IN VITRO CHARACTERIZATION OF NUDT16 ORTHOLOGS BY SEQUENCE AND FUNCTION PROVIDE EVIDENCE SUPPORTING A NUCLEAR RNA DEGRADATION PATHWAY

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

MELISSA TAYLOR-HOLOU

Dr. Brenda Peculis, Dissertation Supervisor

May 2009

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

IN VITRO CHARACTERIZATION OF NUDT16 ORTHOLOGS BY SEQUENCE AND FUNCTION PROVIDE EVIDENCE SUPPORTING A NUCLEAR RNA DEGRADATION PATHWAY

presented by Melissa Taylor-Holou,

a candidate for the degree of doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Brenda Peculis

Professor Linda Randall

Professor Dennis Lubahn

Professor Donald Burke-Aguero

Professor Chris Lorson

DEDICATIONS

God has blessed me with many people that have helped me throughout the years. Various friends and family members encouraged me and helped me realize my potential. Though I am not listing them here, they are never far from my thoughts.

One of the biggest blessings in my life was my family. My parents, Richard and Marjorie Taylor supported me in my decision to go to an expensive private school and then to graduate school. They provided both financial help and encouragement. My sister, Stacey Rucker, was my best friend while growing up. Her husband, Seth, and daughter, Kayla, have been a joy to get to know. Lastly, my "little" brother Michael Taylor who puts up with a lot of teasing from me because he is the baby of the family. I can never repay their support and love, but I can dedicate this step in my career to them.

God's newest blessing is my new husband, Roland Holou, who deserves his own page. He has had to help read this dissertation, wait for me while I work late, and even bring me food to make sure I eat when I am staying late. I dedicate this dissertation to him and thank him for his love and encouragement.

Without God's blessings, my family, and my husband, I would not have had the courage to try to get a PhD, much less to reach my goal, and impress those I with whom I studied. I dedicate this dissertation to them and thank them for their love.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Brenda Peculis, for the time and effort she has given me. Without her, I would never have finished all these experiments. She encouraged me and was an excellent teacher, mentor, and friend. I would also like to thank my doctoral committee members: Dr. Donald Burke-Aguero, Dr. Dennis Lubahn, Dr. Chris Lorson, and Dr. Linda Randall. Dr. Frank Schmidt helped with critical reading of the manuscripts and with experiment details. Colin Park, Brittny Davis, Chad Raw, and Brandon Huffman helped with the cloning of the various orthologs and with making reagents. Kristen Reynolds began teaching me the protocols and set up some of the background for the work I am doing. She also repeated some experiments for me as a control. Various clones were obtained from Dr. David Barnes, Dr. Wayne Hunter, and Dr. David Toole.

TABLE OF CONTENTS

APPENDIX

LIST OF ILLUSTRATIONS

Figure **Page**

LIST OF TABLES

Table **Page**

ABSTRACT

Nudt16, previously called X29, is known to be a metal-dependent decapping enzyme expressed in mammals and *Xenopus* . Nudt16 has been proposed to be the catalytic component of a previously unknown nuclear RNA degradation pathway that degrades nuclear-limited RNAs. However, nothing was known about the function of the protein in other species. This work proposes to examine whether the sequence and, more importantly, the RNA decapping function of the Nudt16 protein is conserved across evolution. The hypothesis tested here is that if the protein and decapping function are conserved then the proposed *in vivo* role of a nuclear pathway for RNA degradation is more likely to be true.

The experiments demonstrate that the sequence of the Nudt16 protein is evolutionarily conserved across metazoans. Additionally, there is evidence for a gene duplication that occurred at the level of tetrapods resulting in a cytoplasmic protein lacking RNA decapping function. Functional analysis of putative orthologs of Nudt16 protein from several divergent organisms demonstrated that the ability to act as an RNA decapping protein is well conserved.

More complete characterization of the Nudt16 protein from *Xenopus* as well as several of the other organisms demonstrated that the protein functions as homodimer with up to two RNA binding sites that may act independently of one another. While metal is a required co-factor, characterization of additional orthologs demonstrated that there is broad RNA substrate specificity for RNA decapping in the presence of both Mg^{+2}

and Mn^{+2}. Identification of putative *in vivo* interacting partners of Nudt16 was performed using co-immunoprecipitations. Three proteins were identified but still need to be verified. Collectively, these data demonstrate that Nudt16 is conserved both in sequence and *in vitro* function across evolution and support the existence of a nuclear RNA degradation pathway for nuclear-limited RNAs in Metazoans.

Chapter 1: Introduction and History of the Project

1.1: How did the world begin?

How did the world begin? The question intrigues even the most brilliant scientists. One possibility is that RNA evolved first and became ribozymes. As evolution continued and the conditions of the primordial world changed, RNA and then proteins with distinct functions evolved. RNA itself was not lost, but remained an essential part of many cellular processes. The study of RNA has evolved greatly increasing the awareness of cellular processes that are controlled or regulated by RNAs. Many of the cellular processes regulated by RNAs rely on altered RNA stability and translation to affect gene expression. This dissertation describes an examination of one protein (Nudt16) that may link RNA degradation and ribosome regulation.

This introduction is a brief review of RNA degradation. It is followed by a description of nuclear RNAs and some of the background and reasoning for the hypotheses presented in the rest of the chapters. A brief overview of each chapter is in the next section, describing what the experiments were and how they added to increase the understanding of this protein. The actual experiments are described in Chapters 2-5. Chapter 6 is the end of the story but is in itself a new beginning as it predicts the future of the project.

1.2: RNA degradation

The story begins with regulated RNA degradation, a conserved and very complex pathway (Lehner and Sanderson, 2004; Mitchell and Tollervey, 2000). Pathways have been described in yeast, mammals, and plants (Xu et al., 2006). There is even evidence for RNA degradation machinery encoded by the Vaccina virus (McLennan, 2007; Parrish and Moss, 2007; Parrish et al., 2007). Details between the organisms vary; however, the general mechanisms and principles are shared.

The general mechanisms and principles of RNA degradation can be divided into two distinct pathways. The difference between the two is that one starts at the 3' end and one starts at the 5' end of the RNA. Both involve large complexes of proteins; some proteins function in both pathways.

In one pathway, 3'-5' degradation involves the exosome complex which degrades RNA beginning with the polyA tail, leaving the m^7GpppG cap to be degraded by the scavenger decapping enzyme, DcpS (Chen et al., 2005; Kalek et al., 2006; Lall et al., 2005; Liu et al., 2002; Liu et al., 2004; Liu et al., 2008; Shen et al., 2008; van Dijk et al., 2003; Wang and Kiledjian, 2001; Wierzchowski et al., 2007).

The 5'-3' pathway in the cytoplasm also involves degradation by a complex of proteins. Dcp2, a member of the NUDIX (**NU**cleoside **DI**phosphate linked to some moiety **X**) family of hydrolases, acts as a nucleotide diphosphatase and cleaves the m^7GpppG-RNA to m^7Gpp and pG-RNA (Wang et al., 2002b). Although Dcp2 is the catalytically functional molecule, other proteins are necessary for the function and

probably the regulation of Dcp2 including Dcp1, Hedls, Edc1, and many others (Jacobson, 2004; Khanna and Kiledjian, 2004; Kofuji et al., 2006; Kshirsagar and Parker, 2004; Lehner and Sanderson, 2004; van Dijk et al., 2002; Xu et al., 2006). Interactions with the translation machinery have also been observed, further complicating the understanding of cytoplasmic RNA degradation (Thompson and Parker, 2007; Vilela et al., 2000).

Both degradation pathways occur in cytoplasmic bodies called "P bodies" or "processing bodies" (Cougot et al., 2004; Lall et al., 2005; Sheth and Parker, 2003, 2006; van Dijk et al., 2002). As might be expected, the identification of P bodies by immunoflourescence and co-localization studies is based on colocalization with the catalytic degradation components, such as Dcp2, Upf1, and GW182 (Lall et al., 2005; Rehwinkel et al., 2005; van Dijk et al., 2002; Wang et al., 2002b). P bodies are cytoplasmic but under conditions of antibody excess, some nuclear staining is observed with Dcp2 antibodies. Most literature reports Dcp2 is strictly a cytoplasmic protein. Thus, the proposition that Dcp2 is also present in the nucleus is controversial.

A nuclear RNA degradation pathway has been described in many organisms. In yeast, it is generally accepted that the 3'-5' degradation pathway exists in the nucleus to degrade mRNAs (Kufel et al., 2004). The exosome, which performs 3'-5' degradation, is conserved from yeast to mammals and has been found in both the cytoplasm and the nucleus. In eukaryotes, the exosome complex and other proteins in the degradation machinery in the nucleus process rRNA, snoRNA, and snRNAs. Most of these events affect uncapped or unadenylated RNA.

The hypothesis guiding the research presented in this dissertation is that Nudt16 is an evolutionarily conserved nuclear decapping protein that is involved in a nuclear 5'-3' RNA degradation pathway for nuclear-limited RNAs. This pathway plays an important role in regulation of the various small RNAs found in the nucleus.

1.3: Small RNAs localized to the nucleus

There are many small RNAs known to be localized primarily in the nucleus, including miRNAs, snoRNAs, snRNAs, piRNAs, etc. One cannot read very many journals without learning about the newest discoveries concerning small RNA molecules. Splicing, a critical pathway in making protein encoding mRNAs, requires small nuclear RNAs (snRNAs) U1, U2, U4, U5, and U6.

Another group includes microRNAs (miRNAs). Transcribed and spliced from pre-miRNAs, microRNAs tend to alter gene expression at transcriptional or post-transcriptional levels. These are frequently discussed in relation to diseases (Akao et al., 2006; Cai et al., 2004; John et al., 2004; Kurihara and Watanabe, 2004; Smalheiser, 2003; Soifer et al., 2007; Sullivan and Ganem, 2005; Talmor-Neiman et al., 2006; Wulczyn et al., 2007). In addition to roles in cancers and other diseases, micro RNAs play roles in RNAi mediated silencing and RNA degradation (Allen et al., 2005; Behm-Ansmant et al., 2006; Chan and Slack, 2006; Rehwinkel et al., 2005).

Small nucleolar RNAs (snoRNAs) are involved in regulation of ribosome biogenesis. These short RNAs (about 70 to about 200 nucleotides long) are known to be involved in many aspects of RNA modification or processing (Kiss, 2001; Lafontaine and Tollervey, 1998; Steitz and Tycowski, 1995). Some are independently transcribed polymerase II genes, particularly in yeast, but others are released from intronic regions of mRNA, although the mechanisms surrounding the intronic release are not fully understood (Maxwell and Fournier, 1995). Since 1990, the identification of snoRNAs

has blossomed requiring the generation of new databases to keep track of each snoRNA and homologs in other organisms (Eddy, 1999; Lestrade and Weber, 2006; Peculis and Mount, 1996; Zwieb, 1997).

Three types of snoRNAs have been identified based on sequence conservation, function, localization, and common binding proteins (Bachellerie et al., 2002; Kiss, 2002; Peculis, 1997; Smith and Steitz, 1997; Weinstein and Steitz, 1999). One class is the box C/D class of snoRNAs which is generally involved in 2'-O-methylation of the rRNA and will be discussed in more detail below.

The second class of snoRNAs is the Box H/ACA class which converts certain uridine residues on rRNA and other RNAs into pseudouridine (Kiss, 2001; Lafontaine et al., 1998). Most of the C/D and H/ACA snoRNAs have conserved sequence RNA regions that are complementary to and directly base pair with the target RNAs (Fatica and Tollervey, 2003; Henras et al., 2004). These and other conserved elements help the snoRNA fold into the correct structure, which involves at least one loop-stem structure. Formation of at least one loop is critical for activity of most snoRNAs; the position of the target RNA in the loop determines which residue is modified (Eddy, 2001; Kiss-Laszlo et al., 1998; Reichow et al., 2007).

The third class of snoRNAs is unique because they can have multiple copies of the C/D of H/ACA sequence elements, or one each of the Box C/D and Box H/ACA conserved elements. The scaRNAs, small Cajal body specific snoRNAs, are a subset of snoRNAs and are much more diverse in both function and sequence (Darzacq et al., 2002; Henras et al., 2004; Kiss et al., 2002). The function of these RNAs can be either

methylation or pseudo-uridylation or both depending on whether the C/D or H/ACA sequence elements are found in the sequence; all are localized to the Cajal body by a sequence called the CAB box (Darzacq et al., 2002; Henras et al., 2004; Richard et al., 2003).

1.4: How the project evolved

One of the unique Box C/D snoRNAs is U8 which is a 140 nucleotide snoRNA identified first in the rat in the mid-1980s (Reddy et al., 1985; Tyc and Steitz, 1989). It was later identified in humans (Tyc and Steitz, 1989) and *Xenopus* (Peculis and Steitz, 1993). U8 has a conserved secondary structure and is retained in small foci in the nucleolus (Reddy et al., 1985; Tyc and Steitz, 1989). U8 has a tri-methyl guanosine cap (m^{227}GpppG-RNA). As a C/D snoRNA, U8 has the Box C/D sequence elements (Jacobson and Pederson, 1998; Lange et al., 1998a; Lange et al., 1998b; Matera et al., 1994; Speckmann et al., 2000) and binds the 4 common core proteins which are conserved in Archaea and Eukarya: Nop1p/ fibrillarin, Nop58p, Snu13p/15.5kDa, and Nop56p (Aittaleb et al., 2003; Peculis, 2000; Rashid et al., 2003). In the nucleolus, U8 functions in pre-rRNA processing, cleaving the 28S and 5.8S rRNAs incorporated into the large subunit (Peculis, 2001; Peculis and Steitz, 1993, 1994).

U8 is conserved in many organisms. Although not identified in yeast, putative sequence orthologs are present in *Drosophila*, fish, reptiles, and amphibians (Gardner et al., 2009; Griffiths-Jones, 2005; Griffiths-Jones et al., 2003; Griffiths-Jones et al., 2005; Peculis et al., 2001; Peculis and Steitz, 1994). Although the primary sequence varies between the different organisms, the predicted structure is conserved and the sequence of the 5' end is important for function. In studies where U8 is injected into oocytes, the 5' end was shown to be critical for U8 function (Lange et al., 1998b; Peculis and Steitz, 1994).

U8 snoRNA was necessary but not sufficient for pre-rRNA processing *in vivo* (Peculis, 2001; Peculis et al., 2001; Peculis and Steitz, 1994); the Peculis laboratoryidentified proteins that bound U8 with high affinity and specificity to better understand how this snoRNA functions (Tomasevic and Peculis, 1999; Tomasevic and Peculis, 2002). A protein, originally called X29 was shown to be a metal-dependent NUDIX hydrolase that can cleave either m^7GpppG or m^{227}GpppG caps from a variety of RNAs that are 50 nucleotides or longer (Tomasevic and Peculis, 1999). The protein was known to be conserved in *Xenopus* and mammals. The *in vitro* function of the X29 protein was known for the *Xenopus* and human orthologs (Ghosh et al., 2004; Peculis et al., 2007).

Small RNAs are necessary for many processes and roles in the nucleus. Being involved in so many pathways, these RNAs have the ability to dramatically affect many cellular processes. No known pathway for degradation of nuclear-limited small RNAs like U8 snoRNA has been described. The identification of X29, a protein capable of binding U8 snoRNA and hydrolyzing the 5' cap, was the first indication of a nuclear RNA degradation pathway specific for nuclear-limited RNAs.

The hypothesis of the current work is that this nuclear protein, found and characterized in two vertebrates, is the catalytic, decapping protein in a previously undescribed nuclear (5'-3') degradation pathway for nuclear-limited RNAs. The goal of this document is to show that Nudt16, formerly called X29, is evolutionarily conserved and may function similarly in all organisms. The thought is that if Nudt16 is conserved both by sequence and *in vitro* function, it is likely, though not certain, that the *in vivo*

function of the protein is similar in organisms containing the Nudt16 gene and protein. Since evidence from the *Xenopus* (Ghosh et al., 2004) and human (SB and BP, unpublished data) show Nudt16 is a nuclear protein in those organisms, it is thought that a nuclear RNA degradation pathway exists in those organisms. Demonstrating that other organisms have the same gene supports the possibility of an RNA degradation pathway in those organisms. Furthermore, if the *in vitro* function of the Nudt16 protein in these other organisms is the same as the *in vitro* function of the human and *Xenopus* proteins, that provides additional support for the existence of a nuclear RNA degradation pathway.

1.5: Description of the following chapters

The hypothesis of this dissertation is that analysis of the evolutionary conservation of the sequence and function of this nuclear decapping protein will provide evidence for a conserved nuclear 5'-3' RNA degradation pathway for nuclear-limited RNAs. This hypothesis is explored over the next few chapters.

Chapter 2 describes the sequence conservation of X29, now called Nudt16, and its homologs in a variety of organisms (Taylor and Peculis, 2008). The paralog syndesmos is introduced and the phylogenetic relationships of Nudt16 to syndesmos are traced. Functional conservation of Nudt16 is briefly described in an invertebrate ortholog and examined more fully in Chapter 4. This chapter demonstrates Nudt16 is a nuclear decapping protein of ancient evolutionary origin with function conserved across Metazoans.

Chapter 3 is a more detailed analysis of the *Xenopus* Nudt16 protein. The mechanism of Nudt16 function is studied using purified *Xenopus* protein over expressed in bacteria. Observed reaction rates under a variety of conditions are found; the observed rates under conditions of protein excess and RNA excess are discussed. Inhibition of Nudt16 by the nonspecific competitor tRNA is analyzed. These data provide a model for Nudt16 function.

Chapter 4 develops the conclusions from both Chapters 2 and 3 by characterizing the function of several Nudt16 orthologs. Three additional Nudt16 proteins are characterized and compared to the *Xenpus* and human proteins. The

observed rates of cap hydrolysis demonstrate specificity of the enzymes. Using temperature and pH conditions optimized for the *Xenopus* Nudt16 ortholog, the decapping activity of each ortholog in three different metals and using multiple substrate RNAs is measured. Functional conservation of decapping activity and RNA binding by Nudt16 is confirmed. Of great significance, the model of Nudt16 function *in vivo* is refined based on data indicating the ability of Nudt16 in a variety of organisms to hydrolyze caps from diverse RNA substrates in a metal dependent reaction.

Chapter 5 further develops the proposed *in vivo* model of Nudt16 function by looking at *in vitro* protein:protein interactions. Using co-immunoprecipitations and identifying co-precipitating proteins by mass spectrometry, three proteins are identified that are putative interactors of Nudt16. The final section discusses possible future experiments designed to confirm these preliminary observations.

The last section of each chapter proposes some additional experiments, but the primary "future experiments" chapter is Chapter 6. Experiments proposed in each chapter are not duplicated rather other experiments directed towards *in vivo* studies are suggested. One of the biggest sections is discussing the plan to begin working in the zebra fish genetic model system.

Taken together, this dissertation demonstrates the Metazoan Nudt16 gene duplicated at the tetrapod lineage to produce two distinct genes. The Nudt16 protein is of ancient evolutionary origin and its function is conserved across Metazoans providing support for the existence of a nuclear RNA degradation pathway for nuclear-limited RNAs in Metazoans.

Chapter 2: Evolutionary Conservation of Nudt16 and Syndesmos

2.1: Introduction to the chapter and article

Ghosh *et al*. demonstrated that Nudt16 was conserved in humans and *Xenopus* (2004). Initial exploration for orthologs in other species identified a similar gene, annotated as syndesmos in the database (Baciu et al., 2000; Denhez et al., 2002). According to the information at the time, syndesmos was present in the human and chicken. The major goal of this part of the project was to identify sequence homologs of the Nudt16 gene in a variety of organisms and establish the phylogenetic relationships between the genes.

The following pages provide results published in *Nucleic Acids Research* in 2008 (Taylor and Peculis). After a comprehensive introduction of the methods involved in the phylogenetic analyses used in this article, a copy of the published article follows. The conclusions listed in the discussion section will then be further explained. The supplementary figures and table are included in Appendix A (pages 175-193) which also includes two phylogenetic trees with updated sequences. One is based on the amino acid sequences, the other on nucleic acid sequences. An up-to-date (February 2009) list of the 65 organisms identified as containing Nudt16 orthologs and their sequences is included in Appendix A (see Table A.2 page 196).

2.2: An introduction to basic phylogenetics

Phylogenetics has long been an important area of study. However, with the new technologies and fields of science (molecular biology, genomics, and proteomics) the classification of organisms has become in some ways easier and in other ways more complex. In the early 1960s, the cladistic approach was developed as a way to answer some of those questions using morphological or phenotypical traits shared by each organism to compare and group them (Stevens and Schofield, 2003). Using morphological or phenotypical traits did not provide enough information and became confusing because organisms that were morphologically different in some ways were also similar in other ways. Therefore, the classification of an organism was hotly debated. With the advent of DNA sequences, many more complex algorithms and programs have been generated to provide more detailed studies of each sequence, taking it apart piece by piece, with a variety of different methods, each based on a different part of the sequence information (Baldauf, 2003; Brown et al., 1998a; Chenna et al., 2003; Drummond and Rambaut, 2007; Felsenstein, 1989; Fredslund, 2006; Gertz et al., 2006; Gu and Vander Velden, 2002; Higgins et al., 1996; Jeanmougin et al., 1998; Lawrence et al., 2004; Lin et al., 2005; Page, 1996; Retief, 2000; Thompson et al., 1997; Thompson et al., 1994; Zmasek and Eddy, 2001). Many different approaches currently exist for classifying the relationships between organisms.

The information gained with each approach and the techniques involved have provided a vast amount of information about the phylogenetic relationships between

the organisms. With the advent of molecular sequence data, many genes have been identified with conserved regions that are recognized only in a specific species. The resulting DNA barcoding, or DNA taxonomy, is used to classify various organisms, even leading to the identification of new species in some cases (Balch et al., 1977; Bhupathiraju et al., 1994; Burggraf et al., 1994; Field et al., 1988; Gosink et al., 1998; Graham et al., 2000; Henry et al., 1994; Meyer and Zardoya, 2003; Ni et al., 1994; Rouviere et al., 1992; Shimkets and Woese, 1992; Weisburg et al., 1985; Winnepenninckx et al., 1998; Woese et al., 1990; Zhao et al., 1993). New international databases now exist with information on barcodes for each species (Hajibabaei et al., 2007; Marshall, 2005; Ratnasingham and Hebert, 2007). In fact, following the trends of various genes or proteins through evolution has become a very useful study. From the phylogenetic relationships demonstrated by a particular gene, scientists can determine relatedness of species, hypothesize a possible function for a novel gene or protein, and identify probable gene duplication, mutation, or lateral gene transfer events (Cannarozzi et al., 2007; Clark et al., 2007; Doerks et al., 2002; Duret and Mouchiroud, 2000; Edwards et al., 2005; Glasner et al., 2006; Larsen et al., 1999; Meyer and Zardoya, 2003; Stark et al., 2007; Wu et al., 2007; Wu and Liu, 2005).

While the field of phylognetics may initially appear to be limited to identifying and classifying organisms, some scientists have applied phylogenetics to the development and evolution of gene networks or genome expansions and duplications (Clark et al., 2007; Evans et al., 2004; Peel, 2008; Thornton, 2001). Another popular area is the search for domain swapping or conserved motifs (Gonnet and Lisacek, 2002;

Massingham, 2008; Massingham and Goldman, 2005). Preferential use of codons and codon substitution has blossomed into an important field as well (Diaz-Lazcoz et al., 1995; Duret and Mouchiroud, 2000; Epstein et al., 2000; Fedorov et al., 2002; Saier, 1995; Schneider et al., 2005). Finally, one of the largest areas of study in phylogenetics is simply improving the speed and accuracy of the current programs and tools for phylogenetic analyses (Dessimoz et al., 2006a; Gonnet et al., 2000a; Korostensky and Gonnet, 2000).

In the analyses described below, phylognetics is used to trace the evolutionary history of the Nudt16 gene. No assumptions about gene networks are made; at this point, these methods were used to identify gene orthologs from a variety of organisms.

2.3: Looking at the programs used to study phylogenetics

Looking at all the uses of phylogenetic information, the importance of this area of science is easily seen. However, what is harder to understand are the various techniques for each type of experiment. For instance, what are the differences between the programs that align the sequences? Once sequences are aligned, what is the next step? What are the most statistically sound algorithms for predicting phylogenetic relationships? All these questions are the focus of a variety of reviews, some of which were used in the summary in the following pages[1]. As the application for this research was to focus on the identification of a particular sequence from a database, determine its relatedness to other sequences, and determine a phylogenetic analysis of the gene through evolution, techniques related to these areas of phylogenetics will be discussed in detail. Other programs will be mentioned, but not described. (see Table 2.1 for summary).

[1] The information obtained here will be compiled from two main sources. For more detailed information, the reader is referred to these books. The first is Krane and Raymer's Fundamental Concepts of Bioinformatics (2003), followed by Hall's Phylogenetic Trees Made Easy (2001). Both books provide more comprehensive discussions of the various matrices and programs.

Table 2.1: A brief summary of possible phylogenetic programs. Each phylogenetic program discussed in the following pages is included in this table. Short descriptions and brief summaries of what each program can do as well as comparisons to other methods are also included. Finally, a distinction between programs used in the analyses and those briefly discussed is also provided.

Task	Method	Used?	Benefits and Drawbacks
Genome Searches	BLAST	Y	Most commonly used Many variations
	BLAT	Y	Faster than BLAST Better for searching through introns
Alignment	Clustal W or Clustal X	Y	X is the most recent version that was published Can change parameters to use different scoring matrices
Scoring matrices	Dayhoff	N	Basis for most other scoring matrices
	PAM	N	Compares substitutions among characters based on mutability and substitution rate Forms a guide tree to initiate alignment
	Blosum	N	More statistically based Does not form guide tree Does not consider the sequences to be related
	Gonnet	Y	Default for Clustal X Forms guide tree Matrix compares every character in sequence to every other character in every other sequence
Making Trees	Neighbor-Joining (NJ)	Y	Computationally simple Distance based matrix Pairs two neighboring branches
	Least squares	N	Computationally simple Distance based matrix Bases trees on squares of sequence instead of sequence itself
	Minimum Evolution	N	Computationally simple Distance based matrix Starts with NJ tree and optimizes it based on length of branches
	Parsimony	Y	Character state matrix Looks for fewest number of changes/evolutionary events Complexity increases as number of sequences increase

	Branch and Bound	N	Character state matrix Used to simplify parsimony analyses Finds one “upper bound” tree and disregards any tree that is less parsimonious
Making Trees	Branch swapping	N	Character state matrix Used to simplify parsimony analyses Begins with a complete tree but swaps branches around to make the most parsimonious tree
	Maximum Likelihood	Y	Character state matrix Statistical method Uses tree with highest probability of occurrence Used with many models
	Jukes-Cantor	Y	Character state matrix Model for maximum likelihood analyses Uses substitution rate as the only parameter
	Kimura	Y	Character state matrix Model for maximum likelihood analyses Transversion and transition rates of nucleotides are the parameters
	Time – reversible	Y	Character state matrix Model for maximum likelihood analyses Uses 6 rates to calculate the time of change
	Bayesian	Y	Character state matrix Computationally very difficult for more than a few sequences Similar to maximum likelihood but is able to move the branches around better so that it does not get “stuck” on one tree
Tree Viewing Options	TreeView	Y	Branches on trees can be moved Trees can be rooted and adjusted to display in the preferred way
	PhyFi	Y	Trees are in color Not as many options to adjust tree

The first step of any phylogenetic analysis is to obtain the sequences of interest. In this case, BLAST (Basic Local Alignment Search Tool) searches were used to identify potential orthologs. The BLAST algorithm searches sequence databases in order to match the query sequence with the fewest number of gaps possible to a sequence in

the database (Krane and Raymer, 2003). Briefly, the query sequence is broken down into smaller fragments (about 4 characters each), scanned against the database, then matched (Hall, 2001). Once matched to a particular sequence, the algorithm scans the rest of the sequence to find out how many other characters match the query sequence. The result is a list of sequences matching the query sequence, each having two scores: the bit score and expect value also called the E-value (Campbell and Heyer, 2003)[2].

The bit score measures the number of mismatches between the query and the aligned sequence. A number is given to a particular character if the two characters match between the query and the aligned sequence. If a mismatch occurs, a zero is given. Summing these numbers yields the bit score; higher bit scores are preferable. The E-value measures how likely that sequence is expected to match the query if matched by chance. Lower E-values show that the similarity is real, unlikely to have occurred by chance. Results with E-values of $1x10^{-5}$ or higher are discarded when looking for sequences of interest.

The next step is to use an alignment program to find out how well the candidate sequences align together. For instance, alignments of candidate Nudt16 genes with the *Xenopus*, human, rat, and mouse genes were used to analyze the sequence conservation of the Nudt16 gene. Three regions of similarity were identified; these were used to narrow future BLAST results and identify novel candidates. As more possible Nudt16 sequences were found, they were added to those in the initial alignment.

[2] Although a brief overview of the bit score and the expect value follows, if the actual equations and mathematics behind the two values are desired, refer to Campbell and Heyer (2003).

There are many different alignment matrices and programs. One of the most popular is CLUSTAL, whether one looks at CLUSTAL W or CLUSTAL X (Chenna et al., 2003; Higgins et al., 1996; Jeanmougin et al., 1998; Thompson et al., 1997; Thompson et al., 1994). The CLUSTAL algorithm, first developed in 1988, aligns each sequence in the alignment to all other sequences in the alignment, producing a guide tree which is then used to help with the multiple alignments. The guide tree is made from a matrix of distances. The algorithm chooses one sequence and compares the distance between that sequence and the next sequence. Then it compares that value to the value obtained with every other sequence. If the distance of the first two sequences is lower, those sequences are considered to be the most closely related sequences and are not used in further searches. In this way, the algorithm builds a guide tree which then the multiple alignment algorithm uses to find the two sequences most closely related. The multiple alignment algorithm aligns them, adds the next similar sequence (again based on its guide tree), aligns it, and then continues until all sequences have been aligned. CLUSTAL W and CLUSTAL X are very similar; CLUSTAL X is the newest version of the algorithm and is thought to be better at global alignments.

Alignment programs operate using scoring matrices to determine what each non-gap position is worth. Multiple scoring matrices exist; many are based on the Dayhoff matrices pioneered in the early 1970s (Dessimoz et al., 2006a; Dessimoz et al., 2006b; Gonnet et al., 1992; Gonnet et al., 2000b; Hall, 2001; Henikoff and Henikoff, 1992; Krane and Raymer, 2003). Each weights gaps and branches differently so many people use multiple scoring matrices for each set of sequences. The default matrix for

CLUSTAL X is the Gonnet matrix so it was used in the analyses below. The Gonnet matrix bases its algorithm on a sequence relationship tree built by an initial guide tree. After the initial scoring, the sequences are aligned based on a second alignment program: the Needleman-Wunsch alignment (Gonnet et al., 1992).

This alignment creates a matrix of every character with every other character in the sequence. Matching characters are given a number value; mismatches are given no value and gaps are penalized. Once the matrix of numbers is generated, the relationships between each character are summed. Finally, the sequences with the highest score are considered the most closely aligned. The process is computationally difficult which is why the Gonnet matrix builds a guide tree before attempting to create all the necessary matrices.

Within each of these matrices are more complex mathematics and algorithms which will not be discussed. Each program, or algorithm, has its own advantages and disadvantages; thus, many scientists use multiple alignment matrices and programs to align sequences. Once obtaining the initial alignments, scientists may manually edit the sequences to achieve a better alignment. However, it is also important not to bias the sequences towards a particular perceived relationship so very few edits on the alignments of the Nudt16 sequences were made. Specifically, the conserved regions were recognized and aligned together as needed but nothing else was changed.

With a suitable alignment in hand, the next step is to determine a phylogenetic relationship usually shown as a tree. This is where the options increase and have changed drastically over the 4 years that this research has taken. Originally, few options

existed: parsimony, likelihood, and distance based methods. Since then, Bayesian and other options have appeared which use more statistical approaches (see Table 2.1 for a brief summary of most of them).[3]

It is evident that choosing the correct program for a particular set of sequences can be difficult. Some phylogeneticists forgo the exactness of the Bayesian analysis for the speed of a distance-based matrix; others prefer the more complete analysis. Most phylogeneticists use multiple analyses, combining various aspects of each algorithm to generate phylogenetic relationships based upon their sequences. Then they show only one figure from one method.

The sequences here have been analyzed with nearly every model and algorithm that was available to the public at no cost. The reported trees are based on the distance matrix and the Fitch-Margoliash method, both included in a popular program called PHYLIP (Felsenstein, 1989). These trees were generated more quickly and still contained the same information as the more complex trees. In Appendix A, the amino acid tree of the Nudt16 protein (Figure A.4, page 191) was created with the maximum likelihood algorithm. The nucleic acid tree (Figure A.5, page 192) was made with the maximum parsimony algorithm. Some variations between the trees can be associated with changes in the method. Sequence information for the additional orthologs is included in Table A.2 (page 196).

[3] As might be expected a number of reviews of the phylogenetic methods have been generated. One of the most useful was *Phylogenetic analysis on molecular evolutionary genetics* by Nei (1996) and *Phylogeny estimation: traditional and Bayesian approaches* by Holder and Lewis (2003).

2.4: *Evolutionary conservation supports ancient origin for Nudt16, a nuclear-localized, RNA-binding, RNA-decapping enzyme*. Taylor MJ and Peculis BA (2008). Nucleic Acids Res. 36:6021-6034.

This paper is included in its entirety here, with the figures and images as in the journal. The references and figures have been renumbered to integrate into the dissertation as a whole; supplementary figures are included in Appendix A (pages 175-193). In Section 2.5, a conclusion and summary section discussing the most recent data of the searches and phylogenetic analyses done since the publication of this paper is included.

Abstract

Nudt16p is a nuclear RNA decapping protein initially identified in *Xenopus* (X29) and known to exist in mammals. Here we identified putative orthologs in 57 different organisms ranging from humans to Cnidaria (anemone/coral). *In vitro* analysis demonstrated the insect ortholog can bind to RNA and hydrolyze the m^7G cap from the 5' end of RNAs indicating the Nudt16 gene product is functionally conserved across metazoans. This study also identified a closely related paralogous protein, known as Syndesmos, which resulted from a gene duplication that occurred in the tetrapod lineage near the amniote divergence. While vertebrate Nudt16p is a nuclear RNA decapping protein, Syndesmos is associated with the cytoplasmic membrane in tetrapods. Syndesmos is inactive for RNA decapping but retains RNA binding activity. This structure/function analysis demonstrates evolutionary conservation of the ancient Nudt16 protein suggesting the existence and maintenance of a nuclear RNA degradation pathway for RNAs in metazoans.

Introduction

Modulation of RNA stability *in vivo* is a major regulatory level of control of gene expression in eukaryotes. In the cytoplasm, two major pathways for RNA turnover are known to exist, depending on whether RNA degradation occurs from the 5' end or from the 3' end (Coller and Parker, 2004; Decker and Parker, 1994; Gu et al., 2004; Meyer et al., 2004; Parker and Song, 2004). Cytoplasmic degradation of RNA from the 5' end requires the removal of the 5' m^7GpppG cap that confers stability to mRNAs in the cell. In the cytoplasm this cap-removal activity is performed by the decapping enzyme, Dcp2. Dcp2 is a NUDIX protein with RNA-dependent cap hydrolysis activity responsible for initiating RNA turnover. The Dcp2 protein is conserved among eukaryotes from yeast to human and functions in coordination with a variety of proteins (Bail and Kiledjian, 2006; Dunckley and Parker, 1999; Dunckley et al., 2001; Piccirillo et al., 2003; van Dijk et al., 2002; Wang et al., 2002b). Dcp2 is a member of the NUDIX family; these proteins catalyze the hydrolysis of **nu**cleoside **di**phosphates that are linked to some moiety **X** (Bessman et al., 1996; Galperin et al., 2006; McLennan, 1999; Mildvan et al., 2005). In the cytoplasm, Dcp2 cleaves the cap releasing m^7GDP and leaving a 5' monophosphate on the 5' end of the RNA. This decapped RNA is a substrate for the 5'-3' exonuclease Xrn1. While Dcp2 has the catalytic hydrolysis activity for cap removal, little is known about the specificity directing the binding activity of this protein.

RNA turnover also occurs in the nucleus (Das et al., 2003; Das et al., 2006; Kuai et al., 2005). There are nuclear exonucleases with both 5'-3' and 3'-5' degradatory activity

such as the exosome and Xrn2/Rat1p (Cougot et al., 2004; Johnson, 1997; Kastenmayer and Green, 2000; Lejeune et al., 2003; West et al., 2004). Since 5' caps are added to RNAs very soon after transcription is initiated (Aguilera, 2005; Lindstrom et al., 2003; Mandal et al., 2004; Moteki and Price, 2002; Speckmann et al., 2000), there must be nuclear decapping machinery that parallels the cytoplasmic machinery. Nudt16, first characterized as X29, is a nuclear-localized NUDIX protein with metal-dependent RNA decapping activity (Ghosh et al., 2004; Peculis et al., 2007; Tomasevic and Peculis, 1999). It has high affinity and specificity for binding U8 snoRNA based electrophoretic mobility shift assays (EMSAs) and can be crosslinked to U8 snoRNA (Ghosh et al., 2004; Peculis et al., 2007; Tomasevic and Peculis, 1999).

The X29/Nudt16 protein is distinct from Dcp2; other than sharing the catalytic NUDIX domain, the two proteins show very little sequence similarity. While both proteins display RNA decapping activity *in vitro*, the decapping repertoire of Nudt16 is broader than that of Dcp2. Dcp2 can use Mg^{+2} or Mn^{+2} to decap m^7GpppG capped RNAs, but X29/Nudt16 protein can utilize Mg^{+2}, Mn^{+2} or Co^{+2} to remove both the m^7G cap or the m^{227}G hypermethylated cap, present on many of the small nuclear-limited RNAs comprising nuclear RNPs (Ghosh et al., 2004; Peculis et al., 2007).

The family of NUDIX proteins eliminates potentially toxic nucleotide derivatives from cells. These proteins have been described in bacteria (MutT was a primary member of this family) and over 20 different family members with different cellular functions exist in humans (McLennan, 1999; Mildvan et al., 2005). Nudt16p was identified in vertebrates but the evolutionary distribution of the protein was not known.

Phylogenetic relationships and protein conservation data have been useful in providing functional and structural information which have assisted in assigning biological roles *in vivo*. Since the information within various available databases has expanded tremendously in the past few years we undertook a search for Nudt16 orthologs using publicly available databases, including those for cDNAs, ESTs, genes and genomes. We identified putative Nudt16 orthologs in 57 different organisms, including vertebrates and invertebrates. Nudt16 ortholog sequences were aligned and characterized by phylogenetic trees, which showed the relationship between organisms with respect to this gene. The phylogenetic trees presented here are consistent with Nudt16 being well conserved in the metazoan lineage (Meyer and Zardoya, 2003; Stuart et al., 2002; Wu et al., 2007). Biochemical analysis indicated there was functional conservation of the Nudt16 protein as well; the insect ortholog was active for RNA decapping *in vitro*. The searches also revealed evidence for a gene duplication followed that resulted in closely related genes on different chromosomes, followed by functional divergence. The paralog of Nudt16p, previously described as Syndesmos (Baciu et al., 2000; Denhez et al., 2002) is a cytoplasmic protein incapable of nucleoside hydrolysis but which can still bind RNA. All metazoans examined here have Nudt16 protein but only a subset have Syndesmos indicating that Nudt16p is the more ancient and functionally conserved protein that plays an essential biological role.

Material and Methods

Identifying orthologs

Putative Nudt16 orthologs were identified using the BLAST algorithm against a variety of databases including GenBank[4] (http://www.ncbi.nlm.nih.gov/), Ensembl (http://www.ensembl.org/index.html), and the Joint Genome Institute (http://www.jgi.doe.gov/). Initially genomes were analyzed with the sequence for X29 (NCBI: NP_001084713.1). As more Nudt16 sequences were identified in other organisms, those sequences were also used to screen each database and identify more diverged Nudt16 orthologs. Detailed sequence information and accession numbers are included in Supplementary Table A.1 (page 190-193).

Phylogenetic Analysis:

Sequences were aligned using the ClustalX1.83 alignment program. Alignments were analyzed with the PHYLIP 3.67 package. Initially, Seqboot was used to generate bootstrap the alignment generating 100 random data sets. Those data sets were then analyzed by applicable programs. Although all programs available in the package were used to analyze the phylogenetic relationships, reported data were generated using distance algorithms, ProtDist and DNADist. Distance matrices were then analyzed by Fitch (in the PHYLIP package), which uses the Fitch-Margoliash algorithm without a molecular clock to generate possible trees from the datasets. PHYLIP program Consense was used to generate a consensus tree using the anemone sequence as the outgroup for each tree. The trees were visualized by TreeView or PhyFi. Branch lengths are included

[4] Although many versions were used, the most recent version of GenBank included in this article was GenBank Version 167.0

as a ruler at the bottom of the figures. Bootstrap values are placed at each node where the bootstrap was less than 80 out of 100 trees.

Cloning of cDNAs – Sharpshooter Nudt16

Template encoding full length cDNA for sharpshooter (NCBI # : DN199400.1) was a generous gift from Wayne Hunter's laboratoryat the USDA in Florida. Deoxyoligonucleotide primers were generated to PCR and clone the cDNA into pET19b in frame with the vector-encoded His-tag. Primer sequences were : Sense : 5' - CGCGCATATGTCCTCTGATACAGG - 3' (NdeI) and Anti-sense : 5' - CGCGCTCGAGCTATTATCAGGACTTGAGTGGAAG - 3' (Xho1). The restriction site in parenthesis was designed into the primer and assisted with directed, in-frame cloning. The protein was overexpressed in E.coli and purified on a sequential Ni-NTA resin (Qiagen) followed by Heparin Sepharose (Amersham/Pharmacia)

Cloning of cDNAs - human Syndesmos

Template encoding full length cDNA for human Syndesmos (NCBI # : BE799462) was purchased from Open Biosystems (MHS1011-61470). The Clone ID (or IMAGE ID) was 3944355. Deoxyoligonucleotide primers were generated to PCR and clone the cDNA into pET19b in frame with the vector-encoded His-tag. Primer sequences were : Sense : 5' - CGCGCATATGTCGACGGCGGCGG - 3' (Nde1) and Anti-sense : 5' - CGCGGGATCCCTATTATCAAGAGGAGGCCGGGAGCAAC – 3' (BamH1)

The restriction site in parenthesis was designed into the primer and assisted with directed, in-frame cloning. The protein was over expressed in E.coli and purified on a

sequential Ni-NTA resin (Qiagen) followed by Heparin Sepharose (Amersham/Pharmacia)

RNA-Protein Cross-linking:

Uniformly ^{32}P-labeled, 4-thioU containing RNAs were transcribed as previously described (30). Crosslinking reactions were performed essentially as described (30). To summarize with modifications included, 0.08 pMol of the 4-thioU RNA was combined with cold competitor tRNA (250 molar fold excess), in 45 mM HEPES, pH 8.5. Protein was added in the amounts indicated to yielded crosslinks : Human: 14.2 pMol; X29: 1.64 pMol; Sharpshooter: 11.8 pMol; Syndesmos: 12.9 pMol. The reaction was incubated at room temperature for 15 min, then exposed to 365-nm UV light source at a distance of 3 cm for 15 min on ice. RNase A was then added to each reaction and incubated for 30 min at 37°C. The proteins in the reactions were resolved on a 10% Bis-Tris NU-PAGE gel. Dried gels were exposed to Kodak phosphorimager plates to visualize proteins via ^{32}P label-transfer from the RNA.

Protein-Protein Cross-linking:

Protein-protein cross-linking was performed using the water-soluble, reversible cross-linker 3,3'-dithiobis(sulfosuccinimidylpropionate) (DTSSP) (Pierce) at final concentration of 0.5 mM with 40 pMol of proteins for 30 min. Nu-PAGE 4x loading dye was added to each sample, with or without BME as indicated. Reactions were resolved on NuPAGE gels and detected via coomassie blue staining.

Molecular Modeling

The molecular modeling was performed using the DeepView/Swiss-PdbViewer program. The sequences were aligned to the Xenopus Nudt16p amino acid sequence then the modeling program generated corresponding folded molecules based upon the X29 crystal structure (PDB: 2A8P). Single conserved amino acids and a beta sheet were highlighted in the ribbons diagrams to facilitate correct orientation of the structures before electrostatic potential was calculated.

Results

Identification and functional analysis of Nudt16 in vertebrates

We had previously identified the *Xenopus laevis* and human homologues of the Nudt16 protein (Ghosh et al., 2004). Alignments of these protein sequences demonstrated that X29/Nudt16 and the putative mammalian orthologues of Nudt16 contain three well conserved regions, outlined by boxes in Figure 2.1. (Supplementary Table A.1, page 190-193 provides details about sequences and accession numbers). These three regions were designated the "hallmarks" of authentic Nudt16 sequences because they were well conserved among the *Xenopus*, human, and rat Nudt16 proteins, all of which are functional orthologs ((Ghosh et al., 2004; Peculis et al., 2007); rat: MJT and BAP, unpublished data).

The three conserved regions were mapped onto the crystal structure of the X29 protein (PDB: 2A8P) (Scarsdale et al., 2006) to reveal a probable functional basis for each. The sequence outlined in the green box of Figure 2.1 forms a β-strand passing deep within the interior core of the protein, presumably required for organization and

proper folding of the protein. The sequence in the red box is the NUDIX consensus sequence. This functional domain is critical for hydrolysis activity and defines members of the NUDIX family. The third conserved region is the sequence outlined in the blue box in Figure 2.1. This is the putative protein dimerization domain. In the crystal structure of X29, this domain forms a significant part of the dimerization interface between the monomers in the crystallographic unit of X29 (Scarsdale et al., 2006). All putative Nudt16 orthologs were required to have all three conserved regions and specifically to contain the catalytic residues within the NUDIX sequence.

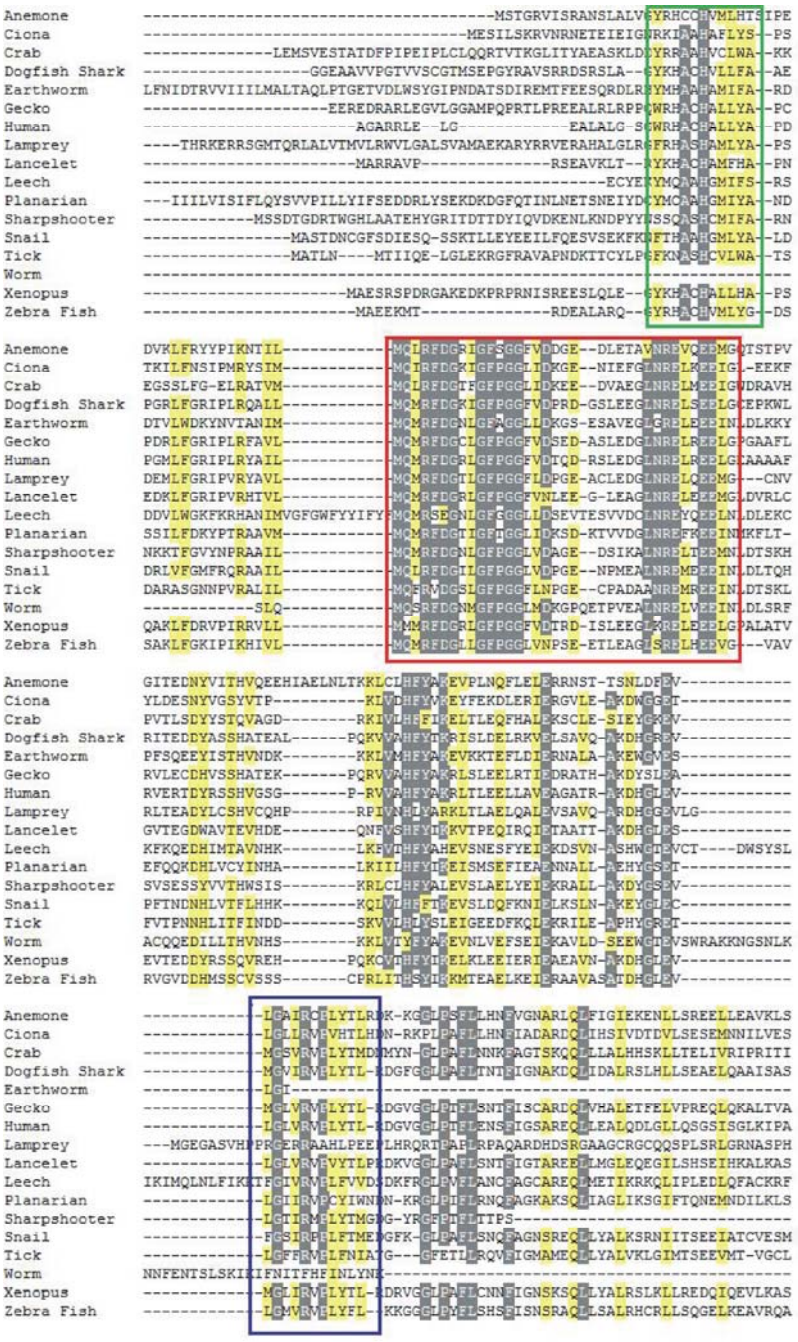


Figure 2.1: Amino acid alignment of putative Nudt16 orthologs from a sampling of the organisms found to have this protein. The alignment reveals three areas of highly conserved sequence, indicated by the green, red and blue boxes. A functional role based on structure for each conserved region is described in more detail in the text. To be designated as a Nudt16 homologue, sequences had to have the NUDIX domain (in red box) and the other two domains. Gray residues indicate identity; yellow highlighted residues show similarity. The accession number for each sequence used is listed in Supplementary Table A.1 (pages 190-193).

Identification of putative Nudt16 orthologs in other organisms

Previous BLAST results revealed Nudt16 existed in *X. laevis* and a variety of mammals, but orthologs were not readily identified in the "model non-vertebrate" organisms including *C. elegans*, yeast, or *D. melanogaster*. Therefore, Nudt16 was initially proposed to be a vertebrate protein (Ghosh et al., 2004). We expanded the search for Nudt16p orthologs using publicly available data bases; we have identified putative Nudt16 orthologs in 57 different organisms, both vertebrate and invertebrate (see below). Approximately half of the represented classes are invertebrates. Curiously, exhaustive searches of the *C. elegans*, yeast, and *D. melanogaster* genomes have not yet revealed orthologs of Nudt16 on the basis of amino acid sequence alone, but convincing sequences have been identified in other worms and insects. Since these organisms do contain several NUDIX proteins but that the Nudt16 proteins are a distinct and identifiable subset, we acknowledge that convincing functional orthologs may not be identified through sequence similarity alone.

Phylogenetic relationships of Nudt16 protein orthologs

To appreciate the relatedness of these proteins phylogenetic trees were generated based on sequence alignments; algorithms were used to generate trees. It was necessary to first identify a standard tree to which all subsequent trees could be compared since different trees can be generated using different algorithms. We created a "generic" tree demonstrating phylogenetic relationships of organisms using information from the Tree of Life Web Project (http://www.tolweb.org/tree/). Taxonomic information for each organism was obtained from TaxBrowser on the NCBI

website (www.ncbi.nlm.nih.gov) and can be found in Supplementary Table A.1 (pages 190-193). This taxonomic tree was the standard for comparison with all trees generated from sequence alignments; it is provided here for reference (Figure 2.4).

Potential Nudt16 orthologs were analyzed using several standard phylogenetic programs and methods (An et al., 2007; Baldauf, 2003; Hall, 2001; Palma et al., 2007). We aligned the sequences of Nudt16 orthologs with ClustalX1.83 and generated trees with the PHYLIP program package (see Methods). A phylogenetic tree (Supplementary Figure A.1, page 176) was generated from aligned protein sequences using the anemone (a cnidarian) Nudt16 ortholog as the out-group. Cnidarians were an appropriate choice for the out-group because they are thought to be the most ancient of the organisms included in this analysis (Abouheif et al., 1998). Sequence alignments were used to generate trees based upon Nudt16p sequence distance, parsimony, or maximum likelihood algorithms. Most results matched the predicted phylogenetic relationships. Species within the same class or phylum were clustered on the Nudt16 tree consistent with phylogenetic relationships predicted by other molecular analyses (Meyer and Zardoya, 2003; Stuart et al., 2002).

Nud16p orthologs are functional for decapping.

Alignments of the Nudt16 orthologs demonstrated this was a conserved metazoan protein and that all orthologs contained a functional NUDIX domain. Phylogenetic conservation of the nuclear decapping protein implied the existence of a conserved nuclear pathway for RNA turnover; thus it was important to confirm that the sequences identified were functional orthologs. Since we had previously characterized

functional (decapping) activity in several vertebrates it was desirable to examine a non-vertebrate protein and chose an insect orthologs for this purpose. The glassy-winged sharpshooter (*Homalodisca coagulata*) was selected as a representative insect because the sequence of a full length cDNA could be obtained. This insect is of economic interest because it is responsible for spreading the bacteria causing Pierce's disease in grapevines and citrus trees (Bi et al., 2007).

We amplified the region encoding the predicted *H. coagulate* (sharpshooter) Nudt16p ortholog via PCR, and cloned the product into an expression vector (see Methods). The His-tagged fusion protein was over-expressed in bacteria, purified and used for biochemical analysis. Decapping reactions combined ^{32}P-cap-labeled U8 snoRNA, metal and purified protein, combined as in our standard assay (Ghosh et al., 2004; Peculis et al., 2007), product formation was examined via thin layer chromatography (Figure 2.2). The results indicated the sharpshooter Nudt16p ortholog is active for cap hydrolysis, releasing m^7GDP from the 5' end of U8 cap-labeled RNA in a metal-dependent manner. Two forms of the human Nudt16 protein and the Xenopus protein are present as positive controls. The product of the decapping reaction was confirmed to be m^7GDP by incubation of the reaction product with nucleotidediphosphate kinase (Ghosh et al., 2004); the kinase activity altered the mobility of this material which then co-migrated with m^7GTP (data not shown). Similar to the other Nudt16 orthologs previously characterized by this laboratory, decapping activity by the insect protein was higher in the presence of Mn^{+2} than in Mg^{+2} or Co^{+2}, but all three metals were capable of mediating cap hydrolysis from U8 snoRNA (data not

shown). Protein concentrations were optimized for the decapping assays; the amount of protein required for 50% decapping within 10 minutes was comparable to that in the better characterized vertebrate orthologs. Thus, the insect Nudt16p protein was catalytically active for decapping and a true functional ortholog of the vertebrate nuclear decapping protein (Figure 2.2).

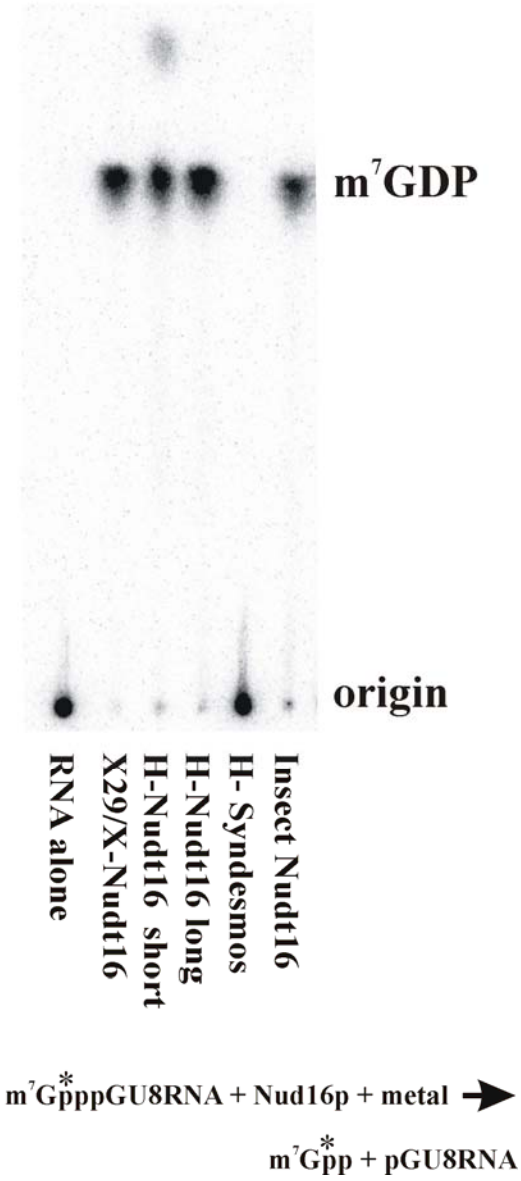


Figure 2.2: Insect Nudt16 protein is active for decapping RNA. Cap labeled U8 RNA was incubated in the presence of buffer, Mn^{+2} and protein (as indicated) for 20 minutes at 37^{o}C (reaction indicated below) Samples of the reactions were spotted on a TLC, which was developed then visualized on a phosphorimager. The Xenopus (X29/X-Nudt16) protein, present as a positive control, released m7GDP from the cap-labeled RNA. The two human proteins, which varied by an unrelated amino terminal extension, were both efficient at decapping the U8 RNA. Insect (sharpshooter) Nudt16 protein hydrolyzed the RNA to release the m^{7}GPD cap comigrating with that cleaved by the other orthologs. The human H-Syndesmos protein displays no decapping activity.

Identification of closely related non-NUDIX proteins in mammals and birds

In addition to identifying functional orthologs of Nudt16p in metazoans, the BLAST searches also identified a closely related family of vertebrate proteins called "Syndesmos" (see Table 2.2, Figure 2.3 and Supplementary Materials, Table A.1, pages 190-193 and Supplementary Figure A.3, pages 178-189). Syndesmos was first reported in chicken and further characterized in human. Syndesmos is a cytoplasmic protein that interacts with Syndecan-4 and the Paxillin family of proteins, including Hic-5 (Baciu et al., 2000; Denhez et al., 2002). When over-expressed in human tissue culture cells Syndesmos localizes to the ventral plasma membrane and enhances cell spreading and focal contact formation (Baciu et al., 2000). There are no *in vivo* functional data describing a biological role for Syndesmos.

Nucleotide (Supplementary Figure A.3, pages 178-189) and amino acid (Figure 2.3) alignments of the Nudt16 proteins and the Syndesmos proteins demonstrated they were closely related but distinct protein families. Within mammals, overall the two proteins were 50-60% identical and 60-75% similar at the amino acid level (Table 2.1). The highest identity occurred within the three regions that were conserved among the Nudt16p proteins (see Figure 2.1 and compare with the partial sequence shown in Figure 2.3). Syndesmos maintains the domain that constitutes the center of the Nudt16 protein (green box, Figure 2.1, not shown in Figure 2.3) and also contains the 'interface interaction" domain (blue boxes, Figures 2.1 and 2.3) found in Nudt16. Interestingly, in Nudt16, the interaction interface is the contact surface for the two monomers seen in the X-ray crystal structure (Scarsdale et al., 2006). In Syndesmos, this protein domain is

reported to be involved in facilitating interactions with other cytoplasmic proteins (Denhez et al., 2002).

Table 2.2: Conservation and divergence of Nudt16 and Syndesmos proteins. Columns 2 and 3 provide the percent identity and similarity (respectively) of Nudt16 compared to the Syndesmos protein within each organism indicated in column 1 (i.e. bat Nudt16 compared to bat syndesmos). The percent identity of the Nudt16 in the indicated organism when compared to human Nudt16 is shown in Column 4. The percent identity of Syndesmos in the indicated organism compared to human Syndesmos is shown in Column 5. These values were calculated by BL2SEQ (Altschul et al., 1997) or MultiView Alignment Display (Brown et al., 1998b) using the Biology Workbench application (workbench.sdsc.edu). Columns 6 and 7 indicate the chromosomal location of the genes encoding Nudt16 or Syndesmos, respectively, obtained through GenBank and Ensembl. Some organisms have not been fully annotated or chromosome information was not available.

Table 2.2: Conservation and divergence of Nudt16 and Syndesmos proteins

Organism	S:Nud % Identity	S:Nud % Similarity	humNudt vs Nudt	HumSynd vs Synd	Chromo# Nudt	Chromo# Synd
Bat	44%	54%	73%	49%	-	-
Bovine	62%	73%	82%	97%	-	25
Chimpanzee	61%	70%	99%	70%	3	16
Dog	53%	60%	78%	88%	23	6
Galago	50%	60%	86%	56%	-	-
Horse	51%	60%	40%	64%	16	-
Human	61%	75%	100%	100%	3	16
Macaque	64%	74%	49%	85%	-	-
Mouse	58%	72%	79%	95%	9	16
Opossum	54%	69%	65%	92%	X	6
Pig	57%	68%	66%	84%	-	-
Platypus	48%	58%	51%	83%	-	-
Rabbit	60%	69%	57%	40%	-	-
Rat	60%	73%	68%	97%	8	10
Rhesus monkey	64%	74%	71%	99.5%	2	16
Averages	**56%**	**67%**	**71%**	**80%**		

The Nudt16 proteins were easily distinguished from Syndesmos orthologs by critical sequence changes within the catalytic NUDIX domain (red box, Figures 2.1 and 2.3). While the NUDIX-like domain is well conserved in Syndesmos compared to Nudt16p (over 90%), the Syndesmos proteins contained a pair of glycine-leucine repeats (outlined in the orange box in Figure 2.3) in place of four amino acids that include the two glutamic acid residues required for catalysis in Nudt16 proteins (outlined in the heavy red box of Figure 2.3) (Abdelghany et al., 2003; Bessman et al., 1996; Galperin et al., 2006; Ghosh et al., 2004; McLennan, 2006). Thus the cytoplasmic Syndesmos family of proteins is clearly a paralog of the catalytic nuclear Nudt16p family of proteins, but was predicted to lack decapping activity. When used in a decapping reaction it demonstrated no cap hydrolysis activity (see Figure 2.2).

