2012

Structure Function Analysis of Adenosine Deaminase Acting on tRNA (ADAT2) from Trypanosoma Brucei

Miller Marshall

STRUCTURE FUNCTION ANALYSIS OF ADENOSINE DEAMINASE ACTING ON TRNA (ADAT2) FROM *TRYPANOSOMA BRUCEI*

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Marshall Miller

June 2012

STRUCTURE FUNCTION ANALYSIS OF ADENOSINE DEAMINASE ACTING ON TRNA (ADAT2) FROM *TRYPANOSOMA BRUCEI*

Marshall Miller, Ph.D.

The Rockefeller University 2012

tRNA is edited at the wobble position of the anticodon to accommodate the degeneracy of the genetic code. The enzymes responsible for deamination of adenosine 34 to inosine are found in all organisms and in all but a few cases are essential for life. In bacteria the reaction is carried out by the homodimeric adenosine deaminase acting on tRNA (ADAT) and modifies only $tRNA^{arg}$, whereas in eukarya the enzyme is the ADAT2/3 heterodimer, which performs the same reaction on all tRNA molecules with adenosine at position 34. Despite a similar evolutionary origin, the characteristics of these two enzymes are significantly different. Substrate binding and catalysis occur within the active site of the bacterial homologue, which can act on both stem-loop substrate analogues and full-length tRNA. Eukaryotic heterodimers function differently with separate substrate recognition and enzymatic components and are only active on full-length tRNA. Furthermore, in trypanosomatids, the ADAT2/3 complex has been shown to also exhibit C-to-U deamination *in vitro*, and the ADAT2 protein alone has been shown to play a role in C-to-U editing *in vivo* at the C32 position of tRNA. The basis for these extra enzymatic activities is unknown. ADAT2 homodimers from trypanosomes and humans have been shown to be stable *in vitro* but these homodimers show dramatically decreased tRNA binding and have no detectable enzymatic activity toward the typical tRNA substrates. While crystallographic data exist for the human ADAT2, the structural underpinnings for both the lack of canonical enzymatic activity and the acquisition of additional

functionality in the trypanosomatid enzymes remain unanswered. In the presented work xray crystallography was used to determine the structure of ADAT2 from *T. brucei*. The solution of the structure was technical, requiring several different crystallographic techniques at each step of the process. These are described in detail. The final structure revealed an enzyme harboring the canonical cytidine deaminase fold and an otherwise normal functional-appearance with no obvious indications as to why this enzyme is inactive. Potential reasons for the loss of substrate binding were indicated by mutations in the regions of ADAT2 analogous to the bacterial enzyme. Further analysis identified three loop regions that are unique to trypanosomatid ADAT2 and are located in regions of structural significance so as to suggest a mechanism for an altered substrate profile for this orphan enzyme.

To everyone that has helped me on my way. Without you none of this would have been possible.

Acknowledgements

This project would not have been possible were it not for the initial protein purification and crystallization screening by Pete Stavropolous working with Nina Papavasiliou and for the extensive biochemical characterization of ADAT2 by the lab of Juan Alfonzo at Ohio State University. X-ray diffraction was helped significantly by Deena Oren at Rockefeller and Wuxian Shi on beamline X29 at Brookhaven. In addition I would like to extend my appreciation to the members of the Stebbins lab for the support and assistance through my years working with them.

Table of Contents

List of Figures

Figure ** Indicates figures for which stero views are provide in appendix I.

List of Tables

Chapter 1 Introduction

The severe impact of *Trypanosoma spp*. on Sub-Saharan Africa and South America cannot be overstated. Trypanosomiases are serious diseases with significant morbidity and mortality not only to humans but also to their domesticated animals resulting in devastating health and economic consequences.[1, 2] These concerns in addition to the complex and fascinating virulence mechanisms exhibited by these parasites have interested researchers in many fields of biology, medicine and even politics for more than a century.[3]

T. bruceii contain two subspecies that cause African trypanosomiasis (commonly, sleeping sickness) in humans: an acute form by *T. b. rhodesiense* in eastern and southern Africa, and a chronic form by *T. b. gambiense* in western and central regions of the continent. The African Union has been joined by the World Health Organization in its efforts to eradicate African trypanomiasis through a variety of techniques, but the complexity of the natural reservoir of these parasites combined with the geopolitical challenges of the region do not foretell complete eradication.[4] Similar efforts have been underway in Central and South America to control *T. cruzii*, which causes American trypanosomiasis (commonly Chagas disease), a chronic disease that if left untreated leads to life threatening gastro-intestinal and/or cardiac conditions.[5] Despite the efforts to control the parasite and its vectors, the primary tool for the control and management of these diseases remains chemotherapeutic treatment, yet the development of new compounds beyond the basic science level has seen a dearth of both interest and funding, something that the Drugs for Neglected Diseases Initiative of the Gates Foundation aims

to specifically address.[6] Regardless, the search for new and more narrowly defined drug targets remains a significant motivator for the continued study of these parasites.

In addition to the health and economic concerns, *Trypanosoma* offer well-suited model systems to explore many questions at the vanguard of immunology, cell-, and molecular biology. While unique and highly adapted to several very specific biological niches, these organisms share a significant amount of cellular machinery with higher eukarya. Modulation of host immunity, the gene regulation of antigenic variation, epigenetic regulation, and post transcriptional gene regulation are all examples of recent reviews highlighting active areas of research in the Trypanosome field that involve or pertain to a broad swath of the higher eukarya.[7-9] Now with the complete genome sequence of *T. brucei*, even more tools will be available for application to these fields of study.[10]

Trypanosoma have been an area of intense interest to those studying gene regulation and RNA editing as these organisms harbor several examples of very unique biology. *Trypanosoma* belong to the class *Kinetoplastida*, so named for the kinetoplast, a large collection of circular DNA found within the single mitochondrion present in these organisms. The kinetoplast genome contains roughly a dozen large circular DNA molecules called maxi-circles harboring the typical genetic information of *mt*DNA and also contain thousands of interlocking mini-circles harboring a variety of genetic elements. The field of RNA editing was founded with the discovery in 1986 that a kinetoplast-derived mRNA transcript for CoxII from *T. brucei* was modified post-transcriptionally to generate a frame shift and eliminate an early stop codon.[11] The mechanism for this type of editing was discovered a few years later when Sturm and Simpson identified small genes they termed guide RNAs on the mini-circles of

kinetoplasts which formed templates by which the mRNA of the maxi-circles were subsequently edited.[12]

Within the nuclear genome, *Kinetoplastida* also exhibit a thymidine modification that is unique to *Kinetoplastida* and *Euglena*, which has been shown to exhibit several hallmarks of epigenetic regulation.[13] This hypermodified thymidine – β-d-glucopyranosyloxymethyluracil – is known as base J and has been identified primarily in telomeric regions across *Kinetoplastida* genomes. In varying degrees base J has also been found in sub-telomeric and/or repetitive DNA sequences in some genera of the class.[14] The formation of base J is a conserved two-step process putatively involving J Binding Proteins 1 and 2 (JBP1/2) and an unidentified glycosyl transferase (GT); however, a conserved mechanism of function for base J remains elusive as have the rules governing initial base J modification, maintenance, and elimination . Data showing both an increase in base J content during the mammalian-host phase of the *Trypanosoma* life cycle and an association of base J content with the variable surface glycoprotein (VSG) repeats in the *Trypanosoma* genome – a family of genes that are variably regulated to affect host immune evasion – have generated significant interest in the possibility that base J in part regulates surface antigen display and variation; however, JBP1-null/JBP2-null mutations in *T. brucei* did not effect viability in either the insect or the mammalian phase of infection.[15] Opposite results wherein JBP1-null mutations are lethal in *Leishmania,* where more than 98% of base J is telomeric, further complicating the a unified function for base J in *Kinetoplastida*.[16, 17] A possibility remains that parasite-specific chemotherapy targeting this pathway could prove effective, particularly in the case of *Leishmania*; however, significant advances in the field will be necessary for such

an approach to become viable. Specifically the regulation of the base J modification pathway is still poorly understood and the role of other factors in addition to JBP1 and 2 cannot be eliminated. While the absence of base J may not compromise Trypanosomatids, the disregulation of the pathway could prove an effective chemotheraputic approach

The genetic organization of trypanosomatids at the chromosome level in the nucleus also highlights the degree to which these organisms differ from other eukarya with regard to gene regulation. A significant portion of the *Trypanosoma* genome is transcribed in polycistronic operons that contain multiple genes belonging to unrelated metabolic and cellular functions.[18] This type of organization then requires a novel type of gene regulation that takes place to a large degree at the level of RNA processing and editing, making *trypanosoma* and other *kinetoplastida* ideal organisms in which to study the process of RNA editing.

There are many ways in which RNA is modified post-transcriptionally. Many of these modifications are best classified as RNA processing and include events such as splicing, capping and polyadenylation. In contrast, RNA editing is the modification of the genetic code through selective enzymatic modifications to specific nucleotides resulting in an RNA molecule that is not just rearranged but actually different from the DNA sequence from which it came. A fascinating example of this type of editing occurs in the anticodon stem-loop domain of tRNA and will be the focus of the discussion to follow.

The role of inosine in the wobble position of the tRNA anticodon was proposed more than 45 years ago by Crick to explain the degeneracy of the genetic code.[19] tRNA

molecules with an inosine in this position are able to translate codons ending in U, A, or C. (Figure 1.1 B) This strategy provides a significant portion of the solution to the problem of translating the 61 codons of the genetic code using no more than 40 tRNA genes, and is essential for life in both prokarya and eukarya.[20]

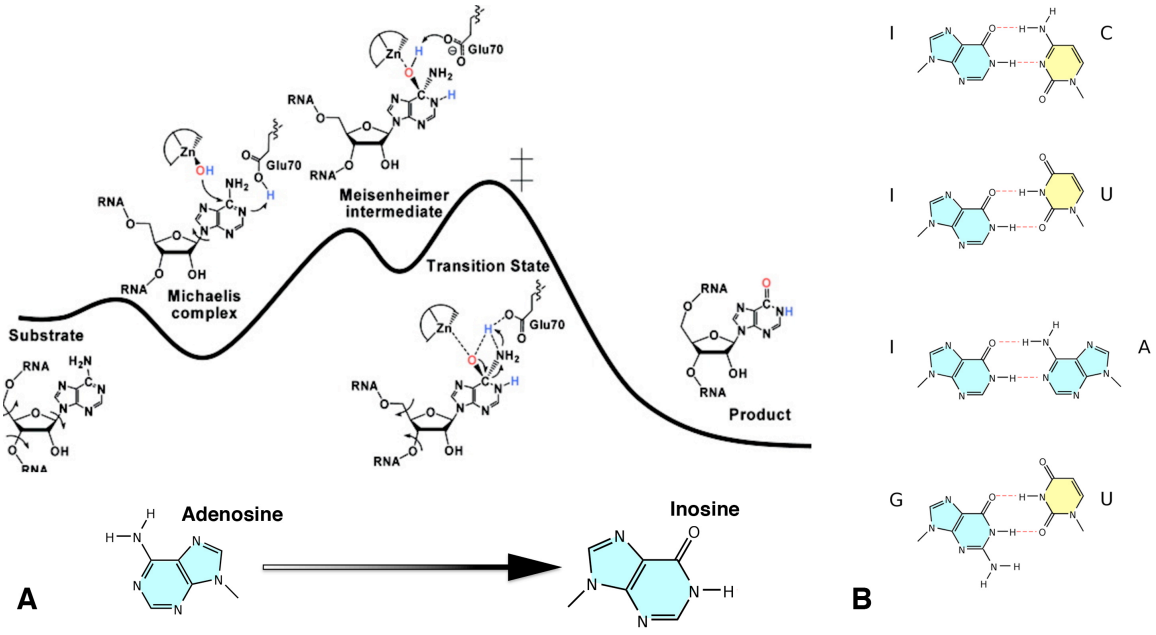


Figure 1.1 The generation and wobble base pairings of inosine in tRNA. **A**. Reaction mechanism for deamination of adenosine to inosine by ADAT from Luo et al.[21] **B**. Wobble base pairing showing the ability of inosine to pair with adenosine, uracil, or cytosine the additional ability of guanine to pair with uracil

Adenosine Deaminases Acting on tRNA (ADAT) in bacteria and plants edit A-to-I at the wobble position (position 34 of the anticodon) only in $tRNA^{Arg}$ (these proteins are also know as tRNA Adenosine Deaminase Arginine, or TADA) whereas in higher eukaryotes this editing occurs in all eight codons (seven in yeast) with an adenosine at position 34 of the anticodon.[22] In bacteria and plants, TADA is active as a homodimer; however, in higher eukaryotes a heterodimer of ADAT2 and ADAT3 is necessary for activity.[23, 24]

Gerber *et al*. have shown that a sequence analysis of these proteins suggests an evolutionary relationship such that ADAT 2 and 3 are paralogues of the original TADA genes [23], which later evolved the necessary mechanisms to accommodate multiple tRNA substrates. Further evidence for this argument is provided by the fact that original studies on yeast ADAT2/3 showed activity toward *E. coli* tRNA, but the converse was not true: *E. coli* ADAT would only modify tRNA with the arginine anticodon ACG.[25] The evolutionary relationship between these enzymes is of significant interest as it suggests a relationship wherein ADAT2 and ADAT3 found in lower eukaryotes could provide significant insight into the origin and function of cytidine deaminases in higher eukarya. Activation Induced Cytodine Deaminase (AID) and the Apo-lipoprotein B mRNA-editing enzyme catalytic polypeptide (APOBEC) family of enzymes are members of the cytidine deaminase family that have been intensely studied for their roles in epigenetic regulation and DNA repair in higher vertebrates; however, a structural understanding of their function is still incomplete and awaits they type of detail that has been elucidated in the prokaryotic enzymes and is beginning to be established for ADAT2/3 in lower eukarya.

