter 5). The Falk *et al*. manuscript did however identify that residues 26 and 29 form hydrophobic interactions with Mtr4 Met1016 (domain4; ratchet domain). *In vitro* pull-down studies further confirmed this interaction, as substitution of Mtr4 Met1016 (M1016E) disrupted Mtr4 interactions with GST-Air2. Therefore, Air2 residues 26 and 29 likely form the binding interactions described by Falk *et al*, and do not bind the RecA domains.

Because TRAMP complexes are found throughout eukaryotes it's likely that specific residues important for mediating important protein-protein interactions are also conserved. Therefore, primary sequence analysis of various Air2 homologues might serve as a useful tool to identify conserved residues that are important for binding to Mtr4. The sequence alignment of the N-termini of yeast Air homologues highlights that several residues within the first 22 amino acids are conserved or semi-conserved (Figure 6-2).

```
            1                    10                 20
            .                    .                  .
S.cerevisae Air2 ....MEK...........NTAPFVV..DTAPT....TPPDKLVAPSIEEV
S.cerevisae Air1 ....MSTLLSEVES...IDTLPYVK..DTTPTGSDSSSFNKLLAPSIEDV
   S. Pombe Air1 MDMTVSNQKAESDDGSDIDDAALLQKINSLPIDQSITNSVSLEKHDFQGS
```

Figure 6-2. N. terminus sequence alignment of Air protein homologues in yeast. Conserved sequences are highlighted orange. Semi-conserved sequences are highlighted light yellow. Residues are numbered according to Air2 sequence.

In future research, these conserved sequences should be considered for site directed mutagenesis followed by additional binding studies with Mtr4 RecA domains to analyze their contributions to binding. In addition, if it is also confirmed that Air2$^{1-29}$ can modulate Mtr4 function, then these conserved residues are also great candidates for site directed mutagenesis to analyze their contribution to regulating Mtr4 activity.

Obtaining a crystal structure of Air2 Air2$^{1-29}$ and Mtr4 would be an ideal way to identify specific binding residues and understand how binding interactions could impact Mtr4 function. Efforts to obtain crystals of Air2$^{1-29}$ and Mtr4 have been carried out with promising success. I have initiated sitting drop co-crystallization trials with a non-labeled Air2$^{1-29}$ peptide and both full length Mtr4 and an archless version of Mtr4. To date, several different crystal hits have been identified only in the trials containing archless Mtr4 (details can be found in Chapter 4). These crystals have not yet been analyzed for diffraction quality nor have they been confirmed to be protein. However, it's likely that these crystals are indeed protein crystals, as none of the crystal hits contain crystals in the reservoir solution and most of the crystallization solutions do not contain compounds that typically form salt crystals. Furthermore, Regardless of diffraction quality, these initial crystallization hits provide a great foundation for future trials.

Do Air1 and Air2 exhibit a common binding mode for interaction with Mtr4?

In addition to fluorescence anisotropy studies which showed Air2$^{1-29}$ can bind full length and truncated forms of Mtr4 with similar affinity, I also demonstrated recombinant Air1 residues 1-41 (Air1$^{1-41}$) can interact with full length Mtr4 in co-IP experiments (Chapter 5), suggesting that this region of Air1 binds to Mtr4 in a similar fashion as Air2$^{1-29}$. Notably, Air1$^{1-41}$ consists of sequences that are homologous to Air2$^{1-29}$. In order to more clearly verify that this region of Air1 and Air2 contains a common Mtr4 binding interface, additional binding studies are needed to determine if Air1$^{1-41}$ also associates with Mtr4 RecA domains, and if the binding affinity (K_d) between Air1$^{1-41}$ and Mtr4 is comparable to that observed for Air2$^{1-29}$ and Mtr4. To this end, I have initiated preliminary studies using florescence anisotropy to analyze binding between an N-terminal labelled Air1$^{1-41}$ peptide (synthesized by the Price lab at Brigham Young University) and recombinant full length Mtr4. All of the experimental parameters (i.e. instrumentation, buffer, temperature, protein concentrations, etc.) were analogous to previous anisotropy studies using Air2$^{1-29}$ (Chapters 4 and 5) However, unlike anisotropy studies with Air2$^{1-29}$, binding interactions between Air1$^{1-41}$ and Mtr4 was not detected. The reason for this observed difference between the Air1$^{1-41}$ and Air2$^{1-29}$ anisotropy results, and the discrepancy between Air1$^{1-41}$ anisotropy and Air1$^{1-41}$ co-IP results is not clear. One possibility could be that during anisotropy studies the Air1 peptide experiences non-specific binding interactions with itself that do not occur in the Air2 peptide or during co-IP studies. This reasoning is based on the fact that Air1$^{1-41}$ contains 12 additional residues (residues 4-11, 24-27) that are not observed in Air2$^{1-29}$ (Figure 6-2) It is possible that during anisotropy experiments, where the Air1 peptide has free-motion