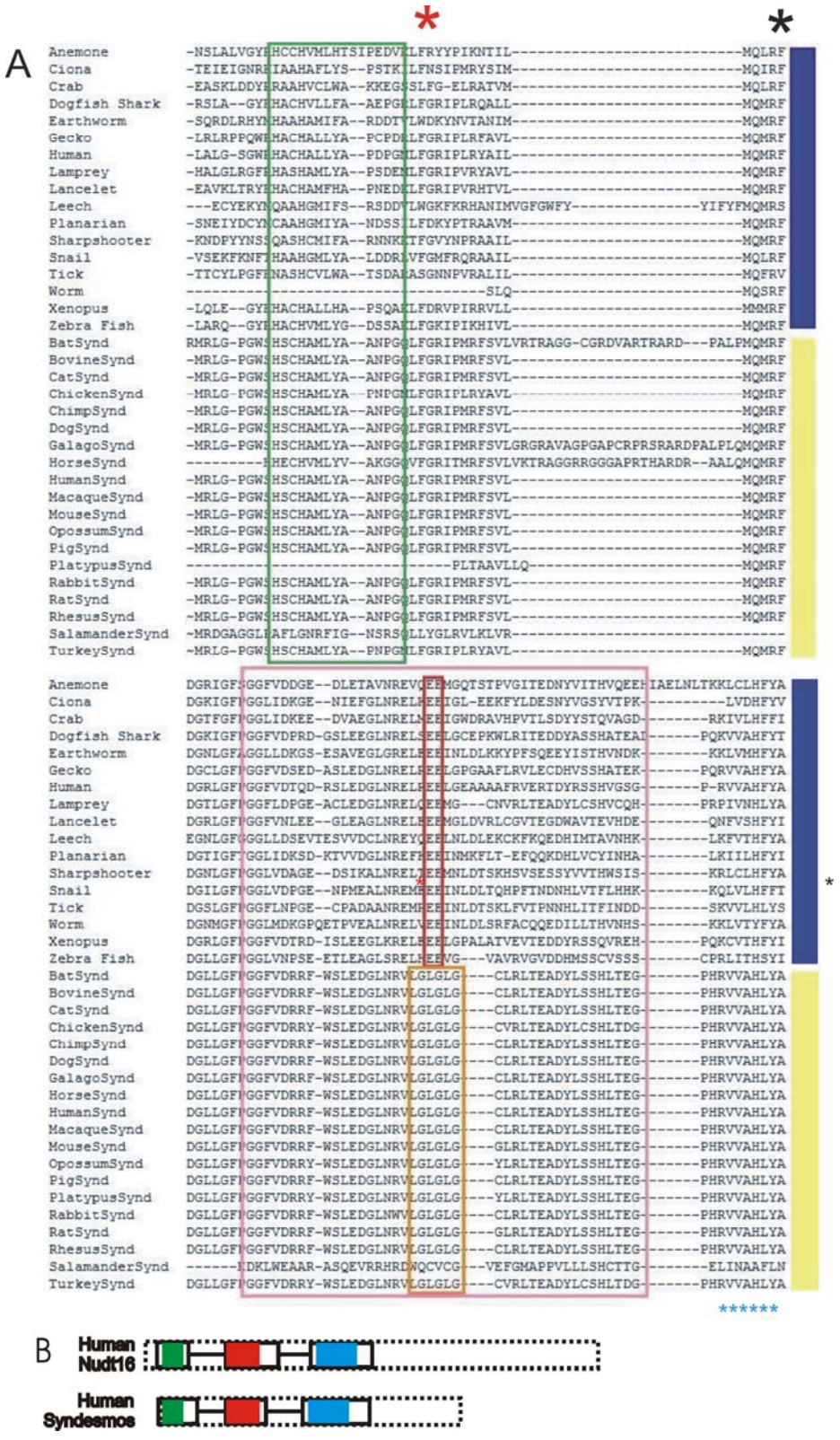

A
B
Human Nudt16
Human Syndesmos

Figure 2.3: Nudt16 and Syndesmos are closely related members of a gene family. A. Alignment of a portion of the Syndesmos protein from 17 organisms (denoted by yellow box on right edge) and the corresponding region of Nudt16 from 10 organisms (marked with the blue box on right edge). All Syndesmos orthologs have a repeated glycine/leucine sequence (orange box) in place of the glutamic acid residues (heavy red box) required for catalysis in the NUDIX domain (red box). The green box is the conserved region in Nudt16 proteins, equivalent to the green box in Figure 2.1. Syndesmos paralogs have all three conserved regions but lack a functional NUDIX domain. The central part of the proteins are aligned here. The accession numbers are in Supplementary Table A.1 (pages 190-193). B. Genomic organization of the orthologs in human. Dashed boxes indicate the transcription unit, solid boxes are exons, horizontal lines are introns and the conserved domains are color coded as per Figure 2.1. Note Nudt16 has a longer 5'UTR and 3'UTR than Syndesmos.

Phylogenetic relationship between Nudt16p and Syndesmos

To examine the phylogenetic relationship between Syndesmos and Nudt16p, we aligned the sequences of these proteins and generated phylogenetic trees from the available data. Although the nucleic acid sequences were used in most of these analyses, Figure 2.5 shows the amino acid sequences of the representative subset of those organisms listed in Supplementary Table A.1 (pages 190-193). The trees did not differ significantly when nucleotide sequences were aligned. When arranged by degree of sequence divergence, the data indicated that the Syndesmos orthologs were grouped together as a distinct branch, diverging from the amniota branch of the Nudt16 tree (Figure 2.5). Figure 2.4 is a schematic indicating the branches where Syndesmos (S) and Nudt16 (N) proteins have been found on a more simplified phylogenetic tree.

One striking observation arising from the alignment of Nudt16p and Syndesmos paralogs is that to date, Syndesmos has been reported only in birds and mammals (Figure 2.4B; Supplementary materials, pages 175-193). Our analysis identified several cases of incorrectly annotated proteins ("NUDIX-like" proteins that are Syndesmos and

vice versa). There are birds, amphibians, reptiles and mammals for which only one paralog has been annotated. Since Nudt16p is found across metazoans, we presume that incomplete coverage/sequencing is the explanation where only Syndesmos had been identified. The salamander sequencing project (www.abystoma.org) yielded a partial salamander Syndesmos sequence. Similarly, sequencing of PCR products amplified from chicken genomic DNA demonstrated that chickens have the Nudt16 gene (M.J.T. and B.A.P. data not shown). Clearly, another possible explanation for organisms where Nudt16p is present but the Syndesmos paralog is absent is that the organism (or its evolutionary predecessor) never underwent the gene duplication event or stochastic genetic loss has occurred.

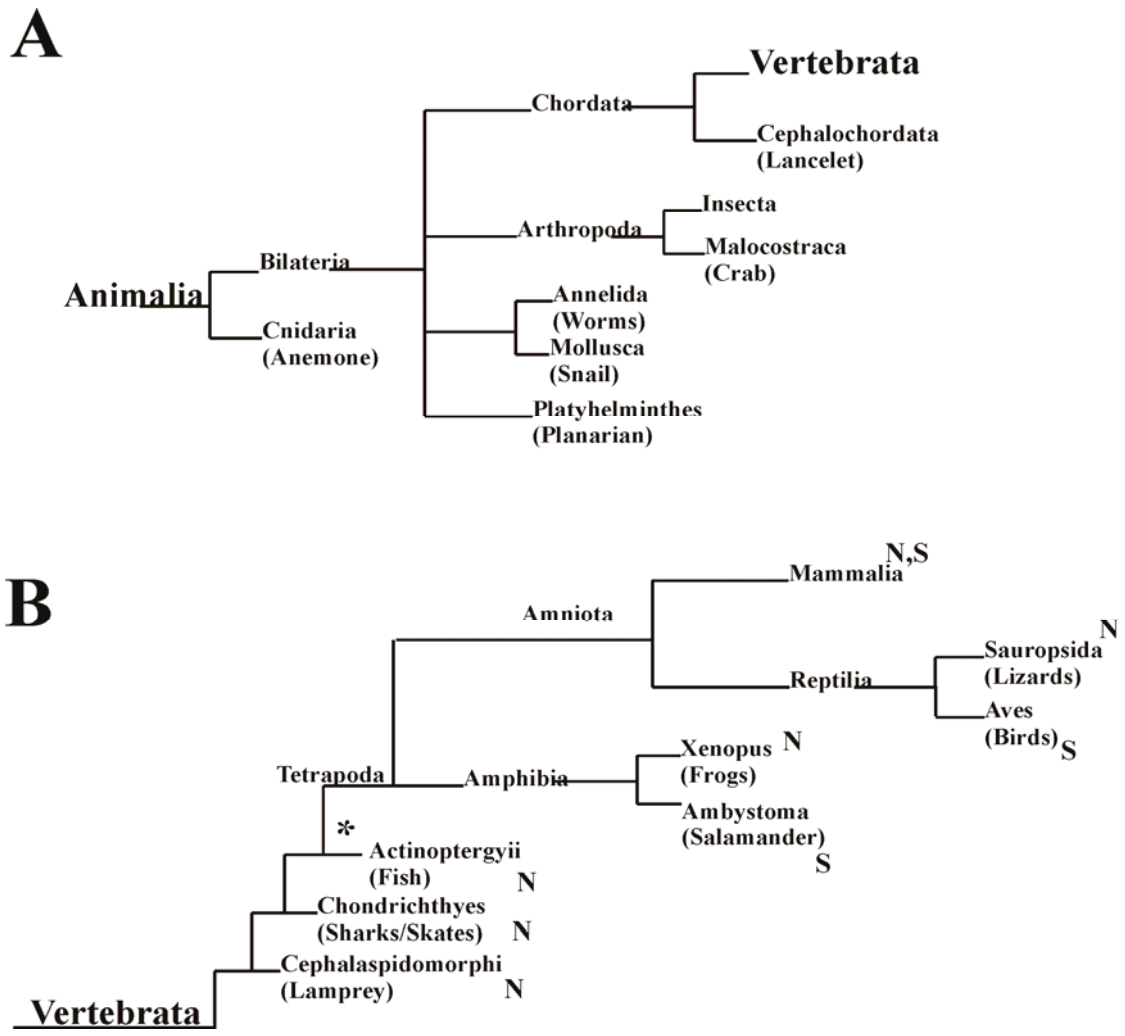


Figure 2.4: A "standard" phylogenetic tree. This tree illustrates the relative position of the organisms examined here and is provided as a quick reference for taxonomic relationships, based on the Tree of Life Web Project. A) The Kingdom Animalia/Metazoa was searched. Only those branches for organisms containing a Nudt16 ortholog are shown. B) An expansion of the vertebrate branch reveals the conservation of Nudt16 and appearance of Syndesmos. The "N" indicates branches with organisms containing Nudt16 orthologs, "S" indicates branches containing Syndesmos, and "*" indicates the likely gene duplication event; acknowledging the possibility that it may have occurred earlier with high rates of subsequent gene loss. Sequences 'above' the "*" have only the Nudt16 protein.

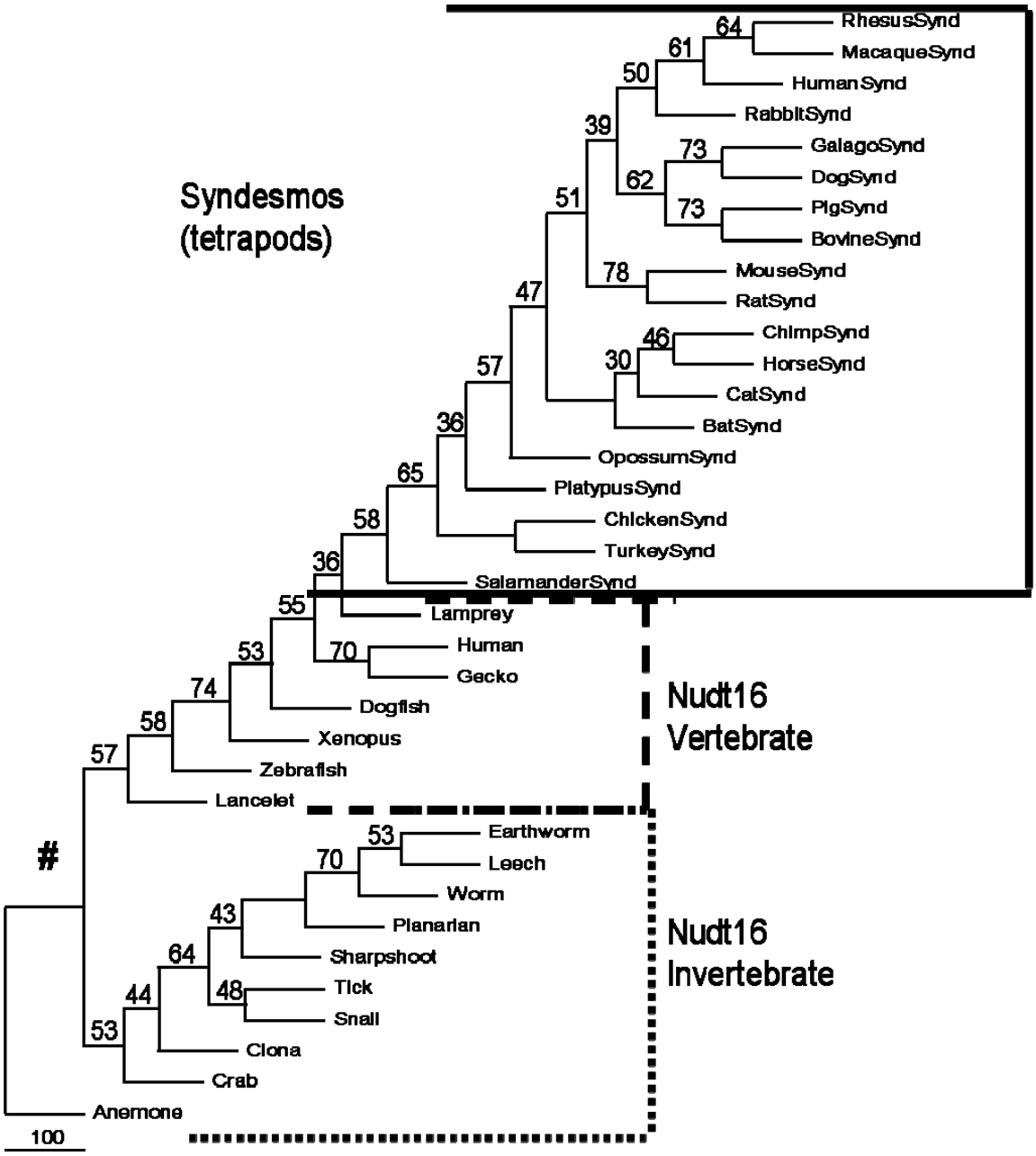


Figure 2.5: A tree of Nudt16 and Syndesmos identifies them as paralogs. A. Amino acid sequences in were aligned, graphed, and analyzed as in Methods. The tree was visualized with TreeView. Bootstrap values are shown. Vertebrates and invertebrates are separated and phylogenetic relationships are comparable to the generic tree of evolution. Syndesmos sequences branch from the tetrapod lineage of the Nudt16 sequences.

Phylogenetic relationships at the nucleic acid level

The genetic code is redundant; most amino acids are encoded by more than one triplet. Thus comparisons of proteins on the amino acid level may mask some of the evolutionary drift or changes that occur at the genomic level (Simmons and Ochoterena, 2000; Simmons et al., 2001; Simmons et al., 2002). Since we are interested in the phylogeny and determining the relatedness of these genes across a variety of organisms, we also performed phylogenetic comparisons at the nucleic acid level. The programs within the PHYLIP package were used, including DNAdist, DNAml, DNAmlk, and DNApars, to obtain phylogenetic trees using alignments of the nucleic acid sequences with the same methods and constraints that were used for the proteins. Different programs resulted in very slight rearrangements of the branches (compared to the amino acid trees) but demonstrated a similar phylogeny. Vertebrates were separated from the invertebrate organisms and the appearance of Syndesmos appeared at the same place on the trees generated from either the amino acid or the nucleotide sequences.

While single nucleotide polymorphism is typically a standard for determining relatedness between organisms, both nucleotide and amino acid alignments were examined because of the very broad spectrum of organisms compared. The ability to obtain similar trees despite the diversity of the organisms strengthens the conclusions about relatedness and demonstrates the ancient origin of the Nudt16 protein and the relatively recent appearance of Syndesmos.

Nudt16 and Syndesmos are related by gene duplication, not RNA processing

Searches of the genomic databases provided additional supportive data about the gene organization and chromosomal localization of the protein paralogs. In mammals the two proteins are encoded by single copy genes located on distinct chromosomes. Both genes contain two introns; the relative location of each intron was conserved between the two gene families (Figure 2.3B). Each exon contains one of the functional/conserved domains identified in Figure 2.1. The conserved introns positions indicate that the loss in Syndesmos of the amino acids critical for cap hydrolysis was not be due to alternative splicing but from mutation of the DNA (see Figure 2.3). Likewise, the nucleotide sequences indicate that the sequence changes altering the amino acids required for catalysis in the NUDIX domain can not be due to frame shifting or an alternative reading frame. Table 2.2 lists the chromosomal locations of the genes for these two proteins. In the cases where chromosomal synteny is known the genes have maintained similar relative positions (human and chimp).

Functional similarity between Nudt16p and Syndesmos

To assess which (if any) of the functional biochemical characteristics of Nudt16p proteins were shared by the Syndesmos paralogs, we cloned the human Syndesmos protein. The Syndesmos protein sequences lacked the amino acids critical for catalysis so they were not expected to have decapping activity but the other properties could be tested. We used PCR to amplify the coding region of the human Syndesmos cDNA. The purified protein was assayed for decapping activity, ability to homodimerize and for RNA binding. Since Syndesmos proteins lack the amino acids critical for cap hydrolysis our

results indicating the human Syndesmos protein did not have decapping activity (Figure 2.2) were not surprising.

Xenopus Nudt16p protein forms a homodimer in solution, readily visible in protein-protein chemical crosslinking assays (Peculis et al., 2007) and in the X-ray crystal structure. It is not yet clear whether the monomer or dimer is the functional form of the protein. The human Nudt16p protein can also form a dimer but less readily; a smaller portion of the human Nudt16p protein is involved in dimer formation (Figure 2.6A) demonstrated by more of the monomer form being present despite incubation with crosslinker relative to the Xenopus protein. Human Syndesmos did not form detectable homodimers under the conditions used in this assay (Figure 2.6A).

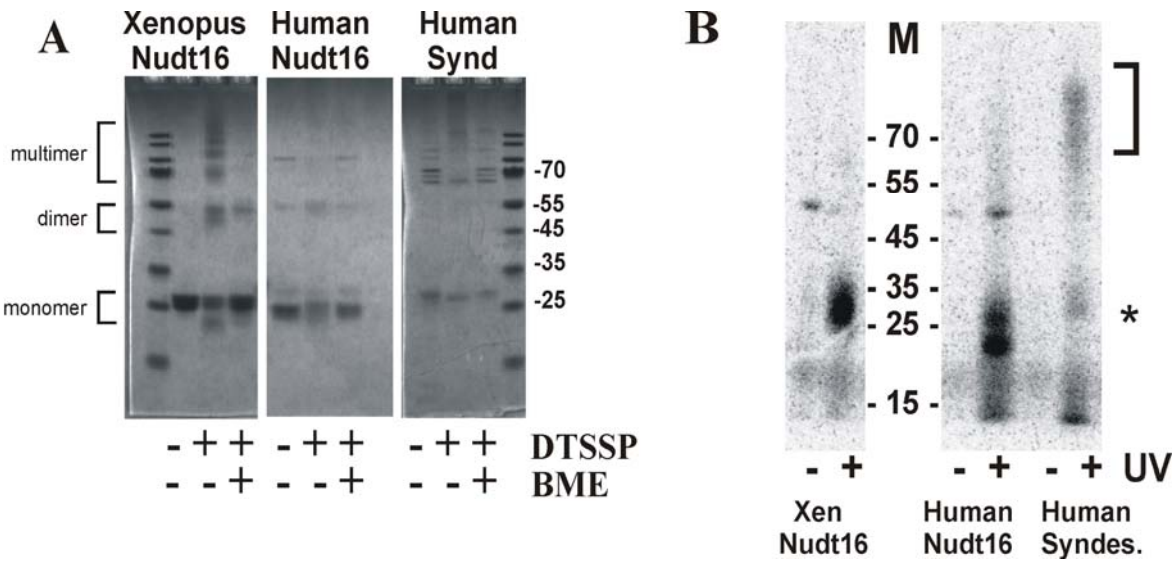


Figure 2.6: Functional comparison of paralogous proteins. A. Chemical crosslinking to examine homodimer formation. Over-expressed, purified proteins were untreated or incubated with the reversible chemical crosslinker DTSSP (as indicated). One half of the treated samples were heated in the presence of reducing agent, β-mercaptoethanol (BME) to reverse the crosslink. All samples resolved on a Nu-PAGE gel. The 30 kDa *Xenopus* protein crosslinked to form a ~60 kDa dimer and larger multimers, as indicated. These resolved to monomers under reducing conditions. Human Nudt16 formed a dimer which also resolved to a monomer under reducing conditions. Human syndesmos did not appreciably alter mobility in the presence of the crosslinker, although some of the higher molecular weight contaminating proteins did alter mobility. Molecular weight markers are present on the far right and left with values indicated on the right. B. RNA-protein crosslinking examined RNA binding. RNA and protein were incubated, then exposed (+) or not (-) to UV. After RNaseA digestion samples were resolved on Nu-PAGE gels and exposed to a phosphorplate. U8 snoRNA was crosslinked to *Xenopus* Nudt16 protein in a UV-dependent manner, resulting in labeled protein at 30 kDa. Human Nudt16 forms UV-dependent crosslinks, migrating slightly faster, consistent with the smaller size of the human protein (see Panel A). The human Syndesmos protein forms a faint UV-dependent crosslink migrating at 30K, indicated by the asterisk and a higher order crosslink (indicated by the bracket) over 70 kDa. Molecular weight markers (M) are as indicated.

Molecular modeling

Since Nudt16p and Syndesmos displayed differential ability to form dimers in the crosslinking assay, the question of electrostatic surface charges of each of the proteins of interest was explored by molecular modeling. Amino acid sequences for the human and sharpshooter Nudt16 and the human Syndesmos protein were aligned to the Xenopus Nudt16p (X29) sequence (PDB: 2A8P). Conserved amino acids or short sequences were used to ensure correct alignment and orientation of the modeled structures (Figure 2.7A). The blue strand is the CVTHFY sequence conserved in the Nudt16p sequences (see Figure 2.3, blue asterisks). The red residue was F49 in *Xenopus*, a well conserved residue that is located very near the catalytic site (red asterisk in Figures 2.3 and 2.7A). A second phenylanine residue near the active site was highlighted in black (black asterisk at F62 in X29 in Figures 2.3 and 2.7A). The Nudt16 and Syndesmos proteins were modeled as monomers. X-ray crystal data revealed the X29 protein was a dimer but the structure is shown with a yellow line to indicate the dimer interface. The monomer unit of the Nudt16p protein above the yellow line is positioned and aligned with the orthologs in the ribbon diagrams (Figure 2.7A). Once the molecules were rotated to correctly orient them relative to X29, the electrostatic potential surface charge was calculated for each structure (Figure 2.7B).

The alignments shown by ribbons in Figure 2.7A provided further evidence that the proteins are orthologs/paralogs and closely related. The calculated electrostatic potential of X29 (Figure 2.7B) indicated that the protein was highly polarized with the predicted RNA binding site on the positively (blue) charged face (Scarsdale et al., 2006).

The predicted surface of the human Nudt16p protein was not as extensively charged but positive patches were present. The corresponding face of the Sharpshooter Nudt16 protein was even less charged but there was a strongly positively charged patch just proximal to the NUDIX site where the RNA would bind. Interestingly, the human Syndesmos protein had an electrostatic charge potential that most resembled that of the *Xenopus* Nudt16p protein monomer, X29.

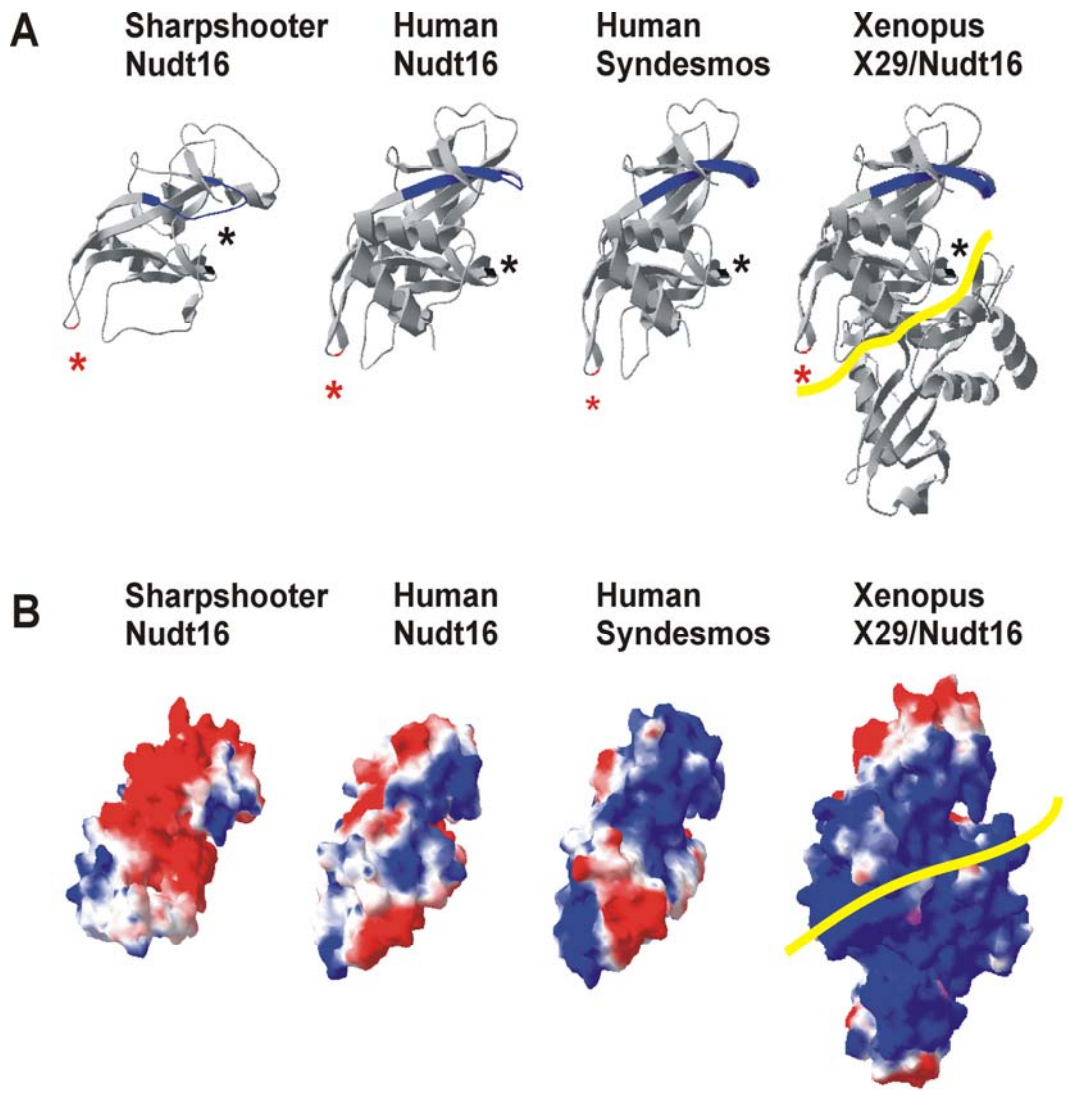


Figure 2.7: Structural comparison of paralogous proteins. A. Molecular modeling of the orthologs and paralogs of Xenopus Nudt16p. The Xray crystal structure of *Xenopus* X29/Nudt16p (PDB: 2A8P) was used to model the human and sharpshooter Nudt16p proteins and human Syndesmos. Selected conserved residues were highlighted in the ribbon models to correctly align the models relative to each other. The blue strand (CVTHFY) is indicated with blue asterisks in Figure 2.3. The black asterisk is a conserved Phe residue near the catalytic site while the red asterisk is a second conserved Phe residue indicated in Figure 2.3. The yellow line in the *Xenopus* protein is oriented across the dimer interface; the monomer above the line is oriented and positioned relative to the other 3 proteins. B. Electrostatic charge potential of structures. Once the proteins were correctly oriented, electrostatic charge potential was calculated. Blue is positively charged surface and red is negatively charged. The yellow line in the Xenopus protein is oriented across the dimer interface as in Panel B.

Protein-RNA Cross-linking :

Another characteristic of the *Xenopus* protein was that it could efficiently bind U8 snoRNA, as evidenced both by electrophoretic mobility shift assay (EMSA) and by UV-mediated RNA-protein crosslinking (Tomasevic and Peculis, 1999). The *Xenopus* Nudt16p protein was first found because of its high affinity for U8 snoRNA and ability to generate an EMSA (Tomasevic and Peculis, 1999). However with the possible exception of the weak affinity demonstrated by the rat ortholog (data not shown, MT and BAP), no other Nudt16p protein has been found to have sufficiently high affinity for detectable complex formation in EMSA.

Another method for examining U8 RNA-X29 protein interactions involved the use of 4-thioU labeled RNA (Tomasevic and Peculis, 1999). This resulted in a ~30kDa UV-dependent crosslink. The RNA-protein crosslink assay was performed using the human and sharpshooter Nudt16 proteins and the human Syndesmos protein. Proteins were pre-incubated with U8 RNA labeled with 4-thioU and ^{32}P-rATP. The reaction was then exposed to UV light to crosslink molecules interacting within 4-10 angstroms. Samples were treated with RNA, resolved on SDS gels and exposed to phosphorplates. The expected 30 kDa crosslink was seen for the *Xenopus* protein (Tomasevic and Peculis, 1999) (Figure 2.6B). The human Nudt16p protein is slightly smaller; a UV-dependent crosslink was detected migrating near 25 kDa. The sharpshooter protein also generated a UV-dependent crosslink at the predicted size. The human Syndesmos protein generated a UV-dependent RNA-protein crosslink at a size consistent with the full length Syndesmos protein. The crosslinks were not seen in comparable reactions containing

large amounts of bovine serum albumen and could be competed with 100 molar fold excess unlabeled U8 RNA or 2,000-molar fold excess tRNA, but not by lower amounts of tRNA (data not shown).

Collectively these data indicated that the Syndesmos protein and Nudt16 protein are closely related. Syndesmos, like the Nudt16p protein orthologs, can bind RNA. Unlike the Nudt16p proteins, the Syndesmos protein can not decap RNA (Figure 2.2) nor form homodimers under the conditions used here (Figure 2.6).

Discussion

We found orthologs of Nudt16 in 57 metazoans, both invertebrates and vertebrates. All putative orthologs contained an intact NUDIX domain plus the two additional structural/functional regions that are conserved among vertebrate Nudt16 proteins. Phylogenetic trees were generated from sequence alignments of the orthologs. Over-all the trees were consistent with the accepted phylogeny: the division between vertebrates and non-vertebrates was apparent, the two Cnidarians were grouped together closest to the ancient ancestor, and the mammals were grouped with the marsupials and monotremes, being more closely related than to the eutherian mammals.

There were a few notable exceptions in the placement of some organisms on the tree. Some of these sequences were known to have partial or incomplete cDNA or EST sequences. The others possibly resulted from imprecise manual editing of whole genome shotgun sequences. Trees generated from the alignment of protein orthologs do not always mirror trees obtained by phenotypic traits or rRNA sequence (Baldauf,

2003; Hillis, 1987; Meyer and Zardoya, 2003; Philippe et al., 2005; Woese, 2000). Sequence drift is informative, however, as it can identify the natural sequence variation tolerated at any single amino acid residue since the selective parameter is the ability to function efficiently *in vivo*. The low bootstrap values, represented by the numbers on the schematics, reflect the exceedingly wide variety of organisms being aligned and compared (most of one of the three arms of the Tree of Life). As such, these values are not inconsistent with other proteins that are conserved across such a wide diversity of organisms.

Functional Orthologs of Nudt16

Nudt16 orthologs were identified in many organisms. A representative of the insects, the Sharpshooter Nudt16 protein, was cloned, purified and characterized here. The insect Nudt16 protein was shown to be active for both RNA binding and 5' cap hydrolysis *in vitro*, demonstrating it to be a functional ortholog of the vertebrate nuclear decapping proteins.

To date, convincing orthologs have not been identified in yeast or plants; this is not to say the protein does not exist there. Database searches indicated that Nudt16 was rapidly diverging at the amino acid level; true functional homologues may not be distinguished by sequence comparisons alone. Putative orthologs in other species may have been missed in these searched because of their high dissimilarity in regions that we would have allowed to be "variable" but which the programs disallowed because the stringency of the search parameters employed. On-going and future searches with custom search parameters dictating high stringency requirements within the

'conserved' regions and lesser stringency in 'variable' regions may lead to the identification of orthologs of Nudt16 in other organisms. Functional homologues may be more effectively identified through structural similarity and validated in *in vitro* functional assays.

Syndesmos is a paralog of Nudt16p

Previous data base searches for Nudt16-related proteins identified the avian and mammalian protein Syndesmos, alternately (and sometimes incorrectly) annotated in databases as both Syndesmos and Nudt16-like. This work demonstrates Syndesmos is closely related to but differs from Nudt16 primarily in that Syndesmos is not a NUDIX protein (see Figure 2.3), and is reported to be cytoplasmic.

The similarity between Nudt16 and Syndesmos protein sequence, gene organization and evolutionary distribution suggests that syndesmos probably originated as a gene duplication event of the more ancient Nudt16. Subsequent divergence of the Syndesmos protein allowed it to acquire new cellular functions (due to or perhaps assisted by the loss of the NUDIX domain). Data supporting this hypothesis include the different chromosomal location of the genes but shared number and placement of introns within the coding region (see Table 2.2 and Figure 2.3B). The presence of introns makes retroposition via an RNA intermediate unlikely. Intron/exon borders that flank but do not interrupt the NUDIX domain makes alternative splicing less likely (Figure 2.3B). Finally, careful sequence analysis indicated that neither alternate splicing nor frame-shifting are feasible explanations for the origin of the two proteins from a single gene. Thus, Syndesmos arose from a gene duplication of the more ancient Nudt16

hydrolase gene early in the tetrapod line, resulting in the cytoplasmic Syndesmos protein.

The X-ray crystal structure of the *Xenopus* X29/Nudt16p protein (Scarsdale et al., 2006) demonstrated that the residues comprising the conserved Domain 3 (in blue on Figures 2.1 and 2.3) were involved in the dimerization interface. The comparable, conserved region in the Syndesmos protein is involved in the interactions with the Paxillin family of proteins and Syndecan-4 in vivo (Baciu et al., 2000; Denhez et al., 2002). Thus it is likely that while the *Xenopus* Nudt16 can homodimerize *in vitro*, this interaction surface may be involved in heterologous protein interactions *in vivo*. Additional insights into particle assembly and *in vivo* binding partners could be gleaned from examining the structural interactions of Syndesmos and its *in vivo* binding partners.

The human Syndesmos protein, like the Nudt16 proteins, displays RNA binding activity, although we have not yet mapped the residues involved in RNA interactions on either protein. The biological role – if any – of the RNA binding activity for Syndesmos is not known, in part because the *in vivo* function of Syndesmos is not clearly defined.

This phylogenetic study provided strong evidence that the Nudt16 gene is the more ancient protein, and that Syndesmos is a newer acquisition, having lost both the catalytic residues in the NUDIX domain and its nuclear/nucleolar location. The lack of Syndesmos in reptiles, such as the gecko and the anole, was somewhat surprising and may reflect a lack of data (rather than a "real" evolutionary trait) (Philippe et al., 2004) or may indicate a stochastic loss of the Nudt16 gene in some – or perhaps all organisms

in some branches of the evolutionary tree. As more reptile genomes are fully sequenced and released, it will be interesting to discover whether other reptiles and amphibians have a Syndesmos ortholog.

An evolutionary conserved nuclear decapping machinery

Our early studies identified the X29/Nudt16 protein through its ability to bind U8 snoRNA with high affinity and specificity (Tomasevic and Peculis, 1999). Once cloned and over-expressed, the nuclear localization and decapping activity of the protein was discovered and a role in negative regulation of ribosome biogenesis proposed (Ghosh et al., 2004). Recent studies have demonstrated that several mammalian orthologs (Ghosh et al., 2004; Peculis et al., 2007) and the insect ortholog of this protein (this work) can remove the m^7G cap from a variety of RNA substrates in a metal-dependent manner. The nuclear enzyme has a broader substrate specificity.

This work was initiated to determine the extent of evolutionary conservation of the Nudt16 protein. The findings revealed unexpected information about the divergence of an ancient protein family. The cytoplasmic protein Syndesmos, present in many vertebrates, was demonstrated to have RNA binding activity. Syndesmos is a paralog of the more ancient nuclear hydrolase, Nudt16 based on sequence and functional data. How – or whether – this RNA-binding activity is used by the protein *in vivo* remains to be demonstrated.

The distribution of Nudt16 across metazoans implies an evolutionary conserved biological role for this protein. This nuclear decapping protein has the ability to remove the m^7G cap (present on pre-mRNAs and mature mRNA prior to export) and the $m^{227}G$

cap (present on nuclear-limited snRNAs and snoRNAs). Thus it is likely to be the catalytic component in an evolutionary conserved (but as yet uncharacterized) nuclear pathway present in all metazoans. Nudt16 may act upon a variety of RNA substrates including the stable, nuclear-limited RNAs (snoRNA, snRNA and other hypermethylated RNAs) and m^7G capped pre-mRNAs and mRNA prior to their export to the cytoplasm giving it a very early position in the complex pathway for control of gene expression. These evolutionary data have facilitated new analyses of the structure-function relationship within the protein family and can provide additional insight into other biological roles for these proteins.

Funding

This work was funded in part by an NIH training grant (T90 DK070105 to MJT) an NSF grant (#0718256 to BAP) and internal funding from MU (to BAP).

Acknowledgements

We would like to thank Wayne Hunter for his generous donation of the construct encoding Sharpshooter Nudt16p and Chad Raw for assistance in cloning. We are grateful to Kristen Reynolds and Frank Schmidt for scientific discussions and to Frank Schmidt for critical reading of the manuscript.

2.5: Up-to-date information and newest conclusions

In the article, 57 Nudt16 genes were identified and included in the analysis (Taylor and Peculis, 2008). Syndesmos genes were also identified, 19 were recorded, and analyzed in a variety of organisms. After publication, additional sequences were found increasing the number of Nudt16 homologs to 65 (see Table A.2, page 196 for more details on the sequences). Alignments were completed using CLUSTAL X1.83 as discussed in the article. The Bayesian analysis was too computationally difficult for the computers. Therfore, the phylogenetic relationships were generated using PHYLIP.

The amino acid tree was generated using the maximum likelihood method (see Figure A.4, page 194). It follows the general pattern of evolution closely. The nucleic acid tree was made using the maximum parsimony algorithm (see Figure A.5, page 195). The nucleic acid tree is more divergent. This is not unexpected as several sequences were generated by manual editing of whole genome shotgun sequences which is not 100% error proof.

These data do indicate that the Nudt16 is evolutionarily conserved. The most recent common ancestor for the Nudt16 gene is the Cnidarians; three Cnidarians have the gene. Other invertebrates have the gene including insects, crustaceans, and worms. Finally, vertebrate organisms contain Nudt16; some also contain the paralog syndesmos. These analyses support the hypothesis that the Nudt16 gene is conserved suggesting the existence of a 5'-3' nuclear RNA degradation pathway for nuclear-limited RNAs could exist in these organisms.

Chapter 3: Mechanisms of Nudt16 Function in *Xenopus*

3.1: Why is studying the mechanism of Nudt16 function important?

The study of enzyme kinetics and mechanism of action are important topics to address when characterizing any enzyme. The previous analyses of *Xenopus* Nudt16 function demonstrated a metal dependent protein that catalyzed an RNA decapping reaction (Ghosh et al., 2004; Peculis et al., 2007). Nenad Tomasevic, a previous post-doc in the Peculis laboratory, generated data supporting a two-site binding model with one high affinity and one low affinity site (NT & BP, unpublished results). These data, obtained with protein purified from *Xenopus* ovary were intriguing, but at the time there was no mechanistic model to fully appreciate them.

The goal of this chapter is that by learning more details about the mechanism of the *Xenopus* Nudt16 protein's function, a better model for Nudt16 function can be tested in other organisms. To ascertain the function, a detailed kinetic and mechanistic analysis of the *Xenopus* Nudt16 protein is discussed in the following pages.[5] The tests begin with decapping assays including assays of protein titrations to determine the functional enzymatic unit, RNA titrations to determine the observed reaction rate, and competitions with the non-specific competitor tRNA. EMSAs are also used to draw

[5] Chapter 4 addresses the function of the other organisms, but none were as thoroughly characterized as the *Xenopus* ortholog.

some conclusions about the number of binding sites on the protein. These data are combined to build a mechanism of Nudt16 function using a 2-site binding model that is based upon new understanding from the details discovered in this work. The last section discusses other questions that come from the model and possible experiments designed to address those questions.

3.2: Introduction to mechanisms, models, and equations of kinetics and Nudt16 function

Before approaching the mechanism of a reaction, it is necessary to determine the overall equation of the reaction. Simply stated the chemical reaction of Nudt16 mediated RNA decapping is as follows:

$m^7GpppG\text{-}RNA + Metal + Nudt16 \rightarrow m^7Gpp + pG\text{-}RNA + Metal + Nudt16$ (1)

Both the binding of the RNA to the protein and the catalyzed reaction have kinetic properties that need to be analyzed. The experiments and results discussed in the subsequent sections will address different parts of the reaction as individual steps. The final reaction mechanism with each known step is presented in Section 3.5 (pages 95-98).

To begin the kinetic analysis, consider Equation 2 shown below

$$Nudt16 + m^7GpppG\text{-}RNA \underset{k-1}{\overset{k1}{\rightleftharpoons}} Nudt16{:}m^7GpppG\text{-}RNA \xrightarrow{k2} Nudt16 + m^7Gpp + pG\text{-}RNA \quad (2)$$

In this reaction scheme, Nudt16 is the enzyme and m^7GpppG-RNA is the cap-labeled RNA or substrate; the metal is not shown here. After catalysis, the cap and the monophosphorylated RNA are the products. The goal of the chapter is to ascertain the mechanism of function; the experiments start with the decapping assay which measures the activity of the enzyme but not the individual steps of RNA binding and product release.

The first experiments are protein titrations that address whether the functional enzymatic unit is a dimer or a monomer. From previous work it is known that the

crystalline protein formed a homodimer (Peculis et al., 2004; Scarsdale et al., 2006). Chemical crosslinking analyses and the presence of a ~60kDa band in purified protein samples showed that the protein can form homodimers *in vitro* (Peculis et al., 2007). Finally, when assayed by gel filtration or size exclusion chromatography, the protein had an apparent Mr (relative molecular mass) of ~60 kDa (data not shown). These data suggest that the protein can form a homodimer. Other proteins, including some NUDIX proteins, are known to require homodimerization for function (Chen et al., 2006; Coseno et al., 2008; Cowley et al., 1997; Dobrzanska et al., 2002; Graziano et al., 2006a; Graziano et al., 2006b; Holbrook et al., 2003; Scott et al., 1991; Wang et al., 2002a). The protein titrations will determine the amount of protein needed to detect activity and whether the protein is mostly monomer or mostly homodimer at those amounts of protein.

To compare each of those concentrations, the protein titration data will be fitted to an equation for a hyperbolic curve. Typically, the Michaelis-Menten equation is used; however, the data presented here are pre-steady state kinetics under single-turn over conditions. For that reason, typical Michaelis-Menten analyses are impossible. Instead, the hyperbolic curve equation uses apparent kinetic constants with similar meanings but different terms:

$$v = \frac{B_{max} * [E]}{K_{app\ prot} + [E]} \quad (3)$$

In the equation, v is the velocity of the reaction (plotted on the y-axis) and [E] is the concentration of enzyme (plotted on the x-axis). *Bmax* is the maximum velocity of

the reaction under these conditions which have been optimized for the *Xenopus* Nudt16 protein. *Bmax* can then be used to find an observed rate of the reaction because

$$k_{obs.rxn} = \frac{B_{max}}{[S]} \quad (4)$$

The $K_{app\ prot}$ is the concentration of enzyme at which 50% *Bmax* is reached. After completing this analysis with the protein titrations, RNA titrations are used to confirm the results. Data from RNA titrations are fitted to the hyperbolic equation shown in Equation 5 and then used to calculate the observed reaction rate using Equation 6.

$$v = \frac{B_{max} * [S]}{K_{app.rxn} + [S]} \quad (5)$$

$$k_{obs.rxn} = \frac{B_{max}}{[E]} \quad (6)$$

Instead of the concentration of enzyme ([E]), the independent variable (x) is the concentration of substrate ([S]). The *Bmax* and $k_{obs\ rxn}$ are defined the same way but will not have the same numerical value because the reaction conditions have changed; these variables depend on the reaction conditions. Finally the specificity constant is the ratio k_{obs} / $K_{app\ rxn}$ which should be a property of the enzyme.

After addressing the functional enzymatic unit by both RNA and protein titrations, the next experiments study how many RNAs bind to the functional enzymatic unit. Does cooperativity exist in this system? To test this, RNA titrations were fitted to a hyperbolic curve equation that addressed cooperativity (Kurganov et al., 2001)

$$v = B_{max} * \frac{([S]/K_{app})^{\wedge} n}{1 + ([S]/K_{app})^{\wedge} n} \quad (7)$$

In this scenario, [S] is the concentration of substrate, K_{app} is the concentration of RNA at which the reaction reaches half the maximum velocity. It should be the same as the $K_{app\ rzn}$ value obtained by Equation 5. The coefficient n is a coefficient describing cooperativity, and v and *Bmax* are the velocity and maximum velocity, respectively, of the reaction.

The coefficient, n, describes cooperativity because if n is greater than one, it implies that binding is cooperative; with n being equal to the minimum number of molecules binding to the protein. If n is less than 1, the system is negatively cooperative. If n is equal to 1, no cooperativity is evident in these conditions.

Knowing whether the binding of the RNA to the enzyme is cooperative adds to the model, details of the binding kinetics cannot be demonstrated in this assay. To study the RNA binding in more detail, the gel shift or electrophoretic mobility shift assay (EMSA) is used. The assay studies binding but does not address catalysis. Instead, the affinity of the RNA to the protein is tested by looking for a shift of the free RNA to the bound RNA band which migrates more slowly than the free RNA band.

To add more detail to the model, other areas of interest are discussed. One area is the lack of turn over exhibited by this protein (Peculis et al., 2007). Many reasons could explain this lack of catalytic turnover; one explanation is product inhibition with two likely possibilities being inability of the enzyme to release the cap or the inability of the enzyme to release the RNA. Either scenario could cause single-turnover characteristics. The cap microdialysis experiment addresses this question by following

the cap through the decapping reaction to see whether it is released or stays bound to the protein.

In addition to product inhibition, non-specific inhibition can also alter the reaction mechanism. Nonspecific competitors are known to inhibit Nudt16 decapping activity; no mechanistic details are known. By understanding the effects of inhibition, a model will be able to give more information about previously unexplained results.

The goal for this project was to understand the mechanism of Nudt16 function and apply it to the current model of Nudt16 function. Using decapping assays, EMSAs, and competitions, the model was adjusted to account for the data. The initial binding model was simplistic: a model with two binding sites; one high affinity and one low affinity site. The protein is a homodimer which can bind multiple RNA molecules. The non-specific competitor tRNA[6] demonstrates a non-competitive mode of inhibition, supporting the two-site binding mechanism. In one site the substrate RNA binds, while tRNA binds in the second site. Binding of one RNA does not affect the binding of a second RNA because no cooperativity exists. Activity is affected because only one site is active; the other site can only bind RNA. The overall effect is an inhibition of the Nudt16 protein.

The hypothesis stated in the first page of this chapter included the need to understand the mechanism of function in order refine the model of Nudt16 function so that it can be tested in experiments with other orthologs (see Chapter 4 of this document, pages 101-141). The *in vitro* conservation of Nudt16 decapping activity and

[6] tRNA is not a substrate for the Nudt16 reaction. Further explanation of tRNA inhibition is found in Section 3.4 (pages 87-94).

mechanism would support a scenario in which Nudt16 function is conserved *in vivo.* The experiments discussed in the next sections were used to describe a proposed mechanism of function that can be used as the standard to which all future tests of function are compared.

3.3: Materials and methods used to determine the kinetic mechanism of Nudt16 activity

Protein Titrations

Decapping reactions were carried out essentially as described in Chapter 2 and elsewhere except that the amounts of protein were altered in each experiment (Taylor and Peculis, 2008). Briefly, 0.025 pmol (5 nM) cap-labeled U8 snoRNA was incubated in the presence of 300 μM Mn^{2+} and varying levels of protein (see Table 3.1 for concentrations of protein). Reactions were incubated at 37°C for 3 minutes in a total volume of 5 μL. Reactions were stopped by addition of 0.5 μL 0.5 M EDTA and analyzed by TLC. Percent decapping was quantitated using the Quantity One software (Bio-Rad). Data were fitted using GraphPad Prism 5 software using Equations 3 and 4 above.

Table 3.1: Concentrations of Nudt16 protein in protein titration experiments. Decapping reactions with 0.025 pmol (5 nM) U8 snoRNA and various amounts of protein were incubated 3 minutes at 37°C. Amount of protein in both grams (g) and picomoles (pmol) are shown. Reactions containing excess enzyme as well as reactions with limiting enzyme were included in the analysis as is shown by the protein:RNA ratios in the last column.

Amount of Protein (g)	Amount of Protein (pmol)	Concentration (nM)	Protein:RNA Ratio
$6.0x10^{-7}$	24.64	4930	986:1
$4.0x10^{-7}$	16.43	3290	658:1
$2.0x10^{-7}$	8.21	1640	328:1
$1.2x10^{-7}$	4.93	986	197:1
$6.0x10^{-8}$	2.46	493	97:1
$4.0x10^{-8}$	1.64	329	66:1
$2.0x10^{-8}$	0.82	164	33:1
$1.0x10^{-8}$	0.41	82.1	16:1
$7.5x10^{-9}$	0.31	61.6	12:1
$5.0x10^{-9}$	0.21	41.1	8:1
$2.0x10^{-9}$	0.08	16.4	3:1
$1.0x10^{-9}$	0.04	8.21	2:1
$6.1x10^{-10}$	0.03	5.00	1:1
$3.2x10^{-10}$	0.01	2.63	0.5:1
$2.0x10^{-10}$	0.008	1.64	0.3:1
$1.0x10^{-10}$	0.0004	0.82	0.2:1

RNA Titrations

Decapping reactions where the concentration of substrate RNA varied but the concentration of protein was stable were assayed for decapping activity in the presence of 300 μM Mn^{2+}. The reactions were completed as described for the protein titrations except that the substrate concentration changed as shown in Table 3.2. RNA titrations were fitted to Equations 5-7 in the preceding section using GraphPad Prism 5 software.

Table 3.2: Concentration of RNA for RNA titration experiments. RNA was added to reactions containing 329 nM Nudt16 protein and allowed to decap for 3 minutes. Reactions were quantitated and analyzed as described in the above section.

Amount of RNA (pmol)	Concentration (nM)	Protein:RNA Ratio
0.1	20	17:1
0.075	15	22:1
0.05	10	33:1
0.025	5	66:1
0.01	2	165:1
0.0075	1.5	219:1
0.005	1	329:1
0.0025	0.5	658:1
0.001	0.2	1654:1

Electrophoretic Mobility Shift Assays (EMSAs)

Gel shift reactions with varying concentrations of substrate RNA (see Table 3.3) and 27.4 nM protein were allowed to reach binding equilibrium by incubating at room temperature for 15 minutes. Samples were resolved on a 4% native acrylamide gel at 12 mA for 3.5 hours to ensure the bands had clearly separated (Tomasevic and Peculis, 1999). Gels were dried, exposed to a phosphor screen for 16 hours, and analyzed using the Quantity One software (BioRad). The data were graphed using Graphpad Prism 5.

Table 3.3: Concentration of RNA for EMSA experiments. RNA was added to reactions containing 27.4 nM Nudt16 protein and allowed to equilibrate for 15 minutes. Reactions were resolved on a 4% native polyacrylamide gel and analyzed as described in the above section.

Amount of RNA (pmol)	Concentration (nM)	Protein:RNA Ratio
0.007	1.3	21:1
0.01	2.7	10:1
0.03	6.8	4:1
0.07	13	2:1
0.10	20	1.4:1
0.17	34	0.8:1
0.25	50	0.5:1
0.34	68	0.4:1
0.67	133	0.2:1
1.0	200	0.1:1
1.7	333	0.08:1

Cap Release Microdialysis Experiments

Four samples, each containing 0.025 pmol radioactive cap-labeled U8 snoRNA substrate and a final concentration of 150 µM $MnCl_2$ were prepared. Tube 1 contained only RNA and no protein so it received 4 µL Buffer B1 (50 mM Tris pH 8.5, 150 mM NaCl) in order to bring the total volume up to 20 µL. Tube 2 included RNA and 2.46 pmol (493 nM) Nudt16 protein. Immediately upon addition of the protein, 1 µL 0.5 M EDTA was added to the sample so no decapping would occur. Tubes 3 and 4 contained both protein and RNA in a total of 20 µL. Reactions were incubated at 37°C for 30 minutes as in the standard reactions described previously. Tubes 1-3 were stopped with 0.5 µL 0.5 M EDTA; Tube 4 did not receive EDTA. Reactions were then filtered with Microcon® YM-10 centrifugation filters (Millipore #42406) by spinning 20 minutes at 15000 rpm; flow through was called the eluate. Filters were washed twice with 20 µL Buffer B1 and flow

throughs were saved as part of the eluate. Finally, the filters and eluates were counted by Cerenkov counting to determine the distribution of the radioactively labeled cap.

tRNA titrations at constant substrate and protein concentrations

For tRNA titrations where the substrate RNA (cap-labeled U8 snoRNA) was kept constant at 0.025 pmol (5 nM), decapping reactions were set up with 4.93 nM, 82.1 nM, or 246 nM Nudt16 as described in the preceding sections. These correlate to protein:substrate RNA ratios of 1:1, 16:1, and 49:1, respectively. In each reaction, various amounts of tRNA were added to the U8 substrate RNA prior to addition of protein. The amount of tRNA extended from 0 to 2.5 pmol (500 nM) tRNA which is 100 molar fold excess (MFE) as shown in Table 3.4. RNA decapping was quantitated with the Quantity One software (Bio-Rad) and graphed with GraphPad Prism.

Table 3.4: Concentration of tRNA for tRNA titration and time course experiments. Decapping reactions with Nudt16 protein and 0.025 pmol (5 nM) U8 snoRNA were incubated with various amounts of tRNA. Amounts of tRNA and concentrations are included in the table. MFE is molar fold excess of tRNA over U8 snoRNA.

Amount of tRNA (pmol)	Concentration of tRNA (nM)	MFE
2.5	500	100
1.25	250	50
0.25	50	10
0.125	25	5
0.05	10	2
0.025	5	1
0.0125	2.5	0.5

tRNA competition titrations when substrate concentrations change

RNA titrations in the presence of tRNA were used to ascertain the type of inhibition exhibited by a nonspecific competitor. The decapping reactions were set up as described in the RNA titrations except that 1 µL of tRNA was added to each reaction so the final [tRNA] was 0, 20, 100, or 200 nM. Titrations of substrate in the presence of the nonspecific competitior tRNA were fitted to the equations for noncompetitive inhibition and mixed model inhibition provided in the Graphpad Prism software.

tRNA competition over time

For time courses with tRNA, decapping reactions with 0.025 pmol cap-labeled U8 snoRNA, 150 µM $MnCl_2$, 493 nM, 246 nM, or 82.1 nM Nudt16 protein, and 0, 0.625, or 2.5 pmol tRNA were incubated at 37°C for 0, 1, 2, 5, 10, 20, or 30 minutes. At each time point, reactions were immediately stopped with EDTA, analyzed by TLC, quantitated, and graphed as with the tRNA titrations.

3.4: Results from the study of Nudt16 mechanism in *Xenopus*

Protein Titrations

In order to understand the function of this protein, as is the hypothesis guiding this project, one of the first things to discern is whether the protein functions (defined as whether the protein exhibits decapping activity) as a monomer or as a homodimer. Protein titrations were performed to determine the effects of protein concentration on the rate of decapping. Decapping reactions with a constant concentration (5 nM) of RNA were set up with varying concentrations of Nudt16 protein (see Table 3.1). The reactions were initiated by incubation at 37°C. At 3 minutes the reaction was stopped with 0.5 µL of 0.5M EDTA. This three minute time point was chosen to ensure that the rates were initial rates. The reaction did not have enough time to go to completion and the time point was well within the linear part of the time course hyperbolas (see Figure 4.4, page 123). Reactions were analyzed as described in Section 3.3 (page 70) with Quantity One (Bio-Rad).

These data were plotted to learn whether the dimer is the functional unit of the protein or if the dimer represents two functional subunits that are interacting. At the lowest concentrations of protein (below $1x10^{-9}$), the protein is in the monomeric form; homodimers would not be likely to be present. At the highest concentrations of protein (above $1x10^{-7}$), the protein is in the homodimer state; all monomers would form dimers or multimers. Between those two extremes, a fraction of the protein is in the monomer state and the rest is in the homodimer state. If the titrations show a higher specific

activity at lower concentrations of protein, the monomer is the functional enzymatic unit. If the titrations show higher specific activity at the higher concentrations of decapping, the homodimer is the functional unit.

Data from at least 5 replicates are shown in Figure 3.1 below. As Figure 3.1 shows, the rate of the reaction increases as more protein is added until the maximum amount of cap is released, shown here by the plateau of the curve. Quantifiable decapping activity was not visible until at least 16 molar fold excess of protein (82.1 nM) was used in the decapping reaction (see Figure 3.1b). In addition, the S-shape curve at the lowest concentrations supports the conclusion that the homodimer is the functional enzyme unit. Thus at protein concentrations over 20 nM, the curve is a single hyperbola. At concentrations below this range, the data fit an S-shape curve consistent with a dimer with a K_d of approximately 6 nM.

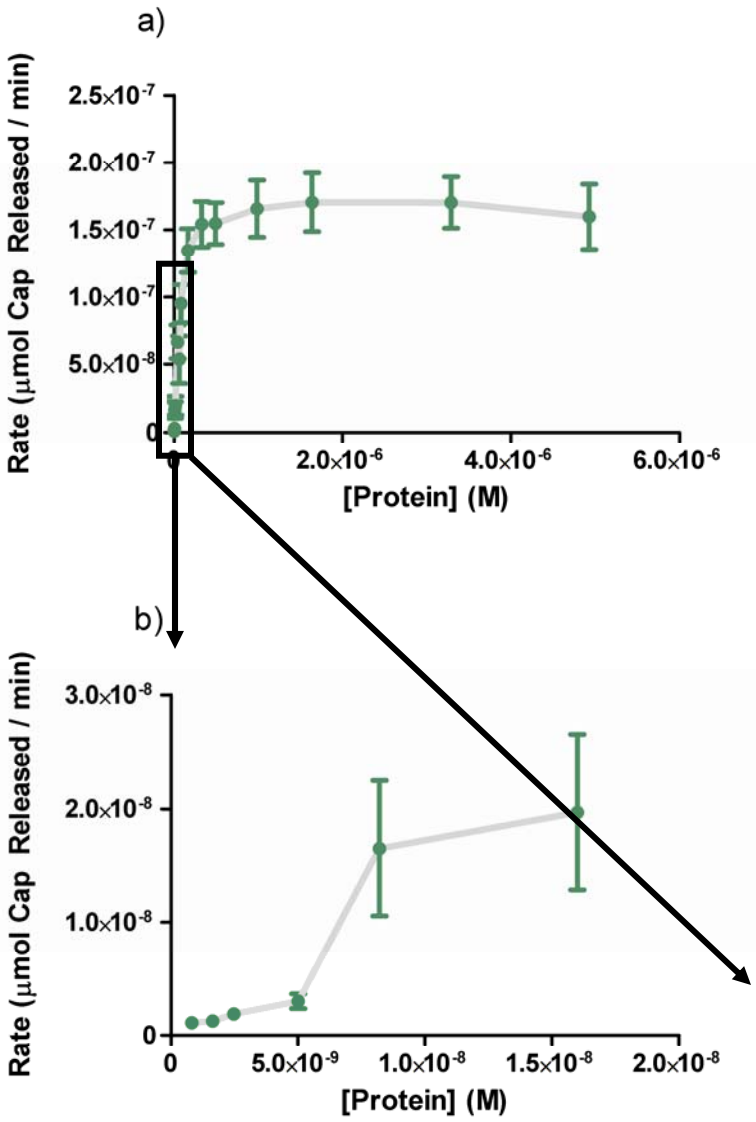


Figure 3.1: Protein titration of Nudt16. a) Plotting the protein titrations with 5 nM U8 snoRNA reveals a hyperbolic curve which was fitted to Equation 3 as discussed in Section 3.2 to find the protein titration constants of the reaction shown in Table 3.5. b) The box in a is expanded in b and illustrates the low decapping activity in lowest concentrations of protein. The curve is S-shaped, another indication that the homodimer is necessary for function.

Fitting the protein titration data to Equation 3 yielded the constants in Table 3.5. The maximal velocity is $1.8x10^{-7}$ µmol cap released / minute (1.8 pmol cap released per minute); the $K_{app\ prot}$ is 78 nM. The high $K_{app\ prot}$ supports the previous conclusion that the functional enzymatic unit is a homodimer in the conditions used for standard decapping reactions. Dividing the maximum velocity by the concentration of substrate (5 nM) yields a $k_{obs\ prot}$ of 35.4 min^{-1} (2121.6 s^{-1}).

In summary, the protein titration data support a model in which the functional unit is a homodimer. At lower concentrations of protein (below $1x10^{-9}$), no decapping occurs because the concentration of homodimers is very low. At higher concentrations of protein (above $1x10^{-7}$), the protein is fully active.

Table 3.5: Constants derived from fitting protein titration data to a hyperbolic curve equation. The maximum velocity and concentration of protein at which half maximal velocity ($K_{app\ prot}$) is reached are provided in this table. The maximum velocity units are µmol cap released / minute. The concentration of protein at half maximal velocity has units of nanomolar. The $k_{obs\ prot}$ is in inverse minutes. The specificity constant units are $min^{-1}\ M^{-1}$.

