

# **STRUCTURAL AND FUNCTIONAL STUDIES OF hAPTX**

**STRUCTURAL AND FUNCTIONAL STUDIES OF DNA REPAIR PROTEIN**

**APRATAXIN (Aptx)**

**By**

**Arin Refat Khan, B.Sc.**

**A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the  
Requirements for the Degree Master of Science**

**McMaster University**

**© Copyright by Arin R Khan, November 2011**

Master of Science (2011)

Biochemistry and Biomedical Sciences

McMaster University

Hamilton, Ontario

Title: Structural and Functional Studies of DNA Repair protein, Aprataxin (Aptx)

Author: Arin R Khan, B.Sc.

Supervisor: Dr. Murray Junop

Number of Pages: 68

## ABSTRACT

DNA integrity is continuously compromised by cellular metabolic activity and environmental factors resulting in many lesions per cell per day (Lindahl et al., 2009). If left unrepaired or if repaired improperly, these lesions pose an obstacle to cellular transcription and replication. In due course, DNA damage may lead to mutations and genomic instability that will eventually threaten the viability of the cell and organism and can lead to tumour formation and neurological disorders (Rass et al., 2006). Aprataxin (Aptx) is a conserved factor in DNA repair required for the repair of damaged DNA following abortive DNA ligation (Date et al., 2001; Moreira et al 2001). Aptx initiates repair by processing adenylated DNA ligation events via its unique AMP lysine hydrolase activity (Rass et al 2006). Inability of Aptx to carry out these functions has been shown to be the causative agent in the neurological diseases, ataxia with oculomotor apraxia (Takahashi et al 2007, Yoon et al 2008, and Ferrarini et al., 2007). Nevertheless, the exact mechanism of Aptx has not been established and the full extent of its function(s) (i.e. binding to DNA or other proteins) is not fully understood. In order to achieve a comprehensive understanding of the mechanism governing Aptx, further functional and structural analysis is required. Here, it was found that Aptx has deadenylation activity that is further supported by the crystal structure of ATPX-DNA-AMP-Zn<sup>2+</sup>. Importantly, work reported in this thesis describes for the first time methods for expression and purification of large quantities of high quality human Aptx from bacterial cells. This protein is shown to possess robust deadenylation activity, suitable for further *in vitro* screening of small molecule inhibitors.

## ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Murray Junop in providing me with the opportunity to carry out my exciting project, guidance and supporting me throughout my Masters. Murray had an innovative approach to science that fostered creative thinking and the ability to work in an open and creative atmosphere ideal for novice scientists. Murray's suggestions provided for thought provoking ideas and discussions that turned into successful experiments. He is a truly a great mentor, both inside and outside of the laboratory and I wish him every possible future success.

I would also like to thank my supervisory committee, Dr. Justin Nodwell and Dr. Xu-Dong Zhu for their years of guidance. I am ever appreciative to their invaluable support and advice throughout my project. Xu-Dong has aided in giving me a more biological perspective on my project that allowed me to think beyond only biochemistry. I truly enjoyed working as a TA with Justin and have learned significantly from him while taking the biochemistry colloquium course.

I want to thank Dr. Alba Guarne for her support throughout the colloquim and helping me to become a better scientific writer as well as providing me with access to materials from her laboratory when required. I would like to show appreciation to Dr. Ivan Ahel, who provided us with a DNA construct of full length Aptx that was used to compare between our constructs.

I would also like to thank Dr. Sara Andres and Dr. Tracy Tieffenbach for their constant support throughout my project, being available for any questions even after hours and for teaching me radioactive experiments. I want to show my appreciation to Kritica Arora, my previous student, for her great work with cloning and purification of Aptx constructs as well as

studying the interaction of Aptx with other proteins. I am thankful to Olesya Petranko, another student in the lab, who has aided me in optimizing the Aptx activity assay required to study the function of Aptx. Furthermore, I would also like to thank Rimpay Cheema for her help in purifying Aptx constructs. I would like to thank all past and present members of the Junop lab, including, Kun, Sara, Tracy, Mac, Sean, Seiji, Kritica, Ting-Wai, Meghan, and undergraduate students who have all contributed to making the Junop lab a great place to work. I will truly miss being a member of the team, but wish the lab further success and greatness.

Finally, I would like to show my deepest gratitude to my loving husband, Saad Kabir, who stayed up those many countless nights assessing my experiments from a chemical perspective, keeping me company while concentrating protein and providing me with wonderful scientific debates that created a deep appreciation and love of science that will always be within me. I would like to thank my family, my wonderful parents and parents in-laws, my caring sister, Ashin, budding scientist, Tawsif and my friends for supporting, encouraging and rooting me on throughout my graduate studies.

# TABLE OF CONTENTS

ABSTRACT.....	iv
ACKNOWLEDGEMENTS.....	v
TABLE OF CONTENTS.....	vii
LIST OF FIGURES AND TABLES.....	x
LIST OF ABBREVIATIONS.....	xii
DECLARATION OF ACADEMIC ACHIEVEMENT.....	xiv

## CHAPTER 1- INTRODUCTION: DNA DAMAGE and REPAIR

1.1 Sources of DNA Damage .....	1
1.2 Mechanisms of DNA Double Strand Break Repair: Homologous Recombination and Non Homologous End Joining.....	5
1.3 DNA Base Excision Repair.....	9
1.4 DNA Single Stranded Break Repair.....	10
1.5 Abortive Ligation.....	12
1.6 Aprataxin.....	15
1.7 Thesis Objectives.....	18

## CHAPTER 2- MATERIALS AND METHODS

2.1 Cloning, Overexpression and Purification of Aptx and Xrcc4.....	20
2.1.1 Cloning.....	20
2.1.2 Overexpression and Purification.....	21

2.1.2.1	AptxQ.....	21
2.1.2.2	Aptx FL.....	22
2.1.2.3	Aptx FHA.....	23
2.1.2.4	Xrcc4.....	23
2.2	Crystallization of Aptx.....	24
2.3	Activity of Aptx.....	24

### CHAPTER 3- Results

3.1	Design of Protein Expression Constructs.....	26
3.2	Optimization of AptxQ Protein Purification.....	27
3.3	Optimization of Aptx FL Protein Purification.....	29
3.4	Optimization of Aptx FH Protein Purification.....	30
3.5	Optimization of Aptx Xrcc4 Protein Purification.....	31
3.6	Xrcc4-Aptx Protein Protein Interactions.....	32
3.7	Crystallization of Aptx.....	32
3.8	Aptx Activity Assay.....	33

### CHAPTER 4- DISCUSSION

4.1	Crystallization of Aptx-DNA-AMP-ZN <sup>2+</sup> .....	36
4.2	Aptx Domain Mapping and Structure Determination.....	37
4.3	Catalytic Mechanism and Structure/Function Relationships of Yeast Aptx.....	40

CHAPTER 5- SUMMARY and CONCLUSIONS.....	43
SUPPLEMENTARY and APPENDICES.....	47
REFERENCES.....	65

## LIST OF FIGURES

Figure 1.1	DNA Damage Response.....	2
Figure 1.2	NHEJ Mechanism.....	6
Figure 1.3	Xrcc4 Crystal.....	8
Figure 1.4	BER Mechanism.....	10
Figure 1.5	SSBr Mechanism.....	12
Figure 1.6	Resolution of Abortive DNA Ligation Intermediates by Aprataxin.....	14
Figure 1.7	Aptx and AOA1 Mutation.....	16
Figure 2.1	Constructs of Aptx and Xrcc4.....	19
Figure 3.1	Aptx FL Nickel Column Purification.....	28
Figure 3.2	APT X FL Sp-sepharose Column Purification.....	28
Figure 3.3	APT X FH Purification.....	30
Figure 3.4	Xrcc4 Purification.....	31
Figure 3.5	Aptx Removes Abortive Ligation Products.....	35
Figure 4.1	Domain Comparison Between Aptx human and <i>S. Pombe</i> .....	38
Figure 4.2	Overall Aptx Crystal Structure.....	39
Figure 4.3	Structure-based mechanism for two-step direct reversal of DNA 5' end adenylation.....	40
Figure 5.1	Aptx Complex Model During DNA Repair.....	46

## LIST OF SUPPLEMENTARY FIGURES AND TABLES

Table 1	MSc. Goals and Achievements.....	47
Table 2	Aptx FL Buffers.....	47

Table 3	Aptx FL Buffers.....	47
Table 4	Aptx FHA Buffers.....	47
Table 5	Aptx FHA Buffers.....	47
Figure 1	DNA protein complex formation in the crystal.....	50
Figure 2	yAptx structure specific DNA binding. YAptx structure-specific DNA binding..	52
Table 6	Data collection and refinement statistics.....	53
Figure 3	yAptx structure-based sequence alignment.....	55
Figure 4	yAptx adenylate access and catalytic mechanism.....	57
Figure 5	yAptx and Aptxc has comparable catalytic activity.....	58
Figure 6	Structure of the HIT-Znf interface.....	58
Table 7	XRCC4 peptides used in SPR experiments cover residues 229-242.....	61
Table 8	The varying FHA concentrations used for studying the interaction with the XRCC4 peptides.....	62
Table 9	Calculated $K_a$ ( $k_{on}$ ), $K_d$ ( $k_{off}$ ) and $K_D$ (equilibrium dissociation constant) obtained for the nine multiply phosphorylated XRCC4 peptides following heterogeneous analyte modeling.....	63

## LIST OF ABBREVIATIONS

AMP	Adenosine Monophosphate
AOA1	Ataxia with Oculomotor Apraxia 1
AOA2	Ataxia with Oculomotor Apraxia 2
Aptx	Aprataxin
ATP	Adenosine Triphosphate
BER	Base Excision Repair
CK2	Casein Kinase 2
DSB	Double Stranded Break
DSBr	Double Stranded Break Repair
DTT	Dithiothreitol
EDTA	Ethylenediaminetetracetic acid
Fc	Calculated Structure Factors
FHA	Forkhead domain
Fo	Observed Structure Factors
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIT	Histidine Triad Domain
HR	Homologous Recombination
LDAO	N-lauryl-N, N-dimethylamine N-oxide
IPTG	Isopropyl-thio- $\beta$ -D-galactopyranoside
kDa	Kilodalton
Lig III	DNA ligase III
Lig IV	DNA ligase IV

NHEJ	Non Homologous End Joining
NLS	Nuclear Localization Signal
OD	Optical Density
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
PNK	Polynucleotide Kinase
SDS-PAGE	Sodium-dodecyl sulphate-polyacrylamide gel electrophoresis
SSB	Single Stranded Break
SSBr	Single Stranded Break Repair
TRIS	2-amino-2hydroxymethyl-1,3-propanediol
TEV	Tobacco Etch Virus
Xrcc1	X-ray Cross Complementation Group 1
Xrcc4	X-ray Cross Complementation Group 4
V(D)J	Variable Diverse Joining
Znf	Zinc Finger Domain

## **DECLARATION OF ACADEMIC ACHIEVEMENT**

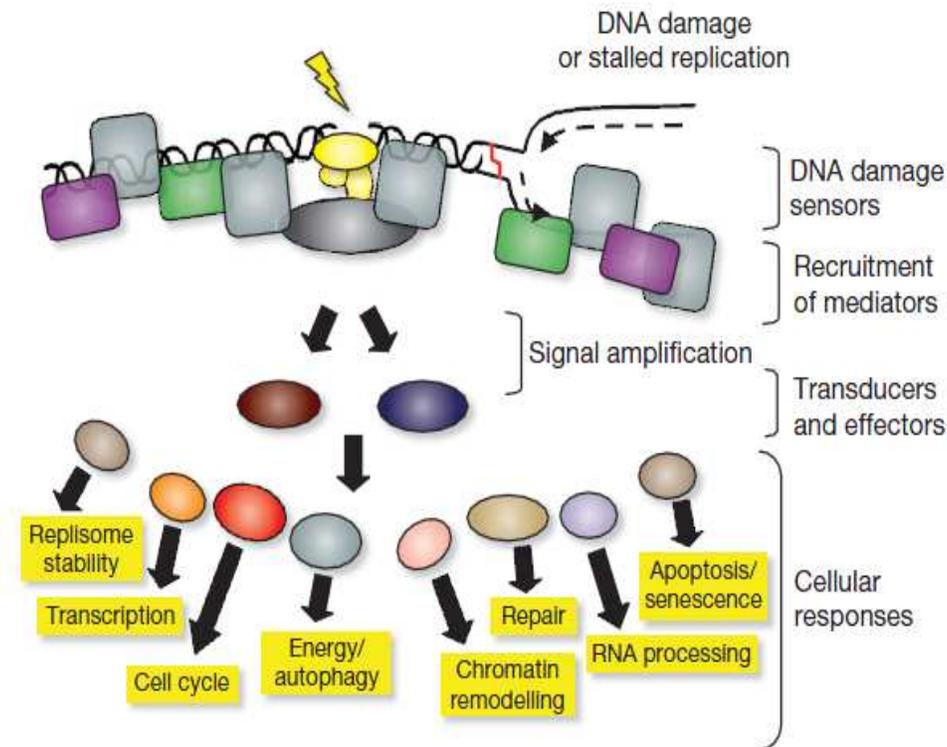
All experiments unless stated otherwise has been performed by the author. This thesis was written by the author.

## **CHAPTER 1**

### **INTRODUCTION: DNA DAMAGE and REPAIR**

#### **1.1 Sources of DNA Damage**

DNA harbours essential genetic information required for the foundation of life. However, DNA can be compromised through various forms of damage by a variety of DNA damaging agents. Protecting the integrity of DNA is vital for cellular function. Such damage must be sensed and subsequently repaired within the context of various stages of the cell cycle. The success of a given repair event is in part dependant on the nature and severity of the type of DNA damage. Eukaryotes have evolved a number of mechanisms important for successful repair including sensing of damage, activation of cell cycle checkpoints, arrest of cell cycle progression as well as authorizing the correct DNA repair pathway for a given type of damage (Figure 1.1). Nevertheless, if the DNA has been damaged severely, apoptosis may occur. Apoptosis is an important option which permits cells to circumnavigate the possibility of serious pathological consequences should a cell with damaged DNA continue to divide and multiply with altered genetic content.



**Figure 1.1:** The presence of a lesion in DNA is first detected by sensor proteins with the downstream effects of for example, replication stalling, repair or apoptosis. These sensors proteins are responsible for recruiting mediators that results in a signal amplification through transducers and effectors that ultimately leads to a cellular response (modified and adapted from West et al., 2009).

DNA can be damaged by exposure to endogenously generated agents such as by spontaneous hydrolysis of DNA glycosyl bonds at rates of  $10^4$  events per day per cell (Dinner et al., 2001). Spontaneous deamination occurs via the primary amine group on cytosine. The cytosine is then converted into uracil through a hydrolytic deamination reaction. There can also be a natural methylation of cytosines at the 5'- carbon position, which results in 5-methylcytosine that is then susceptible to spontaneous deamination. The end result is conversion of cytosine to uracil and of 5-methylcytosine to thymine. 5-methylcytosine is prone to the

formation of a G-C to A-T transition. This type of point mutation accounts for one third of single site mutations found in hereditary diseases.

Reactive Oxygen Species (ROS) are generated endogenously in mitochondria through the action of cytochrome P450 enzymes as well as in macrophages/inflammatory cells. Exogenous sources of ROS have been attributed to compounds such as lindane, TCDD, phenobarbital and ionizing radiation to mention just a few. Enzymes that counteract the activity of ROS include superoxide dismutase, catalase and glutathione peroxidase. Oxidation of bases may result in DNA damage and creation of thymine glycols, formamidopyrimidines and 8-hydroxyguanines. Oxidation of guanine residues to 8-hydroxyguanine accounts for major mutagenic base lesion generated by hydroxyl radicals, which then base-pair preferentially with adenine rather than cytosine. This results in a transversion mutation following replication.

Alkylating compounds can be found endogenously within the cell as well as in the environment such as in antineoplastic agents. O-alkylations (O6-alkylG and O4-alkylT) are highly mutagenic and genotoxic. N-alkylations (7-alkylG, 3-alkylA and 1-alkylA) are genotoxic, but are less mutagenic with the most common site for base alkylation being 7-alkylG. Alkylation of DNA bases may block DNA replication. Primary alkylation products such as methylations and chloroethylations are repaired by direct damage reversal or base excision repair. Secondary alkylation products that result in abasic sites, strand breaks and interstrand cross links are repaired through nucleotide excision repair, mismatch repair, base excision repair and homologous recombination (see section on DNA repair mechanisms for further details). If left unrepaired, these kinds of lesions can result in SSBs. Every cell that is exposed to 1 Gray of ionizing radiation will incur about 1500 damaged bases, 900 damaged sugars, 1000 single strand breaks, 40 double strand brakes and 150 DNA-protein cross links (Kao et al., 2005). DNA

double strand breaks (DSBs) are especially detrimental for the cell and it can be generated both internally and externally of the cell such as through radiomimetic agents.