in solution, the extra Air1 residues form non-specific binding interactions that prevent binding to Mtr4. In contrast, the co-IP experiments were conducted by first binding the FLAG-tagged N-terminus of Air1[1-41] to anti-FLAG resin prior to addition of Mtr4. This immobilization of the N-terminus of Air1 may have prohibited non-specific binding interactions, allowing Mtr4 to bind and co-immunoprecipitate with Air1[1-41]. To circumvent such a problem, future researchers should consider using other methods for characterizing Air1-Mtr4 binding interactions in which the N-terminus of Air1 can be attached to a solid surface and remain stationary. One such method is SPR (Surface Plasmon Resonance), which is currently available at USU.

**Future directions: Air proteins and Hmt1 regulation**

In addition to their role in TRAMP mediated RNA degradation, Air1 and Air2 have also been reported to function as regulators of nuclear mRNA export. This regulation is attained by directly modulating the activity of the arginine methyltransferase protein Hmt1 [21]. In budding yeast, methylation by Hmt1 is required to activate the nuclear mRNA transport protein Npl3. In its non-methylated state, Npl3 no longer effectively delivers mRNA to the cytoplasm. Air1 has been shown to directly inhibit Hmt1 methylation of Npl3 *in vitro* and *in vivo*; suggesting that binding of Air1 to Hmt1 inhibits methylation, and thus inhibits mRNA export. Although it is clear that Air1 can function to inhibit Hmt1, details of how protein-protein interactions facilitate Hmt1 inhibition are unknown. In addition, Because Air2 can bind to Hmt1 *in vitro,* it has been assumed that Air2 also inhibits Hmt1 activity. However, inhibition of Hmt1 by Air2 has never been reported.

Over the last two decades, the functional importance of protein arginine methylation has become more recognized. This post-translational modification has been shown to regulate many critical cellular processes including DNA repair, chromatin maintenance, gene expression, and translation [23]. Surprisingly, there are few examples of how methylation itself is regulated within the cell. Hmt1 is the primary arginine methyl transferase in *S. cerevisiae*, and Air1 is the only protein that has been shown to regulate its activity. Understanding the mechanistic details of how Air1 inhibits Hmt1 activity would provide valuable insight to how arginine methylation is regulated in eukaryotes. As part of my dissertation work I have initiated research to address outstanding questions regarding Air protein mediated inhibition of Hmt1. In the text below, I describe preliminary data and future directions aimed at answering these questions.

How do the Air proteins bind Hmt1?

The first 54 amino acids of Hmt1 have been identified to be required for binding the Air proteins, but which regions of the Air proteins are involved in associating with Hmt1 have not been described. Therefore, to further identify the Air-Hmt1 binding interface my work has focused on identifying the protein regions of Air1 and Air2 required for Hmt1 binding. As described in previous chapters, each Air protein contains five zinc knuckles that are flanked on each side by extended N- and C-terminal sequences. To identify which regions are important for binding Hmt1, I have designed several expression constructs of Air1 and Air2 truncation mutants (details in Chapter 4) which have been used to qualitatively assess Air-Hmt1 binding using *in vitro* pull-down

and gel filtration analysis. Preliminary data has indicated that the extended C-terminus of

Air1 and Air2 is not required for Hmt1 binding. Pulldown experiments using FLAG-tag

Air1 indicate that an Air1 protein containing the N-terminus through zinc knuckle 5 (Air1

N-ZnK5) can bind Hmt1. Binding is also observed with a more highly truncated Air1

protein consisting of only zinc knuckles 4-5 (Air1 ZnK 4-5) (Figure 6-3A). Interestingly,

Air1 ZnK4-5 appears to pulldown Hmt1 more effectively than the longer Air1 N-ZnK5

protein. The reason for this increased binding by ZnKs 4-5 is currently not known.