Parameter	Hyperbolic Fitting
Maximum Velocity (µmol Cap released / min)	$1.8x10^{-7} \pm 7.7x10^{-9}$
$K_{app\ prot}$ (nM)	78 ± 13
$k_{obs\ prot}$ (min^{-1})	35.4 ± 1.5
Specificity constant ($min^{-1}\ M^{-1}$)	$4.6x10^{+8} \pm 1.2x10^{+8}$

RNA Titrations

After demonstrating the functional enzyme unit is a homodimer, the second step to generate a mechanism of function for the Nudt16 protein is to look at RNA binding. Substrate RNA titrations were performed to determine the number of RNA molecules able to bind to each functional homodimer. If multiple RNAs bind to the homodimer, the data would be expected to have an S-shaped curve similar to the one exhibited by the protein titrations (see Figure 3.1b).

Decapping reactions in which the substrate concentrations changed were allowed to decap for 3 minutes at 37°C as described in Section 3.2 (see Table 3.2, page 72 for RNA concentrations). The data were fitted to Equations 5 and 7. The results are shown in Figure 3.2 and Table 3.6.

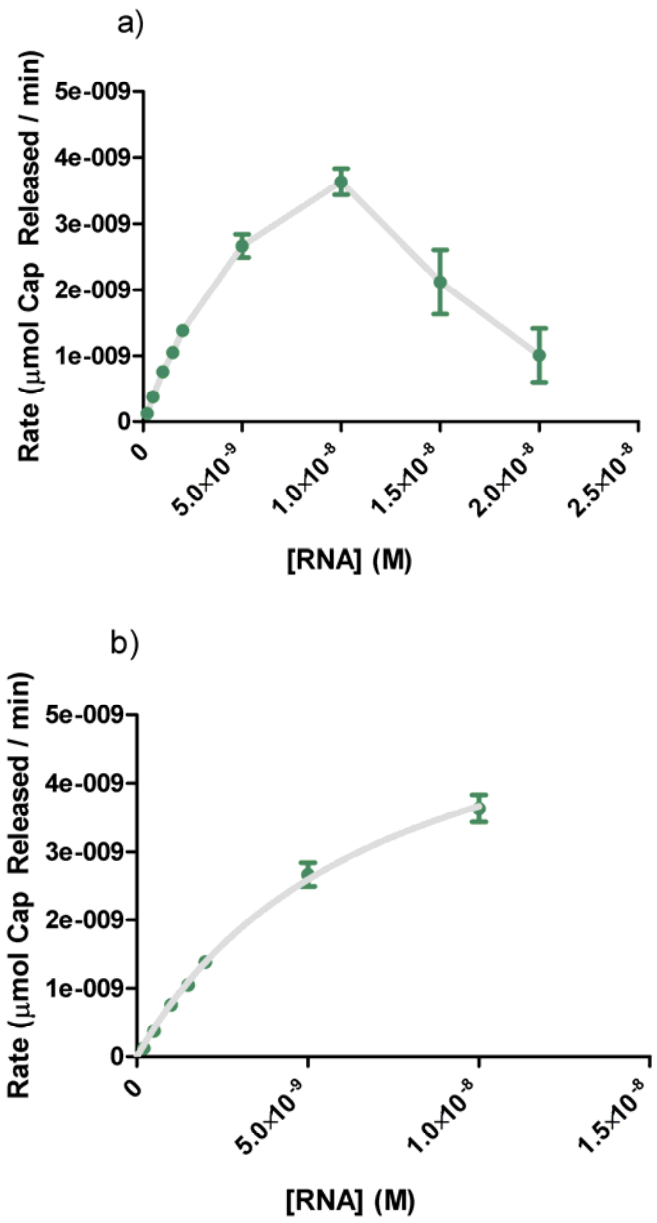


Figure 3.2: Titration of U8 snoRNA with a constant concentration of Nudt16 protein. a) Decapping reactions with various levels of RNA but a constant concentration of Nudt16 (329 nM) were incubated for 3 minutes, stopped, and analyzed by TLC. At the lowest concentrations tested, starting from 0.2 nM U8, there is an increase in decapping which peaks at 10 nM U8. Adding more RNA after this point does not increase the decapping ability, even though enzyme is still in excess. Decapping ability drops at the 15 and 20 nM concentrations of U8. b) Equations 5 and 7 were fitted to the first 7 substrate RNA concentrations (all except the 15 and 20 nM RNA concentrations).

Table 3.6: RNA titrations yield *Bmax* and $K_{app\ rxn}$ for the decapping reaction. The maximum velocity and concentration of RNA at which half maximal velocity ($K_{app\ rxn}$) is reached are given in this table. The maximum velocity units are µmol cap released / min. The concentration of RNA at half maximal velocity has units of nM. The $k_{obs\ rxn}$ is in inverse minutes. The specificity constant units are in $min^{-1}\ M^{-1}$. The cooperativity coefficient has no units.

Parameter	Hyperbolic Fit	Cooperativity Fit
Maximal Velocity (µmol Cap Released / min)	$6.3x10^{-9} \pm 4.8x10^{-10}$	$5.3x10^{-9} \pm 8.1x10^{-10}$
$K_{app\ rxn}$ (nM)	7.1 ± 0.1	4.9 ± 1.5
$k_{obs\ rxn}$ (min^{-1})	$0.02 \pm 1.5x10^{-3}$	-
Specificity constant ($min^{-1}\ M^{-1}$)	$2.7x10^{+6} \pm 1.5x10^{+6}$	-
Cooperativity Coefficient	-	1.1 ± 0.1

At first, increasing the concentration of substrate (U8 snoRNA) while maintaining a steady protein concentration increased decapping in a hyperbolic manner. At 10 nM substrate, the level of decapping reached a maximum and then decreased at the 15 and 20 nM concentrations (see Figure 3.2a). This supports the requirement for homodimer formation. At 15 and 20 nM substrate concentrations the ratio of protein:RNA drops below the 16 molar fold excess necessary for activity shown in Figure 3.1.

Fitting the data to Equation 5 yielded the hyperbolic fit described in Table 3.6. The $K_{app\ rxn}$ of the reaction was 7.1 nM; higher concentrations of RNA would only yield higher amounts of decapping if the protein:RNA ratio were adjusted. The maximum velocity of the reaction was $6.3x10^{-9}$ µmol cap released per minute (6.3 fmol cap released per minute) resulting in a $k_{obs\ rxn}$ of 0.02 min^{-1} (1.20 s^{-1}). Since this number is small, something has slowed the reaction; 3000.0 s (50.0 min) are required for RNA to

be cleaved and released as opposed to $4.7x10^{-4}$ s (0.03 min) under conditions in which the protein is titrated.

Fitting the data to Equation 7 supported the previously determined *Bmax* and $K_{app\ rxn}$. In addition to these data, the cooperativity coefficient was calculated to be 1.1 indicating the system does not demonstrate cooperativity. In this scenario, when one RNA binds, another RNA can bind but the binding of the second RNA is not influenced by the binding of the first RNA.

From the protein titrations, the mechanism was modified to include a homodimer as the functional unit. These data support that mechanism, further demonstrating the importance of considering the protein:RNA ratio in other experiments. The RNA titrations also provide evidence that multiple RNAs may bind to the protein but decapping of one RNA does not affect decapping of a second RNA.

RNA binding analysis using EMSAs

The study of RNA binding was continued using electrophoretic mobility shift assays (EMSAs) to analyze the binding of Nudt16 to the RNA. Previously, Nenad Tomasevic had determined two potential *Kd* values for the reaction using the X29 protein purified from *Xenopus* ovary (NT and BP, unpublished). Here, the *Xenopus* protein was over-expressed in bacteria and purifed. RNA titrations were done using the EMSA. In these analyses, the reactions are at equilibrium. Binding of the RNA in the presence of 27.4 nM of protein was measured (see Table 3.3, page 73 for RNA concentrations).

Data were graphed using Scatchard plots of the concentration of bound RNA versus the ratio of bound over free RNAs (see Figure 3.3). Figure 3.3a shows that at the lower concentrations of bound RNA (below 20nM) there is a curve that then plateaus. When only concentrations lower than 20 nM are included in the figure (see Figure 3.3b), the curve is more visible.

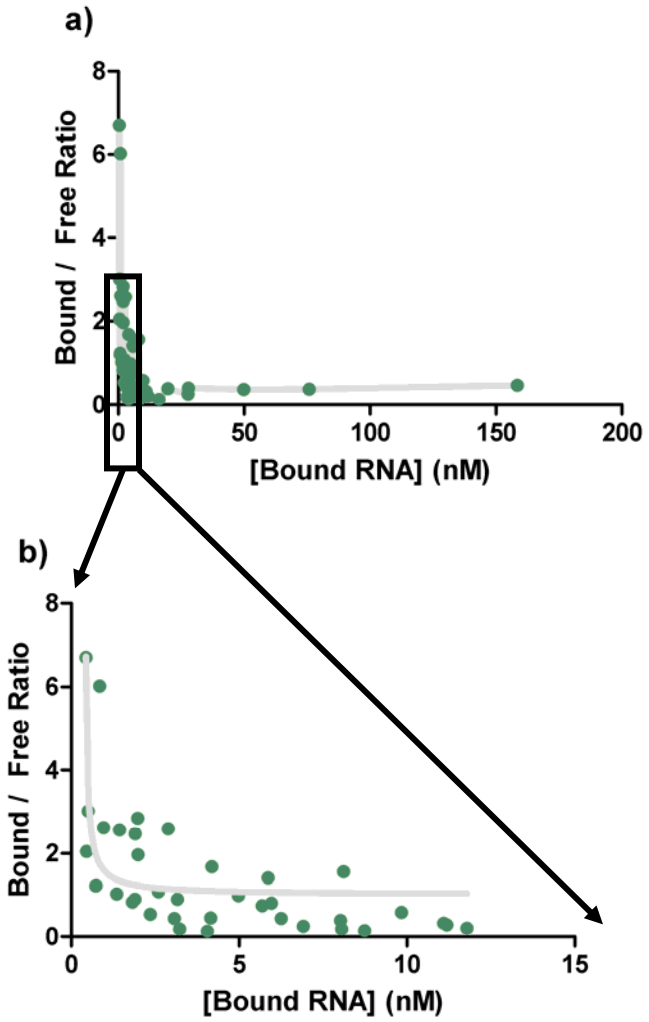


Figure 3.3: EMSA data of RNA titrations show two distinct binding sites. a) A Scatchard plot of the [Bound] RNA vs the [Bound]/[Free] yields a hyperbola which can be used to determine two binding sites for the enzyme, each with different values for the K_d of RNA binding. b) Box a is expanded in b and shows a Scatchard plot of the lower concentrations of RNA.

These data add complexity to the mechanism showing that there are two RNA binding sites on the homodimer of protein. If only one binding site was present, the Scatchard plots would be straight lines. Instead, these data are a curve indicating two binding sites with two different RNA binding affinities.

Cap-Release Experiments

The protein titrations, RNA titrations, and the EMSA experiments collectively support a mechanism in which the Nudt16 homodimer functions as a single unit with two potential RNA binding sites. Only one of those binding sites is able to decap RNA and no turn over is displayed (Peculis et al., 2007). The lack of catalytic turn over could be due to inhibition of the reaction by m^7Gpp or by an inability to release the RNA. The next experiments described test whether cap release or RNA release prevents the next round of catalysis.

Cap release was tested first. In this reaction, four different decapping reactions were made as described in the Methods. Tube 1 was the RNA only control, no protein was included in the reaction. Tube 2 contained both protein and RNA but also included EDTA to terminate the reaction. Tubes 3 and 4 contained both protein and RNA. All tubes were incubated at 37°C for 30 minutes until all reactions were completed. Tubes 1-3 were stopped with EDTA; tube 4 did not receive any EDTA. The resulting samples were then loaded onto a Microcon filter and spun which would allow only molecules with molecular weights less than 10 kDa to pass through the filter. The protein monomer is 25 kDa while the RNA is 46 kDa so neither would be eluted. Eluate was

collected and the column was washed with a Tris-based buffer. Eluates and filters were counted by Cerenkov counting to follow the P^{32} present in the cap of the RNA.

Table 3.7 shows that in the absence of protein, RNA primarily binds to the membrane with <40% of the counts in the eluate. In tube 2, when EDTA was added to the decapping reaction before the incubation, the result is the same as if no protein were added to the reaction. A complete decapping reaction regardless of whether EDTA was used to quench the reaction, as in tubes 3 and 4, demonstrated that >90% of the counts eluted through the filter. These data are consistent with a scenario where the cap is released from the protein after decapping. Thus the inability of the protein to turn over is not due to failure of the cap to be released. It could be due to RNA failing to release or inability of the protein to reset itself after the reaction.

Table 3.7: The cap does not inhibit turnover of Nudt16. The percent of cpms retained on the filters and in the flow through was measured by Cerenkov counting. The reactions are as described in the text and methods.

Tube #	Reaction Conditions	Membrane	Flow-Through
1	RNA Only	69%	31%
2	EDTA in Decapping Reaction	64%	36%
3	Complete Decapping Reaction Terminated with EDTA	7%	93%
4	Complete Decapping Reaction Not Terminated with EDTA	5%	95%

Titrations of tRNA competitor

The mechanism now indicates a functional homodimer with two RNA binding sites that is able to release the cap but still cannot turn over. To find out how competitors fit into the mechanism, alterations in decapping activity in the presence of a non-specific inhibitor (tRNA) was studied. The tRNA was titrated into decapping reactions containing 5 nM U8 substrate RNA (see Table 3.4, page 75, for ratios of tRNA to cap-labeled U8). These U8 substrate tRNA titrations were performed with three concentrations of Nudt16.

The data in Figure 3.4 indicate an increase of inhibition (a decrease in the rate of decapping) in the presence of increasing amounts of tRNA. Consistent with the protein titrations, if the protein monomer:RNA ratio is 1:1 (as it is in the 4.93 nM Nudt16 concentration), no decapping occurs as Figure 3.4a demonstrates.[7] At higher protein:RNA ratios, decapping can occur but it is competed by increasing amounts of tRNA. By 10 nM tRNA (2 molar fold excess) over cap-labeled substrate RNA, the decapping activity dropped below 50% of the original activity at all protein concentrations.

[7] The tRNA titrations include a reaction with no tRNA as well as the concentrations of tRNA given in Table 3.4 (page 73). When equal molar ratios of RNA to protein were used, this reaction had no decapping activity.

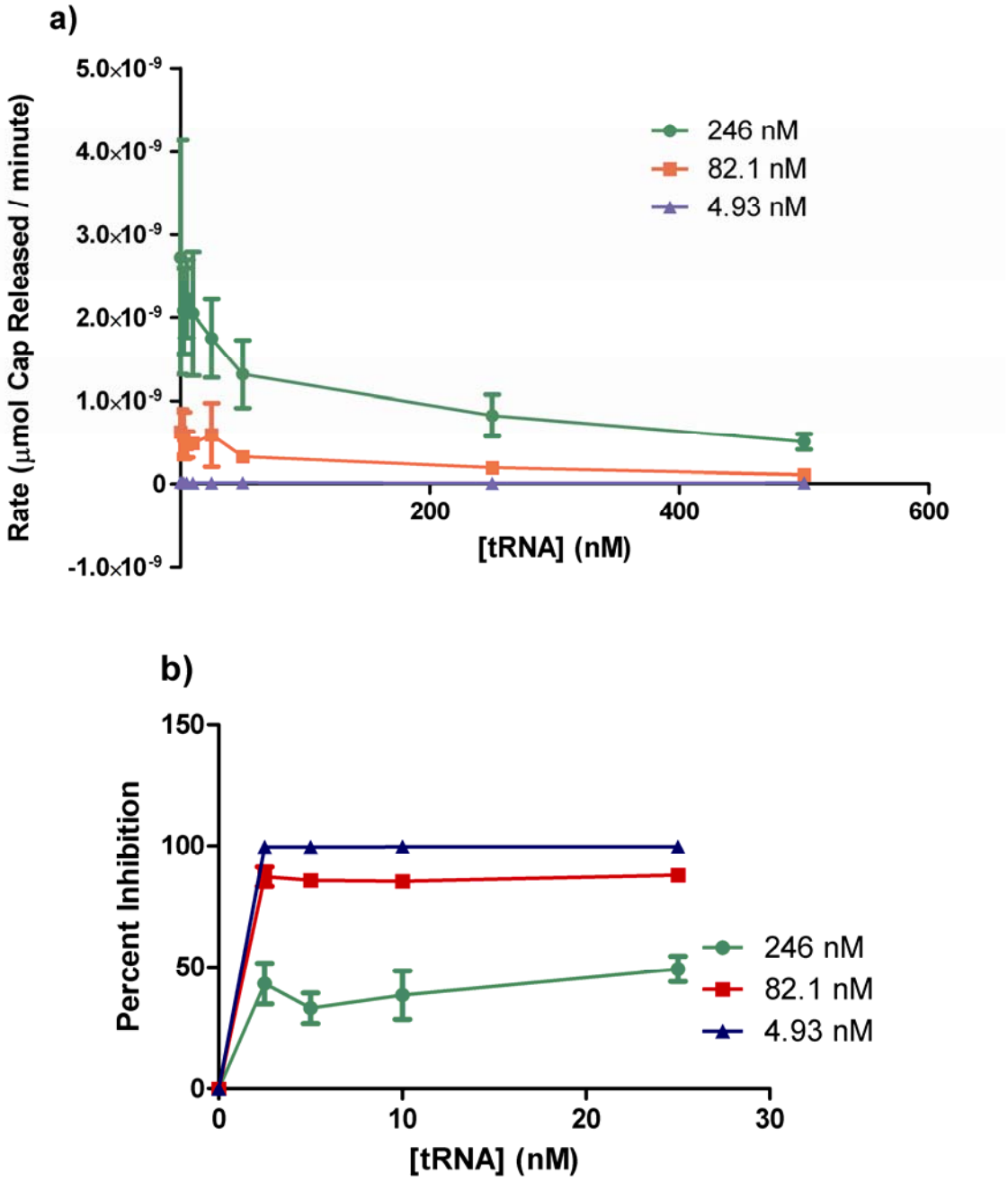
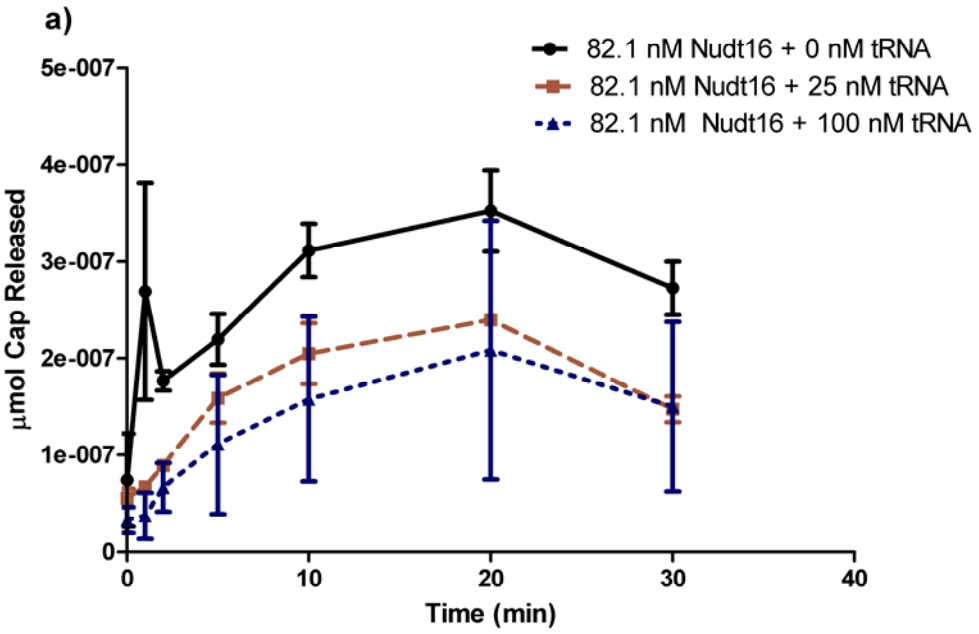


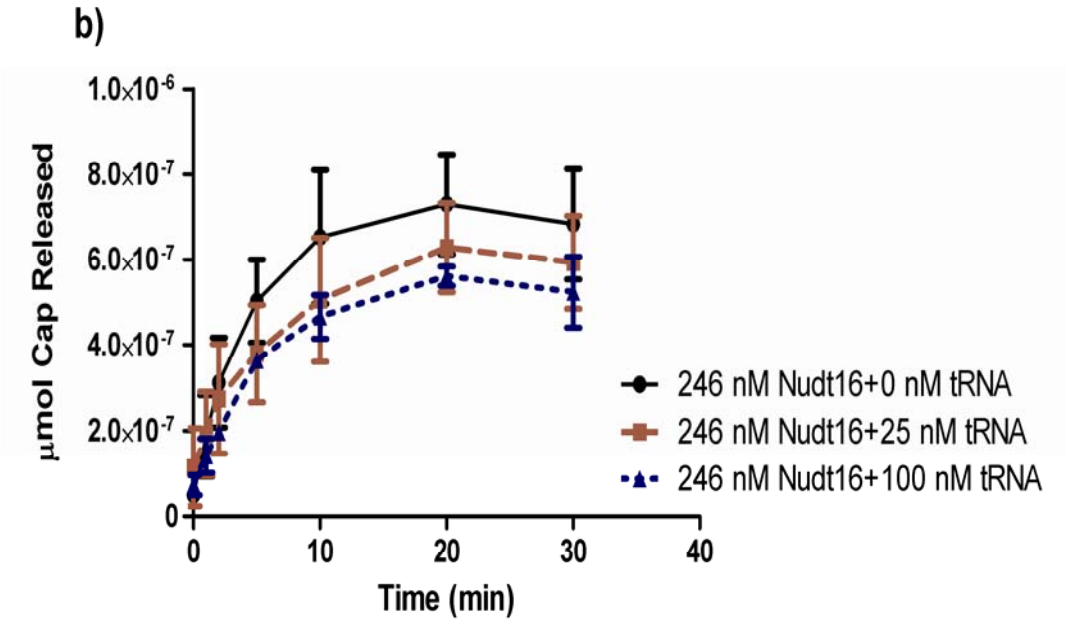
Figure 3.4: tRNA inhibits Nudt16 function. a) The rate of decapping of 5 nM substrate U8 RNA by three concentrations of Nudt16 is shown in this figure. Plots show the decrease in decapping activity when cold tRNA competitor is titrated into the reaction. b) The percent inhibition by tRNA for each of the concentrations of protein is shown. In order to see the effects of increasing the amount of tRNA more closely, the highest competitor concentrations were not included.

The previous experiments demonstrated that tRNA did inhibit the reaction, but did not describe the mechanism of that inhibition. There are two ways to determine the mechanism of inhibition; both were used in the following experiments.

The first method to determine the mechanism of inhibition is to use a time course of tRNA inhibition. A hyperbola will be formed no matter which mechanism of inhibition is used but the differences will lie in the plateau value for the curve and the slope of the linear part of the curve. If the slope changes but the plateau value does not change, the inhibition is uncompetitive inhibition. If the slope changes and the plateau value changes, the demonstrated inhibition is competitive inhibition. If the slope does not change but the plateau value changes, the result is noncompetitive inhibition.

To determine the mechanism for tRNA inhibition using the first method, a time course was performed in the presence of various levels of tRNA shown in Figure 3.5a-c. The plateau values decreased in response to the addition of tRNA. The slope of the linear part of the curve, in the first 5 minutes, did not change significantly as shown in Figure 3.5d. Based on the preceding explanation, these data indicate the inhibition exhibited by tRNA is an example of noncompetitive inhibition. This supports the previous mechanism which included two RNA binding sites.


a)
82.1 nM Nudt16 + 0 nM tRNA
82.1 nM Nudt16 + 25 nM tRNA
82.1 nM Nudt16 + 100 nM tRNA
µmol Cap Released
5e-007
4e-007
3e-007
2e-007
1e-007
0
0
10
20
30
40
Time (min)


b)
µmol Cap Released
1.0x10-6
8.0x10-7
6.0x10-7
4.0x10-7
2.0x10-7
0
0
10
20
30
40
Time (min)
246 nM Nudt16+0 nM tRNA
246 nM Nudt16+25 nM tRNA
246 nM Nudt16+100 nM tRNA

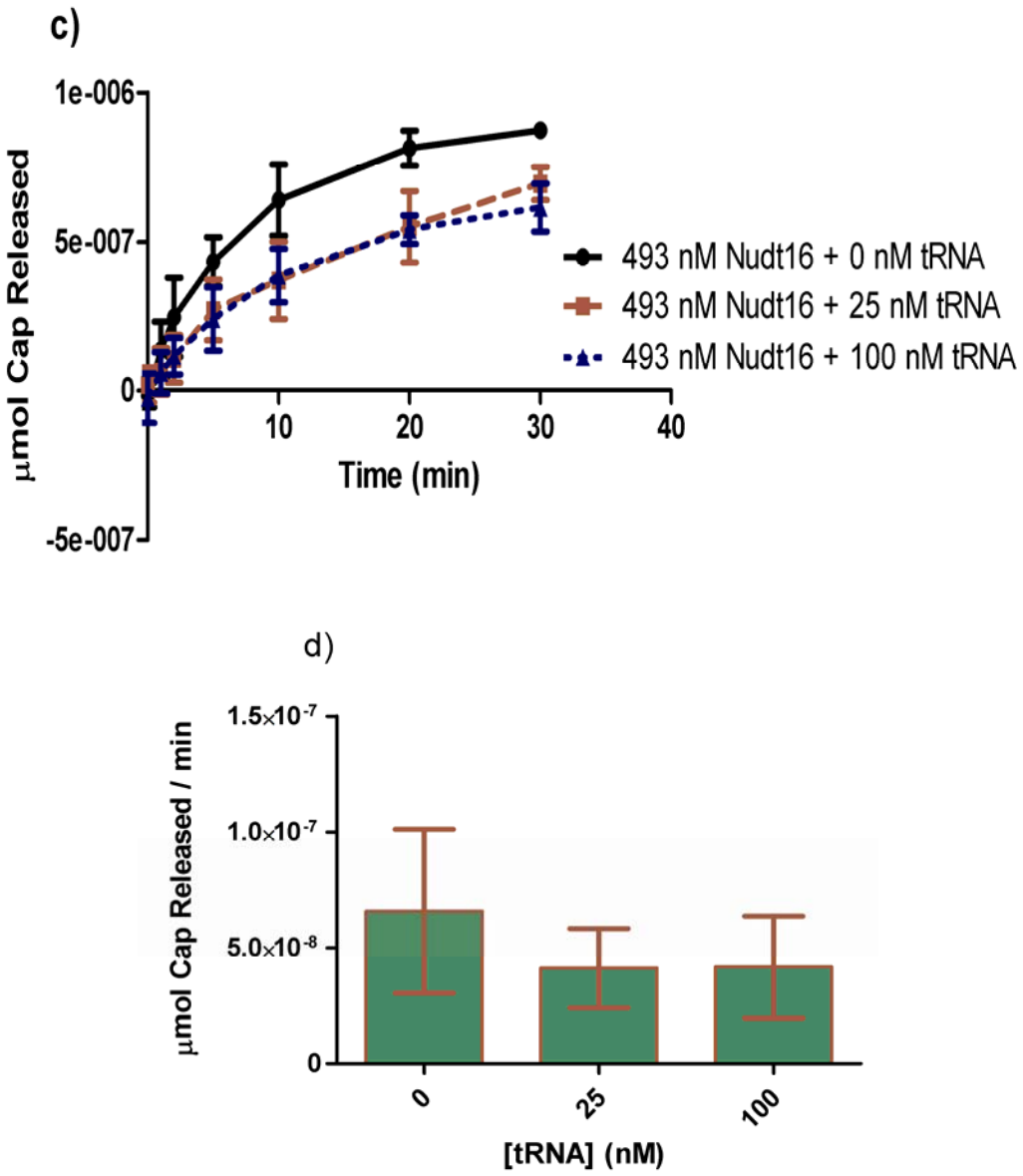


Figure 3.5: Addition of tRNA to time course of Nudt16 activity. a) In the presence of 82.1 nM Nudt16, addition of tRNA drops the total µmol of cap released by 50% even with the lower concentration of tRNA. Addition of more tRNA does not change the result significantly. b) With Nudt16 at 246 nM, the drop in activity is less significant at the lower tRNA concentration but definitely apparent at the higher tRNA concentration. c) The highest concentration of Nudt16 (493 nM) was inhibited by tRNA, but the inhibition took longer to be visible and perhaps with longer time points would seem insignificant. d) Using only data from the time points within the initial rates (the first 5 minutes), a linear relationship is seen whose slope does not change upon addition of tRNA. This indicates non competitive inhibition.

The second method to determine the mechanism of inhibition is by titrating the substrate in the presence of tRNA. In these experiments, the substrate RNA was titrated in the presence of four different concentrations of a constant amount of tRNA. Curves were fitted with the mixed model and noncompetitive models of inhibition, as described in Section 3.2 (see Figure 3.6 and Table 3.8). Based on the α value obtained in the mixed-model of inhibition curve fitting, the reaction could be either non or uncompetitive. The data in Figure 3.5 demonstrated noncompetitive inhibition so the model for noncompetitive inhibition was used to fit the data. The values obtained for each constant was similar to that used with the mixed model of inhibition. Comparison of the fit of the curve with the noncompetitive model of inhibition to the models for uncompetitive and competitive showed the noncompetitive model fit the data better (data not shown). The maximimum velocity of the reaction based on the noncompetitive model was 1.7×10^{-8} µmol of cap released per minute (18 fmol of cap released per minute). The $K_{app\ inhib}$ is 276.7 nM and the $K_{app\ rxn}$ increased to 25.2 nM. The $k_{obs\ rxn}$ increased to 0.05 min^{-1} (3 s^{-1}) corresponding to a time for RNA release of 0.3 s (20 min) which could possibly be observed in assays with longer time points. Assays with longer time points have not demonstrated turnover (data not shown); inhibition is still occurring.

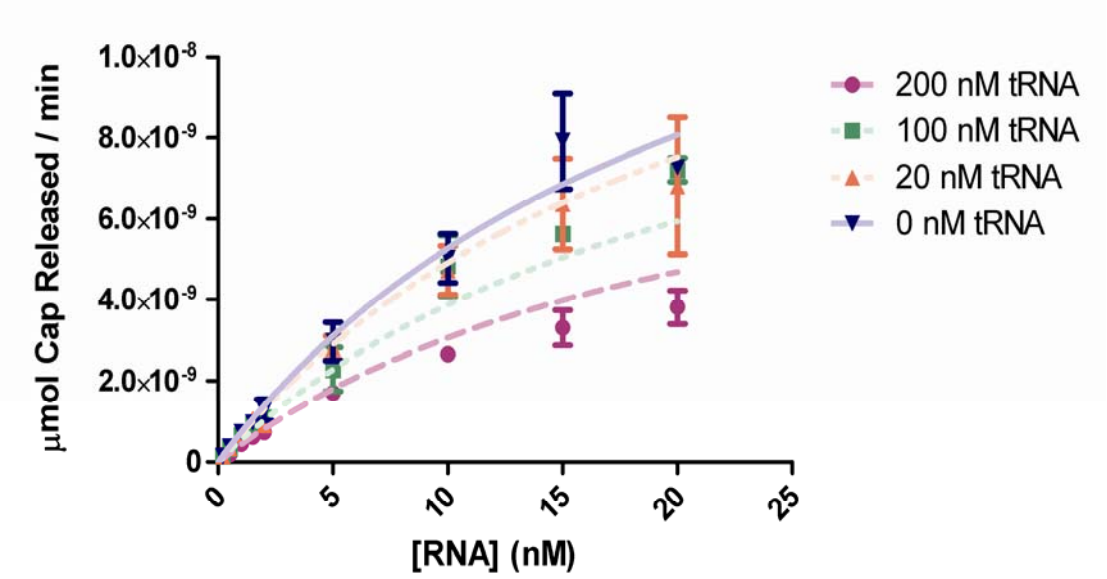


Figure 3.6: RNA titrations in the presence of tRNA affectNudt16 decapping activity. RNA titrations in the presence of 4 concentrations of tRNA were completed using the same concentrations of enzyme as described in Section 3.3. The concentrations of the substrate RNA are provided in Table 3.2 (page 72). The plotted lines are the curve fit of the noncompetitive inhibition equation to each of these curves.

Table 3.8: Titration of RNA substrate in the presence of tRNA. The maximum velocity and binding affinity of tRNA ($K_{app\ inhib}$) are provided in this table. The maximum velocity units are µmol cap released / min. The concentrations of substrate RNA at half maximal velocity, as well as the $K_{app\ inhib}$ have units of nanomolar. The $K_{app\ rxn}$, the concentration of substrate at half the maximum velocity, increases with the addition of tRNA. The $k_{obs\ rxn}$ is in inverse minutes. The specificity constant changes little in the presence of tRNA with units in $min^{-1}\ M^{-1}$.

Parameter	Mixed-Model	Non-Competitive Model
Maximal Velocity (µmol Cap Released / min)	$1.8x10^{-8} \pm 4.0x10^{-9}$	$1.7x10^{-8} \pm 2.6x10^{-9}$
alpha	0.5 ± 0.7	-
$K_{app\ inhib}$ (nM)	380 ± 310	208 ± 46
$K_{app\ rxn}$ (nM)	25 ± 8	23 ± 5
$k_{obs\ rxn}$ (min^{-1})	0.06 ± 0.01	$0.05 \pm 8.0x10^{-3}$
Specificity constant ($min^{-1}\ M^{-1}$)	$2.4x10^{+6} \pm 1.2x10^{+6}$	$2.2x10^{+6} \pm 1.5x10^{+6}$

Together these tRNA titration data support the previous mechanism with two RNA binding sites. One RNA binding site is specific for U8 (or any other substrate RNA). The second site is likely a site of nonspecific binding; tRNA and other competitor RNAs can bind to the second site but are not cleaved.

3.5: Discussion and conclusions from the analysis of Nudt16 function

The goal of these experiments was to understand the mechanism of Nudt16 function. The hypothesis was that by understanding the mechanism of function, other orthologs could then be studied using similar experiments to show they function similarly. Decapping assays, as well as EMSAs, were used to determine a mechanism for Nudt16 action, which will be described below. Before describing the mechanism, a summary of the questions guiding each experiment and brief discussion of how the answers contributed to the proposed mechanism is included.

The first question was whether the functional enzymatic unit was the homodimer or the monomer. To answer this question, protein titrations in the presence of 5 nM substrate U8 RNA were completed. Curve fitting of the data shown in Figure 3.1 resulted in the parameters shown in Table 3.5. Analysis of the data provided convincing evidence for the homodimer as the functional unit.

After studying the functional enzymatic unit, the second experiment was designed to address RNA binding to the protein. RNA titrations in the presence of 329 nM Nudt16 protein supported a mechanism with homodimer function (Figure 3.2 and Table 3.6). In addition, the data provided evidence of a system where multiple RNAs could bind. Binding of one RNA did not affect the binding of a second RNA according to these data.

Electrophoretic mobility shift assays studied RNA binding to the protein in more detail. The results showed two possible RNA binding sites (see Figure 3.3). This supported previously unpublished results in the Peculis laboratoryusing protein purified from *Xenopus* ovary (NT and BP, unpublished).

From this point, the mechanism is two Nudt16 monomers form a homodimer. At least two RNAs can bind; binding of one RNA does not influence binding of the second RNA. The

decapping reaction occurs, but no turn over is apparent. In order to investigate this more fully, cap release was studied using decapping and microdialysis experiments. The results, shown in Table 3.7, reveal that the cap is released, which suggests that the lack of catalytic turn over is not because of lack of cap release.

The final experiments examined the inhibition of Nudt16 by the nonspecific competitor tRNA. Figure 3.4 proved that Nudt16 was inhibited by tRNA by 10 nM tRNA (2 MFE over substrate RNA). The data in Figures 3.5 and 3.6 revealed that inhibition of Nudt16 by tRNA was due to noncompetitive inhibition. This supported the EMSA results in which two RNA binding sites were proposed. The parameters given in Table 3.8 show changes in the $K_{app\ rxn}$ in the presence of tRNA and in the $k_{obs\ rxn}$ but those changes are not reflected in the specificity constant.

The initial model for Nudt16 function included a two-site binding model with the RNA binding to a homodimer of the Nudt16 protein. It was also known that the protein was not active for a second round of catalysis based on "turn over" assays (Peculis et al., 2007). The details of the mechanism and the kinetics of the reaction had not been established. The results presented in the preceding sections can be used to add to the old model of Nudt16 function. Recall the simplistic reaction scheme in Equation 1:

$$m^7GpppG\text{-RNA} + \text{Metal} + \text{Nudt16} \rightarrow m^7Gpp + pG\text{-RNA} + \text{Metal} + \text{Nudt16} \quad (1)$$

The number of molecules involved was not addressed. It was known that the Nudt16 protein could homodimerize. Combining all the new data presented in Figures 3.1-3.2 and Table 3.5, the first step in the new reaction scheme is binding of the two Nudt16 monomers:

$$2\ \text{Nudt16} \underset{k-1}{\overset{k1}{\rightleftharpoons}} \text{Nudt16}_2 \quad (8)$$

The next is RNA binding to the Nudt16 homodimer. In some cross-linking and EMSA experiments (Ghosh et al., 2004; Peculis et al., 2007; Tomasevic and Peculis, 1999), it was shown that binding of the RNA to the protein was both cap and metal independent (data not shown) therefore metal is not required at this step. The reaction mechanism proceeds:

$$m^7GpppG\text{-}RNA + Nudt16_2 \underset{k-2}{\overset{k2}{\rightleftharpoons}} m^7GpppG\text{-}RNA{:}Nudt16_2 \quad (9)$$

The gel shift assays and experiments with nonspecific competitor tRNA support a model with two-binding sites (see Figures 3.3-3.6). The RNA substrate binds in the catalytically active site. The second site is not catalytically active, but RNAs can bind there. Binding of one RNA does not affect binding of a second RNA.

From other gel shift evidence involving competition assays (data not shown), $k_2 \approx k_{-2}$; the RNA can bind and dissociate from the protein at approximately equal rates when no metal is present. In the proposed reaction scheme, once metal is added to the reaction, catalysis occurs. Previously, Scarsdale *et al.* used the Hill equation to show that three metal ions are required to bind the phosphates, correctly positioning the cap in the active site (2006). The next steps of the reaction would then be:

$$m^7GpppG\text{-}RNA{:}Nudt16_2 + 3\ Metal \xrightarrow{k3} m^7GpppG\text{-}RNA{:}Nudt16_2{:}Metal_3 \quad (10)$$

$$m^7GpppG\text{-}RNA{:}Nudt16_2{:}Metal_3 \xrightarrow{k4} m^7Gpp + pG\text{-}RNA{:}Nudt16_2{:}Metal_3 \quad (11)$$

The $m^7GpppG\text{-}RNA{:}Nudt16_2$ complex binds metal in an essentially irreversible reaction. Catalysis occurs resulting in the removal of m^7Gpp from the RNA. From the time course data given in Chapter 4 of this document, it is known that some decapping has taken place 30 seconds into the reaction (see Figure 4.4, page 123). The time for product accumulation suggests that the rate of these two reactions is relatively quick and cannot be distinguished using the standard decapping assay.

As shown in Equation 11, m^7Gpp is released. This was studied in the cap microdialysis experiment (see Table 3.6). The next steps are dissociation of the metal and RNA from the protein. Assuming that the order of dissociation is the same as the order of association the reaction scheme is shown below:

$$m^7Gpp + pG\text{-}RNA:Nudt16_2:Metal_3 \Rightarrow m^7Gpp + pG\text{-}RNA + Nudt16_2 + 3\ Metal \quad (12)$$

At this time, no data exists addressing the order of release of the metal, the cap, and the RNA. Experiments to address this are well beyond the scope of this work.

Attempts to study RNA release have been technically problematic; the experiments have not been able to follow both the cap released and the RNA molecule. It is known that the protein cannot catalyze a second reaction; one possibility is the RNA is not released. The second possibility is the enzyme cannot reset itself after the first catalysis. Two other NUDIX proteins (MutM and MutY) are both known to linger with their products or need modification in order to release the product based on *in vitro* experiments (Koval et al., 2004; Porello et al., 1998). It is likely that Nudt16 requires another factor for product release and turnover *in vivo*.

In summary this more detailed mechanism includes a homodimer as the functional enzymatic unit to which at least one and maybe two RNAs can bind. Although metal does not affect binding it is required for catalysis. When metal is included one RNA is decapped and because of the lack of catalytic turn over of the protein, the other RNA cannot be decapped. For this reason, even with two RNA binding sites on the protein, there is half-site reactivity. With this more detailed mechanism of function to study, comparison of the function of other Nudt16 proteins to the *Xenopus* Nudt16 mechanism will be able determine which parts of the mechanism of Nudt16 function is conserved in other organisms.

3.6: Future projects and questions to consider

Though many answers have been obtained from this analysis many questions still remain. One question involves the K_d of homodimer dissociation. From the decapping assays, it was possible to indirectly estimate a K_d of homodimer dissociation of 6 nM. However, it would be beneficial to use a direct method to confirm this estimate using a standard technique such as analytical centrifugation or sedimentation equilibrium.

Another project is repeating RNA titrations in the presence of specific competitors. Cold U8 is known to inhibit the Nudt16 protein much better than tRNA. However, it is not known whether U8 is a competitive or noncompetitive inhibitor. The answer to the question could add to the mechanism showing whether U8 RNA binds in one site or two sites. If the inhibition is competitive, it suggests a scenario in which U8 competes for one binding site, probably the catalytic site. If the inhibition is noncompetitive or uncompetitive, that would indicate the U8 can bind both sites. Repeating these with other RNAs such as U3 or 5S will determine whether they bind the same sites as U8.

Knowing that there are two binding sites, the RNA binding mechanism becomes of more interest. There are a couple of possible RNA binding sites visible on the crystal structure of Nudt16. At most, only a few nucleotides would fit in the site but this would be enough if the secondary structure of the substrate RNA is linear after the cap. It is also known that Nudt16 will decap a variety of different RNAs (see Chapter 4, pages 125-126). Where are the RNAs binding? How does Nudt16 recognize those RNAs or differentiate between them? Why does it recognize a particular RNA over another?

How do the RNAs interact with the protein? These questions are more difficult to answer.

To answer where the RNAs bind, it is possible to use an RNAse protection assay. Once the samples have been treated with RNAse and resolved on a sequencing gel, one could see the nucleotides that were protected from the RNAse; those nucleotides probably bind Nudt16. Using more than one RNA substrate would be necessary. Then a study of the secondary structures of those RNA substrates would determine the structure of each RNA at the protected location. This could then be compared to other known RNA secondary structures. Perhaps a common pattern will emerge that can be used to define RNAs that bind to Nudt16.

Comparison of the function of other Nudt16 proteins to the *Xenopus* Nudt16 mechanism will be able determine which parts of the mechanism of Nudt16 function is conserved in other organisms. These experiments are the basis of the next chapter.

Chapter 4: Comparison of Various Nudt16 Proteins

4.1: Introduction to the project

After identifying the multitude of putative Nudt16 homologs (Taylor and Peculis, 2008), interest grew in knowing whether these were functional orthologs or just sequence "look–a-likes". The following chapter describes the characterization of three homologs and comparison with the previously mentioned paralog: the human syndesmos protein (see Chapter 2, pages 24-60). The data from these proteins are compared to the *Xenopus* Nudt16 ortholog analyzed and discussed in detail in Chapter 3 (pages 62-100).

Nudt16 was identified in *Xenopus* and mammals (Ghosh et al., 2004), thus it was known to be a vertebrate gene. The evolutionary distribution of Nudt16 was demonstrated to be much wider with the discovery of three insect Nudt16 sequences, the glassy-winged sharpshooter, peach aphid, and pea aphid, which grew into multiple invertebrate sequences as described in Chapter 2 (Taylor and Peculis, 2008). Though the sequences clearly demonstrated these were part of the same gene family, the question of conservation of function was as yet unanswered. Do other putative Nudt16 genes encode functional decapping proteins or just share similar sequences? The results with the insect Nudt16 ortholog from the glassy-winged sharpshooter

demonstrated the ability of a distantly related ortholog to decap RNA (Taylor and Peculis, 2008); the hydrolysis activity was conserved in invertebrates. Though the function was conserved, it was apparent that the insect ortholog was different from the *Xenopus* protein as the insect protein required a higher concentration of protein for decapping activity. RNA binding properties between orthologs and the *Xenopus* protein needed to be examined. Investigation of other Nudt16 homologs[8] would determine whether those homologs functioned like the *Xenopus* ortholog.

Five Nudt16 orthologs were chosen for further analysis: rat, skate, zebra fish, sharpshooter, and crab. Each was chosen for a specific reason. The rat cDNA sequence was the same length as the *Xenopus* cDNA, but was much closer in terms of sequence conservation and evolutionary relationship to the human Nudt16 gene, previously characterized by the Peculis laboratory (Peculis et al., 2007). Secondly, the rat could be a good mammalian model system with which to study Nudt16 depletion *in vivo*.

Another potential model system in which to study the effects of Nudt16 could be the zebra fish. The zebra fish has been well studied as a model of vertebrate development. In addition, the zebra fish is a bony fish lying closer to the most recent common ancestor than either the rat, human, or *Xenopus*. Even closer to the most recent common ancestor than the bony fish are the cartilaginous fish. The little skate was chosen to represent this class and a full-length cDNA clone was readily obtainable.

Finally, two invertebrates were chosen for analysis: crab and sharpshooter. The insect, the glassy-winged sharpshooter, was shown to decap RNA in Mn^{2+} and to bind to

[8] The use of the word "homologs" here instead of "orthologs" is because the syndesmos protein, a paralog, is also discussed in the following chapter.

RNA based on RNA:protein crosslinking assays (Taylor and Peculis, 2008). No further details about the mechanism of the function were included. A more thorough analysis is detailed here and provides a more complete description of the mechanism of the function of the sharpshooter protein. The crab was included in order to have another uncharacterized invertebrate in the study.

cDNA clones for each of these Nudt16 orthologs were obtained, sub-cloned into the protein expression vector, over-expressed in *E. coli*, and purified using a two-column purification method. The proposal tested was that each of these Nudt16 proteins would be a functional Nudt16 ortholog as determined by the ability to decap U8 snoRNA in a metal dependent reaction. In addition to being able to decap U8 snoRNA, all functional orthologs were examined to determine their ability to bind RNA in RNA:protein crosslinking assays. Most proteins were stable and maintained activity for months. The exception was the skate protein which was active for decapping immediately after purification but lost activity within two weeks. The activity of each protein was compared to the results of the *Xenopus* and human Nudt16 proteins which were assayed in parallel (Peculis et al., 2007).

The results will show that all orthologs decap RNA with different efficiencies as determined by the specific activities. Each has a different $K_{app\ rxn}$ and observed rate value ($k_{obs\ rxn}$), and with the exception of the sharpshooter protein, all values were within an order of magnitude. Metals and RNA substrates affected the efficiency of decapping but all homologs could decap all tested substrate RNAs using Mn^{2+}, Mg^{2+}, or Co^{2+}. Significantly, a variety of RNAs were decapped in the presence of Mg^{2+},

demonstrating the *Xenopus* protein may have a narrower RNA substrate specificity in this purified system than other species. Together these data support functional conservation of the Nudt16 gene in a variety of different organisms, across Metazoans including insects and crustaceans. This provides further evidence that the Nudt16 gene is an important gene whose sequence and function are conserved throughout evolution, supporting the existence of the proposed nuclear RNA degradation pathway.

4.2: Experimental protocols for the functional conservation studies

Obtaining cDNA clones for the various Nudt16 orthologs

The cDNAs for the putative Nudt16 proteins were obtained from various sources. A truncated version of the rat (NCBI: XM_343464) was obtained from ATCC (#9110368). The 22 amino acid amino terminal extension was added during cloning using overlapping oligos (see Appendix B, page 197-198, for description of oligos and cloning strategy). While the extended rat protein was larger, it did not over-express or purify as well as the truncated rat protein which had similar activity. The data here describes the extended version of the rat protein.

The zebra fish cDNA (NCBI: BC083482) clone came from ATCC (# 1062166). The crab cDNA (NCBI: CX993972.1) was a generous gift from Dr. David Towle (Mount Desert Island Biological Laboratory, Salsbury Maine). The glassy-winged sharpshooter (NCBI: DN199400.1) clone was a generous gift from Dr. Wayne Hunter (US Horticulture Research Laboratory, Fort Pierce Florida). The little skate cDNA (NCBI: EE992989.1) was a generous gift from Dr. David Barnes (Mount Desert Island Biological Laboratory, Salsbury Maine).

PCR and cloning of Nudt16 clones

Each cDNA was used as template for PCR with oligos containing unique restriction sites to facilitate in-frame cloning into pET19b (see Table B.1, page 198, for a list of oligos for each organism). The resulting PCR product was gel purified using the QIAquick Gel Extraction Kit (Qiagen #28704) and ligated into the pGEM-TEasy vector

system (Promega #A1360) for 16-18 hours. Ligations were transformed in DH5α *E. coli* competent cells and allowed to grow on LB plates top-layered with ampicilin, β-galactosidase and Isopropyl β-D-1-thiogalactopyranoside (IPTG) for 16 hours. DNA was isolated from cultures of white colonies and digested to release insert. Clones with a putative insert were sequenced to make sure the clone was the correct sequence. Clones with the correct sequence were digested with the restriction enzymes shown in Table B.1 (page 198) for 16 hours and gel purified with the QIAquick Gel Extraction Kit. Purified cDNAs were ligated for 16-18 hours into pET19b (Novagen #69677-3) which had been digested with the same restriction enzymes as the insert fragment. Ligations were transformed into DH5α cells and then plated on LB plates with ampicilin. Colonies were randomly chosen; DNA was isolated from cultures of these colonies and sequenced to ensure the orientation and reading frame of the clone was correct.

Protein expression and purification

Correctly oriented Nudt16 clones in pET19b were transformed into BL21 protein expression *E. coli* cells. Colonies were grown in 2 mL LB with ampicilin and chloramphenicol for 6 hours and then transferred to 50 mL LB with ampicilin and chloramphenicol for 16 hours. Large scale (1L) cultures were generated. A 100 mL overnight culture was divided into 3 equal proportions, one 30 mL aliquot was added per 1L of LB broth, grown for 2 hours at 37°C, and induced with 0.6 mM IPTG for 3 hours at room temperature or 4 hours at 37°C. Cells were harvested by centrifugation at 6000 rpm for 7 minutes at 4°C to pellet cells. Pellets were frozen and stored at -80°C.

To isolate protein, frozen cell pellets were resuspended in 20 mL cold lysis buffer (50 mM Tris pH 8.5, 1 M NaCl, 15% glycerol) plus 50 µL protease inhibitor cocktail (Sigma #P8849) and passed through the French press twice to ensure complete lysis. Lysis buffer with no protease inhibitor cocktail was added to a final volume of 50 mL and cell lysate was spun at 8000 rpm for 20 minutes to pellet debris. The induced protein was purified from the cell lysate using a Ni-NTA His (Qiagen) column on the BioLogic DuoFlow 10 System (BioRad). Briefly, the column was equilibrated with 3 column volumes of low imidazole buffer (20 mM Tris, pH 8.5, 20 mM Imidazole, 250 mM NaCl) before supernatant was loaded onto the column. Cell lysate supernatant was loaded on the column at a flow rate of 1.5 mL/ min. The column was washed with 2 column volumes of lysis buffer, followed by 3 column volumes of low imidazole buffer. The final wash was 1 column volume of 99% low imidazole buffer, 1% high imidazole buffer (final concentration of imidazole in the wash was 22.3 mM). Protein was eluted using a gradient of the high imidazole buffer (20 mM Tris pH 8.5, 250 mM imidazole, 250 mM NaCl). Typically the protein eluted at a final concentration of 112 mM imidazole. Fractions were collected with the BioFrac fraction collector (BioRad) at 1 mL intervals.

Eluted fractions with increased absorbance at 280 nm were then analyzed by SDS-PAGE on 10% Bis-Tris NuPAGE gels (Invitrogen #NP0301Box). Fractions containing the protein of interest, as determined by migration of the bands on SDS-PAGE, were combined and diluted with 20 mM Tris pH 8.5 until the concentration of NaCl was below 100 mM. Diluted samples were purified on an ion-exchange column such as Heparin, Q-sepharose, or UnoQ1 using the BioLogic DuoFlow 10 system. Briefly, the column was

equilibrated with 60 mM B1 buffer (20 mM Tris pH 8.1, 60 mM NaCl) for 3 column volumes. The eluted fractions that had been combined and diluted were loaded onto the column at a flow rate of 1.25 mL/min. Washing of the column for 3 column volumes with 60 mM B1 followed sample loading. A second wash of 99% 60 mM B1 and 1% 1M B1 (20 mM Tris pH 8.5, 1 M NaCl) was 2 column volumes (the final concentration of salt in this wash was 60.4 mM). Elution occurred using a gradient of 1M B1; typically the protein eluted at salt concentrations of 624 mM. Fractions were collected and analyzed as with the Ni-NTA column.

Protein titrations

Protein titrations were carried out as described previously in Chapter 3 for the Xenopus ortholog (pages 70-71). The amount of protein used in Chapter 3 for the protein titrations with the Xenopus ortholog was used with all other orthologs (see Table 4.1). U8 snoRNA was used as the substrate at a final concentration of 0.025 pmol (5 nM) RNA; the metal was Mn^{2+} at a final concentration of 300 µM. Other reaction conditions are as described in Chapter 3 (pages 70-71). Cap release was measured by migration of the cap through a PEI thin layer chromatography plate (Sigma #Z122882-25EA) developed in 1 M Formic Acid, 0.5 M LiCl, dried then exposed to phosphor screen for 1-16 hours. Screens were scanned with the Personal Molecular Imager (BioRad #170-9400) and Quantity One software to quantitate the percent decapping which was converted to the µmoles of cap released and used in curve fitting (see Section 4.3, pages 114-134).

Table 4.1: Concentrations of Nudt16 protein in protein titration experiments. Decapping reactions with 0.025 pmol (5 nM) U8 snoRNA and various amounts of protein were incubated 3 minutes at 37°C. Amount of total protein in grams (g) are shown. Concentrations of each Nudt16 ortholog are nM concentrations and assume that all protein in Column #1 is Nudt16 which is clearly an overestimation. Both reactions above and below the concentration of substrate were tested. The other proteins (the rat and crab) were also characterized with similar values and had comparable results (data not shown). The skate was not included in these analyses.

Total Amount of Protein (g)	Zebra fish (nM)	Sharp-shooter (nM)	Human (nM)	Frog (nM)
$6.0x10^{-7}$	5190	4710	5680	4930
$4.0x10^{-7}$	3460	3410	3780	3290
$2.0x10^{-7}$	1730	1570	1890	1640
$1.2x10^{-7}$	1040	943	1140	986
$6.0x10^{-8}$	591	471	568	493
$4.0x10^{-8}$	346	341	378	329
$2.0x10^{-8}$	173	157	189	164
$1.0x10^{-8}$	86.4	78.6	94.6	82.1
$7.5x10^{-9}$	64.8	58.9	70.9	61.6
$5.0x10^{-9}$	43.2	39.3	47.3	41.1
$2.0x10^{-9}$	17.3	15.7	18.9	16.4
$1.0x10^{-9}$	8.64	7.86	9.46	8.21
$6.1x10^{-10}$	5.27	4.71	5.68	5.00
$3.2x10^{-10}$	2.77	2.51	3.03	2.63
$2.0x10^{-10}$	1.73	1.57	1.89	1.64
$1.0x10^{-10}$	0.86	0.79	0.95	0.82

RNA titrations

As with the protein titrations, the concentrations of RNA tested in Chapter 3 were used in the RNA titrations with the other Nudt16 orthologs (see Table 4.2). Detailed methods for this experiment were outlined in Chapter 3 and under the methods for the protein titrations in this chapter (page 108). The metal was Mn^{2+} at a final concentration of 300 µM.

Table 4.2: Concentration of RNA for RNA titration experiments. RNA was added to reactions containing Nudt16 protein (40 ng for the zebra fish, human and frog orthologs; 300 ng for the sharpshooter ortholog) and allowed to decap for 3 minutes. Reactions were quantitated and analyzed as described in the above section and in Chapter 3 (page 70-72). The last four columns are the protein:RNA molar ratios of the organisms in the column heading.

Amount of RNA (pmol)	Concentration (nM)	Zebra:RNA	Human:RNA	Frog:RNA	Sharp:RNA
0.1	20	17:1	19:1	16:1	125:1
0.075	15	23:1	25:1	22:1	166:1
0.05	10	35:1	38:1	33:1	249:1
0.025	5	69:1	76:1	66:1	498:1
0.01	2	173:1	189:1	165:1	1245:1
0.0075	1.5	231:1	252:1	219:1	1660:1
0.005	1	346:1	378:1	329:1	2490:1
0.0025	0.5	692:1	756:1	658:1	4980:1
0.001	0.2	1730:1	1890:1	1645:1	12450:1

Time course decapping assays

In this assay, a time course was examined in order to better compare the efficiency of decapping of each organism. The standard decapping reaction contained 0.025 pmol (5 nM) cap-labeled U8 RNA, 45 mM Tris pH 8.5, and 300 µM Mn^{2+}. To initiate the reactions, 100 ng of protein was added. Reactions were incubated at 37°C for 0, 0.5, 1, 2, 5, 7, 10, 20, or 30 minutes. At each time point, 2 µL of the original reaction were removed and stopped with 0.05 M EDTA. Released cap was analyzed and quantitated as described in the preceding section. Product release over time was graphed in GraphPad Prism 5 and fitted to a single exponential decay model based on the model suggested by Campbell *et al*. (2002)

$$y = B_{max} * (B\text{-}exp^{k_{obs}x})$$

where y is the released product in µmol cap released; *Bmax* is the plateau value where the maximum activity is reached; B is the amplitude for the burst before the plateau[9], and k_{obs} is the observed rate constant of that reaction in µmol of cap released per minute. The specific activity of each protein under these conditions was found by multiplying the plateau value (*Bmax*) by the observed rate constant (k_{obs}), then dividing by the grams or µmol of protein in each reaction.

Metal dependent RNA decapping

Decapping reactions with constant concentrations of protein and RNA were divided into 3 equal portions to which one metal was added to the following final concentrations: 2 mM Mg^{2+}, 300 µM Mn^{2+}, or 300 µM Co^{2+}. Reactions were incubated at 37°C for 30 minutes to ensure complete decapping activity and then stopped with 0.05 M EDTA. Cap was separated from RNA by TLC and analyzed as above using Quantity One. Graphs were made using GraphPad Prism 5 software.

Stability of Nudt16 homologs at 37°C

To test whether the Nudt16 proteins were stable at 37°C, a series of three different decapping reactions were made. One contained protein alone – no metal or RNA was added to the reactions. A second series had both protein and 0.015 pmol (3.3 nM) substrate RNA but no metal. The third set had protein and 300 µM Mn^{2+} but no substrate RNA. These reactions were pre-incubated for 0, 15, or 30 minutes at 37°C. Decapping was initiated by addition of the missing reactants: metal and/or RNA.

[9] It has been argued that the term "B" is unnecessary in this analysis. Instead, the value 1 should be used. This is the ideal case; however, an argument has also been made that the protein has two phases. If that is the case, B would have a value other than 1. To support the argument that there is only 1 phase, the B term was included. Both this equation and the exponential decay equation where B = 1 resulted in the same B_{max} and k_{obs} values.

Reactions were allowed to decap for 15 minutes at 37°C and stopped with EDTA. Standard analysis occurred via TLC and quantitation by Quantity One.

Gel shift of Nudt16 orthologs to assess RNA binding

The protocol for gel shifts has been described previously (Tomasevic and Peculis, 1999). Briefly, RNA was *in vitro* transcribed in the presence of rGTP – αP^{32} which yields uniformly labeled RNA. RNA was incubated with Nudt16 in the presence of buffer, 93 nM tRNA, and glycerol for 15 minutes at room temperature. Reactions were loaded on a 4% polyacrylamide gel and run for 3.5 hours at 12 mA. Gels were dried and exposed to a phosphor screen for 1-16 hours. Screens were scanned and visualized with Quantity One (BioRad) software.

RNA:protein crosslinking assay to assess RNA binding

In vitro transcribed 4-thio-U RNA was made with limiting cold UTP, addition of 4-thio-UTP, and rATP-αP^{32} as described previously (Taylor and Peculis, 2008). For the assay, 0.04 pmol (2.67 nM) 4-thio-U RNA was incubated with 7.1 pmol (355 nM) protein and 0.19 pmol (12.7 nM) tRNA for 15 minutes at room temperature to allow binding. Afterwards, protein:RNA crosslinks were irradiated for 15 minutes with 365 nm UV light at 3 cm distance from samples on ice. In parallel a no UV control was incubated on ice for 15 minutes. Pancreatic RNAseA was added to each reaction and incubated 30 minutes at 37°C. NuPAGE 4x loading dye (Invitrogen #NP0008) + 10% βME was then added and reactions were incubated at 65°C for 5 minutes to denature protein before loading on a 10% Bis-Tris NuPAGE gel (Invitrogen #NP0301Box). Gels were run for 45 minutes at 200 V (225 mA), dried 2 hours, and exposed to phosphor screen for 1-16

hours. Screens were analyzed as with decapping reactions by the personal molecular imager and Quantity One software (BioRad).

Protein:Protein dimerization assay

Proteins were dialyzed against 20 mM Hepes pH 8.1, 200 mM NaCl overnight at 4°C to remove Tris which inhibits the crosslinker. Reactions (50 µL volume) contained 2 µg of protein and crosslinker at a final concentration of 0.5 mM. Two crosslinking agents were used: the hydrolysable crosslinker 3,3´-Dithiobis[sulfosuccinimidylpropionate] (DTSSP) (Pierce #21578) and the non-hydrolysable crosslinker BS^3 (*Bis*[sulfosuccinimidyl] suberate) (Pierce #21580). Reactions were incubated for 30 minutes at room temperature, after which 4x NuPAGE loading dye (Invitrogen) was added to the reaction with or without the addition of 10% βME and proteins were denatured at 65°C for 5 minutes. Samples were then analyzed by running 10% Bis-Tris NuPAGE gels (Invitrogen) at 200 V for 45 minutes, then stained with coomassie (PageBlue Fermentas #R0571) as per manufacturer's directions.