These proteins all belong to the large Cytidine Deaminase (CD) superfamily of enzymes. Members of this family all perform similar deamination chemistry on a variety of substrates including nucleotides, nucleosides, small-molecule metabolites, tRNA, dsRNA, and DNA. The signature sequence of the superfamily, H/C-X-E…PCXXC, contains three resides that coordinate a tetrahedral zinc ion (two cysteines and a histidine, or three cysteines). The fourth ligand to the zinc is an activated water molecule that performs the nucleophilic attack on the carbon atom that is being deaminated. Proton

extraction by the glutamate in the above motif results in oxidation of the hydroxyl and release of ammonia. (Fig. 1.1 A)

The conserved core fold of the cytidine deaminase superfamily has been identified by sequence and structure analysis. The fold consists of a 5-strand β-sheet core sandwiched on either side by α-helix bundles – a two-strand bundle on one side, and a three-strand bundle on the other. The three-strand α-helix bundle contains the residues of the zinc binding motif as well as a majority of the residues involved in the most commonly seen dimerization interface. The quaternary structure varies across the superfamily, and a variety of subunit conformations have been shown or proposed including putative monomers, homo-dimers, homo-trimers, tetramers, and as in the case of ADAT2/3, heterodimers. The ADAT enzymes at the core of this discussion have all been shown to exist as either homo- or hetero-dimers with an interface created by the same helicies that contain the active site zinc as described above. (Figure 1.2 A-B)

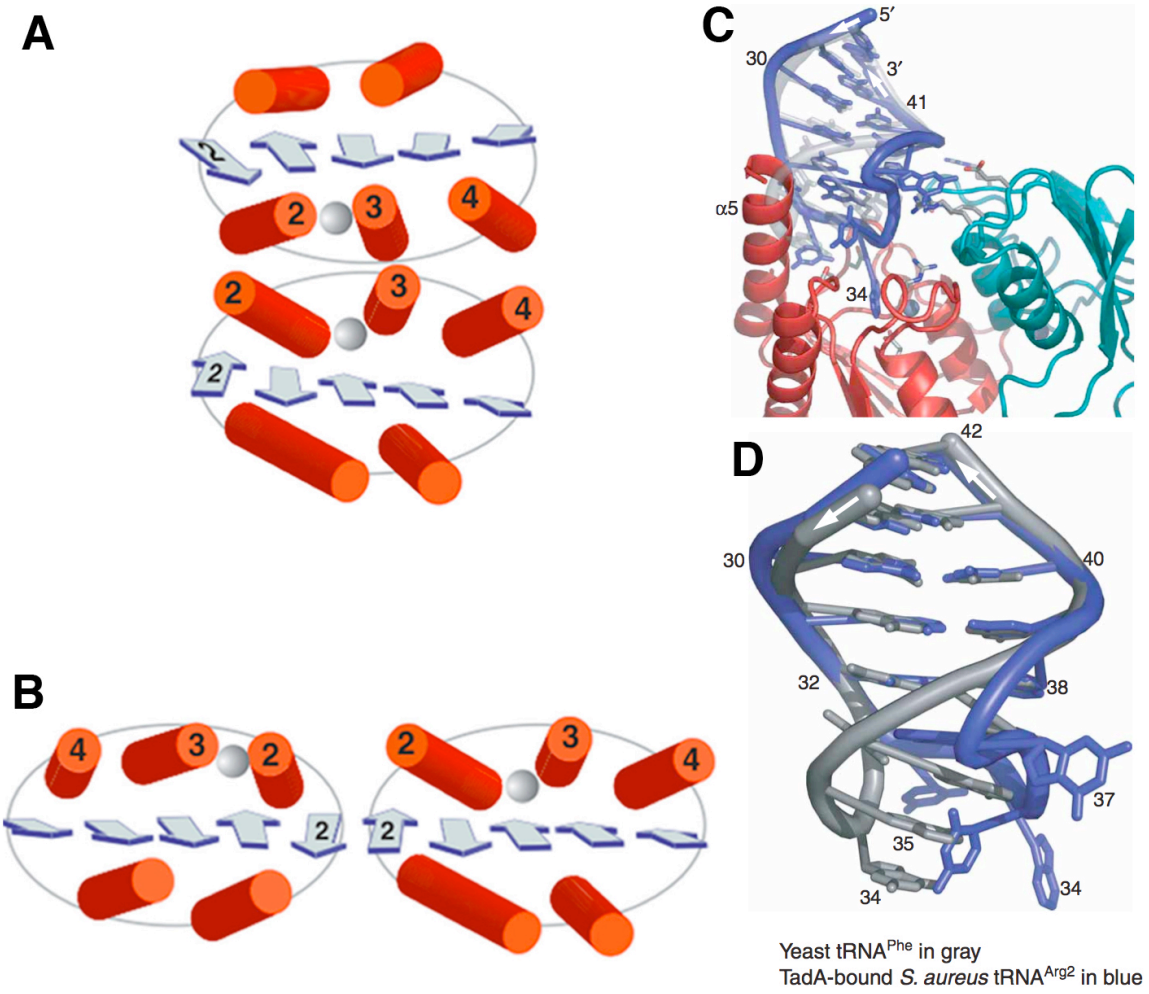


Figure 1.2 The core cytidine deaminase fold as exemplified in the ADAT family of enzymes. **A**. Topology of the α-helices (red) and β-sheets (blue), dimerization interface, and zinc coordination for ADAT **B**. Alternate orientation that has been described for APOBEC2. **C**. Active site of saADAT2 (monomers A and B are red and cyan) bound to tRNA substrate analogue (blue). **D**. Alignment of the tRNA anti-codon step loops in the native (gray) and saADAT-bound (blue) conformations showing the splaying out of the bases induced by substrate binding, specifically at bases 34-37 of the anticodon loop. Figures A and B from Silvestro *et al*.[26], C and D from Losey *et al*.[27]

Multiple crystal structures of bacterial ADAT homologues have been reported, as has the structure of the human ADAT2 homodimer (hsADAT2) all at medium to high resolution.[27-32] Structural and sequence alignments of these homologues show a very

highly conserved core fold with r.m.s. values between 1.2-1.5Å. All structures to date show a consistent quaternary structure comprising a homodimer with each monomer related to the other by a rotational axis of symmetry parallel to the dimer interface. This orientation generates two equivalent active sites on opposite faces of the protein with each monomer contributing to each site. The structural basis for tRNA binding in bacterial ADAT has been demonstrated via the co-crystalization of *S. aureus* TadA (saADAT2) with a non-hydrolyzable step-loop tRNA substrate analogue.[27] (Figure 1.2 C) This structure demonstrates how $tRNA^{arg}$ recognition is accomplished primarily through sequence-specific interactions that clearly discriminate for the $tRNA^{arg}$ anticodon sequence and seems to rely very little on non-specific structural elements within the tRNA molecule. tRNA binding to saADAT2 induces a conformational change in the tRNA molecule such that the anticodon bases are splayed out facilitating sequence-specific interactions with highly conserved saADAT2 residues, but interestingly, there is only minimal structural change in saADAT induced upon substrate binding compared to other bacterial homologues. (Figure 1.2 D) Interestingly, the structure suggests that the histidine participating in the zinc-coordination is also critical in providing important sequence specific contacts to A34. The structure highlights the fact that few non-specific interactions occur between the protein and the stem-loop RNA and suggests that the predominant characteristics of substrate binding in bacterial ADATs relate to the anticodon loop and are sequence specific. Additional analysis of the substrate-bound saADAT2 reveals two equivalent binding sites per dimer with very minimal change relative to other structures of bacterial ADAT without substrate, which suggests this enzyme lacks significant allosteric or coopertive substrate regulation. Given the nature

of the stem-loop used for crystallization in the saADAT structure, the conclusions drawn from this structural analysis are preliminary and await further data with full-length substrate. Of particular interest is the degree to which the remainder of the tRNA interacts with ADAT, how exactly that interaction is shared between the two monomers of the functional homodimer, and to what degree those interactions confer any type of specificity.

Of the five ADAT structures to date, none have been able to show ordered density in the most C-terminal region of ADAT, which has been shown biochemically to interact with tRNA in ADAT2/3 and, based on the saADAT-tRNA structure, would be oriented adjacent to the bound tRNA molecule. The precise nature of the interaction between tRNA and these unstructured regions of ADAT is relatively unknown. This interaction has been shown to be critical for activity in both prokaryotic ADAT and eukaryotic ADAT2/3; however, biochemical data indicate that there are potential differences in the function and importance of this region. Elias *et al*. have demonstrated that truncation of the 8 C-terminal residues in *A. tumefaciens* ADAT (atADAT) decreases enzymatic activity toward both a stem-loop and a native $tRNA^{arg}$ substrate by 2000-fold.[31] Similarly, C-terminal deletions as well as specific point-mutations in a C-terminal lysine-rich domain in the eukaryotic *T. brucei* ADAT2 (tbADAT2) have been shown by Ragone *et al*. to significantly affect binding and catalysis of tRNA substrates in the context of the tbADAT2/3 heterodimer [33]; however these authors also note several differences between tbADAT2/3 substrate binding and that of saADAT. They have noted several non-conservative mutations in residues of the tbADAT2 sequence that align to residues required for $tRNA^{arg}$ binding in the saADAT structure described above, and they have

reported that tRNA binding in tbADAT2/3 is unaffected by single alanine point-mutations to conserved residues in tbADAT2 that align to residues important for $tRNA^{arg}$ binding in the saADAT structure. For example, whereas E92 is makes important contacts to $tRNA^{arg}$ in saADAT, E92A in tbADAT2/3 does not affect substrate binding. Furthermore, in ADAT2/3 from *T. brucei* and *S. cerevisiae* it has been shown that the eukaryotic tRNA deaminases require full-length substrates and cannot catalyze the stem-loop analogues used for much of the biochemical characterization of the bacterial ADAT enzymes [23, 25, 31, 33], which suggests that although critical in both prokaryotic and eukaryotic ADAT, the specific nature of the interaction between the C-terminal tail and the full tRNA molecule also seems altogether different in the two groups of ADAT. These results indicate that both bacterial and eukaryotic ADATs utilize regions of the C-terminus to facilitate binding and/or catalysis, but they also demonstrate how there is a significant divergence in substrate binding behavior both within the active site with regard to sequence-specific substrate binding and outside the active site with regard to the size and nature of the substrate being recognized. These results point to many of the gaps in our understanding of the different mechanisms of action across the family of ADATs and how these differences could account for the narrow versus broad substrate profiles exhibited by the prokaryotic and eukaryotic homologues, respectively. Many questions on this topic remain. Among them, why is the C-terminus important for catalysis – is it substrate affinity alone, or possibly an induced fit mechanism? What is different about narrow versus broad substrate recognition? Are the differences due to a different active site alone or to additional factors outside the active site? As will be discussed below, what is the contribution of ADAT3 and does this account for some of these differences?

The core fold of these enzymes has not been shown to change significantly between the substrate bound and unbound states, and as mentioned above, the c-terminus has not been ordered in any of the homologous structures to date. How this crystallographically disordered region participates in substrate recognition and catalysis remains to be seen. One likely explaination would be a conserved feature within the substrate tRNAs for the eukaryotic heterodimers – either a specific sequence or motif – that could that would prove essential for substrate recognition thus allowing the eukaryotic enzymes to screen a larger set of molecules for a compatible anti-codon sequence on which to perform chemistry. Direct evidence for such a mechanism will require a more detailed understanding of the eukaryotic ADAT2 and ADAT3 enzymes and is one of the motivations for the studies initiated below. To date only the human ADAT2 structure has been reported in the protein data bank and there remains a paucity of structural data for the ADAT3 protein.

As mentioned above, the formation of an ADAT2/3 heterodimer is essential for A-to-I activity in both yeast and trypanosomes and both ADAT2 and ADAT3 are necessary and essential for this activity.[23, 24] The features of this heterodimer in *T. brucei* have been the focus of significant biochemical characterization in by the Alfonzo group at Ohio State University, which has revealed that tbADAT2/3 exhibits exhibits the expected A-to-I editing all tRNA with A34 in the wobble position of the anti-codon, but in addition they have identified a much wider substrate profile in addition to additional enzymatic activity.

The most striking observation thus far is the ability of tbADAT2/3 to perform C-to-U editing on DNA substrates *in vitro* as well as evidence that ADAT2 plays a role in C-to-U

editing on C32 of tRNA *in vivo* [24], which stimulates additional A34-I editing of the anticodon.[34] The ability to deaminate cytosine and adenosine is unique to the trypanosomatid enzymes and has not been reported for any other members of the CD superfamily, raising the possibility that ADAT2/3 represents an evolutionary step between the A-to-I editing enzymes and the C-to-U editing enzymes. In higher eukaryotes the search for demethylases involved in epigenetic regulation has identified Activation Induced cytidine Deaminase (AID) as a potential candidate. The purported mechanism of AID entails methyl-cytidine to uridine deamination, which then prompts removal of the methyl-uridine via mismatch repair resulting in unmethylated C at the same position.[35] The hypothesis that ADAT2/3 is an evolutionary precursor to AID is not supported by a genetic analysis [36]; however the likelihood that both families evolved similar solutions to the same enzymatic problems remains possible. Certainly the relationship between these families will be an area of future interest. Methylases have been identified in *T. brucei* that will allow this question to be probed in much greater detail as it is expected to soon be possible to generate the 3-methyl-cytosine substrates that are essential for epigenetic regulation in higher eukaryotes. Such substrates, whether tRNA or DNA, will then be available for biochemical analysis with ADAT2/3 to more completely understand the link between methylation and deamination that early evidence has indicated. Specifically, it has recently been shown that the C/U32 base is methylated in *T. bruci* (Alfonzo, unpublished). This result together with the evidence that ADAT2 plays a role in C-to-U editing at this position raises the interesting possibility that ADAT2/3 editing is linked to methylation. This result parallels the activity of ADAT1 that has been shown to deaminate A-37 of the anticodon loop, which

is subsequently methylated by an unknown mechanism.[22] Similar to the suggestions above, ADAT1 has also been suggested as an ancestral precursor to the eukaryotic adenosine deaminase acting on RNA (ADAR) family of enzymes.[37] The adaptation of the cytidine deaminase fold to many of the biochemical tasks of cell biology is a broad field and will by no means be sufficiently explained by one or two key crystal structures. Rather it is a field that is advancing as examples of each type enzymatic activity is investigated in relation to each of the various substrates acted upon by these enzymes. A significant appeal, therefore, is found in the ADAT2/3 enzyme in *T. brucei* as it would provide a look at an enzyme that has been shown to be involved in several different enzymatic activites on methylated substrates, which is a field that is being actively investigated but has yet to yield significant structural results.