Although this data clearly indicates that ZnKs 4-5 of Air1 can effectively bind Hmt1, it

does not exclude the possibility of additional binding by other regions of Air1.

A

B
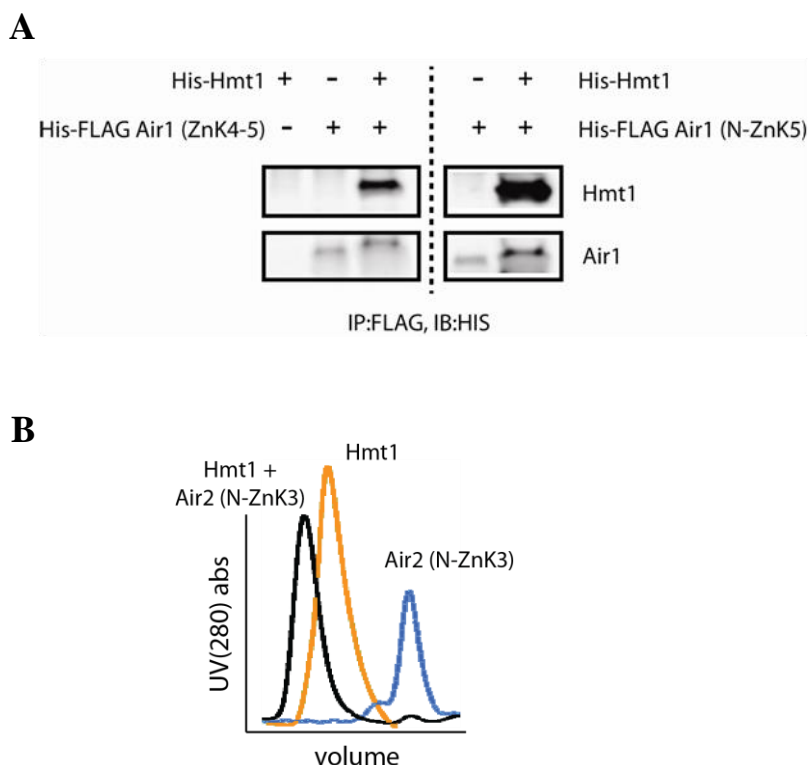
Figure 6-3. Protein binding analysis of Air proteins and Hmt1. (A) FLAG-tagged Air1 (N-ZnK5) and Air1 (ZnK4-5) pull down Hmt1 on anti-FLAG resin. Samples were run on SDS-PAGE and visualized by anti-His antibody. (B) Purified Hmt1 and Air2 (N-ZnK3) co-elute on gel filtration, forming a complex with apparent 1:1 stoichiometry.

Binding of Hmt1 by Air2 has been analyzed by gel filtration. This analysis was conducted using individually purified Air2 (Air2 N-ZnK3) and Hmt1 proteins. Notably, due to protein availability, Air2 N-ZnK3 was the only isolated Air2 protein used in this experiment. Each protein was run separately and together over a gel filtration column (Superdex 200). Analysis of the chromatogram in Figure 6-3B shows that Air2 N-ZnK3 and Hmt1 co-elute as a higher molecular weight peak compared to the individual proteins, indicating that Air2 N-ZnK3 and Hmt1 form a stable complex *in vitro*. It is still unknown whether Air2 ZnK4-5 can also bind to Hmt1. Future binding studies need to be conducted using additional truncation mutants of both Air1 and Air2 to assess (1) if they share common Hmt1 binding regions and (2) identify the smallest portion of each protein that binds Hmt1. Once minimal binding regions have been identified, other binding analyses could be used (such as ITC, SPR, fluorescence anisotropy, or EMSAs) to quantify the various binding interactions and assess the binding contribution of particular regions (or residues) of the Air proteins. Such studies could additionally be used to identify tight-binding complexes that may hold promise for crystallization trials and x-ray diffraction.