4.3: Results of the functional conservation study

Cloning and purification of Nudt16 homologs

Nudt16 protein homologs were successfully cloned, over-expressed, and purified from *E. coli* cells. Most purified proteins had some to several contaminating bands. When a visible band co-migrated at the point presumed for a homodimer it was assumed to be the homodimer seen in other experiments (see Figure 4.1). The rat and crab had multiple bands, with only 10% of the protein being Nudt16. Some of the assays were done with these batches of purififed proteins, but because of the impurities, not all analyses were completed with them. The skate protein was quite pure but unstable; activity decreased within one week and was completely gone within two weeks although full-length protein was visible on a gel. This ortholog was not further characterized.

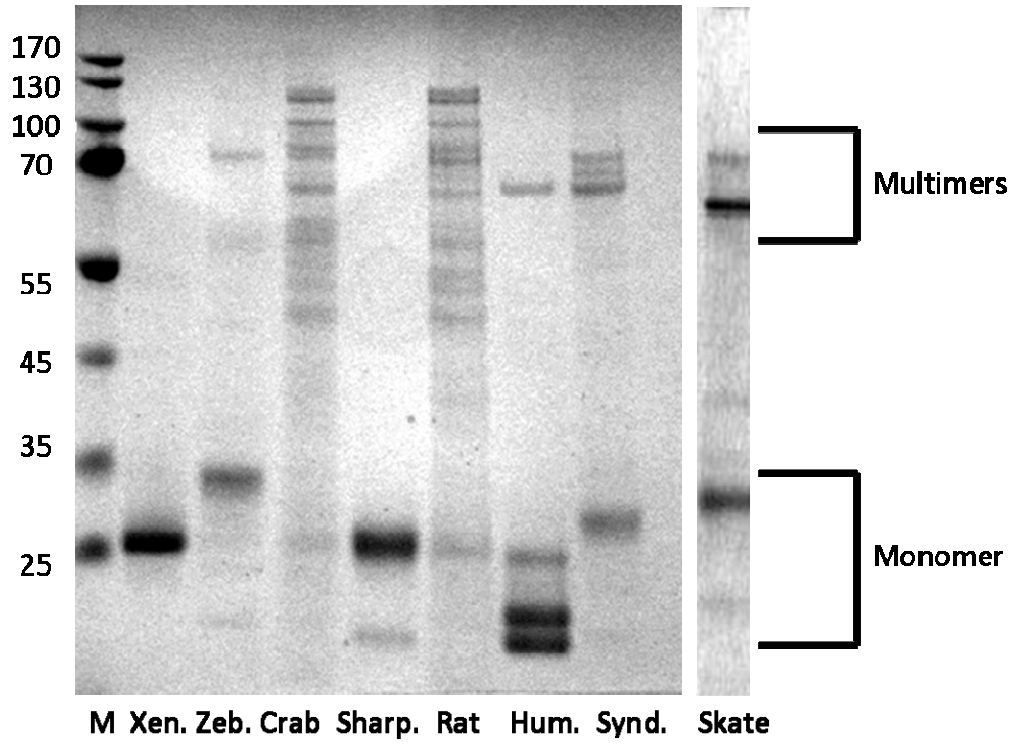


Figure 4.1: Nudt16 orthologs were successfully purified. The protein profiles of the Nudt16 homologs are shown in this figure. Lane 1 is the marker lane; the approximate molecular weight in kDa is given on the left side of lane 1. The other lanes are in the following order: *Xenopus* (Xen.), zebra fish (Zeb.), crab (Crab), sharpshooter (Sharp.), rat (Rat), human Nudt16 (Hum.), human syndesmos (Synd.), and skate (Skate). The Nudt16 band is between the 25 and 35 kDa bands; the exact molecular weight varies between organisms. The homodimer is between 55 and 70 kDa.

To show conservation of Nudt16 function, all Nudt16 orthologs were tested for decapping activity in a standard decapping assay. All proteins were able to decap U8 snoRNA (data not shown); Nudt16 function is conserved in all orthologs.

Protein titrations indicate all Nudt16 orthologs function as homodimers

Having demonstrated that all Nudt16 orthologs are active for decapping, studies to determine whether the mechanism described in Chapter 3 was conserved in these

orthologs were examined. To determine whether the functional enzymatic unit of these orthologs is the homodimer or monomer, protein titrations using the same amount of protein as those used in Chapter 3 (see Table 4.1) were incubated 3 minutes at 37°C. They were analyzed as described in Chapter 3 (page 70-71). These data were graphed using the same program so different colors were used to differentiate between the various organisms as shown in Table 4.3.

Table 4.3: Guide to interpreting subsequent graphs and tables. Each Nudt16 ortholog was coded based on the genus and species of the organism from which it came. In this document, the species are called by their common name; for completeness the genus and species are given here. Each organism has been colored in all subsequent graphs using the same color shown in the table.

Designation	Organism Name	Color of Line/Graph
Crab	*Carcinus maenas*	Black
Zebra fish	*Danio rerio*	Blue
Sharpshooter	*Homalodisca coagulate*	Pink
Human	*Homo sapiens*	Purple
Rat	*Rattus norvegicus*	Gold
Xenopus	*Xenopus laevis*	Green

All organisms were able to decap U8 snoRNA. Data from the protein titrations were hyperbolic curves (see Figure 4.2) which were fitted to the hyperbolic curve equation given in Chapter 3 (page 65-66) to give the parameters shown in Table 4.4. The *Xenopus* Nudt16 had the lowest $K_{app\ rxn}$ suggesting this ortholog had a higher affinity

for the U8 substrate and thus required less protein to reach half-maximal activity. The zebrafish was next, followed by the human and the sharpshooter orthologs. The $k_{obs\ rxn}$ varied as well; the sharpshooter had the lowest rate, followed by the *Xenopus*, zebra fish, and human orthologs.

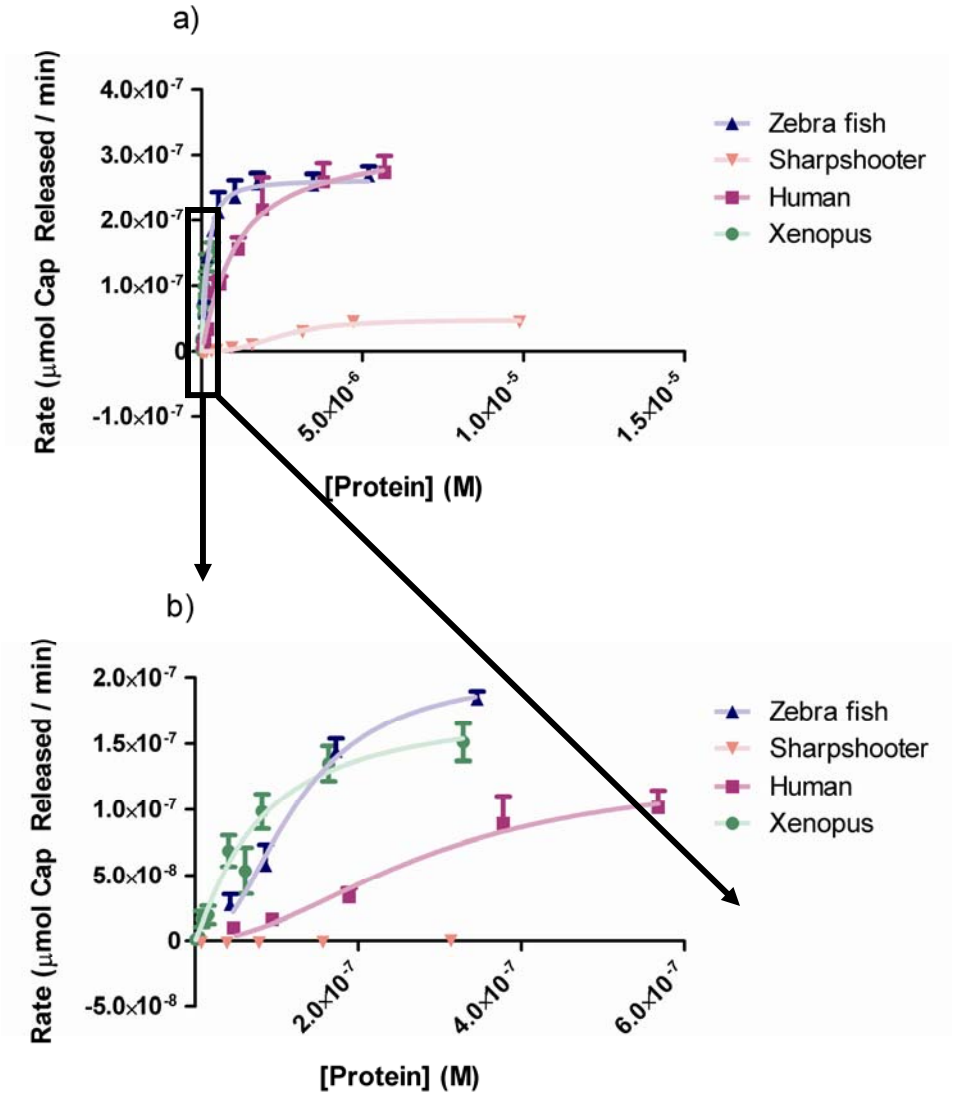


Figure 4.2: Protein titrations of Nudt16 homologs show different levels of decapping. a) Protein titrations with various Nudt16 homolog proteins were fitted to the hyperbolic curve equation to determine whether the protein functioned as a monomer or a homodimer. b) The box in a is expanded in b.

Perhaps the best comparison of all the parameters was the specificity constant which was used to compare the efficiency of protein activity in different proteins (Blikstad et al., 2008; Eisenthal et al., 2007; Taylor et al., 2007). The specificity constant varied depending on the organism; lower specificity constants indicated higher efficiency. The zebra fish Nudt16 ortholog had the lowest specificity constant with the closest ortholog (*Xenopus)* being about half as efficient. The human and sharpshooter were much less efficient being 9 and 197 fold, respectively above the *Xenopus* ortholog.

Table 4.4: Nudt16 parameters from protein titrations fitted to the hyperbolic equation. The results from the curve fitting using the hyperbolic curve equation suggest most Nudt16 orthologs have a similar mechanism to the *Xenopus* ortholog. The parameters obtained by fitting the hyperbolic curve to the equation are shown below. The *Bmax* is in µmol cap released. The $K_{app\ rxn}$, the concentration of protein at which the reaction is half maximal, is in nM. The observed rate of the reaction, k_{obs}, is in min^{-1}. Finally, the specificity constant is in min^{-1} M^{-1}.

Organism	*Bmax*	$K_{app\ rxn}$ (nM)	$k_{obs\ rxn}$ (min^{-1})	Specificity Constant
Zebra fish	$2.8x10^{-7}$ ± $1.2x10^{-8}$	210.5 ± 37.0	56.6 ± 2.4	$2.7x10^{+8}$ ± $6.4x10^{+7}$
Sharpshooter	$8.3x10^{-8}$ ± $1.8x10^{-8}$	6706 ± 2.540	16.6 ± 3.5	$2.5x10^{+6}$ ± $1.4x10^{+6}$
Human	$3.4x10^{-7}$ ± $3.1x10^{-8}$	1273 ± 310	68.7 ± 6.2	$5.4x10^{+7}$ ± $2.0x10^{+7}$
Xenopus	$1.7x10^{-7}$ ± $6.7x10^{-9}$	71.2 ± 11.2	34.7 ± 1.3	$4.9x10^{+8}$ ± $1.2x10^{+8}$

Further analysis confirmed that all Nudt16 orthologs functioned as a homodimer; high amounts of decapping activity only occurred after the proteins had reached a concentration of 200 nM (see Figure 4.2b). At this concentration the protein

was in the homodimer form; some monomeric protein may also exist but as the concentration of protein increased the monomer form is converted to the homodimer. The sharpshooter ortholog demonstrated the need for a homodimer as it required 1 μM of protein to function (see Figure 4.2a).

RNA titration results of Nudt16 orthologs are similar to the *Xenopus* results

Based on the protein titrations, the Nudt16 orthologs form homodimers; the model of function is conserved to this point. The second question is whether the RNA binding is conserved. To study this question, RNA titrations using the same concentrations of RNA and the same amount of protein (based on the number of grams of protein) were done. The data are plotted in Figure 4.3 and summarized in Table 4.5. The resulting data indicated that once the protein:RNA ratio dropped below 16 molar fold excess, the decapping activity also dropped (data not shown). This result is the same as the decrease in decapping activity exhibited by the *Xenopus* ortholog at the 15 and 20 nM RNA concentrations (see Figure 3.2a, page 81). As with the Xenoups ortholog, in order to curve fit the data, only concentrations of substrate RNA below 10 nM were used (see Figure 4.3 and Table 4.5).

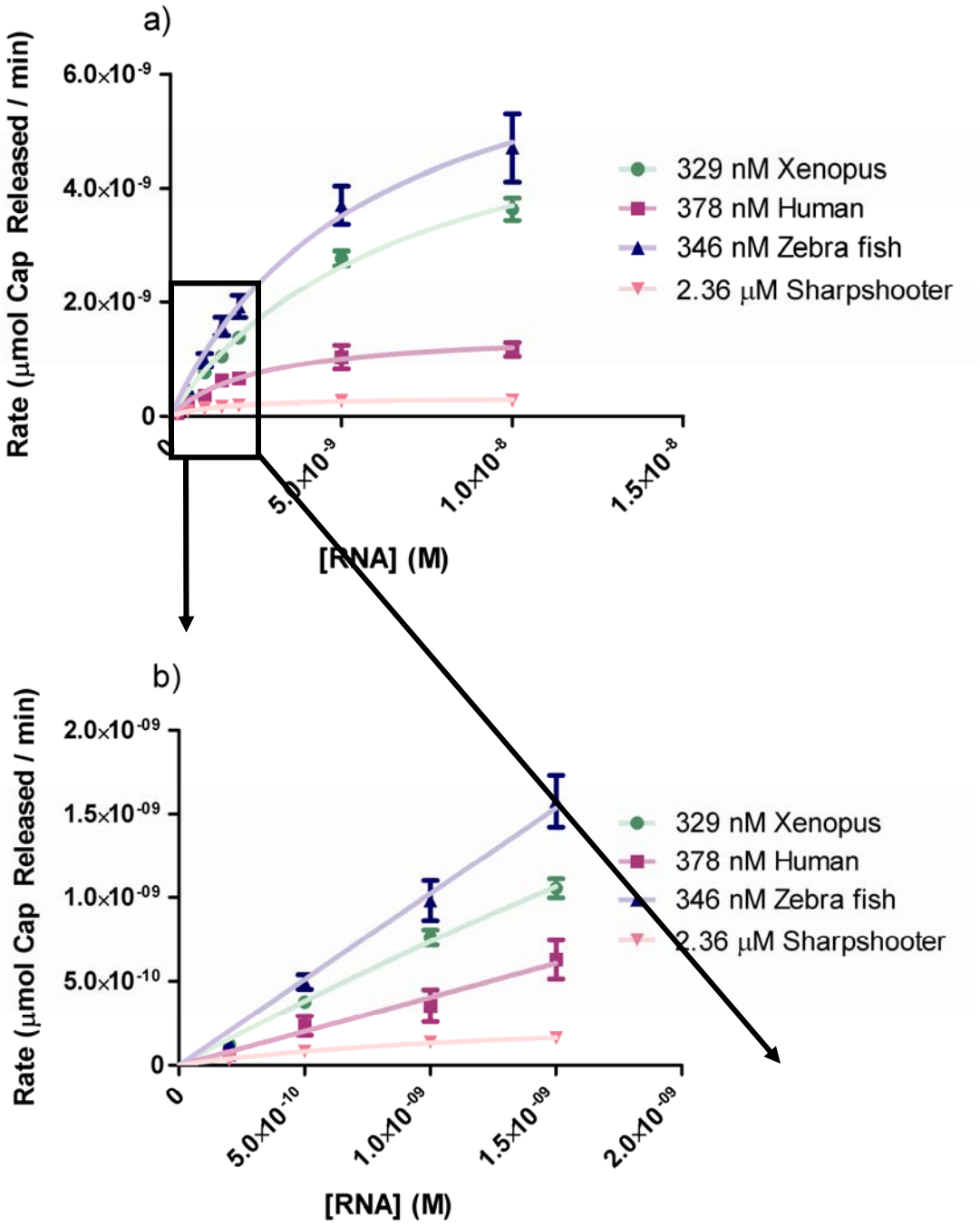


Figure 4.3: RNA titrations of Nudt16 orthologs show the mechanism of RNA binding is also conserved. a) RNA titration data were fitted to a hyperbolic curve equation and the equation for cooperativity as described in Chapter 3 (page 66). The resulting parameters are located in Table 4.5. b) The box in a is expanded in b to show the linearity of decapping activity at lower concentrations of RNA.

Table 4.5: Hyperbolic curve analysis of the RNA titrations from all organisms. Using the same concentrations of RNA as in Chapter 3, RNA was titrated with constant amounts of protein. The $K_{app\ rxn}$ and $k_{obs\ rxn}$ were used to compare all orthologs to the *Xenopus* Nudt16 ortholog. The *Bmax* is in µmol cap released, $K_{app\ rxn}$ is in nM, $k_{obs\ rxn}$ is in min^{-1}, and the specificity constant is in $min^{-1}\ M^{-1}$.

Organism	*Bmax*	$K_{app\ rxn}$ (nM)	$k_{obs\ rxn}$ (min^{-1})	Specificity Constant
Zebra fish	$7.6x10^{-9} \pm 1.0x10^{-9}$	5.9 ± 1.5	$0.02 \pm 2.99E^{-3}$	$3.8x10^{+6} \pm 2.0x10^{+6}$
Sharpshooter	$3.1x10^{-10} \pm 4.1x10^{-11}$	1.4 ± 0.5	$1.4E^{-4} \pm 1.7E^{-5}$	$10.0x10^{+4} \pm 3.4x10^{+4}$
Human	$1.5x10^{-9} \pm 2.1x10^{-10}$	2.6 ± 0.8	$4.0E^{-3} \pm 5.5E^{-4}$	$1.6x10^{+6} \pm 6.6x10^{+5}$
Xenopus	$6.3x10^{-9} \pm 4.4x10^{-10}$	6.9 ± 0.9	$0.02 \pm 1.3E^{-3}$	$2.8x10^{+6} \pm 1.5x10^{+6}$

The first equation to which the data were fitted was the hyperbolic curve; the results are shown in Table 4.5. The observed rate of the reaction ($k_{obs\ rxn}$) and the concentration at which half-maximal velocity is reached under these conditions ($K_{app\ rxn}$) were used to compare each ortholog to the *Xenopus* Nudt16 protein. The sharpshooter was much slower than all other orthologs ($1.4x10^{-4}$ compared to 0.01 min^{-1}), consistent with a scenario with less than optimal conditions. The zebra fish and *Xenopus* orthologs were relatively close together (0.01 – 0.02 min^{-1}). The human observed rate was between the sharpshooter and zebra fish/*Xenopus* rates ($4.0x10^{-3}$ min^{-1}). This is consistent with previous observations that the human Nudt16 ortholog was less active than the *Xenopus* ortholog (Peculis et al., 2007). The concentration of RNA at which half maximal velocity was reached ($K_{app\ rxn}$) was 1.4 nM for the sharpshooter, 2.6 nM for the human, but 5.9-6.9 nM for the zebra fish and *Xenopus* Nudt16 orthologs.

The second equation to which the RNA titration data were fitted is the cooperativity equation (see Chapter 3, page 66). The results of this curve fit are included in Table 4.6. All Nudt16 orthologs have cooperativity coefficients close to 1. The RNA binding mechanism is conserved; when one RNA binds it does not affect binding of a second RNA.

Table 4.6: RNA titrations fitted to the cooperativity equation demonstrate similar RNA binding mechanisms. The parameters obtained by curve fitting the RNA titration data to the cooperativity equation are shown in this table. The maximum velocity units are µmol cap released / min. The concentration of RNA at half maximal velocity ($K_{app\ rxn}$) has units of nanomolar (nM). The cooperativity coefficient has no units.

Parameter	Zebra fish	Sharpshooter	Human	Frog
Maximal Velocity (µmol Cap Released / min)	$5.9x10^{-9} \pm 1.2x10^{-9}$	$3.1x10^{-10} \pm 6.1x10^{-11}$	$1.3x10^{-9} \pm 2.5x10^{-10}$	$4.8x10^{-9} \pm 5.5x10^{-10}$
$K_{app\ rxn}$ (nM)	5.9 ± 1.3	1.3 ± 0.6	1.8 ± 0.7	4.1 ± 0.9
Cooperativity Coefficient	1.3	1.1	1.3	1.2

Taken together, these data indicate the RNA binding mechanism and the requirement for homodimerization are conserved in all organisms. It supports the hypothesis that both the sequence and the function of the Nudt16 protein is conserved in all these organisms. It also provides evidence that the model for Nudt16 function proposed in Chapter 3 (see pages 95-98) is true for all Nudt16 orthologs.

Time Course of each protein determines specific activity

The previous experiments showed that the decapping activity of Nudt16 function was conserved. It was apparent that some Nudt16 proteins required higher concentrations of protein to have the same decapping activity as the other organisms (sharpshooter ortholog). A time course was performed in order to obtain the specific activity for each Nudt16 ortholog.

For the time course, decapping reactions were set up and incubated from 0 – 30 minutes. At specific times, 2 µL of sample was removed and immediately stopped with EDTA, as described in Section 4.2. Product release was analyzed and the data were fitted to the exponential decay model as described (see Figure 4.4). From these data, the specific activity was for each ortholog was calculated.

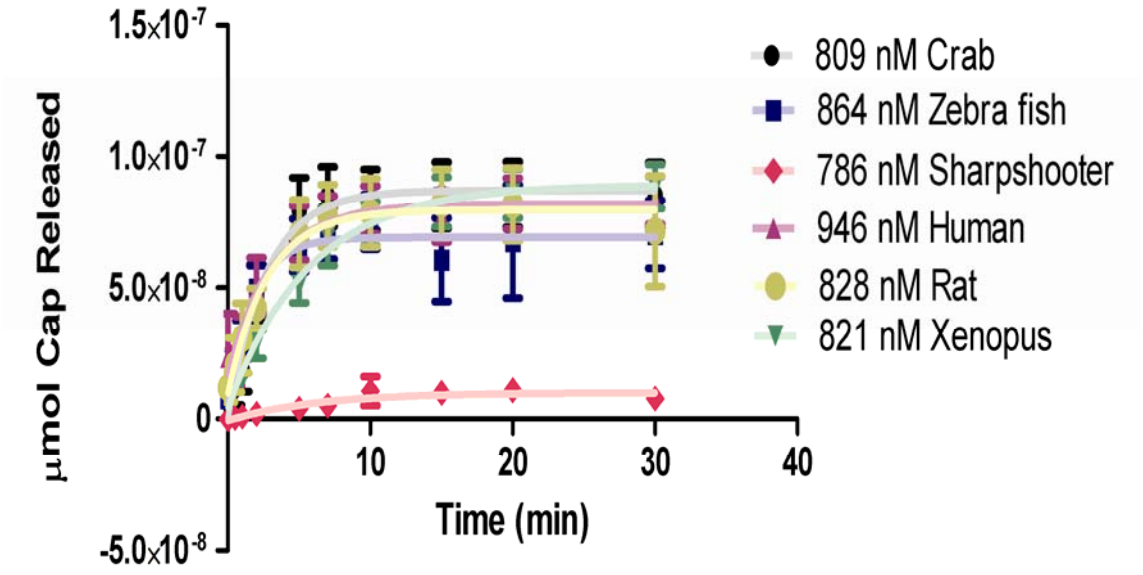


Figure 4.4: Time course of Nudt16 homologs. Bacterially expressed and purified Nudt16 proteins were analyzed by decapping assays. This time course of each Nudt16 homolog using 5 nM U8 snoRNA and 100 ng of protein; the concentration of protein in each assay is given in the figure. Table 4.7 gives the specific activities from this assay.

Based on these data, the zebra fish had the highest specific activity. It was followed by the crab, rat, human, frog, and sharpshooter (see Table 4.7). Most organisms had similar specific activities (within 0.2-0.4 µmoles cap released min^{-1} g^{-1}). The sharpshooter had a much lower specific activity (0.02 µmoles cap released min^{-1} g^{-1}) consistent with the previous observations of lower reaction rates for this ortholog.

Table 4.7: Specific activities for each organism from the time course data. Curves of the decapping activity of each Nudt16 organism were fitted to a single exponential decay model (see Figure 4.4). The specific activity was calculated by multiplying the maximum product released (*Bmax*) by the observed rate constant (k_{obs}) and dividing by the grams of protein used in the assay. The *Bmax* is in µmol cap released, k_{obs} is in min^{-1}, and specific activity is in units of µmol cap released / min / gram of protein or per µmol of protein.

Organism	*Bmax* (µmol cap released)	k_{obs} (min^{-1})	Specific Activity (µmoles cap released / min / g of protein)	Specific Activity (µmoles cap released / min / µmol of protein)
Crab	$8.7E^{-8} \pm 5.0E^{-9}$	0.4 ± 0.1	0.3 ± 0.006	$8.1E^{+9} \pm 1.4E^{+8}$
Zebra fish	$6.9E^{-8} \pm 5.1E^{-9}$	0.6 ± 0.2	0.4 ± 0.012	$8.9E^{+9} \pm 2.8E^{+8}$
Sharpshooter	$1.0E^{-8} \pm 2.0E^{-9}$	0.2 ± 0.1	0.02 ± 0.002	$4.1E^{+8} \pm 4.9E^{+7}$
Human	$8.2E^{-8} \pm 5.3E^{-9}$	0.3 ± 0.1	0.3 ± 0.006	$5.5E^{+9} \pm 1.4E^{+8}$
Rat	$8.0E^{-8} \pm 6.1E^{-9}$	0.4 ± 0.2	0.3 ±0.009	$7.0E^{+9} \pm 2.3E^{+8}$
Xenopus	$8.9E^{-8} \pm 5.2E^{-9}$	0.2 ± 0.04	0.2 ± 0.002	$3.8E^{+9} \pm 4.6E^{+7}$

These data indicate that although the decapping activity of Nudt16 was conserved, the efficiency of each ortholog (as determined by the specific activity) varied. It is possible that these reaction conditions and substrates used are not those preferred by all Nudt16 orthologs although preliminary results with the sharpshooter protein

indicate that the temperature, pH, and salt concentrations are acceptable (KR, MT, and BP, unpublished).

Metal and RNA Preferences

With a quantifiable difference in the specific activity of each ortholog under defined but possibly sub-optimal conditions, efforts to determine metal and RNA preferences seemed necessary. Previously, it had been shown that the *Xenopus* protein could decap all RNAs in Mn^{2+}, but could decap only U8 snoRNA in the presence of Mg^{2+} (Ghosh et al., 2004). All other Nudt16 homologs were analyzed for decapping ability in the presence of various metals and RNAs.

The data from the crab protein are shown in Figure 4.5; the data of proteins from other organisms are included as supplemental figures (see Figures C.1-C.6, pages 199-203). All Nudt16 orthologs were most active in Mn^{2+}; some RNAs were better substrates. For the *Xenopus* and zebra fish proteins, reactions in which U3 snoRNA was the substrate RNA had lower decapping activity. For the sharpshooter, 5S rRNA was the poorest substrate. The crab and rat could decap all RNAs equally well in Mn^{2+}.

Surprisingly, all RNAs were also decapped in the presence of Mg^{2+} for most organisms. In the presence of Mg^{2+}, non-U8 snoRNAs were generally decapped with lower efficiency. Lower decapping efficiency of non-U8 snoRNA substrates would explain the previously published observation that only U8 snoRNA could be decapped in the presence of Mg^{2+} (Ghosh et al., 2004). U8 snoRNA and its truncated counterpart (5'40 RI) were decapped with the same efficiency in all organisms.

Reactions using Co^{2+} as the metal were not decapped efficiently. In some organisms, a low level of decapping was observed.

Together these data show that all three metals, Co^{2+}, Mg^{2+}, and Mn^{2+} can be used in decapping reactions. However, the level of decapping activity changes based on which substrate RNA and metal combination is used. This suggests the *in vivo* RNA and metal conditions for each protein may be part of the regulation of this protein.

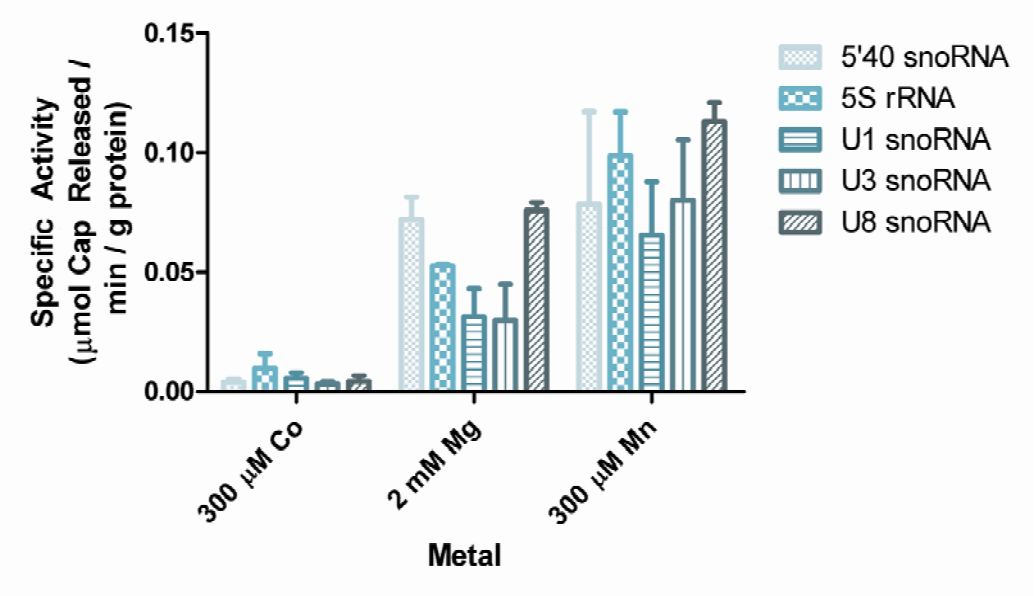


Figure 4.5: The crab Nudt16 ortholog can decap multiple RNAs in different metals although Mn^{2+} exhibits the highest activity. In the presence of Mn^{2+}, all RNAs were decapped by the crab Nudt16 protein (214 nM); no RNA was a better substrate than others. Reactions incubated with Co^{2+} did not decap well at all. All RNAs were decapped in the presence of 2 mM Mg^{2+} though with varying efficiencies. This result suggests that even if Mg^{2+} is the metal present in the *in vivo* conditions, Nudt16 would be able to decap many different types of RNAs.

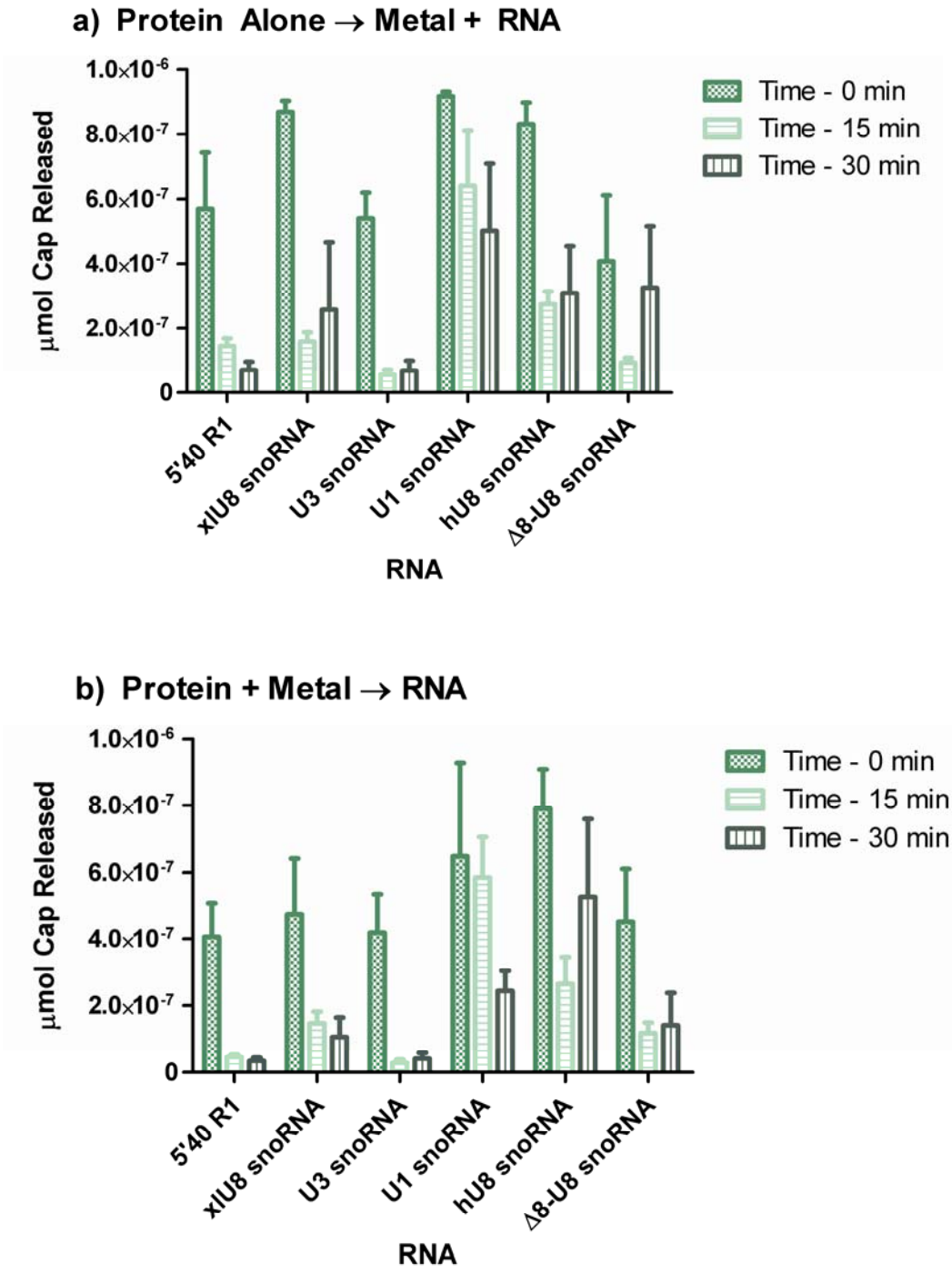
Stability of Nudt16 homologs was dependent on pre-incubation with RNA

Though all Nudt16 orthologs were able to decap different RNAs under variable conditions, none of the orthologs demonstrated convincing catalytic turn over. The question of whether the stability of Nudt16 played a role in the lack of turn over had not been investigated. To test this, proteins were pre-incubated at 37°C for 0, 15, or 30 minutes before all components needed for catalysis were added and changes in activity were monitored. Reactions contained protein alone, protein pre-incubated with metal, and protein pre-incubated with RNA (see methods on pages 111-112). At each time point, the missing reactant was added to the reaction and the incubation continued for 15 minutes.

The results with the *Xenopus* Nudt16 ortholog are shown below; the other Nudt16 orthologs are included in Figures C.7-C.9 (pages 204-209). Pre-incubation of the protein in buffer or the protein + metal resulted in decreased activity of the protein when added to a complete decapping reaction (see Figure 4.6a and b). However, pre-incubation of the Nudt16 protein with any RNA substrate for up to 30 minutes allowed the protein to maintain full decapping activity (see Figure 4.6c). The RNA may confer some sort of stability to the Nudt16 protein, independent of RNA identity.

The data from the Nudt16 orthologs were similar and indicated that pre-incubation with RNA was able to maintain activity of some Nudt16 orthologs, particularly the *Xenopus* and rat orthologs (Figure 4.6 and Figure C.9, pages 208-209). The stabilization of the orthologs was not dependent on a specific substrate; all RNAs were able to stabilize the protein to a similar extent. The zebra fish and human proteins

were not stabilized in the presence of RNA or metal; the activities of the proteins decreased whether the protein was incubated alone, with metal, or with RNA (see Figures C.7 and C.8, pages 204-207).



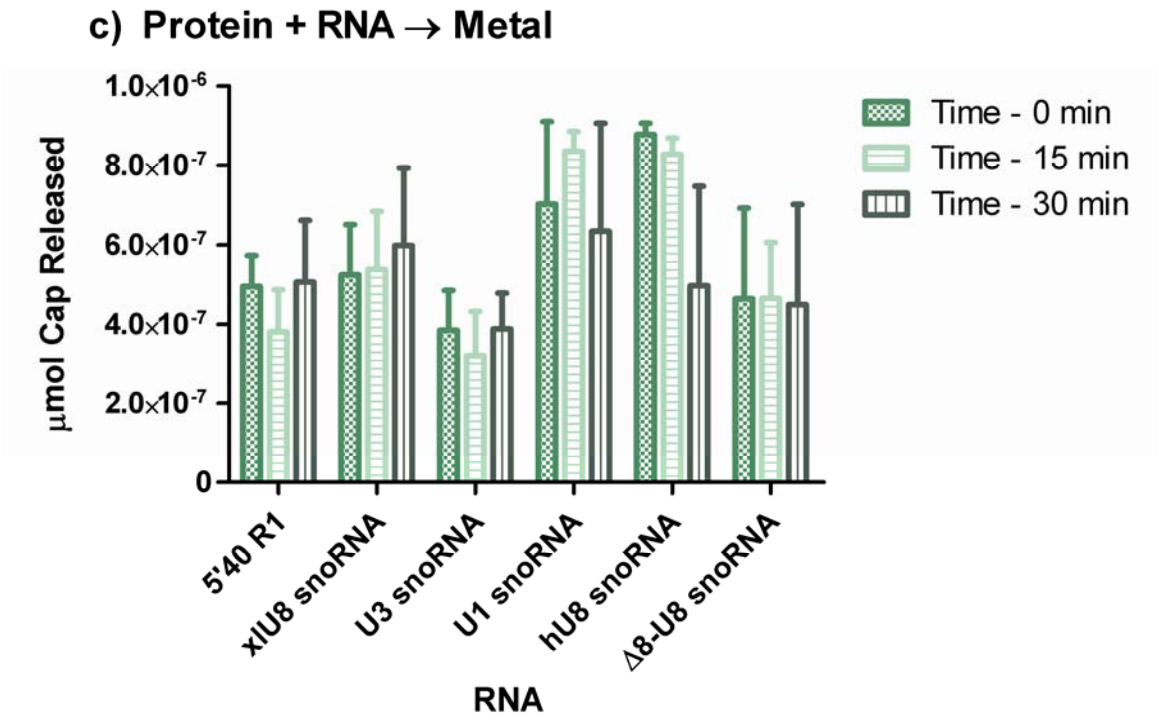


Figure 4.6: Pre-incubation of the frog Nudt16 ortholog without RNA or metal affects ability to decap over time. a) Nudt16 activity decreased over time when the protein was pre-incubated at 37 °C for 15 or 30 minutes. b) When Nudt16 protein was pre-incubated in the presence of metal, the activity of Nudt16 decreased. c) When Nudt16 was pre-incubated in the presence of RNA, the activity of Nudt16 did not significantly decrease.

Can all Nudt16 orthologs bind RNAs?

It is clear that all Nudt16 orthologs can decap RNA. They must be able to bind RNA yet, with the differences in decapping ability amongst the different Nudt16 orthologs as shown in the above experiments, the question remains whether binding of RNA to the Nudt16 proteins is related to the decapping ability in these organisms. The *Xenopus* Nudt16 ortholog was initially identified because it could gel shift U8 snoRNA in an EMSA assay. However, the human Nudt16 ortholog has not been able to demonstrate EMSA gel shifts.

To test whether other Nudt16 orthologs would result in shifted complexes via EMSA with U8 snoRNA, the laboratory's now standard EMSA was performed with each ortholog. Uniformly labeled U8 snoRNA was transcribed as described previously (Ghosh et al., 2004; Peculis et al., 2007; Tomasevic and Peculis, 1999). Protein and RNA were combined in our standard binding buffer with 93 nM tRNA and allowed to bind for 15 minutes at room temperature; binding of RNA to the *Xenopus* protein was in equilibrium after 15 minutes so the same method was used for the other Nudt16 orthologs. Equilibrated reactions were analyzed by electrophoresis on a 4% native polyacrylamide gel. Gels were dried and exposed to a phosphor screen for analysis with Quantity One software (BIO-RAD).

Bands were identified based on the *Xenopus* Nudt16 ortholog observable gel shift. No other orthologs gel shifted the RNA under standard conditions or conditions in which salt and tRNA are added to the reaction (see Figure 4.7).

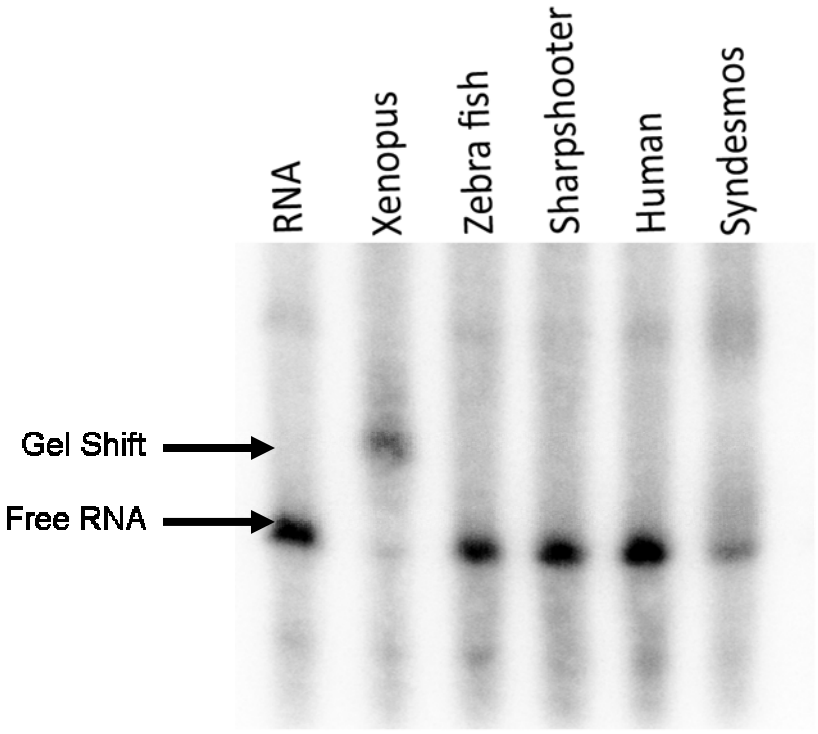


Figure 4.7: Only the frog Nudt16 ortholog gel shifts U8 snoRNA. Under the same conditions, only the *Xenopus* Nudt16 ortholog has a convincing shift denoted "gel shift" in the above figure. All other Nudt16 homologs have similar free RNA bands but no gel shift.

The gel shift results were negative but there must be some transient binding of RNA as each ortholog could decap RNA. Another assay used by the laboratoryto examine direct RNA binding is a UV-dependent RNA:protein crosslinking assay. U8 snoRNA was *in vitro* transcribed in the presence of 4-thio-Uridine and ATP-αP^{32}. This photoactive crosslinker has a radius of 4-10 Å. Nudt16 protein and U8 RNA were allowed to bind under EMSA conditions for 15 minutes and then irradiated for 15 minutes as described in Section 4.2 (pages 112-113). Samples were digested with Pancreatic RNAse A for 30 minutes and then resolved on a denaturing SDS polyacrylamide gel. All Nudt16 orthologs bound U8 snoRNA in a cap-independent,

metal-independent manner demonstrated by the presence of a 20-30 kDa band in the UV crosslinks (see Figure 4.8). These bands were not seen in the absence of protein or UV.

All Nudt16 orthologs also bound U3 and 5S rRNAs, but to different efficiencies in these crosslinking assays. Controls included BSA which did not bind U8 snoRNA and that the crosslinks were competed by addition of unlabeled RNA prior to UV exposure (data not shown). Several bands are visible with the lower bands migrating between 20-30 kDa and the upper bands between 55-70 kDa. These data suggest all Nudt16 orthologs can form homodimers that bind RNA.

Collectively, these data support the hypothesis that the Nudt16 orthologs can bind U8 snoRNA tightly enough to be crosslinked. The crosslinking assay traps interacting molecules, making transient interactions more visible. However, the negative controls, UV-dependence, and competitions of crosslinks argue the interaction is authentic. Since the *Xenopus* ortholog can be used in the EMSA assay, it is likely the interaction between the *Xenopus* ortholog and U8 snoRNA is unusually tight.

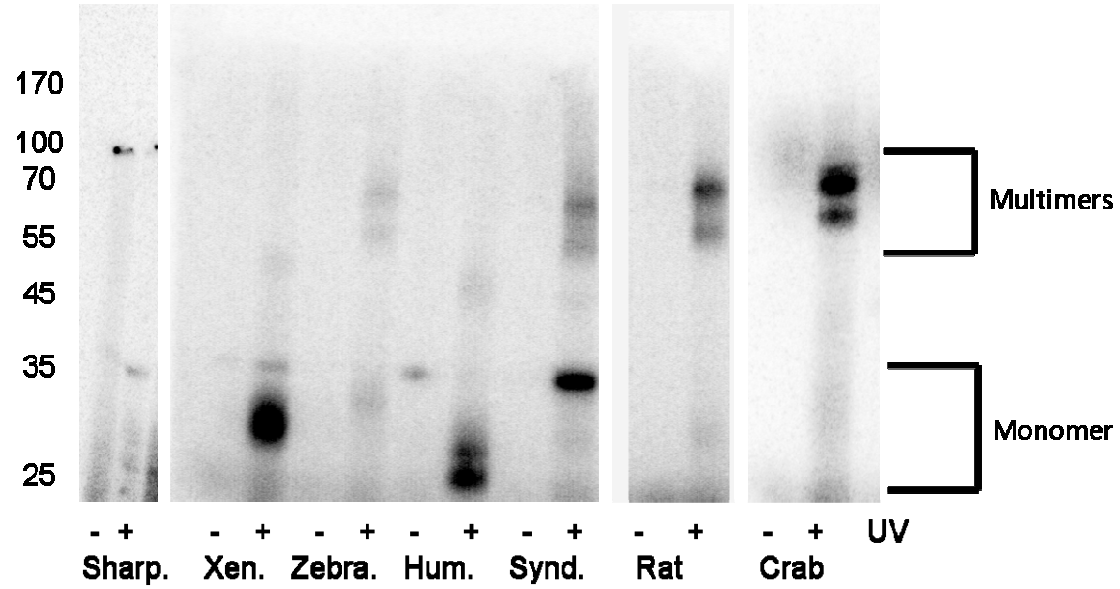


Figure 4.8: All organisms form UV-dependent crosslinks on U8 snoRNA. 4-thio-U U8 snoRNA was crosslinked to Nudt16 orthologs in a UV-dependent manner. The monomeric form of the Nudt16 protein and other multimers are visible. Molecular weight markers run on the gel (left) allow estimation of protein sizes. The lanes are minus UV (-) or plus UV (+) of each ortholog. The orthologs are sharpshooter (Sharp.), *Xenopus* (Xen.), zebra fish (Zebra.), human Nudt16 (Hum.), human syndesmos (Synd.), rat (Rat), and crab (Crab).

Homodimerization of all proteins produces more questions than answers

The crosslinking demonstrated that all Nudt16 orthologs were able to bind RNA and be analyzed as something resembling the monomer state. In addition, most of the Nudt16 orthologs formed 50-70 kDa crosslinks consistent with a homodimer binding RNA, although this was not usually the predominant band. To test whether homodimers could form in Nudt16 orthologs in the absence of RNA, chemical crosslinkers were used. Two micrograms of the Nudt16 orthologous proteins were incubated with 0.5 mM DTSSP or BS^3 as described in Section 4.2 (page 113). Crosslinked

proteins were resolved on denaturing gels and proteins were visualized by Coomassie staining.

The representative gel shown in Figure 4.9 shows that most orthologs were able to form homodimers; RNA was not required for homodimerization of any Nudt16 ortholog. Different orthologs had different efficiencies of homodimer formation. The human Nudt16 protein does not homodimerize well; the other orthologs have higher homodimerization efficiencies. Addition of 10% βME to crosslinked reactions reduced the crosslink to the monomer form of the protein (data not shown).

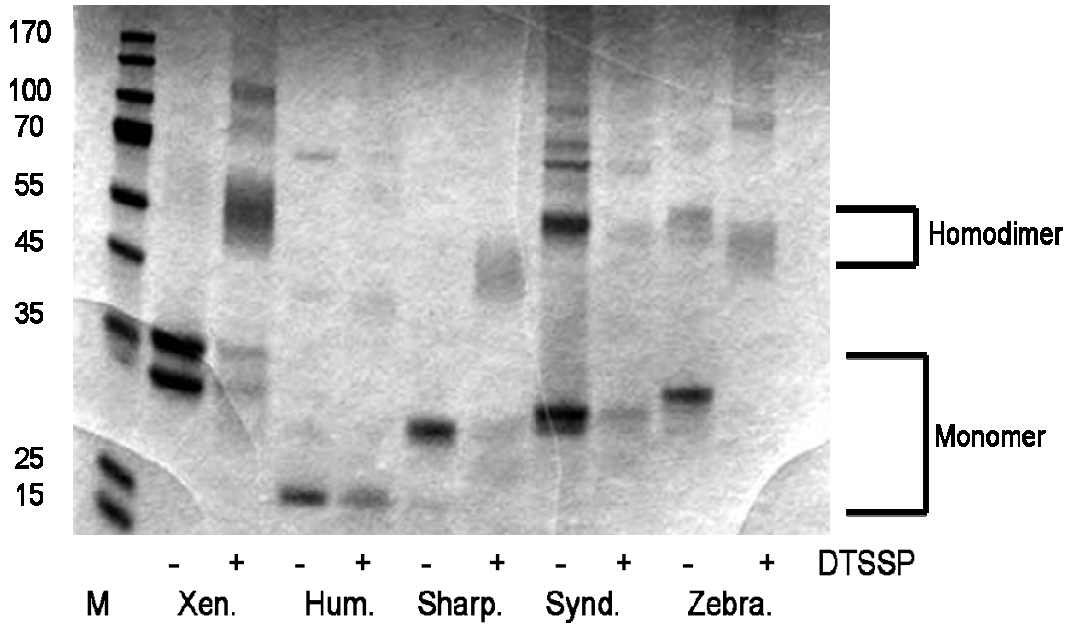


Figure 4.9: Nudt16 orthologs can homodimerize in the presence of DTSSP. Nudt16 orthologs were assayed in the presence and absence of DTSSP for ability to homodimerize. The lanes are marker (M), *Xenopus* (Xen.), human Nudt16 (Hum.), sharpshooter (Sharp.), human syndesmos (Synd.), and zebra fish (Zebra.). The homodimer and monomers are indicated on the right of the figure; molecular weights of the marker bands are displayed on the left side. Nudt16 orthologs do homodimerize based on decreased monomer and/or formation of a homodimer band. The paralog syndesmos has a lot of homodimer already; addition of more crosslinker did not increase the formation of the homodimer. Reducing the Nudt16 crosslinks by addition of βME was observed (data not shown).

4.4: Discussion and conclusions from the conservation of function experiments

Five Nudt16 orthologs were successfully purified (see Figure 4.1). All proteins were functional for decapping indicating these were functional orthologs. The crab protein was not purified to homogeneity and had multiple bands but still exhibited decapping activity. The skate protein was less stable but was still active in decapping. Thus the Nudt16 protein is of ancient evolutionary origin and the primary sequence, structure, and most importantly the catalytic function is conserved in all orthologs.

In addition to determining activity of the orthologs, additional characterization was performed. Over all, the orthologs all bound RNA (assayed by RNA:protein crosslinking), formed homodimers, and decapped a wide variety of substrates in both Mg^{2+} and Mn^{2+}. The significance of each discovery will be addressed below.

To demonstrate conservation of function, the first experiments were protein titrations to determine whether the orthologous proteins could decap in a protein-dependent manner (see Figure 4.2 and Table 4.4). These data indicated that all orthologs were functional and that the homodimer was the functional unit in all Nudt16 orthologs. The RNA titrations supported these results demonstrating that an excess of protein was necessary for decapping activity (see Figure 4.3 and Table 4.5) and also demonstrated that binding of one RNA did not affect binding of a second RNA (see Table 4.6).

Curve fitting of the RNA titration data demonstrated some interesting patterns. The observed rate and the $K_{app\ rxn}$ of the decapping reaction depended on the organism.

The sharpshooter had the lowest $k_{obs\ rxn}$, followed by the human. The zebra fish and the *Xenopus* Nudt16 orthologs exhibited higher reaction rates. The order for the $K_{app\ rxn}$ paralleled the order of the observed reaction rates. The zebra fish and *Xenopus* orthologs were highest, followed by the human, and lastly the sharpshooter.

These data indicated that some orthologs had lower decapping efficiencies. To verify the differences and to compare the efficiency of each ortholog, time courses measuring the decapping activity of the enzyme at various time points were completed. These data, shown in Figure 4.4, were then fitted to an exponential decay model; the fitted values were used to determine a specific activity for each ortholog under the same conditions (see Table 4.7).

The human, rat, crab, and zebra fish orthologs showed the same decapping efficiencies. The *Xenopus* ortholog had a slightly lower specific activity but the sharpshooter specific activity was at least 10 fold lower than all other organisms.

It is possible that bacterially expressed protein functions differently than the *in vivo* protein; however, that is difficult to test without an *in vivo* model system. It is also possible that the sharpshooter (and all insects) have a modified complex required for RNA degradation. Equally likely is that the lower decapping activity is a result of not optimizing the conditions for the sharpshooter protein. It is likely that other RNA substrates are used by the sharpshooter Nudt16 ortholog *in vivo*.

To test the decapping activities of Nudt16 orthologs on different substrates in different metals, five RNAs and three metals were chosen (see Figure 4.5 for the crab results and Figures C.1-C.6, pages 199-203, for all homologs). These five RNAs included

two snoRNAs (U3 and U8), one snRNA (U1), one rRNA (5S), and an artificially truncated construct of U8 snoRNA (5'40R1). The decapping ability of each Nudt16 ortholog on these substrate RNAs in the presence of Mg^{2+}, Mn^{2+}, and Co^{2+} was studied.

In the presence of Mn^{2+}, all RNAs were decapped confirming previous results (Peculis et al., 2007). In the presence of Co^{2+}, much lower levels of decapping were observed for all orthologs. The data from the analyses with Mg^{2+} were of particular interest. Previously, data demonstrated that in the presence of Mg^{2+}, the *Xenopus* ortholog would only decap U8 snoRNA; no other RNAs would act as substrates (Ghosh et al., 2004). The results presented here found that any RNA would act as substrate for the Nudt16 orthologs in the presence of 2 mM Mg^{2+}. Consistent with the previous results, the decapping activity of all orthologs in Mg^{2+} was lower than in the presence of Mn^{2+}. Together these data suggest that the *in vivo* metal could be either Mg^{2+} or Mn^{2+}. Co^{2+} could also be used *in vivo* if the lower levels of decapping in the presence of Co^{2+} were sufficient for the biological system and relatively high levels of Co^{2+} were present in cells.

Another interesting aspect of this study was the discovery that the sharpshooter ortholog did not exhibit increases in decapping activity in any of the metal or RNA conditions used. It is possible that another variable needs to be changed in order to increase the activity of the sharpshooter. Preliminary optimization studies indicate that the pH, salt, and temperature conditions are acceptable for this ortholog (KR, MT, and BP, unpublished); the RNA substrates are not sharpshooter RNAs so may be part of the reason for the sub-optimal activity of this ortholog.

Altered stability of the protein and thus variations in decapping activity in the presence of certain metals and RNAs could explain the differences in specific activity in the protein from different organisms. To test the ability of various RNAs to alter the stability of the protein at 37°C, order of addition experiments were performed, pre-incubating protein with or without metal and RNA. The results shown in Figure 4.6 and Figures C.7-C.9 (pages 204-209) indicated that all Nudt16 orthologs lost decapping activity when pre-incubated at 37°C in the absence of RNA. For most orthologs, pre-incubation with any RNA would preserve activity when metal was added to initiate the reaction.

The decapping assays showed that function was conserved in the Nudt16 orthologs. However, differences were observed between each ortholog. Some possibilities for those differences, like sub-optimal conditions, have been discussed. One additional explaination is differences in RNA binding ability of each ortholog. RNA binding was measured by EMSA and RNA:protein crosslinking assays. The *Xenopus* ortholog was the only Nudt16 protein able to bind RNA in the EMSA assay (see Figure 4.7); it must have a lower K_d of RNA dissociation. The other Nudt16 orthologs could bind RNA demonstrated by a UV-dependent RNA:protein crosslinking assay (see Figure 4.8). Both monomer and homodimer forms of the protein were observed. With so many positive results, it is possible nonspecific crosslinks were being formed but competition experients using cold competitor RNA and reactions containing bovine serum albumin as a negative control did not result in visible crosslinks (data not shown).

Based on the RNA:protein crosslinking assay, the RNA can bind to protein; when resolved on SDS gels, the monomer and a presumed covalently linked homodimer are seen. To test whether the RNA was necessary for formation of the homodimer, the Nudt16 orthologs were studied using chemical crosslinking assays with DTSSP. Nudt16 orthologs were found to form homodimers which were reduced to monomers upon addition of βME in an SDS gel (see Figure 4.9 and data not shown). Since all proteins can homodimerize, with or without the presence of substrate RNA, it supports the mechanism for Nudt16 function outlined in Chapter 3 (pages 95-98). The model of function is also supported by the decapping experiments in this chapter.

Together, these assays show the *in vitro* similarity of the Nudt16 orthologs. All Nudt16 orthologs function as metal-dependent decapping enzymes; the metal can alter the efficiency of the decapping. All Nudt16 orthologs can crosslink RNA and homodimerize. Although variations in decapping activity occurred based on metal and RNA conditions, the general mechanism is still the same: the protein must be a homodimer and binding of RNA to the protein is not cooperative. The evidence points to the conservation of *in vitro* function of all Nudt16 orthologs. It further corroborates the hypothesis that Nudt16 plays an important role in nuclear RNA degradation of RNAs *in vivo* in these organisms.

4.5: Future projects and final summary of these data

The question of RNA binding and protein:RNA interactions is of interest to further understanding of the mechanism of Nudt16 function. It is possible that the observed specific activity differences may be due to the lower K_d of the RNA dissociation. For that reason, trying to analyze the RNA binding of the sharpshooter, human, and *Xenopus* could be useful. However, since EMSAs cannot be used for these analyses, other techniques must be used. These are complex enough and difficult enough that one could easily have another thesis based on the biophysical properties of RNA binding to the Nudt16 protein.

A more biologically relevant question is to look at the localization of each Nudt16 ortholog in the organism. The *Xenopus* Nudt16 ortholog is localized to the nucleolus of the nucleus in punctuate spots (Ghosh et al., 2004). If this is true for other Nudt16 orthologs, it would support the proposed model of nuclear RNA degradation in all these organisms. This project is not feasible at this time; none of the antibodies or reagents to do this project is available. Everything would have to be generated.

Finally, although several Metazoans have been identified with the Nudt16 gene, the absence of this gene and protein in other organisms is equally interesting. Roundworms and earthworms and leeches have a recognizable Nudt16 gene but no *C. elegans* gene looks like a Nudt16 gene. Likewise, aphids, the sharpshooter, the cricket, and the tick are known to have the Nudt16 ortholog but no such ortholog has been recognized in *D. melanogaster.* If the protein plays such an important role in a nuclear

RNA degradation pathway, it is highly unlikely that these organisms do not have a functional Nudt16-like gene.

There are three possibilities to explain the apparent lack of Nudt16. The first is that there is a Nudt16-like gene that has not been found because the constraints used to identify Nudt16 genes by sequence analysis were too strict. It is possible that using a structural analysis instead of a strict sequence analysis to determine Nudt16-like genes will provide potential Nudt16 orthologs in those species. Structural comparisons of NUDIX proteins in *D. melanogaster* to that of the sharpshooter Nudt16 ortholog demonstrated some proteins containing a Nudt16-like structure (GC and BAP, unpublished). It remains to be seen whether those proteins are functional Nudt16-like genes.

A second possibility to explain the apparent lack of the Nudt16 gene in these organisms is that Nudt16 did not evolve until later so nuclear RNA degradation occurs through a different nuclear pathway. It is known that a 3'-5' nuclear RNA degradation pathway exists in yeast (Das et al., 2003; Das et al., 2006). It is possible that the 3'-5' degradation pathway is used in other organisms that do not have a Nudt16-like gene.

Finally, the third possibility is that the cytoplasmic RNA degradation pathway is the only pathway used in organisms lacking the Nudt16 gene. It is possible that the Nudt16 gene was lost in *D. melanogaster* and *C. elegans*. Instead of having two RNA degradation pathways, these organisms have only the cytoplasmic pathway.

Although many things remain unknown about the Nudt16 orthologs, the data shown in this chapter have added to the knowledge of Nudt16 function. The gene is

conserved (shown in Chapter 2 pages 24-60) and this chapter showed that the function is also conserved. RNA binding and homodimerization indicate some differences between the orthologs which may contribute to the differences in activity. From these results, the model of Nudt16 function has been refined to include decapping activity on all substrate RNAs in the presence of Mg^{2+} as well as Mn^{2+} with the metal co-factor still one way to regulate the efficiency of Nudt16 decapping activity.

Chapter 5: Protein Co-Immunoprecipitations and Interactions

5.1: Why are protein interactions of Nudt16 important?

In the previous two chapters, experiments were designed to detail the functional mechanism for the Nudt16 protein from a wide variety of organisms. Several common themes occurred, but one of the most prevailing was the fact that none of the proteins would catalyze a second round of decapping, based on several different types of evidence (see Chapters 3 and 4 for more details, pages 62-142). It was proposed that other factors might be necessary *in vivo* for Nudt16 to release the RNA or to reset the protein for a second round of catalysis.

Discerning which protein bound both Nudt16 and U8 is difficult because, like all snoRNAs, U8 exists as snoRNPs *in vivo* with various proteins bound to the RNA. At any point in the cell cycle, it is known that many proteins to bind to U8 RNA. Looking for any partners to bind U8 RNA is likely to produce false positives, especially if Nudt16 has broad RNA specificity. It is possible to approach the problem using Nudt16.

Section 5.2 details the procedure used to investigate potential interactors of Nudt16. Co-immunoprecipitations using a flag-tagged construct of Nudt16 and tissue culture cells isolated functionally active protein. Three proteins were identified that bound to Nudt16 in at least two independent replicates; an explanation of the role

those proteins might have in a complex with Nudt16 is briefly mentioned. Finally, possible experiments and future directions for this project are discussed in Section 5.5.