A striking difference between the bacterial and eukaryotic ADAT enzymes involves the architecture of the active site and the necessity for an ADAT2/3 heterodimer. Secondary structure predictions and computational modeling suggest ADAT3 exhibits the same cytidine deaminase fold as ADAT2, and consequently is predicted to form a dimer with an orientation and interface similar to the prokaryotic homodimer.[38] This orientation raises several questions since the two identical biding sites observed in the bacterial ADATs would no longer be equivalent in the heterodimer, i.e. the zinc binding site of ADAT2 would open toward a surface created in part by ADAT3, and the reciprocal active site in ADAT3 would open towards a surface created in part by ADAT2 that could potentially be completely different. Additionally, the HxE…CxxC motif is mutated in eukaryotic ADAT3 with a Glu-Val mutation at the key catalytic glutamate.[22] Since the histidine and two cysteines required for zinc coordination were present in this sequence,

it was generally accepted that ADAT2/3 heterodimers contained two zinc ions, one in each monomer, and that only the ADAT2 zinc was catalytically active.[23] Spears *et al* have recently called this assumption into question through a series of mutagenesis studies demonstrating that the zinc-coordinating cysteines of ADAT2 were not essential for either substrate binding or activity whereas ADAT2 glutamate and the equivalent ADAT3 cysteines were essential for enzymatic activity. As a potential explaination for this data, the authors have proposed a domain-swapped model in which the HxE motif of ADAT2 combines with the CxxC motif of ADAT3 to form a complete active site.[38] A domain-swapped heterodimer would still exhibit two different active sites, one comprising the essential glutatmate 92 from ADAT2 with the essential zinc coordinating cystines from ADAT3. The precedent of a domain swap in a member of the CD superfamily has been shown in *B. subtilis* guanosine deaminase in which one β-strand and the terminal α-helix of one monomer exchange in a symmetrical fashion with the other protomer of the dimer pair.[39] In the case of ADAT2/3, the proposed swap would be more significant and entail an exchange of 3 β-strands and 2 helicies, essentially one half of the ADAT2 protein. (Figure 1-3 A and B)

The results of Spears *et al*. have modified the understanding of ADAT2/3 in several significant ways: they confirm a role for ADAT3 in catalysis, they demonstrate a 1:1 stoichiometry (tRNA:dimer) as opposed to 2:1 in bacteria, and they provide evidence for an intricate interplay between ADAT2 and ADAT3 to form a shared catalytic zinc site. It is worth noting that these results do seem to conflict to a degree with other results mentioned above regarding the importance the ADAT2 C-terminus. This so-called KR-domain of ADAT2 (for its amino acid sequence KRKRK) has been shown to be essential

for tRNA binding and catalysis [33]; however, an analysis of the domain swap structure would suggest that the C-terminus of ADAT2 in the domain-swapped model is positioned on the opposite side of the dimer from the putative active site. (Figure 1-3 B) How these models will evolve under the light of further scrutiny remains to be seen.

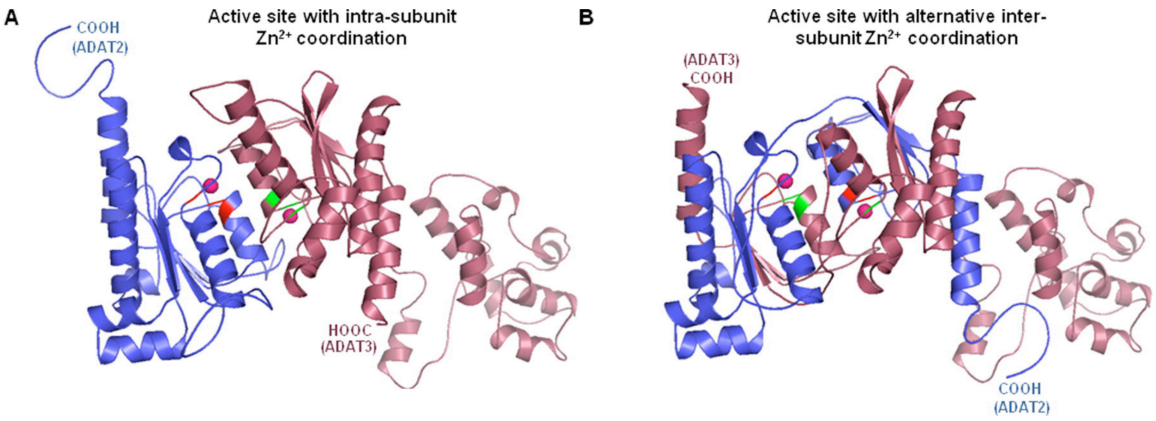


Figure 1-3. Proposed configurations for *T. brucei* ADAT2/ADAT3 heterodimer (blue and maroon, respectively with zinc ions depicted as pink spheres) from Spears *et al*.[40] ADAT2 is modeled from the saADAT crystal structure and ADAT3 is modeled from an *ab initio* simulation. **A**. Non-swapped heterodimer where the HxE and CxxC motifs for ADAT2 (red) and ADAT3 (green) coordinate the zinc ion of the respective monomer. **B**. Domain swapped orientation that generates one zinc-coordination site (on the left in B) containing the HxE of ADAT2 and CxxC of ADAT3, which is accessible from the top of the molecule in this orientation, directly adjacent to the C-terminus of ADAT3.

One final discovery that has been observed along-side the developments in ADAT2/3 biochemistry has been the identification of a stable species *in vitro* of a tbADAT2 homodimer [38], which is consistent with the recent crystal structure of hsADAT2. There has been very little characterization of these eukaryotic ADAT2 homodimers. Based on the hsADAT2 crystal structure (PDB:3DH1), they share a common fold and domain architecture with bacterial ADAT; however, the tbADAT2 sequence shows several

distinct insertions relative to hsADAT2 that have remained uncharacterized and could provide interesting alternate substrate profiles (discussed below).

Binding of tRNA substrates has been reported for tbADAT2 homodimer that is one to two orders of magnitude lower than for the tbADAT2/3 heterodimer; however, no *in vitro* A-to-I activity has been detected in tRNA substrates by tbADAT2 in the absence of ADAT3, despite extensive incubations under saturating conditions.[38] This result has prompted speculation that inactivity is not due simply to low affinity but may involve structural reorganizations involving, for example, incorrect zinc-coordination in the homodimer and/or occlusion of the active site, which are somehow rescued by ADAT3 through heterodimer formation, perhaps through the domain-swap mechanism discussed above. The presence of a stable tbADAT2 homodimer that is inactive toward tRNA substrates raises many interesting possibilities and provides an important tool with which to study the increased complexity observed when moving from prokaryotes to eukaryotes, particularly in regard to the involvement of ADAT2 in potential methylation pathways in *T. brucei*. Given that ADAT3 rescues the inactivity of the tbADAT2 homodimer toward tRNA, an interesting question is whether there are features of tbADAT2 that have been lost, so to speak, in comparison to bacterial ADAT and what specifically these differences reveal about the rules for substrate binding in the ADAT2/3 heterodimer. An analysis of the human ADAT2 structure reveals a dimer interface nearly identical to those observed in bacterial ADAT suggesting there is no clear impediment for the same tbADAT2 dimer seen *in vitro* to also exist *in vivo*; however, the function of this dimer remains unknown. Perhaps the tbADAT2 homodimer is simply a reservoir for the ADAT2/3 heterodimer and that transition between the homo- and hetero-dimer forms

allows some level of regulation of activity within the cell. Evidence that ADAT2 is involved in C-to-U editing in DNA and also is involved in potential methylation pathways suggests the intriguing possibility that tbADAT2 as a homodimer is not inactive so much as the actual substrate has yet to be identified. In losing activity toward tRNA perhaps tbADAT2 has opened the possibility to act on DNA or other types of RNA. The role of the tbADAT2 homodimer will be important to determine in order to further the understanding of the role this enzyme plays in the biology of *T. brucei* and other higher eukaryotes.

As an initial step to address the questions raised above, the following discussion presents the structure of tbADAT2 by means of x-ray crystallography. Chapters describing the results from the major phases of the project are as follows: 2, materials and methods; 3, protein purification; 4, crystal optimization and data collection; 5, structure determination; and 6, structure analysis, in which the tbADAT2 homodimer reveals several interesting characteristics that show a departure from the typical structural features of known homologues. Chapter 7 will discuss the implications of the finding that the ADAT2 structure reveals no clear mechanism of enzymatic inactivity and could provide several possible mechanisms of alternate substrate specificity.

Chapter 2: Materials and Methods

Protein purification

Recombinant N-terminal 6xHis-tagged tbADAT2 (C117S) was expressed off a modified pET28 (6xHis-3C-AUG) vector in BL21 RIL (Stratagene) cells as follows: 16 x 1.5 L cultures of bacteria were grown at 37°C in Luria Broth to an OD_{600} of 0.5-0.8 and induced at 0.5 mM IPTG for 18 hours at 18°C. Cells were harvested by centrifugation, resuspended in 100-120 mL of resuspension buffer (20 mM HEPES pH 7.6 @4°C, 200 mM NaCL, 20 mM Imidazole, and 20 mM β-mercaptoethanol) and lysed via Emulsiflex (Avestin) until sample was no longer viscous. Cleared lysate (25,000xg) was filtered (0.65 μ) and loaded onto a 14 ml Ni-NTA (Quiagen) column equilibrated in resuspension buffer. The loaded column was washed with 10 column volumes (CV) of resuspension buffer and bound protein was eluted by a linear imidazole gradient (20-500 mM imidazole in resuspension buffer) over 8 CV. Fraction contents were visualized via SDS-PAGE, pooled, and dialyzed overnight with 3C protease in dialysis buffer (20mM HEPES pH 7.6 @4°C, 200 mM NaCl, and 5 mM DTT). Following dialysis, the protein sample was diluted to 100 mM NaCl with Buffer A (20 mM HEPES pH 7.6, 5 mM DTT) and loaded on a 1.0 mL Hi-Trap Heparin Agarose FPLC column (GE Lifesciences) equilibrated in 2.5% buffer B (20 mM HEPES pH 7.6 at 4°C, 2.0M NaCl, 5mM DTT). The loaded column was washed with 10 CV 2.5% buffer B and bound protein was eluted by a linear NaCl gradient (2.5-25% buffer B) over 40CV. Fraction contents were visualized via SDS-PAGE and desired fractions were pooled, spin concentrated (30 kDa Amicon Ultra), and further purified first with Superdex 200 16/60 and then with Superdex 75 16/60 gel filtration (GE Healthcare).

Crystallization

Purified tbADAT2 was concentrated to 100-110 mg/mL by spin concentration (30 kDa Amicon Ultra) and crystallized at 4°C in 2-4 μL drops in 1:1 ratio with a well solution (WS) of 944 mM K/Na tartrate, 200 mM $LiSO_4$, and 100 mM HEPES pH 6.9 via hanging drop vapor diffusion (total well volume 1.0 mL) for 2-3 weeks. Crystals were cryo protected using 20% glycerol or dehydration and flash cooled by plunging into liquid nitrogen (LN2): crystals were incubated in WS containing increasing glycerol concentration (5% increments to 20%, 2 minutes per step); or crystals were dehydrated 12-16 hrs in 10 μl drops comprised 1:1 of WS:WS with saturated with Na/K tartrate (WS-sat) by vapor diffusion over 1.0 ml WS-sat, and then further dehydrated by allowing the drop to evaporate 5-6 minutes (just before the formation of salt crystals) and then flash cooled as above. Crystals were stored in LN2 until data collection.

Data Collection

Crystals were screened for quality assessment and successful cryo-protection using an in-house Raxis-IV++ rotating Cu-anode x-ray source, and complete native and Zn-SAD datasets were collected on beamline X-29 at Brookhaven National Synchrotron Light Source (NSLS) at wavelengths of 1.0750Å (360° by 1° oscillations, 1.0 s exposures,) and 1.2826Å (360° by 1° oscillations, 2.0 s exposures), respectively, utilizing a robotic auto-mounter for high sample throughput and a ADSC-Quantum 315r CCD detector. The wavelength for the Zn-SAD dataset was determined empirically using CHOOCH to analyze the X-ray fluorescence spectra from a representative tbADAT2 crystal.[41] All data were collected at LN2 cold-stream temperatures (~110K).

Structure Determination

Diffraction data were indexed, integrated, and scaled using HKL2000.[42] Asymmetric unit contents were estimated using Matthew's coefficients.[43, 44] Molecular replacement, phasing, and initial model building were performed using AutoMR, Auto-Sol, and AutoBuild, respectively in PHENIX (**P**ython based **H**ierarchical **EN**vironment for **I**ntegrated **X**tallography)[45], followed by rounds of manual model building and refinement in COOT[46] and REFMAC5[47], with additional application of anisotropic scaling[48] and TLS refinement.[49] The final protein data bank (pdb) files resulting from TLS refinement were analyzed with TLSANL[50] to recover thermal B-factors. Molecular replacement search models for tbADAT2 were created with the program CHAINSAW[51] in CCP4[52] using the hsADAT2[32] as a starting model. Additional model building for unstructured loops was attempted with ARP/wARP[53] and fit_loops in PHENIX. Overall structural stastics including rotomer, Cβ-, geometric, and ramachandran analyses were performed during model building in COOT and following refinement using MOLPROBITY[54] and ERRAT.[55] Structural analyses were conducted with SWISS-Pdb Viewer[56], COOT, and CCP4mg[57] as well as using web servers DALI[58], PHYRE[59] and ERRAT[55] for structural comparisons and homology searches, and ESBRI[60] for identifying salt bridges in the structure.

Disulphide cross-linking

The final structure of tbADAT2 was used to model the ADAT2/ADAT3 heterodimer.[38] The webserver Disulphide by Design ®[61] was used to analyze both the tbADAT2 homodimer and ADAT2/3 heterodimer for permissive sites to introduce cysteine point

mutations for the formation of engineered disulfide bonds. These mutations were then modeled using CCP4mg[57] to determine which cysteine pairs would give definitive SDS-PAGE mobility shifts as described by Sorensen *et al*.[62]

Site-directed mutagenesis

	Single AA substitution	**Insert Mutagenesis**	**Sub-cloning/LIC**
Forward Primer	0.5 µl 2.5 pmoles/µl	1.0 µl of 50 µM	1 µl 50 µM
Reverse Primer	0.5 µl 2.5 pmoles/µl	1.0 µl of 50 µM	1 µl 50 µM
dNTP mix	1.25 µl (2.5 mM each dNTP)	5 µl of 2.5 mM	2 µl 25 mM
Buffer (10X)	1.25 µl PFU Buff. (Stratagene)	5 µl PFU Buff. (Stratagene)	10 µl Thermopol (NEB)
Template DNA	1.0 µl (2ng/µl) miniprep DNA	2-10 ng miniprep DNA	50-90ng miniprep DNA
Polymerase*	0.25 µl PFU (Stratagene)	1 µl PFU (Stratagene)	1 µl Deep Vent (NEB)
ddHOH	to 12.5 µl	To 50 µl	to 100 µl for 2 reactions
Incubation (°C)	95° 5’	95° 5’	95° 5’
Melting (°C)	95° 1’15’	95° 1’15’	95° 30”
Annealing (°C)	T_m of Primer - 5°C	T_m of Primer - 5°C	56°
Extension (°C)	68° 1’ + 1’/1kb	68° 1’ + 1’/1kb	72° 1’30”
Cycles	20-22	12-25	25-35
2° Annealing °C	T_m -5°C	T_m -5°	N/A
2° Extension°C	68° 10’	68 30’	N/A

Table 2.1 PCR conditions for single amino-acd (AA) site-directed mutagenesis, restriction endonuclease sub-cloning, and ligation independent cloning (LIC).