Obtaining high resolution crystal structures of Air1/2 and Hmt1 would be the most informative piece of data to characterize the Air-Hmt1 binding interface. I have initiated crystallization trials with Air2 truncation mutants and Hmt1 with success. These successful crystallization trials were performed using Air2 N-ZnK5 and an Hmt1 mutant which contained a point mutation at the thirteenth amino acid (K13S). As described in Chapter 4, both proteins were co-expressed using a pET-Duet1 plasmid and co-purified using affinity chromatography and gel filtration. Crystallization screening produced

crystals in many conditions. The most promising of these hits produced protein crystals that diffracted to 7 Å using a home-source x-ray generator at USU (Figure 6-4). An SDS-PAGE gel confirmed that these crystals contain both Air2 and Hmt1. Furthermore, another crystal from that same condition was sent to the synchrotron source SSRL for diffraction analysis. Unfortunately that particular crystal had poor x-ray diffraction (~20Å). However, the diffraction pattern suggested that the crystal was indeed protein. Additionally, an x-ray excitation scan was conducted on that crystal at SSRL which detected the presence zinc ions, suggesting that the crystal contained zinc knuckles. Further optimization of this crystal condition may allow for x-ray diffraction of crystals to a resolution greater than 7Å. Future researchers are encouraged to *chase* this crystal hit and other crystal hits indicated in Chapter 4.
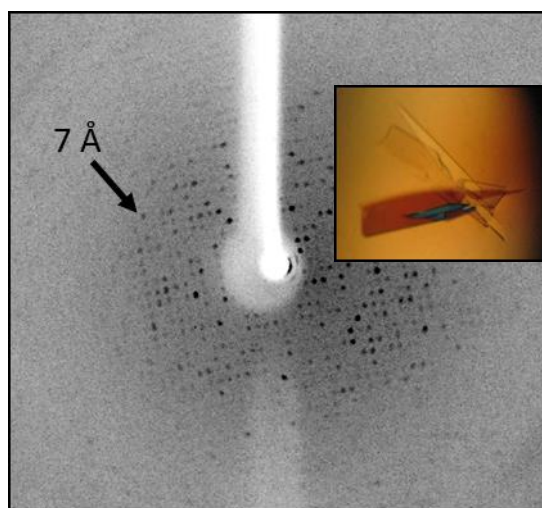

Figure 6-4. Air2 (N-ZnK5)-Hmt1 crystals and X-ray diffraction.

How do the air proteins regulate Hmt1 activity?

Although it has been established that Air1 inhibits Hmt1 mediated methylation of Npl3, the mechanism of inhibition is unclear. It has also remained unclear whether Air2 also inhibits Hmt1 activity. A collaborative effort between our group (Johnson lab) and the Hevel lab group has begun preliminary studies to identify the mechanistic detail of Hmt1 inhibition by the Air proteins. These preliminary studies have involved using an *in vitro* methyltransferase assay (developed by the Hevel lab) to monitor Hmt1 methylation of Npl3. Using that assay it was shown that an Air1 protein consisting of N-ZnK5 can effectively inhibit methylation by Hmt1, but an Air1 protein consisting of N-ZnK3 could not. This observation indicates that ZnKs 4-5 of Air1 are required for inhibition of Hmt1. This result is consistent with the pull-down binding analysis (mentioned above) which indicated Air1 ZnK4-5 could bind to Hmt1. Interestingly, additional methyltransferase assays showed no inhibition of Hmt1 when Air2 N-ZnK-5 was used in the reaction setup, indicating for the first time that Air2 does not inhibit Hmt1 methylation. This result highlights a critical flaw in the original Air-Hmt1 manuscript which concluded that Air1 and Air2 are functionally redundant in their effect of Hmt1 activity. The molecular detail of the difference in Hmt1 inhibition by Air1 and Air2 is still unknown and is the subject of future work.