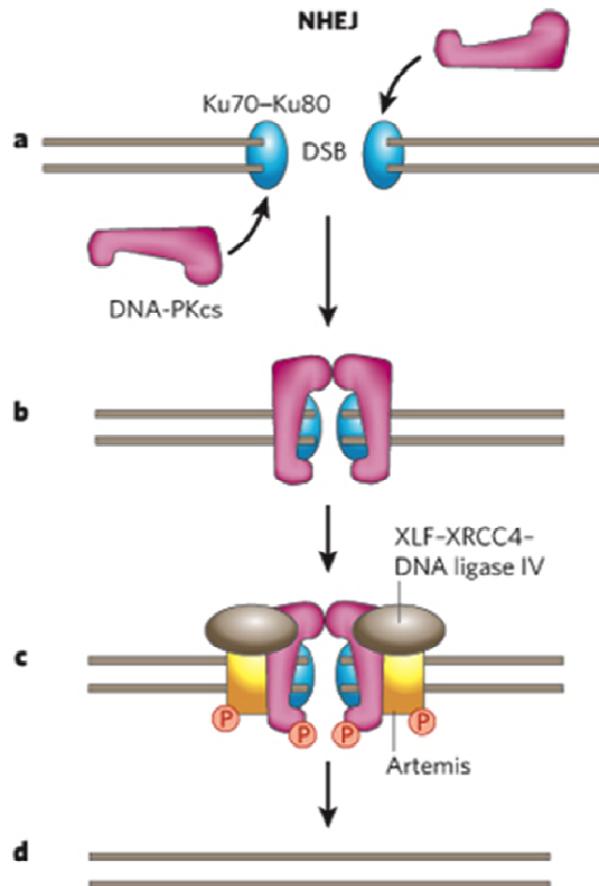
Interestingly, double strand breaks can occur naturally within the cell through programmed events including yeast mating-type switching, immunoglobulin class switching recombination and DNA rearrangement during V(D)J recombination. Immunoglobulin class switching recombination results in different isotypes of immunoglobulin where as V(D)J recombination facilitates the production of a large repertoire of antigen receptors with varying specificities. Both of these processes are required for the essential functioning of the cell. Double strand breaks can also be induced following exposure to ionizing radiation or chemicals such as those that cause alkylation of phosphates or oxidative damage to sugars. Other major causes of DNA double stranded breaks have been attributed to the stalling/collapse of the replication fork and mechanical shearing of daughter chromosomes during segregation.

It is therefore apparent that DNA DSBs can be routinely generated through numerous routes. If left un-repaired or if repaired incorrectly, DSBs can cause cell death, chromosomal fragmentation, chromosomal rearrangements and chromosomal instability that can lead to activation of oncogenes and inactivation of tumour suppressor genes that may in turn result in tumour formation. As such it becomes of great interest and medical importance to not only understand the formation of such damage, but also the mechanisms by which cells repair such damage. The following sections provide a closer description of several important DNA repair pathways that all make use of a processing enzyme called Aprataxin (Aptx).

## **1.2 Mechanisms of DNA Double Strand Break Repair: Homologous Recombination and Non Homologous End Joining**

DSBs are repaired by two competing pathways, homologous recombination (HR) and non-homologous end joining (NHEJ). HR involves the exchange of genetic information via recombination between homologous sequences and is considered the most error free mechanism for repair of DSBs occurring during the presence of a homologous chromosome. Through the utilization of diploidy during S and G2 phases of the cell, HR is able to resolve DSBs. HR is thus a form of homology directed repair that conserves genetic material and is the major form of DSB repair in lower organisms and at those times in the cell cycle when two copies of a chromosome are present. A minor form of repair is non-conservative in its method and involves annealing of a single stranded tail at a nearby direct repeat sequence (Lin et al., 1990).

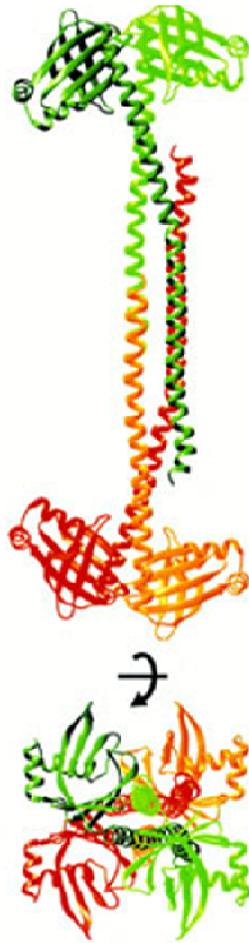
NHEJ simply involves the rejoining of two ends of DNA in much the same way that T4 ligase is used to join DNA fragments in a standard plasmid construction. NHEJ repairs DSBs in an end-to-end fashion and does not require any sequence homology (Figure 1.2). NHEJ is conserved evolutionarily from bacteria to mammals but is predominant in higher organisms such as mammals due to the complexity of the genome, which hinders easy access to homologous chromosomes. As such, there is large preference for NHEJ over HR in higher mammals. Interestingly, NHEJ is imprecise and may lead to the loss of several nucleotides. It has been suggested that this small alteration of the genome may provide an avenue for genetic evolution. Unlike HR, NHEJ occurs throughout the cell cycle during G0, G1 and early S phase.



**Figure 1.2: NHEJ. a, A DSB is recognized by the Ku dimer (Ku70–Ku80) and DNA-PKcs. b, The two DNA ends are synapsed. c, DNA-PKcs and Artemis are phosphorylated, and the DNA ends are processed by a complex consisting of XLF (Xrcc4-like complex; also known as NHEJ1), Xrcc4 (X-ray-repair cross-complementing protein 4) and DNA lig IV, and by Artemis. d, The DNA ends are ligated by DNA ligase IV, and the DNA-repair factors dissociate (modified and adapted from Downs et al., 2007).**

Recent studies have shown that there may be two types of NHEJ; D-NHEJ and B-NHEJ. D-NHEJ represents a faster, more predominant form that is dependant on DNA-PK and Xrcc4/Lig IV. B-NHEJ appears to act independently of DNA-PK, is slower in nature and requires Lig III (Wang et al., 2005).

For efficient NHEJ to occur, a minimum of 6 key players are required; Ku70/Ku80, DNA-PKcs, Artemis, Xrcc4, DNA lig IV and XLF. Current research shows that the Ku70/Ku80 heterodimer is the first DNA repair protein that is involved in binding to and sensing broken DNA in order to protect the DNA breaks from being degraded by nucleolytic activity. Ku binding is also suggested to aid in the formation of a bridge between broken DNA ends (Bliss et al., 1992, Carry et al., 1997, Ramsden et al., 1998). The DNA binding affinity of Ku is not sequence specific (2.4 nM) (Blier et al., 1993) and allows DNA to weave through an opening in a ring like structure that is formed by both subunits (Walker et al., 2001). Ku together with DNA-PKcs form the DNA-PK complex, which is thought to be responsible for aligning and bridging DNA ends (Spagnolo et al., 2006). Following DNA damage, DSBs often have unprocessed termini such as a 3' phosphate or 5' hydroxyl that requires the DNA ends to be processed prior to ligation. End processing enzymes such as Artemis which contains both endonuclease and exonuclease activity (Ma et al., 2002) become phosphorylated by DNA-PKcs, and can subsequently process a variety of DNA ends including gaps, flaps and hairpins (Ma et al., 2005) such as those generated during V(D)J recombination. Once the ends are processed, ligation can occur. Ligation requires the activity of ATP-dependant DNA Ligase IV in complex with Xrcc4 (Junop et al., 2000; Modesti et al., 2003). Xrcc4 may be required to help stabilize and align DNA ends for Lig IV activity (Grawunder et al., 1997; Modesti et al., 1999). Consistent with this idea, Lig IV has weak DNA binding activity where as Xrcc4 can bind nicked or linear DNA in a co-operative manner and with high affinity. Analysis of the Xrcc4 crystal structure revealed that it exists in an unusually elongated molecule in a dimeric conformation with each monomer composed of an N-terminal "head" domain consisting of a  $\beta$ -sandwich subsequently followed by an extended C-terminal  $\alpha$ -helical tail (Figure 1.3).



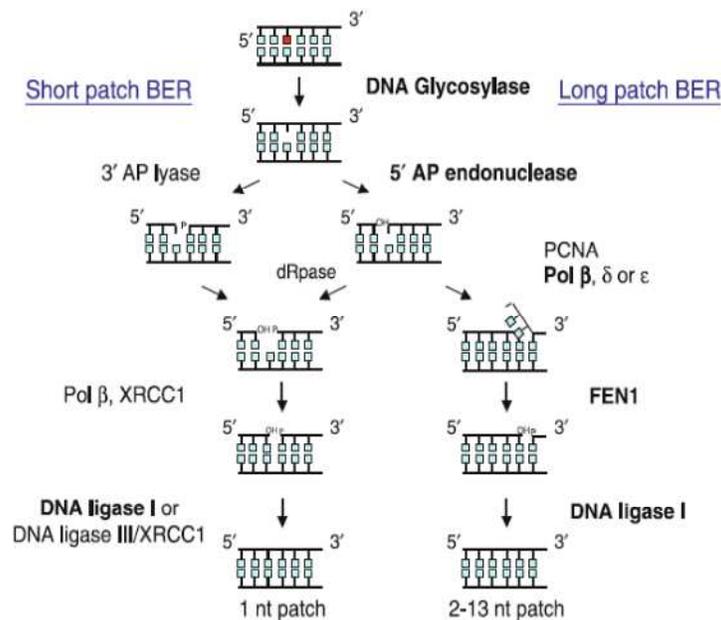
**Figure 1.3: Two orthogonal views of Xrcc4 tetramer. The crystallographic dyad axis is marked. The two dimers are coloured in green and red: a darker color for the L subunit and a lighter color for the S subunit (modified and adapted from Junop et al., 2000).**

Although not absolutely required for NHEJ, several other proteins play an important supportive role depending on the nature of the DNA damage. PNK (polynucleotide Kinase) represents such a protein and consists of both 5' DNA kinase and 3' DNA phosphatase activities (Karimi-Busheri et al., 1999). Fen-1 is another processing enzyme implicated in cleaving 5' flaps (Wu et al., 1999), while DNA polymerase  $\mu$  and  $\lambda$  has been shown to have gap filling activity during rejoining of two non-complementary DNA ends (Daley et al., 2005). Terminal

deoxynucleotidyl transferase, or TDT is involved in the addition of non-templated nucleotides to DNA ends and is responsible for increasing the variety of junctions during V(D)J recombination. Other factors such as WRN and BLM are helicases thought to support the unwinding and processing of DNA ends (Chen et al., 2003); Lengland et al., 2002). Finally, Aprataxin (Aptx) has been implicated in both HR and NHEJ following abortive ligation and is required for removal of the adenylate group from the 5'phosphate position.

### **1.3 DNA Base Excision Repair**

Base excision repair (BER) is a pathway of DNA repair that targets hydrolytic depurination, hydrolytic deamination of cytosine/5-methylcytosine bases, formation of covalent adducts with DNA and oxidative damage to bases (Figure 1.4). DNA glycosylases serve an important function in BER by recognizing abnormal bases and catalyzing the hydrolytic cleavage of the C1'-N-glycosyl bond that links the base to sugar. This release of the base from the DNA generates an apurinic/aprimidinic (AP) site. Several classes of DNA glycosylases recognize and initiate repair of distinct base lesions such as those targeted towards repair of 'inappropriate' bases or repair of oxidized bases. Simple glycosylases, catalyze only the hydrolytic removal of the base from an AP site where as others cleave off the base by a lyase mechanism and catalyze a subsequent AP lyase reaction. In general, lyase reactions are associated with glycosylases that remove oxidized bases, but not with those that remove normal or alkylated bases. Like repair of double strand breaks, the final step in repair involves ligation. Consequently, APTX activity is frequently required in this repair pathway following abortive ligation.

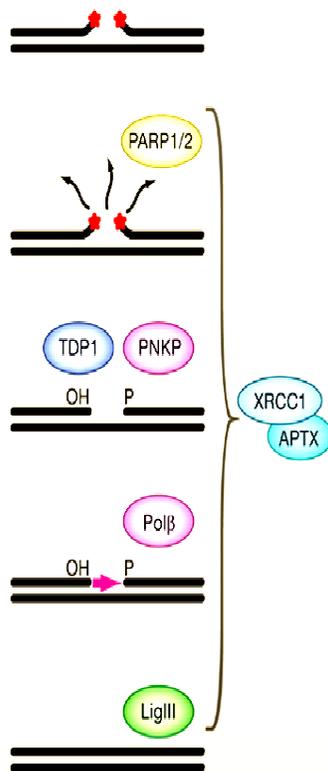


**Figure 1.4: Aptx acts in the short patch of the BER pathway of SSB repair. The first step of BER requires DNA glycosylase which recognizes damaged DNA and hydrolytically cleaves the glycosidic bond on the apurinic (AP) site of the SSB. In short patch BER, only the base is removed where as, during long patch BER AP lyase cleaves the base and a piece of the sugar backbone. Following this step of short patch BER, AP endonuclease nicks 5' to the AP site which results in a 5' phosphate and 3' OH. Phosphodiesterase removes fragmented sugars and finally, DNA polymerase and ligase fills and seals the gap. Aptx is involved in the short patch of BER where it cleaves adenyl groups from DNA. It also forms a complex with Xrcc1 and DNA ligase III which joins ends of DNA (Figure adapted and modified from McPherson, 2008).**

#### 1.4 DNA Single Stranded Break Repair

DNA single strand breaks (SSBs) are primarily detected by the enzyme, poly (ADP-ribose) polymerase1 (PARP1). At the time PARP1 encounters a DNA strand break, it activates itself and other proteins by poly ADP-ribose modifications (Figure 1.5). However, this modification is only transient such that PARP becomes de-ribosylated in order for subsequent rounds of sensing and repair to occur. Most, if not all SSB sites contain damaged 5' and/or 3' end termini. These damaged ends must be restored to traditional 3' hydroxyl and 5' phosphate

ends for ligation and gap filling to occur by processing enzymes. The DNA end processing stage is the step of DNA repair that involves the use of enzymatically diverse proteins. This simply reflects a need to process many different types of SSB termini generating in response to various forms of damaging events. Termini at break sites range in their cytotoxic effects on the cell. For example, one of the most cytotoxic breaks is the 5' deoxyribose phosphate and its removal is dependant on the enzyme, DNA polymerase  $\beta$ . A common SSB site includes a 3' phosphate and 5' hydroxyl end where the processing of this type of termini is dependant on the catalytic activity of PNKP. Another type of termini is the 3'-phosphoglycolate end, which is repaired by the DNA repair protein, APE1. Once damaged termini are restored to their conventional 5' and 3' ends, gap filling and sealing of the SSB may occur. This is primarily accomplished by DNA polymerase  $\beta$  and ligase III, but other proteins may also play a role such as Apx which processes DNA ends following abortive ligation.

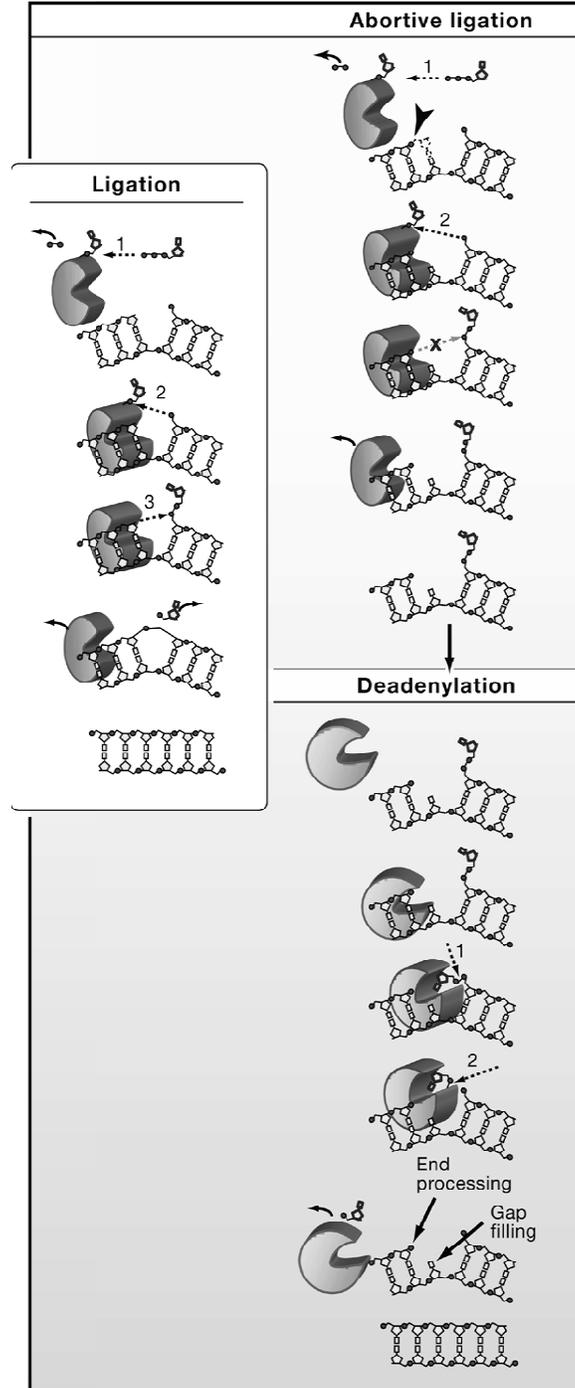


**Figure 1.5: DNA single strand breaks resulting from sugar damage or enzymatic cleavage are detected by PARP1 and PARP2. Binding of the break activates PARP for poly (ADP-ribosylation), thereby signalling to other DNA-repair proteins. PARP recruits Xrcc1, setting up a molecular scaffold for SSB repair, which is coordinated through interactions of individual repair enzymes. DNA end-processing factors such as PNKP and TDP1 restore conventional 5'-phosphate and 3'-hydroxyl termini at the break site. Most gapped single-strand breaks are then reduced to a nick by the action of DNA polymerase  $\beta$ , before being sealed by DNA ligase III. Aprataxin interacts with Xrcc1 and may participate in DNA end processing of abortive ligation events. Red asterisks indicate unconventional DNA termini (modified and adapted from West et al., 2007).**

## 1.5 Abortive Ligation

The final stage in repair of many types of DNA damage involves resealing of one or more DNA strands. In mammals, this is accomplished by ATP-dependent DNA ligase (I, III or IV). These DNA ligases join the ends of DNA in three general steps (Figure 1.6): (1) Formation of a covalent enzyme-AMP intermediate linked to a lysine side-chain in the enzyme, (2) Transfer of the AMP nucleotide to the 5' phosphate of the nicked DNA strand and (3) Attack on the AMP-

DNA bond by the 3'-OH of the nicked DNA, resulting in the sealing of the phosphate backbone and subsequently releasing AMP. If the ligation reaction is initiated before the attacking the 3'OH strand is fully processed, DNA ligase is able to catalyze steps 1 and 2 of ligation but is unable to complete the reaction resulting in adenylated DNA. Ligase enzymes play an important role in multiple DNA repair pathways such as DSB repair including HR and NHEJ, BER and SSB repair. Ahel et al.,(2006) suggested that the accumulation of damaged DNA due to mutated Aptx, and the inability of Aptx to act in DNA repair may be the underlying cause of ataxia with oculomotor apraxia.