*Polymerase added once samples reached 95°C

Site directed mutagenesis was performed using two different protocols: For single amino acid substitutions, fully overlapping primers (25-45 nucleotides) were designed using the PrimerX webserver[63] with T_m values optimized between 68-80°C; for inserts – restriction sites and FLAGtag – primer design was performed as described in Liu *et al*. with primer length not exceeding 45 nucleotides.[64] (Table 2.1)

Polymerase reactions were assayed by 1% agarose gel electrophoresis (80V x 45', constant potential) and incubated with 0.4 U/μl (typically 0.25μl in 10μ PCR reaction) and incubated at 37°C 4-8 hrs. Digested pcr material was used for transformation reactions.

Restriction Endonuclease Sub-cloning.

Generation of sub-clones of a given construct, i.e. N- or C- terminal truncations, were performed by pcr amplification (Table 2.1) with primers harboring the desired restriction endonuclease sequence constructed 5'AAATAATATTGT-Restriction_Site_ABC-Gene_sequence (5-7 codons, ending in G or C)—3'. PCR products (inserts) from these primers were purified with ion-exchange spin columns (Quiagen) and digested in 30-40 μl reaction volumes containing 5-10 U of each restriction endonuclease (NEB) (as with the desired vectors) for 8-16 hrs at 37°C. Digested vector and insert samples were purified by1% agarose gel electrophoresis; desired DNA bands were excised, and purified as above with ion-exchange spin columns. DNA concentrations were quantified using Nanodrop spectrophotometry, and ligation reactions were performed using Quick Ligase (NEB) following the manufactures protocol.

Ligation Independent Cloning.

PCR amplification of desired gene fragments was performed with forward and reverse primers bearing LIC-compatible sequences at the 5' ends. (Table 2.1) PCR products were purified by 1% agarose gel electrophoresis as above and measured for concentration by Nanodrop spectrophotometry. LIC-compatible vectors were digested with the appropriate restriction endonuclease at 37° 8-12 hrs and similarly gel purified. Vector

and insert samples were digested with T4 polymerase in the presence of only dGTP (insert) or dCTP (vector) in 35 µl reactions: 30 µl gel purified vector or insert, 0.8 µl 100 mM dNTP (Fischer Scientific), 0.34 µl 100x BSA (NEB), 3.5µl 10x T4 Polymerase buffer (NEB restriction endonuclease buffer #2), 1 µl T4 DNA polymerase (3000 U/ml, NEB). The reaction was incubated at 24°C for 20' and heat inactivated at 75°C for 30'. Annealing of the digested insert and vector was performed by mixing 30-50 ng of vector with a 5-fold molar excess of digested insert in a 10µl reaction volume (1x NEB restriction endonuclease buffer #2). These reactions incubated 10' at 42°C and allowed to cool to room temperature for 10', after which 1µl of 110 mM EDTA was added (10mM final concentration). This annealed material (2-10 µl) was directly transformed into chemically competent cells.

Competent cell production and transformation

Glycerol stocks of untransformed chemically competent cells were used to inoculate 10-15 ml cultures and grown 12-16 hours at 37°C with 180-200 rpm rotary shaking in Terrific Broth (Novagen) or Luria Broth (EMD) without antibiotic selection. Large cultures (1-1.5 L) were started by inoculation with these overnight cultures and grown as above to an OD_{600} between 0.2-0.3. Cells were harvested (3800g x 5' at 4°C) and re-suspended in 1/3 original culture volume with sterile 100 mM $CaCl_2$ (pre-chilled on ice). Cell suspensions were incubated on ice for 8-12 hrs, harvested as above, re-suspended in 1/100 original culture volume with sterile 100mM $CaCl_2$ with 15% glycerol (pre-chilled on ice), and snap frozen in LN2 in desired aliquots. Cells were stored at -80°C until use. For cell lines harboring the pLysSRARE-Cam^R plasmid, the protocol above was modified

to include the appropriate selection throughout the overnight and culture expansion phases, but not the $CaCl_2$ incubations.

Transformation of chemically competent cells:

Cells were thawed on ice, diluted four-fold with sterile, prechilled 100 mM CaCl2 (typically 20 µl cells: 60 µl CaCl2), and gently mixed with plasmid DNA: 0.2-1 µl miniprep DNA, 2-4 µl site-directed mutagenesis reactions, 2-10 µl ligation/LIC annealing reactions. Transformation reactions were incubated 30' on ice, 2' in 42°C water bath, 5-10' on ice, and mixed with 350-400 µl sterile LB or TB without antibiotic selection and incubated at 37°C with 220 rpm rotary shaking for 1-3 hrs before plating on LB agar plates, pre-warmed and dried 30-60'. For all transformations of DNA from site-directed mutagenesis, ligation-, and annealing reactions, DH5α cell lines were used. For protein expression, BL21 DE3, BL21-AI, and BL21-Rosetta-pLysS strains were used. All prepared as described above.

Alignments

Alignment of DNA and amino acid sequences were performed using PRALINE[65], CLUSTAL[66], T-Coffee[67], and BLAST.[68]

Limited Proteolysis

Samples of desired protein (2-8 mg/mL) were digested with empirically determined concentrations of Subtilisin (usually 0.01-.1% protein:subtilisin) at room temperature and on ice for 15-30 minutes (also an optimized variable). Digests were terminated by adding SDS-PAGE loading dye and boiling the samples for 5 minutes. SDS-PAGE analysis of

proteolysis was performed and transferred to PVFD membranes. Bands of interest were assayed for N-terminal sequencing by the Rockefeller Proteomics Resource Center.

Ni-NTA Purification

All constructs of SopB, YpkA, and Gelsolin were purified similarly: desired constructs were subcloned into modified Duet expression vectors (Novagen) containing a 3C-protease cleavable 6xHis tag and grown under appropriate antibiotic selection in Terrific Broth (Novagen) to an OD600 0.5-08 and induced with 0.05-0.1 mM IPTG and grown at 16-20°C for 36-38 hours. Cells were pelleted by centrifugation at 4°C, resuspended in lysis buffer (4°C)with vigorous stirring (20-40 mM HEPES 7.4, 250 mM NaCl, and 10 mM imidazole) and flash frozen in liquid nitrogen. For purifications, cells were thawed on ice and lysed by Emulsiflex (Avestin) 15-17,000 psi, 2 passes. Lysed material was brought to 0.1 mg/mL DNAseI, 5 mM $MgCl_2$, and 0.1 mM PMSF. Lysates were incubated 30 minutes on ice and centrifuged. Cleared supernates were loaded 3x over Ni-NTA matrix (typically 30 mL Ni-NTA for 6L worth of cell culture) with gravity flow, washed with 10 column volumes of lysis buffer, 10 column volumes of wash buffer (lysis buffer + 30 mM imidazole), and eluted with 4-5 column volumes elution buffer (lysis buffer + 500 mM inidazole). Eluted fractions were analyzed on SDS-PAGE and desired fractions pooled, assayed for protein concentration, and mixed 200:1 protein:His-tagged-3C protease in dialysis tubing, which was dialyzed 16-48 hours with 2-3 changes agains 5L of gel-filtration buffer (20 mM HEPES pH 7.4, 200 mM NaCl, 1 mM DTT). Proteins were then purified from the 3C-cleaved His tag by another pass over Ni-NTA and purified further as desired using gel-filtration and ion exchange chromatography following manufacturer's protocols.

Culturing *Dictyostelium discoidium*

A generic laboratory strain of *D. discoidium* was kindly provided by Larua Macro in the lab of Sandy Simon at Rockefeller University. All stocks were grown axenically and passaged in HL5 medium with glucose (referred to as HL-5g heretofore): HL-5 purchased as a dry mix from Formedium (ref# HLD0102) and prepared as per manufacturer's instructions supplemented with 20 mL 50% sterile glucose and 10 ml Penn/Strep (10,000 Units/mL Penicillin and 10 mg/mL Streptomycin). The techniques for culturing *D. discoidium* have changed very minimally from the original paper by Ashworth and Watts in 1970, and much of the protocol followed for the below experiments can be found therein.[71] Frozen 1.0 mL stocks of *D. discoidium* were thawed quickly by holding the tube in latex-gloved hands until the stock became slushy and partly thawed whereupon the stock was pipetted into 10 mL of room temperature (20-22°C) HL-5g in a sterile 100 mm plastic petri dish. Cells were allowed to adhere to the bottom of the dish 4-8 hrs and the media was decanted and replaced with fresh HL-5g. Cells were monitored for confluence and were split when the dish was 70-80% confluent and cells were seen detaching from the surface. Splitting cells was performed by gentle pipetting of the culture medium across the bottom of the plate to dislodge cells and then adding 0.2-0.5 mL of the resulting cell suspension to a new plate and repeating as above. When culture expansion in liquid phase was necessary, the cell density was determined by manual hemocytometer cell counting methods and dilutions were made such that cell densities were never below $1x10^4$ cells/mL. Liquid cultures were grown at 19-20°C with gentle shaking (100-140 rpm). Cells were harvested by centrifugation, washed in fresh medium, centrifuged again and flash frozen as a pellet. For actin purifications from *D.*

discoidium, cell pellets were thawed on ice and lysed in Emulsiflex (Avestin) on minimal pressure settings (200-500 psi.). Lysate was analyzed by microscope to assess completeness of cell lysis.

Culture of *Acanthamoeba*

Stocks of Acanthamoeba castellani were obtained from the lab of Dyche Mullins at University California San Francisco and grown axenically as per his lab's published protocol.[70] A second stock of *Acanthamoeba castellani* was purchased from the American Type Culture Collection (ATCC). Empiric testing with the recommended culture medium for the ATCC strain, the Mullin's strain, and for the D. discoidium culture discussed above, it was determined that Acanthamoeba grew as well if not better on the HL-5g medium used for Dictyostelium. To simplify lab stocks, all cultures of Acanthamoeba were grown in HL-5g in a manner very similar to that of Dictyostelium. Maintained on plates, expanded in liquid, and grown in large batches when necessary.

Actin Purification

All attempts to purify actin strictly followed established protocol by Zuchero for *Acanthamoeba* actin and Trasak *et al*. for *Dictyostelium* actin, personal communication.[69, 70] Important elements of the *Dictyostelium* protocol are noted here as they are not published, but it should be stressed that the expected results of the published protocols were never reproduced. 10-30 Liters of Dictyostelium discoidium culture were harvested with low speed centrifugation (1500 rpm/10 min), washed with 10 mM Tris pH 8.0 150 mM NaCl twice. Cells were resuspended in lysis buffer (2ml/g wet pellet; 30mM Tris, pH8, 4 mM EGTA, 2mM DTT, 30% sucrose, 5 mM benzamidine, 2mM EDTA, 0.2

mM ATP) and lysed in Emulsiflex (Avestin) ~ 500 psi, 1 pass). Lysate was cleared at 100,000xg for 90 min, pellet washed with lysis buffer, same spin, keeping the cleared lysate, which was adjusted pH to 8.0 and loaded onto equilibrated DE52 (2/3 of the resin in the column (3cm x 50 cm), 1/3 of the resin stirred into the homogenate for 30 min, slurry then loaded onto the fresh DE52 in the column); the reason for this procedure: if one has large amounts of homogenate the loading takes a long time and actomyosin has time to flocculate, to build a considerable layer of fluffy material on top of the resin, and clogg the tubing. A batch procedure was an attempt to avoid this difficulty. After the slurry (dark yellow) is nicely packed, the column is washed with at least two column volumes of 10mM Tris, pH 8.0, 1mM EGTA, 1mM DTT, 0.02% NaN3, 1mM benzamidine, 0.5 mM PMSF, 0.1 mM ATP (called 'TEDABPA' also column equilibration buffer). Elution of bound proteins was performed with a 0-400mM linear NaCl gradient over 10 column volumes. This step was quite difficult as the column would routinely clog and require very slow flow rates (0.1 ml/min), which equated to days of running the column. Actin elutes at about 120-175 mM NaCl. The column profile was checked by SDS PAGE and actin fractions pooled. After a 0-35% ammoniumsulfate precipitation the pellet is well dialyzed (48 hrs with frequent changes at 4°C) against normal G-buffer. The dialysate was ultracentrifuged (100,000 g x 60 min), pellet discarded, and supernatant treated with 2 mM MgCl2, 50mM KCl, 1mM ATP, 2 h, at 4°C. After this step of actin polymerization, the sample was ultracentrifuged at 100,000g for 3h. The glassy pellet was homogenized with a glass homogenizer, and dialyzed thoroughly against G-buffer for depolymerization. Another ultracentrifugation spin was performed to clear

polymerized actin (100,000g x 3h at 4°C). Only the upper 2/3 of the supernatant in the centrifugation tubes were recovered from final ultracentrifugation step to try and obtain actin monomers and avoid multimers that aren't sufficiently large to pellet upon centrifugation.

Chapter 3: Protein Purification and Initial Crystallization

** The work presented below that preceeded initial x-ray data collection was performed by Pete Stavropoulos prior to- and independent of- the author's involvement with this project.

Initial attempts to recombinantly express tbADAT2 (wt) were hampered by low yields and high molecular weight species likely resulting from cysteine cross-linking. Site-directed mutagenesis of individual cystine residues identified C117S, (heretofore called tbADAT2 as opposed to tbADAT2 (wt), as a key mutation that greatly increased expression, solubility, and stability. tbADAT2 was subsequently expressed with a 3C-protease cleaveable N-terminal 6xHistidine tag via a protocol involving an initial Ni-NTA purification followed by 3C digestion, Heparin-Agarose ion-exchange, and two rounds of gel filtration (Fig. 3.1 A-D). Purified tbADAT2 was well-behaved, soluble, and non-aggregating, even at concentrations in excess of 100 mg/ml.