It's very intriguing that both Air1 and Air2 can bind Hmt1 but only Air1 inhibits methylation. A promising avenue for identifying the specific amino acids of Air1 involved in Hmt1 inhibition is to identify sequence differences within ZnKs 4-5 of Air1 and Air2. All current knowledge indicates that both Air1 and Air2 form a tight complex with Trf4 and Trf5 *in vivo*, and there is no evidence to suggest that either Air1 or Air2

exists independent of Trf4 or Trf5 in the cell. Therefore, Air1 sequences that should be

considered as contributors to Hmt1 inhibition are those that differ from sequences in Air2

and those that are not characterized as Trf4/Trf5 binding residues. Figure 6-5 highlights

that the majority of such residues reside in the linker region (L4) between ZnK4 and

ZnK5. In future studies those residues should be the target of site directed mutagenesis

and construction of chimeric proteins. Those mutant constructs should then be tested in

Hmt1 methylation assays to define the sequences required for inhibition of Hmt1.
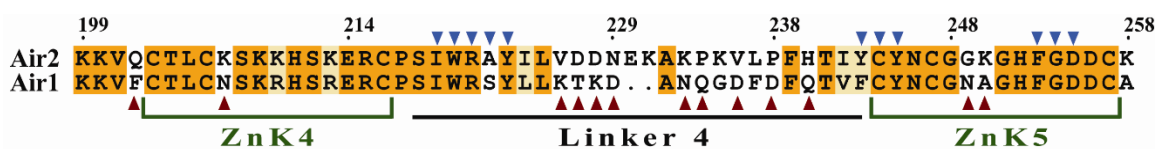
Figure 6-5. Sequence alignment of Air1 and Air2 zinc knuckles 4-5. Conserved residues are colored orange. Semi-conserved residues are colored light yellow. Blue triangles indicate residues that interact with Trf4 according to Air2-Trf4 structure (pdb 3NYB). Red arrows indicate Air1 residues that differ from Air2, and are potential targets for mutagenesis and functional studies. Residues are numbered according to Air2 sequence.

Conclusion

The research presented in this dissertation has provided a promising foundation

for future research efforts regarding Air proteins and their various binding interactions.

Air-Mtr4 studies have identified an important binding interaction that can be explored

using known techniques that are familiar to researchers in the Johnson lab. I have also

narrowed down another binding interface between Air1 and Hmt1. My efforts in that

project have provided a new and continuing avenue of research for both the Johnson and

Hevel labs. In addition, I have developed purification protocols and initiated

crystallization studies for both Air2-Mtr4 and Air2-Hmt1. Crystallization hits have been

identified in each case, and future repetition and modification of those studies has great

potential for leading to high resolution crystal structures. Future researchers are strongly encouraged to conduct follow-up crystallization studies. Furthermore, several crystals that were produced from Air2-Mtr4 and Air2-Hmt1 crystallization trials currently remain undisturbed in crystallization trays, and are awaiting x-ray diffraction analysis.

REFERENCES

[1]	D.A. Sliwa, A.M. Krishnakumar, J.W. Peters, S.A. Ensign, Biochemistry 49 (2010) 3487-3498

[2]	J. Houseley, J. LaCava, D. Tollervey, Nature Reviews 7 (2006) 529-539.

[3]	J. LaCava, J. Houseley, C. Saveanu, E. Petfalski, E. Thompson, A. Jacquier, D. Tollervey, Cell 121 (2005) 713-724.

[4]	S. Vanacova, J. Wolf, G. Martin, D. Blank, S. Dettwiler, A. Friedlein, H. Langen, G. Keith, W. Keller, PLoS Biology 3 (2005) e189.

[5]	F. Wyers, M. Rougemaille, G. Badis, J.C. Rousselle, M.E. Dufour, J. Boulay, B. Regnault, F. Devaux, A. Namane, B. Seraphin, D. Libri, A. Jacquier, Cell 121 (2005) 725-737. (2010) 3487-3498.

[6]	S. Falk, J.R. Weir, J. Hentschel, P. Reichelt, F. Bonneau, E. Conti, Molec. Cell 55 (2014) 856-867.

[7]	P. Holub, J. Lalakova, H. Cerna, J. Pasulka, M. Sarazova, K. Hrazdilova, M.S. Arce, F. Hobor, R. Stefl, S. Vanacova, Nucl. Acids Res. 40 (2012) 5679-5693.

[8]	S. Hamill, S.L. Wolin, K.M. Reinisch, Proc. Natl. Acad. Sci. USA 107 (2010) 15045-15050.

[9] F. Gnad, L.M. de Godoy, J. Cox, N. Neuhauser, S. Ren, J.V. Olsen, M. Mann, Proteomics 9 (2009) 4642-4652.