5.2: Methods to search for *in vivo* partners of Nudt16

Preparation of tissue culture cells

Constructs of both the *Xenopus* and human Nudt16 cDNAs were previously cloned into the flag-tagged pCMV vector (TG and BAP, unpublished). Flag-tagged constructs were over-expressed in *Xenopus* (XLKWG) and human (HeLa) tissue culture cells using Lipofectamine 2000 (Invitrogen #11668-019) for 16-24 hours post-transfection. One 150 mm dish of cells was harvested by scraping in 10 mL phosphate buffer solution (PBS). Cells were pelleted by centrifugation at 1500 rpm for 5 minutes. PBS was decanted and cells were resuspended in 10 mL PBS and pelleted again to wash the cells; wash was repeated a second time.

Pellets were resuspended in 600 µL NET-2 buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% NP-40 with 6 tablets of protease inhibitor cocktail (Roche #1836170) in every 50 mL buffer). Suspended pellets were lysed by sonicating twice at 30% amplitude with a microprobe for twenty seconds each with at least 30 seconds on ice between each round of sonication. Sonicated cell debris was pelleted by spinning at 15000 rpm for 10 minutes at 4°C. The supernatant (the "pre" fraction) was removed and stored at 4°C for the next step.

Co-immunoprecipitation using anti-FLAG beads

Anti-FLAG beads (Sigma #F2426) were washed three times with NET-2 buffer. 50 µL of 50% anti-flag bead slurry were added to the supernatant of the pelleted tissue culture cells. Lysate bound at 4°C for 2 hours, rotating gently. Supernatant was

removed and kept as a control to ensure depletion of the Flag-tagged protein; this is the "post" fraction. Beads were washed 10 times with 1 mL NET-2 buffer. The first elution was in the addition of 50 µL 1 mg/mL Flag peptide (Signma #F3290) and rotation at 4°C for 1 hour. Three subsequent 10 minute elutions with 50 µL 0.5 mg/mL Flag peptide were pooled with the first elution to ensure all proteins were completely eluted from the beads.

Decapping reactions with extracts from co-immunoprecipitations

Standard decapping reactions with 10 nM RNA and 1 µL of each immunoprecipitation fraction were incubated as previously described (Ghosh et al., 2004; Peculis et al., 2007). Metal was added to a final concentration of 150 µM Mn^{2+} to ensure optimal performance of the Nudt16 protein. Samples were incubated at 37°C for 30 minutes; the reaction was stopped with 0.5 µL 0.5 M EDTA. One µL of each decapping reaction was spotted on the TLC plates and developed in 0.5 M LiCl and 1 M Formic Acid. Data were analyzed and quantitated with ImageGauge software (Fujifilm).

Western blot of co-immunoprecipitation extracts

Pre, post, eluate, and wash fractions were denatured in 4x NuPAGE dye + 10% BME for 5 minutes at 65°C. Fractions were resolved on a 10% Bis-Tris NuPAGE gel at 200 V for 45 minutes. Gel was transferred to a nitrocellulose membrane (Invitrogen #LC2001) for 1.5 hours at 35 V. Membrane was enhanced 2 minutes with 20 mL Millenium Enhancer Solution (Biochain Institute #K3172120), washed with TTBS (50 mM Tris pH 7.5, 150 mM NaCl, 0.1% (v/v) Tween-20) for 5 minutes, and blocked for 2 hours in a 5% blocking solution. Blocked membrane was incubated over night at 4°C in

antibody solution containing an anti-Flag antibody (Abcam #ab21536-100) diluted at 1:1000 (final concentration of antibody was 1 μg/mL).

The next day, the membrane was washed four times with TTBS for 5 minutes each. The secondary antibody was anti-rabbit HRP (Sigma #A-6154) diluted 1:1000. Membrane was incubated in secondary antibody for 2 hours and then washed again in TTBS for 5 minutes. Three subsequent washes in TTBS for 5 minutes each ensured all non-specific antibody interactions were removed from the membrane. Membrane was washed a final time for 5 minutes with TBS (50 mM Tris pH 7.5, 150 mM NaCl). Membrane was developed by adding 5 mL of ECL Reagent and 2.5 μL 30% H_2O_2 and exposed to film.

TCA precipitation and analysis of eluted fractions

To precipitate the proteins, 0.02% (v/v) sodium deoxycholate was added to the eluted samples for 15 minutes at room temperature to help facilitate precipitation. Trichloroacetic acid (TCA) was then added to a final concentration of 7% to initiate precipitation.[10] Samples were incubated on ice for at least 2 hours. Precipitates were collected by centrifugation at 4°C for 15 minutes at 15000 rpm. TCA was removed and 100 μL of ice cold 100% acetone was added to remove TCA. Samples were spun, washed a second time in acetone and pelleted. Residual acetone was removed and pellets were air dried for at least 10 minutes.

[10] Sodium deoxycholate is added as a base and binds to the protein through hydrophobic interactions. This helps the proteins precipitate better. TCA is a strong acid which then precipates both the sodium deoxycholate and the protein.

Dried pellets were redissolved in 10 µL 2x standard SDS gel loading sample buffer (125 mM Tris, pH 7, 4% SDS, 20% Glycerol, 10% βME, 10% bromophenol blue), denatured for 5 minutes at 65°C, and run on a 15 cm x 15 cm 10% SDS-glycine-PAGE gel until the bromophenol blue dye was at the bottom of the gel. The gel was washed and stained with PageBlue (Fermentas #R0571) as per manufacturer's directions. Visible bands were excised and subjected to trypsin digestion.

Trypsin Digest of Excised Bands

Excised gel pieces were sliced into 1 mm^3 pieces and equilibrated 15 minutes at room temperature in 500 µL 100 mM ammonium bicarbonate. Equilibrated gel pieces were washed with 500 µL Wash Solution 1 (50% 100 mM ammonium bicarbonate, 50% acetonitrile) for 15 minutes at room temperature. Wash was repeated three times in Wash Solution 1. Gel pieces were rinsed with 500 µL 100% acetonitrile and then dehydrated for 20 minutes at room temperature with 500 µL 100% acetonitrile. After air-drying for 20 minutes, the protein was reduced with 150 µL 10 mM dithiothreitol (DTT) in 100 mM ammonium bicarbonate for 30 minutes at 56°C. Samples were cooled to room temperature and the DTT solution was discarded. Reduced protein was alkylated with 100 µL 50 mM iodoacetamide (IAA) in 100 mM ammonium bicarbonate for 30 minutes at room temperature in darkness. IAA was removed and gel pieces were rinsed with 500 µL 100% acetonitrile. Gel slugs were dehydrated for 20 minutes at room temperature with 500 µL 100% acetonitrile. Gel pieces were re-hydrated with 30 µL 0.2 µg/µL trypsin (Promega #V5111) diluted in trypsin buffer (40% 100 mM ammonium bicarbonate, 10% acetonitrile, 50% water). After digesting 2 hours at 4°C, the trypsin

was removed and gel pieces were resuspended in 50 µL trypsin buffer, wrapped in foil, and incubated 16 hours at 37°C.

Trypsin buffer was removed and 13 µL of extraction solution (50% acetonitrile, 5% formic acid, 45% water) was added to acidify the eluted peptides. This was stored on ice until the end of the extractions. To finish extraction, 25 µL of extraction solution was added for 10 minutes and incubated at room temperature. A total of 3 extractions were done. Extractions were combined and both extracts and gel pieces were immediately frozen in liquid nitrogen and stored at -80°C for analysis. Samples were analyzed by the Proteomics Center at the University of Missouri Columbia for protein identification.

5.3: Results of the search for proteins that interact with Nudt16

Differential decapping of co-immunoprecipitations extracts

Each of the different factions collected from the co-immunoprecipitations were assayed in standard decapping reactions. The incubation time was 30 minutes to ensure the maximum amount of enzyme activity was detected. The initial cleared extract (pre) and the depleted extract containing proteins that did not bind the flag beads (post) were examined in both human (HeLa) and *Xenopus* (XLKWG) cells for decapping activity (Figure 5.1a). All "pre" were active for decapping and the "post" generally had some decapping activity. Two possible explanations exist for the observed decapping activity in these fractions. It is possible that the Nudt16 protein was not completely depleted during the co-immunoprecipitation. Alternatively, the decapping activity could be due to other enzymes in the preparation (such as Dcp2).

To determine whether Nudt16 was depleted in the "post" fraction, a Western blot of all fractions was done using an anti-Flag antibody (see Figure 5.1b). The "pre" samples had some protein; "post" fractions had no protein. Wash and eluates both contained protein.

Decapping assays performed on the combined washes showed negligible decapping activity (see Figure 5.1a). Eluates showed a variety of decapping activities. Elutates from the XLKWG cells decapped U8 snoRNA, whether the over-expressed cDNA was the clone for the *Xenopus* ortholog of Nudt16 or the human ortholog. In contrast, HeLa cell extract eluates were unable to decap RNA, probably due to the low amount of

protein present in those samples (see Figure 5.1b). The lower amount of protein in HeLa extracts may be due to inefficient transient transfection of the plasmid, to poor in vivo translation, to inefficient co-immunoprecipitation or perhaps suggest the presence of a negative regulator of Nudt16 expression in mammalian cells.

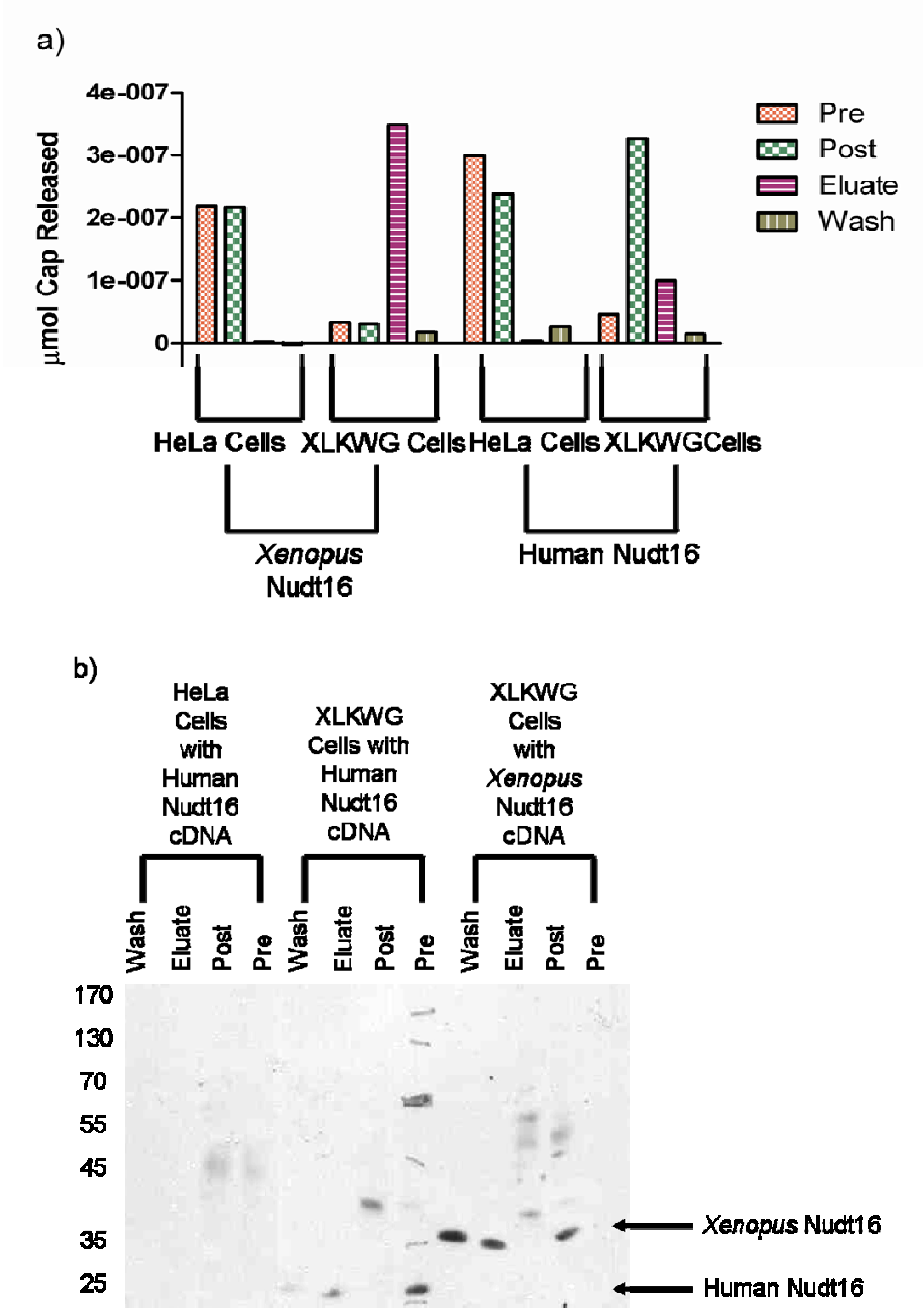


Figure 5.1: Decapping of immunoprecipitation extracts is dependent on cell type. a) Fractions from each step of the immunoprecipitation experiment were set up in decapping reactions with 10 nM (0.05 pmol) U8 snoRNA. Pre, post, eluate, and wash are as described in the text. b) Pre, post, eluate, and wash fractions used in the decapping reactions were analyzed by Western blotting using an anti-Flag antibody. In XLKWG cells, the protein was evident in the pre, eluate, and wash fractions. The protein was depleted in the post fractions. The HeLa cells did have protein based on a silver stain, but the level of protein was below the limit of Western blot detection.

Co-immunoprecipitation experiments showed distinct band patterns

Co-immunoprecipitations of *in vivo* assembled complexes with Nudt16 using a flag-tagged construct of the human and *Xenopus* Nudt16 proteins were examined. Negative controls consisting of vector only samples were run in parallel. Examining all samples in a denaturing SDS-glycine-PAGE gel identified multiple bands (Figures 5.2-5.3). Initially, silver stain was used to visualize the bands (Figure 5.2). Nudt16 was co-immunoprecipitated only in those cells over-expressing the Nudt16 cDNA. Some bands were the same between extracts from cells over-expressing the vector alone and extracts from the cells over-expressing the flag-tagged human Nudt16 (see Figure 5.2).

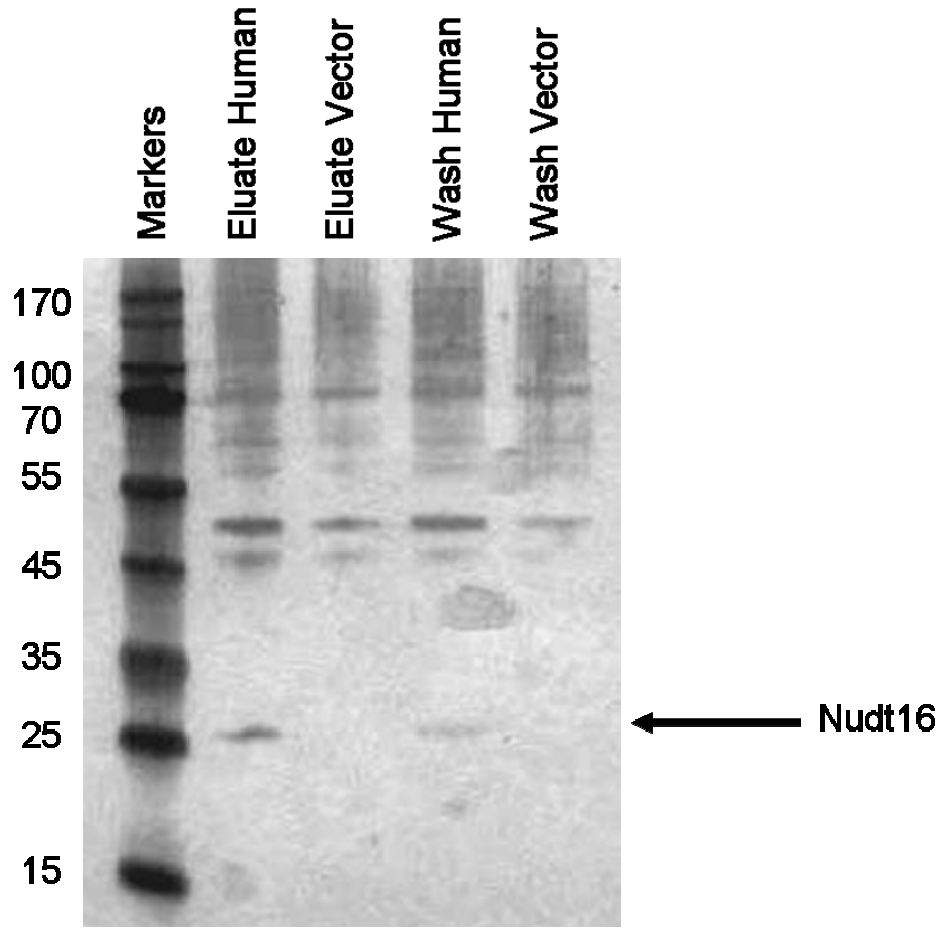


Figure 5.2: Silver stain of gel with immunoprecipitation eluates and washes. Eluates and washes from HeLa cells transformed either with the human Nudt16 cDNA or the empty pCMV vector are shown as indicated. The band identified as Nudt16 migrated just above the 25 kDa molecular weight band in the marker lane which is where it would be expected to run based on the predicted molecular weight from the amino acid sequence. It is identified in both the eluate and wash of the human Nudt16 samples but not the vector samples.

The co-immunoprecipitation was repeated or some fraction of the previous sample was resolved on a 10% SDS-glycine-gel stained with coomassie blue (see Methods). Bands that were different between cells expressing the vector and cells expressing the Nudt16 cDNA were excised for mass spectrometry. The resulting protein identifications are labeled in Figure 5.3.

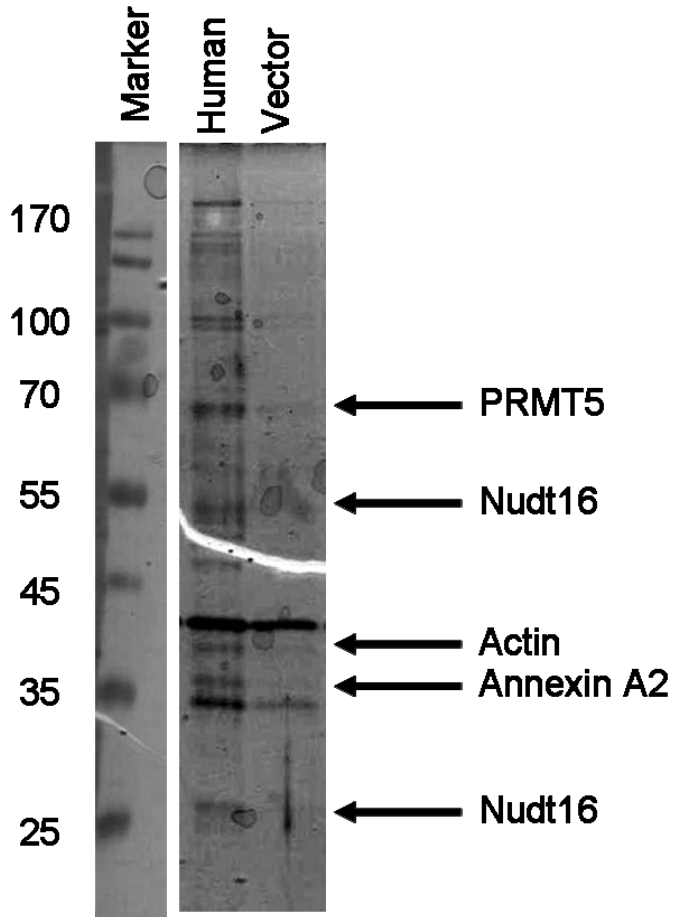


Figure 5.3: Gel of Nudt16 co-immunoprecipitaions with bands identified. Gel of Nudt16 co-immunoprecipitations from which bands were excised for mass spectrometry analysis. Those with confident IDs are labeled. Unmarked bands either were not excised or did not have confident IDs because the concentration of protein was too low. The "Marker" lane contains molecular weight markers whose approximate molecular weight in kDa is shown on the left. The "Human" lane is HeLa cells over expressing the human Nudt16 cDNA. The "Vector" lane is HeLa cells transfected with the vector construct.

Mass spectrometry results of the excised bands clearly identified three proteins

To find potential regulators of Nudt16, bands were excised from Coomassie stained gels, trypsin digested, and sent to the mass spectrometry facility at the University of Missouri – Columbia for further analysis. The analysis revealed several trypsin peaks. Some were identified as Nudt16; others were unidentified or unknown proteins (see Table 5.1). Those with no confident ID were likely a result of too little protein in the sample or a mix of proteins in the isolated band. Three proteins were identified in multiple analyses: actin, PRMT5, and annexin A2 (see Figure 5.3).

Table 5.1: Results of mass spectrometery analysis revealed three interacting proteins. The mass spectrometry analysis was done on the trypsin digests using tandem MS/MS. An example of one of the results is shown below. Skb1Hs is another name for PRMT5. Actin, Annexin A2, and Nudt16 were also identified (refer to Figure 5.3 for their locations on the gels).

PMF-based identification using MASCOT vs NCBI-mammals (score >71 = p<0.05; 95% confidence)	Matched peptides
Skb1Hs [Homo sapiens] (127) - arginine methyltransferase	20
Gamma-actin [Mus musculus] (189)	19
Annexin A2 [Pan troglodytes] (166)	23
Annexin A2 isoform 1 [Homo sapiens] (129)	18
NUDT16 protein [Homo sapiens] (228) gi|16306954	20

5.4: Discussion of the results and proposed model for *in vivo* Nudt16 function

The goal of this part of the project was to identify proteins that interacted with Nudt16 *in vivo*. Nudt16 expression *in vivo* is tightly regulated; transformation of the cDNA construct for the Nudt16 protein into cells results in abundant over-expression of the Nudt16 protein initially but this decreases to basal levels within 2-4 weeks of over-expression, even when under selection to maintain the plasmid (Peculis Laboratory, unpublished data).

The initial experiments showed that there was decapping activity in both the cleared and the depleted extracts of the co-immunoprecipitations (Figure 5.1a). Since decapping enzymes such as Dcp2 would be present but not expected to bind to the Flag beads, they could be responsible for this activity. In addition, a Western blot of the fractions showed low levels of Nudt16 in the pre sample; that may also account for the decapping activity in that fraction. Nudt16 protein was depleted in the post samples; Nudt16 activity would not explain the decapping activity in the post fraction. Wash samples had negligible decapping activity though at least for the *Xenopus* cDNA in XLKWG cells, equal amounts of protein were present in the wash and eluate.

What was surprising was the difference in activity based on the host species of the protein production. Eluates from the XLKWG cells were able to decap, whether the over expressed cDNA encoded the *Xenopus* homolog or the human homolog of Nudt16 though the human ortholog was less active due to lower protein expression (see Figure 5.1b). Correspondingly, the Western blot showed much less Nudt16 in those cells. The HeLa cell eluates were unable to decap, probably due to the low levels of protein being

expressed (see Figure 5.1b). Many things could contribute to the low protein expression including difficulties in the transfection, poor *in vivo* translation of the cDNA construct, or inefficient co-immunoprecipitation.

In order to learn more about the interacting proteins *in vivo,* co-immunoprecipitations were used to isolate *in vivo* assembled Nudt16 protein complexes. Distinct differences between cells over-expressing a Nudt16 flag-tagged construct and cells over-expressing the vector alone were observed (Figures 5.2-5.3). The 30 kDa Nudt16 band was confirmed by mass spectrometry. Other isoforms of Nudt16 were also found, particularly the homodimer migrating just under the 70 kDa molecular weight band and potential multimers migrating closer to 100 kDa (data not shown).

Some bands were more distinct than others, possibly due to their abundance in the cell and their affinity for the resin. Actin is ubiquitously expressed and is highly abundant so that band was easily distinguished. The Nudt16 band was less abundant and more difficult to visualize. Other indistinct bands were of such low concentration that not only was it difficult to cut out an individual band, the amount of protein was below the limit of detection of the mass spectrometer. Several bands were recognized multiple times so future work might involve trying to identify these less abundant bands which are most likely to contain authentic interactors.

To identify these putative regulators and interacting proteins, bands that differed between the cell extracts from HeLa cells expressing the human Nudt16 cDNA were excised and sent for sequencing at the University of Missouri – Columbia mass

spectrometry core. The results are shown in Table 5.1. Some hits had no match or matched with hypothetical or unidentified proteins. Three proteins were identified 3-4 times in independent replicates: actin, annexin A2, and PRMT5.

Together these data indicate that co-immunoprecipitations could be useful in determining interactors of Nudt16. It must be stressed that these are only preliminary results, but there is some interest in these proteins. Many nuclear and nucleolar proteins are regulated by methylation; one of the identified proteins is a methyltransferase. PRMT5, previously called Skb1, forms symmetric dimethylarginine residues in a number of proteins (Branscombe et al., 2001; Pollack et al., 1999). Boisvert *et al*. identified several protein complexes that were methylated; many were involved in pre-mRNA processing (2003). It might be possible that Nudt16 is also methylated and that allows the enzyme to turn over. As these are preliminary results, the hypothesis is interesting, but more evidence must be obtained before a model can be built.

The identification of actin and Annexin A2 are interesting because they have been shown to interact together in the cell (Filipenko and Waisman, 2001; Hayes et al., 2004; Hayes et al., 2006). In addition, Annexin A2 can bind mRNAs recognizing a consensus sequence in the 3'UTR: AA(C/G) (A/U)G. (Aukrust et al., 2007). A possible model in which Annexin A2 is regulated by nucleo-cytoplasmic transport has been proposed but is incomplete, missing at least one factor that is involved in the pathway (Liu and Vishwanatha, 2007); perhaps Nudt16 is the missing factor. Again, the data presented in this dissertation are only preliminary so it is necessary to do more experiments before these pathways can be considered *in vivo* roles of Nudt16.

5.5: Future Projects

The first project is to confirm the protein interactions of Nudt16 and the three proteins identified in these experiments. Clones for PRMT5 and Annexin A2 have been obtained from ATCC. They have been sub-cloned into the pGEM T-Easy vector system (Promega #A3610) but will need to be cloned into a protein expression vector (pET19b; Novagen #69677-3) before further experiments can be done. Once those proteins are cloned into pET19, they need to be expressed and purified probably using the His-tag affinity purification procedures described in Chapter 4 (pages 106-108) for the Nudt16 orthologs.

Once the proteins are expressed and purified, the first test is to see whether they will affect the decapping activity of Nudt16 *in vitro*. The simplest experiment is to add increasing concentrations of the protein of interest to a standard decapping reaction with a constant amount of Nudt16 and U8 snoRNA. For reactions with PRMT5, it is necessary to add some amount of S-adenosyl–L- methionine so that the enzyme has the substrate required for methylation of Nudt16. For reactions with Annexin A2, calcium will need to be added for binding of annexin to the RNA. Incubate all reactions 30 minutes at 37°C and analyze by TLC in the standard methods described in previous chapters.

Whether there is an observable effect on Nudt16 decapping activity, it would still be important to study the ability of Nudt16 to form complexes with these proteins. Co-immunoprecipitations can have false positives; proteins that bind to the Flag-tag of the

cDNA construct or that bind to the resin by itself are likely to contaminate any immunoprecipitaion experiment. For that reason, it is necessary to make sure the proteins really are interacting with Nudt16. One way to tests protein:protein interactions is using crosslinking reagents.

Each protein would have to be tested individually as well as in combination with Nudt16; Table 5.2 lists the interactions to test. The experiment described in the last row of Table 5.2 is of particular interest. It is possible that Annexin A2, actin, and Nudt16 form a trimeric complex; Nudt16 forms the base of the complex through an interaction with either Annexin A2 or actin. Evidence that all 3 proteins are able to form a heterotrimer would support a model in which the trimeric complex is involved in nuclear-cytoplasmic transport of RNA. Combining all three proteins in a crosslinking reaction and comparing the bands of the combination of three proteins to the bands with Nudt16 plus Annexin A2 or Nudt16 plus actin or Nudt16 alone should provide some clues.

Table 5.2: Reactions of interest in determining the ability of Nudt16 to heterodimerize. The first four rows are controls to know whether each individual protein can form multimeric complexes. The next three test Nudt16 heterodimerization with a protein of interest. Following that, two samples testing PRMT5 heterodimerization appear. Annexin A2 and actin are combined next to see whether they function as it would be expected; if annexin A2 can bind to actin the proteins are functioning correctly. Finally, a reaction looking for a proposed trimer when Nudt16, actin, and Annexin A2 bind together.

Reaction	Protein 1	Protein 2	Protein 3	Reason
Nudt16 Only	Nudt16	-	-	Standard Control
PRMT5 Only	PRMT5	-	-	Standard Control
Annexin A2 Only	Annexin A2	-	-	Standard Control
Actin Only	Actin A2	-	-	Standard Control
Nudt16 + PRMT5	Nudt16	PRMT5	-	Does Nudt16 heterodimerize with this protein?
Nudt16 + Annexin A2	Nudt16	Annexin A2	-	Does Nudt16 heterodimerize with this protein?
Nudt16 + Actin	Nudt16	Actin	-	Does Nudt16 heterodimerize with this protein?
PRMT5 + Annexin A2	PRMT5	Annexin A2	-	Does PRMT5 heterodimerize with this protein?
PRMT5 + Actin	PRMT5	Actin	-	Does PRMT5 heterodimerize with this protein?
Annexin A2 + Actin	Annexin A2	Actin	-	Do annexin A2 and actin function as expected?
Nudt16 + Annexin +Actin	Nudt16	Annexin A2	Actin	Will Nudt16, annexin A2, and actin form trimer?

Chapter 6: Future Projects and Directions

6.1: Introduction to the chapter on proposed experiments

This chapter proposes hypotheses that come directly from the data presented in this dissertation. As such, these questions and the experiments proposed will direct some of the future research of the Peculis laboratory.

The first two sections of this chapter ask about the *in vivo* role of the Nudt16 protein. There are several questions involved in this section. How do the expression levels of the Nudt16 mRNA and protein change during development? Are cells where Nudt16 is depleted or absent still viable? What are the target RNAs? How do hormones change these answers?

The first section proposes the development of a genetic model system in zebra fish. The questions of target RNAs, Nudt16 expression in development, and viability of Nudt16-depleted cells could be addressed by these experiments.

The second section of this chapter asks the questions of changes in Nudt16 expression in the presence of hormones. These experiments could address the regulation of Nudt16 expression which has been shown to be linked to estrogen and progesterone expression in sheep (Ing et al., 2006). Instead of looking at the effects of hormones in tissue culture cells or tissues, using the zebra fish model system developed in the first section, the effect in the organism could be monitored.

Hormones are only one way of regulating protein expression or activity. It is also possible to regulate protein activity by evolving complexes of proteins. Does Nudt16 have a regulator? Are other factors involved in the proposed nuclear decapping complex? The third section discusses some experiments that might address these questions. The initial discoveries in Chapter 5 could be expanded.

The final section in this chapter is the summary of the dissertation. The hypothesis is reviewed and the data from each chapter put in perspective with respect to the larger questions guiding the research in the Peculis laboratory.

6.2: *In vivo* model system of Nudt16 function

In studying the *in vivo* role of Nudt16, there are many questions that could be addressed. One question is the role of Nudt16 in development. It is known that the Nudt16 protein is expressed in *Xenopus* in both somatic cells (XLKWG) and oocytes. However, in somatic cells, the endogenous levels of Nudt16 protein are lower than in oocytes which may either be using the protein during the development of the oocyte into the embryo or stockpiling the protein for future use. It is possible that Nudt16 is important in *Xenopus* for degrading maternal RNA transcripts prior to transcription of embryonic RNAs. However, it would be useful to get a different genetic model system in which to study the Nudt16 expression during embryogenesis and into adulthood.

From Chapter 2, the mouse and rat are potential model systems. However, databases show a pseudo-gene exists in mammals which could be transcribed and play a role in regulation. This would complicate experiments designed to address Nudt16 function. It is no doubt that some time the laboratory will return to mammalian systems, but it will be after the *in vivo* role of Nudt16 is better understood. A better model organism to begin the investigation is the zebra fish. The Nudt16 gene from the zebra fish has been cloned and is active for decapping (see Figure 4.2, page 117). Zebra fish have been used for genetic manipulation in many different fields of biology and their relatively quick growth rate and large hatchings make embryo manipulations easier for biochemistry, molecular biology, and cell biology.

Before doing any experiments, the tools required for this project need to be obtained. A sensitive and specific antibody for the zebra fish Nudt16 ortholog needs to be made and tested to make sure it recognizes *in vivo* levels of the Nudt16 protein. Morpholinos for testing the effects of deletion of Nudt16 must be obtained and tested. Morhpolinos are designed to inhibit translation of the mRNA of a gene of interest. They bind at the 5'UTR or at various splice sites blocking the ribosome from translation of the mRNA.

The first experiment is to inject morpholinos into zebra fish embryos which have been fertilized. Embryos will be monitored throughout development, noting different phenotypic effects in all parts of the organism. At each stage, the protein and mRNA levels of Nudt16 will also be tested using Western and Northern blots, respectively. Ideally, phenotypic defects will occur after a decrease in the Nudt16 RNA and/or protein is evident.

After determining a phenotypic effect, Northern blots or RT-PCR to test for levels of various nuclear limited RNAs would be tested. U8 snoRNA would be tested first because it has been shown bind to the *Xenopus* Nudt16 protein *in vivo* (Tomasevic and Peculis, 1999). Although there is no known expressed U8 snoRNA in zebra fish at this time, the RFAM database (http://rfam.sanger.ac.uk/) (Gardner et al., 2009; Griffiths-Jones, 2005; Griffiths-Jones et al., 2003; Griffiths-Jones et al., 2005) revealed potential U8 sequences in a variety of organisms, conserved throughout much of Eukaryota, including the zebra fish, anole, various mammals, X*enopus*, and other fish (MJT & BAP unpublished data). It is likely U8 snoRNA is expressed in zebra fish; these experiments

would verify the expression of U8 snoRNA in the zebra fish and also demonstrate whether there is an effect of Nudt16 depletion on U8 levels.

In addition to U8 snoRNA, other nuclear-limited RNAs that are transcribed by RNA Polymerase II would be followed. This would address the question of target RNAs. If a particular nuclear limited RNA changed expression levels based on Nudt16 expression levels, it is likely that RNA is a Nudt16 target RNA.

6.3: Effect of Nudt16 function in the presence of hormones

After monitoring the expression of Nudt16 and discovering the target RNAs, another question to answer is whether the protein expression levels change in response to hormones. It is known that sex steroids, or hormones, affect mRNA stability and proteins (Acconcia et al., 2006; Atwater et al., 1990; Farnell and Ing, 2002; Fretz and Spindler, 1999; Guhaniyogi and Brewer, 2001; Gustafson et al., 1987; Hsu et al., 2006; Ing, 2005; Katzenellenbogen, 1996; Leavitt et al., 1987; Leavitt et al., 1986; Lee et al., 2003; Liu et al., 2006a; Liu et al., 2006b; McInerney et al., 1996; Ntukidem et al., 2008; Pal et al., 1997; Staton et al., 2000; Takeda, 1988; Wu et al., 2006; Yanagihara et al., 2006; Zhang et al., 2006; Zhao et al., 2007). Of the organisms identified with the Nudt16 gene, most of them either demonstrate direct evidence for responsiveness to various hormonal signals or have an organism in the same family which responds to hormonal signals, including either estrogen or progesterone or both (Bayaa et al., 2000; Callard and Mak, 1985; Callard et al., 1985; Cuevas and Callard, 1992; Custodia-Lora et al., 2004; Hamad et al., 2007; Jin et al., 2007; Jorgensen et al., 2007; Kausch et al., 2008; Kimura et al., 1987; Kohler et al., 2007; Lee et al., 2003; Lowartz et al., 2003; Mak and Callard, 1985; Mullinix et al., 1976; Pang et al., 2008; Pudney et al., 1985; Puri and Toft, 1986; Roy et al., 2007; Sahu et al., 1981; Tada et al., 2007; Takeda et al., 2003; Yi et al., 2002). In cases in which neither exists, it is also possible a similar hormone is used to increase transcription of the Nudt16 gene.

In 2006, a paper was published demonstrating an increase in Nudt16 mRNA levels in endometirial epithelium in response to the estrus cycle in sheep (Ing et al.). Since then, no further evidence has appeared for hormonal regulation of the Nudt16 gene. Can the addition of estradiol or progesterone affect the expression of Nudt16? How much estradiol or progesterone do you need to see an effect of Nudt16 mRNA levels? How long does the treatment need to be? Answering these questions about the hormonal regulation of Nudt16 could yield new information on already well-studied hormone pathways.

The first experiments would need to demonstrate an effect of hormones on Nudt16 mRNA and protein expression. Once a significant change in protein level is detected, one can then look for specific effects on RNA. It would be feasible to use the target RNAs identified in the experiments outlined in Section 6.2 to determine whether there are effects in target RNA stability from alterations in Nudt16 expression.

6.4: Testing Nudt16 interactions with other proteins

Chapter 5 dealt with the identification of interacting proteins and showed three proteins present in co-immunoprecipitation samples with Flag-tagged Nudt16. To confirm that these proteins do interact with Nudt16 *in vivo*, co-localization studies would be useful.

In the co-immunoprecipiations, it is important to show that results from immunoprecipitations of the Actin, Annexin A2, and PRMT5 proteins include the Nudt16 protein. Positive results would support the proposed interactions between Nudt16 and the protein of interest; negative results would suggest the observed interactions were non-specific or unable to be observed using this method due to the low concentration of protein complexes.

Localization experiments would further support the observed interactions by showing that actin, annexin A2, PRMT5, and Nudt16 can all be in the nucleus and interact. The fluorescent antibodies for these proteins have been obtained and the immunofluoresence protocol has been established in the laboratory. Optimization of the conditions for the immunofluoresence studies will be necessary but the results could provide evidence that the proteins are together at the same time in the same place; it would support the hypothesis of interactions between the proteins.

Finally, although three proteins were identified, many other potential interacting proteins of Nudt16 were visible as faint bands on the gels (see Figures 5.2 and 5.3 pages 154-155). Further optimization of the co-immunoprecipitation protocol, such as using

more cells or excising bands from silver stained gels where those bands are more visible, may identify some of those fainter bands as putative Nudt16 interacting proteins.

6.5: Recapitulation, conclusions, and summary

The hypothesis stated on page 4 was as follows "Nudt16 is an evolutionarily conserved nuclear decapping protein that is involved in a nuclear 5'-3' RNA degradation pathway for nuclear-limited RNAs." The goal of this dissertation was to demonstrate that Nudt16 is conserved based upon both sequence analysis and functional decapping activity.

Chapter 2 showed the conservation of Nudt16 in Metazoa and discussed the phylogenetic relationship between the Nudt16 genes and the paralog syndesmos genes. The article proposed a nuclear RNA degradation pathway conserved throughout Metazoa and suggested that organisms without Nudt16 had evolved a different way to degrade nuclear RNAs.

Chapters 3 and 4 discussed the conservation of Nudt16's RNA-decapping function, which would be necessary for the proposed 5'-3' nuclear RNA degradation pathway. Chapter 3 studied the function of the *Xenopus* ortholog in detail. The homodimer was shown to be the functional enzymatic unit. The presence of two RNA binding sites on the protein was demonstrated. Binding of the first RNA did not have an effect on binding of a second RNA in the second site; catalysis could occur only in one site because of the inability of the protein to catalyze a second reaction. This chapter showed that the apparent lack of turn over was not due to inability of the protein to release the cap.

Conservation of Nudt16 decapping function was studied in Chapter 4. Orthologs of Nudt16 were functional for RNA binding and decapping activity, supporting the existence of a nuclear decapping pathway in these organisms. In addition, it was shown that Nudt16 could decap a variety of substrate RNAs in the presence of Mg^{2+}, Mn^{2+}, or Co^{2+}; the metal used affected the level of decapping activity. Pre-incubation of the Nudt16 protein with RNA was shown to have higher levels of decapping activity than pre-incubations of the protein alone or the protein with metal. RNA binding of all Nudt16 orthologs was demonstrated by UV-dependent RNA:protein crosslinks. Protein homodimerization of all Nudt16 orthologs was shown in experiments with chemical crosslinkers. Together these data showed the conservation of Nudt16 function, providing more evidence that Nudt16 could be involved in a nuclear RNA degradation pathway.

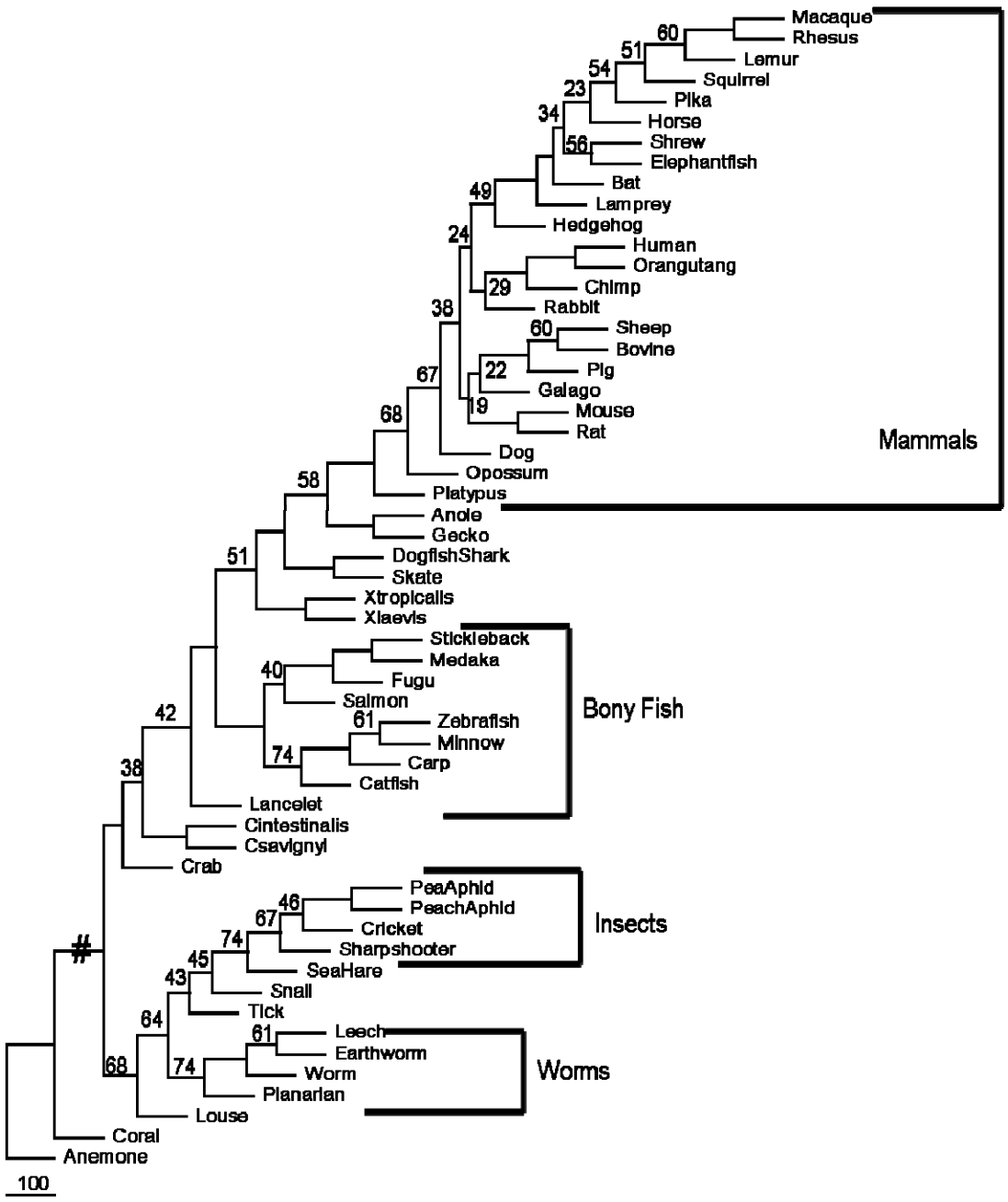
Part of the above hypothesis is that an entire 5'-3' RNA degradation pathway exists in the nucleus; Nudt16 is only one of the proteins that would be involved in this pathway. The experiments in Chapter 5 looked for potential *in vivo* interactors of Nudt16. Three proteins were identified as putative Nudt16 interacting proteins actin, Annexin A2, and PRMT5. Although still preliminary results, the co-immunoprecipitations can be used to guide other research that occurs in the Peculis laboratory.

The end of this document has come but the great thing is that in many ways, the end is a beginning. The experiments outlined in the above pages and chapters are a good place to start the next set of experiments. As each set of experiments yields new information, it will be added to the current models of Nudt16 function. The models will

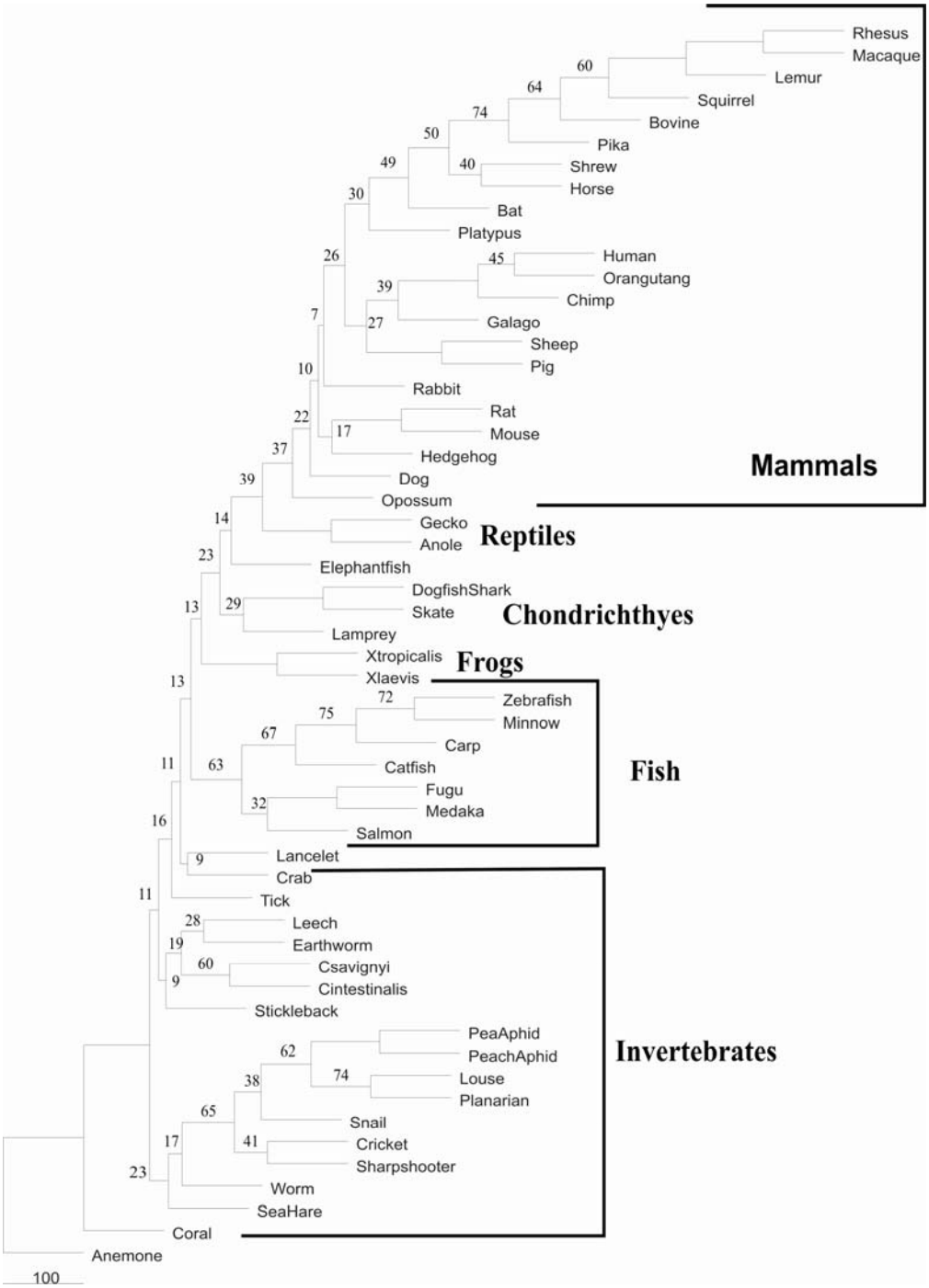
be refined and improved based on the new data. It is likely that as the models are refined and more data becomes available, more questions will arise which can be tested.

Appendix A: Supplementary Materials and Updated Information

The evolution of the Nudt16 gene is the focus of Chapter 2 (pages 13-61) of this dissertation. Supplementary Figures A.1-A.3 and Supplementary Table A.1 are from the original paper. However, since the publication of that article (Taylor and Peculis, 2008), other organisms with the Nudt16 gene have been identified. With the updated information, newer phylogenetic relationships were obtained and are depicted below in Figures A.4 and A.5. The information for each updated sequence is included in Table A.2.

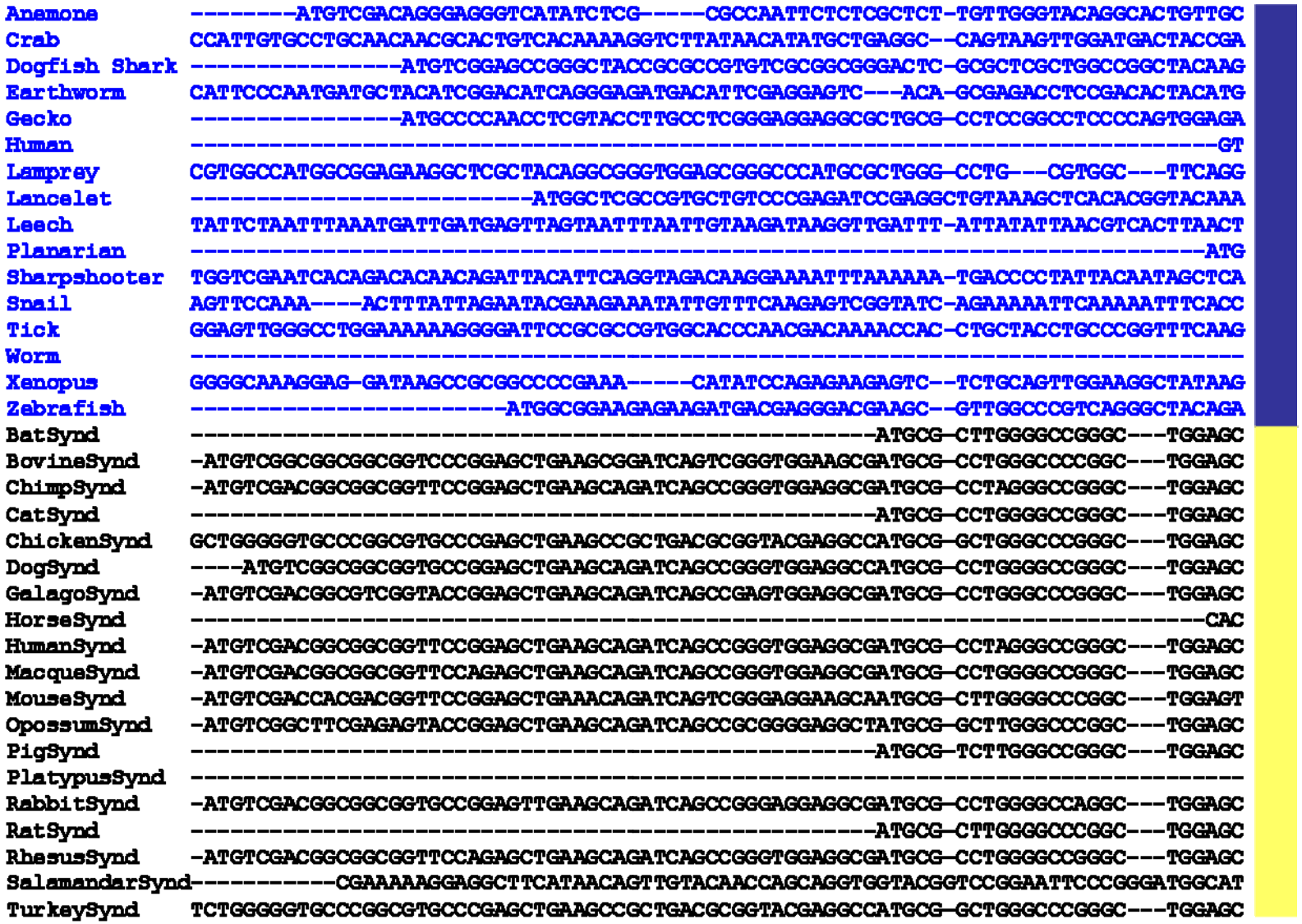


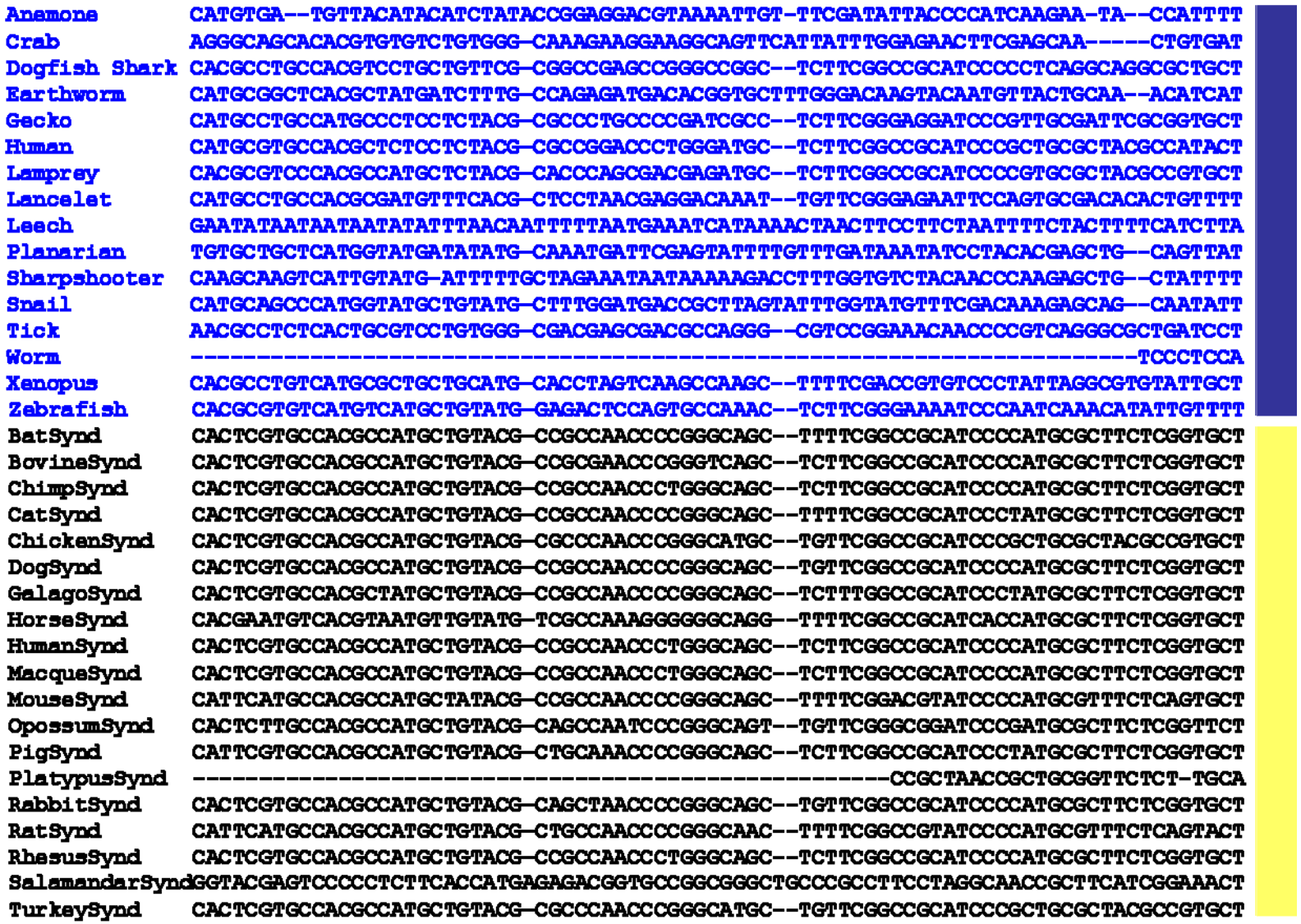
Supplementary Figure A.1: Phylogenetic analysis of amino acid sequence of Nudt16 orthologs. Amino acid sequences were aligned, graphed, and analyzed as described in the Methods Section. TreeView was used to visualize the tree. The numbers on each branch are the bootstrap values out of 100 trees; only bootstrap values under 80 are presented. The hash mark (#) indicates the branching between vertebrate and invertebrate sequences.



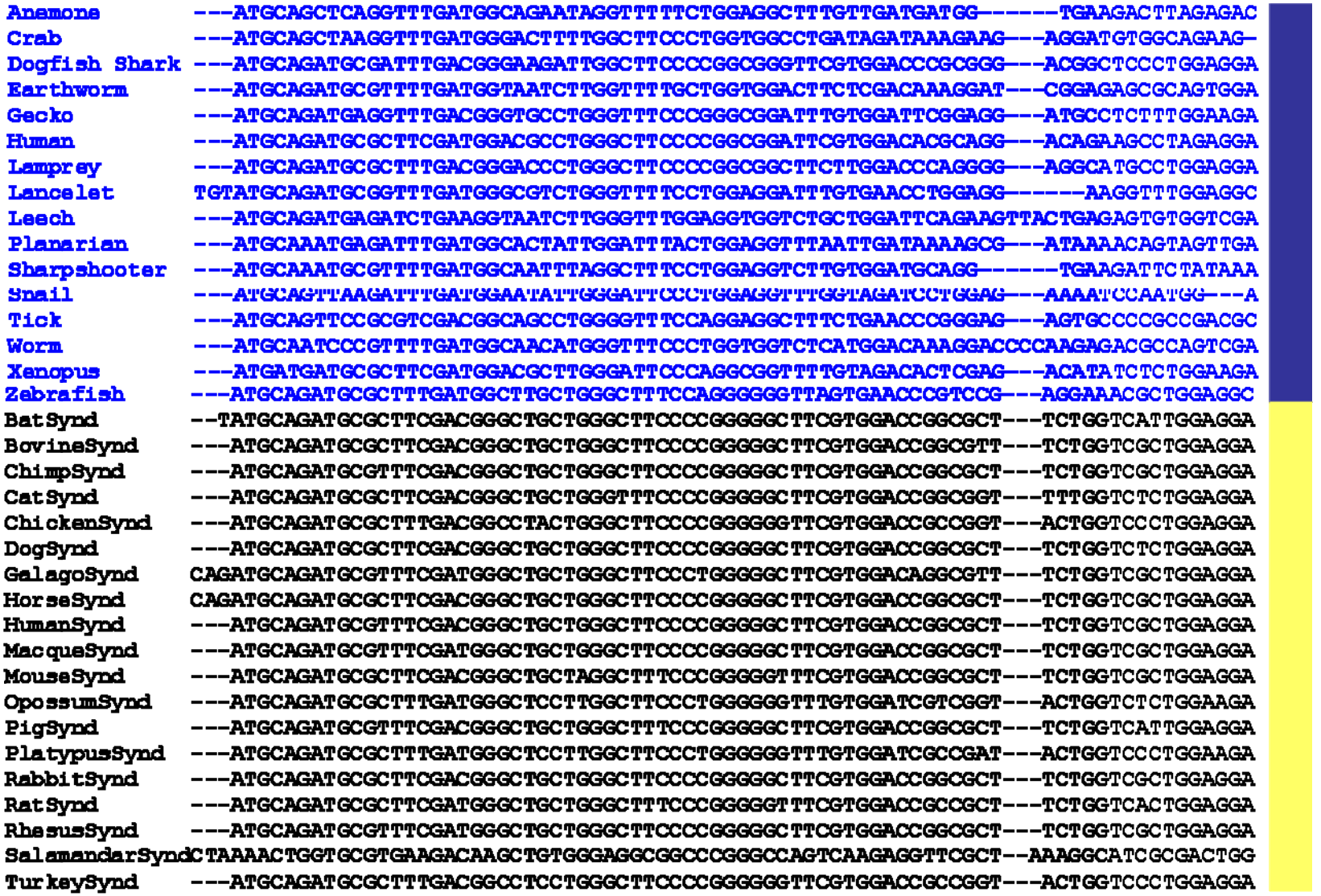
Supplementary Figure A.2: Phylogenetic analysis of nucleic acid sequence of Nudt16 orthologs. Nucleic acid sequences were aligned, analyzed, and graphed as described in the Methods section. PhyFi was used to visualize the tree; branch lengths are indicated by the ruler, and bootstrap values are included at each node. The Anemone sequence was used as the outgroup because of its proximity to a proposed common ancestor.