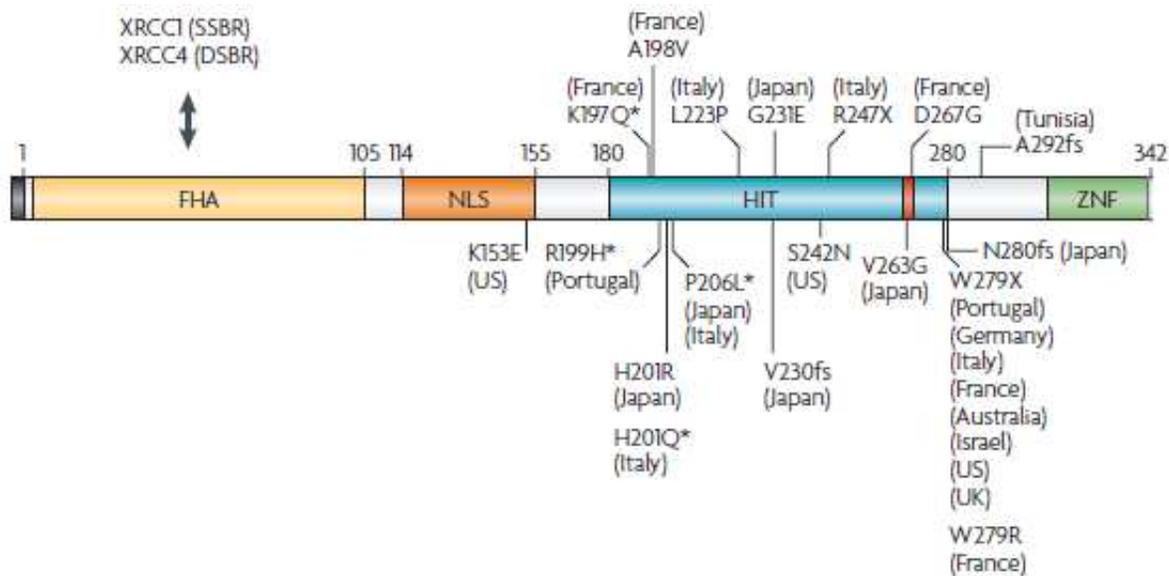
**Figure 1.6: Resolution of Abortive DNA Ligation Intermediates by Aprataxin (Inset) ATP-dependent DNA ligases catalyze ligation through three distinct reaction steps: (1) cleavage of ATP into AMP and pyrophosphate and transfer of the AMP moiety to its active site lysine; (2) attack of the 5' phosphate at the break site on the activated phosphoryl group of the AMP at the active site of the ligase; (3) attack of the 3' hydroxyl of the SSB on the**

**adenylated 5' phosphate, which restores the continuity of the DNA backbone and releases AMP. (Top panel) The action of DNA ligase at a gapped SSB with unconventional DNA termini can lead to abortive ligation. A typical oxidative SSB contains phosphates at both the 5' and the 3' end (indicated by an arrowhead) and a one-nucleotide gap (indicated by the dotted outline of a missing nucleotide). At such a dirty break the DNA ligase can catalyze steps 1 and 2 of the ligation reaction but then aborts, leading to the formation of a 5'-DNA adenylate. (Lower panel) Aprataxin resolves 5' adenylates by releasing AMP and restoring the 5' terminus of the SSB. This proofreading DNA repair function probably proceeds through two catalytic steps: (1) attack of the adenylated 5' phosphate leading to release of the deadenylated DNA and transfer of the AMP onto Aprataxin's active site histidine triad, and (2) hydrolysis of the 5' phosphohistidine to release AMP and restore Aprataxin. Subsequently, the SSB can be processed and ligated (modified and adapted from West et al., 2007).**

## 1.6 Aprataxin

Damaged DNA has been implicated in an array of hereditary diseases, many of which are characterized by neurological disorders and/or amplified genetic instability and cancer. The cumulative evidence suggests that diseases may arise from a single mutation in a DNA repair factor (Ahel et al., 2006). AOA1 is neurodegenerative disorders caused by defects in Aptx. Recent findings demonstrate that neuronal cell death occurs in AOA1 (Moreira et al., 2001) due to defects in Aptx leading to an accumulation of SSB.

Repair is a dynamic process involving choreography of numerous proteins and not surprisingly it does not always transpire as planned. DNA repair proteins may inadvertently further the extent of damage, when encountering an unprocessed termini. A classic example of this scenario occurs when DNA ligase become involved in abortive activity that can result in a 5'-AMP being abandoned at a single or double strand break site. Aptx removes these abortive ligation products in order for repair to be successfully reinitiated. However, if left unrepaired, this type of break can accumulate and lead to the neurodegenerative disorder, AOA1. AOA1 is characterized by progressive neurodegeneration, problematic motor coordination, loss of reflexes, hypoalbuminemia and hypercholesterolemia.



**Figure 1.7: Aptx and AOA1 mutation.** The Aptx protein consists of three domains; The Forkhead domain (FHA), the histidine triad domain (HIT), and the Zinc-finger domain (ZNF). Aptx also contains a nuclear localization signal (NLS). The FHA domain binds to DNA repair factors Xrcc1 during SSB repair and Xrcc4 during double-strand break repair (DSBr). Mutations which render Aptx inactive are mainly localized to the HIT domain and mutations from populations around the world are shown. Asterisks indicate mutations which are associated with late onset or mild clinical syndromes of AOA1. Black box indicates 13 extra nucleotides that are present with the longest variant of Aptx. Red box represents the position of the catalytic triad (modified and adapted from West et al., 2007).

Aprataxin was first identified by two independent researchers using positional cloning. Date et al. (2001) and Moreira et al. (2001) identified a novel gene encoding a member of the histidine triad superfamily which was previously identified as FLJ20157, the causative gene mutated in AOA1. The predominant form of the nuclear protein Aptx is 39.1 kD, and is found in human cell lines from the brain, heart, lung, kidney liver, and trachea. Aptx has three domains (Figure 1.7): a zinc finger, a histidine triad (HIT) and a forkhead associated domain (FHA). Based on sequence homology, the zinc finger domain has been predicted to consist of two antiparallel beta strands and a single alpha helix. The zinc cation provides stability within the domain and is required for the domain to remain folded. It has been suggested that the zinc finger of Aptx may be important for binding to DNA, as this function is often associated with zinc

finger domains in DNA binding proteins. The HIT domain is highly conserved and has unique AMP-lysine hydrolase activity. Mutations in this region are found in patients with AOA1. The FHA domain of Apx closely resembles the FHA domain of the DNA repair protein PNKP. Both PNKP and Apx share mutual binding partners, Xrcc1 and Xrcc4, and interact with these proteins in a phosphoserine-dependent manner. The N terminus of the predominant form of Apx interacts with the C terminal regions of both Xrcc1 and Xrcc4. During SSB repair, Apx binds to Xrcc1 in complex with DNA polymerase beta and DNA ligase III and Apx. During DSB repair, Apx has been shown to bind Xrcc4 within the Xrcc4-LigIV complex.

The involvement for Apx in SSB repair was first demonstrated by Ahel et al.,(2006). This group purified aprataxin protein and extracts from defective Apx DT40 cells or Apx <sup>-/-</sup> mouse primary neural cells to show that Apx resolves abortive DNA ligation intermediates. In particular it was found that Apx catalyzes the nucleophilic release of adenylate groups covalently linked to 5' phosphate termini at single- strand nicks and gaps. This results in 5' phosphate termini, which are able to undergo ligation. Ahel et al.,(2006) also suggested that neurological disorders that arise from mutations in aprataxin are caused by the gradual accumulation of unrepaired DNA that results from abortive ligation events. Mouse primary vertebrate extracts were analyzed for their activity of Apx when treated with SSB inducing agents. Repair was able to occur in these cells as seen by the reversal of AMP-DNA to DNA. However, under these conditions cells from AOA1 patients retained 5' adenylates. This indicates that the causative agent of AOA1 is the mutated Apx protein. Nevertheless, the question remains how exactly does accumulation of SSBs with 5' adenylates lead to AOA1? The authors proposed an interesting possibility in which neuronal cells are exquisitely sensitive to oxygen radicals and due to the fact that they are non-proliferative in nature, they are unable to utilize alternate repair

pathways such as homologous recombination. Hence, neuronal cells were proposed to be especially susceptible to malfunctions in SSB repair. Work by Ahel and colleagues, greatly advanced the understanding of the mechanistic connection between hereditary neurological disorder and DNA repair.

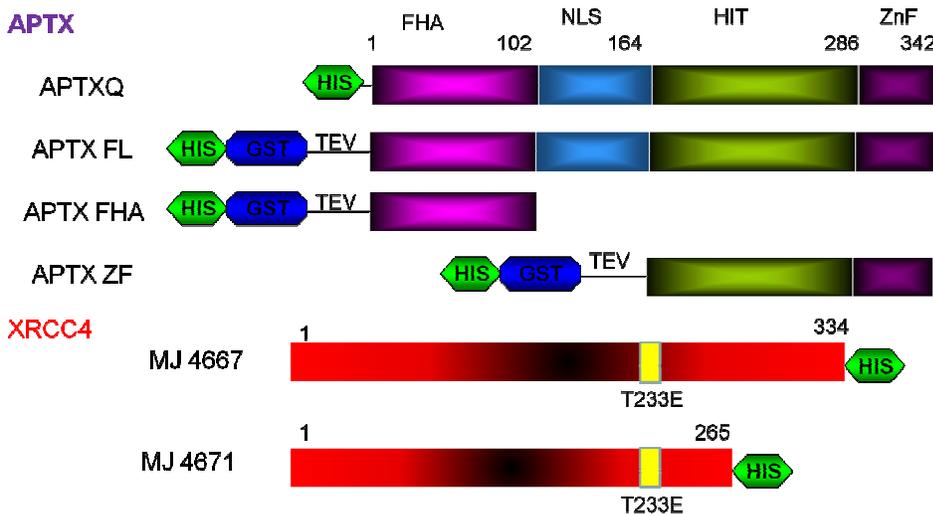
## **1.7 Thesis Objective**

Aprataxin is the causative gene mutated in ataxia with oculomotor apraxia (AOA1). Mutated Aptx causes a defect in several repair pathways: DSB repair, NHEJ, BER and SSB repair. It is now clear that Aptx is responsible for removal of abortive ligation products in order for repair of DNA to occur. However, its enzymatic and mechanistic role in DNA repair is not yet fully understood. How Aptx interacts with DNA or other repair factors has yet to be established. A crystal structure of Aptx will provide an in depth look into the mechanism governing Aptx. To achieve high-resolution structural information, milligram quantities of highly purified Aptx must be available for structural studies. Thus far, an expression system has not yet been described that can meet these requirements. A major objective therefore will be generating sufficient material of high quality. Development of assays, which test for the activity of Aptx will allow us to not only determine if exogenously expressed Aptx is active but also afford the opportunity to further explore the mechanism of Aptx in DNA repair. Using this knowledge, we can start to understand the role of Aptx in DSB<sub>r</sub> and SSB<sub>r</sub>. Studies of Aptx and more generally several DNA repair pathways will remain incomplete without a crystal structure to bridge structural knowledge with functional understanding. In addition to providing a mechanistic basis of action, a crystal structure of human Aptx may also lead to exploitation of Aptx as a novel target for chemotherapeutic development, since Aptx has been found to be upregulated in cancer cell lines (Dopeso et al., 2010).

## CHAPTER 2

### MATERIALS AND METHODS

In order to obtain a comprehensive understanding of the functional and structural properties of Aptx, a crystal structure of Aptx is required. To improve the likelihood of obtaining sufficient quantities of stable protein to achieve this goal, a number of approaches were taken. First, there was extensive design of domain boundaries for cloning, overexpression and purification of various forms of protein. Since Aptx interacts with Xrcc4, this design was extended to include both Aptx and Xrcc4. In addition to these considerations, the methods used to monitor enzymatic activity of Aptx and to assess a crystal structure are discussed below.



**Figure 2.1: Constructs of Aptx and Xrcc4.** Versions of Aptx include that of AptxQ; full length human protein with a non cleavable N-terminal HIS tag, Aptx FL; cleavable N-terminal HIS and GST tag of the full length human protein, Aptx FHA; cleavable N-terminal HIS and GST tag of the FHA domain of human Aptx and Aptx ZF; cleavable N-terminal HIS and GST tag of the HIT-ZF domain of human Aptx protein. MJ4667 is the

**full length human Xrcc4 protein with a C-terminal non cleavable HIS tag containing a single site mutation of T233E. MJ4671 comprises of amino acid residues 1-265 with a C-terminal HIS tag and containing a single site mutation of T233E.**

## **2.1 Cloning, Overexpression and Purification of Aptx and Xrcc4**

Human homologues of Aptx and Xrcc4 were chosen for structural studies in order to provide maximal information on repair in the context of human disease. Prior to this work Aptx was expressed in small quantities from either human or insect cells. To generate sufficient quantities of Aptx for structural studies, expression was carried out in bacterial cells. In general, Aptx protein was difficult to work with due to proteolytic degradation within the disordered region joining FHA and HIT domains. Nevertheless, with considerable optimization it was possible to obtain very pure Aptx protein, which was of sufficient quality to conduct crystallization trials. Xrcc4 was used in functional binding studies as well as a positive control for the Aptx activity assay.

### **2.1.1 Cloning**

In order to assess the functional and structural properties of Aptx, several constructs were designed for both Aptx and its binding partner, Xrcc4. Proteins containing disordered regions frequently do not lend themselves readily to formation of an ordered lattice required for crystal formation. As such, it is desirable to attempt structural studies by exploring multiple version of a protein that encompasses different domain boundaries. Domain boundaries were determined based on sequence conservation and secondary structure prediction algorithms. In addition to exploring domain boundaries, various fusions tags are often assessed for their ability to promote expression, solubility and efficiency in purification. Finally, to improve expression in a heterologous host it is often advantageous to codon optimize an open reading frame to match the expression host. Aptx-Q was commercially codon optimized and prepared by whole gene

synthesis. Expression constructs prepared for Aptx include: codon optimized full length human protein with a non-cleavable N-terminal HIS tag, Aptx-Q; full length human protein with a non-cleavable N-terminal HIS tag, Aptx FL; cleavable N-terminal HIS and GST tag of the full length human protein, Aptx FHA; cleavable N-terminal HIS and GST tag of the FHA domain of human Aptx and Aptx ZF; cleavable N-terminal HIS and GST tag of the HIT-ZF domain of Aptx protein (Figure 2.1). Various constructs exist for Xrcc4 but two will be discussed here. Plasmid MJ4667 has an expression cassette for the open reading frame corresponding to the full-length human Xrcc4 protein with a C-terminal non cleavable HIS tag containing a single amino acid substitution of T233E. Plasmid MJ4671 is identical to MJ4667 except the open reading frame is mutated to generate a c-terminal truncation (residues 1-265). Several of these constructs, including AptxQ, Aptx FL, Aptx FH and MJ4667 have been overexpressed and purified.

### **2.1.2 Overexpression and Purification**

Overexpression and purification of several Aptx and Xrcc4 constructs, as listed in Figure 2.1, were completed in order to conduct functional and structural studies of APTX. The details of the protocols used for purification of these Aptx protein constructs are listed below

#### **2.1.2.1 AptxQ**

The full-length version of human Aptx open reading frame was PCR amplified for Gateway Cloning. In order to express Aptx, Rosetta (DE3) cells were grown to an  $OD_{600}$  of 0.5 and induced with 1 mM IPTG for 6 hours at 30°. During purification, cells were re-suspended in Ni A buffer (20mM Tris pH 8, 1.5 M KCL, 0.03% LDAO, 1 mM BME, 10% glycerol, 10 mM Imidazole). Cells were lysed in the presence of protease inhibitors (1 mM each of leupeptine,

benzamidine, aprotinin) using 4 successive passages through the French Press at 10,000 psi. The lysate was clarified by centrifugation at 10,000 x g for 40 min to separate soluble protein from insoluble cellular components. The supernatant was subsequently filtered to remove any small particulates. Cell lysate was applied to a nickel-NTA column at a flow rate of 1 ml/min. Equilibration and subsequent washes were performed at 24.5 mM imidazole (5% Ni B) and 39 mM of imidazole (10% Ni B, where Ni B contains the same components as Ni A buffer + 300 mM imidazole). Fractions from a 10% to 100% gradient over 40 mL of NiB buffer were collected. Fractions containing Aptx protein were then diluted to 350 mM salt in order to bind an ion-exchange SP-sepharose column further purify the protein. A dual buffer system, one with low salt (Sa- contained 20 mM MES pH 6.5, and 10% glycerol) and another with high salt (Sb- 500 mM KCl, 20 mM MES pH 6.5, and 10% glycerol), was used. The ion-exchange column was pre-equilibrated at 200 mM KCl and washed between 200 mM and 300 mM KCl. Protein was eluted off a gradient between 300 mM and 500 mM KCl over 150 ml. Two peaks were collected off the SP-sepharose column and analyzed by SDS-PAGE. Each peak was then separately injected onto a buffer exchange column. Although the protein was relatively pure at this point, a buffer exchange column allowed us to obtain more suitable buffer conditions for crystallization trials. Aptx protein from both peaks was separately exchanged into 275 mM KCl, 20 mM MES pH 6.5, 10% glycerol and 1 mM DTT. Protein was concentrated to final concentrations varying between 6 mg/ml and 20 mg/ml.