[10] M.A. Erce, D. Abeygunawardena, J.K. Low, G. Hart-Smith, M.R. Wilkins, Mol. Cell Proteomics 12 (2013) 3184-3198.

[11] M.B. Smolka, C.P. Albuquerque, S.H. Chen, H. Zhou, Proc. Natl. Acad. Sci. USA 104 (2007) 10364-10369.

[12] M.B. Fasken, S.W. Leung, A. Banerjee, M.O. Kodani, R. Chavez, E.A. Bowman, M.K. Purohit, M.E. Rubinson, E.H. Rubinson, A.H. Corbett, J. Biol. Chem. 286 (2011) 37429-37445.

[13] K. Schmidt, Z. Xu, D.H. Mathews, J.S. Butler, RNA 18 (2012) 1934-1945.

[14] H. Jia, X. Wang, J.T. Anderson, E. Jankowsky, Proc. Natl. Acad. Sci. USA 109 (2011) 7292-7297.

[15] H. Jia, X. Wang, F. Liu, U.P. Guenther, S. Srinivasan, J.T. Anderson, E. Jankowsky, Cell 145 (2012) 890-901.

[16] Z.Y. Dossani, C.S. Weirich, J.P. Erzberger, J.M. Berger, K. Weis, Proc. Natl. Acad. Sci. USA 106 (2009) 16251-16256.

[17] N.L. Korneeva, E.A. First, C.A. Benoit, R.E. Rhoads, J. Biol. Chem. 280 (2005) 1872-1881.

[18] J.P. Staley, J.L. Woolford, Jr., Current opinion in Cell Biology 21 (2009) 109-118.

[19] N. Tanaka, A. Aronova, B. Schwer, Genes & Development 21 (2007) 2312-2325.

[20] T. Tanaka, T. Mizukoshi, K. Sasaki, D. Kohda, H. Masai, J. Biol. Chem. 282 (2007) 19917-19927.

[21]   K. Inoue, T. Mizuno, K. Wada, M. Hagiwara, J. Biol. Chem. 275 (2000) 32793-32799.

[22]   L.L. Taylor, R.N. Jackson, M. Rexhepaj, A.K. King, L.K. Lott, A. van Hoof, S.J. Johnson, Nucl. Acids Res. 42 (2014) 13861-13872.

[23]   M.T. Bedford, S. Richard, Molec. Cell 18 (2005) 263-272.

APPENDIX

Permissions

Personal (scholarly) purposes

Authors can use their articles, in full or in part, for a wide range of scholarly, non-commercial purposes as outlined below:

- Share copies of the article and distribute them via email to colleagues for their research use (also known as 'scholarly sharing').

- Share the article for personal use or for the author's own classroom teaching.

- Use the article at a conference, meeting or for teaching purposes.

- Allow the author's employers to use the article for other internal purposes (such as training).

- Include the article in a printed compilation of the author's works, such as collected writings and lecture notes.

- Inclusion the article in a thesis or dissertation

- Use the article in full or in part to prepare other derivative works, including expanding the article to book-length form, with each work to include full acknowledgement of the article' original publication.

These rights apply for all Elsevier authors who publish their article as either a subscription article or an open access article. In all cases we require that all Elsevier authors always include a full acknowledgement and, if appropriate, a link to the final published version hosted on Science Direct.

CURRICULUM VITAE

Jeremy W. Bakelar

PERSONAL

Date of Birth: July 3, 1981
Place of Birth: Salt Lake City, Utah

ADDRESSES

Work: Department of Chemistry and   Home: 1150 N 360 W
       Biochemistry   Logan, Utah 84341
       0300 Old Main Hill
       Utah State University
       Logan, Utah 84322-0300
       (435) 797-8039
       (435) 797-3390 fax
       jwbakelar@gmail.com

EDUCATION

Ph.D.   Biochemistry   Utah State University, Logan, UT 84322
        (expected completion   2008-2015
         January 2015)
B.S.   Biology   Utah State University, Logan, UT 84322
        Molecular/Cell emphasis   2003-2007
        Chemistry minor

RESEARCH EXPERIENCE

2008-present   Ph.D. Research Assistant in the laboratory of Dr. Sean J. Johnson,
                 Utah State University. Thesis: Structural Characterization of
                 Stereospecificity in R- and S-hydroxypropyl-coenzyme M
                 dehydrogenases / Structural and biochemical characterization of
                 protein components of the nuclear TRAMP complex involved in RNA
                 surveillance and RNA quality control.