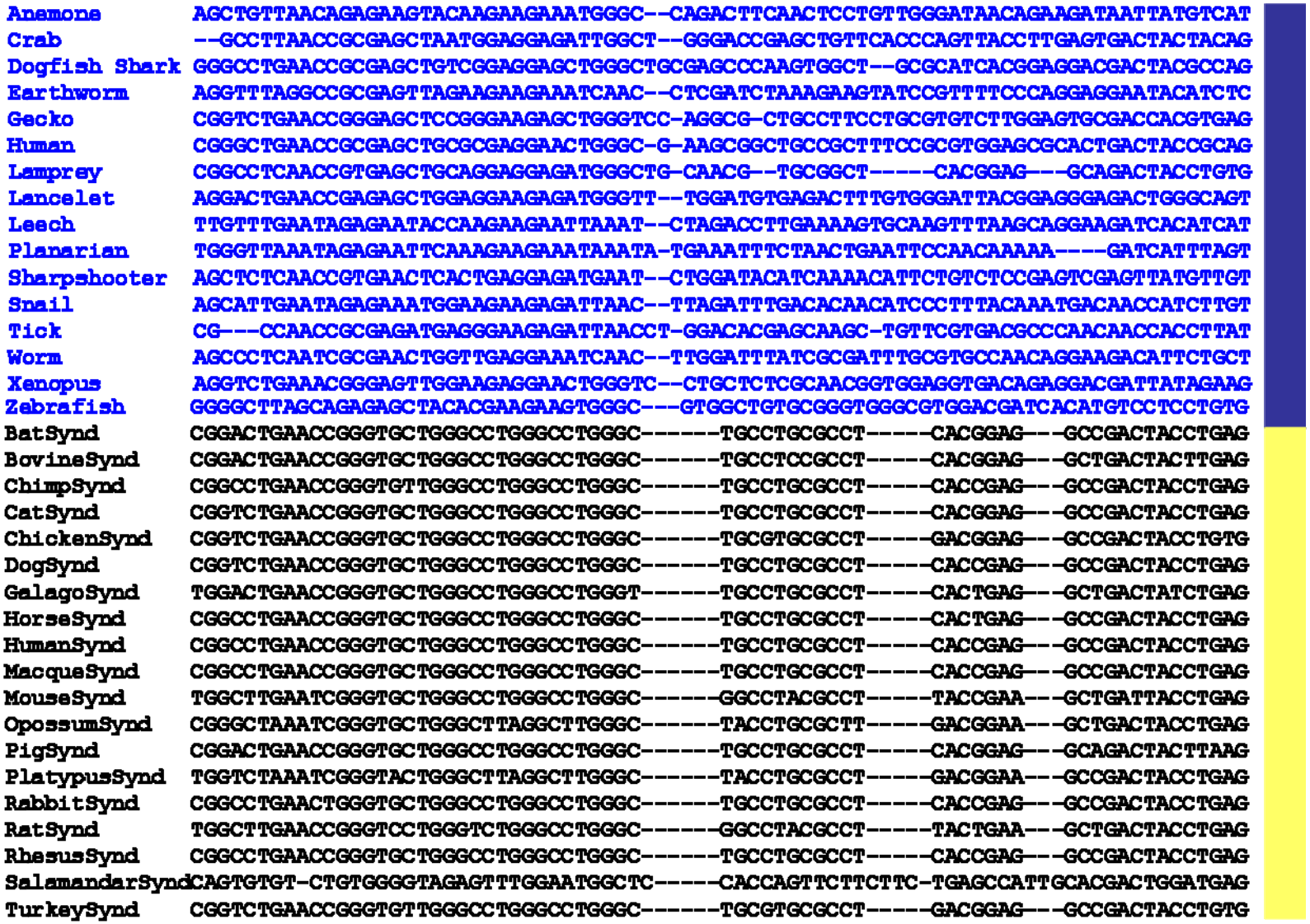
Supplementary Figure A.3: Nucleic acid sequence alignment for Nud16 and Syndesmos for selected organisms. Alignment of the majority of the open reading from for selected organisms, corresponding to those shown in Figure 2.3 (page 42) of the text. Blue vertical bar on the right indicates the Nud16 sequences, shown in blue text. The yellow vertical bar indicates Syndesmos sequences, provided in black text. Highlighted residues are positions that are better than 75% conserved among the organisms aligned here. The green box, red box and blue box indicate the nucleotide sequences corresponding to the amino acid sequences annotated identically in Figure 2.3 (page 42) in the text and generate the central core of the protein, the NUDIX domain and the interaction face of the proteins, respectively.

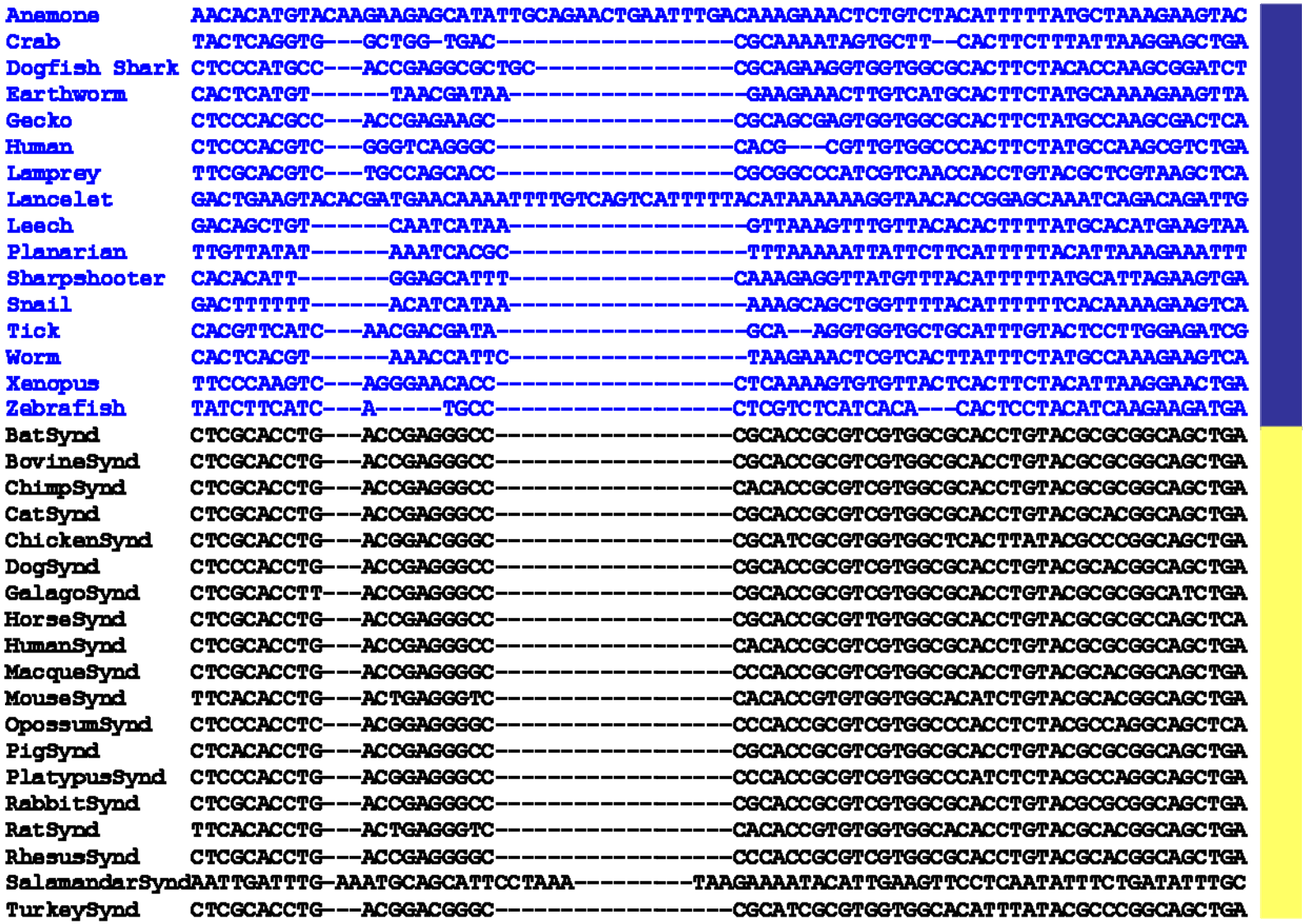

Anemone --------ATGTCGACAGGGAGGGTCATATCTCG-----CGCCAATTCTCTCGCTCT-TGTTGGGTACAGGCACTGTTGC
Crab CCATTGTGCCTGCAACAACGCACTGTCACAAAAGGTCTTATAACATATGCTGAGGC--CAGTAAGTTGGATGACTACCGA
Dogfish Shark ----------ATGTCGGAGCCGGGCTACCGCGCCGTGTCGCGGCGGGACTC-GCGCTCGCTGGCCGGCTACAAG
Earthworm CATTCCCAATGATGCTACATCGGACATCAGGGAGATGACATTCGAGGAGTC---ACA-GCGAGACCTCCGACACTACATG
Gecko ----------ATGCCCCAACCTCGTACCTTGCCTCGGGAGGAGGCGCTGCG-CCTCCGGCCTCCCCAGTGGAGA
Human ----------GT
Lamprey CGTGGCCATGGCGGAGAAGGCTCGCTACAGGCGGGTGGAGCGGGCCCATGCGCTGGG-CCTG---CGTGGC---TTCAGG
Lancelet ----------ATGGCTCGCCGTGCTGTCCCGAGATCCGAGGCTGTAAAGCTCACACGGTACAAA
Leech TATTCTAATTTAAATGATTGATGAGTTAGTAATTTAATTGTAAGATAAGGTTGATTT-ATTATATTAACGTCACTTAACT
Planarian ----------ATG
Sharpshooter TGGTCGAATCACAGACACAACAGATTACATTCAGGTAGACAAGGAAAATTTAAAAAA-TGACCCCTATTACAATAGCTCA
Snail AGTTCCAAA----ACTTTATTAGAATACGAAGAAATATTGTTTCAAGAGTCGGTATC-AGAAAAATTCAAAAATTTCACC
Tick GGAGTTGGGCCTGGAAAAAAGGGGATTCCGCGCCGTGGCACCCAACGACAAAACCAC-CTGCTACCTGCCCGGTTTCAAG
Worm ----------
Xenopus GGGGCAAAGGAG-GATAAGCCGCGGCCCCGAAA-----CATATCCAGAGAAGAGTC--TCTGCAGTTGGAAGGCTATAAG
Zebrafish ----------ATGGCGGAAGAGAAGATGACGAGGGACGAAGC--GTTGGCCCGTCAGGGCTACAGA
BatSynd ----------ATGCG-CTTGGGGCCGGGC---TGGAGC
BovineSynd -ATGTCGGCGGCGGCGGTCCCGGAGCTGAAGCGGATCAGTCGGGTGGAAGCGATGCG-CCTGGGCCCCGGC---TGGAGC
ChimpSynd -ATGTCGACGGCGGCGGTTCCGGAGCTGAAGCAGATCAGCCGGGTGGAGGCGATGCG-CCTAGGGCCGGGC---TGGAGC
CatSynd ----------ATGCG-CCTGGGGCCGGGC---TGGAGC
ChickenSynd GCTGGGGGGTGCCCGGCGTGCCCGAGCTGAAGCCGCTGACGCGGTACGAGGCCATGCG-GCTGGGCCCGGGC---TGGAGC
DogSynd ----ATGTCGGCGGCGGTGCCGGAGCTGAAGCAGATCAGCCGGGTGGAGGCCATGCG-CCTGGGGCCGGGC---TGGAGC
GalagoSynd -ATGTCGACGGCGTCGGTACCGGAGCTGAAGCAGATCAGCCGAGTGGAGGCGATGCG-CCTGGGCCCGGGC---TGGAGC
HorseSynd ----------CAC
HumanSynd -ATGTCGACGGCGGCGGTTCCGGAGCTGAAGCAGATCAGCCGGGTGGAGGCGATGCG-CCTAGGGCCGGGC---TGGAGC
MacqueSynd -ATGTCGACGGCGGCGGTTCCAGAGCTGAAGCAGATCAGCCGGGTGGAGGCGATGCG-CCTGGGGCCGGGC---TGGAGC
MouseSynd -ATGTCGACCACGACGGTTCCGGAGCTGAAACAGATCAGTCGGGAGGAAGCAATGCG-CTTGGGGCCCGGC---TGGAGT
OpossumSynd -ATGTCGGCTTCGAGAGTACCGGAGCTGAAGCAGATCAGCCGCGGGGAGGCTATGCG-GCTTGGGCCCGGC---TGGAGC
PigSynd ----------ATGCG-TCTTGGGCCGGGC---TGGAGC
PlatypusSynd ----------
RabbitSynd -ATGTCGACGGCGGCGGTGCCGGAGTTGAAGCAGATCAGCCGGGAGGAGGCGATGCG-CCTGGGGCCAGGC---TGGAGC
RatSynd ----------ATGCG-CTTGGGGCCCGGC---TGGAGC
RhesusSynd -ATGTCGACGGCGGCGGTTCCAGAGCTGAAGCAGATCAGCCGGGTGGAGGCGATGCG-CCTGGGGCCGGGC---TGGAGC
SalamandarSynd----------CGAAAAAGGAGGCTTCATAACAGTTGTACAACCAGCAGGTGGTACGGTCCGGAATTCCCGGGATGGCAT
TurkeySynd TCTGGGGGGTGCCCGGCGTGCCCGAGCTGAAGCCGCTGACGCGGTACGAGGCCATGCG-GCTGGGCCCGGGC---TGGAGC

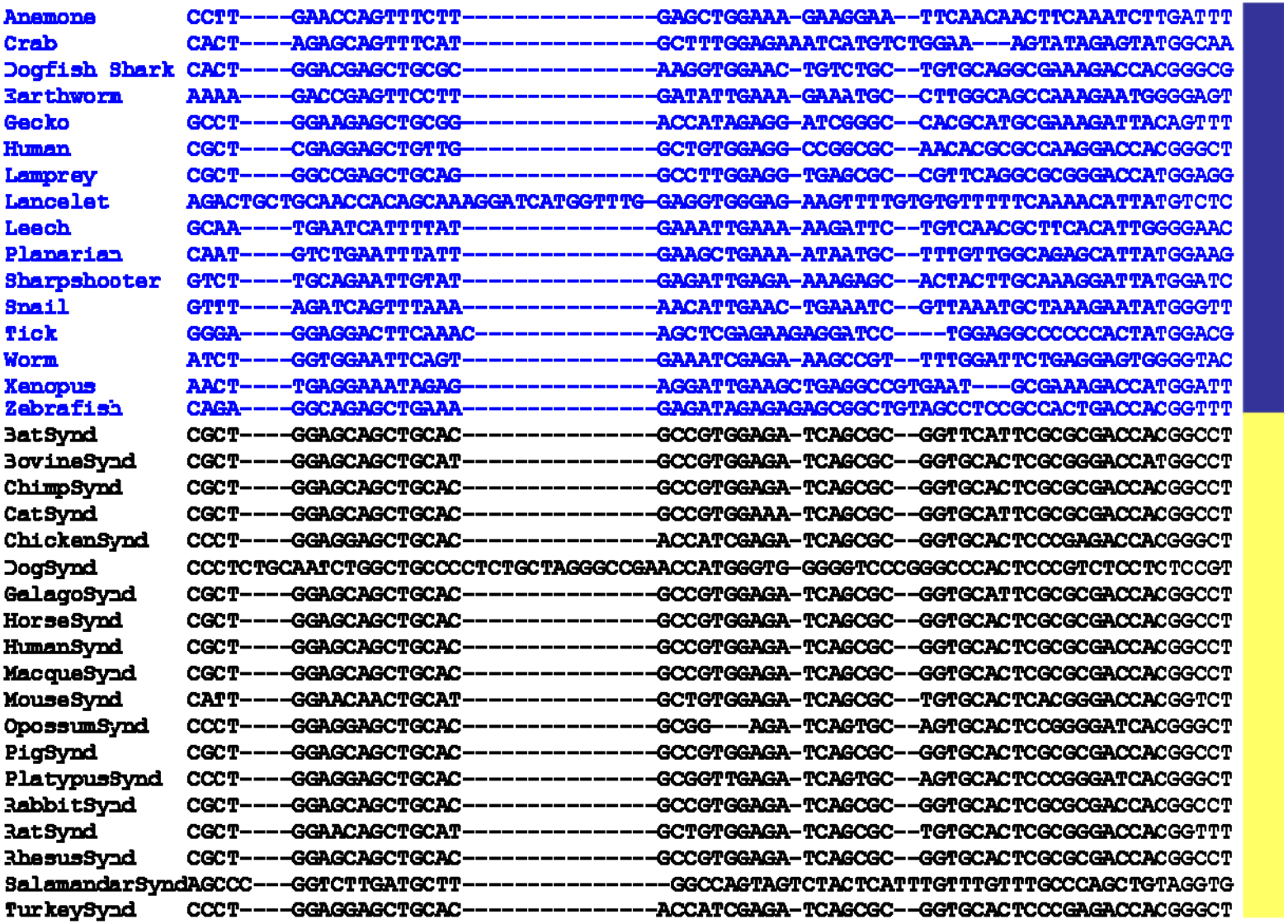

Anemone CATGTGA--TGTTACATACATCTATACCGGAGGACGTAAAATTGT-TTCGATATTACCCCATCAAGAA-TA--CCATTTT
Crab AGGGCAGCACACGTGTGTCTGTGGG-CAAAGAAGGAAGGCAGTTCATTATTTGGAGAACTTCGAGCAA-----CTGTGAT
Dogfish Shark CACGCCTGCCACGTCCTGCTGTTCG-CGGCCGAGCCGGGCCGGC--TCTTCGGCCGCATCCCCCTCAGGCAGGCGCTGCT
Earthworm CATGCGGCTCACGCTATGATCTTTG-CCAGAGATGACACGGTGCTTTGGGACAAGTACAATGTTACTGCAA--ACATCAT
Gecko CATGCCTGCCATGCCCTCCTCTACG-CGCCCTGCCCCGATCGCC--TCTTCGGGAGGATCCCGTTGCGATTCGCGGTGCT
Human CATGCGTGCCACGCTCTCCTCTACG-CGCCGGACCCTGGGATGC--TCTTCGGCCGCATCCCGCTGCGCTACGCCATACT
Lamprey CACGCGTCCCACGCCATGCTCTACG-CACCCAGCGACGAGATGC--TCTTCGGCCGCATCCCCGTGCGCTACGCCGTGCT
Lancelet CATGCCTGCCACGCGATGTTTCACG-CTCCTAACGAGGACAAAT--TGTTCGGGAGAATTCCAGTGCGACACACTGTTTT
Leech GAATATAATAATAATATATTTAACAATTTTTAATGAAATCATAAAACTAACTTCCTTCTAATTTTCTACTTTTCATCTTA
Planarian TGTGCTGCTCATGGTATGATATATG-CAAATGATTCGAGTATTTTGTTTGATAAATATCCTACACGAGCTG--CAGTTAT
Sharpshooter CAAGCAAGTCATTGTATG-ATTTTTGCTAGAAATAATAAAAAGACCTTTGGTGTCTACAACCCAAGAGCTG--CTATTTT
Snail CATGCAGCCCATGGTATGCTGTATG-CTTTGGATGACCGCTTAGTATTTGGTATGTTTCGACAAAGAGCAG--CAATATT
Tick AACGCCTCTCACTGCGTCCTGTGGG-CGACGAGCGACGCCAGGG--CGTCCGGAAACAACCCCGTCAGGGCGCTGATCCT
Worm ----------TCCCTCCA
Xenopus CACGCCTGTCATGCGCTGCTGCATG-CACCTAGTCAAGCCAAGC--TTTTCGACCGTGTCCCTATTAGGCGTGTATTGCT
Zebrafish CACGCGTGTCATGTCATGCTGTATG-GAGACTCCAGTGCCAAAC--TCTTCGGGAAAATCCCAATCAAACATATTGTTTT
BatSynd CACTCGTGCCACGCCATGCTGTACG-CCGCCAACCCCGGGCAGC--TTTTCGGCCGCATCCCCATGCGCTTCTCGGTGCT
BovineSynd CACTCGTGCCACGCCATGCTGTACG-CCGCGAACCCGGGTCAGC--TCTTCGGCCGCATCCCCATGCGCTTCTCGGTGCT
ChimpSynd CACTCGTGCCACGCCATGCTGTACG-CCGCCAACCCTGGGCAGC--TCTTCGGCCGCATCCCCATGCGCTTCTCGGTGCT
CatSynd CACTCGTGCCACGCCATGCTGTACG-CCGCCAACCCCGGGCAGC--TTTTCGGCCGCATCCCTATGCGCTTCTCGGTGCT
ChickenSynd CACTCGTGCCACGCCATGCTGTACG-CGCCCAACCCGGGCATGC--TGTTCGGCCGCATCCCGCTGCGCTACGCCGTGCT
DogSynd CACTCGTGCCACGCCATGCTGTACG-CCGCCAACCCCGGGCAGC--TGTTCGGCCGCATCCCCATGCGCTTCTCGGTGCT
GalagoSynd CACTCGTGCCACGCTATGCTGTACG-CCGCCAACCCCGGGCAGC--TCTTTGGCCGCATCCCTATGCGCTTCTCGGTGCT
HorseSynd CACGAATGTCACGTAATGTTGTATG-TCGCCAAAGGGGGGCAGG--TTTTCGGCCGCATCACCATGCGCTTCTCGGTGCT
HumanSynd CACTCGTGCCACGCCATGCTGTACG-CCGCCAACCCTGGGCAGC--TCTTCGGCCGCATCCCCATGCGCTTCTCGGTGCT
MacqueSynd CACTCGTGCCACGCCATGCTGTACG-CCGCCAACCCTGGGCAGC--TCTTCGGCCGCATCCCCATGCGCTTCTCGGTGCT
MouseSynd CATTCATGCCACGCCATGCTATACG-CCGCCAACCCCGGGCAGC--TTTTCGGACGTATCCCCATGCGTTTCTCAGTGCT
OpossumSynd CACTCTTGCCACGCCATGCTGTACG-CAGCCAATCCCGGGCAGT--TGTTCGGGCGGATCCCGATGCGCTTCTCGGTTCT
PigSynd CATTCGTGCCACGCCATGCTGTACG-CTGCAAACCCCGGGCAGC--TCTTCGGCCGCATCCCTATGCGCTTCTCGGTGCT
PlatypusSynd ----------CCGCTAACCGCTGCGGTTCTCT-TGCA
RabbitSynd CACTCGTGCCACGCCATGCTGTACG-CAGCTAACCCCGGGCAGC--TGTTCGGCCGCATCCCCATGCGCTTCTCGGTGCT
RatSynd CATTCATGCCACGCCATGCTGTACG-CTGCCAACCCCGGGCAAC--TTTTCGGCCGTATCCCCATGCGTTTCTCAGTACT
RhesusSynd CACTCGTGCCACGCCATGCTGTACG-CCGCCAACCCTGGGCAGC--TCTTCGGCCGCATCCCCATGCGCTTCTCGGTGCT
SalamandarSyndGGTACGAGTCCCCCTCTTCACCATGAGAGACGGTGCCGGCGGGCTGCCCGCCTTCCTAGGCAACCGCTTCATCGGAAACT
TurkeySynd CACTCGTGCCACGCCATGCTGTACG-CGCCCAACCCGGGCATGC--TGTTCGGCCGCATCCCGCTGCGCTACGCCGTGCT

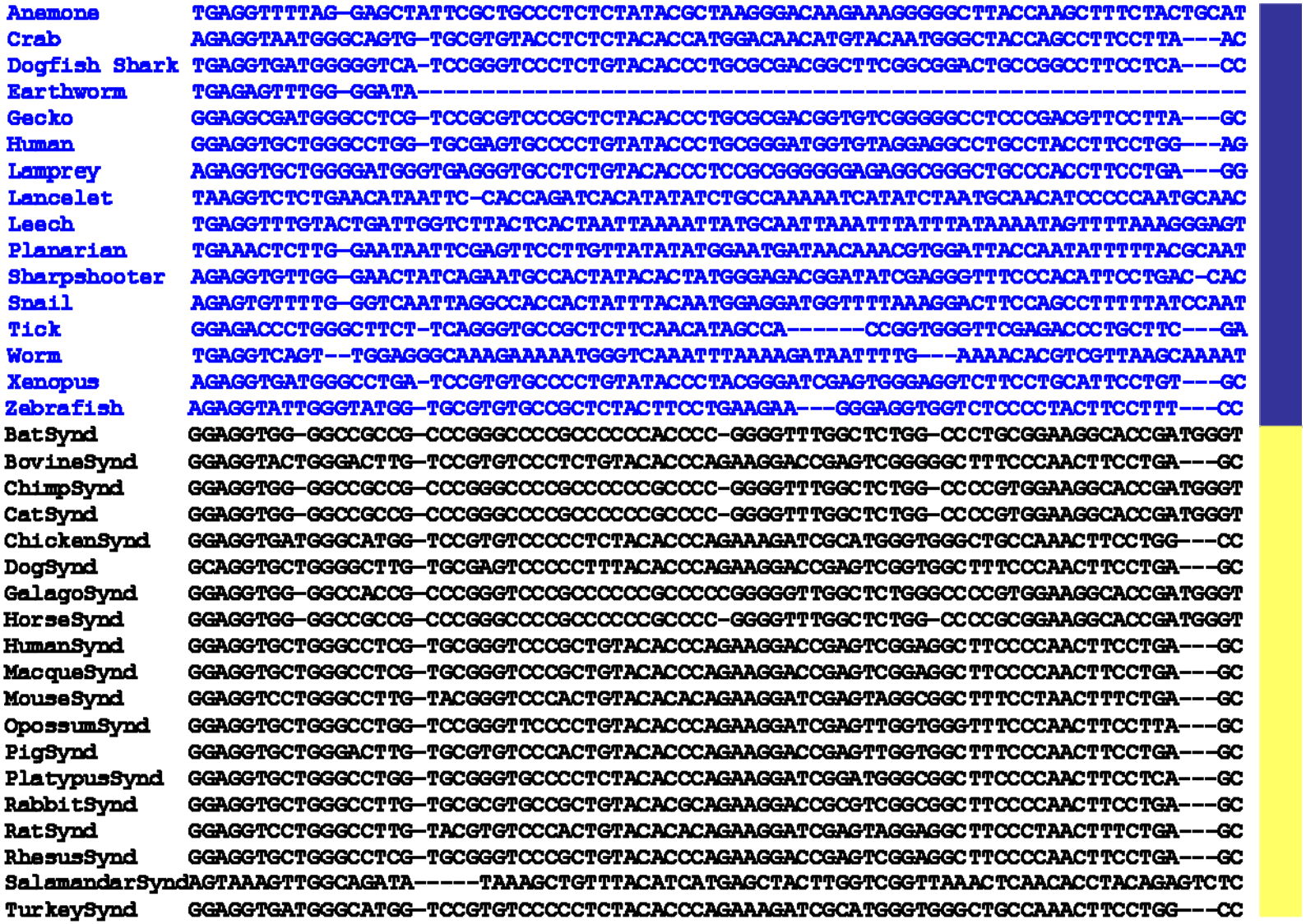

Anamone G----------
Crab G----------
Dogfish Shark G----------
Earthworm G----------
Gecko G----------
Human G----------
Lamprey G----------
Lancelet GGTAAGGCTTTTATTAGGGGGCGGG----------GCGACTAAAGTTTACATACATGTAACTAGACTCAAACCCTGGTGG
Leech G----------
Planarian G----------
Sharpshooter G----------
Snail G----------
Tick C
Worm G----------
Xenopus G----------
Zebrafish G----------
BatSynd GGTGAGGACGCGGGCGGGCGGATGT-----GGGCGGGACGTC--GCTCGGACCCGCGCCCGTGACCCGGCGCTCCC----
BovineSynd G----------
ChimpSynd G----------
CatSynd G----------
ChickenSynd G----------
DogSynd G----------
GalagoSynd GGGACGGGGGCGGGCAG-TGGCCGGGCCGGGGGGCACCGTGCCGACCCCGCTCCCGCGCCCGTGACCCGGCGCTCCCCTTG
HorseSynd GGTGAAGACGCGGGCGGGCGGCCGG--CGGGGGGGGGGCGCC--CCTCGGACCCACGCCCGTGACC--GCGCTGCC-TTG
HumanSynd G----------
MacqueSynd G----------
MouseSynd G----------
OpossumSynd G----------
PigSynd G----------
PlatypusSynd G----------
RabbitSynd G----------
RatSynd G----------
RhesusSynd G----------
SalamanderSyndCCCGGAGCCAGCTGCTCTACGGTTT----------GCGGGTC
TurkeySynd G----------

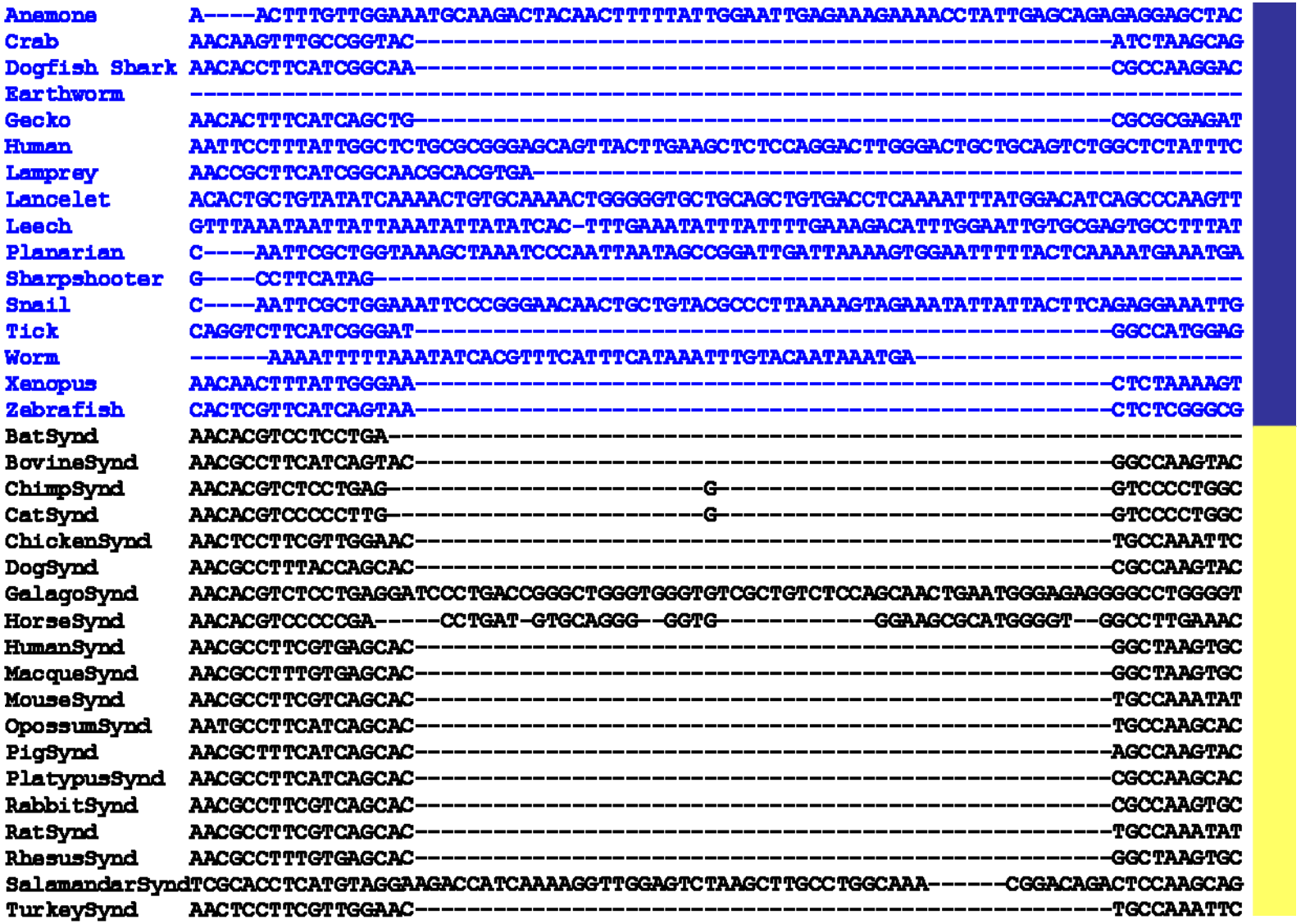

Anemone ---ATGCAGCTCAGGTTTGATGGCAGAATAGGTTTTTCTGGAGGCTTTGTTGATGATGG------TGAAGACTTAGAGAC
Crab ---ATGCAGCTAAGGTTTGATGGGACTTTTGGCTTCCCTGGTGGCCTGATAGATAAAGAAG---AGGATGTGGCAGAAG-
Dogfish Shark ---ATGCAGATGCGATTTGACGGGAAGATTGGCTTCCCCGGCGGGTTCGTGGACCCGCGGG---ACGGCTCCCTGGAGGA
Earthworm ---ATGCAGATGCGTTTTGATGGTAATCTTGGTTTTGCTGGTGGACTTCTCGACAAAGGAT---CGGAGAGCGCAGTGGA
Gecko ---ATGCAGATGAGGTTTGACGGGTGCCTGGGTTTCCCGGGCGGATTTGTGGATTCGGAGG---ATGCCTCTTTGGAAGA
Human ---ATGCAGATGCGCTTCGATGGACGCCTGGGCTTCCCCGGCGGATTCGTGGACACGCAGG---ACAGAAGCCTAGAGGA
Lamprey ---ATGCAGATGCGCTTTGACGGGACCCTGGGCTTCCCCGGCGGCTTCTTGGACCCAGGGG---AGGCATGCCTGGAGGA
Lancelet TGTATGCAGATGCGGTTTGATGGGCGTCTGGGTTTTCCTGGAGGATTTGTGAACCTGGAGG------AAGGTTTGGAGGC
Leech ---ATGCAGATGAGATCTGAAGGTAATCTTGGGTTTGGAGGTGGTCTGCTGGATTCAGAAGTTACTGAGAGTGTGGTCGA
Planarian ---ATGCAAATGAGATTTGATGGCACTATTGGATTTACTGGAGGTTTAATTGATAAAAGCG---ATAAAACAGTAGTTGA
Sharpshooter ---ATGCAAATGCGTTTTGATGGCAATTTAGGCTTTCCTGGAGGTCTTGTGGATGCAGG------TGAAGATTCTATAAA
Snail ---ATGCAGTTAAGATTTGATGGAATATTGGGATTCCCTGGAGGTTTGGTAGATCCTGGAG---AAAATCCAATGG---A
Tick ---ATGCAGTTCCGCGTCGACGGCAGCCTGGGGTTTCCAGGAGGCTTTCTGAACCCGGGAG---AGTGCCCCGCCGACGC
Worm ---ATGCAATCCCGTTTTGATGGCAACATGGGTTTCCCTGGTGGTCTCATGGACAAAGGACCCCAAGAGACGCCAGTCGA
Xenopus ---ATGATGATGCGCTTCGATGGACGCTTGGGATTCCCAGGCGGTTTTGTAGACACTCGAG---ACATATCTCTGGAAGA
Zebrafish ---ATGCAGATGCGCTTTGATGGCTTGCTGGGCTTTCCAGGGGGGTTAGTGAACCCGTCCG---AGGAAACGCTGGAGGC
BatSynd --TATGCAGATGCGCTTCGACGGGCTGCTGGGCTTCCCCGGGGGCTTCGTGGACCGGCGCT---TCTGGTCATTGGAGGA
BovineSynd ---ATGCAGATGCGCTTCGACGGGCTGCTGGGCTTCCCCGGGGGCTTCGTGGACCGGCGTT---TCTGGTCGCTGGAGGA
ChimpSynd ---ATGCAGATGCGTTTCGACGGGCTGCTGGGCTTCCCCGGGGGCTTCGTGGACCGGCGCT---TCTGGTCGCTGGAGGA
CatSynd ---ATGCAGATGCGCTTCGACGGGCTGCTGGGTTTCCCCGGGGGCTTCGTGGACCGGCGGT---TTTGGTCTCTGGAGGA
ChickenSynd ---ATGCAGATGCGCTTTGACGGCCTACTGGGCTTCCCCGGGGGGTTCGTGGACCGCCGGT---ACTGGTCCCTGGAGGA
DogSynd ---ATGCAGATGCGCTTCGACGGGCTGCTGGGCTTCCCCGGGGGCTTCGTGGACCGGCGCT---TCTGGTCTCTGGAGGA
GalagoSynd CAGATGCAGATGCGTTTCGATGGGCTGCTGGGCTTCCCTGGGGGCTTCGTGGACAGGCGTT---TCTGGTCGCTGGAGGA
HorseSynd CAGATGCAGATGCGCTTCGACGGGCTGCTGGGCTTCCCCGGGGGCTTCGTGGACCGGCGCT---TCTGGTCGCTGGAGGA
HumanSynd ---ATGCAGATGCGTTTCGACGGGCTGCTGGGCTTCCCCGGGGGCTTCGTGGACCGGCGCT---TCTGGTCGCTGGAGGA
MacqueSynd ---ATGCAGATGCGTTTCGATGGGCTGCTGGGCTTCCCCGGGGGCTTCGTGGACCGGCGCT---TCTGGTCGCTGGAGGA
MouseSynd ---ATGCAGATGCGCTTCGACGGGCTGCTAGGCTTTCCCGGGGGTTTCGTGGACCGGCGCT---TCTGGTCGCTGGAGGA
OpossumSynd ---ATGCAGATGCGCTTTGATGGGCTCCTTGGCTTCCCTGGGGGGTTTGTGGATCGTCGGT---ACTGGTCTCTGGAAGA
PigSynd ---ATGCAGATGCGTTTCGACGGGCTGCTGGGCTTTCCCGGGGGCTTCGTGGACCGGCGCT---TCTGGTCATTGGAGGA
PlatypusSynd ---ATGCAGATGCGCTTTGATGGGCTCCTTGGCTTCCCTGGGGGGTTTGTGGATCGCCGAT---ACTGGTCCCTGGAAGA
RabbitSynd ---ATGCAGATGCGCTTCGACGGGCTGCTGGGCTTCCCCGGGGGCTTCGTGGACCGGCGCT---TCTGGTCGCTGGAGGA
RatSynd ---ATGCAGATGCGCTTCGATGGGCTGCTGGGCTTTCCCGGGGGTTTCGTGGACCGCCGCT---TCTGGTCACTGGAGGA
RhesusSynd ---ATGCAGATGCGTTTCGATGGGCTGCTGGGCTTCCCCGGGGGCTTCGTGGACCGGCGCT---TCTGGTCGCTGGAGGA
SalamandarSynd CTAAAACTGGTGCGTGAAGACAAGCTGTGGGAGGCGGCCCGGGCCAGTCAAGAGGTTCGCT--AAAGGCATCGCGACTGG
TurkeySynd ---ATGCAGATGCGCTTTGACGGCCTCCTGGGCTTCCCCGGGGGGTTCGTGGACCGCCGGT---ACTGGTCCCTGGAGGA

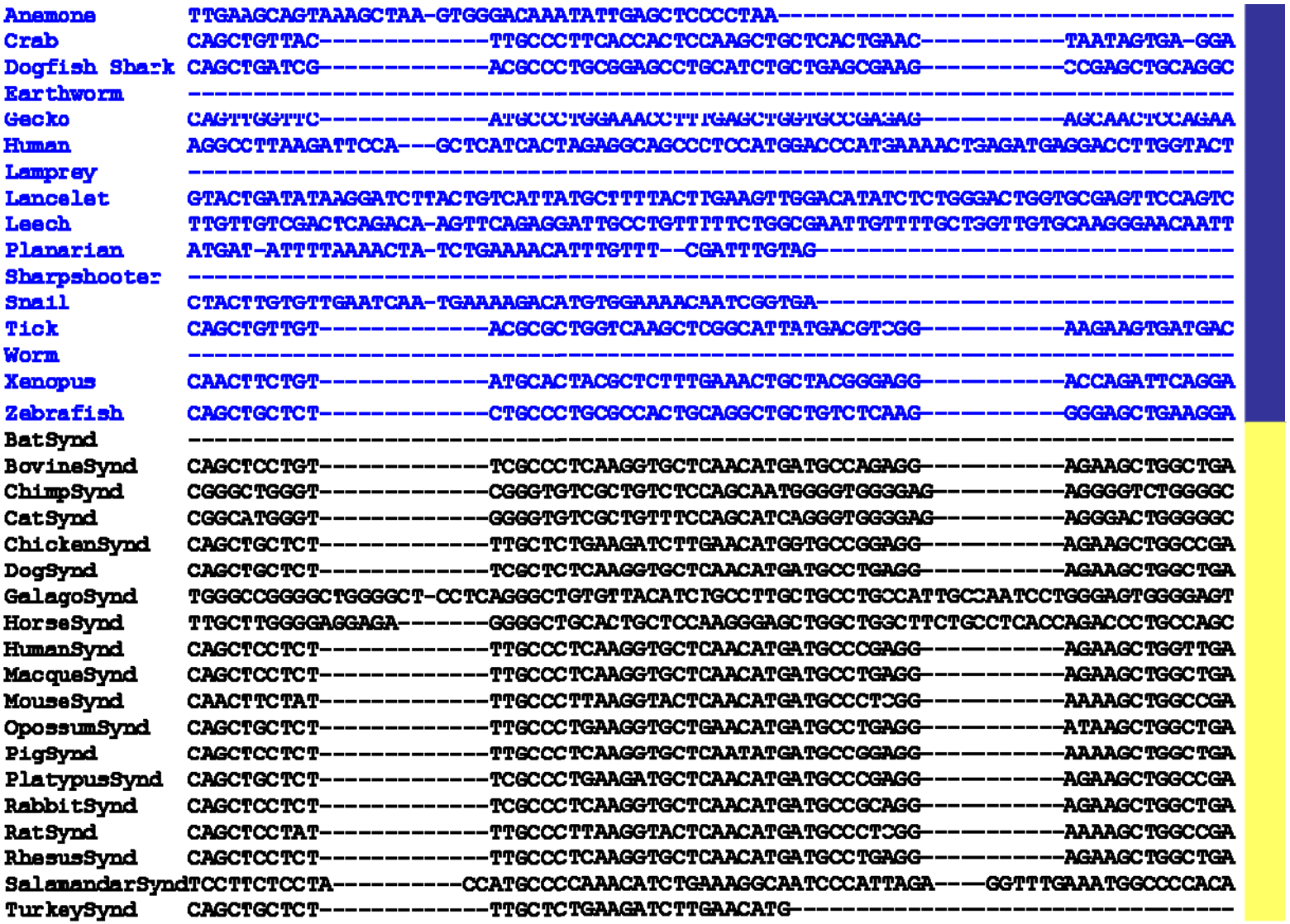

Anemone AGCTGTTAACAGAGAAGTACAAGAAGAAATGGGC--CAGACTTCAACTCCTGTTGGGATAACAGAAGATAATTATGTCAT
Crab --GCCTTAACCGCGAGCTAATGGAGGAGATTGGCT--GGGACCGAGCTGTTCACCCAGTTACCTTGAGTGACTACTACAG
Dogfish Shark GGGCCTGAACCGCGAGCTGTCGGAGGAGCTGGGCTGCGAGCCCAAGTGGCT--GCGCATCACGGAGGACGACTACGCCAG
Earthworm AGGTTTAGGCCGCGAGTTAGAAGAAGAAATCAAC--CTCGATCTAAAGAAGTATCCGTTTTCCCAGGAGGAATACATCTC
Gecko CGGTCTGAACCGGGAGCTCCGGGAAGAGCTGGGTCC-AGGCG-CTGCCTTCCTGCGTGTCTTGGAGTGCGACCACGTGAG
Human CGGGCTGAACCGCGAGCTGCGCGAGGAACTGGGC-G-AAGCGGCTGCCGCTTTCCGCGTGGAGCGCACTGACTACCGCAG
Lamprey CGGCCTCAACCGTGAGCTGCAGGAGGAGATGGGCTG-CAACG--TGCGGCT-----CACGGAG---GCAGACTACCTGTG
Lancelet AGGACTGAACCGAGAGCTGGAGGAAGAGATGGGTT--TGGATGTGAGACTTTGTGGGATTACGGAGGGAGACTGGGCAGT
Leech TTGTTTGAATAGAGAATACCAAGAAGAATTAAAT--CTAGACCTTGAAAAGTGCAAGTTTAAGCAGGAAGATCACATCAT
Planarian TGGGTTAAATAGAGAATTCAAAGAAGAAATAAATA-TGAAATTTCTAACTGAATTCCAACAAAAA----GATCATTTAGT
Sharpshooter AGCTCTCAACCGTGAACTCACTGAGGAGATGAAT--CTGGATACATCAAAACATTCTGTCTCCGAGTCGAGTTATGTTGT
Snail AGCATTGAATAGAGAAATGGAAGAAGAGATTAAC--TTAGATTTGACACAACATCCCTTTACAAATGACAACCATCTTGT
Tick CG---CCAACCGCGAGATGAGGGAAGAGATTAACCT-GGACACGAGCAAGC-TGTTCGTGACGCCCAACAACCACCTTAT
Worm AGCCCTCAATCGCGAACTGGTTGAGGAAATCAAC--TTGGATTTATCGCGATTTGCGTGCCAACAGGAAGACATTCTGCT
Xenopus AGGTCTGAAACGGGAGTTGGAAGAGGAACTGGGTC--CTGCTCTCGCAACGGTGGAGGTGACAGAGGACGATTATAGAAG
Zebrafish GGGGCTTAGCAGAGAGCTACACGAAGAAGTGGGC---GTGGCTGTGCGGGTGGGCGTGGACGATCACATGTCCTCCTGTG
BatSynd CGGACTGAACCGGGTGCTGGGCCTGGGCCTGGGC------TGCCTGCGCCT-----CACGGAG---GCCGACTACCTGAG
BovineSynd CGGACTGAACCGGGTGCTGGGCCTGGGCCTGGGC------TGCCTCCGCCT-----CACGGAG---GCTGACTACTTGAG
ChimpSynd CGGCCTGAACCGGGTGTTGGGCCTGGGCCTGGGC------TGCCTGCGCCT-----CACCGAG---GCCGACTACCTGAG
CatSynd CGGTCTGAACCGGGTGCTGGGCCTGGGCCTGGGC------TGCCTGCGCCT-----CACGGAG---GCCGACTACCTGAG
ChickenSynd CGGTCTGAACCGGGTGCTGGGCCTGGGCCTGGGC------TGCGTGCGCCT-----GACGGAG---GCCGACTACCTGTG
DogSynd CGGTCTGAACCGGGTGCTGGGCCTGGGCCTGGGC------TGCCTGCGCCT-----CACCGAG---GCCGACTACCTGAG
GalagoSynd TGGACTGAACCGGGTGCTGGGCCTGGGCCTGGGT------TGCCTGCGCCT-----CACTGAG---GCTGACTATCTGAG
HorseSynd CGGCCTGAACCGGGTGCTGGGCCTGGGCCTGGGC------TGCCTGCGCCT-----CACTGAG---GCCGACTACCTGAG
HumanSynd CGGCCTGAACCGGGTGCTGGGCCTGGGCCTGGGC------TGCCTGCGCCT-----CACCGAG---GCCGACTACCTGAG
MacqueSynd CGGCCTGAACCGGGTGCTGGGCCTGGGCCTGGGC------TGCCTGCGCCT-----CACCGAG---GCCGACTACCTGAG
MouseSynd TGGCTTGAATCGGGTGCTGGGCCTGGGCCTGGGC------GGCCTACGCCT-----TACCGAA---GCTGATTACCTGAG
OpossumSynd CGGGCTAAATCGGGTGCTGGGCTTAGGCTTGGGC------TACCTGCGCTT-----GACGGAA---GCTGACTACCTGAG
PigSynd CGGACTGAACCGGGTGCTGGGCCTGGGCCTGGGC------TGCCTGCGCCT-----CACGGAG---GCAGACTACTTAAG
PlatypusSynd TGGTCTAAATCGGGTACTGGGCTTAGGCTTGGGC------TACCTGCGCCT-----GACGGAA---GCCGACTACCTGAG
RabbitSynd CGGCCTGAACTGGGTGCTGGGCCTGGGCCTGGGC------TGCCTGCGCCT-----CACCGAG---GCCGACTACCTGAG
RatSynd TGGCTTGAACCGGGTCCTGGGTCTGGGCCTGGGC------GGCCTACGCCT-----TACTGAA---GCTGACTACCTGAG
RhesusSynd CGGCCTGAACCGGGTGCTGGGCCTGGGCCTGGGC------TGCCTGCGCCT-----CACCGAG---GCCGACTACCTGAG
SalamandarSynd CAGTGTGT-CTGTGGGGTAGAGTTTGGAATGGCTC-----CACCAGTTCTTCTTC-TGAGCCATTGCACGACTGGATGAG
TurkeySynd CGGTCTGAACCGGGTGTTGGGCCTGGGCCTGGGC------TGCGTGCGCCT-----GACGGAG---GCCGACTACCTGTG


Anemone AACACATGTACAAGAAGAGCATATTGCAGAACTGAATTTGACAAAGAAACTCTGTCTACATTTTTATGCTAAAGAAGTAC
Crab TACTCAGGTG---GCTGG-TGAC----------CGCAAAATAGTGCTT--CACTTCTTTATTAAGGAGCTGA
Dogfish Shark CTCCCATGCC---ACCGAGGCGCTGC----------CGCAGAAGGTGGTGGCGCACTTCTACACCAAGCGGATCT
Earthworm CACTCATGT------TAACGATAA----------GAAGAAACTTGTCATGCACTTCTATGCAAAAGAAGTTA
Gecko CTCCCACGCC---ACCGAGAAGC----------CGCAGCGAGTGGTGGCGCACTTCTATGCCAAGCGACTCA
Human CTCCCACGTC---GGGTCAGGGC----------CACG---CGTTGTGGCCCACTTCTATGCCAAGCGTCTGA
Lamprey TTCGCACGTC---TGCCAGCACC----------CGCGGCCCATCGTCAACCACCTGTACGCTCGTAAGCTCA
Lancelet GACTGAAGTACACGATGAACAAAATTTGTCAGTCATTTTACATAAAAAAGGTAACACCGGAGCAAATCAGACAGATTG
Leech GACAGCTGT------CAATCATAA----------GTTAAAGTTTGTTACACACTTTTATGCACATGAAGTAA
Planarian TTGTTATAT------AAATCACGC----------TTTAAAAATTATTCTTCATTTTTACATTAAAGAAATTT
Sharpshooter CACACATT-------GGAGCATTT----------CAAAGAGGTTATGTTTACATTTTTATGCATTAGAAGTGA
Snail GACTTTTTT------ACATCATAA----------AAAGCAGCTGGTTTTACATTTTTTCACAAAAGAAGTCA
Tick CACGTTCATC---AACGACGATA----------GCA--AGGTGGTGCTGCATTTGTACTCCTTGGAGATCG
Worm CACTCACGT------AAACCATTC----------TAAGAAACTCGTCACTTATTTCTATGCCAAAGAAGTCA
Xenopus TTCCCAAGTC---AGGGAACACC----------CTCAAAAGTGTGTTACTCACTTCTACATTAAGGAACTGA
Zebrafish TATCTTCATC---A-----TGCC----------CTCGTCTCATCACA---CACTCCTACATCAAGAAGATGA
BatSynd CTCGCACCTG---ACCGAGGGCC----------CGCACCGCGTCGTGGCGCACCTGTACGCGCGGCAGCTGA
BovineSynd CTCGCACCTG---ACCGAGGGCC----------CGCACCGCGTCGTGGCGCACCTGTACGCGCGGCAGCTGA
ChimpSynd CTCGCACCTG---ACCGAGGGCC----------CACACCGCGTCGTGGCGCACCTGTACGCGCGGCAGCTGA
CatSynd CTCGCACCTG---ACCGAGGGCC----------CGCACCGCGTCGTGGCGCACCTGTACGCACGGCAGCTGA
ChickenSynd CTCGCACCTG---ACGGACGGGC----------CGCATCGCGTGGTGGCTCACTTATACGCCCGGCAGCTGA
DogSynd CTCCCACCTG---ACCGAGGGCC----------CGCACCGCGTCGTGGCGCACCTGTACGCACGGCAGCTGA
GalagoSynd CTCGCACCTT---ACCGAGGGCC----------CGCACCGCGTCGTGGCGCACCTGTACGCGCGGCATCTGA
HorseSynd CTCGCACCTG---ACCGAGGGCC----------CGCACCGCGTTGTGGCGCACCTGTACGCGCGCCAGCTCA
HumanSynd CTCGCACCTG---ACCGAGGGCC----------CACACCGCGTCGTGGCGCACCTGTACGCGCGGCAGCTGA
MacqueSynd CTCGCACCTG---ACCGAGGGGC----------CCCACCGCGTCGTGGCGCACCTGTACGCACGGCAGCTGA
MouseSynd TTCACACCTG---ACTGAGGGTC----------CACACCGTGTGGTGGCACATCTGTACGCACGGCAGCTGA
OpossumSynd CTCCCACCTC---ACGGAGGGGC----------CCCACCGCGTCGTGGCCCACCTCTACGCCAGGCAGCTCA
PigSynd CTCACACCTG---ACCGAGGGCC----------CGCACCGCGTCGTGGCGCACCTGTACGCGCGGCAGCTGA
PlatypusSynd CTCCCACCTG---ACGGAGGGCC----------CCCACCGCGTCGTGGCCCATCTCTACGCCAGGCAGCTGA
RabbitSynd CTCGCACCTG---ACCGAGGGCC----------CGCACCGCGTCGTGGCGCACCTGTACGCGCGGCAGCTGA
RatSynd TTCACACCTG---ACTGAGGGTC----------CACACCGTGTGGTGGCACACCTGTACGCACGGCAGCTGA
RhesusSynd CTCGCACCTG---ACCGAGGGGC----------CCCACCGCGTCGTGGCGCACCTGTACGCACGGCAGCTGA
SalamandarSyndAATTGATTTG-AAATGCAGCATTCCTAAA---------TAAGAAAATACATTGAAGTTCCTCAATATTTCTGATATTTGC
TurkeySynd CTCGCACCTG---ACGGACGGGC----------CGCATCGCGTGGTGGCACATTTATACGCCCGGCAGCTGA


Anemone CCTT----GAACCAGTTTCTT--------------GAGCTGGAAA-GAAGGAA--TTCAACAACTTCAAATCTTGATTT
Crab CACT----AGAGCAGTTTCAT--------------GCTTTGGAGAAATCATGTCTGGAA---AGTATAGAGTATGGCAA
Dogfish Shark CACT----GGACGAGCTGCGC--------------AAGGTGGAAC-TGTCTGC--TGTGCAGGCGAAAGACCACGGGCG
Earthworm AAAA----GACCGAGTTCCTT--------------GATATTGAAA-GAAATGC--CTTGGCAGCCAAAGAATGGGGAGT
Gecko GCCT----GGAAGAGCTGCGG--------------ACCATAGAGG-ATCGGGC--CACGCATGCGAAAGATTACAGTTT
Human CGCT----CGAGGAGCTGTTG--------------GCTGTGGAGG-CCGGCGC--AACACGCGCCAAGGACCACGGGCT
Lamprey CGCT----GGCCGAGCTGCAG--------------GCCTTGGAGG-TGAGCGC--CGTTCAGGCGCGGGACCATGGAGG
Lancelet AGACTGCTGCAACCACAGCAAAGGATCATGGTTTG-GAGGTGGGAG-AAGTTTTGTGTGTTTTTCAAAACATTATGTCTC
Leech GCAA----TGAATCATTTTAT--------------GAAATTGAAA-AAGATTC--TGTCAACGCTTCACATTGGGGAAC
Planarian CAAT----GTCTGAATTTATT--------------GAAGCTGAAA-ATAATGC--TTTGTTGGCAGAGCATTATGGAAG
Sharpshooter GTCT----TGCAGAATTGTAT--------------GAGATTGAGA-AAAGAGC--ACTACTTGCAAAGGATTATGGATC
Snail GTTT----AGATCAGTTTAAA--------------AACATTGAAC-TGAAATC--GTTAAATGCTAAAGAATATGGGTT
Tick GGGA----GGAGGACTTCAAAC--------------AGCTCGAGAAGAGGATCC----TGGAGGCCCCCCACTATGGACG
Worm ATCT----GGTGGAATTCAGT--------------GAAATCGAGA-AAGCCGT--TTTGGATTCTGAGGAGTGGGGTAC
Xenopus AACT----TGAGGAAATAGAG--------------AGGATTGAAGCTGAGGCCGTGAAT---GCGAAAGACCATGGATT
Zebrafish CAGA----GGCAGAGCTGAAA--------------GAGATAGAGAGAGCGGCTGTAGCCTCCGCCACTGACCACGGTTT
BatSynd CGCT----GGAGCAGCTGCAC--------------GCCGTGGAGA-TCAGCGC--GGTTCATTCGCGCGACCACGGCCT
BovineSynd CGCT----GGAGCAGCTGCAT--------------GCCGTGGAGA-TCAGCGC--GGTGCACTCGCGGGACCATGGCCT
ChimpSynd CGCT----GGAGCAGCTGCAC--------------GCCGTGGAGA-TCAGCGC--GGTGCACTCGCGCGACCACGGCCT
CatSynd CGCT----GGAGCAGCTGCAC--------------GCCGTGGAAA-TCAGCGC--GGTGCATTCGCGCGACCACGGCCT
ChickenSynd CCCT----GGAGGAGCTGCAC--------------ACCATCGAGA-TCAGCGC--GGTGCACTCCCGAGACCACGGGCT
DogSynd CCCTCTGCAATCTGGCTGCCCCTCTGCTAGGGCCGAACCATGGGTG-GGGGTCCCGGGCCCACTCCCGTCTCCTCTCCGT
GalagoSynd CGCT----GGAGCAGCTGCAC--------------GCCGTGGAGA-TCAGCGC--GGTGCATTCGCGCGACCACGGCCT
HorseSynd CGCT----GGAGCAGCTGCAC--------------GCCGTGGAGA-TCAGCGC--GGTGCACTCGCGCGACCACGGCCT
HumanSynd CGCT----GGAGCAGCTGCAC--------------GCCGTGGAGA-TCAGCGC--GGTGCACTCGCGCGACCACGGCCT
MacqueSynd CGCT----GGAGCAGCTGCAC--------------GCCGTGGAGA-TCAGCGC--GGTGCACTCGCGCGACCACGGCCT
MouseSynd CATT----GGAACAACTGCAT--------------GCTGTGGAGA-TCAGCGC--TGTGCACTCACGGGACCACGGTCT
OpossumSynd CCCT----GGAGGAGCTGCAC--------------GCGG---AGA-TCAGTGC--AGTGCACTCCGGGGATCACGGGCT
PigSynd CGCT----GGAGCAGCTGCAC--------------GCCGTGGAGA-TCAGCGC--GGTGCACTCGCGCGACCACGGCCT
PlatypusSynd CCCT----GGAGGAGCTGCAC--------------GCGGTTGAGA-TCAGTGC--AGTGCACTCCCGGGATCACGGGCT
RabbitSynd CGCT----GGAGCAGCTGCAC--------------GCCGTGGAGA-TCAGCGC--GGTGCACTCGCGCGACCACGGCCT
RatSynd CGCT----GGAACAGCTGCAT--------------GCTGTGGAGA-TCAGCGC--TGTGCACTCGCGGGACCACGGTTT
RhesusSynd CGCT----GGAGCAGCTGCAC--------------GCCGTGGAGA-TCAGCGC--GGTGCACTCGCGCGACCACGGCCT
SalamanderSyndAGCCC---GGTCTTGATGCTT--------------GGCCAGTAGTCTACTCATTTGTTTGTTTGCCCAGCTGTAGGTG
TurkeySynd CCCT----GGAGGAGCTGCAC--------------ACCATCGAGA-TCAGCGC--GGTGCACTCCCGAGACCACGGGCT


Anemone TGAGGTTTTAG-GAGCTATTCGCTGCCCTCTCTATACGCTAAGGGACAAGAAAGGGGGCTTACCAAGCTTTCTACTGCAT
Crab AGAGGTAATGGGCAGTG-TGCGTGTACCTCTCTACACCATGGACAACATGTACAATGGGCTACCAGCCTTCCTTA---AC
Dogfish Shark TGAGGTGATGGGGGTCA-TCCGGGTCCCTCTGTACACCCTGCGCGACGGCTTCGGCGGACTGCCGGCCTTCCTCA---CC
Earthworm TGAGAGTTTGG-GGATA----------
Gecko GGAGGCGATGGGCCTCG-TCCGCGTCCCGCTCTACACCCTGCGCGACGGTGTCGGGGGCCTCCCGACGTTCCTTA---GC
Human GGAGGTGCTGGGCCTGG-TGCGAGTGCCCCTGTATACCCTGCGGGATGGTGTAGGAGGCCTGCCTACCTTCCTGG---AG
Lamprey AGAGGTGCTGGGGATGGGTGAGGGTGCCTCTGTACACCCTCCGCGGGGGGAGAGGCGGGCTGCCCACCTTCCTGA---GG
Lancelet TAAGGTCTCTGAACATAATTC-CACCAGATCACATATATCTGCCAAAAATCATATCTAATGCAACATCCCCCAATGCAAC
Leech TGAGGTTTGTACTGATTGGTCTTACTCACTAATTAAAATTATGCAATTAAATTTATTTATAAAATAGTTTTAAAGGGAGT
Planarian TGAAACTCTTG-GAATAATTCGAGTTCCTTGTTATATATGGAATGATAACAAACGTGGATTACCAATATTTTTACGCAAT
Sharpshooter AGAGGTGTTGG-GAACTATCAGAATGCCACTATACACTATGGGAGACGGATATCGAGGGTTTCCCACATTCCTGAC-CAC
Snail AGAGTGTTTTG-GGTCAATTAGGCCACCACTATTTACAATGGAGGATGGTTTTAAAGGACTTCCAGCCTTTTTATCCAAT
Tick GGAGACCCTGGGCTTCT-TCAGGGTGCCGCTCTTCAACATAGCCA------CCGGTGGGTTCGAGACCCTGCTTC---GA
Worm TGAGGTCAGT--TGGAGGGCAAAGAAAAATGGGTCAAATTTAAAAGATAATTTTG---AAAACACGTCGTTAAGCAAAAT
Xenopus AGAGGTGATGGGCCTGA-TCCGTGTGCCCCTGTATACCCTACGGGATCGAGTGGGAGGTCTTCCTGCATTCCTGT---GC
Zebrafish AGAGGTATTGGGTATGG-TGCGTGTGCCGCTCTACTTCCTGAAGAA---GGGAGGTGGTCTCCCCTACTTCCTTT---CC
BatSynd GGAGGTGG-GGCCGCCG-CCCGGGCCCCGCCCCCCACCCC-GGGGTTTGGCTCTGG-CCCTGCGGAAGGCACCGATGGGT
BovineSynd GGAGGTACTGGGACTTG-TCCGTGTCCCTCTGTACACCCAGAAGGACCGAGTCGGGGGCTTTCCCAACTTCCTGA---GC
ChimpSynd GGAGGTGG-GGCCGCCG-CCCGGGCCCCGCCCCCCGCCCC-GGGGTTTGGCTCTGG-CCCCGTGGAAGGCACCGATGGGT
CatSynd GGAGGTGG-GGCCGCCG-CCCGGGCCCCGCCCCCCGCCCC-GGGGTTTGGCTCTGG-CCCCGTGGAAGGCACCGATGGGT
ChickenSynd GGAGGTGATGGGCATGG-TCCGTGTCCCCCTCTACACCCAGAAAGATCGCATGGGTGGGCTGCCAAACTTCCTGG---CC
DogSynd GCAGGTGCTGGGGCTTG-TGCGAGTCCCCCTTTACACCCAGAAGGACCGAGTCGGTGGCTTTCCCAACTTCCTGA---GC
GalagoSynd GGAGGTGG-GGCCACCG-CCCGGGTCCCGCCCCCCGCCCCCGGGGGTTGGCTCTGGGCCCCGTGGAAGGCACCGATGGGT
HorseSynd GGAGGTGG-GGCCGCCG-CCCGGGCCCCGCCCCCCGCCCC-GGGGTTTGGCTCTGG-CCCCGCGGAAGGCACCGATGGGT
HumanSynd GGAGGTGCTGGGCCTCG-TGCGGGTCCCGCTGTACACCCAGAAGGACCGAGTCGGAGGCTTCCCCAACTTCCTGA---GC
MacqueSynd GGAGGTGCTGGGCCTCG-TGCGGGTCCCGCTGTACACCCAGAAGGACCGAGTCGGAGGCTTCCCCAACTTCCTGA---GC
MouseSynd GGAGGTCCTGGGCCTTG-TACGGGTCCCACTGTACACACAGAAGGATCGAGTAGGCGGCTTTCCTAACTTTCTGA---GC
OpossumSynd GGAGGTGCTGGGCCTGG-TCCGGGTTCCCCTGTACACCCAGAAGGATCGAGTTGGTGGGTTTCCCAACTTCCTTA---GC
PigSynd GGAGGTGCTGGGACTTG-TGCGTGTCCCACTGTACACCCAGAAGGACCGAGTTGGTGGCTTTCCCAACTTCCTGA---GC
PlatypusSynd GGAGGTGCTGGGCCTGG-TGCGGGTGCCCCTCTACACCCAGAAGGATCGGATGGGCGGCTTCCCCAACTTCCTCA---GC
RabbitSynd GGAGGTGCTGGGCCTTG-TGCGCGTGCCGCTGTACACGCAGAAGGACCGCGTCGGCGGCTTCCCCAACTTCCTGA---GC
RatSynd GGAGGTCCTGGGCCTTG-TACGTGTCCCACTGTACACACAGAAGGATCGAGTAGGAGGCTTCCCTAACTTTCTGA---GC
RhesusSynd GGAGGTGCTGGGCCTCG-TGCGGGTCCCGCTGTACACCCAGAAGGACCGAGTCGGAGGCTTCCCCAACTTCCTGA---GC
SalamandarSyndAGTAAAGTTGGCAGATA-----TAAAGCTGTTTACATCATGAGCTACTTGGTCGGTTAAACTCAACACCTACAGAGTCTC
TurkeySynd GGAGGTGATGGGCATGG-TCCGTGTCCCCCTCTACACCCAGAAAGATCGCATGGGTGGGCTGCCAAACTTCCTGG---CC