#### **2.1.2.2 Aptx FL**

Full-length Aptx contained of a TEV protease-cleavable N-terminal HIS/GST tag. The full-length construct was cloned into a gateway destination vector and expressed in BL21 cells

for 6 hours at 30 ° C following induction with IPTG at an optical density of 0.5. Aptx FL was first purified using a nick-NTA column with buffers that were used to purify AptxQ, except that the NiB buffer contained 450 mM imidazole, instead of 300 mM imidazole. A total of 252 mg of protein was obtained following the nickel column and frozen in 1 ml aliquots. Ten mg of Ni column protein was injected onto a pre-equilibrated S column following dilution to 100 mM salt. Buffers for ion-exchange were as described for AptxQ. A gradient to 2M KCl over 30 ml was carried out. This resulted in the ability to obtain 8 mg of full-length protein but contaminated with some break down product of Aptx FL.

### **2.1.2.3 Aptx FHA**

The coding region for the Aptx forkhead domain, also known as Aptx FHA, was cloned into a gateway destination vector and expressed in BL21 cells for 6 h at 30 °C. Aptx FHA was purified using a nickel-NTA column and further purified using an SP-sepharose column as described for Aptx FL. Following Ni column purification, 240 mg of Aptx FHA was recovered and frozen in 1 ml aliquots for long term storage at -80 ° C. 40 mg of protein from the nickel column purification was injected onto an S ion-exchange column with pre-equilibrated 100 mM KCl. 15 mg of pure protein was recovered following ion-exchange chromatography.

### **2.1.2.4 Xrcc4**

Plasmid pMJ4667 encodes the full length open reading frame for human Xrcc4 with a single amino acid substitution (T233E) that mimics the phosphorylation required for Aptx association. This construct contained a C-terminal HIS tag and was purified again using a nickel-NTA column. Details of the buffer condition are listed in the supplementary material, buffer

conditions. Further purification was not required as this single column generated protein of greater than 95% purity.

## 2.2 Crystallization of Aptx

Crystallization trials were performed using the hanging drop vapour diffusion method with varying concentrations of ammonium sulphate placed in wells as dehydrant. Precipitants included crystal magic I/II, Nextel kits, and wizard screens. In total, ~3000 different crystallization conditions were tested for each protein construct. Crystallization trials of AptxQ did not yield diffraction quality crystals and further trials need to be explored with different protein concentrations, temperatures, pHs and salt solutions. Further, crystallization trials were set using a combination of peptides of Xrcc4, purified Xrcc4 and Aptx protein constructs. A successful crystal from the *S. Pombe* homolog of Aptx in complex with DNA-AMP-Zn<sup>2+</sup> was recently reported by Ahel et al. (2011).

## 2.3 Activity of Aptx

In order to assess the activity of purified Aptx protein, a 5'-deadenylation assay was developed. Plasmid DNA was subjected to iron and hydrogen peroxide in order to induce single strand breaks. This was achieved by mixing 150 nM of plasmid DNA in a buffer containing 0.1 M NaCl, 100 mM NADH, 20 uM H<sub>2</sub>O<sub>2</sub>, 0.2 mM FeCl<sub>3</sub> and 0.2 mM EDTA incubated for 1.5 hour at room temperature. The reaction was subsequently stopped by adding 5mM of EDTA. The DNA was treated with T4 ligase and 50 µCi [ $\alpha$ -<sup>32</sup>P]-ATP in ligation buffer overnight at 37° C. Because single strand breaks generated in this fashion do not typically have an intact 3'-OH

group available, abortive ligation occurs frequently under these conditions. Abortive ligation products at sites of breaks occurred as a result of the incorporation of  $^{32}\text{P}$ -labelled AMP at 5'-termini and reactions were ceased by the addition of 40 mM EDTA. Aptx protein was incubated with the abortive ligation product at room temperature and deadenylation activity monitored by removal of  $^{32}\text{P}$  signal from plasmid. Full deadenylation activity, therefore results in loss of radioactively labeled adenly groups from plasmid substrate. Reactions were resolved on 1% agarose gels run in TBE buffer. Gels were imaged using imagequant.

## CHAPTER 3

### RESULTS

#### 3.1 Design of Protein Expression Constructs

Often, full-length proteins are unable to form crystals due to disordered regions which are subject to proteolytic degradation. The region between the FHA and HIT domain of Aptx was found to contain such a large disordered, proteolytically sensitive site. AptxQ corresponds to a non-cleavable N-terminal HIS tag fusion of the full-length human version of the Aptx protein. This construct proved to be a difficult protein to purify and crystallize. During purification, the protein degraded rapidly at the region spanning the FHA and HIT domains. Furthermore, crystallization of full-length protein was equally challenging, most likely due to the presence of a large unstructured region. Moreover, the non-cleavable N-terminal HIS tag may be responsible for difficulty in both crystallization and interaction with Xrcc4. Thus, additional expression constructs were generated (Figure 2.1) that permitted bypassing several of these obstacles.

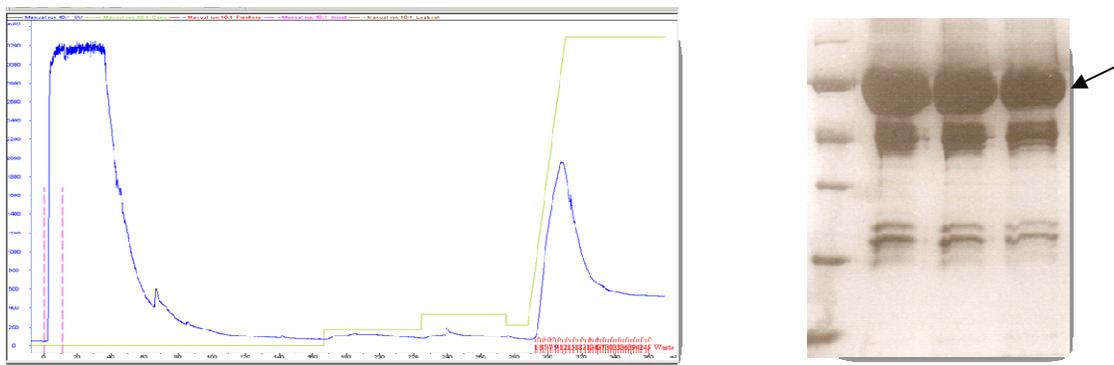
Given that Aptx has been reported to stably bind phosphorylated Xrcc4, it seemed reasonable to attempt structure determination with Aptx-Xrcc4 complex, as complex formation is often reported to improve stability. However, Xrcc4 protein is required to be phosphorylated for stable complex formation with Aptx. Using this rationale, we had two options; either phosphorylate Xrcc4 protein using CK2 or alter amino acid residues such that phosphorylation could be mimicked. CK2 was successfully used in experiments to phosphorylate Xrcc4 that resulted in stable association with Aptx; however, this method was very inefficient. I therefore turned to the use of a pseudophosphorylation mutant (T233E). By altering a threonine on position

233 to a glutamic acid it is often possible to mimic the effect of phosphorylation. In this way it would be possible to efficiently produce Xrcc4 that would form stable complex with Aptx. Since glutamic acid side chains can mimic the negative charge associated with a phosphor-threonine group, Xrcc4 constructs harbouring T-to-E substitutions can be termed pseudo-phosphorylated. Again, using the rationale that a full-length protein may be difficult to crystallize; synthetic peptides of Xrcc4 containing the pseudo-phosphorylation sites were also tested for complex formation and structural studies. However, binding studies of Xrcc4 did not show interaction with Aptx and therefore this approach was abandoned. The difficulty in obtaining diffraction quality crystals of full-length human protein is very likely related to the ‘floppy’ nature of the FHA domain being held to the HIT domain by a long disordered tethering span of amino acids. This notion is further supported by the fact that the recently reported crystal form of Aptx was taken from *S. Pombe* that lacked a FHA domain (Ahel et al., 2011).

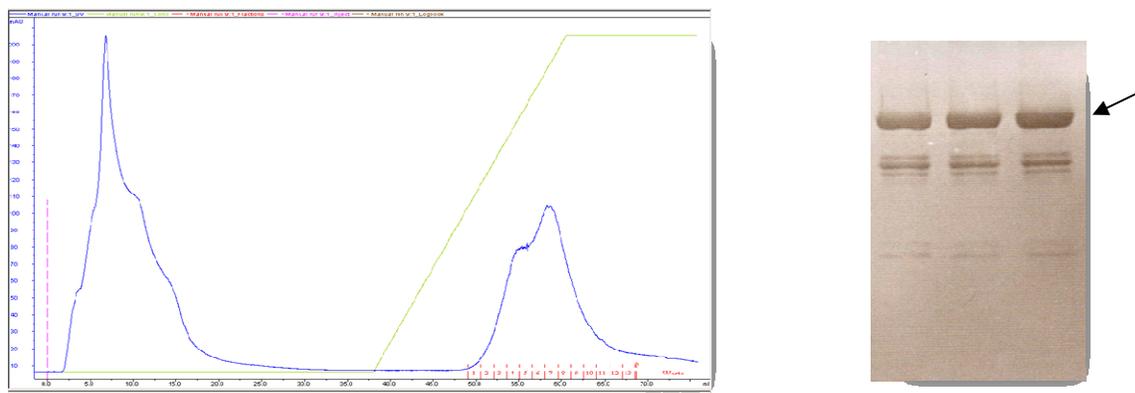
### **3.2 Optimization of APTXQ Protein Purification**

Purification of large quantities of Aptx is necessary in order to achieve a crystal structure. Often, full-length proteins are unable to undergo nucleation needed for crystallization due to the presence of disordered regions that do not easily pack into an ordered lattice needed for the formation and growth of a crystal. Thus, it is often beneficial to create truncations of the protein of interest to remove such regions. Based on this rationale several truncations were created as indicated in Figure 2.1. The first construct tested, AptxQ, overexpressed well in bacteria and was found to be very soluble. Fast protein liquid chromatography (FPLC) was employed in conjunction with affinity, ion exchange and buffer exchange columns to generate highly purified

protein (Figure 3.1). Unfortunately, Aptx protein degraded quickly and therefore proved to be a challenge for crystallization. Truncated versions of Aptx (FHA domain, Zn finger domain) with/without HIS-tag fusions as well as Aptx bound to other DNA/protein aided in achieving a more stable protein that did not degrade readily.



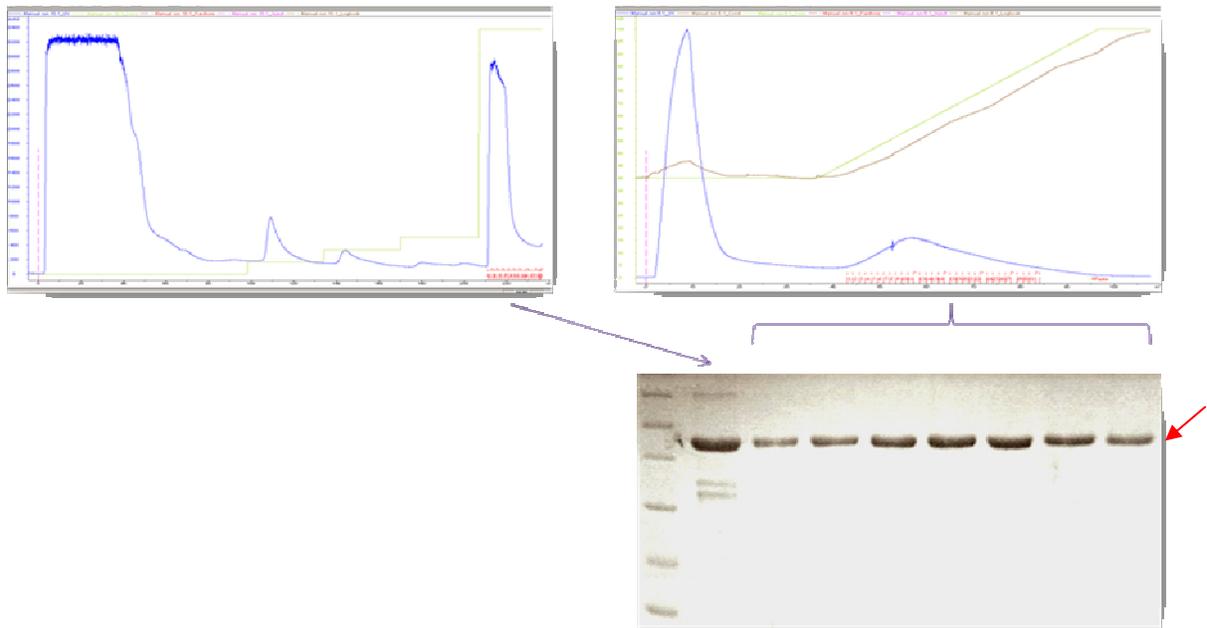
**Figure 3.1: Aptx FL nickel column chromatogram and corresponding SDS-PAGE gel of fractions 9, 12 and 15. Arrow indicates Aptx FL protein.**



**Figure 3.2: Aptx FL SP-sepharose column chromatogram and corresponding SDS-PAGE gels of fractions 5, 7 and 8. Arrow indicates Aptx FL protein.**

### **3.3 Optimization of Aptx FL Protein Purification.**

The Aptx FL construct contained a removable tandem HIS-GST tag. The protein was soluble and expressed well in BL21 cells. Fast protein liquid chromatography was again employed to purify the construct using a nickel-NTA column followed by an SP-sepharose column. Protein was injected onto the nickel column following lysis by the French press. Two washes were performed at 30 mM imidazole and 45 mM imidazole where very little to no protein eluted. Pure Aptx FL protein eluted from a steep gradient between 33 mM and 450 mM imidazole. A total of 252 mg of protein was obtained following the nickel column purification. This protein was ~ 50% pure as indicated by the SDS-PAGE presented in figure 3.1. Protein was stored in 1 ml aliquots at the -80 ° C until further purification was conducted. Further purification was carried out using an ion-exchange SP-sepharose column. Starting with 10 mg of protein, it was possible to obtain ~ 8 mg of Aptx FL protein which was relatively impure as indicated by the SDS-PAGE from Figure 3.2, and contained contaminants which were hypothesized to be break down products. Aptx FL was further used to assess protein-protein interaction and binding studies between itself and Xrcc4.

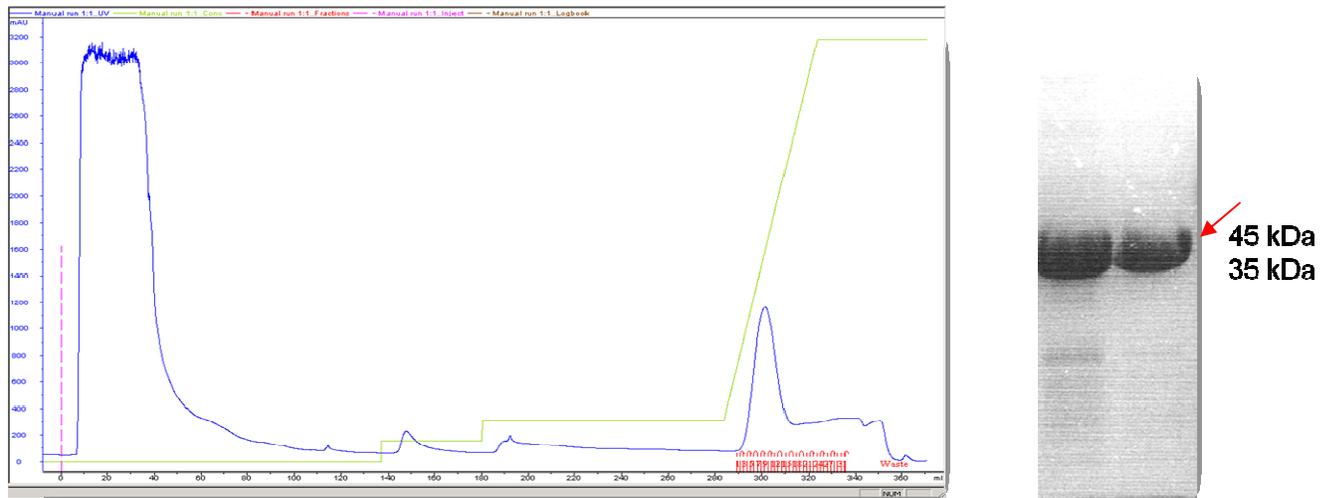


**Figure 3.3: Aptx FH purification.** Left chromatogram corresponds to the nickel column where as right is of the SP-sepharose column. Arrow represents nickel column protein injected onto SP-sepharose column, lanes 3-10 represent fractions from the SP-sepharose column. Red arrow represents location of Aptx FH protein.

### 3.4 Optimization of Aptx FH Protein Purification

Aptx FH was also able to be expressed at high levels and a soluble form in BL21 cells. FPLC purification was used to purify the protein first with a Ni-NTA column followed by an SP-sepharose column. Protein was injected onto the nickel column and was subjected to three washes at 30, 45 and 60 mM imidazole. There was a total of 240 mg of protein following the nickel column. Nevertheless, in order to remove the contaminants, 80 mg of the nickel column protein was further purified on an SP-sepharose ion-exchange column. This step yielded ~15 mg of > 95% pure Aptx FH protein (Figure 3.3). This represented a significant step forward in the

ability to obtain large amounts of high quality protein. Consequently, functional and structural studies were conducted using this construct.



**Figure 3.4: Full-length Xrcc4 nickel column purification showing chromatogram and corresponding SDS-PAGE gel. Red arrow represents location of Apx FH protein.**

### 3.5 Optimization of Human Xrcc4 Protein Purification

Xrcc4 protein expressed from pMJ4667 was soluble and obtained at high yields from BL21 cells. In order to purify the protein, a Ni-NTA column was used. Two washes were performed at 30 mM and 45 mM imidazole where little to no protein eluted. Xrcc4 protein eluted at 88 mM imidazole and continued to elute across a gradient to 300 mM imidazole. A total of 132 mg of protein eluted from the nickel column at about ~80% purity. Further purification was not required for the purpose of using this protein to conduct preliminary binding studies with Apx (Figure 3.4). Comparatively, Xrcc4 protein was more stable than Apx protein as there was

little to no break down over time. Xrcc4 protein was further used to assess protein-protein interactions with Aptx.