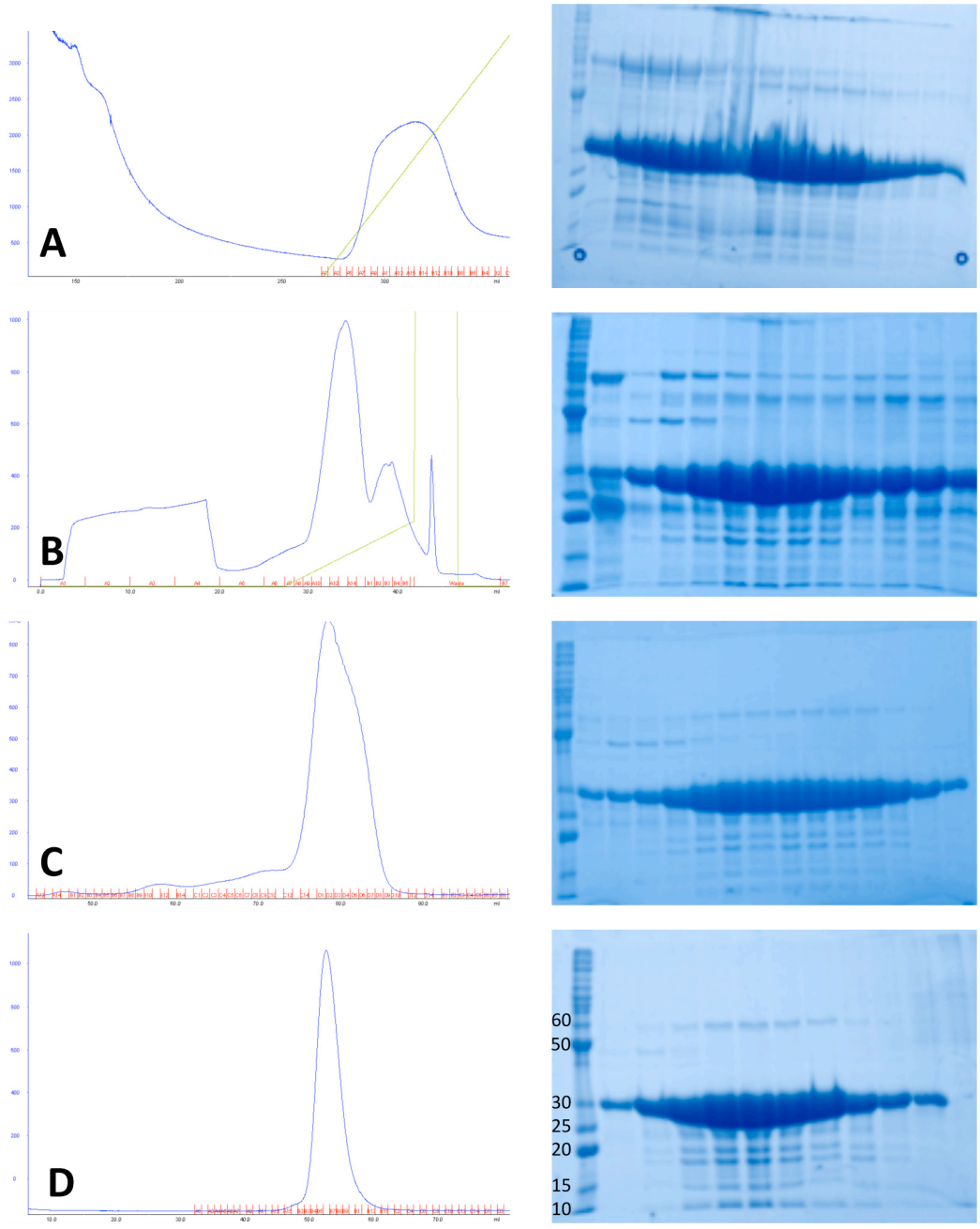


Figure 3.1. Purification steps of tbADAT2 shown as FPLC elution profiles (uv trace in blue, gradient in yellow) and SDS-PAGE analysis of peak fractions, left and right panels, respectively. A. Ni-NTA elution (20-500 mM Imidazole) B. Hi-Trap Heparin-agarose (50-500 mM NaCl) C. Superdex 200 (Elution Peak ~78.0 mL) D. Superdex 75 (Elution Peak 52.0 mL)

Crystallization screens of tbADAT2 at 100-105 mg/ml yielded a positive crystallization condition at 4°C that was optimized within the following range of conditions: 880-1120 mM Na/K tartrate, 200 mM LiSO4, and 100 mM HEPES pH 6.5-7.2. Crystals were elongated square dipyrimids (~0.3 x 0.1 mm) with rounded edges such that they appeared more like an ellipse and lacked a sharp, facetted appearance. (Fig. 3.2)

Figure 3.2 tbADAT2 crystals grown 14 days at 4°C by vapor diffusion over 944 mM Na/K Tartrate, 200 mM LiSO4, 10 mM HEPES pH 6.9. The largest crystals are approximately 0.3 mm in the longest dimension.

Initial tbADAT2 crystals were cryoprotected in 20% glycerol and complete datasets were collected on several samples. The best of these datasets was successfully indexed, integrated, and scaled in space group p41212 with unit cell dimensions 62.74x62.74x228.80, 90,90,90 (see table 1); however, the data had a mosaicity ranging

from 1.1-1.8 (also evidenced by the "smeary" appearance of the diffraction data) and showed strong anisotropic diffraction with resolution limits of 3.6x3.6x3.2 Å [48]. (Fig. 3.3)

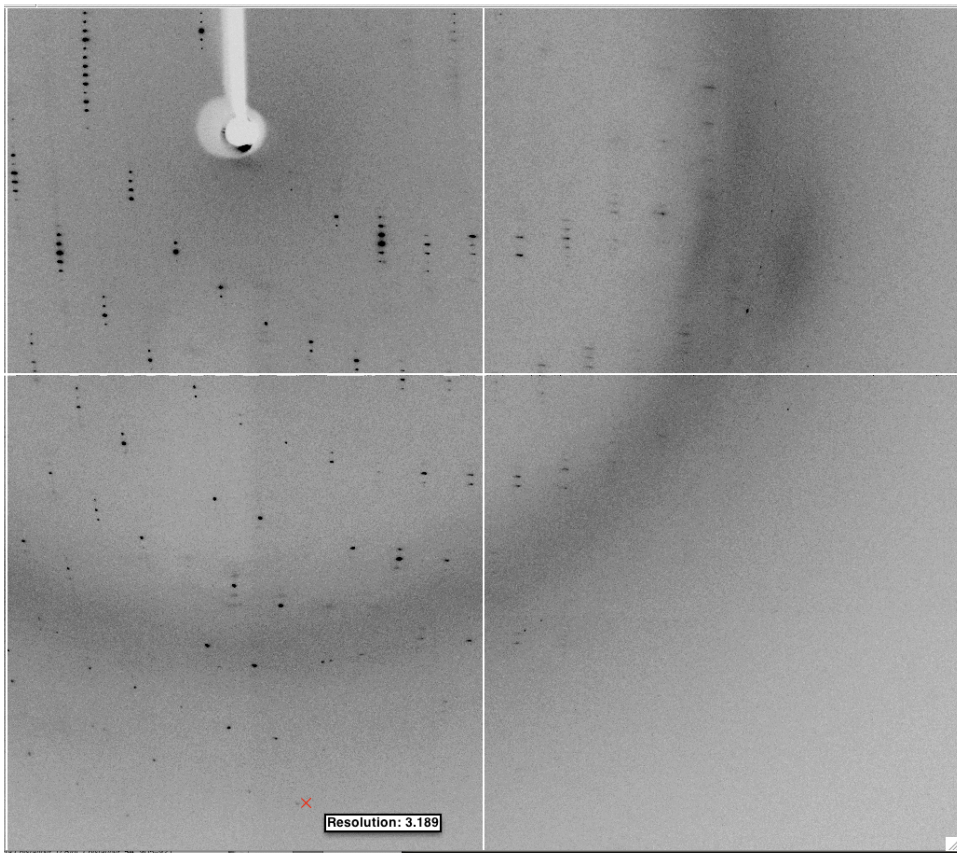


Figure 3.3. Diffraction data for tbADAT2 crystals in 20% glycerol showing roughly one quadrant of the detector field. The resolution of diffraction at the decector edge is ~2.9Å. The best diffraction on this image is ~3.1 and ~3.7Å in the vertical and horizontal axes, respectively.

The Matthew's coefficient for these data suggested an asymmetric unit comprising 2 monomers with a solvent content of 48.5%. Using these parameters and a homologous monomer from the hsADAT2 structure[32] as a search model (36% identity), a molecular replacement solution was obtained from these data with an R-work/R-Free of (35.0/42.0) for an auto-built structure comprising 324 of the total 450 amino acids in the dimer. Maps of this structure revealed general fold information, including interpretable regions

of β-sheet and α-helix, and the density for the coordinated zinc was clear; however, side-chain detail for key residues in this region was minimal if not absent. (Fig. 3.4) In total 46% of residues in this model were categorized as having acceptable density while the remaining residues were categorized as having weak to very weak density by the autobuild module of PHENIX.

Attempts at *de novo* structure determination with Zn-MAD/SAD and with Se-methionine derrivatized tbADAT2 crystals were unsuccessful. Fluorescence scans of the crystals around the Zn-peak for the purpose of anomolous phasing failed to detect a useable signal, and data quality Se-methionine tbADAT2 crystals were not obtained.

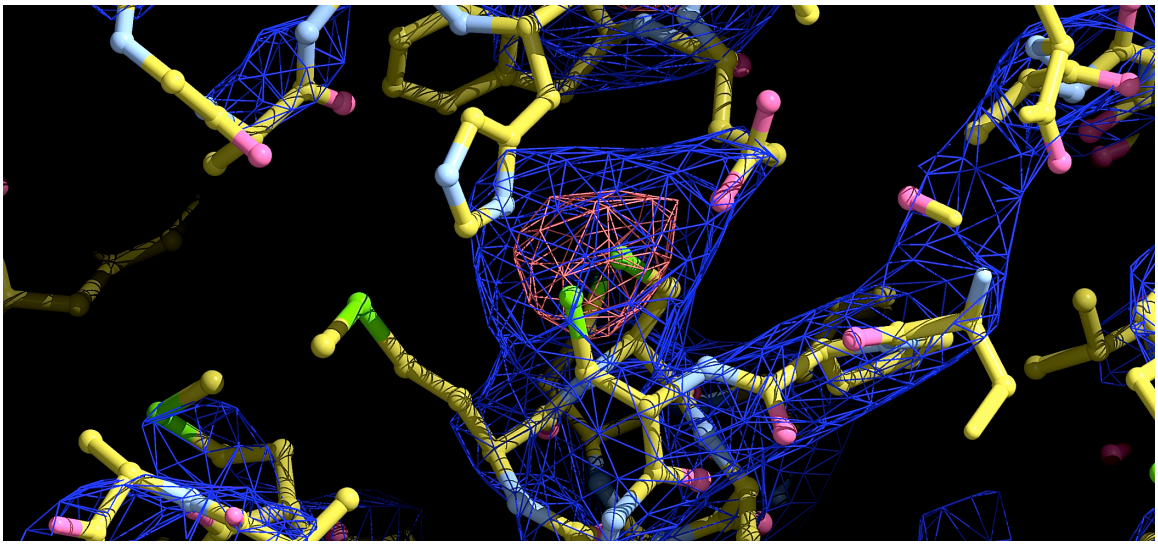


Figure 3.4. View of the Zn-binding motif in the auto-built model of the molecular replacement solution of tbADAT2 with density modified maps contoured at 1.0 sigma (blue) and 5 sigma (red) showing the overall lack of detail in this solution.

Chapter 4 Crystal Optimization, Data Collection, and Processing

Crystallization trials were repeated as above at 4°C using a more detailed grid of Na/K Tartrate concentration (880-960 mM) in 8 mM steps with 200 mM LiSO4 and 100 mM HEPES pH 6.6-6.9 (0.1pH unit steps). Data quality crystals were harvested from all conditions of this grid, with optimal conditions lying between 930-960 mM N/K tartrate and pH 6.8. Below and above these points growth decreased and nucleation increased, respectively. Using these crystals, several cryoprotection strategies were attempted to improve the diffraction data quality. Solutions containing 2-3 M Na acetate and Na malate were tried both alone and in combination with each other in an attempt to find a suitable cryoprotectant.[72] These solutions vitrified well and had very low background diffraction, but no transfer method into these solutions was found for the tbADAT2 crystals that did not deleteriously affect diffraction. An alternative approach to stabilize the crystals before cryoprotection through glutaraldehyde cross-linking[73] also failed to improve upon the diffraction described in Chapter 3. It was noted during cryoprotection trials that using the crystallization well solution as a cryoprotectant did preserve diffraction as well as- or better than- that observed previously, but significant ice-ring diffraction remained problematic. Na/K tartrate can serve as a cryoprotectant at concentrations near 2.0 M, which is very close to its saturation point.[72] Such concentrations were unattainable when considering the dilution factor resulting from the addition of the other reagents in the well solution and the necessity to chill all buffers to 4°C. A dehydration protocol was developed using vapor diffusion wherein the well solution with saturated Na/K tartrate served as a dehydrant followed by an evaporation step of the crystal-containing drop, which drove the Na/K tartrate to saturation. Crystals

were cryo-cooled in L2N directly from the evaporating drops. Dehydrated tbADAT2 crystals showed improved resolution and decreased mosaicity.