Anemone A----ACTTTGTTGGAAATGCAAGACTACAACTTTTTATTGGAATTGAGAAAGAAAACCTATTGAGCAGAGAGGAGCTAC
Crab AACAAGTTTGCCGGTAC----------ATCTAAGCAG
Dogfish Shark AACACCTTCATCGGCAA----------CGCCAAGGAC
Earthworm ----------
Gecko AACACTTTCATCAGCTG----------CGCGCGAGAT
Human AATTCCTTTATTGGCTCTGCGCGGGAGCAGTTACTTGAAGCTCTCCAGGACTTGGGACTGCTGCAGTCTGGCTCTATTTC
Lamprey AACCGCTTCATCGGCAACGCACGTGA----------
Lancelet ACACTGCTGTATATCAAAACTGTGCAAAACTGGGGGTGCTGCAGCTGTGACCTCAAAATTTATGGACATCAGCCCAAGTT
Leech GTTTAAATAATTATTAAATATTATATCAC-TTTGAAATATTTATTTTGAAAGACATTTGGAATTGTGCGAGTGCCTTTAT
Planarian C----AATTCGCTGGTAAAGCTAAATCCCAATTAATAGCCGGATTGATTAAAAGTGGAATTTTTACTCAAAATGAAATGA
Sharpshooter G----CCTTCATAG----------
Snail C----AATTCGCTGGAAATTCCCGGGAACAACTGCTGTACGCCCTTAAAAGTAGAAATATTATTACTTCAGAGGAAATTG
Tick CAGGTCTTCATCGGGAT----------GGCCATGGAG
Worm ------AAAATTTTTAAATATCACGTTTCATTTCATAAATTTGTACAATAAATGA----------
Xenopus AACAACTTTATTGGGAA----------CTCTAAAAGT
Zebrafish CACTCGTTCATCAGTAA----------CTCTCGGGCG
BatSynd AACACGTCCTCCTGA----------
BovineSynd AACGCCTTCATCAGTAC----------GGCCAAGTAC
ChimpSynd AACACGTCTCCTGAG----------G----------GTCCCCTGGC
CatSynd AACACGTCCCCCTTG----------G----------GTCCCCTGGC
ChickenSynd AACTCCTTCGTTGGAAC----------TGCCAAATTC
DogSynd AACGCCTTTACCAGCAC----------CGCCAAGTAC
GalagoSynd AACACGTCTCCTGAGGATCCCTGACCGGGCTGGGTGGGTGTCGCTGTCTCCAGCAACTGAATGGGAGAGGGGCCTGGGGT
HorseSynd AACACGTCCCCCGA-----CCTGAT-GTGCAGGG--GGTG----------GGAAGCGCATGGGGT--GGCCTTGAAAC
HumanSynd AACGCCTTCGTGAGCAC----------GGCTAAGTGC
MacqueSynd AACGCCTTTGTGAGCAC----------GGCTAAGTGC
MouseSynd AACGCCTTCGTCAGCAC----------TGCCAAATAT
OpossumSynd AATGCCTTCATCAGCAC----------TGCCAAGCAC
PigSynd AACGCTTTCATCAGCAC----------AGCCAAGTAC
PlatypusSynd AACGCCTTCATCAGCAC----------CGCCAAGCAC
RabbitSynd AACGCCTTCGTCAGCAC----------CGCCAAGTGC
RatSynd AACGCCTTCGTCAGCAC----------TGCCAAATAT
RhesusSynd AACGCCTTTGTGAGCAC----------GGCTAAGTGC
SalamandarSyndTCGCACCTCATGTAGGAAGACCATCAAAAGGTTGGAGTCTAAGCTTGCCTGGCAAA------CGGACAGACTCCAAGCAG
TurkeySynd AACTCCTTCGTTGGAAC----------TGCCAAATTC


Anemone TTGAAGCAGTAAAGCTAA-GTGGGACAAATATTGAGCTCCCCTAA----------
Crab CAGCTGTTAC----------TTGCCCTTCACCACTCCAAGCTGCTCACTGAAC----------TAATAGTGA-GGA
Dogfish Shark CAGCTGATCG----------ACGCCCTGCGGAGCCTGCATCTGCTGAGCGAAG----------CCGAGCTGCAGGC
Earthworm ----------
Gecko CAGTTGGTTC----------ATGCCCTGGAAACCTTTGAGCTGGTGCCGAGAG----------AGCAACTCCAGAA
Human AGGCCTTAAGATTCCA---GCTCATCACTAGAGGCAGCCCTCCATGGACCCATGAAAACTGAGATGAGGACCTTGGTACT
Lamprey ----------
Lancelet GTACTGATATAAGGATCTTACTGTCATTATGCTTTTACTTGAAGTTGGACATATCTCTGGGACTGGTGCGAGTTCCAGTC
Leech TTGTTGTCGACTCAGACA-AGTTCAGAGGATTGCCTGTTTTTCTGGCGAATTGTTTTGCTGGTTGTGCAAGGGAACAATT
Planarian ATGAT-ATTTTAAAACTA-TCTGAAAACATTTGTTT--CGATTTGTAG----------
Sharpshooter ----------
Snail CTACTTGTGTTGAATCAA-TGAAAAGACATGTGGAAAACAATCGGTGA----------
Tick CAGCTGTTGT----------ACGCGCTGGTCAAGCTCGGCATTATGACGTCGG----------AAGAAGTGATGAC
Worm ----------
Xenopus CAACTTCTGT----------ATGCACTACGCTCTTTGAAACTGCTACGGGAGG----------ACCAGATTCAGGA
Zebrafish CAGCTGCTCT----------CTGCCCTGCGCCACTGCAGGCTGCTGTCTCAAG----------GGGAGCTGAAGGA
BatSynd ----------
BovineSynd CAGCTCCTGT----------TCGCCCTCAAGGTGCTCAACATGATGCCAGAGG----------AGAAGCTGGCTGA
ChimpSynd CGGGCTGGGT----------CGGGTGTCGCTGTCTCCAGCAATGGGGTGGGGAG----------AGGGGTCTGGGGC
CatSynd CGGCATGGGT----------GGGGTGTCGCTGTTTCCAGCATCAGGGTGGGGAG----------AGGGACTGGGGGC
ChickenSynd CAGCTGCTCT----------TTGCTCTGAAGATCTTGAACATGGTGCCGGAGG----------AGAAGCTGGCCGA
DogSynd CAGCTGCTCT----------TCGCTCTCAAGGTGCTCAACATGATGCCTGAGG----------AGAAGCTGGCTGA
GalagoSynd TGGGCCGGGGCTGGGGCT-CCTCAGGGCTGTGTTACATCTGCCTTGCTGCCTGCCATTGCCAATCCTGGGAGTGGGGAGT
HorseSynd TTGCTTGGGGAGGAGA-------GGGGCTGCACTGCTCCAAGGGAGCTGGCTGGCTTCTGCCTCACCAGACCCTGCCAGC
HumanSynd CAGCTCCTCT----------TTGCCCTCAAGGTGCTCAACATGATGCCCGAGG----------AGAAGCTGGTTGA
MacqueSynd CAGCTCCTCT----------TTGCCCTCAAGGTGCTCAACATGATGCCTGAGG----------AGAAGCTGGCTGA
MouseSynd CAACTTCTAT----------TTGCCCTTAAGGTACTCAACATGATGCCCTCGG----------AAAAGCTGGCCGA
OpossumSynd CAGCTGCTCT----------TTGCCCTGAAGGTGCTGAACATGATGCCTGAGG----------ATAAGCTGGCTGA
PigSynd CAGCTCCTCT----------TTGCCCTCAAGGTGCTCAATATGATGCCGGAGG----------AAAAGCTGGCTGA
PlatypusSynd CAGCTGCTCT----------TCGCCCTGAAGATGCTCAACATGATGCCCGAGG----------AGAAGCTGGCCGA
RabbitSynd CAGCTCCTCT----------TCGCCCTCAAGGTGCTCAACATGATGCCGCAGG----------AGAAGCTGGCTGA
RatSynd CAGCTCCTAT----------TTGCCCTTAAGGTACTCAACATGATGCCCTCGG----------AAAAGCTGGCCGA
RhesusSynd CAGCTCCTCT----------TTGCCCTCAAGGTGCTCAACATGATGCCTGAGG----------AGAAGCTGGCTGA
SalamandarSyndTCCTTCTCCTA----------CCATGCCCCAAACATCTGAAAGGCAATCCCATTAGA----GGTTTGAAATGGCCCCACA
TurkeySynd CAGCTGCTCT----------TTGCTCTGAAGATCTTGAACATG----------


Anemone
Crab
Dogfish Shark
Earthworm
Gecko
Human
Lamprey
Lancelet
Leech
Planarian
Sharpshooter
Snail
Tick
Worm
Xenopus
Zebrafish
BatSynd
BovineSynd
ChimpSynd
CatSynd
ChickenSynd
DogSynd
GalagoSynd
HorseSynd
HumanSynd
MacqueSynd
MouseSynd
OpossumSynd
PigSynd
PlatypusSynd
RabbitSynd
RatSynd
RhesusSynd
SalamandarSynd
TurkeySynd

Supplementary Table A.1: Organisms containing putative Nudt16 orthologs are found across many phyla and classes. Nudt16 and Syndesmos homolog sequence information from the various databases. Accession numbers are provided where known. Hash marks (#) indicate a sequence from the JGI website. Astericks (*) indicate a sequence from the Salmander sequencing project. Ensembl sequences begin with ENS. The 57 putative Nudt16 orthologs are found in 5 phyla and among 17 classes as indicated. A sequence alignment of one representative organism from each class is shown in Figure 2.1 (page 32). All accession numbers used to obtain these data are provided.

Organism	NCBI # (nt)	Phylum	Class	Nudt16 or Syndesmos ?
Gastropod Snail (*Lottia gigantea*)	#	Mollusca	Gastropoda	Nudt16
American Pika (*Ochotona princeps*)	AAYZ01408159.1	Chordata	Actinoptergyii	Nudt16
Atlantic Salmon (*Salmo salar*)	EG762083.1	Chordata	Actinoptergyii	Nudt16
Bat (*Myotis lucifugus*)	AAPE01386445.1	Chordata	Mammalia	Nudt16
Bat (*Myotis lucifugus*)	AAPE01098161.1	Chordata	Mammalia	Syndesmos
Black-legged Tick (*Ixodes scapularis*)	EW827927.1	Arthropoda	Arachnida	Nudt16
Bovine (*Bos taurus*)	NM_001075560	Chordata	Mammalia	Nudt16
Bovine (*Bos taurus*)	NM_001034526	Chordata	Mammalia	Syndesmos
California Sea Hare (*Aplysia californica*)	AASC01167384	Mollusca	Gastropoda	Nudt16
Cat (*Felis catus*)	AANG01147609.1	Chordata	Mammalia	Syndesmos
Channel Catfish (*Ictalurus punctatus*)	CK419605.1	Chordata	Actinoptergyi	Nudt16
Chicken (*Gallus gallus*)	NM_204780	Chordata	Aves	Syndesmos
Chimpanzee (*Pan troglodytes*)	ENSPTRG00000015395	Chordata	Mammalia	Nudt16
Chimpanzee (*Pan troglodytes*)	XM_510787	Chordata	Mammalia	Syndesmos
Ciona savignyi	AACT01065371.1	Chordata	Ascidiacea	Nudt16
Common Carp (*Cyprinus carpio*)	CA968258.2	Chordata	Actinoptergyii	Nudt16
Coral (*Acropora millepora)*	DY583358	Cnidaria	Anthozoa	Nudt16
Cotton-melon aphid (*Aphis gossypii*)	DR907224.1	Arthropoda	Insecta	Nudt16
Crab-eating Macaque (*Macaca fascicularis*)	AB171049.1	Chordata	Mammalia	Nudt16
Crab-eating Macaque (*Macaca fascicularis*)	CJ440098.1	Chordata	Mammalia	Syndesmos
Dog (*Canis familiaris*)	XM_848264	Chordata	Mammalia	Nudt16

Dog (*Canis familiaris*)	XM_536987	Chordata	Mammalia	Syndesmos
Elephantfish (*Callorhinchus milii)*	AAVX01477340.1	Chordata	Chondricthyes	Nudt16
Fat Heat Minnow (*Pimephales promelas*)	DT148789	Chordata	Actinoptergyii	Nudt16
Florida Lancelet (*Branchiostoma floridae*)	#	Chordata	Leptocardii	Nudt16
Freshwater Planarian (*Schmidtea mediterranea*)	AAWT01049381.1	Platyhelminthes	Turbellaria	Nudt16
Gecko gecko	EB170884.1	Arthropoda	Insecta	Nudt16
Glassy-winged Sharpshooter (*Homalodisca coagulata*)	DN199400.1	Arthropoda	Insecta	Nudt16
Gray Mouse Lemur (*Microcebus murinus*)	ABDC01200929.1	Chordata	Mammalia	Nudt16
Green Anole (*Anolis carolinensis*)	AAWZ01026618.1	Chordata	Sauropsida	Nudt16
Green Crab (*Carcinus maenas*)	CX993972.1	Arthropoda	Malocostraca	Nudt16
Green Peach Aphid (*Myzus persicae*)	EC387416.1	Arthropoda	Insecta	Nudt16
Ground Squirrel (*Spermophilus tridecemlineatus*)	AAQQ01740884	Chordata	Mammalia	Nudt16
Horse (*Equus caballus*)	NW_001799668.1	Chordata	Mammalia	Nudt16
Horse (*Equus caballus*)	AAWR01029567.1	Chordata	Mammalia	Syndesmos
Human (*Homo sapiens*)	AK055827	Chordata	Mammalia	Nudt16
Human (*Homo sapiens*)	AK092642	Chordata	Mammalia	Syndesmos
Human Body Louse (*Pediculus humanus corporis*)	AAZO01000475.1	Arthropoda	Insecta	Nudt16
Humus Earthworm (*Lumbricus rubellus*)	CF798519.1	Annelida	Haplotaxida	Nudt16
Japanese medaka (*Oryzias latipes*)	ENSORLG00000001613	Chordata	Actinoptergyii	Nudt16
Leech (*Helobdella robusta*)	#	Annelida	Hirundi	Nudt16
Little skate (*Leucoraja erinacea*)	EE992989.1	Chordata	Chondricthyes	Nudt16
Mouse (*Mus musculus*)	BC064048	Chordata	Mammalia	Nudt16
Mouse (*Mus*	AK007897	Chordata	Mammalia	Syndesmos

musculus)				
Opossum (*Monodelphis domestica*)	ENSMODT00000001 4692	Chordata	Mammalia	Nudt16
Opossum (*Monodelphis domestica*)	ENSMODT00000008 035	Chordata	Mammalia	Syndesmos
Orangutang (*Pongo pygmaeus*)	CR791406	Chordata	Mammalia	Nudt16
Pea Aphid (*Acyrthosiphon pisum*)	CN758496.1	Chordata	Mammalia	Nudt16
Pig (*Sus scrofa*)	BX918030	Chordata	Mammalia	Nudt16
Pig (*Sus scrofa*)	DY418313	Chordata	Mammalia	Syndesmos
Platypus (*Ornithorhynchus anatinus*)	AC164651.3	Chordata	Mammalia	Nudt16
Platypus (*Ornithorhynchus anatinus*)	AAPN01225323	Chordata	Mammalia	Syndesmos
Rabbit (*Oryctolagus cuniculus*)	AAGW01335664.1	Chordata	Mammalia	Nudt16
Rabbit (*Oryctolagus cuniculus*)	EB379926	Chordata	Mammalia	Syndesmos
Rat (*Rattus norvegicus*)	XM_343464	Chordata	Mammalia	Nudt16
Rat (*Rattus norvegicus*)	BC105819	Chordata	Mammalia	Syndesmos
Rhesus Monkey (*Macaca mulatta*)	XM_001115614	Chordata	Mammalia	Nudt16
Rhesus Monkey (*Macaca mulatta*)	XM_001097827	Chordata	Mammalia	Syndesmos
Salamander (*Ambystoma mexicanum*)	*	Chordata	Sauropsida	Syndesmos
Sea Lamprey (*Petromyzon marinus*)	CO548416	Chordata	Cephalaspidomorphi	Nudt16
Sea Squirt (*Ciona intestinalis*)	ENSCING000000148 76	Chordata	Ascidiacea	Nudt16
Sheep (*Ovis aries*)	NM_001038013	Chordata	Mammalia	Nudt16
Shrew (*Tupaia belangeri*)	AAPY01639450.1	Chordata	Mammalia	Nudt16
Small Madascar Hedgehog (*Echinops telfairi*)	ENSETEG000000041 18	Chordata	Mammalia	Nudt16
Small-eared Galago (*Otolemur garnettii*)	AAQR01162051.1	Chordata	Mammalia	Nudt16
Small-eared Galago (*Otolemur garnettii*)	AAQR01647052	Chordata	Mammalia	Syndesmos

Spiny Dogfish Shark (*Squalus acanthias*)	EE721764.1	Chordata	Chondricthyes	Nudt16
Starlet Sea Anemone (*Nematostella vectensis*)	#	Cnidaria	Anthozoa	Nudt16
Takifugu rubripes	CA330048.1	Chordata	Actinoptergyii	Nudt16
Three Spined Stickleback (*Gasterosteus aculeatus*)	AANH01002833.1	Chordata	Chondricthyes	Nudt16
Turkey (*Meleagris gallopavo*)	EH283769.1	Chordata	Aves	Syndesmos
Two-spotted Cricket (*Gryllus bimaculatus*)	DC440739.1	Arthropoda	Insecta	Nudt16
Worm (*Capitella sp. I*)	#	Annelida	Polychaeta	Nudt16
Xenopus laevis	AY423379	Chordata	Amphibia	Nudt16
Xenopus tropicalis	CU075346.1	Chordata	Amphibia	Nudt16
Zebrafish (*Danio rerio*)	NM_001006008	Chordata	Actinoptergyii	Nudt16

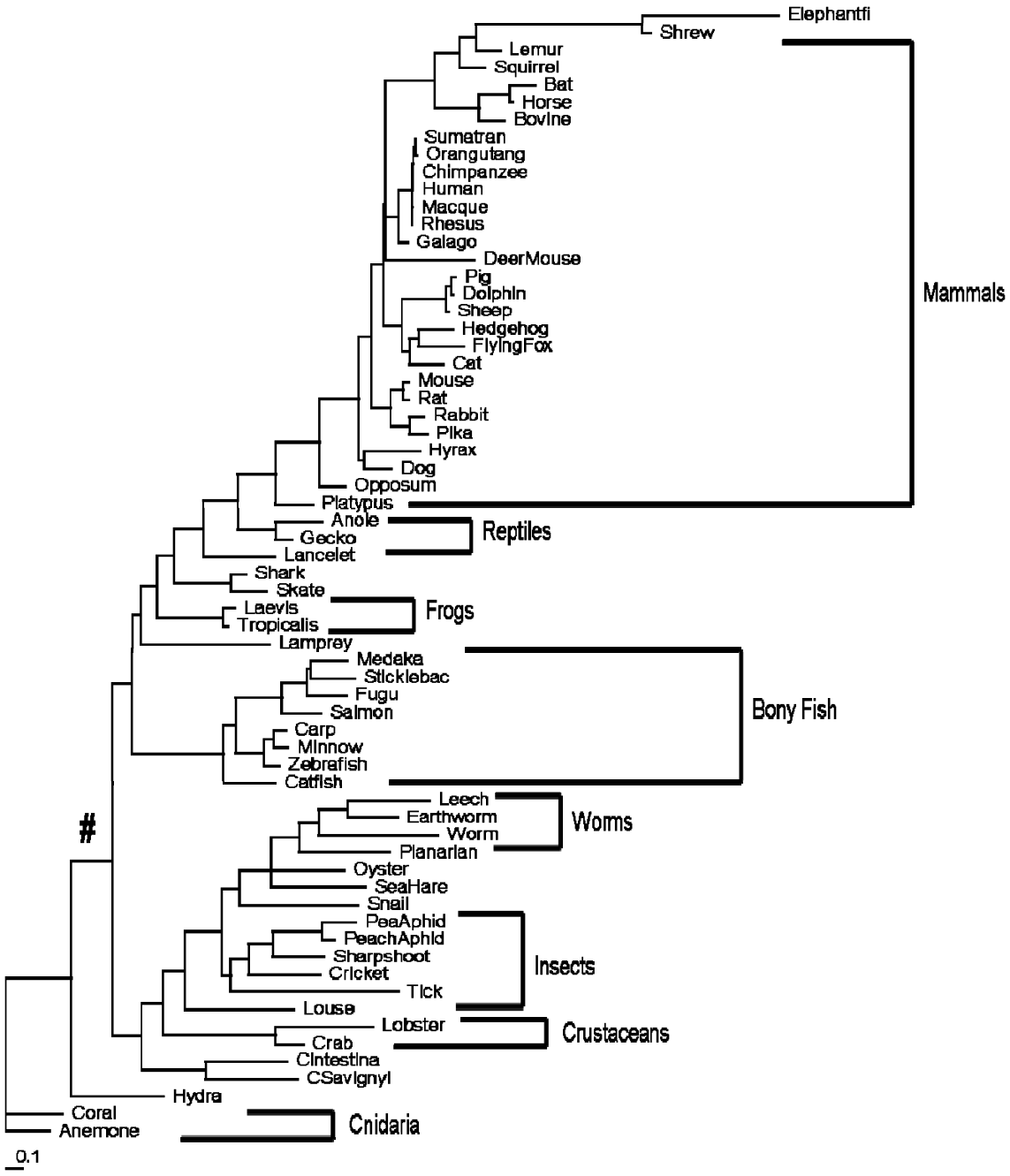


Figure A.4: A maximum likelihood tree of all 65 Nudt16 orthologs. Amino acid sequences were aligned using ClustalX and graphed using the PHYLIP maximum likelihood algorithm. TreeView was used to visualize the tree. More sequences made it harder for the program to determine the phylogenetic relationships. The hash mark (#) indicates the division between invertebrate and vertebrate sequences.

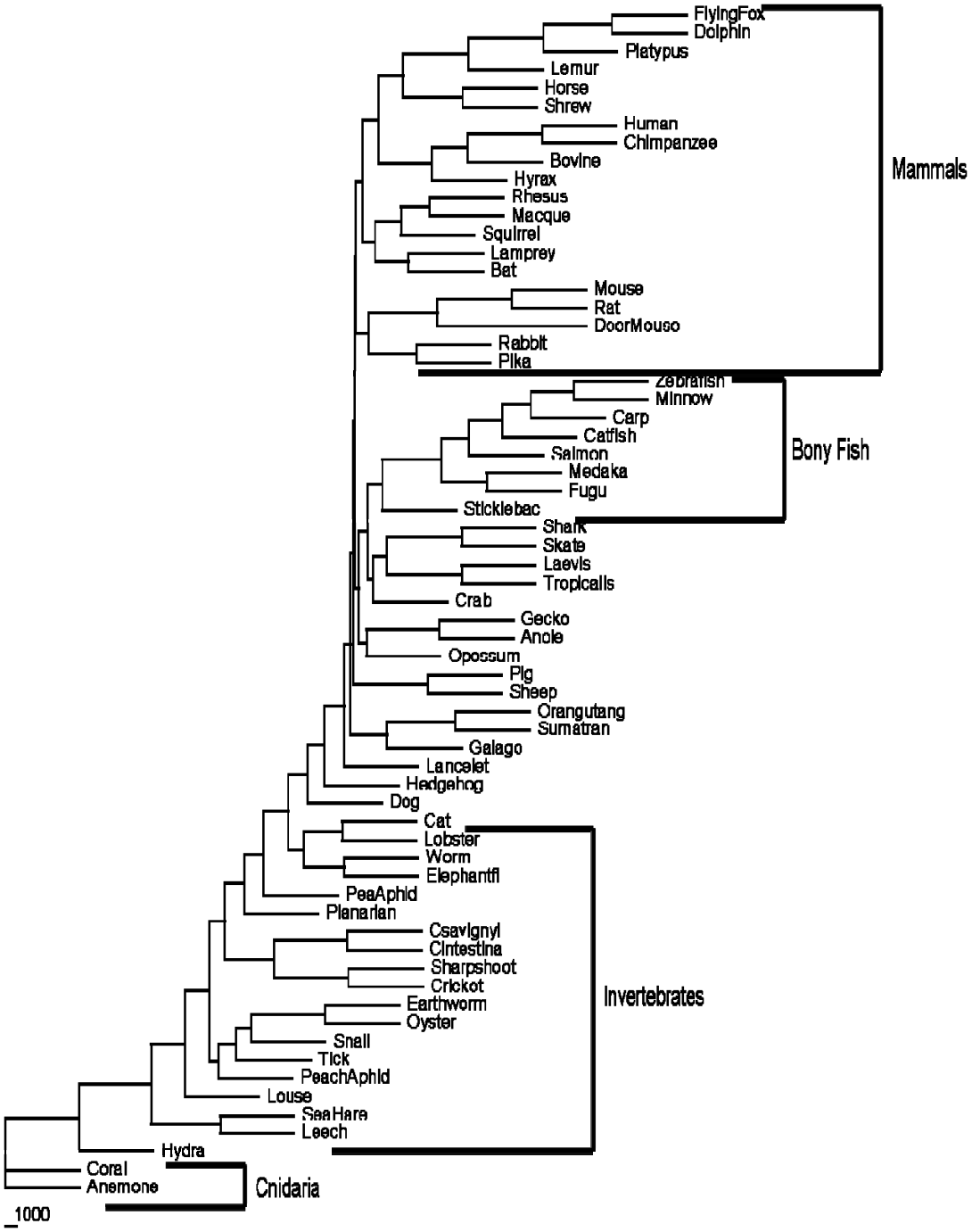


Figure A.5: A maximum parsimony tree of all 65 Nudt16 nucleic acid sequences. Nucleic acid sequences of all Nudt16 orthologs were aligned using Clustal X and graphed using the PHYLIP maximum parsimony algorithm. TreeView was used to draw the tree. The division between vertebrate and invertebrate sequences is unclear as many are scattered; this is not unexpected for such divergent sequences.

Table A.2: Sequence information for the up-dated Nudt16 orthologs. The updated list of Nudt16 orthologs includes mostly mammals, but also some more distant orthologs. All were included in the trees in Figures A.4 and A.5.

Organism	NCBI # (nt)	Phylum	Class	Nudt16 or Syndesmos?
Bottlenosed Dolphin (*Tursiops truncatus*)	ABRN01220060.1	Chordata	Mammalia	Nudt16
Cape Rock Hyrax (*Procavia capensis*)	ABRQ01363315.1	Chordata	Mammalia	Nudt16
Cat (*Felis catus*)	AANG01274351.1	Chordata	Mammalia	Nudt16
Hydra magnipapillata	ABRM01001513	Cnidaria	Hydrozoa	Nudt16
Large Flying Fox (*Pteropus vampyrus*)	ABRP01019892.1	Chordata	Mammalia	Nudt16
Lobster (*Homarus americanus*)	EX487804.1	Arthropoda	Malocostraca	Nudt16
Pacific Oyster (*Crassostrea gigas*)	AM859738.1	Mollusca	Bivalvia	Nudt16
Prairie Deer Mouse (*Peromyscus maniculatus bairdii*)	EV469951.1	Chordata	Mammalia	Nudt16
Sumatran Orangutan (*Pongo abelii*)	ABGA01292878.1	Chordata	Mammalia	Nudt16

Appendix B: Cloning Strategy for Nudt16 Orthologs

The cloning of each Nudt16 ortholog was described in Chapter 4 (pages 105-106) of this dissertation. The oligo descriptions were not included as a way to condense the text. For completeness, a description of each oligo is included below.

The first oligos ordered were for cloning the amino-terminal extended rat ortholog. Figure B.1 shows the sequence of the oligos and their positions on the sequence of the full-length cDNA sequence as well as the ATCC clone sequence. The figure was made using BioEdit (Hall, 1999).

Once the amino-terminus was added to the rat cDNA clone, the full-length cDNA was used as a template for a PCR reaction with two organism-specific oligos. All organisms had oliogs that were specific for that organism which contained a CGCG clamp and a restriction site. Oligos for the 5' end included the initator codon (ATG); oligos for the 3' end included all three stop codons. Table B.1 includes the sequence of each oligo as well as the restriction site used to clone the PCR products as described in Chapter 4 (pages 105-106).

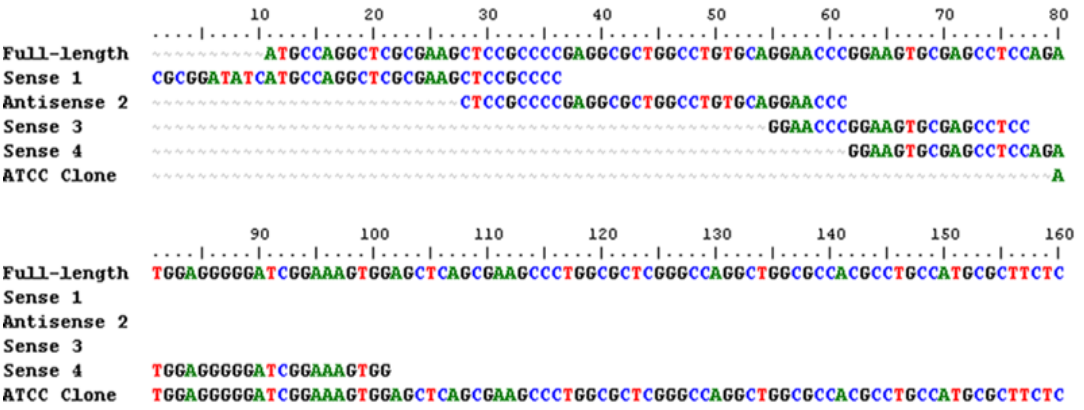


Figure B.1: Alignment of the oligos used to add the 5' – amino terminus to the ATCC clone. Oligos 1-4 were added to a PCR reaction containing the ATCC clone template DNA. The final PCR product was then analyzed based on the increased size to see whether the PCR had added the 5' amino terminus. The figure was created using BioEdit software.

Table B.1: Nudt16 orthologs were cloned using these oligos for PCR. Each oligo is designed to contain a particular restriction site (colored in **bold red**) as well as a CG clamp (colored in blue). Each 5' amino terminal end oligo had an initatiator methionine (underlined); each 3' carboxy terminus had all three stop codons (colored in *purple italics*).

Organism	5' End Oligo	3' End Oligo	Restriction Sites Used
Rat	5' CGCG**CATATG**ATGCCA GGCTCGCGAAGCTC 3'	5' CGCG**CTCGAG***TCACTATTAC* TTAGAATCTGGGATCTTGAG 3'	Nde & Xho1
Danio rerio	5' CGCG**CTCGAG**ATGGCGGA AGAGAAG 3'	5' CGCG**GGATCC***TCATTACTA*CT AGCAGCTCTGG 3'	Xho1 & BamH1
Little Skate	5' GGCGGACGCCG**CATAT G**GCGGAGCCC 3'	5' CGCG**CTCGAG***TCATTACTA*G CCCACCCGACTG 3'	Xho1
Crab	5' CGCG**CTCGAG**ATGTCT GTGGAATCAAC 3'	5' CGCG**CTCGAG***TCACTATTA* GTTCAGTGAGCAGCTTG 3'	Xho1
Sharp-shooter	5' CGCG**CATATG**TCCTCTG ATACAGG 3'	5' CGCG**CTCGAG***CTATTATCA* GGACTTGAGTGGAAG 3'	Nde & Xho1

Appendix C: Nudt16 Activity in the Presence of Metals, RNAs, and with Pre-incubation

C.1: Activity of Nudt16 in various metals and using different substrate RNAs

Nudt16 orthologs decapped multiple RNAs in the presence of three metals but not the paralog syndesmos. The text in Chapter 4 (pages 125-126) briefly describes each the over-all results of this analysis and includes the graph for the crab Nudt16 ortholog. The graphs for all other Nudt16 homologs are included in this appendix for completeness. The legends describe the results which are summarized in Chapter 4 (pages 125-126).

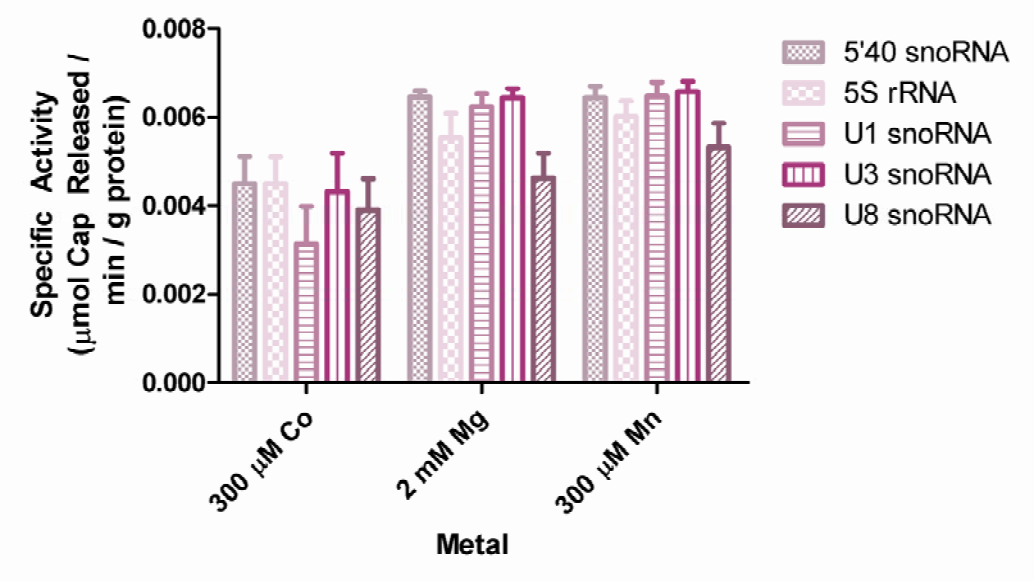


Figure C.1: The human Nudt16 protein decaps multiple RNAs in all metals. The human Nudt16 ortholog was tested as described in Chapter 4 (page 111) for its decapping activity using a variety of substrate RNAs and metal co factors. The concentration of protein in this experiment was 4.5 µM; the RNA concentrations were 5 nM. All RNAs were decapped in all metals; decapping was slightly lower in $CoCl_2$ than the other metals.

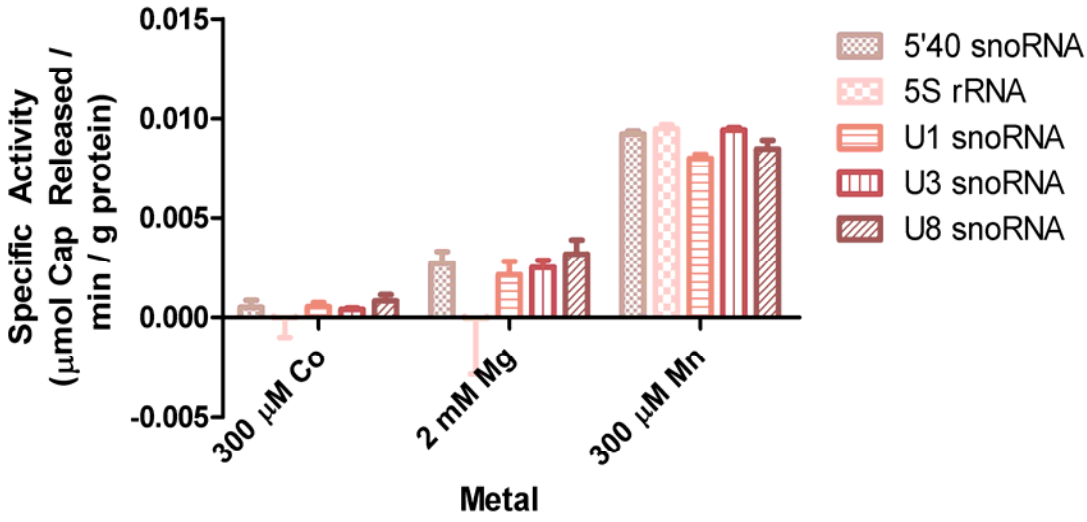


Figure C.2: The sharpshooter ortholog decaps multiple RNAs in many metals. The sharpshooter Nudt16 ortholog was used to test for its decapping ability in various metal and RNA conditions. At concentrations of protein at 2.5 µM, decapping activity was the highest in Mn, slightly lower in Mg, and virtually non-existent in Co. The only RNA that had significantly different decapping activity was 5S rRNA.

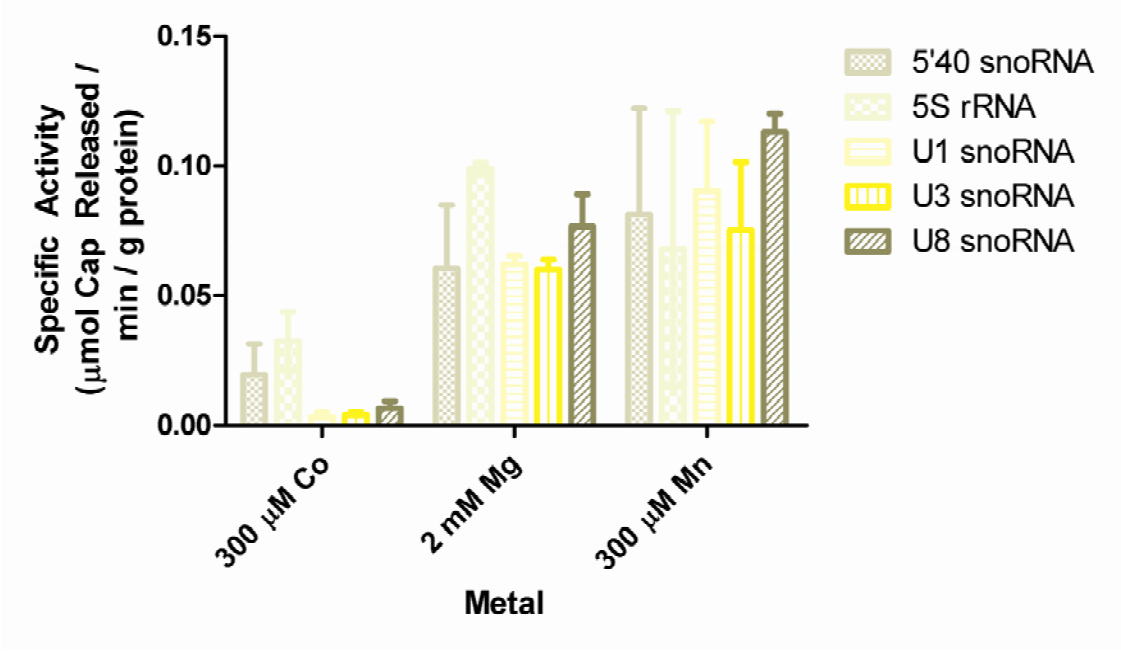


Figure C.3: The rat Nudt16 ortholog decaps all RNAs with equal efficiencies in Mn but shows preference in Mg and Co. The rat Nudt16 ortholog (238 nM) decapped all RNAs with equal efficiency when Mn was used as the metal. In the presence of Mg, 5S rRNA was decapped with higher efficiency than the other RNAs though the difference between it and the 5'40 R1 and U8 snoRNAs is likely not significant. In the presence of Co, the truncated form of U8 and 5S rRNA yielded higher decapping activities.

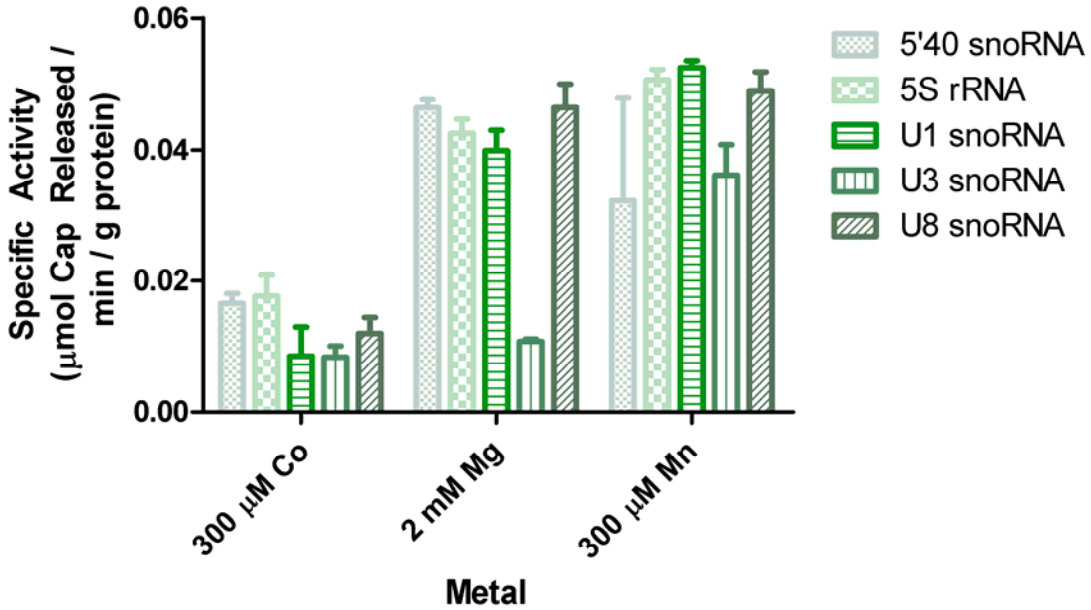


Figure C.4: The *Xenopus* Nudt16 ortholog shows some substrate specificity in Mg but not in Co or Mn. Decapping of the *Xenopus* ortholog at concentrations of 493 nM demonstrated lower decapping of U3 snoRNA whether Mn or Mg was the metal used in the reaction. The same effect is not observed for reactions containing Co, but decapping activity in Co was lower for all RNAs.

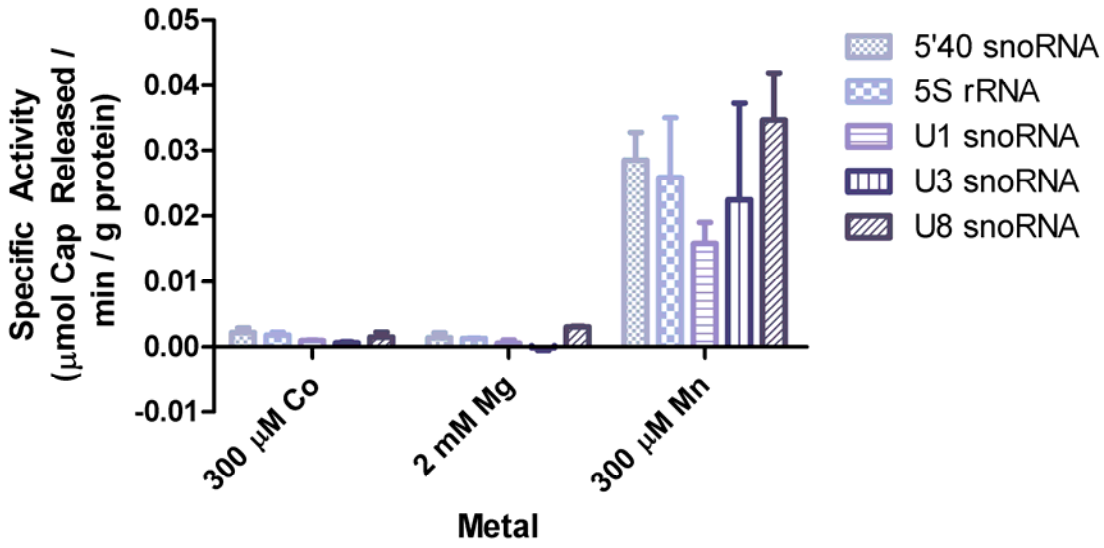


Figure C.5: The zebra fish Nudt16 ortholog did not function as well in Mg but could still decap most RNAs. The zebra fish Nudt16 ortholog (519 nM) was able to decap all RNAs in the presence of Mn but did not decap RNAs in the presence of Mg or Co. Similarly to the *Xenopus* ortholog, use of U3 snoRNA yielded lower levels of decapping.

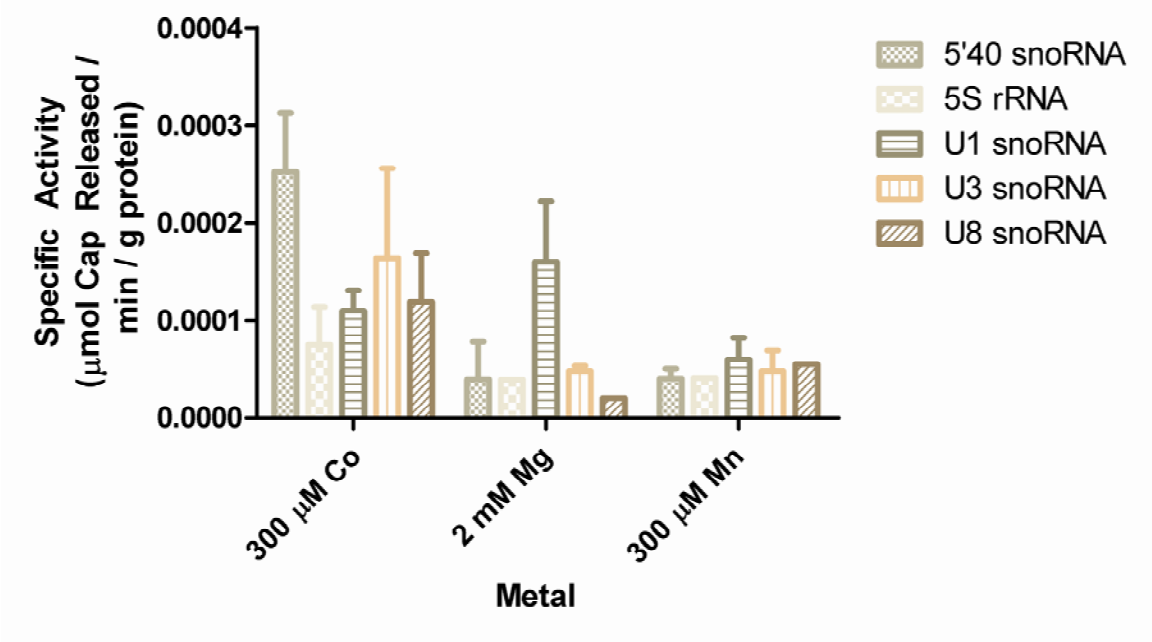
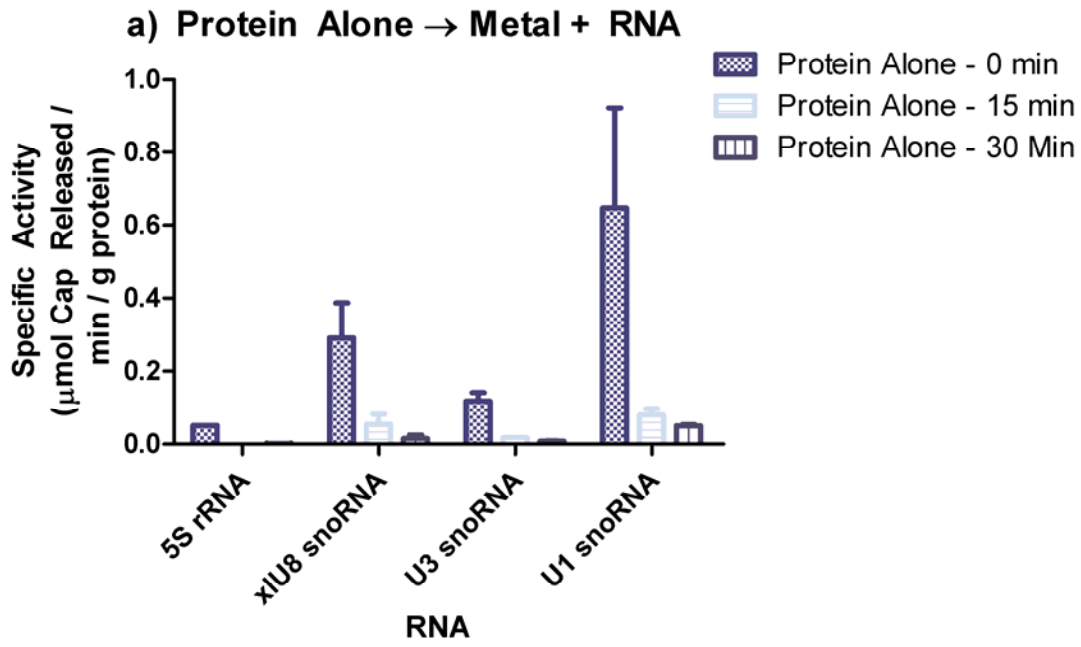


Figure C.6: The paralog syndesmos did not decap any RNAs in any metals. The human syndesmos protein was used at a final concentration of 3.1 µM protein to test whether decapping would occur when other RNAs and metals are used. Although the figure and graph show variations in decapping activity, the units of the y-axis are far below the units used in any of the above analyses. The observed variations are well within pipetting error.

C.2: Nudt16 instability at 37°C contributes to lack of turn over

Nudt16 ortholog stability at 37°C was studied by pre-incubating the protein in the absence of both RNA and metal or in the presence of one but not the other reactant (see Chapter 4 pages 112 and 127-129). The results for the *Xenopus* ortholog were discussed in Chapter 4 (pages 127-129). The data for the other organisms is included below. In all cases, Figure a) is the Nudt16 protein alone, Figure b) is protein plus metal, and Figure c) is protein plus RNA. The sharpshooter was not tested for instability at 37°C because of its low specific activity.



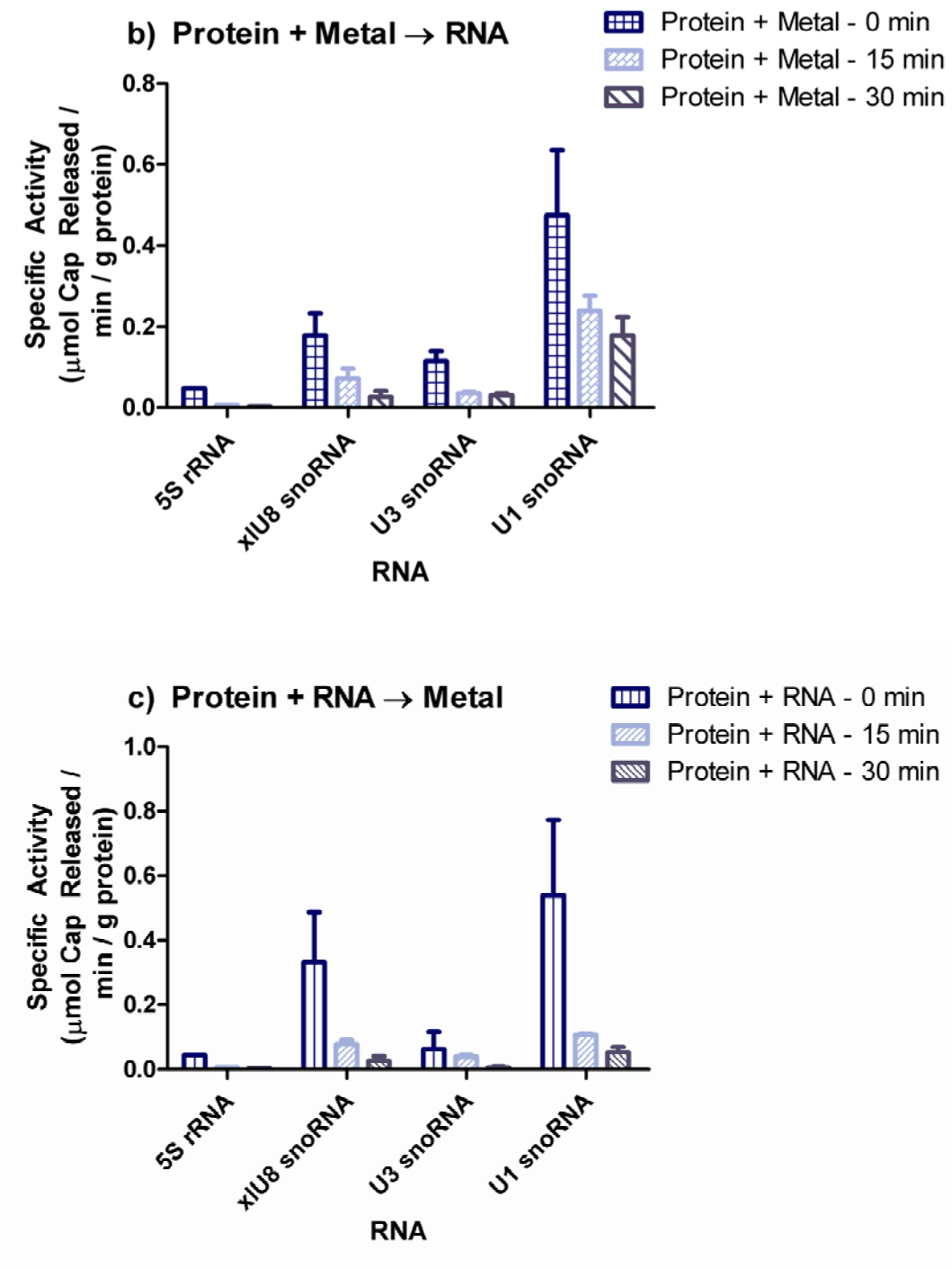
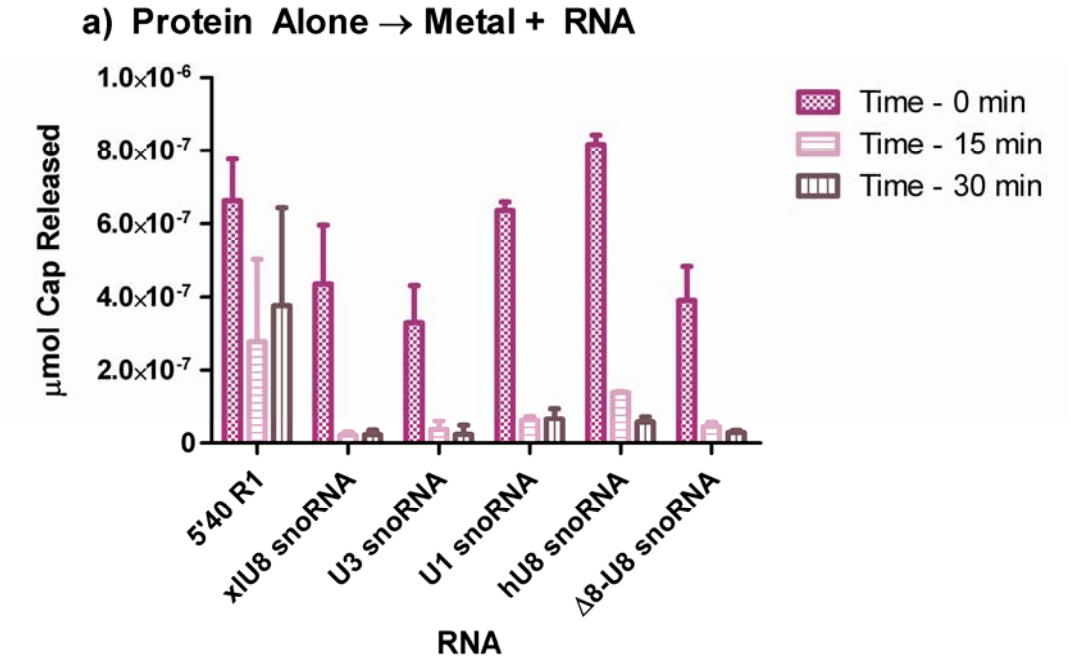
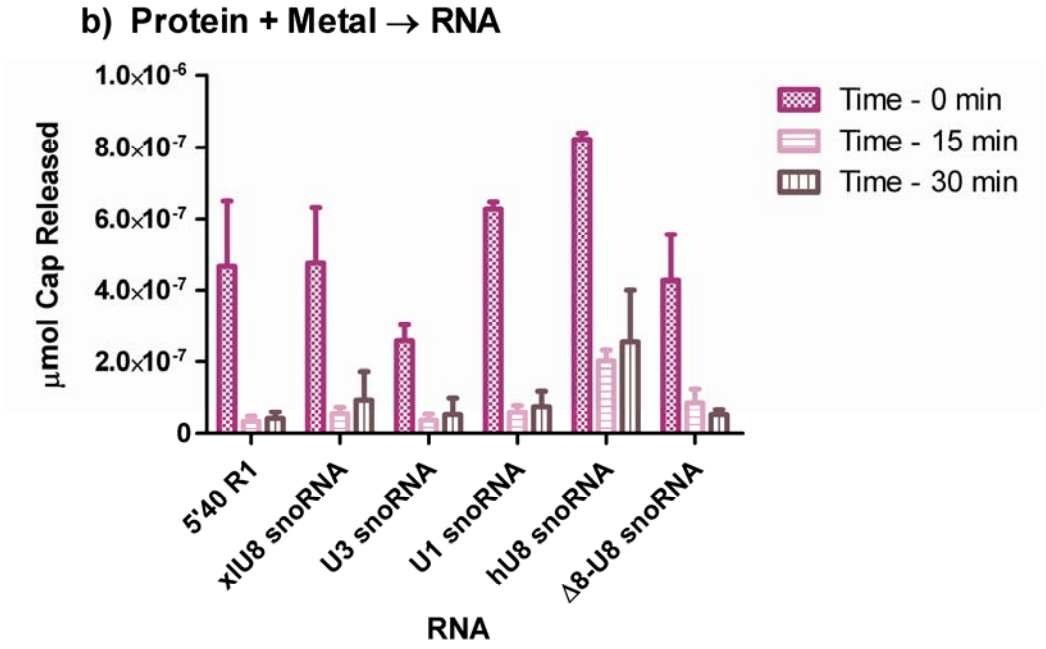


Figure C.7: The zebra fish ortholog is not stabilized by pre-incubation with RNA or metal. The zebra fish ortholog was pre-incubated alone (a), in the presence of metal (b), or in the presence of RNA (c). Unlike the *Xenopus* ortholog, the zebra fish ortholog is not stabilized by the presence of RNA.


a) Protein Alone → Metal + RNA
µmol Cap Released
1.0×10-6
8.0×10-7
6.0×10-7
4.0×10-7
2.0×10-7
0
Time - 0 min
Time - 15 min
Time - 30 min
5'40 R1
xlU8 snoRNA
U3 snoRNA
U1 snoRNA
hU8 snoRNA
Δ8-U8 snoRNA
RNA


b) Protein + Metal → RNA
µmol Cap Released
1.0×10-6
8.0×10-7
6.0×10-7
4.0×10-7
2.0×10-7
0
Time - 0 min
Time - 15 min
Time - 30 min
5'40 R1
xlU8 snoRNA
U3 snoRNA
U1 snoRNA
hU8 snoRNA
Δ8-U8 snoRNA
RNA

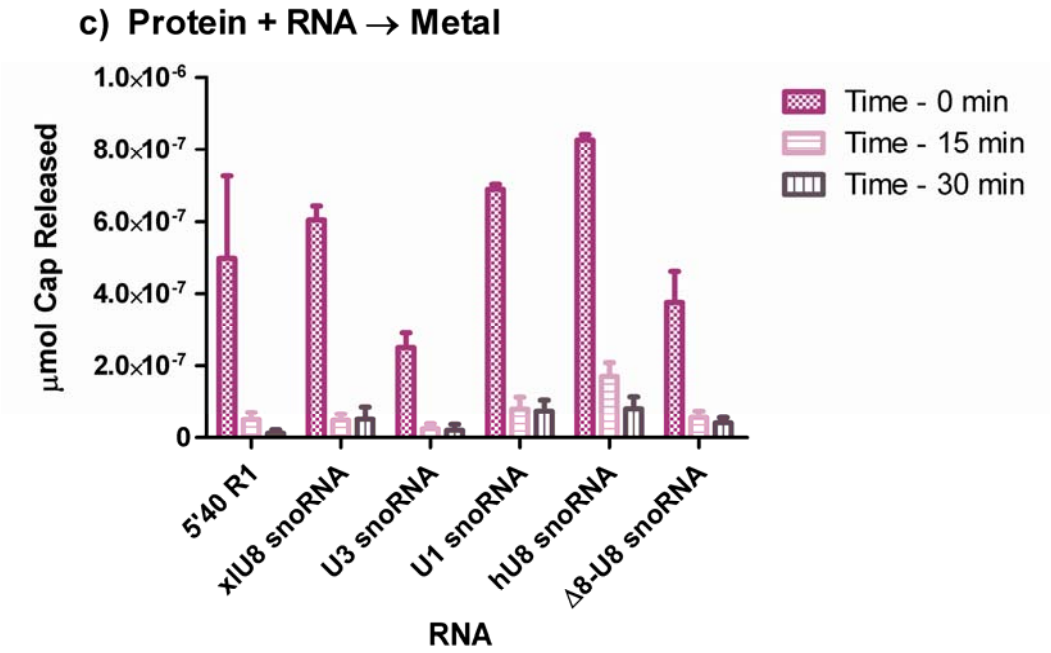
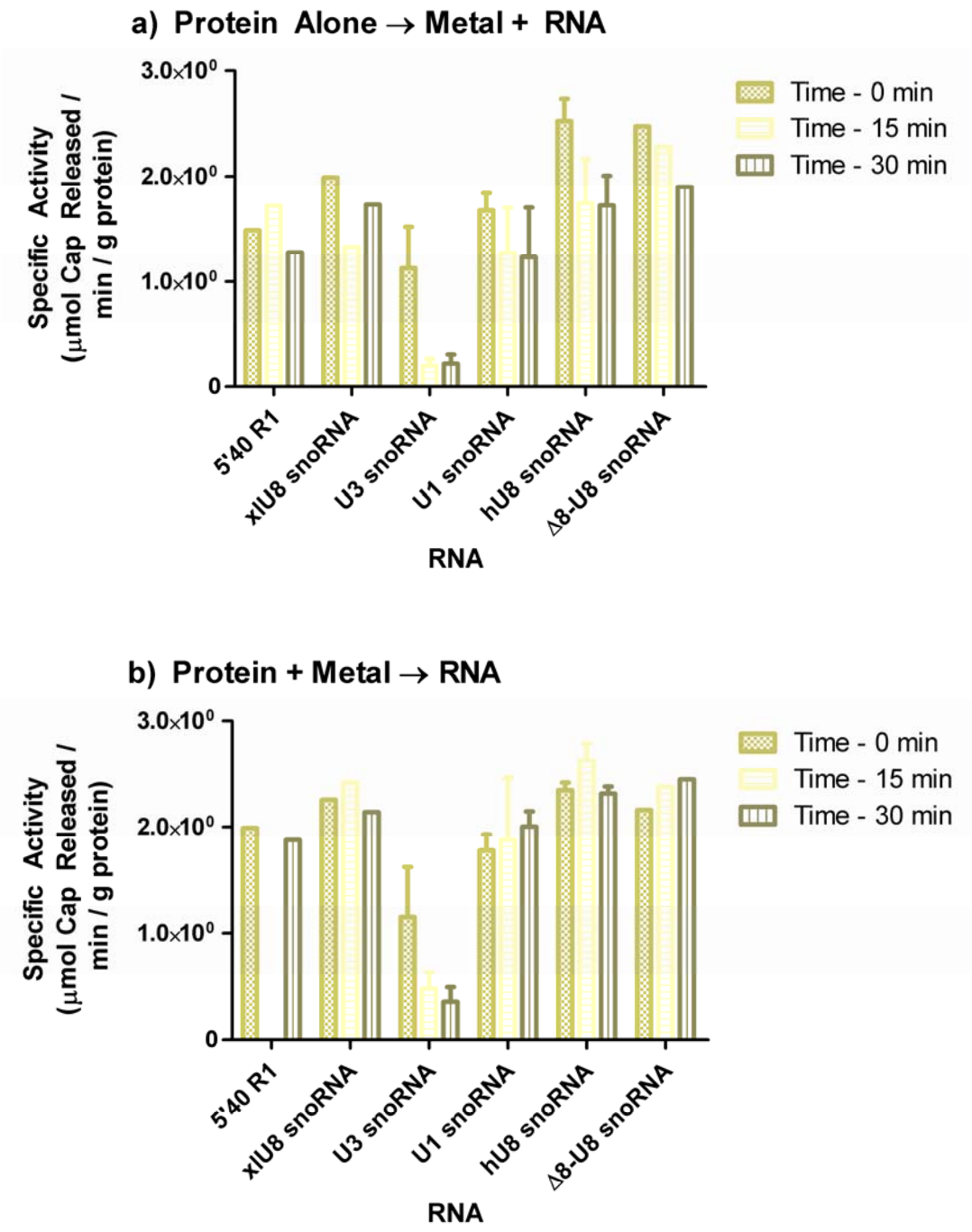


Figure C.8: The human Nudt16 ortholog is more like the zebra fish Nudt16 protein than the *Xenopus* Nudt16 ortholog. The human Nudt16 protein was incubated alone (a), in the presence of metal (b), or in the presence of RNA (c). Like the zebra fish, the human protein was not stabilized by the addition of RNA; it was destabilized in all conditions.


a) Protein Alone → Metal + RNA
Specific Activity (μmol Cap Released / min / g protein)
3.0×10⁰
2.0×10⁰
1.0×10⁰
0
Time - 0 min
Time - 15 min
Time - 30 min
5'40 R1
xlU8 snoRNA
U3 snoRNA
U1 snoRNA
hU8 snoRNA
Δ8-U8 snoRNA
RNA
b) Protein + Metal → RNA
Specific Activity (μmol Cap Released / min / g protein)
3.0×10⁰
2.0×10⁰
1.0×10⁰
0
Time - 0 min
Time - 15 min
Time - 30 min
5'40 R1
xlU8 snoRNA
U3 snoRNA
U1 snoRNA
hU8 snoRNA
Δ8-U8 snoRNA
RNA

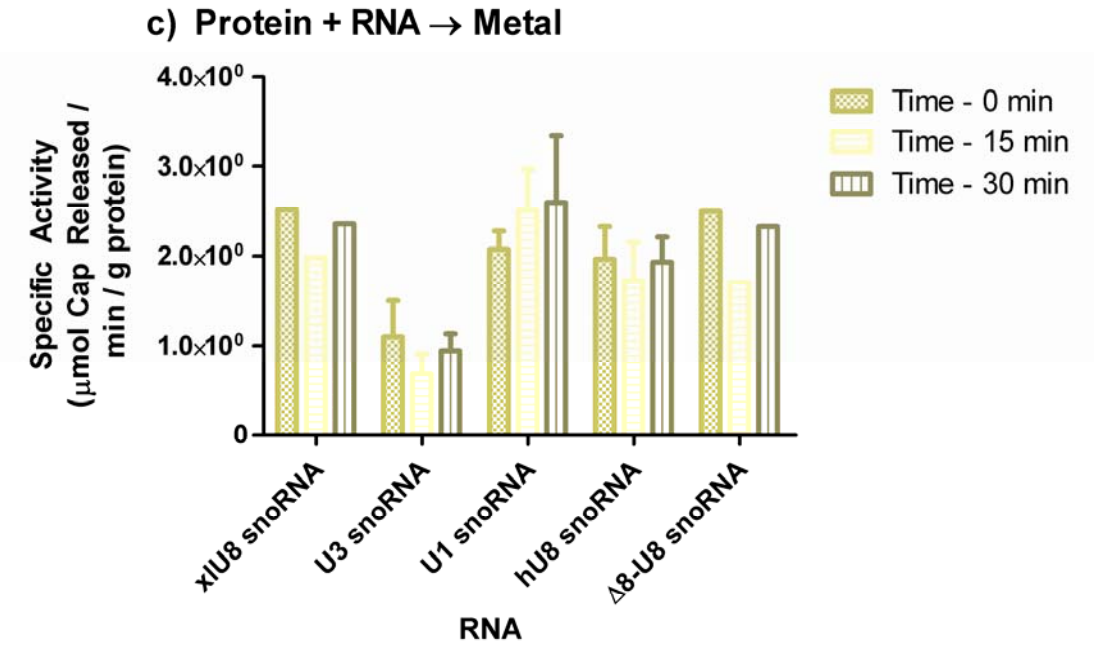


Figure C.9: The rat Nudt16 ortholog is not destabilized at 37°C. Unlike all other Nudt16 orthologs, the rat Nudt16 ortholog is not greatly affected by destabilization of the protein. In cases where a decrease in activity exists (for instance the U3 snoRNA), RNA stabilizes the protein.