### **3.6 Xrcc4-Aptx Protein- Protein Interactions**

Aptx has been suggested to bind Xrcc4 in the region spanning amino acids 230-260. In order to investigate this possibility and perhaps make use of increased stability afforded by an Xrcc4-Aptx complex, we tested binding of Aptx to native Xrcc4, as well as pseudo-phosphorylated versions of Xrcc4, peptides of Xrcc4 with and without phosphorylation and finally CK2 phosphorylated Xrcc4. While binding was observed with variants that contained phosphothreonine modifications, the ability to fully phosphorylate Xrcc4 *in vitro* proved inefficient and therefore this avenue was not pursued further for preparing complex for structural studies. Since phosphorylated peptide was able to form complex (as judged by SPR analysis- see appendix, work conducted by Kritica Arora) it may still be possible to carry on this work with synthetic phosphorylated peptide fragments of Xrcc4.

### **3.7 Crystallization of Aptx**

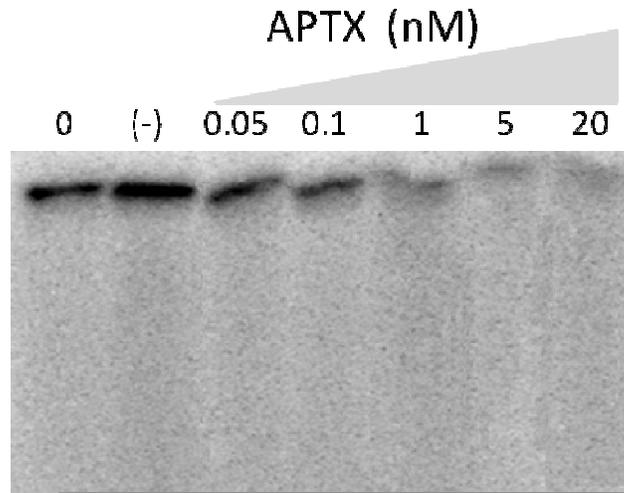
Crystallization trials were performed using the hanging drop vapour diffusion method at room temperature and 4 degrees Celsius with varying concentrations of ammonium sulphate dehydrant. In addition, 6 different protein (3-20 mg/ml) and buffering salt (300-1000 mM KCl) concentrations were screened. Precipitants included crystal magic I/II, Nextel kits, and wizard screens. In total over 3,000 crystallization conditions were performed for each protein construct, starting salt concentration and incubation temperature tested. Crystals were set in a methodical manner where each drop was assessed for possible crystal formation. If drops were clear the

concentration of the well solution and/or Aptx protein was increased in a systematic style. No conditions were identified which provided diffraction quality protein crystals. Crystallization trials will need to be continued by expanding the number of starting buffering conditions tested. Although the protein appeared homogeneous on a Coomassie-stained SDS-PAGE gel, it is still possible that it is heterogeneous in terms of quaternary structure. Prior to further crystallization trials an extensive screen of buffering conditions that promote homogeneity should be conducted using Dynamic Light Scattering to assess sample homogeneity. Once an optimal starting buffer has been determined further screening of crystallization conditions could be carried out, perhaps with better success. A structural analysis of a related Aptx protein from *S. Pombe* was recently reported by Ahel et al., providing new insight into the enzymatic mechanism of Aptx. Despite this structure having been reported, it will still be important to determine the structure of human Aptx to fully understand this protein in a more disease relevant context.

### **3.8    Aptx Activity Assay**

In order to assess activity of purified Aptx protein, it was necessary to develop an assay capable of monitoring 5'-deadenylation activity. A number of different approaches using short DNA oligos were explored. Although several of these had been reported in the literature, in general these methods suffered from inefficient labelling, low overall yield and high levels of background DNA oligo contamination. To circumvent many of these problems, a more simplified approach was taken to generate deadenylation substrate based on an assay similar to that reported by Steven West's group (Ahel et al., 2006). Plasmid DNA was used as a starting substrate. This plasmid was exposed to damaging conditions that generate random single strand breaks. Since these breaks do not normally retain an intact 3'-OH group, this substrate

undergoes abortive ligation when incubated with T4 ligase is *in vitro*. As such, an abortive ligation was simulated in which there was incorporation of <sup>32</sup>-P labeled AMP at 5' termini. Following substrate preparation by this method, three forms of plasmid DNA were obtained; unreacted cold super coiled, and radioactively labelled open circular and linear plasmid. Initially, we identified DNA damaging conditions that enabled us to generate primarily open circular DNA. As damage becomes more extensive, the ratio of linear to open circular to super coiled DNA increased. Purified full-length Aptx protein constructs was used to assess the ability to carry out deadenylation of radioactive AMP groups incorporated with the plasmid substrate (Figure 3.8). As deadenylation occurred radioactively labelled 5'-adenylated groups were removed resulting in regeneration of cold open circular DNA. The extent of deadenylation was detected by autoradiography. The full-length human version of Xrcc4 protein and 'no Aptx protein' were used as negative controls. Unfortunately, a positive control was not available. Nevertheless, as expected, Xrcc4 controls did not generate any deadenylation activity. At Aptx concentrations of 50 – 100 pM, there was little to no removal of abortive ligation products. However, a dose dependant removal of AMP was seen starting at ~1 nM. This is a significant result as it demonstrates for the first time that it is possible to not only produce human Aptx in high yield and purity from bacteria, but importantly that the protein is highly active. This not only ensures that subsequent structural analysis can be interpreted with greater confidence but also provides the basis for production of protein needed to conduct screening of small molecule inhibitor libraries *in vitro*.



**Figure 3.4: Aptx removes abortive DNA ligation products. Products were analyzed by native PAGE and visualized by autoradiography. Reactions contained plasmid harbouring radioactivity labelled 5'-DNA-adenylate groups. The amounts of Aptx are indicated with (-) lane indicating Xrcc4 only as a control. All reactions were 10 ul total with 150 uM of plasmid DNA. Aptx protein used was full length HIS-GST Aptx and expressed in BL21 cells, purified by Nickel and S column chromatography. Aptx protein effectively removes abortive ligation products from DNA in a concentration dependant manner. Results indicate that Aptx protein promotes pyrophosphate bond hydrolysis on adenylated DNA intermediates, which brings about the removal of covalently bound AMP.**

## CHAPTER 4

### DISCUSSION

The investigation of optimal protein construct design reported in this thesis has been successful in not only identification of Aptx expression constructs capable of producing large quantities of highly purified human Aptx from bacteria, but has demonstrated that this protein is highly active and suitable for further inhibitor screening and *in vitro* structure-function analysis. Recently the structure of a related protein from *S. pombe* was reported by Ahel et al. (2011). The following discussion considers the approach taken by Ahel et al. (2011) and provides comparative analysis with that described in the current thesis work.

#### 4.1 Crystallization of Aptx-DNA-AMP-ZN<sup>2+</sup>

During crystallographic studies of Aptx, a crystal structure of a related protein found in *S. pombe* was published by Ahel et al. (2011). The following section relates their structure to the functional work described previously. Ahel et al. describe an unanticipated ‘wedge-pivot-cut’ mechanism for DNA 5'-adenylate access and processing. Structure-specific DNA binding by the HIT-Zn finger composite surface was further shown to be critical for engagement of DNA end or nicked substrates. It was suggested that this mode of substrate interaction may reflect the diversity found within the HIT superfamily.

The structure of yeast Aptx (yAptx) reveals a C2HE structural zinc-coordinating element that is related to the originally described C2H2 zinc-finger fold. Vertebrate Aptx homologs

consist of a C2H2 zinc-binding core, underscoring the potential for structural and functional interchangeability of C2H2 and C2HE zinc finger scaffolding in the aprataxin family. Importantly, the yeast Aptx structure expands the collection of C2H2-like zinc fingers to include sequence-independent DNA binding architectures. The ubiquity of the C2H2 DNA binding class, which is widespread throughout nature and predicted to be found in ~3.0% of human proteins, may be much larger than that which was previously estimated, given the expansion of the fold to include C2HE zinc binding motifs (Ahel et al., 2011).

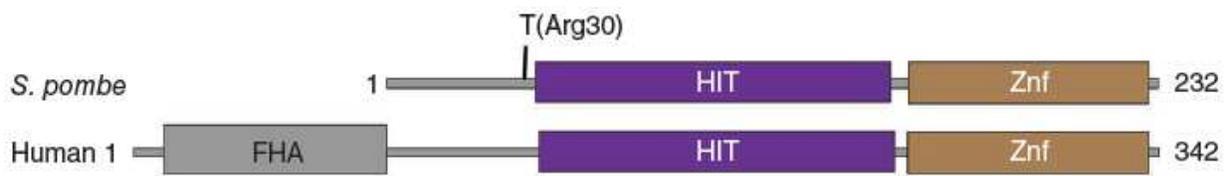
#### **4.2 Aptx Domain Mapping and Structure Determination**

The crystal structure which used a catalytically active domain of APTX was found using limited trypsin proteolysis. This construct spanned residues 30-232 and encompassed the HIT and Zn finger domains, known as Aptxc hereafter. Both full-length Aptx and Aptxc had comparable 5'-deadenylation activity on a nicked 5'-adenylated plasmid substrate (see Appendix). Interestingly, yeast Aptx does not contain an N-terminal FHA domain suggesting that some of the difficulty observed in crystallization of full-length human Aptx may be associated with 'flexibility' of the FHA domain within the context of an otherwise stable HIT-Zn finger core domain. Prior to this report, another student in the lab had already begun work on structure determination of the HIT-Zn finger region from human Aptx.

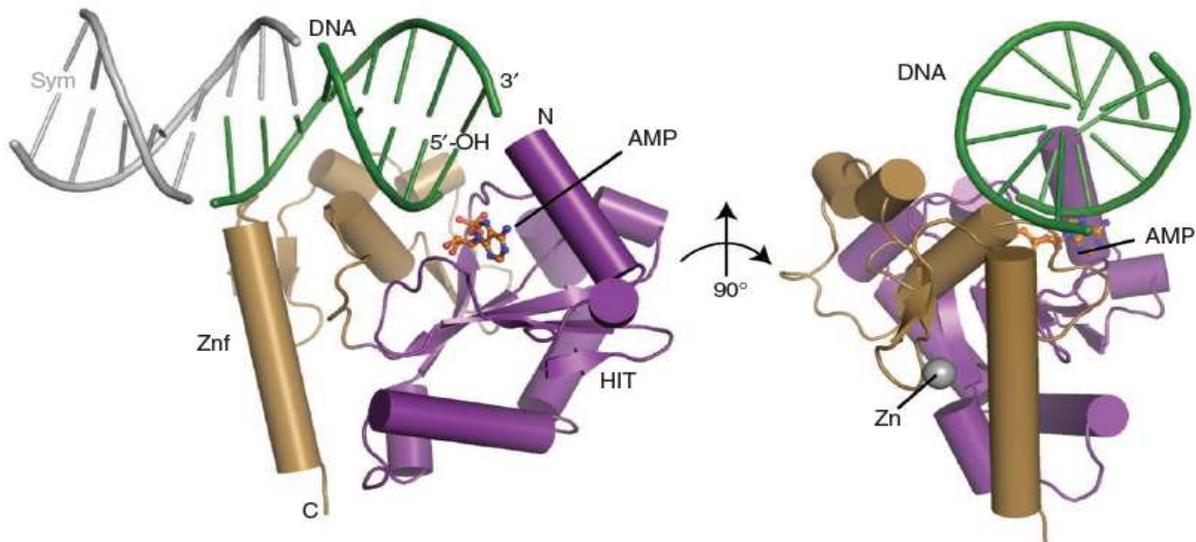
In order to understand Aptx binding to DNA and subsequent deadenylation of DNA, a DNA end-bound ATPX-DNA-AMP-Zn<sup>2+</sup> quaternary complex was crystallized. The presence of substrate DNA may provide further stability of the HIT-Zn finger domain and make it more amenable to structural studies. This possibility should be investigated further in ongoing

attempts to gain structural information of human Apx. A recent technical advancement reported by New England Biolabs, suggests that generating large quantities of adenylated substrate DNA oligo may no longer be a limiting aspect of pursuing Apx-substrate structural studies.

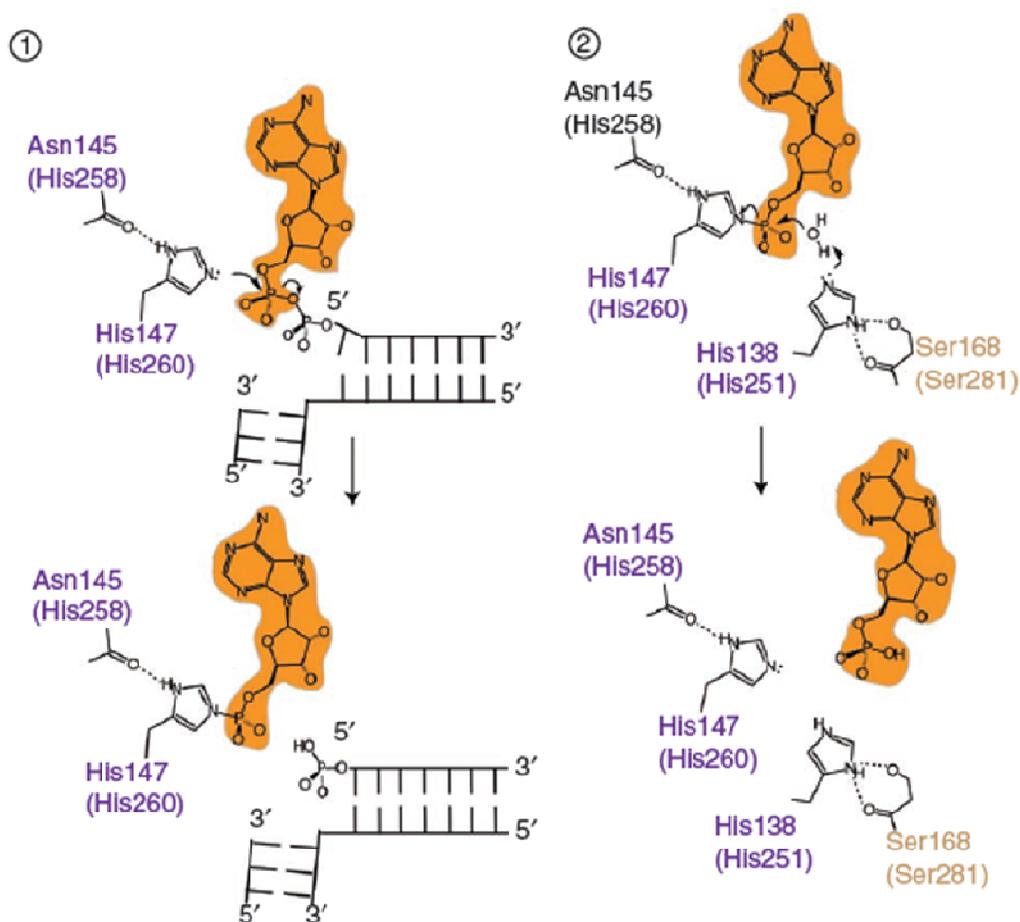
Vertebrate Apx orthologs are proposed to have  $C_2H_2$  zinc-binding motifs. However, the second histidine of the yeast Apx protein is replaced with a glutamate in fungal aprataxins (Date et al., 2001; Moreira et al., 2001). The Zn-binding capacity of Apx had only been proposed but not confirmed until recently through the crystal structure of yeast Apx. The crystallized complex included a duplex with two strands, with Apx bound at the DNA end bearing a 5'-hydroxyl group (Figure 4.1 and 4.2). Nevertheless, there is at present, still no crystal structure of the human version of Apx. This may be possibly due to the N-terminal region of human Apx, which contains disordered segments that most likely impeded ability to pack adjacent protein molecules required for crystallization.



**Figure 4.1: Domain comparison between Apx human and *S. Pombe*. The yeast version of the protein does not contain the FHA domain found in humans which has been implicated in binding other DNA repair factors (modified and adapted from Ahel et al., 2011).**



**Figure 4.2: Overall  $\gamma$ ATPX-DNA-AMP- $Zn^{2+}$  complex; orthogonal views of the fused Aptx HIT-Znf domain, HIT domain shown in purple, Znf in brown. The two domains are shown here collaborating to assemble the onto DNA (green) binding surface (modified and adapted from Ahel et al., 2011)**



**Figure 4.3: Structure-based mechanism for two-step direct reversal of DNA 5' end adenylation. Step 1 involves nucleophilic attack of the adenylate adduct with formation of an enzyme-AMP adduct. In step 2, the enzyme-AMP adduct is resolved by hydrolysis (modified and adapted from Ahel et al., 2011).**

### 4.3 Catalytic Mechanism and Structure/Function Relationships of Yeast Aptx

The function of Aptx is at least in part, if not entirely, related to its ability to remove 5'adenyl groups generated via abortive ligation during the final step of several DNA repair pathways. It has been suggested that the HIT domain of Aptx would catalyze 5'-deadenylation through a two-step mechanism requiring transfer of the AMP group to Aptx forming a covalent Aptx-AMP intermediate. This would be followed by the hydrolysis of the intermediate and

release of AMP. Surprisingly, however, both the HIT and Zn finger domains were found to contribute amino acid residues to the catalytic active site. Interestingly, the first residue of the HIT catalytic motif (Asn145 in *S. Pombe* Aptx) was not found to be conserved among all members of the aprataxin superfamily. Instead, this side chain projects away from the active site and mainly contributes to interaction with a carbonyl group in the main chain of the active site. Conserved His residues include, His147, His149 and His71. These histidines along with a unique His-Ser pair (His138 and Ser168 of the Zn finger) were found to be adjacent to the AMP product.