Diffraction data for these dehydrated crystals showed diffraction resolution in-house that was nearly equivalent to the resolution glycerol treated crystals had shown using a synchrotron x-ray source. Additionally, these data showed decreased anisotropy and decreased mosaicity. (Fig. 4.1A)

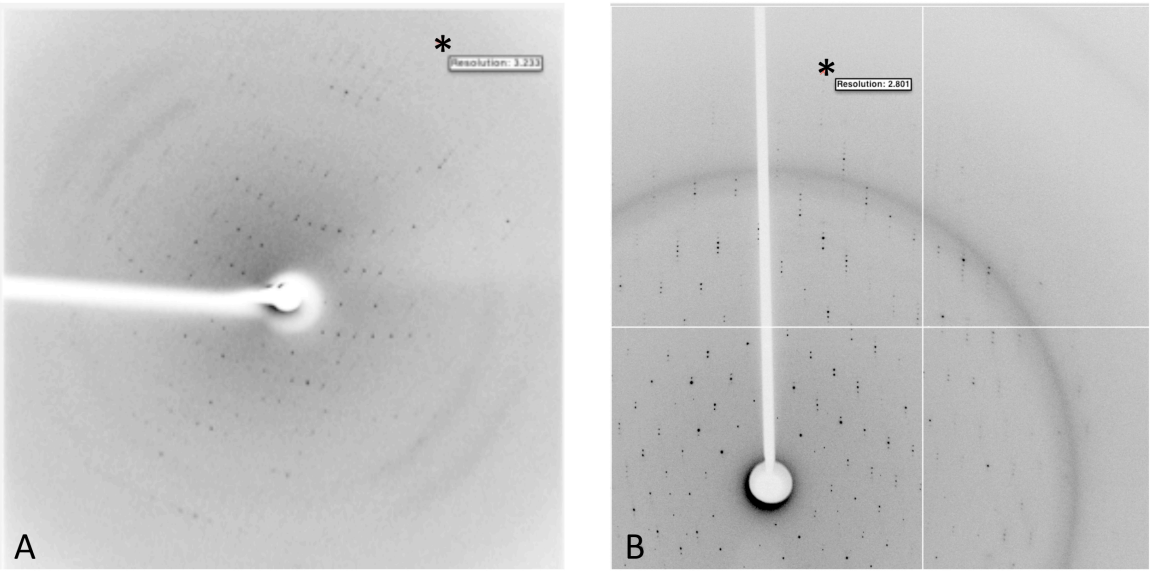


Figure 4.1. Diffraction data on the identical crystal of tbADAT2 croyo-protected in saturated Na/K tartrate following dehydration: A. 15 minute exposure using in-house rotating Cu-anode x-ray source with diffraction to 3.2Å. B. 1.0 second exposure on beamline X29 at National Synchrotron Light Source with diffraction to 2.8. Asterisk (*) indicates visually appreciable diffraction limit.

Data collection on the same crystal using in-house and then synchrotron x-ray sources showed a significant increase in resolution from 3.2 to 2.8Å, respectively. (Fig. 4.1)

Using a robotic auto-mounter, 34 crystals were screened on the synchrotron x-ray source. These crystals exhibited a range of data quality (sharp to moderately mosaic) and

resolution (2.4-3.1 Å). The four best crystals were selected for full dataset collection. Two native datasets were collected at a wavelength of 1.0750 Å (the optimal wavelength for beamline X29). Following a fluorescence scan of a test crystal that showed a small but measurable zinc signal (Fig 4.2A), two full datasets were collected at the empirical Zn-peak of 1.2826 Å for use in SAD phase determination, which is in close agreement to the theoretical Zinc edge. (Fig 4.2B B) It is important to note that Zn was never added to buffers during purification nor during crystallization, suggesting a very tight association with the structure and consequently a high occupancy in the crystal lattice.

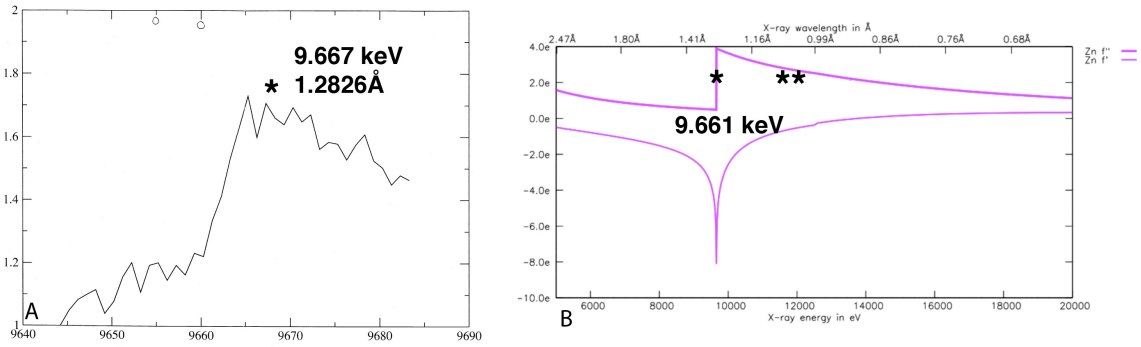


Figure 4.2 X-ray fluorescence spectra of tbADAT2 showing A. the empirical Zn-edge determination for SAD/MAD data collection compared to B. the theoretical Zn-edge.

The mosaicity for these data was between 0.48 and 0.98, which, when combined with the long unit cell (228.6Å) required modification to the typical spot-finding and integration parameters to accurately measure individual reflections against background. The orientation of the long axis of diffraction (i.e. the dimension with the tightest spacing) was oriented such that a rectangular spot-finding profile allowed for an increased area of background measurement while avoiding overlapping data. (Fig. 4.3)

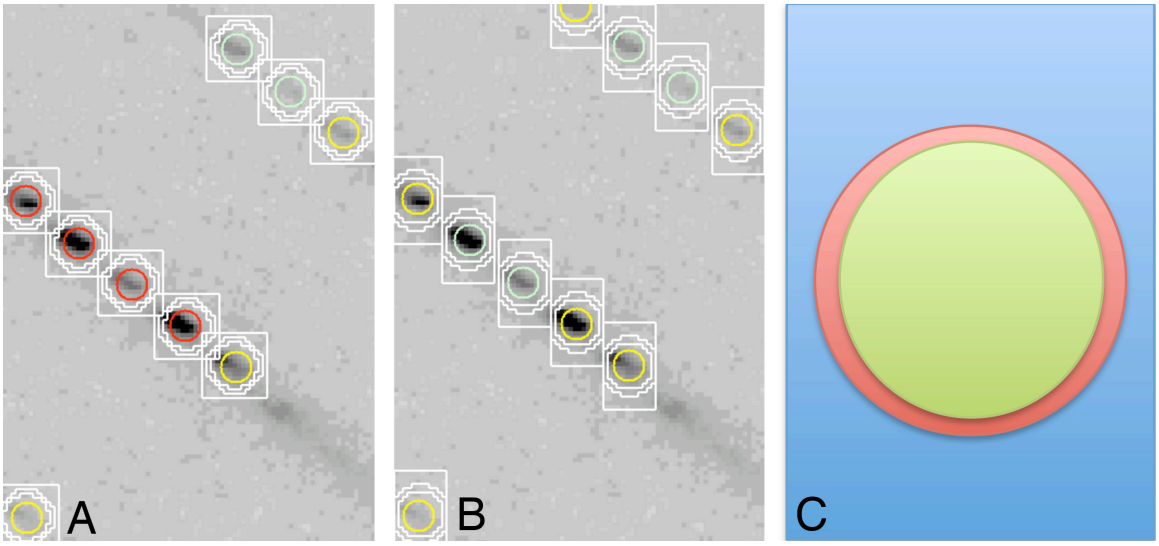


Figure 4.3 Integration profiles used for data processing. A. showing overlapping profiles that were remedied by a the rectangular profile in B. A schematic of the three regions comprising the generic spot profile is shown in C: background (blue), disregarded buffer (red), and integrated diffraction spot (green).

Diffraction data were scaled to 2.95Å (I/σI = 2.44) and the final space group of P4(1)2(1)2 was determined empirically by inspection of systematic absences during scaling as well as by comparison to the results of the molecular replacement solution discussed above. Individual scale error factors were then adjusted to optimize overall χ^2 to 1.0. At this point a clear, resolution dependent deviation of χ^2 for merged vs. un-merged Friedel pairs were seen above ~3.9Å indicating the presence of an anomalous signal for the Zn-SAD dataset. (Fig 4.4A) Data analysis with the X-triage module of PHENIX confirmed the presence of a measureable anomalous signal (Fig. 4.4B) while also ruling out the presence of twinning and other crystallographic features that can complicate structure determination (not shown).

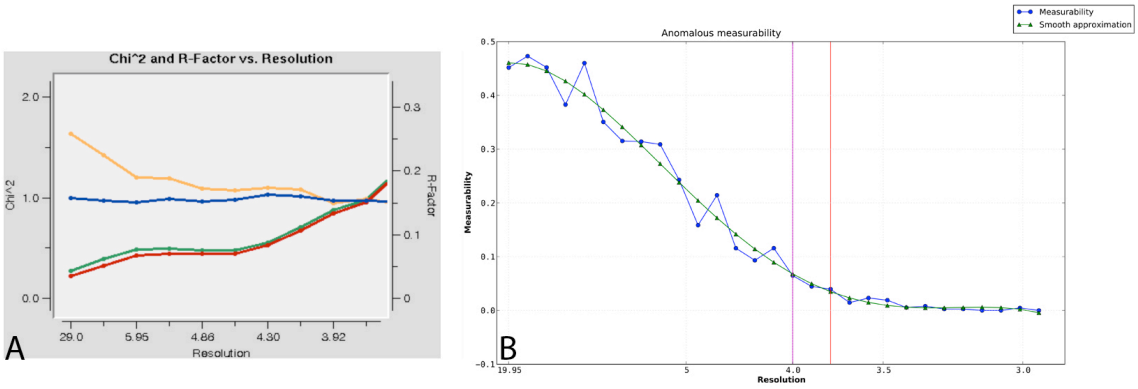


Figure 4.4 Measurable anomalous signal from tbADAT2 crystal diffraction collected at the Zn-peak as assessed during scaling (A) and initial phasing (B) by HKL2000 and PHENIX, respectively.

Native high-resolution tbADAT2 diffraction data were scaled using several different approaches to generate data to be used to extend the resolution of the *de novo* phase solution from the Zn-SAD data. Briefly, these data were scaled with and without regard for the potential anomalous signal of this data and each of these datasets were then truncated for either isotropic or anisotropic diffraction.

Initial data processing was performed as above to address the issue of elevated mosaicity with a long unit cell, and the data was additionally scaled with or without the presence of an anomalous signal given the wavelength of data collection was still within the region of the x-ray spectrum that interacts with Zn atoms. (Fig. 4.4B) Both the anomalously and non-anomalously scaled native data were truncated for isotropic/spherical diffraction at 2.56 Å (I/σI = 2.58), or the data were first truncated at 2.5Å (I/σI = 1.0) and subsequently truncated anisotropically/elliptically to a resolution of 2.8x2.8.2.5Å.

Table 4.1: Crystallographic statistics

Data Collection	tbADAT2 native anomalous and anisotropic	tbADAT2 native	tbADAT2 (Zn-peak)	tbADAT2 -20%gly.
Space Group	P4(1)2(1)2	P4(1) 2(1) 2	P4(1) 2(1) 2	P4(1) 2(1) 2
Cell Dimensions				
a, b, c (Å)	62.49, 62.49, 228.37	62.49, 62.49, 228.368	62.437, 62.437, 228.637	62.747, 62.747, 228.843
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	50-2.5 (2.54-2.5)	50-2.69 (2.6-2.5)	50-2.95 (3.0-2.95)	50.0-3.16 (3.21-3.16)
R_{sym} (%)	10.5	10.5	11.6	12.1
Anisotropy*	2.8x2.8x2.5	N/A	N/A	3.6x3.6x3.2
Mosaicity	0.492-1.096	0.492-1.096	0.485-0.981	0.649-1.749
I/σI	27.754	27.919 (2.58)	32.375 (2.44)	38.625 (2.2)
Completeness (%)	NA	99.9 (39.3)	98.5 (49)	99.2 (35)
Completeness Anisotropic (%)	80.74 (7.44)	N/A	N/A	N/A
Redundancy (anomalous)	14.2 (10.5)	26.7	13.5 (11.2)	47.3 (33.7)
Refinement				
Resolution (Å)	N/A	41.22-2.56	44.1-2.95	48.46-3.16
Resolution anisotropic (Å)	41.22-2.8x2.8x2.5	N/A	N/A	N/A
No. reflections	11992	415,761	243626	410,013
Unique Reflections		11992	8617	8172
Rwork / Rfree	0.2579 / 0.2958	0.2941 / 0.3582	0.2854 / 0.3609	0.3534 / 0.4189**
No. Atoms				
Protein	2375 (313 Cα)	2375 (329 Cα)	2224 (327 Cα)	2270 (324 Cα)
Water	0	0	0	0
Zinc	2	2	2	0
Overall Score	96th percentile	17th percentile	4th percentile	0th percentile
Clash score all atoms	96th percentile	85th percentile	13th percentile	3rd percentile
Poor rotamers (%)	1.23	17.0	22.53	44.39
Ramachandron plot				
Most favored (%)	94.31	68.89	66.03	55.12
Allowed (%)	5.69	86.35	18.73	35.98
Outliers (%)	0	13.65	15.24	19.14
R.m.s deviations				
Bond lengths (Å)	0.0177	0.015	0.013	0.017
Bond angles (°)	1.646	1.923	1.73	2.266

* Anisotropically truncated data were initially scaled to 2.5 Å so as not to eliminate useful data. These data were then subsequently analyzed for anisotropic diffraction and truncated elliptically as described above.

**Rwork/Rfree for this models is for the unrefined autobuilt model with no manual model building/correction.

Chapter 5 Structure Determination

Phases were determined for the Zn-SAD dataset using Auto-Sol in PHENIX. An initial calculation of the measurability of the Zn-anomalous signal was 0.0637 and extended to 4.0Å (acceptable >0.05). For this data, two phase solutions were found with p4(1)2(1)2 receiving the highest scores for electron density skew (0.08), figure of merit (0.29) and estimated map correlation coefficient (40.6 +/- 29.0). Acceptable values for skew and figure of merit are 0.10 and 0.3, respectively, so the values for the Zn-anomalous solution were slightly less than acceptable levels. Solvent flattening performed with RESOLVE in PHENIX yielded initial electron density maps with clear protein/solvent boundaries (Fig. 5.1A-B), which, when contoured at 5σ (Fig 5.1C), showed very clear density for the individual Zinc atoms with a pattern of eight pairs of zinc atoms per unit cell corresponding to eight dimers (one dimer/asymmetric unit).