Bibliography

Abdelghany, H.M., Bailey, S., Blackburn, G.M., Rafferty, J.B., and McLennan, A.G. (2003). Analysis of the catalytic and binding residues of the diadenosine tetraphosphate pyrophosphohydrolase from Caenorhabditis elegans by site-directed mutagenesis. J Biol Chem *278*, 4435-4439.

Abouheif, E., Zardoya, R., and Meyer, A. (1998). Limitations of metazoan 18S rRNA sequence data: implications for reconstructing a phylogeny of the animal kingdom and inferring the reality of the Cambrian explosion. J Mol Evol *47*, 394-405.

Acconcia, F., Barnes, C.J., and Kumar, R. (2006). Estrogen and tamoxifen induce cytoskeletal remodeling and migration in endometrial cancer cells. Endocrinology *147*, 1203-1212.

Aguilera, A. (2005). Cotranscriptional mRNP assembly: from the DNA to the nuclear pore. Curr Opin Cell Biol *17*, 242-250.

Aittaleb, M., Rashid, R., Chen, Q., Palmer, J.R., Daniels, C.J., and Li, H. (2003). Structure and function of archaeal box C/D sRNP core proteins. Nat Struct Biol *10*, 256-263.

Akao, Y., Nakagawa, Y., and Naoe, T. (2006). MicroRNAs 143 and 145 are possible common onco-microRNAs in human cancers. Oncol Rep *16*, 845-850.

Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell *121*, 207-221.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res *25*, 3389-3402.

An, D.J., Roh, I.S., Song, D.S., Park, C.K., and Park, B.K. (2007). Phylogenetic characterization of porcine circovirus type 2 in PMWS and PDNS Korean pigs between 1999 and 2006. Virus Res.

Atwater, J.A., Wisdom, R., and Verma, I.M. (1990). Regulated mRNA stability. Annu Rev Genet *24*, 519-541.

Aukrust, I., Hollas, H., Strand, E., Evensen, L., Trave, G., Flatmark, T., and Vedeler, A. (2007). The mRNA-binding site of annexin A2 resides in helices C-D of its domain IV. J Mol Biol *368*, 1367-1378.

Bachellerie, J.P., Cavaille, J., and Huttenhofer, A. (2002). The expanding snoRNA world. Biochimie *84*, 775-790.

Baciu, P.C., Saoncella, S., Lee, S.H., Denhez, F., Leuthardt, D., and Goetinck, P.F. (2000). Syndesmos, a protein that interacts with the cytoplasmic domain of syndecan-4, mediates cell spreading and actin cytoskeletal organization. J Cell Sci *113 Pt 2*, 315-324.

Bail, S., and Kiledjian, M. (2006). More than 1 + 2 in mRNA decapping. Nat Struct Mol Biol *13*, 7-9.

Balch, W.E., Magrum, L.J., Fox, G.E., Wolfe, R.S., and Woese, C.R. (1977). An ancient divergence among the bacteria. J Mol Evol *9*, 305-311.

Baldauf, S.L. (2003). Phylogeny for the faint of heart: a tutorial. Trends Genet *19*, 345-351.

Bayaa, M., Booth, R.A., Sheng, Y., and Liu, X.J. (2000). The classical progesterone receptor mediates Xenopus oocyte maturation through a nongenomic mechanism. Proc Natl Acad Sci U S A *97*, 12607-12612.

Behm-Ansmant, I., Rehwinkel, J., and Izaurralde, E. (2006). MicroRNAs silence gene expression by repressing protein expression and/or by promoting mRNA decay. Cold Spring Harb Symp Quant Biol *71*, 523-530.

Bessman, M.J., Frick, D.N., and O'Handley, S.F. (1996). The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. J Biol Chem *271*, 25059-25062.

Bhupathiraju, V.K., Oren, A., Sharma, P.K., Tanner, R.S., Woese, C.R., and McInerney, M.J. (1994). Haloanaerobium salsugo sp. nov., a moderately halophilic, anaerobic bacterium from a subterranean brine. Int J Syst Bacteriol *44*, 565-572.

Bi, J.L., Castle, S.J., and Toscano, N.C. (2007). Amino acid fluctuations in young and old orange trees and their influence on glassy-winged sharpshooter (Homalodisca vitripennis) population densities. J Chem Ecol *33*, 1692-1706.

Blikstad, C., Shokeer, A., Kurtovic, S., and Mannervik, B. (2008). Emergence of a novel highly specific and catalytically efficient enzyme from a naturally promiscuous glutathione transferase. Biochim Biophys Acta *1780*, 1458-1463.

Boisvert, F.M., Cote, J., Boulanger, M.C., and Richard, S. (2003). A proteomic analysis of arginine-methylated protein complexes. Mol Cell Proteomics *2*, 1319-1330.

Branscombe, T.L., Frankel, A., Lee, J.H., Cook, J.R., Yang, Z., Pestka, S., and Clarke, S. (2001). PRMT5 (Janus kinase-binding protein 1) catalyzes the formation of symmetric dimethylarginine residues in proteins. J Biol Chem *276*, 32971-32976.

Brown, N.P., Leroy, C., and Sander, C. (1998a). MView: a web-compatible database search or multiple alignment viewer. Bioinformatics *14*, 380-381.

Brown, N.P., Leroy, C., and Sander, C. (1998b). MView: a web-compatible database search or multiple alignment viewer. Bioinformatics *14*, 380-381.

Burggraf, S., Mayer, T., Amann, R., Schadhauser, S., Woese, C.R., and Stetter, K.O. (1994). Identifying members of the domain Archaea with rRNA-targeted oligonucleotide probes. Appl Environ Microbiol *60*, 3112-3119.

Cai, X., Hagedorn, C.H., and Cullen, B.R. (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. RNA *10*, 1957-1966.

Callard, G.V., and Mak, P. (1985). Exclusive nuclear location of estrogen receptors in Squalus testis. Proc Natl Acad Sci U S A *82*, 1336-1340.

Callard, G.V., Pudney, J.A., Mak, P., and Canick, J.A. (1985). Stage-dependent changes in steroidogenic enzymes and estrogen receptors during spermatogenesis in the testis of the dogfish, Squalus acanthias. Endocrinology *117*, 1328-1335.

Campbell, A.M., and Heyer, L.J. (2003). Discovering genomis, proteomics, and bioinformatics (Pearson Education, Inc.).

Campbell, F.E., Jr., Cassano, A.G., Anderson, V.E., and Harris, M.E. (2002). Pre-steady-state and stopped-flow fluorescence analysis of Escherichia coli ribonuclease III: insights into mechanism and conformational changes associated with binding and catalysis. J Mol Biol *317*, 21-40.

Cannarozzi, G., Schneider, A., and Gonnet, G. (2007). A phylogenomic study of human, dog, and mouse. PLoS Comput Biol *3*, e2.

Chan, S.P., and Slack, F.J. (2006). microRNA-mediated silencing inside P-bodies. RNA Biol *3*, 97-100.

Chen, H., Wei, P., Huang, C., Tan, L., Liu, Y., and Lai, L. (2006). Only one protomer is active in the dimer of SARS 3C-like proteinase. J Biol Chem *281*, 13894-13898.

Chen, N., Walsh, M.A., Liu, Y., Parker, R., and Song, H. (2005). Crystal structures of human DcpS in ligand-free and m7GDP-bound forms suggest a dynamic mechanism for scavenger mRNA decapping. J Mol Biol *347*, 707-718.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., and Thompson, J.D. (2003). Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res *31*, 3497-3500.

Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., Kaufman, T.C., Kellis, M., Gelbart, W., Iyer, V.N.*, et al.* (2007). Evolution of genes and genomes on the Drosophila phylogeny. Nature *450*, 203-218.

Coller, J., and Parker, R. (2004). Eukaryotic mRNA decapping. Annu Rev Biochem *73*, 861-890.

Coseno, M., Martin, G., Berger, C., Gilmartin, G., Keller, W., and Doublie, S. (2008). Crystal structure of the 25 kDa subunit of human cleavage factor Im. Nucleic Acids Res *36*, 3474-3483.

Cougot, N., Babajko, S., and Seraphin, B. (2004). Cytoplasmic foci are sites of mRNA decay in human cells. J Cell Biol *165*, 31-40.

Cowley, S.M., Hoare, S., Mosselman, S., and Parker, M.G. (1997). Estrogen receptors alpha and beta form heterodimers on DNA. J Biol Chem *272*, 19858-19862.

Cuevas, M.E., and Callard, G. (1992). Androgen and progesterone receptors in shark (Squalus) testis: characteristics and stage-related distribution. Endocrinology *130*, 2173-2182.

Custodia-Lora, N., Novillo, A., and Callard, I.P. (2004). Effect of gonadal steroids on progesterone receptor, estrogen receptor, and vitellogenin expression in male turtles (Chrysemys picta). J Exp Zoolog A Comp Exp Biol *301*, 15-25.

Darzacq, X., Jady, B.E., Verheggen, C., Kiss, A.M., Bertrand, E., and Kiss, T. (2002). Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. EMBO J *21*, 2746-2756.

Das, B., Butler, J.S., and Sherman, F. (2003). Degradation of normal mRNA in the nucleus of Saccharomyces cerevisiae. Mol Cell Biol *23*, 5502-5515.

Das, B., Das, S., and Sherman, F. (2006). Mutant LYS2 mRNAs retained and degraded in the nucleus of Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *103*, 10871-10876.

Decker, C.J., and Parker, R. (1994). Mechanisms of mRNA degradation in eukaryotes. Trends Biochem Sci *19*, 336-340.

Denhez, F., Wilcox-Adelman, S.A., Baciu, P.C., Saoncella, S., Lee, S., French, B., Neveu, W., and Goetinck, P.F. (2002). Syndesmos, a syndecan-4 cytoplasmic domain interactor,

binds to the focal adhesion adaptor proteins paxillin and Hic-5. J Biol Chem *277*, 12270-12274.

Dessimoz, C., Boeckmann, B., Roth, A.C., and Gonnet, G.H. (2006a). Detecting non-orthology in the COGs database and other approaches grouping orthologs using genome-specific best hits. Nucleic Acids Res *34*, 3309-3316.

Dessimoz, C., Gil, M., Schneider, A., and Gonnet, G.H. (2006b). Fast estimation of the difference between two PAM/JTT evolutionary distances in triplets of homologous sequences. BMC Bioinformatics *7*, 529.

Diaz-Lazcoz, Y., Henaut, A., Vigier, P., and Risler, J.L. (1995). Differential codon usage for conserved amino acids: evidence that the serine codons TCN were primordial. J Mol Biol *250*, 123-127.

Dobrzanska, M., Szurmak, B., Wyslouch-Cieszynska, A., and Kraszewska, E. (2002). Cloning and characterization of the first member of the Nudix family from Arabidopsis thaliana. J Biol Chem *277*, 50482-50486.

Doerks, T., Copley, R.R., Schultz, J., Ponting, C.P., and Bork, P. (2002). Systematic identification of novel protein domain families associated with nuclear functions. Genome Res *12*, 47-56.

Drummond, A.J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol *7*, 214.

Dunckley, T., and Parker, R. (1999). The DCP2 protein is required for mRNA decapping in Saccharomyces cerevisiae and contains a functional MutT motif. EMBO J *18*, 5411-5422.

Dunckley, T., Tucker, M., and Parker, R. (2001). Two related proteins, Edc1p and Edc2p, stimulate mRNA decapping in Saccharomyces cerevisiae. Genetics *157*, 27-37.

Duret, L., and Mouchiroud, D. (2000). Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. Mol Biol Evol *17*, 68-74.

Eddy, S.R. (1999). Noncoding RNA genes. Curr Opin Genet Dev *9*, 695-699.

Eddy, S.R. (2001). Non-coding RNA genes and the modern RNA world. Nat Rev Genet *2*, 919-929.

Edwards, S.V., Bryan Jennings, W., and Shedlock, A.M. (2005). Phylogenetics of modern birds in the era of genomics. Proc Biol Sci *272*, 979-992.

Eisenthal, R., Danson, M.J., and Hough, D.W. (2007). Catalytic efficiency and kcat/KM: a useful comparator? Trends Biotechnol *25*, 247-249.

Epstein, R.J., Lin, K., and Tan, T.W. (2000). A functional significance for codon third bases. Gene *245*, 291-298.

Evans, B.J., Kelley, D.B., Tinsley, R.C., Melnick, D.J., and Cannatella, D.C. (2004). A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. Mol Phylogenet Evol *33*, 197-213.

Farnell, Y.Z., and Ing, N.H. (2002). Estradiol and a selective estrogen receptor modulator affect steroid hormone receptor messenger RNA levels and turnover in explant cultures of sheep endometrium. In Vitro Cell Dev Biol Anim *38*, 595-600.

Fatica, A., and Tollervey, D. (2003). Insights into the structure and function of a guide RNP. Nat Struct Biol *10*, 237-239.

Fedorov, A., Saxonov, S., and Gilbert, W. (2002). Regularities of context-dependent codon bias in eukaryotic genes. Nucleic Acids Res *30*, 1192-1197.

Felsenstein, J. (1989). PHYLIP - Phylogeny Inference Package (Version 3.2). . Cladistics *5*, 164-166.

Field, K.G., Olsen, G.J., Lane, D.J., Giovannoni, S.J., Ghiselin, M.T., Raff, E.C., Pace, N.R., and Raff, R.A. (1988). Molecular phylogeny of the animal kingdom. Science *239*, 748-753.

Filipenko, N.R., and Waisman, D.M. (2001). The C terminus of annexin II mediates binding to F-actin. J Biol Chem *276*, 5310-5315.

Fredslund, J. (2006). PHY.FI: fast and easy online creation and manipulation of phylogeny color figures. BMC Bioinformatics *7*, 315.

Fretz, A., and Spindler, K.D. (1999). Hormonal regulation of actin and tubulin in an epithelial cell line from Chironomus tentans. Arch Insect Biochem Physiol *41*, 71-78.

Galperin, M.Y., Moroz, O.V., Wilson, K.S., and Murzin, A.G. (2006). House cleaning, a part of good housekeeping. Mol Microbiol *59*, 5-19.

Gardner, P.P., Daub, J., Tate, J.G., Nawrocki, E.P., Kolbe, D.L., Lindgreen, S., Wilkinson, A.C., Finn, R.D., Griffiths-Jones, S., Eddy, S.R.*, et al.* (2009). Rfam: updates to the RNA families database. Nucleic Acids Res *37*, D136-140.

Gertz, E.M., Yu, Y.K., Agarwala, R., Schaffer, A.A., and Altschul, S.F. (2006). Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. BMC Biol *4*, 41.

Ghosh, T., Peterson, B., Tomasevic, N., and Peculis, B.A. (2004). Xenopus U8 snoRNA binding protein is a conserved nuclear decapping enzyme. Mol Cell *13*, 817-828.

Glasner, M.E., Gerlt, J.A., and Babbitt, P.C. (2006). Evolution of enzyme superfamilies. Curr Opin Chem Biol *10*, 492-497.

Gonnet, G.H., Cohen, M.A., and Benner, S.A. (1992). Exhaustive matching of the entire protein sequence database. Science *256*, 1443-1445.

Gonnet, G.H., Hallett, M.T., Korostensky, C., and Bernardin, L. (2000a). Darwin v. 2.0: an interpreted computer language for the biosciences. Bioinformatics *16*, 101-103.

Gonnet, G.H., Korostensky, C., and Benner, S. (2000b). Evaluation measures of multiple sequence alignments. J Comput Biol *7*, 261-276.

Gonnet, P., and Lisacek, F. (2002). Probabilistic alignment of motifs with sequences. Bioinformatics *18*, 1091-1101.

Gosink, J.J., Woese, C.R., and Staley, J.T. (1998). Polaribacter gen. nov., with three new species, P. irgensii sp. nov., P. franzmannii sp. nov. and P. filamentus sp. nov., gas

vacuolate polar marine bacteria of the Cytophaga-Flavobacterium-Bacteroides group and reclassification of 'Flectobacillus glomeratus' as Polaribacter glomeratus comb. nov. Int J Syst Bacteriol *48 Pt 1*, 223-235.

Graham, D.E., Overbeek, R., Olsen, G.J., and Woese, C.R. (2000). An archaeal genomic signature. Proc Natl Acad Sci U S A *97*, 3304-3308.

Graziano, V., McGrath, W.J., DeGruccio, A.M., Dunn, J.J., and Mangel, W.F. (2006a). Enzymatic activity of the SARS coronavirus main proteinase dimer. FEBS Lett *580*, 2577-2583.

Graziano, V., McGrath, W.J., Yang, L., and Mangel, W.F. (2006b). SARS CoV main proteinase: The monomer-dimer equilibrium dissociation constant. Biochemistry *45*, 14632-14641.

Griffiths-Jones, S. (2005). Annotating non-coding RNAs with Rfam. Curr Protoc Bioinformatics *Chapter 12*, Unit 12 15.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S.R. (2003). Rfam: an RNA family database. Nucleic Acids Res *31*, 439-441.

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S.R., and Bateman, A. (2005). Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res *33*, D121-124.

Gu, M., Fabrega, C., Liu, S.W., Liu, H., Kiledjian, M., and Lima, C.D. (2004). Insights into the structure, mechanism, and regulation of scavenger mRNA decapping activity. Mol Cell *14*, 67-80.

Gu, X., and Vander Velden, K. (2002). DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. Bioinformatics *18*, 500-501.

Guhaniyogi, J., and Brewer, G. (2001). Regulation of mRNA stability in mammalian cells. Gene *265*, 11-23.

Gustafson, T.A., Bahl, J.J., Markham, B.E., Roeske, W.R., and Morkin, E. (1987). Hormonal regulation of myosin heavy chain and alpha-actin gene expression in cultured fetal rat heart myocytes. J Biol Chem *262*, 13316-13322.

Hajibabaei, M., Singer, G.A., Hebert, P.D., and Hickey, D.A. (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet *23*, 167-172.

Hall, B.G. (2001). Phylogenetic trees made easy: a how-to manual for molecular biologists (Sunderland, Massachusetts, Sinauer Associates, Inc.).

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT Nucleic Acids Symposium Series *41*.

Hamad, A., Kluk, M., Fox, J., Park, M., and Turner, J.E. (2007). The effects of aromatase inhibitors and selective estrogen receptor modulators on eye development in the zebrafish (Danio rerio). Curr Eye Res *32*, 819-827.

Hayes, M.J., Rescher, U., Gerke, V., and Moss, S.E. (2004). Annexin-actin interactions. Traffic *5*, 571-576.

Hayes, M.J., Shao, D., Bailly, M., and Moss, S.E. (2006). Regulation of actin dynamics by annexin 2. EMBO J *25*, 1816-1826.

Henikoff, S., and Henikoff, J.G. (1992). Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A *89*, 10915-10919.

Henras, A.K., Dez, C., and Henry, Y. (2004). RNA structure and function in C/D and H/ACA s(no)RNPs. Curr Opin Struct Biol *14*, 335-343.

Henry, E.A., Devereux, R., Maki, J.S., Gilmour, C.C., Woese, C.R., Mandelco, L., Schauder, R., Remsen, C.C., and Mitchell, R. (1994). Characterization of a new thermophilic sulfate-reducing bacterium Thermodesulfovibrio yellowstonii, gen. nov. and sp. nov.: its phylogenetic relationship to Thermodesulfobacterium commune and their origins deep within the bacterial domain. Arch Microbiol *161*, 62-69.

Higgins, D.G., Thompson, J.D., and Gibson, T.J. (1996). Using CLUSTAL for multiple sequence alignments. Methods Enzymol *266*, 383-402.

Hillis, D.M. (1987). Molecular versus morphological approaches to systematics. Annual Review of Ecology and Systematics *18*, 23-42.

Holbrook, E.L., Schulze-Gahmen, U., Buchko, G.W., Ni, S., Kennedy, M.A., and Holbrook, S.R. (2003). Purification, crystallization and preliminary X-ray analysis of two nudix hydrolases from Deinococcus radiodurans. Acta Crystallogr D Biol Crystallogr *59*, 737-740.

Hsu, H.H., Cheng, S.F., Chen, L.M., Liu, J.Y., Chu, C.H., Weng, Y.J., Li, Z.Y., Lin, C.S., Lee, S.D., Kuo, W.W.*, et al.* (2006). Over-expressed estrogen receptor-alpha up-regulates hTNF-alpha gene expression and down-regulates beta-catenin signaling activity to induce the apoptosis and inhibit proliferation of LoVo colon cancer cells. Mol Cell Biochem *289*, 101-109.

Ing, N.H. (2005). Steroid hormones regulate gene expression posttranscriptionally by altering the stabilities of messenger RNAs. Biol Reprod *72*, 1290-1296.

Ing, N.H., Wolfskill, R.L., Clark, S., DeGraauw, J.A., and Gill, C.A. (2006). Steroid hormones acutely regulate expression of a Nudix protein-encoding gene in the endometrial epithelium of sheep. Mol Reprod Dev *73*, 967-976.

Jacobson, A. (2004). Regulation of mRNA decay: decapping goes solo. Mol Cell *15*, 1-2.

Jacobson, M.R., and Pederson, T. (1998). A 7-methylguanosine cap commits U3 and U8 small nuclear RNAs to the nucleolar localization pathway. Nucleic Acids Res *26*, 756-760.

Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G., and Gibson, T.J. (1998). Multiple sequence alignment with Clustal X. Trends Biochem Sci *23*, 403-405.

Jin, Y., Wang, W., Sheng, G.D., Liu, W., and Fu, Z. (2007). Hepatic and extrahepatic expression of estrogen-responsive genes in male adult zebrafish (Danio rerio) as biomarkers of short-term exposure to 17beta-estradiol. Environ Monit Assess.

John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., and Marks, D.S. (2004). Human MicroRNA targets. PLoS Biol *2*, e363.

Johnson, A.W. (1997). Rat1p and Xrn1p are functionally interchangeable exoribonucleases that are restricted to and required in the nucleus and cytoplasm, respectively. Mol Cell Biol *17*, 6122-6130.

Jorgensen, A., Andersen, O., Bjerregaard, P., and Rasmussen, L.J. (2007). Identification and characterisation of an androgen receptor from zebrafish Danio rerio. Comp Biochem Physiol C Toxicol Pharmacol *146*, 561-568.

Kalek, M., Jemielity, J., Darzynkiewicz, Z.M., Bojarska, E., Stepinski, J., Stolarski, R., Davis, R.E., and Darzynkiewicz, E. (2006). Enzymatically stable 5' mRNA cap analogs: synthesis and binding studies with human DcpS decapping enzyme. Bioorg Med Chem *14*, 3223-3230.

Kastenmayer, J.P., and Green, P.J. (2000). Novel features of the XRN-family in Arabidopsis: evidence that AtXRN4, one of several orthologs of nuclear Xrn2p/Rat1p, functions in the cytoplasm. Proc Natl Acad Sci U S A *97*, 13985-13990.

Katzenellenbogen, B.S. (1996). Estrogen receptors: bioactivities and interactions with cell signaling pathways. Biol Reprod *54*, 287-293.

Kausch, U., Alberti, M., Haindl, S., Budczies, J., and Hock, B. (2008). Biomarkers for exposure to estrogenic compounds: gene expression analysis in zebrafish (Danio rerio). Environ Toxicol *23*, 15-24.

Khanna, R., and Kiledjian, M. (2004). Poly(A)-binding-protein-mediated regulation of hDcp2 decapping in vitro. EMBO J *23*, 1968-1976.

Kimura, K., Takeda, A., and Uchida, T.A. (1987). Changes in progesterone concentrations in the Japanese long-fingered bat, Miniopterus schreibersii fuliginosus. J Reprod Fertil *80*, 59-63.

Kiss-Laszlo, Z., Henry, Y., and Kiss, T. (1998). Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. EMBO J *17*, 797-807.

Kiss, A.M., Jady, B.E., Darzacq, X., Verheggen, C., Bertrand, E., and Kiss, T. (2002). A Cajal body-specific pseudouridylation guide RNA is composed of two box H/ACA snoRNA-like domains. Nucleic Acids Res *30*, 4643-4649.

Kiss, T. (2001). Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. EMBO J *20*, 3617-3622.

Kiss, T. (2002). Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. Cell *109*, 145-148.

Kofuji, S., Sakuno, T., Takahashi, S., Araki, Y., Doi, Y., Hoshino, S., and Katada, T. (2006). The decapping enzyme Dcp1 participates in translation termination through its interaction with the release factor eRF3 in budding yeast. Biochem Biophys Res Commun *344*, 547-553.

Kohler, H.R., Kloas, W., Schirling, M., Lutz, I., Reye, A.L., Langen, J.S., Triebskorn, R., Nagel, R., and Schonfelder, G. (2007). Sex steroid receptor evolution and signalling in aquatic invertebrates. Ecotoxicology *16*, 131-143.

Korostensky, C., and Gonnet, G.H. (2000). Using traveling salesman problem algorithms for evolutionary tree construction. Bioinformatics *16*, 619-627.

Koval, V.V., Kuznetsov, N.A., Zharkov, D.O., Ishchenko, A.A., Douglas, K.T., Nevinsky, G.A., and Fedorova, O.S. (2004). Pre-steady-state kinetics shows differences in processing of various DNA lesions by Escherichia coli formamidopyrimidine-DNA glycosylase. Nucleic Acids Res *32*, 926-935.

Krane, D.E., and Raymer, M.L. (2003). Fundamental concepts of bioinformatics (San Francisco CA, Pearson Education Inc.).

Kshirsagar, M., and Parker, R. (2004). Identification of Edc3p as an enhancer of mRNA decapping in Saccharomyces cerevisiae. Genetics *166*, 729-739.

Kuai, L., Das, B., and Sherman, F. (2005). A nuclear degradation pathway controls the abundance of normal mRNAs in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *102*, 13962-13967.

Kufel, J., Bousquet-Antonelli, C., Beggs, J.D., and Tollervey, D. (2004). Nuclear pre-mRNA decapping and 5' degradation in yeast require the Lsm2-8p complex. Mol Cell Biol *24*, 9646-9657.

Kurganov, B.I., Lobanov, A.V., Borisov, I.A., and Reshetilov, A.N. (2001). Criterion for Hill equation validity for description of biosensor calibration curves. Analytica Chimica Acta *427*, 11-19.

Kurihara, Y., and Watanabe, Y. (2004). Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. Proc Natl Acad Sci U S A *101*, 12753-12758.

Lafontaine, D.L., Bousquet-Antonelli, C., Henry, Y., Caizergues-Ferrer, M., and Tollervey, D. (1998). The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. Genes Dev *12*, 527-537.

Lafontaine, D.L., and Tollervey, D. (1998). Birth of the snoRNPs: the evolution of the modification-guide snoRNAs. Trends Biochem Sci *23*, 383-388.

Lall, S., Piano, F., and Davis, R.E. (2005). Caenorhabditis elegans decapping proteins: localization and functional analysis of Dcp1, Dcp2, and DcpS during embryogenesis. Mol Biol Cell *16*, 5880-5890.

Lange, T.S., Borovjagin, A., Maxwell, E.S., and Gerbi, S.A. (1998a). Conserved boxes C and D are essential nucleolar localization elements of U14 and U8 snoRNAs. EMBO J *17*, 3176-3187.

Lange, T.S., Borovjagin, A.V., and Gerbi, S.A. (1998b). Nucleolar localization elements in U8 snoRNA differ from sequences required for rRNA processing. RNA *4*, 789-800.

Larsen, L.K., Andreasen, P.H., Dreisig, H., Palm, L., Nielsen, H., Engberg, J., and Kristiansen, K. (1999). Cloning and characterization of the gene encoding the highly expressed ribosomal protein l3 of the ciliated protozoan Tetrahymena thermophila. Evidence for differential codon usage in highly expressed genes. Cell Biol Int *23*, 551-560.

Lawrence, C.J., Zmasek, C.M., Dawe, R.K., and Malmberg, R.L. (2004). LumberJack: a heuristic tool for sequence alignment exploration and phylogenetic inference. Bioinformatics *20*, 1977-1979.

Leavitt, W.W., Cobb, A.D., and Takeda, A. (1987). Progesterone-modulation of estrogen action: rapid down regulation of nuclear acceptor sites for the estrogen receptor. Adv Exp Med Biol *230*, 49-78.

Leavitt, W.W., Takeda, A., and MacDonald, R.G. (1986). Progesterone regulation of protein synthesis and steroid receptor levels in decidual cells. Ann N Y Acad Sci *476*, 136-157.

Lee, B.K., Park, M.R., Srinivas, B., Chun, J.C., Kwon, I.S., Chung, I.M., Yoo, N.H., Choi, K.G., and Yun, S.J. (2003). Induction of phenylalanine ammonia-lyase gene expression by paraquat and stress-related hormones in Rehmannia glutinosa. Mol Cells *16*, 34-39.

Lehner, B., and Sanderson, C.M. (2004). A protein interaction framework for human mRNA degradation. Genome Res *14*, 1315-1323.

Lejeune, F., Li, X., and Maquat, L.E. (2003). Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities. Mol Cell *12*, 675-687.

Lestrade, L., and Weber, M.J. (2006). snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. Nucleic Acids Res *34*, D158-162.

Lin, C.Y., Lin, F.K., Lin, C.H., Lai, L.W., Hsu, H.J., Chen, S.H., and Hsiung, C.A. (2005). POWER: PhylOgenetic WEb Repeater--an integrated and user-optimized framework for biomolecular phylogenetic analysis. Nucleic Acids Res *33*, W553-556.

Lindstrom, D.L., Squazzo, S.L., Muster, N., Burckin, T.A., Wachter, K.C., Emigh, C.A., McCleery, J.A., Yates, J.R., 3rd, and Hartzog, G.A. (2003). Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. Mol Cell Biol *23*, 1368-1378.

Liu, D., Hinshelwood, M.M., Giguere, V., and Mendelson, C.R. (2006a). Estrogen related receptor-alpha enhances surfactant protein-A gene expression in fetal lung type II cells. Endocrinology *147*, 5187-5195.

Liu, H., Rodgers, N.D., Jiao, X., and Kiledjian, M. (2002). The scavenger mRNA decapping enzyme DcpS is a member of the HIT family of pyrophosphatases. EMBO J *21*, 4699-4708.

Liu, J., and Vishwanatha, J.K. (2007). Regulation of nucleo-cytoplasmic shuttling of human annexin A2: a proposed mechanism. Mol Cell Biochem *303*, 211-220.

Liu, S.W., Jiao, X., Liu, H., Gu, M., Lima, C.D., and Kiledjian, M. (2004). Functional analysis of mRNA scavenger decapping enzymes. RNA *10*, 1412-1422.

Liu, S.W., Rajagopal, V., Patel, S.S., and Kiledjian, M. (2008). Mechanistic and Kinetic Analysis of the DcpS Scavenger Decapping Enzyme. J Biol Chem *283*, 16427-16436.

Liu, W., Konduri, S.D., Bansal, S., Nayak, B.K., Rajasekaran, S.A., Karuppayil, S.M., Rajasekaran, A.K., and Das, G.M. (2006b). Estrogen receptor-alpha binds p53 tumor suppressor protein directly and represses its function. J Biol Chem *281*, 9837-9840.

Lowartz, S., Petkam, R., Renaud, R., Beamish, F.W., Kime, D.E., Raeside, J., and Leatherland, J.F. (2003). Blood steroid profile and in vitro steroidogenesis by ovarian follicles and testis fragments of adult sea lamprey, Petromyzon marinus. Comp Biochem Physiol A Mol Integr Physiol *134*, 365-376.

Mak, P., and Callard, G.V. (1985). Characterization of estrogen receptors in the hamster brain. J Steroid Biochem *22*, 355-361.

Mandal, S.S., Chu, C., Wada, T., Handa, H., Shatkin, A.J., and Reinberg, D. (2004). Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. Proc Natl Acad Sci U S A *101*, 7572-7577.

Marshall, E. (2005). Taxonomy. Will DNA bar codes breathe life into classification? Science *307*, 1037.

Massingham, T. (2008). Detecting the presence and location of selection in proteins. Methods Mol Biol *452*, 311-329.

Massingham, T., and Goldman, N. (2005). Detecting amino acid sites under positive selection and purifying selection. Genetics *169*, 1753-1762.

Matera, A.G., Tycowski, K.T., Steitz, J.A., and Ward, D.C. (1994). Organization of small nucleolar ribonucleoproteins (snoRNPs) by fluorescence in situ hybridization and immunocytochemistry. Mol Biol Cell *5*, 1289-1299.

Maxwell, E.S., and Fournier, M.J. (1995). The small nucleolar RNAs. Annu Rev Biochem *64*, 897-934.

McInerney, E.M., Tsai, M.J., O'Malley, B.W., and Katzenellenbogen, B.S. (1996). Analysis of estrogen receptor transcriptional enhancement by a nuclear hormone receptor coactivator. Proc Natl Acad Sci U S A *93*, 10069-10073.

McLennan, A.G. (1999). The MutT motif family of nucleotide phosphohydrolases in man and human pathogens (review). Int J Mol Med *4*, 79-89.

McLennan, A.G. (2006). The Nudix hydrolase superfamily. Cell Mol Life Sci *63*, 123-143.

McLennan, A.G. (2007). Decapitation: poxvirus makes RNA lose its head. Trends Biochem Sci *32*, 297-299.

Meyer, A., and Zardoya, R. (2003). Recent advances in the (molecular) phylogeny of vertebrates. Annu Rev Ecol Evol Syst *34*, 311-338.

Meyer, S., Temme, C., and Wahle, E. (2004). Messenger RNA turnover in eukaryotes: pathways and enzymes. Crit Rev Biochem Mol Biol *39*, 197-216.

Mildvan, A.S., Xia, Z., Azurmendi, H.F., Saraswat, V., Legler, P.M., Massiah, M.A., Gabelli, S.B., Bianchet, M.A., Kang, L.W., and Amzel, L.M. (2005). Structures and mechanisms of Nudix hydrolases. Arch Biochem Biophys *433*, 129-143.

Mitchell, P., and Tollervey, D. (2000). mRNA stability in eukaryotes. Curr Opin Genet Dev *10*, 193-198.

Moteki, S., and Price, D. (2002). Functional coupling of capping and transcription of mRNA. Mol Cell *10*, 599-609.

Mullinix, K.P., Wetekam, W., Deeley, R.G., Gordon, J.I., Meyers, M., Kent, K.A., and Goldberger, R.F. (1976). Induction of vitellogenin synthesis by estrogen in avian liver:

relationship between level of vitellogenin mRNA and vitellogenin synthesis. Proc Natl Acad Sci U S A *73*, 1442-1446.

Ni, S., Woese, C.R., Aldrich, H.C., and Boone, D.R. (1994). Transfer of Methanolobus siciliae to the genus Methanosarcina, naming it Methanosarcina siciliae, and emendation of the genus Methanosarcina. Int J Syst Bacteriol *44*, 357-359.

Ntukidem, N.I., Nguyen, A.T., Stearns, V., Rehman, M., Schott, A., Skaar, T., Jin, Y., Blanche, P., Li, L., Lemler, S*., et al.* (2008). Estrogen receptor genotypes, menopausal status, and the lipid effects of tamoxifen. Clin Pharmacol Ther *83*, 702-710.

Page, R.D. (1996). TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci *12*, 357-358.

Pal, U., Biswas, S.C., and Sarkar, P.K. (1997). Regulation of actin and its mRNA by thyroid hormones in cultures of fetal human brain during second trimester of gestation. J Neurochem *69*, 1170-1176.

Palma, A.C., Araujo, F., Duque, V., Borges, F., Paixao, M.T., and Camacho, R. (2007). Molecular epidemiology and prevalence of drug resistance-associated mutations in newly diagnosed HIV-1 patients in Portugal. Infect Genet Evol *7*, 391-398.

Pang, Y., Dong, J., and Thomas, P. (2008). Estrogen signaling characteristics of Atlantic croaker G protein-coupled receptor 30 (GPR30) and evidence it is involved in maintenance of oocyte meiotic arrest. Endocrinology *149*, 3410-3426.

Parker, R., and Song, H. (2004). The enzymes and control of eukaryotic mRNA turnover. Nat Struct Mol Biol *11*, 121-127.

Parrish, S., and Moss, B. (2007). Characterization of a Second Vaccinia Virus mRNA Decapping Enzyme Conserved in Poxviruses. J Virol.

Parrish, S., Resch, W., and Moss, B. (2007). Vaccinia virus D10 protein has mRNA decapping activity, providing a mechanism for control of host and viral gene expression. Proc Natl Acad Sci U S A *104*, 2139-2144.

Peculis, B. (1997). RNA processing: pocket guides to ribosomal RNA. Curr Biol *7*, R480-482.

Peculis, B.A. (2000). RNA-binding proteins: if it looks like a sn(o)RNA. Curr Biol *10*, R916-918.

Peculis, B.A. (2001). snoRNA nuclear import and potential for cotranscriptional function in pre-rRNA processing. RNA *7*, 207-219.

Peculis, B.A., DeGregorio, S., and McDowell, K. (2001). The U8 snoRNA gene family: identification and characterization of distinct, functional U8 genes in Xenopus. Gene *274*, 83-92.

Peculis, B.A., and Mount, S.M. (1996). Ribosomal RNA: small nucleolar RNAs make their mark. Curr Biol *6*, 1413-1415.

Peculis, B.A., Reynolds, K., and Cleland, M. (2007). Metal determines efficiency and substrate specificity of the nuclear NUDIX decapping proteins X29 and H29K (Nudt16). J Biol Chem *282*, 24792-24805.

Peculis, B.A., Scarsdale, J.N., and Wright, H.T. (2004). Crystals of X29, a Xenopus laevis U8 snoRNA-binding protein with nuclear decapping activity. Acta Crystallogr D Biol Crystallogr *60*, 1668-1669.

Peculis, B.A., and Steitz, J.A. (1993). Disruption of U8 nucleolar snRNA inhibits 5.8S and 28S rRNA processing in the Xenopus oocyte. Cell *73*, 1233-1245.

Peculis, B.A., and Steitz, J.A. (1994). Sequence and structural elements critical for U8 snRNP function in Xenopus oocytes are evolutionarily conserved. Genes Dev *8*, 2241-2255.

Peel, A.D. (2008). The evolution of developmental gene networks: lessons from comparative studies on holometabolous insects. Philos Trans R Soc Lond B Biol Sci *363*, 1539-1547.

Philippe, H., Lartillot, N., and Brinkmann, H. (2005). Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. Mol Biol Evol *22*, 1246-1253.

Philippe, H., Snell, E.A., Bapteste, E., Lopez, P., Holland, P.W., and Casane, D. (2004). Phylogenomics of eukaryotes: impact of missing data on large alignments. Mol Biol Evol *21*, 1740-1752.

Piccirillo, C., Khanna, R., and Kiledjian, M. (2003). Functional characterization of the mammalian mRNA decapping enzyme hDcp2. RNA *9*, 1138-1147.

Pollack, B.P., Kotenko, S.V., He, W., Izotova, L.S., Barnoski, B.L., and Pestka, S. (1999). The human homologue of the yeast proteins Skb1 and Hsl7p interacts with Jak kinases and contains protein methyltransferase activity. J Biol Chem *274*, 31531-31542.

Porello, S.L., Leyes, A.E., and David, S.S. (1998). Single-turnover and pre-steady-state kinetics of the reaction of the adenine glycosylase MutY with mismatch-containing DNA substrates. Biochemistry *37*, 14756-14764.

Pudney, J., Canick, J.A., Clifford, N.M., Knapp, J.B., and Callard, G.V. (1985). Location of enzymes of androgen and estrogen biosynthesis in the testis of the ground squirrel (Citellus lateralis). Biol Reprod *33*, 971-980.

Puri, R.K., and Toft, D.O. (1986). Peptide mapping analysis of the avian progesterone receptor. J Biol Chem *261*, 5651-5657.

Rashid, R., Aittaleb, M., Chen, Q., Spiegel, K., Demeler, B., and Li, H. (2003). Functional requirement for symmetric assembly of archaeal box C/D small ribonucleoprotein particles. J Mol Biol *333*, 295-306.

Ratnasingham, S., and Hebert, P.D. (2007). bold: The Barcode of Life Data System (http://www.barcodinglife.org). Mol Ecol Notes *7*, 355-364.

Reddy, R., Henning, D., and Busch, H. (1985). Primary and secondary structure of U8 small nuclear RNA. J Biol Chem *260*, 10930-10935.

Rehwinkel, J., Behm-Ansmant, I., Gatfield, D., and Izaurralde, E. (2005). A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. RNA *11*, 1640-1647.

Reichow, S.L., Hamma, T., Ferre-D'Amare, A.R., and Varani, G. (2007). The structure and function of small nucleolar ribonucleoproteins. Nucleic Acids Res *35*, 1452-1464.

Retief, J.D. (2000). Phylogenetic analysis using PHYLIP. Methods Mol Biol *132*, 243-258.

Richard, P., Darzacq, X., Bertrand, E., Jady, B.E., Verheggen, C., and Kiss, T. (2003). A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. EMBO J *22*, 4283-4293.

Rouviere, P., Mandelco, L., Winker, S., and Woese, C.R. (1992). A detailed phylogeny for the Methanomicrobiales. Syst Appl Microbiol *15*, 363-371.

Roy, S., De, J., Kundu, S., Biswas, A., Pramanik, M., and Ray, A.K. (2007). Estradiol-17beta: tracing its metabolic significance in female fatbody of fifth instar larvae of silkworm, Bombyx mori L (race: Nistari). Life Sci *80*, 446-453.

Sahu, A., Pal, D., and Maiti, B.R. (1981). Effect of sex-hormones on the uropygial gland of the female domestic duckling. Mikroskopie *38*, 204-212.

Saier, M.H., Jr. (1995). Differential codon usage: a safeguard against inappropriate expression of specialized genes? FEBS Lett *362*, 1-4.

Scarsdale, J.N., Peculis, B.A., and Wright, H.T. (2006). Crystal structures of U8 snoRNA decapping nudix hydrolase, X29, and its metal and cap complexes. Structure *14*, 331-343.

Schneider, A., Cannarozzi, G.M., and Gonnet, G.H. (2005). Empirical codon substitution matrix. BMC Bioinformatics *6*, 134.

Scott, G.K., Kushner, P., Vigne, J.L., and Benz, C.C. (1991). Truncated forms of DNA-binding estrogen receptors in human breast cancer. J Clin Invest *88*, 700-706.

Shen, V., Liu, H., Liu, S.W., Jiao, X., and Kiledjian, M. (2008). DcpS scavenger decapping enzyme can modulate pre-mRNA splicing. RNA *14*, 1132-1142.

Sheth, U., and Parker, R. (2003). Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. Science *300*, 805-808.

Sheth, U., and Parker, R. (2006). Targeting of aberrant mRNAs to cytoplasmic processing bodies. Cell *125*, 1095-1109.

Shimkets, L., and Woese, C.R. (1992). A phylogenetic analysis of the myxobacteria: basis for their classification. Proc Natl Acad Sci U S A *89*, 9459-9463.

Simmons, M.P., and Ochoterena, H. (2000). Gaps as characters in sequence-based phylogenetic analyses. Syst Biol *49*, 369-381.

Simmons, M.P., Ochoterena, H., and Carr, T.G. (2001). Incorporation, relative homoplasy, and effect of gap characters in sequence-based phylogenetic analyses. Syst Biol *50*, 454-462.

Simmons, M.P., Ochoterena, H., and Freudenstein, J.V. (2002). Amino acid vs. nucleotide characters: challenging preconceived notions. Mol Phylogenet Evol *24*, 78-90.

Smalheiser, N.R. (2003). EST analyses predict the existence of a population of chimeric microRNA precursor-mRNA transcripts expressed in normal human and mouse tissues. Genome Biol *4*, 403.

Smith, C.M., and Steitz, J.A. (1997). Sno storm in the nucleolus: new roles for myriad small RNPs. Cell *89*, 669-672.

Soifer, H.S., Rossi, J.J., and Saetrom, P. (2007). MicroRNAs in Disease and Potential Therapeutic Applications. Mol Ther.

Speckmann, W.A., Terns, R.M., and Terns, M.P. (2000). The box C/D motif directs snoRNA 5'-cap hypermethylation. Nucleic Acids Res *28*, 4467-4473.

Stark, A., Lin, M.F., Kheradpour, P., Pedersen, J.S., Parts, L., Carlson, J.W., Crosby, M.A., Rasmussen, M.D., Roy, S., Deoras, A.N.*, et al.* (2007). Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. Nature *450*, 219-232.

Staton, J.M., Thomson, A.M., and Leedman, P.J. (2000). Hormonal regulation of mRNA stability and RNA-protein interactions in the pituitary. J Mol Endocrinol *25*, 17-34.

Steitz, J.A., and Tycowski, K.T. (1995). Small RNA chaperones for ribosome biogenesis. Science *270*, 1626-1627.

Stevens, J.R., and Schofield, C.J. (2003). Phylogenetics and sequence analysis--some problems for the unwary. Trends Parasitol *19*, 582-588.

Stuart, G.W., Moffett, K., and Leader, J.J. (2002). A Comprehensive Vertebrate Phylogeny Using Vector Representations of Protein Sequences from Whole Genomes. Mol Biol Evol *19*, 554-562.

Sullivan, C.S., and Ganem, D. (2005). MicroRNAs and viral infection. Mol Cell *20*, 3-7.

Tada, N., Saka, M., Shiraishi, F., and Kamata, Y. (2007). A field study on serum vitellogenin levels in male Reeves' pond turtles (Chinemys reevesii) from estrogen-contaminated sites and a reference site. Sci Total Environ *384*, 205-213.

Takeda, A. (1988). Progesterone and antiprogesterone (RU 38486) modulation of estrogen-inducible glycoprotein (USP-1) synthesis and secretion in rat uterine epithelial cells. Endocrinology *122*, 1559-1564.

Takeda, N., Kubokawa, K., and Matsumoto, G. (2003). Immunoreactivity for progesterone in the giant Rohde cells of the amphioxus, Branchiostoma belcheri. Gen Comp Endocrinol *132*, 379-383.

Talmor-Neiman, M., Stav, R., Frank, W., Voss, B., and Arazi, T. (2006). Novel micro-RNAs and intermediates of micro-RNA biogenesis from moss. Plant J *47*, 25-37.

Taylor, E.A., Rinaldo-Matthis, A., Li, L., Ghanem, M., Hazleton, K.Z., Cassera, M.B., Almo, S.C., and Schramm, V.L. (2007). Anopheles gambiae purine nucleoside phosphorylase: catalysis, structure, and inhibition. Biochemistry *46*, 12405-12415.

Taylor, M.J., and Peculis, B.A. (2008). Evolutionary conservation supports ancient origin for Nudt16, a nuclear-localized, RNA-binding, RNA-decapping enzyme. Nucleic Acids Res *36*, 6021-6034.

Thompson, D.M., and Parker, R. (2007). Cytoplasmic decay of intergenic transcripts in Saccharomyces cerevisiae. Mol Cell Biol *27*, 92-101.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res *25*, 4876-4882.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res *22*, 4673-4680.

Thornton, J.W. (2001). Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. Proc Natl Acad Sci U S A *98*, 5671-5676.

Tomasevic, N., and Peculis, B. (1999). Identification of a U8 snoRNA-specific binding protein. J Biol Chem *274*, 35914-35920.

Tomasevic, N., and Peculis, B.A. (2002). Xenopus LSm proteins bind U8 snoRNA via an internal evolutionarily conserved octamer sequence. Mol Cell Biol *22*, 4101-4112.

Tyc, K., and Steitz, J.A. (1989). U3, U8 and U13 comprise a new class of mammalian snRNPs localized in the cell nucleolus. EMBO J *8*, 3113-3119.

van Dijk, E., Cougot, N., Meyer, S., Babajko, S., Wahle, E., and Seraphin, B. (2002). Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. EMBO J *21*, 6915-6924.

van Dijk, E., Le Hir, H., and Seraphin, B. (2003). DcpS can act in the 5'-3' mRNA decay pathway in addition to the 3'-5' pathway. Proc Natl Acad Sci U S A *100*, 12081-12086.

Vilela, C., Velasco, C., Ptushkina, M., and McCarthy, J.E. (2000). The eukaryotic mRNA decapping protein Dcp1 interacts physically and functionally with the eIF4F translation initiation complex. EMBO J *19*, 4372-4382.

Wang, S., Mura, C., Sawaya, M.R., Cascio, D., and Eisenberg, D. (2002a). Structure of a Nudix protein from Pyrobaculum aerophilum reveals a dimer with two intersubunit beta-sheets. Acta Crystallogr D Biol Crystallogr *58*, 571-578.

Wang, Z., Jiao, X., Carr-Schmid, A., and Kiledjian, M. (2002b). The hDcp2 protein is a mammalian mRNA decapping enzyme. Proc Natl Acad Sci U S A *99*, 12663-12668.

Wang, Z., and Kiledjian, M. (2001). Functional link between the mammalian exosome and mRNA decapping. Cell *107*, 751-762.

Weinstein, L.B., and Steitz, J.A. (1999). Guided tours: from precursor snoRNA to functional snoRNP. Curr Opin Cell Biol *11*, 378-384.

Weisburg, W.G., Oyaizu, Y., Oyaizu, H., and Woese, C.R. (1985). Natural relationship between bacteroides and flavobacteria. J Bacteriol *164*, 230-236.

West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature *432*, 522-525.

Wierzchowski, J., Pietrzak, M., Stepinski, J., Jemielity, J., Kalek, M., Bojarska, E., Jankowska-Anyszka, M., Davis, R.E., and Darzynkiewicz, E. (2007). Kinetics of C. elegans DcpS cap hydrolysis studied by fluorescence spectroscopy. Nucleosides Nucleotides Nucleic Acids *26*, 1211-1215.

Winnepenninckx, B.M., Van de Peer, Y., and Backeljau, T. (1998). Metazoan relationships on the basis of 18S rRNA sequences: a few years later... J Amer Zool *38*, 888-906.

Woese, C.R. (2000). Interpreting the universal phylogenetic tree. Proc Natl Acad Sci U S A *97*, 8392-8396.

Woese, C.R., Maloy, S., Mandelco, L., and Raj, H.D. (1990). Phylogenetic placement of the Spirosomaceae. Syst Appl Microbiol *13*, 19-23.

Wu, W., Niles, E., Hirai, H., and LoVerde, P. (2007). Evolution of a novel subfamily of nuclear receptors with members that each contain two DNA binding domains. *7*, 27 %U http://www.biomedcentral.com/1471-2148/1477/1427.

Wu, Y.T., and Liu, J.Y. (2005). Molecular cloning and characterization of a cotton glucuronosyltranferase gene. J Plant Physiol *162*, 573-582.

Wu, Z.F., Liu, D.W., Wang, Q.L., Zeng, H.Y., Chen, Y.S., and Zhang, H. (2006). Study on the association between estrogen receptor gene (ESR) and reproduction traits in Landrace pigs. Yi Chuan Xue Bao *33*, 711-716.

Wulczyn, F.G., Smirnova, L., Rybak, A., Brandt, C., Kwidzinski, E., Ninnemann, O., Strehle, M., Seiler, A., Schumacher, S., and Nitsch, R. (2007). Post-transcriptional regulation of the let-7 microRNA during neural cell specification. FASEB J *21*, 415-426.

Xu, J., Yang, J.Y., Niu, Q.W., and Chua, N.H. (2006). Arabidopsis DCP2, DCP1, and VARICOSE form a decapping complex required for postembryonic development. Plant Cell *18*, 3386-3398.

Yanagihara, N., Liu, M., Toyohira, Y., Tsutsui, M., Ueno, S., Shinohara, Y., Takahashi, K., and Tanaka, K. (2006). Stimulation of catecholamine synthesis through unique estrogen receptors in the bovine adrenomedullary plasma membrane by 17beta-estradiol. Biochem Biophys Res Commun *339*, 548-553.

Yi, S., Bernat, B., Pal, G., Kossiakoff, A., and Li, W.H. (2002). Functional promiscuity of squirrel monkey growth hormone receptor toward both primate and nonprimate growth hormones. Mol Biol Evol *19*, 1083-1092.

Zhang, L., Liu, X., Zhang, J., Cao, R., Lin, Y., Xie, J., Chen, P., Sun, Y., Li, D., and Liang, S. (2006). Proteome analysis of combined effects of androgen and estrogen on the mouse mammary gland. Proteomics *6*, 487-497.

Zhao, H., Yang, D., Woese, C.R., and Bryant, M.P. (1993). Assignment of fatty acid-beta-oxidizing syntrophic bacteria to Syntrophomonadaceae fam. nov. on the basis of 16S rRNA sequence analyses. Int J Syst Bacteriol *43*, 278-286.

Zhao, L., Jin, C., Mao, Z., Gopinathan, M.B., Rehder, K., and Brinton, R.D. (2007). Design, synthesis, and estrogenic activity of a novel estrogen receptor modulator--a hybrid

structure of 17beta-estradiol and vitamin E in hippocampal neurons. J Med Chem *50*, 4471-4481.

Zmasek, C.M., and Eddy, S.R. (2001). ATV: display and manipulation of annotated phylogenetic trees. Bioinformatics *17*, 383-384.

Zwieb, C. (1997). The uRNA database. Nucleic Acids Res *25*, 102-103.

VITA

Melissa Taylor was born in Creston, IA in April 1983. During her school years, she prospered becoming known for her intellect and ability to read quickly. She was continually on the High Honor Roll and was inducted to the National Honor Society. She graduated valedictorian of her class with top honors in her classes.

Melissa was accepted into the Oxbridge Honors Program at William Jewell College and began working towards her goal of being a doctor. In 2005, Melissa chose the University of Missouri – Columbia as the place to do her graduate work. There, Melissa began working for Dr. Brenda Peculis in Spring of 2006. Her research focused on the characterization of Nudt16 and brought with it several different types of experiments and techniques. The results of that effort and the work that she did are detailed in this document.