Together, analysis of the yeast Aptx structure suggests that the first step of deadenylation involves a nucleophilic attack of the DNA 5'-adenylate (Figure 4.3) by His147 that results in the formation of a covalent yAptx-AMP intermediate. Stabilization of His147 occurs through hydrogen bonding with the Asn145 main chain carbonyl group. As one would predict, alteration of His147 abolished the ability to carry out the first enzymatic step. The second step in deadenylation is predicted to involve a His149 acting to stabilize the covalent intermediate with the help of Ser168 of the Zn finger further aligning His138 that itself forms a hydrogen bond to the product AMP phosphate oxygen. That is, the bridging oxygen in a 5'-adenylated DNA substrate. This is in line with His147 and the AMP  $\alpha$ -phosphate. Due to this orientation, the yeast Aptx-specific pair, His138 and Ser168 were suggested to contribute to hydrolytic activity. Furthermore, it was proposed that His138 may also contribute as a proton donor to the bridging oxygen during the formation of a covalent AMP-Aptx intermediate. Consistent with these proposals, His138A and S168A mutants were impaired for ability to deadenylate plasmid DNA *in vitro*.

It will be interesting to see if similar findings hold true for human Aptx. The ability to generate and purify human Aptx from bacteria will greatly aid in further studies of human Aptx focused on understanding of structure-function relationships. In addition, the simple and robust activity assay described in this thesis will permit facile quantitative analysis of deadenylation activity necessary for such studies.

## CHAPTER 5

### SUMMARY AND CONCLUSIONS

The importance of Aptx in mammals is represented by the fact that mutations in the human aprataxin gene are linked to the autosomal recessive neurological disorder AOA1 (Date et al., 2001; Moreira et al., 2011) with coenzyme Q10 (coQ10) deficiency (Quinzii et al., 2005) as well as other syndromes that follow a similar clinical manifestation including multiple system atrophy (MSA) (Baba et al., 2007). Cell lines taken from patients with AOA1 have been shown to succumb to DNA damage following exposure to oxidative stress (Harris et al., 2009; Hirano et al., 2007) and the topoisomerase inhibitor, camptothecin (Mosesso et al., 2005). The homologs of human Aptx from budding and fission yeast also exhibit clastogen sensitivity and provide evidence of further conservation of the Aptx gene function. Reports from several groups have demonstrated that the accumulation of 5'-adenylates following single strand breaks can negatively affect cellular transcription (Ahel et al., 2006; Caldecott et al., 2008; Rass et al., 2007) and give rise to neurodegeneration and neural cell damage.

The Aptx gene is a member of the histidine triad family of nucleotide hydrolases and transferases (Lima et al., 1997; Brenner et al., 2002). Biochemical studies support the hypothesis that the Aptx catalytic mechanism involves formation of a covalently bound AMP intermediate followed by subsequent protein-assisted hydrolysis (Rass et al., 2008). Mammalian Aptx differs from the yeast homologs in that it contains an N-terminal FHA domain that mediates interaction with Xrcc1, Xrcc4 (Clements et al., 2004) and MDC1 (Becherel et al., 2010). Reflecting these diverse interactions, Aptx has been implicated to function in the final ligation step in multiple

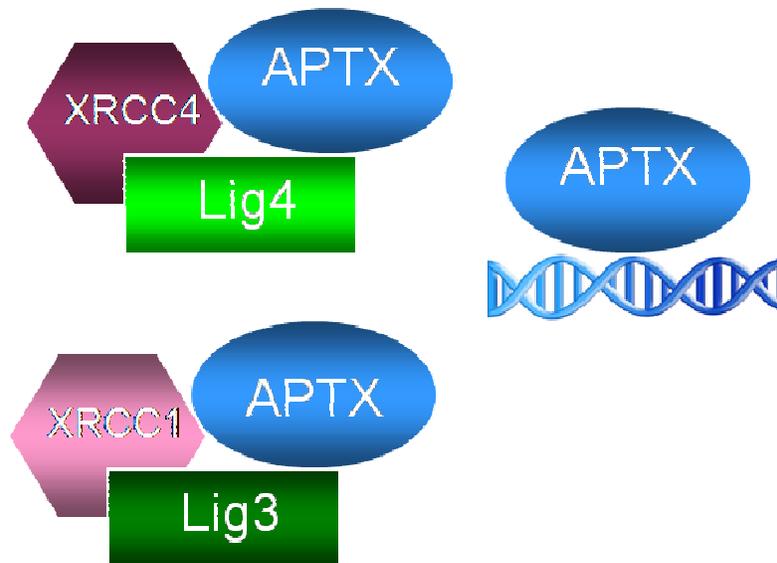
repair pathways including NHEJ, HR, BER and SSB. Apx repairs abortive ligation products and has comparable efficacy towards 5'-adenylated blunt and protruding ends as well as buried nicked DNA substrates (Rass et al., 2007). Although current structural data is limited to the yeast homologue of Apx it is likely that in terms of the catalytic mechanism many features will be conserved across the larger family of Apx proteins. Interestingly, however, mapping studies using DNase I footprinting showed human Apx protects 10-11 base pairs of duplex DNA downstream of the 5'-adenylate, and 2-3 bases of ssDNA upstream of the gap, which is in contrast to what is observed in the crystal structure of yeast Apx. Yeast Apx protects less DNA, suggesting yeast and human proteins may interact somewhat differently with their DNA substrates, perhaps reflecting other differences not apparently obvious from sequence alignments alone. As more structural data becomes available for other members of the Apx family it will become increasingly clear to what extent particular features of substrate recognition and catalysis are shared and unique amongst family members.

Human Apx has three domains (Figure 1): a zinc finger domain (Moreira et al., 2001), a histidine triad domain (HIT) (Moreira et al., 2001) and a forkhead associated domain (FHA) (Date et al., 2004). While the zinc finger domain is expected to participate in DNA binding and the adjacent HIT domain carry out catalysis, it is the N-terminal FHA domain that has been suggested to be responsible for Apx's ability to participate in different repair pathways. The FHA domain from human Apx is able to associate in a phosphorylation dependent manner with proteins from SSB, NER and BER repair (Xrcc1) as well as DSB repair (Xrcc4). Interestingly, the yeast homologue of Apx does not contain a FHA domain and is several hundred amino acids shorter than human Apx protein. The fact that yeast maintain equivalent repair mechanisms, suggests that there may be a number of differences between the way yeast and human Apx

interact with binding partners in various repair pathways and perhaps how they interact with substrate as well. Work in this thesis demonstrated that phosphorylated peptides of Xrcc4 were sufficient for interaction with human Apx. It will be interesting to determine how binding of Xrcc4 to Apx occurs and if this interaction has any regulatory effect on catalytic activity, in addition to interaction with Xrcc1 and Xrcc4.

The repair of adenylated DNA following abortive ligation is dependant on the presence of Apx (Ahel et al.,2006). To further understand the mechanistic basis of this important repair function, structural studies of the human Apx protein were undertaken. Although not all technical challenges associated with this task were overcome during the completion of this thesis, several important technical advancements were made that will contribute to future research with this protein. In particular, work completed in this thesis describes a new method for production of milligram quantities of highly purified human Apx. This method is simple and makes use of a bacterial expression system. Importantly, protein was shown to retain robust 5' deadenylation activity. Together these advancements will permit not only further attempts to gain structural information, but also analysis of function in vitro as well as screening for small molecule inhibitors.

The crystal structure of HIT-Zn finger domain of Apx from *S. pombe* have provided the first structural insight into the enzymatic activity of Apx. It will be very interesting to gain similar insight into the human Apx enzyme. In addition, determining the structure of human Apx in complex with other DNA repair factors (Figure 5.1) will be important to understand how DNA transitions occur within the context of a higher-order nucleoprotein DNA repair complex.



**Figure 5.1: Apx forms a complex with Xrcc4 and Ligase IV during DSB and NHEJ, and Xrcc1 and Ligase III during SSB and BER. Apx also binds directly to abortive ligation products in order to process adenylated DNA products following an abortive ligation. This allows DNA to be processed and ligated during DNA repair.**

In summary, Apx is a novel protein required for repair of abortive DNA ligation products. During SSB repair, Apx forms a complex with Xrcc1 and ligase III. Apx is also involved in DSB repair through the formation of a complex with Xrcc4 and DNA ligase IV. The mechanistic basis of its role in SSB repair and DSB repair has become clearer following the recent successful determination of the crystal structure of yeast Apx. Further studies are nevertheless required in order to understand the protein-protein and protein-DNA interactions of human Apx. Identifying the binding regions of Apx with other repair factors and understanding how these interactions effect DNA repair will give insight into how Apx functions within the context of a higher-order repair complex. Ultimately this knowledge may prove useful in design of new clinically important therapeutics.

**SUPPLEMENTARY MATERIAL AND APPENDICES**

Supplementary Table 1. MSc. Goals and Achievements

<b><u>Goals- Structural Studies</u></b>	<b><u>Achievement Progress</u></b>
<b>1. Full length Aptx +/- His Tag</b>	<b>Expressed and Purified</b>
<b>2. Express, purify, crystallize FHA domain</b>	<b>Expressed and Purified</b>
<b>3. Crystal of Aptx in complex with DNA</b>	<b>Achieved (Ahel et al., 2011)</b>
<b>4. Understand enzymatic mechanism of Aptx</b>	<b>Achieved</b>

**Buffer Conditions**

Supplementary table 2 and 3: Aptx FL Buffers

Nickel Column Buffers

<b>NiA</b>	<b>NiB</b>
1.5 M KCl	1.5 M KCl
20 mM Tris Ph 8	20 mM Tris Ph 8
0.03% LDAO	0.03% LDAO
1 mM BMe	1 mM BMe
10 mM Imidazole	450 mM Imidazole
10% glycerol	10% glycerol

SP-sepharose Column Buffers

<b>Sa</b>	<b>Sb</b>
-	2 M KCL
20 mM MES Ph 6.5	20 mM MES Ph 6.5
10% glycerol	10% glycerol

Supplementary table 4 and 5: APTX FHA Buffers

Aptx FHA

<b>NiA</b>	<b>NiB</b>
50 mM Hepes pH 7.5	50 mM Hepes pH 7.5
200 mM KCL	200 mM KCL
10% glycerol	10% glycerol
10 mM Imidazole	300 mM imidazole

<b>Sa</b>	<b>Sb</b>
-	500 mM KCL
50 mM Hepes Ph 7.5	50 mM Hepes Ph 7.5
10% glycerol	10% glycerol

### Xrcc4 Protein Purification

Nickel column buffer conditions are the same as that of Aptx FL, except with 300 mM Imidazole in NiB.

### **Aptx Activity Assay**

Step 1: To Generate Breaks in DNA

Ingredient	Concentration
DNA	150 nM
NaCl	0.1 M
NADH	100 mM
H <sub>2</sub> O <sub>2</sub>	20 Um
FeCl <sub>3</sub>	0.2 mM
EDTA	0.2 mM

- Incubate Reaction at Room Temperature for 1.5 hour
- Stop Reaction with 5mM EDTA

Step 2: Abortive Ligation

DNA reaction from step 1
10X T4 Ligation buffer
T4 DNA ligase
$\alpha$ - P32 ATP
Nuclease free water

- Keep overnight in 37° waterbath with appropriate shielding
- Stop reaction with 40 nM of EDTA

Step 3: Activity testing

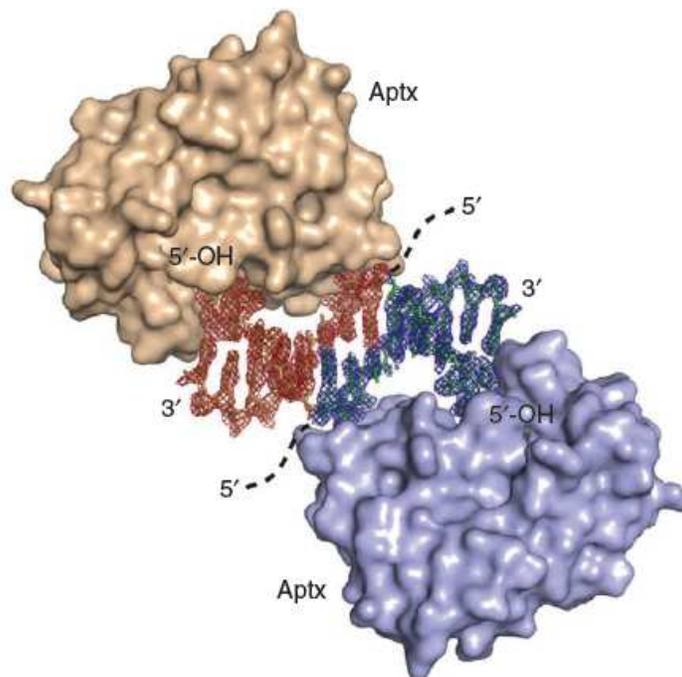
- Add desired amount of Aptx protein, incubate at room temperature for 1 hour and run reaction out on 1% agarose gel

- Image get using autoradiography.

## Crystallographic Analysis of Aptx Ortholog in *S. Pombe*

### Aptx Domain Mapping and Structure Determination

Using limited trypsin proteolysis, a catalytically active ordered domain of Aptx in *S. Pombe* was found. This contained residues 30-232 and encompassed the HIT and Znf domains, known as Aptxc. Both full-length Aptx and Aptxc had comparable 5'-deadenylation activity on a nicked 5'-adenylated plasmid substrate.

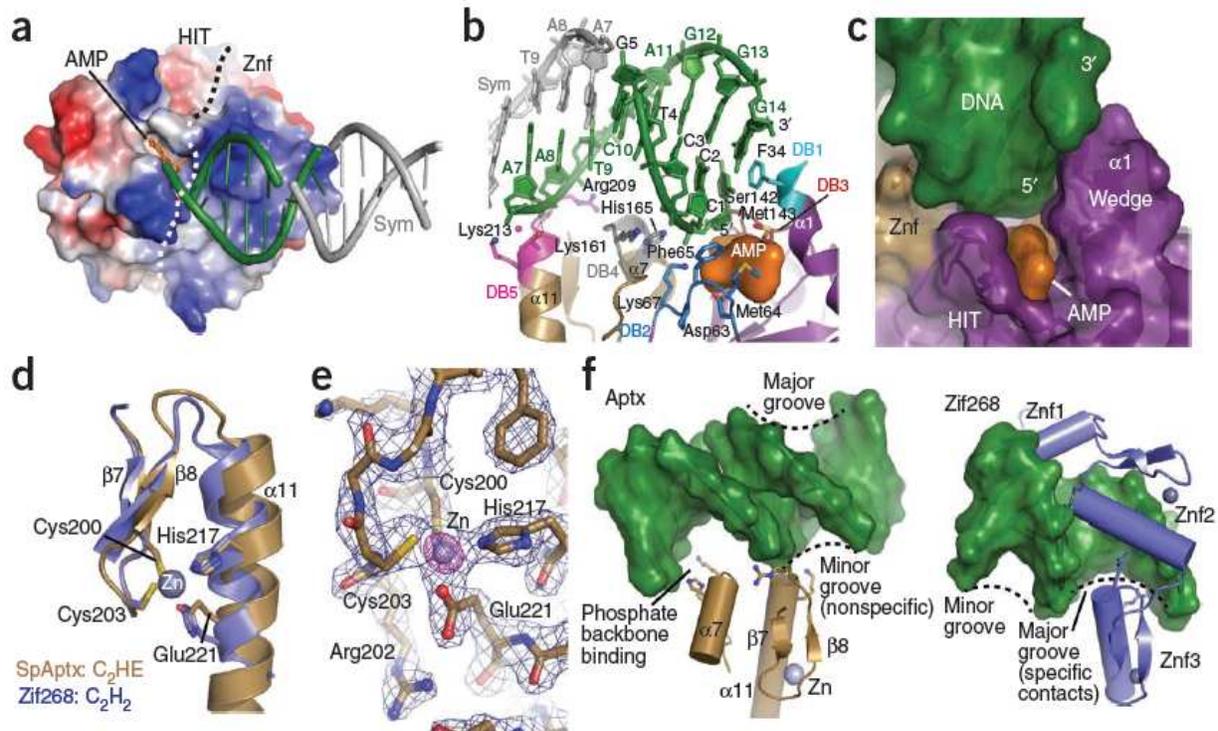


---

**Supplementary figure 1: DNA protein complex formation in the crystal. The  $\sigma_A$ -weighted  $F_0-F_C$  positive difference density map calculated before building the model is shown for symmetry related molecules (red and blue density, contoured at  $2.4 \sigma$ ) (modified and adapted from Ahel et al., 2011).**

### **yAptx-DNA-AMP-Zn<sup>2+</sup> Complex Architecture**

The crystal structure of yAptx-DNA-AMP-Zn<sup>2+</sup> revealed the union of HIT nucleoside hydrolase-related fold, specifically in residues 33-153 with a Znf binding domain between residues 154-232. Together, the two domains make up the Aptxc domain. The HIT domain of Aptx has a central five stranded antiparallel  $\beta$ -sheet ( $\beta$ 1- $\beta$ 5) that is supported on three sides by eight helices ( $\alpha$ 1- $\alpha$ 6,  $3_{10}^1$ ,  $3_{10}^2$ ). The catalytic domain has a positive charged dsDNA interaction scaffold. There is also a network of 20 conserved hydrophobic residues that reinforce the HIT-Znf domain, suggesting that there is a relative alignment and connectivity of the domain that is important for interdomain folding and DNA deadenylation activity. The DNA is found in end-bound complexes within the crystals across a crystallographic two-fold axis, which extends a DNA pseudo-helix through interactions involving a three-base bridge. The footprint of Aptx on the DNA end is also consistent with previous studies done using DNase I mapping that showed Aptx binding to the region at the 5' end of an adenylated DNA substrate (Rass et al., 2008). The quaternary structure of the yAptx-DNA-AMP-Zn<sup>2+</sup> crystal describes how yAptx has evolved a protein surface that is specifically suited for binding and initiating deadenylation activity on a 5'-adenylated DNA substrate. This deadenylation activity is a characteristic that is unique to Aptx. Aptx is monomeric, and a Znf domain  $\beta$ -strand ( $\beta$ 6) caps the Aptx HIT  $\beta$ -sheet which blocks the formation of a canonical HIT dimer. The Znf motif also contributes to a composite HIT-Znf active site that surrounds the HIT catalytic machinery core with strong interaction with the positively charged dsDNA.