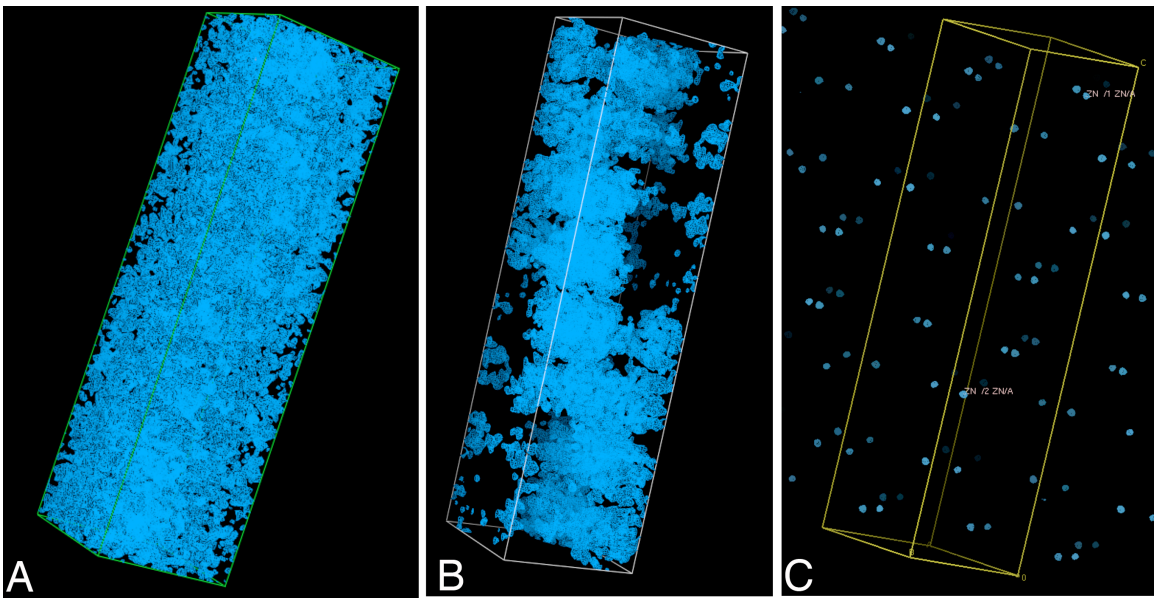


Figure 5.1. Electron density from initial phase solution of tbADAT2 (**A**) and the same density following solvent flattening with RESOLVE contoured at 1σ (**B**) and at 5σ (**C**).

Automated model building was attempted using both ARPWARP and AutoBuild (PHENIX), with only the latter achieving successful results. The resulting model comprised 223 residues in 3 chains with an Rwork/Rfree of 0.3546 / 0.4042 (Fig. 5.2A). Alignment of hsADAT2 to this initial model (RMS 3.69Å for 169 residues) revealed general information about the initial structure allowing assignments to be made for the various fragments with regard to monomer and residue numbering. (Fig 5.2B)

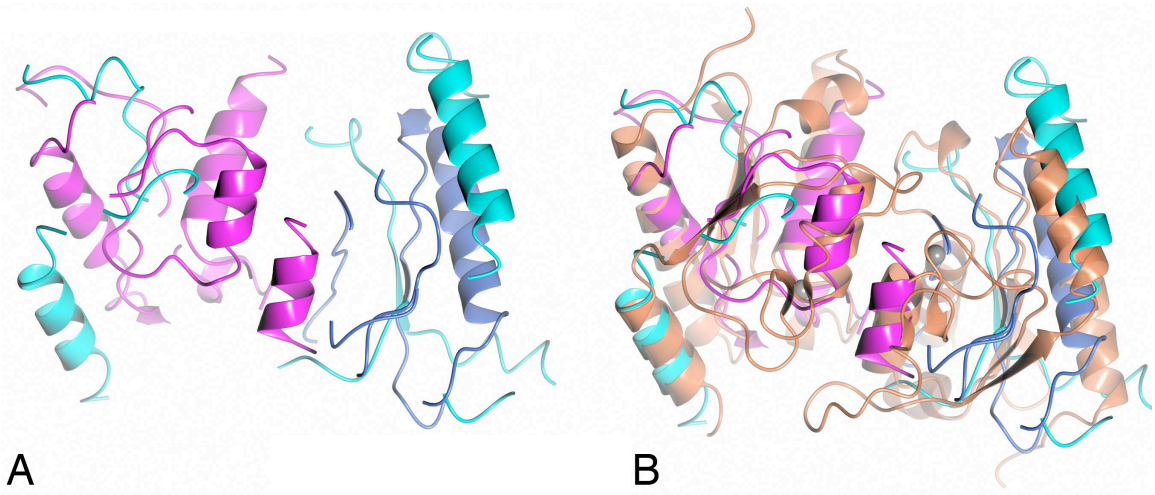


Figure 5.2 Resulting tbADAT2 structure from automated model building (**A**) aligned with hsADAT2 in beige (**B**). In A the three colors represent three different chain assignments made by the autobuild software that can be seen to align to one of two chains in the hsADAT2 structure.

Manual model building initially focused on one of the two monomers, which was built using the initial F_0-F_c density map contoured at 1σ. Of the 225 residues in the monomer, 157 were built into clear, continuous electron density. The model of the initial monomer (A) was then aligned to the opposing monomer (B) and used as a template to finish building the dimer. At this point, the correct dimer interface was distinguished from crystal contacts using PISA in addition to alignments with- and comparisons to-

published protein structures containing the cytidine deaminase fold.[74, 75] Refinement of the completed model against the Zn-SAD data to 2.95Å resulted in an R_{work}/R_{free} of 0.30/0.34.

The refined tbADAT2 model was improved through additional rounds of model building and refinement against the higher resolution native data. In all cases the freeR data from the initial solution were appended to the higher-resolution native data to ensure meaningful R_{free} calculations. Initial refinement against these native data truncated isotropically at 2.56Å ($I/\sigma I = 2.58$) without anomalous scaling showed marginal improvement of the model during refinement and resulted in electron density maps with little increase in detail. Anomalous refinement of the tbADAT2 model against native data scaled anomalously at 2.69Å ($I/\sigma I = 2.48$) showed improved R_{free} values that were lower still when refinement accounted for non-crystallographic symmetry or when the pre-refined model was subjected to TLS analysis. Additionally, the R_{free} was improved by refinement against anisotropically truncated data (2.8x2.8x2.5Å) that included the strongest higher resolution diffraction while simultaneously discarding the noisy, weak diffraction in the higher resolution data shells. The improvement on Rfree was found to be relatively independent and additive for these three approaches, so the best refinement strategy was found to be restrained anomalous refinement on anomalously-scaled, anisotropically truncated native data in combination with a model subjected to TLS analysis prior to rounds of refinement. Throughout model building and refinement, geometric restraints and Ramachandran statistics were monitored within COOT and using MOLPROBITY to assess and remedy problematic regions of the model. This approach yieled a final tbADAT2 with an R_{free}/R_{work} of 29.6/25.8 and a MOLPROBITY scores for

all-atom contacts and protein geometry in the 96th percentile out of 6730 structures with a resolution of 2.51 +/- 0.25Å. The final model had no Ramachandran outliers and had 94.3% of residues in favored regions. (Fig 5.3)

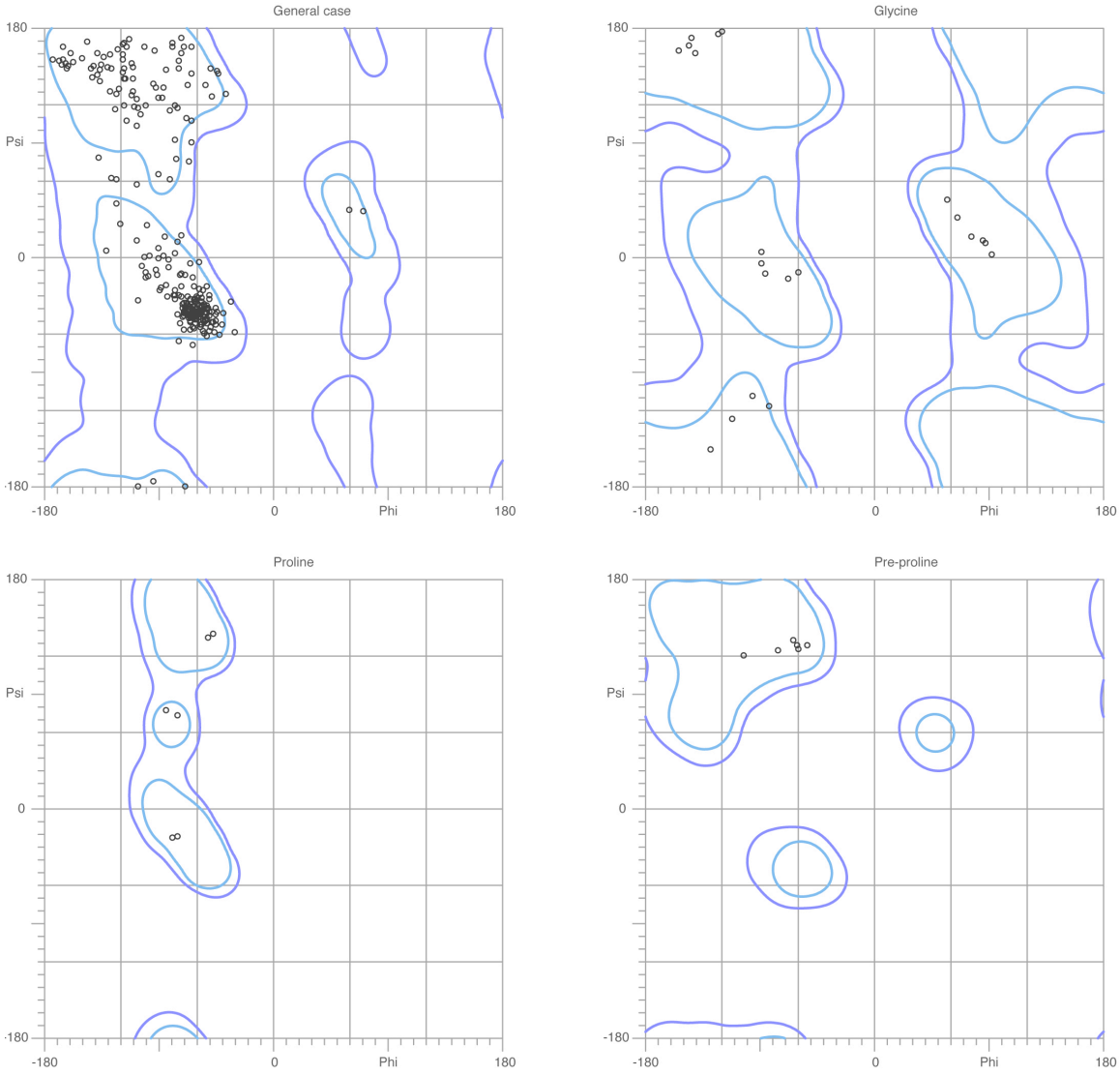


Figure 5.3. Ramachandran plot generated by MOLPROBITY of final tbADAT2 model showing, clockwise from upper left, all residues, glycines, pre-prolines, and prolines: 94.3% of all residues in favored regions (light blue), 100% of residues in allowed regions.

The improved refinement with higher resolution data failed to clarify regions of discontinuous electron density in the initial model: residues 1-18, 108-121, 161-184, and 209-225. (Fig. 5.4 A-B) Several attempts were made to improve the electron density maps in these areas including simulated annealing and the calculation of composite omission maps. Several automated loop-building applications in PHENIX and ARPWARP were also used to improve the model of these unresolved areas of density. Neither of these additional approaches yielded any improvement.

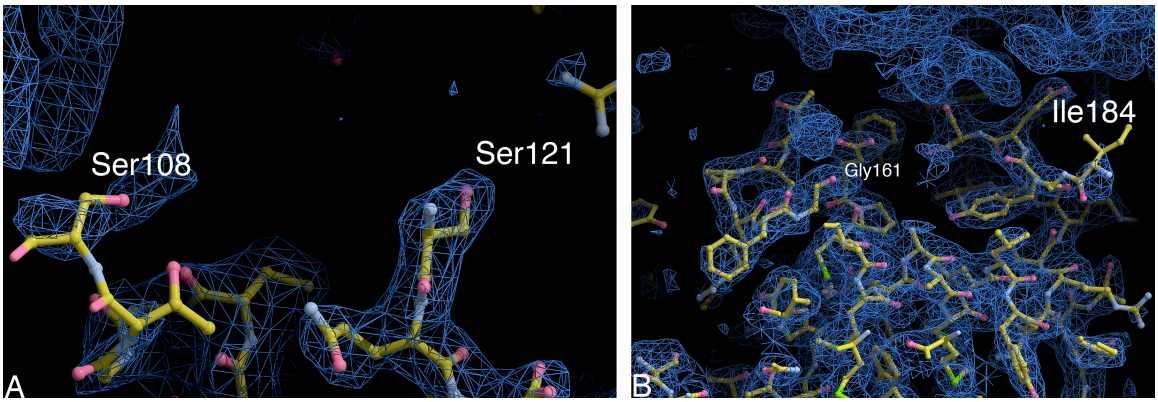


Figure 5.4. Discontinuities in final model of tbADAT2 model and refined 2Fo-Fc electron density map contoured at 1.0σ showing the two regions of discontinuity in the structure: A S108-S121 and B G161-I184.

Finally, attempts were made to add the Zn-coordinated water molecule that has been seen in homologous structures at higher resoltuion. While the density wasn't perfectly clear for the placement of water, it did appear that density surrounding the zinc ion exhibited a bulge in the correct direction to indicate the presence of this water. (Fig. 5.5) Multiple refinement attempts following the addition of this water molecule to the model failed to satisfactorily refine the location of the water, so this water molecule was not included in the final model.