2006-2007   Undergraduate Research associate in the lab of Dr. Dale Barnard,
                 Institute for Antiviral Research, Utah State University. Research area:
                 High throughput screening of potentially antiviral compounds.
                 Maintained animal cell culture lines, and performed *in vivo* and *in
                 vitro* antiviral assays.

2004-2006   Undergraduate research associate in the lab of Dr. Michael Pfrender.
                Research area: Adaptive evolution of the zoo plankton *Daphnia pulex*.

Professional/Academic Awards

2014   *Thomas F. Emery Outstanding Gradute Student in Biochemistry Award*
           Utah State University. Logan, UT.
2013   *Linus Pauling Award* for best poster presentation at the American
           Crystallographic Association Annual meeting, Honolulu, HI.
2013   *Honorable Mention Award* for poster presentation at the Intermountain
           Graduate Research Symposium. Logan, UT.
2012   *Best Poster Award* for poster presentation at the Hansen Biochemistry
           Retreat. Logan, UT.
2012   *Best Poster Award* for poster presentation at the Northwest Protein
           Crystallography Workshop. Bozeman, MT.
2011   *Honorable Mention Award* for poster presentation at the West Coast
           Protein Crystallography Workshop. Monterey, CA.

Training

Molecular Biology/Biochemistry

- Molecular cloning techniques (plasmid construction, site directed mutagenesis)
- Recombinant protein expression (using *E. coli* and yeast expression systems)
- Protein purification/FPLC Chromatography (affinity, ion exchange, HIC , size exclusion)
- Common analytical techniques (i.e. SDS-PAGE, Western blot analysis)
- *In vitro* RNA transcription and purification.
- Protein Binding/kinetic techniques (fluorescence anisotropy, immunoprecipitation, pull-down assays)

Structural Biology

- Crystallization screening and crystal optimization techniques.
- Home-source and synchrotron (SSRL remote) data collection.
- Major crystallography software packages (HKL2000, Phenix, COOT).

Computation

- *Operating system*: Linux and Windows
- *Graphics*: Adobe photoshop, Adobe Illustrator, Pymol
- *Additional software:* KaleidaGraph, Multi Gauge V2.3, Microsoft Word, Excel, and PowerPoint

Leadership and Management

- Trained undergraduate and new graduate students.
- Mentored undergraduate research projects and undergraduate research proposal development.

——————————— *Research Activities* ———————————

Publications

2014   Losh JS, King AK, **<u>Bakelar J,</u>** Taylor L, Loomis J, Rosenzweig JA, Johnson SJ, van Hoof A. Interaction between the RNA-dependent ATPase and poly(A) polymerase subunits of the TRAMP complex is mediated by short peptides and important for snoRNA processing. Nucleic Acids Research. **1**(2015)

2013   **<u>J. Bakelar</u>**, D. Sliwa, S. Johnson, Crystal Structures of *S*-HPCDH reveal determinants of stereospecificity for *R*- and *S*-hydroxypropyl-coenzyme M dehydrogenases. Archives of Biochemistry and Biophysics, **533**(2013) pp. 62-68

2007   L. Latta, **<u>J. Bakelar,</u>** R. Knapp, M. Pfrender, Rapid evolution in response to introduced predators II: the contribution of adaptive plasticity. BMC Evolutionary Biology, **7**:21(2007)

Research Poster Presentations

2013   *American Crystallographic Association Annual meeting*, Honolulu, HI. "Crystal Structures of *S*-HPCDH reveal determinants of stereospecificity for *R*- and *S*-hydroxypropyl-coenzyme M dehydrogenases. **Jeremy W. Bakelar**, Dariusz A. Sliwa, and Sean J. Johnson

2013   *Intermountain Graduate Research Symposium*, Logan, UT. "Crystal Structures of *S*-HPCDH reveal determinants of stereospecificity for *R*- and *S*-hydroxypropyl-coenzyme M dehydrogenases. **Jeremy W. Bakelar**, Dariusz A. Sliwa, and Sean J. Johnson