**Supplementary figure 2: yAptx structure specific DNA binding.** YAptx structure-specific DNA binding. (a) The yAptx HIT and Znf surfaces assemble an electropositive DNA binding platform. AMP (orange) binds in a recessed hydrophobic pocket. A dashed line demarcates surface boundaries of the HIT and Znf domains. (b) Five conserved regions (DB1–DB5) mediate yAptx DNA binding contacts. DNA base stacking from DB1 (turquoise) and DB3 (red) probe and interrogate the dsDNA-end base stack for end recognition. DB2 (blue) and DB4 (gray) bind the incoming 5' strand, whereas DB5 engages the complementary nonadenylated strand. (c) Surface representation of the DNA basestacking wedge. The AMP product is sandwiched between the wedge, the 5' strand and the HIT-Znf active site. (d) The Aptx C2HE zinc finger is structurally related to the C2H2 zinc fingers. YAptx Glu221 replaces the second histidine of a canonical C2H2. (e) A zinc anomalous difference Fourier map calculated using phases from the final protein–DNA complex (purple, shown at  $14 \sigma$ ) reveals the position of the single bound zinc. The blue electron density map is the final  $\sigma_A$ -weighted  $2Fo - Fc$  map and is contoured at  $1.4 \sigma$ . (f) The Aptx C2HE Znf domain engages the DNA minor groove. Left, the Aptx  $\beta\alpha$ -fold and  $\alpha 7$  bind the phosphate backbone in a previously unknown minor groove DNA-binding mode. Right, canonical major groove interaction by the Zif268 C2H2 domains (modified and adapted from Ahel et al., 2011).

The Znf domain of Aptx binds to DNA sequence independently in order to act as a DNA processing enzyme. yAptx has four residues in two zinc binding motifs, Cys200 and Cys 203 from ZB1 in the Znf  $\beta$ 7- $\beta$ 8 sheet and His 217 and Glu221 from the ZB2 of helix  $\alpha$ 11 which tetrahedrally coordinate to bind  $Zn^{2+}$ . Additionally, helical element  $\alpha$ 7 and DNA binding recognition DB4 and DB5 also assemble around the  $C_2HE \beta\beta\alpha$  core (Supplementary Figure 2). The Znf binds DNA through exclusively the minor groove in a non-specific manner and interacts with the sugar-phosphate backbone. This is in contrast to other DNA binding proteins which recognizes the major groove.

Data Collection	Aptx-DNA-AMP- $Zn^{2+}$ complex
Space group	/432
Cell dimensions a,b,c (Å)	157.05 Zn-SAD
Resolution (Å)	50-2.35 (2.43-2.35)
$R_{merge}^a$	10.2 (46.2)
Completeness (%)	100.0 (100.0)
Redundancy	7.7 (7.6)
Refinement	
Resolution (Å)	50-2.35
No. Reflections	25.992
$R_{work} - R_{free}$	16.5/21.1
No. Atoms	2,153
Protein	1,645
DNA	265
AMP	23
Zn	1
Solvent	219
<i>B</i> -factors	
Protein	23.9
DNA	50.6
AMP	23.7
Zn	41.6
Solvent	35.1
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angle (°)	1.062

---

**Supplementary Table 6: Data collection and refinement statistics (Ahel et al., 2011).**

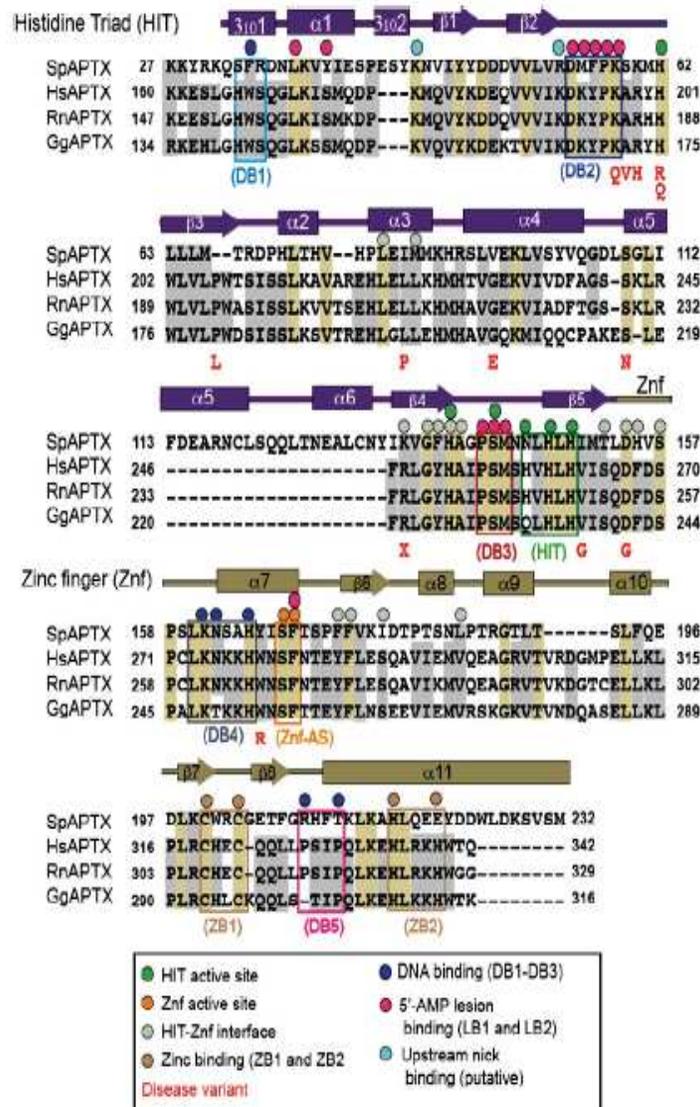
## **yAptx DNA Binding**

The crystal structure of the yAptx complex aided in elucidating the mechanism of 5'-deadenylation activity between DNA and protein. Three characteristics define yAptx binding to DNA: (i) sugar phosphate binding of 5'-adenylated strand by HIT and Znf domains, (ii) binding of the nonadenylated strand by the zinc finger and (iii) DNA end base stacking by helical hydrophobic wedge. The continuous HIT Znf domain bind to the major groove of the dsDNA and about 990 Å<sup>2</sup> of solvent accessible protein surface is buried after binding to DNA and AMP. The DNA binding surface is made up of ten residues which includes DB1-DB3 of the HIT, DB4-DB5 of the Znf. Phe65, Lys67 of DB2 from the HIT domain and Ser142-Met143 of DB3 along with Lys161 and His165 of Znf DB4 directs the 5'-adenylated DNA strand onto the HIT active site. Interestingly, there is a conserved motif shared between yAptx and DNA ligase I known as the PK motif in the DB2 loop. This may be due to convergent evolution which selected these structurally analogous proteins for 5'-adenylated DNA binding and processing.

Lys213 was found to have a solvent bridged hydrogen bond DNA contact and direct DNA phosphate backbone binding by Arg209 in order to engage the complementary nonadenylated strand. Between human and *S. Pombe*, the DB1-DB4 binding regions are conserved but DB5 is not. The *S. Pombe* DB5 contains a conserved ZNF DB4 lysine-lysine pair at the N-terminus of  $\alpha 7$  of vertebrate yAptx that given the position of the loop, can be conjured to have interaction with the non adenylated strand. Hydrogen bonding to the DNA phosphate backbone occurs through the main chain amide of the C<sub>2</sub>HE- $\beta\beta\alpha$  core which is on the electropositive end of the helix  $\alpha 11$  dipole. This also most likely contributes to the yAptx-DNA interaction binding energy.

A helical hydrophobic wedge form by the HIT domain N-terminal helices was found to interact with the dsDNA-end base pair (Supplementary figure 2, part B). Interactions between

the HIT domain wedge with the DNA as well as the AMP contacts comprise 40% of the DNA adduct binding surface. The conserved aromatic residue Phe34 (Trp167 in human ATPX) from loop DB1 forms a  $\pi$ -stack which leads to the distortion of the terminal C-G base pair.



**Supplementary figure 3: yAptx structure-based sequence alignment.** Secondary structure for the HIT domain (purple) and Znf domain (brown) is displayed above ClustalW sequence alignment for selected Aptx homologs. Residue classifications are marked in the legend. AOA1 disease variants are shown in red sequence under the alignment (modified and adapted from Ahel et al., 2011).

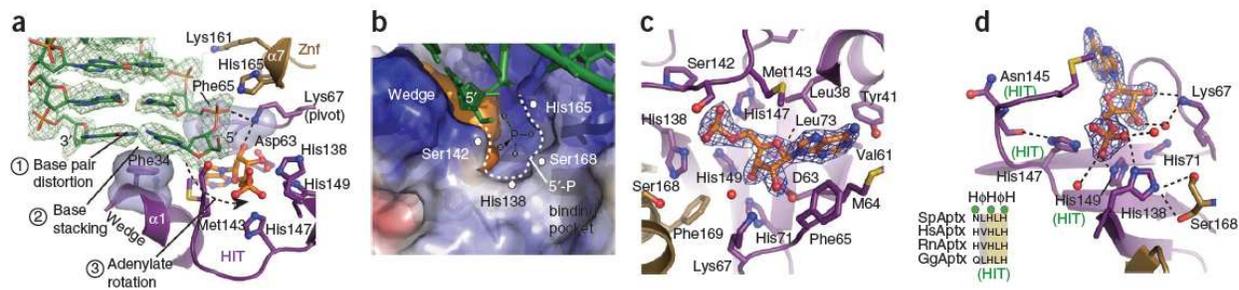
### **yAptx-DNA Interaction Motifs**

In order to understand the binding efficacies of Aptx with DNA, FITC (fluorescein isothiocyanate) conjugated DNA substrates were used in EMSA assays with wildtype and various Aptx mutants. Full-length Aptx had a binding preference for gapped substrates with a 3'-phosphate with a  $K_d$  of 80 nM over dsDNA with  $K_d=$  190 nM or ssDNA with  $K_d=$  330 nM. This type of damage is similar to that undergone following oxidative stress. An FPK mutant with a F65A change had lower affinity for binding the SSB substrate with  $K_d=$  250 nM. Mutants with K67E, K16E and H165E weakened DNA binding even further. Mutations in the Znf domain or FPK domains also impaired plasmid DNA deadenylation activity. This implies that both the domains are required for 5'-deadenylation activity.

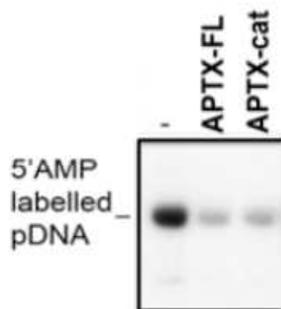
### **Accessing DNA 5'-adenylate**

In order to access double stranded DNA previously undergoing repair, Aptx must overcome steric hindrances, unlike other members of the HIT superfamily. The DNA was found slightly underwound and bent, relative to B-form DNA. The terminal C-G base pair was distorted by the Znf and FPK motifs. The FPK motif acts like a pivot point for adenylation (hereon referred to as the pivot) binds the 5'-phosphate and to the AMP adduct. The catalytic domain of yAptx makes a binding pocket which recognizes AMP that is sandwiched between the  $\alpha$ -wedge, the HIT active site and the 5' strand. Six residues comprising of Leu38, Tyr41, Met64, Phe65, Lys73 and Met143 bind the adenosine base. The pivot creates van der Waals interaction and hydrogen bonding through Lys67 and Phe65 along with the Asp63 side chain that forms hydrogen bonds with the ribose 2' and 3' hydroxyls in order to direct AMP into the active site such that it is orthogonal to the 5' strand.

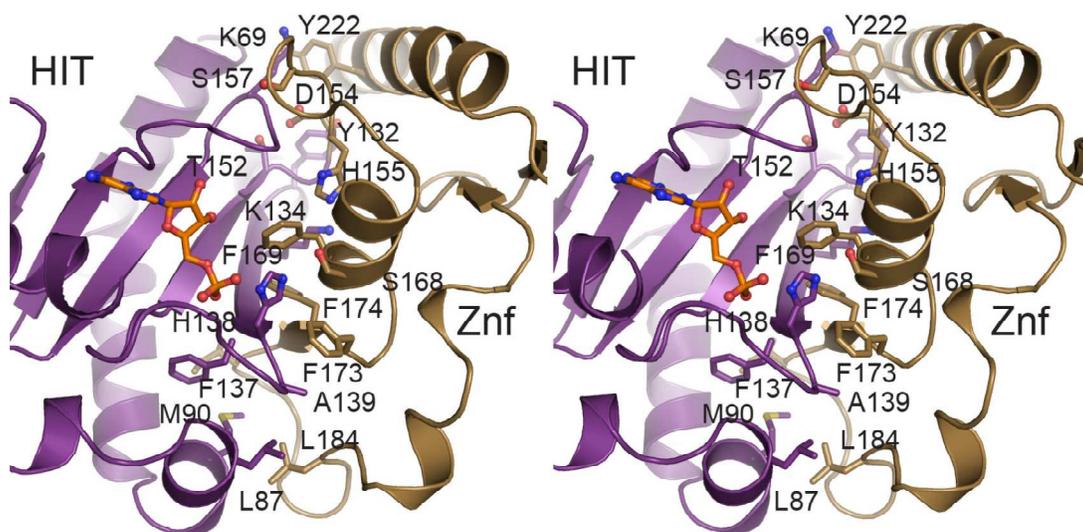
Phe65 aids in stabilized the DNA 5'-adenylate in order to catalyze the reaction. The crystals were found to have an AMP DNA end containing a 5'-hydroxyl. The positively charged pocket formed by the HIT domain and Znf domain containing residues His138, Ser168, His165 and Ser142 is required to bind DNA and is required for deadenylation activity. Mutations which impair these residues results in the inability of yAptx to deadenylate DNA. In summary, the general mechanism for accessing the DNA 5'-adenylates involves a wedge mediated DNA end or nicked DNA base stacking, FPK pivot and wedge driven adenylate rotation and terminal base pair distortion.



**Supplementary figure 4: yAptx adenylate access and catalytic mechanism. (a)** yAptx DNA 1 2 end binding and adenylate rotation is mediated by the HIT domain wedge and pivot (FPK) elements that extract and align 5'-adenylate into the HIT-Znf active site. Unbiased  $\sigma_A$ -weighted  $F_o - F_c$  positive difference density map (green, contoured at  $2.4 \sigma$ ) calculated before building the DNA model is shown. Phe65 and Lys67 of the pivot directly bind AMP and the sugarphosphate backbone of the incoming 5'-strand (b) DNA 5'-phosphate (5'-P) binding pocket. An electrostatic surface potential with electropositive surface (blue) and electronegative surface (red) is shown with an orange surface representation for the AMP. The DNA 5'-phosphate binding pocket is indicated as a dashed line. (c) The AMP binding pocket is shown with bound AMP. Blue electron density is the final  $\sigma_A$  weighted  $2F_o - F_c$  map, contoured at  $1.4 \sigma$ . (d) The Aptx active site catalytic residues. Canonical HIT motif residues are marked (green 'HIT'). Residues His138 and Ser168 are conserved in the aprataxins, and they assemble to form the Aptx HIT-Znf composite active site. Sequence for yeast (*S. pombe*, SpAptx), human (*Homo Sapiens*, HsAptx), rat (*Rattus norvegicus*, RnAptx) and chicken (*Gallus gallus* GgAptx) are displayed (modified and adapted from Ahel et al., 2011).



**Supplementary figure 5: yAptx and Aptxc has comparable catalytic activity (modified and adapted from Ahel et al., 2011).**



**Supplementary figure 6: Structure of the HIT-Znf interface. Residues conserved at the HIT-Znf interface are shown. HIT domain is in purple, Znf in gold. Bound AMP is shown in orange. DNA is omitted for clarity. A stereo view is displayed (modified and adapted from Ahel et al., 2011).**

#### Accession codes

Coordinates for the yAptx-DNA-AMP-Zn<sup>2+</sup> complex have been deposited in the Protein Data Bank (accession code 3SZQ).

## **Protein Protein Interaction Experiments Between Aptx and its Partners**

The experiments and the results found in the following section has been conducted by my student, Kritica Arora

### **Characterization of the interaction between non-phosphorylated XRCC4 peptide and the FHA domain**

There was no interaction found between the non-phosphorylated XRCC4 peptide (see Table 7) and the FHA domain. The RU, following auto-process and double referencing, stayed at baseline. It is proposed that the non-phosphorylated peptide most likely behaved as an ideal control for both processing the data; indicating that the system was working correctly.

### **Classification of the interaction between pseudo-phosphorylated XRCC4 peptide and the FHA domain**

We found no interaction between pseudo-phosphorylated XRCC4 peptide (see Table 7) and the FHA domain. After auto-processing and double referencing, the sensogram showed a similar behavior that was displayed by the non-phosphorylated XRCC4 peptide – at or below baseline RU. This points towards an authentic phosphate group would be required for any interaction between Aptx and Xrcc4.

### **Visualization of the heterogeneous analyte model used to calculate the kinetics of the interactions**

Two varying affinities were seen for the same epitope or peptide resulting in the two sets of  $K_D$  values (Table 9). The results indicated a competition between two analytes. One analyte

bound rapidly to the peptide but was rapidly displaced by the second analyte. However, the association between the ligand and second analyte was brief due to low affinity binding. This eased the way for the first analyte to re-associate and was followed by slow dissociation. The results indicate a high affinity interaction. Due to the competition between the two-analyte species, steady-state was not achieved as was indicated by the lack of a plateau from the slope during the association time period.