2012   *Hansen Biochemistry Retreat*, Logan, UT. "Characterization of the zinc knuckle proteins Air1 and Air2." **Jeremy W. Bakelar,** Emily S. Frampton, Joanie M. Hevel, and Sean J. Johnson

2012   *Northwest Protein Crystallography Workshop*, Bozeman,MT. "The 1.6 Å Crystal Structure of (S)-HPCDH Reveals Determinants of Stereospecificity for *R*- and *S*-hydroxypropyl-coenzyme M dehydrogenases. **Jeremy W. Bakelar**, Dariusz A. Sliwa, and Sean J. Johnson

2012	*Intermountain Graduate Research Symposium*, Logan, UT. "Characterization of the Zinc Knuckle Proteins Air1 and Air2." **Jeremy W. Bakelar,** Joanie M. Hevel, and Sean J. Johnson
2011	*West Coast Protein Crystallography Workshop*, Monterey, CA. "The Structural Basis for Stereospecificity Displayed in (S)-hydroxypropyl-Coenzyme M Dehydrogenase." **Jeremy W. Bakelar**, Dariusz A. Sliwa, and Sean J. Johnson
2011	*Hansen Biochemistry Retreat*, Logan, UT. "Characterization of the Zinc Knuckle Proteins Air1 and Air2." **Jeremy W. Bakelar,** Joanie M. Hevel, and Sean J. Johnson
2010	*Intermountain Graduate Research Symposium*, Logan, UT. "Expression and Purification Strategies for the Poly(A) Polymerase Trf4 and the RNA Binding Protein Air2." **Jeremy W. Bakelar** and Sean J. Johnson

Oral Research Presentations

2014	*Intermountain Graduate Research Symposium*, Logan, UT. "Binding Interactions of Air1 and Air2 proteins from *S. cerevisiae*"
2012	*Hansen Biochemistry Retreat*, Logan, UT. "The Johnson Lab Probes Molecular Mechanisms Through Structural Studies".

Conferences/Professional Meetings Attended

2013	American Crystallographic Association Annual meeting, Honolulu, HI.
2012	Northwest Protein Crystallography Workshop, Bozeman,MT.
2011	West Coast Protein Crystallography Workshop, Monterey, CA.
2009	American Chemical Society National Meeting, Salt Lake City, UT.

Funding/support

2013	Doctoral Dissertation Improvement Grant, Utah State University
2013	American Crystallographic Association - Travel Award
2012	Dept. of Chemistry./Biochemistry, Utah State University - Travel Award
2011	Dept. of Chemistry./Biochemistry, Utah State University - Travel Award

Professional Memberships

2013-present   Member: American Crystallographic Association.

Distinctions

Thomas F. Emery Fellow

────Teaching Activities────

<u>Courses Taught</u>

2014 (Spring)	Advanced Biochemistry I - guest lecture
2013 (Spring)	Advanced Biochemistry I - guest lecture
2012 (Spring)	Principles of Chemistry II - Recitation lectures
2012 (Fall)	Principles of Chemistry I - Recitation lectures
2009-2012	Advanced Biochemistry Laboratory
2008(Spring)	Principles of Chemistry I - Recitation lectures
2008 (Fall)	Principles of Chemistry I Laboratory

<u>Education Enrichment Activities</u>

2013   Grant Writing Workshop, Getting Started as a Successful Proposal Writer and Academician seminar. Logan, UT.

──────── References ────────

Sean J. Johnson, Ph.D.
Associate Professor
Department of Chemistry and Biochemistry
Utah State University
0300 Old Main Hill
Logan, UT 84322
sean.johnson@usu.edu
(435) 797-2089 (phone).

Joan M. Hevel, Ph.D.
Associate Professor
Department of Chemistry and Biochemistry
Utah State University
0300 Old Main Hill
Logan, UT 84322
joan.hevel@usu.edu
(435) 797-1622 (phone) Lance Seefeldt Ph.D.

Lance C. Seefeldt, Ph.D.
Professor
Department of Chemistry and Biochemistry
Utah State University
0300 Old Main Hill
Logan, UT 84322
lance.seefeldt@usu.edu
(435) 797-3964 (phone)