The sensogram for triple phosphorylated XRCC4 peptide indicated a unique problem, as an upward drift was visible in the post-referenced data. This led to an imprecise kinetic fit of the heterogeneous analyte model and thus, no kinetic constants were able to be calculated.

### **Analysis of the kinetics for the phosphorylated XRCC4 peptides**

The highest affinity interaction was seen between pThr-pSer *bis*-phosphorylated XRCC4 peptide (YDESpTDEEpSENQTD) and Aptx, (210nM as the calculated equilibrium dissociation constant;  $K_D1$ ). Ser237 (YDESTDEEpSENQTD) is thought to be required for the interaction between XRCC4 peptide and the FHA domain as  $K_D1$  was 330nM (see Table 9). The FHA domain located in the N-terminal region of Aptx also interacted with the mono-Thr233 phosphorylated XRCC4 peptide (YDESpTDEESENQTD) but with a lower affinity ( $K_D1=8.5\mu\text{M}$ ). Following phosphorylation of Ser232 in the XRCC4 peptide the  $K_D1$  values increased (YDEpSpTDEEpSENQpTD at  $24\mu\text{M}$ , and YDEpSpTDEESENQTD at  $33\mu\text{M}$ ). Thus, we hypothesize that the phosphate group on this residue may be hindering the reaction.

### **Determination of steady-state by equilibrium analysis:**

Equilibrium analysis was conducted on the mono-Ser237 and di-Thr233, Ser237 phosphorylated XRCC4 peptides using SPR. An exponential curve, instead of a plateau at specific RU was found. These results suggest the lack of steady-state interactions between the heterogeneous analytes and the various ligands.

In conclusion, it appears that Xrcc4 and Aptx do interact but further studies are required to optimize binding reactions through identification of binding regions and phosphorylation statuses.

## TABLES

14-1	YDESTDEESENQTD
14-2	YDES <b>E</b> DEESENQTD
14-3	YDE <b>S</b> TDEE <b>S</b> ENQ <b>T</b> D
14-4	YDE <b>S</b> TDEESENQTD
14-5	YDESTDEE <b>S</b> ENQTD
14-6	YDESTDEESENQ <b>T</b> D
14-7	YDE <b>S</b> TDEESENQTD
14-8	YDE <b>S</b> TDEE <b>S</b> ENQTD
14-9	YDE <b>S</b> TDEE <b>S</b> ENQTD

---

**Supplementary Table 7: XRCC4 peptides used in SPR experiments cover residues 229-242. These 14 amino acids were chosen specifically within the 213-250 region that the FHA domain binds to. This is because PNKP's FHA domain was shown to interact with this XRCC4-derived peptide - YDESTDEESEKK. (modified and adapted from Kritica Aroroa).**

XRCC4 peptides		XRCC4 peptides	
14-1, 14-2, 14-3, 14-4, 14-7 and 14-8		First sensor chip: 14-1, 14-2, 14-3, 14-4, 14-7 and 14-8 Second sensor chip: 14-1, 14-5, 14-6 and 14-9	
Sensor chip row	FHA concentration (nM)	Sensor chip row	FHA concentration (nM)
A1	105	A1	50
A2	80	A2	25
A3	54	A3	12.5
A4	29	A4	6.25
A5	3.4	A5	3.125
A6	Running Buffer	A6	Running Buffer

**Supplementary Table 8: The varying FHA concentrations used for studying the interaction with the XRCC4 peptides. Two defined groups of FHA concentrations were used in the SPR experiments to obtain the desired  $R_{max}$  value of 100 RU for the highest [FHA] and achieve equal correlation between [FHA] dilutions and decrease in RU values. The starting [FHA] was lowered to 50 nM for the second set of experiments in an effort to decrease the possible aggregation of the analyte. The sixth injection was simply the running buffer to be used for double referencing the data during kinetic analysis (modified and adapted from Kritica Arora).**

Peptides	$K_{a1}$ (1/Ms)	$K_{d1}$ (1/s)	$K_{D1}$	$K_{a2}$ (1/Ms)	$K_{d2}$ (1/s)	$K_{D2}$
YDEpSpTDEEpSENQpTD	3.26	8.1E -5	24 $\mu$ M	3.26	8.1E -5	24 $\mu$ M
YDESpTDEESENQTD	6.49	5.5E -5	8.5 $\mu$ M	9.49	2.2E -2	2.3m M
YDESTDEEpSENQTD	2.2E4	7.3E -3	330n M	1.6E4	1.9E -5	1.2nM
YDEpSpTDEESENQTD	2.52	8.3E -5	33 $\mu$ M	2.52	8.3E -5	33 $\mu$ M
YDESpTDEEpSENQTD	3.4E4	7.0E -3	210n M	2.9E4	1.8E -5	0.6nM
YDESTDEESENQpTD	N/A					
YDEpSpTDEEpSENQTD	N/A					
YDESTDEESENQTD	No binding					
YDESEDEESENQTD	No binding					

**Supplementary Table 9: Calculated  $K_a$  ( $k_{on}$ ),  $K_d$  ( $k_{off}$ ) and  $K_D$  (equilibrium dissociation constant) obtained for the nine multiply phosphorylated XRCC4 peptides following heterogeneous analyte modeling. The two sets of  $K_D$  values result from the presence of multiple species within the analyte sample competing for a single ligand site. Some values land within the unlikely range – for example:  $K_a$  of 2.2E4 (1/Ms) for the interaction between FHA domain and *bis*-phosphorylated XRCC4 peptide. There is also a lack of proper kinetic fit of the heterogeneous analyte model for the interaction between some XRCC4 peptides and the FHA domain. Further optimization and multiple runs are required to obtain the true set of  $K_D$  values (modified and adapted from Kritica Aroroa).**



---

**Supplementary figure 7: Unrefined image of Aptx activity gel with indicated concentrations of Aptx protein; 0 indicates no Aptx added, (-) indicates lane with only Xrcc4 protein and the last lane, labelled as ATP is of ATP only.**

## References:

- Ahel, I., Rass, U., El-Khamisy, S.F., Katyal, S., Clements, P.M., McKinnon, P.J., Caldecott, K.W. and West, S.C. The neurodegenerative disease protein aprataxin resolves abortive DNA ligation intermediates. *Nature*. 443, 713-715. (2006).
- Ahel et al. Structure of an aprataxin–DNA complex with insights into AOA1 neurodegenerative disease. *Nature Structural & Molecular Biology* 18, 1189–1195 (2011).
- Baba, Y. *et al.* Aprataxin (APTX) gene mutations resembling multiple system atrophy. *Parkinsonism Relat. Disord.* 13, 139–142 (2007).
- P R Blier, A J Griffith, J Craft and J A Hardin. Binding of Ku protein to DNA. Measurement of affinity for ends and demonstration of binding to nicks. *The Journal of Biological Chemistry*, 268, 7594-7601 (1993).
- Bliss TM, Lane DP. Ku selectively transfers between DNA molecules with homologous ends. *J Biol Chem.*272(9):5765-73 (2001).
- Caldecott, Keith W. Single-strand break repair and genetic disease. *Nature Reviews*. 619. (2008).
- Carey, R.B., Peterson, S.R., Wang, J-T., Bear, D.G., Bradbury, E.M. and Chen, D.J. DNA Looping by Ku and the DNA-Dependent Protein Kinase. *Proc. Nat'l. Acad. Sci. USA* 94: 4267-4272 (1997).
- Chen, L, Trujillo, K, Ramos, W, Sung, P, and Tomkinson, AE Promotion of Dn14 mediated DNA End Joining by the Rad50/Mre11/Xrs2 Complex. *Molecular Cell* 8(5):1105-1115 (2001).
- Cleaver, J.E. Defective repair replication in xeroderma pigmentosum. *Nature*, 218, 652 656. (1968)
- Clements, P.M., Breslin, C., Deeks, E.D., Byrd, P.J., Ju, L., Bieganowski, P., Brenner, C., Moreira, M-C., Taylor, A.M.R. and Caldecott, K.W. *DNA Repair* 3, 1493-1502 (2004).
- Dale A. Ramsden and Martin Gellert. Ku protein stimulates DNA end joining by mammalian DNA ligases: a direct role for Ku in repair of DNA double-strand breaks. *The EMBO Journal*. 17, 609 - 614 (1998).
- Daley JM, et al. DNA joint dependence of pol X family polymerase action in nonhomologous end joining. *J Biol Chem*. 280(32):29030-7(2005).
- Date, H., Onodera, O., Tanaka, H., Iwabuchi, K., Uekawa, K. Igarashi, S., Koike, R., Hiroi, T.,

- Yuasa, T., Awaya, Y., Sakai, T., Takahashi, T., Nagamoto, H., Sekijima, Y., Kawachi, I., Takiyama, Y., Nishizawa, M., Fukuhara, N., Saito, K., Sugano, S. and Tsuji, S. *Nature Genetics* 29, 184-188 (2001).
- Date, H., Igarashi, S., Sano, Y., Takahashi, T., Takahashi, T., Takano, H., Tsuji, S., Nishizawa, M. and Onodera, O. *Biochemical and Biophysical Research Communication* 325, 1279-1285 (2004).
- Davies, D.R., Interthal, H., Champoux, J.J., and Hol, W.G.J. Insights into substrate binding and catalytic mechanism of human tyrosyl-DNA phosphodiesterase (TDP1) from vanadate and tungstate-inhibited structures. *J. Mol. Biol.* 324, 917–932. (2002).
- Dinner, A.R., Blackburn, G.M. & Karplus, M. Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature* 413, 752-755 (2001).
- Dopeso, H. *et al.* Aprataxin tumor levels predict response of colorectal cancer patients to irinotecan-based treatment. *Clin. Cancer Res.* 16, 2375–2382 (2010).
- Downs, Jessica A., Nussenzweig, Michel, C. & Nussenzweig, André. Chromatin dynamics and the preservation of genetic information. *Nature* 447, 951-958 (2007).
- El-Khamisy, S. F. *et al.*, Defective DNA single-strand break repair in spinocerebellar ataxia with axonal neuropathy-1. *Nature* 434, 108–113 (2005).
- Ferrarini, M., Squintani, G., Cavallaro, T., Ferrari, S., Rizzuto, N. and Fabrizi, G.M. *Journal of the Neurological Sciences* 260, 219-224. (2007).
- Gueven, N., Becherel, O.J., Kijas, A.W., Chen, P., Howe, O., Rudolph, J.H., Gatti, R., Date, H., Onodera, O., Taucher-Scholz, G. and Lavin, M.F. *Human Molecular Genetics* 13(10), 1081-1093. (2004).
- Harris, J.L. *et al.* Aprataxin, poly-ADP ribose polymerase 1 (PARP-1) and apurinic endonuclease 1 (APE1) function together to protect the genome against oxidative damage. *Hum. Mol. Genet.* 18, 4102–4117 (2009).
- Hirano, M. *et al.* DNA single-strand break repair is impaired in aprataxin-related ataxia. *Ann. Neurol.* 61, 162–174 (2007).
- Interthal, H., Pouliot, J.J., and Champoux, J.J. The tyrosyl- DNA phosphodiesterase TDP1 is a member of the phospholipase superfamily. *Proc. Natl. Acad. Sci. USA* 98, 12009–12014. (2001).
- Interthal, H., Chen, H.J., Kehl-Fie, T.E., Zotzmann, J., Leppard, J.B., and Champoux, J.J. SCAN1 mutant TDP1 accumulates the enzyme-DNA intermediate and causes camptothecin hypersensitivity. *EMBO J.* 24, 2224–2233. (2005).

- Jackson SP and Bartek J. The DNA damage response in human biology and disease. *Nature* 461, 1071-1078 (2009).
- Junop, Murray S., Modesti, Mauro, Guarné, Alba, Ghirlando, Rodolfo, Gellert, Martin and Yang Wei Crystal structure of the Xrcc4 DNA repair protein and implications for end joining. *The EMBO Journal* 19, 5962 - 5970 (2000).
- Kao, J., Rosenstein, B.S., Peters, S., Milano, M.T., Kron, S. J. Cellular Response to DNA Damage., *Ann. N.Y. Acad. Sci.* 1066, 243-258 (2005).
- Kulkarni, A., McNeill, D.R., Gleichmann, M., Mattson, M.P. and Wilson III, D. M. *Nucleic Acids Research* 36(15), 5111-5121 (2008).
- Lima, C.D., Klein, M.G. & Hendrickson, W.A. Structure-based analysis of catalysis and substrate definition in the HIT protein family. *Science* 278, 286–290 (1997).
- Lima, C.D. *et al.* MAD analysis of FHIT, a putative human tumor suppressor from the HIT protein family. *Structure* 5, 763–774 (1997).
- F L Lin, K Sperle, N Sternberg. Intermolecular recombination between DNAs introduced in mouse L cells is mediated by a nonconservative pathway that leads to crossover products. *Mol. Cell Biol.* 10, 103-112. (1990).
- Lindahl, T. & Barnes, D. E. Repair of endogenous DNA damage. *Cold Spring Harb. Symp. Quant. Biol.* 65, 127–134 (2000).
- Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky SL, Darnell J. *Molecular Biology of the Cell*, p963. WH Freeman: New York, NY. 5th ed. (2004).
- Ma Y, Lu H, Schwarz K, Lieber MR. Repair of double-strand DNA breaks by the human nonhomologous DNA end joining pathway: the iterative processing model. *Cell Cycle.* (2005).
- Modesti, M. Junop, M.S., Girlando, R., Van der Rakt, M., Gellert, M., Yang, W. and Kanaar, R. Tetramerization and DNA ligase IV interaction of the DNA double-strand break repair protein XRCC4 are mutually exclusive. *J. Mol. Biol.* 334, 215-228. (2003).
- Modesti M, Hesse JE, Gellert M. DNA binding of Xrcc4 protein is associated with V(D)J recombination but not with stimulation of DNA ligase IV activity. *EMBO J.* 18 (7) (1999).
- Moreira, M-C., Barbot, C., Tachi, N., Kozuka, N., Uchida, E., Gibson, T., Mendonca, P., Costa, M., Barros, J., Yanagisawa, T., Watanabe, M., Ikeda, Y., Aoki, M., Nagata, T., Coutinho, P. Sequeiros, J. and Koenig, M. The gene mutated in ataxia-ocular apraxia 1 encodes the new HIT/Znfinger protein aprataxin. *Nature Genetics* 29, 189-193 (2001).

- Moreira, M-C., Taylor, A.M.R. and Caldecott, K.W. The ataxia-oculomotor apraxia 1 gene product has a role distinct from ATM and interacts with the DNA strand break repair proteins Xrcc1 and Xrcc4. *DNA Repair*, 1493-1502 (2004).
- Mosesso, P. *et al.* The novel human gene aprataxin is directly involved in DNA single-strand break repair. *Cell. Mol. Life Sci.* 62, 485–491 (2005).
- Pouliot, J. J., Yao, K. C., Robertson, C. A. & Nash, H. A. Yeast gene for a Tyr-DNA phosphodiesterase that repairs topoisomerase I complexes. *Science* 286, 552–555 (1999).
- Quinzii, C.M. *et al.* Coenzyme Q deficiency and cerebellar ataxia associated with an aprataxin mutation. *Neurology* 64, 539–541 (2005).
- Quioco, F.A., Roa, B.B., Nakagawa, M., et al., Mutation of *TDPI*, encoding a topoisomerase I-dependent DNA damage repair enzyme, in spinocerebellar ataxia with axonal neuropathy. *Nat. Genet.* 32, 267–2 (2002).
- Rass, U, Ahel, I., and West, C. Actions of Aprataxin in multiple DNA repair pathways. *Journal of Biological Chemistry.* 282 (13), 9469-9474. (2007).
- Rass, U., Ahel, I. and West, C. Defective DNA repair and neurodegenerative disease. *Cell* 130, 991-1001 (2007).
- Takahashi, T., Tada, M., Igarashi, S., Koyama, A., Date, H., Yokoseki, A., Shiga, A., Yoshida, Y., Tasuji, S., Nishizawa, M. and Onodera, O. *Nucleic Acids Research* 35(11), 3797-3809 (2007).
- Takashima, H., Boerkoel, C.E., John, J., Saifi, G.M., Salih, M.A.M., Armstrong, D., Mao, Y., Quioco, F.A., Roa, B.B., Nakagawa, M., et al., Mutation of *TDPI*, encoding a topoisomerase I-dependent DNA damage repair enzyme, in spinocerebellar ataxia with axonal neuropathy. *Nat. Genet.* 32, 267–27 (2002).
- Takiyama, Y., Nishizawa, M., Fukuhara, N., Saito, K., Sugano, S. and Tsuji, S. Early onset ataxia with ocular motor apraxia and hypoalbuminemia is caused by mutations in a new HIT superfamily gene. *Nature Genetics* 29, 184-188 (2001).
- Tranchant, C., Fleury, M., Moreira, M.C., Koenig, M. & Warter, J.M. Phenotypic variability of aprataxin gene mutations. *Neurology* 60, 868–870 (2003).
- Walker JR, Corpina RA, Goldberg J/ Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* 412 (6847): 607–14. (2001).
- Wang H, Zeng ZC, Perrault AR, Cheng X, Qin W, Iliakis G. Genetic evidence for the

involvement of DNA ligase IV in the DNA-PK-dependent pathway of non-homologous end joining in mammalian cells. *Nucleic Acids Res.* 15;29(8):1653-60 (2001).

Wilson, S.H. & Kunkel, T.A. Passing the baton in base excision repair. *Nat. Struct. Biol.* 7, 176–178 (2000).

Yoon, G., Westmacott, R., MacMillan, L., Quercia, N., Koutsou, P., Georghiou, A., Christodoulou, K and Banwell, B. J. *Neurol Neurosurg Psychiatry* 79(2), 234-23 (2008).