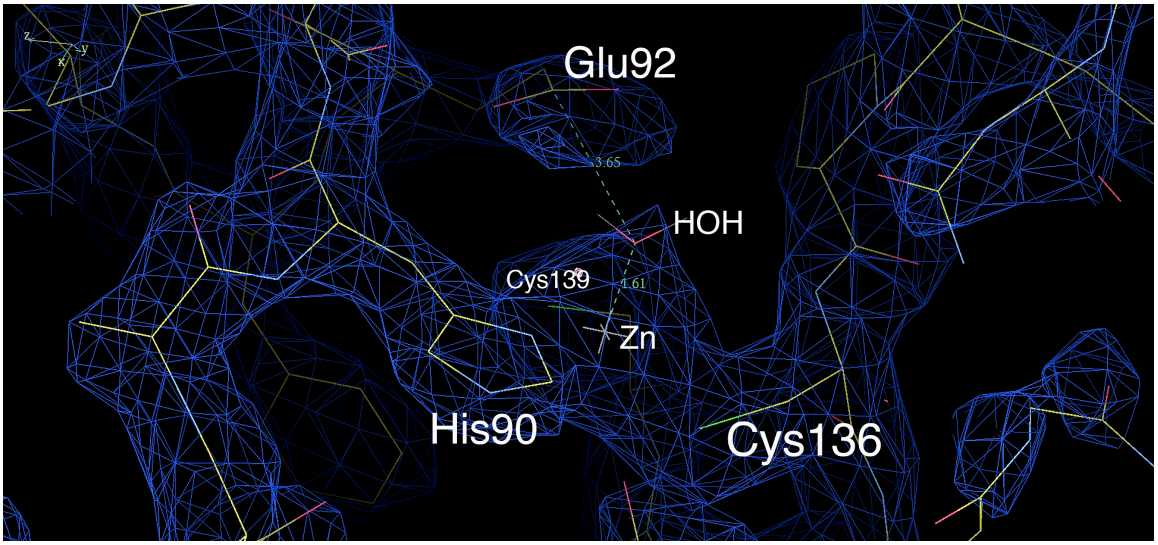


Figure 5.5. Final tbADAT2 model and refined 2Fo-Fc electron density contoured at 2σ showing the active-site zinc and surrounding zin-coordinating residues. A water molecule is shown modeled (not refined) into the bulge in electron density surrounding the Zinc with distances to the Zn and E92 of 1.61 and 3.65Å, respectively.

Chapter 6 Structure Analysis

The structure of tbADAT2 was solved to 2.8x2.8x2.5Å resolution in space group p4(1)2(1)2 by SAD phasing off one natively bound zinc atom per monomer, demonstrating a phasing effectiveness of one anomalous scatterer for 225 amino acid residues. The structure contains clear density for 313 of the 450 residues (69.6%) in the asymmetric unit, which comprises a dimer formed across an extensive, primarily hydrophobic interface burying 2285 $Å^2$ between two monomers related by non-crystallographic two-fold rotational symmetry. (Fig 6.1 A) Two additional interfaces are seen in the crystal which burry 892.8 and 689.8 $Å^2$, both of which have been classified as crystal packing interactions based on comparisons to homologous structures and structural analysis with PISA (Protein Interfaces Surfaces and Assemblies).[76]

The structure of tbADAT2 exhibits a topology closely resembling other members of the cytidine deaminase family to which it belongs. Each monomer contains a single domain with the canonical cytidine deaminase fold comprising a mixed 5-strand β-sheet core sandwiched on one side with two long α-helices forming the outer border of the monomer and on the other side with a set of three α-helices that form both the Zn-binding cleft and the dimer interface. (Fig. 6.1A) Each monomer contains a single coordinated zinc ion with apparent tetrahedral coordination geometry consistent with other members of the cytidine deaminase family. Three residues coordinate the zinc, His90 (1.83Å), Cys136 (2.39Å) and Cys139 (2.18Å). A fourth coordination is believed to be provided by an activated water as in many zinc metallo-enzymes. Evidence of this water is suggested by the electron density around the zinc (Fig 5.5); however the presence of a coordinating water is not certain as various attempts to refine this water

proved inconclusive. The B-factors for the final structure follow the degree of conservation of the cytidine deaminase fold with the most conserved core region exhibiting B-factors in the 20s-30s smoothly transition up the the mid 50s-60s for boundary regions of the protein that are less highly conserved. Residues 54-76 form an unstructured, non-conserved loop that exhibit the highest B-factors of the structure indicating a high degree of flexibility. This loop was particularly difficult to refine successfully and contains some of the weakest density in the structure. The B-factors for the middle of the loop (125) are at the same level of residues immediately preceding and following other breaks in the electron density (see below). Therefore, while the loop did refine successfully as built, any serious structural conclusions made regarding its structure and orientation must be made with the above discussion in mind.

Other structural features of the monomer are eight distinct intra-chain salt bridges that serve to tie the outer and inner bundles of α-helices together as well as to tie the helices to the β-sheet core. No disulfide bonds are seen between any of the six cysteines that are not involved in coordinating the zinc.

Structural alignments of tbADAT2 with other ADAT homologues reveal a high degree of conservation of the cytidine deaminase fold. (Fig. 6.2D) Sequence identities for these alignments were 32, 34, and 42% with r.m.s. deviations of 1.3, 1.5, and 1.9Å. for atADAT, saADAT, and hsADAT2 homologues, respectively. (Fig. 6.2B) This high degree of conservation is exemplified at the active site zinc where the structural conservation of the HAE…PCIMC domain is quite high with the coordination geometry remaining nearly identical for all 4 examples compared. (Fig. 6.1C) Several of the homologues contain coordinated water, and the location matched the orientation of the

indicative bulge in electron density discussed above for the zinc-coordinated water the tbADAT2 structure.

The dimer interfaces for the hsADAT2, saADAT, and atADAT structures were all within 100Å^2 of the tbADAT2 structure and were all equally hydrophobic. While alignments of individual monomers from these homologues show the high structural homology mentioned above, a fair degree of structural diversity is seen across the dimer interface with monomers rotated 20-30° across the dimer interface relative to the tbADAT2 structure. This rotation is seen between structures both with and without substrate bound, so it does not appear to be a substrate-related conformation switch, although the differences certainly alters the substrate binding cleft, as discussed below.

A sequence alignment of tbADAT2 with hsADAT2, saADAT, and atADAT identified three regions in the tbADAT2 sequence that contained insertions and/or regions of low homology vis-à-vis other members of this family, referred to here as loops A (residues 54-71), B (105-120), and C (172-190). (Fig 6.2 A) Of these, only loop A is seen in the tbADAT2 structure and protrudes away from the bulk of the core fold at the base of the β-sheet, opposite the opening to the active site of the monomer to which it belongs. It is loop A that was referenced above for exhibiting high B-factors suggesting this loop is likely more mobile than the surrounding structure. (Fig. 6.2 B-C) The loop itself was initially built as a small helix turn helix sub-domain, but attempts to model the loop as such yielded consistently poorer results than the final orientation of the loop modeled without α-helical restraints, as shown here. (Fig. 6.2C)

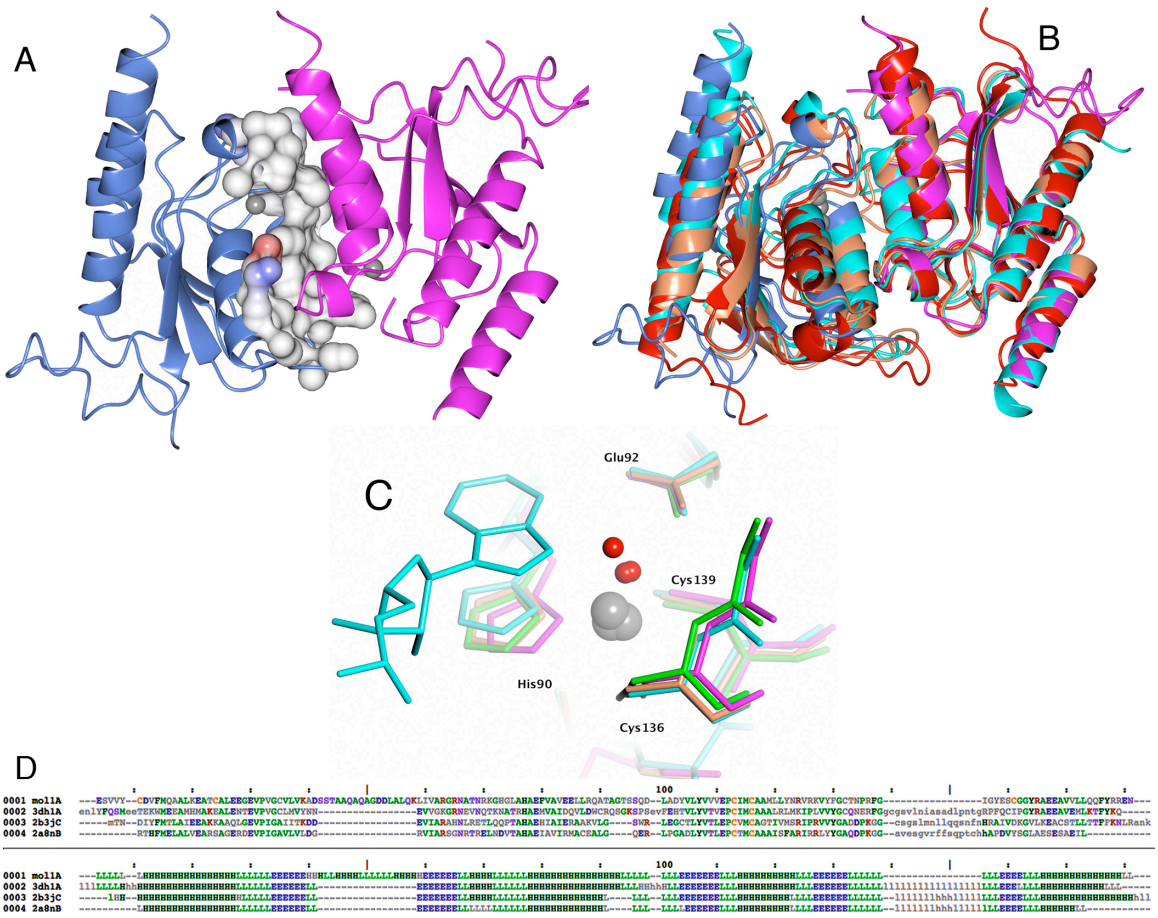
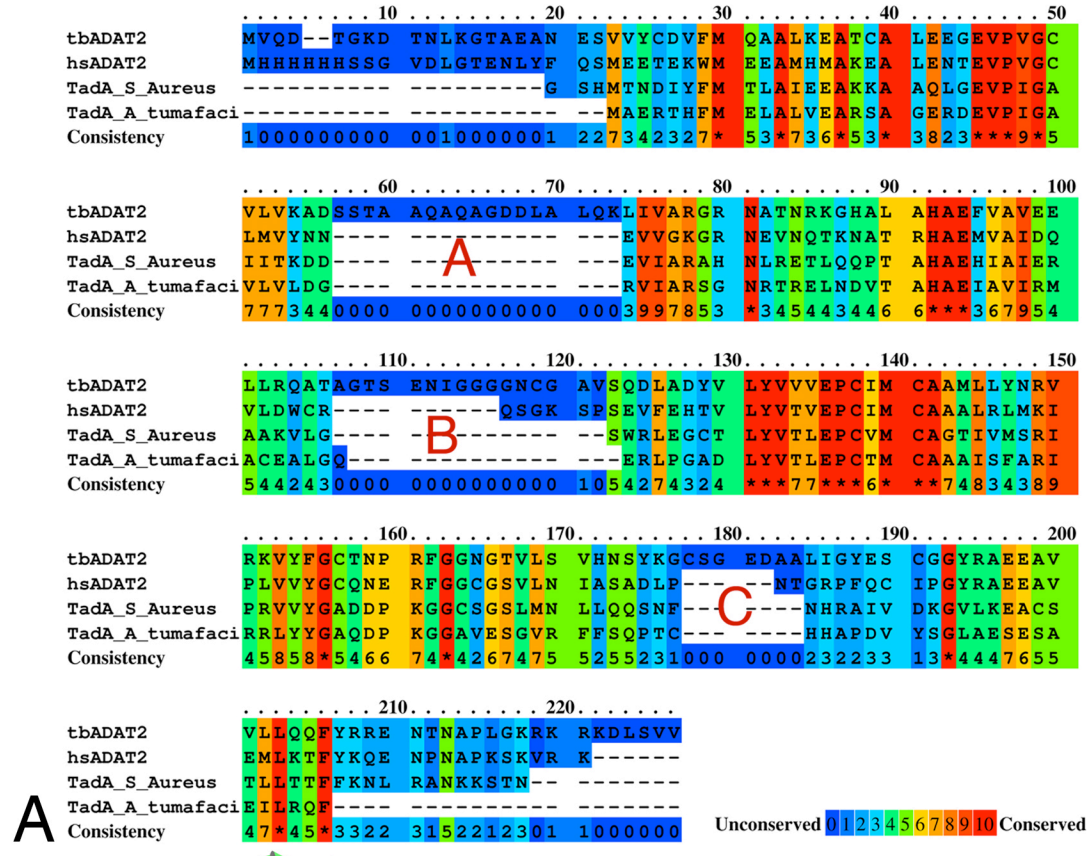
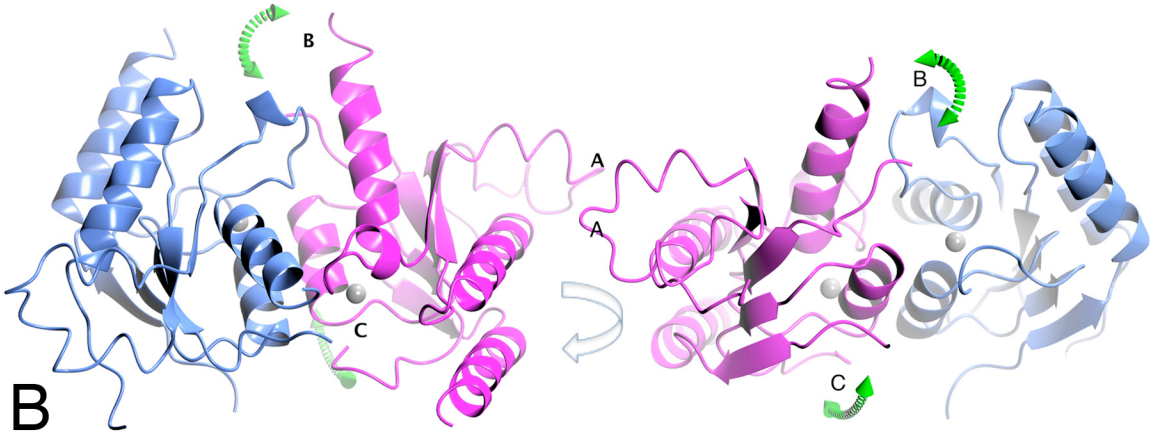


Figure 6.1. Structure of tbADAT2 showing (**A**) the dimer comprising the asymmetric unit with a space-filling representation of the dimer interface colored for electrostatic potential. (**B-D**) Structural alignment of tbADAT2 with homologous structures hsADAT2 (pdb3DH1 in red), *S. aureus* TadA (pdb2B3J in cyan B, and *A. tumefaciens* (pdb2A8N in coral). The alignments in (**B**) were performed using rigid models such that the high structural homology can be observed for the monomers (right side in **B**, in **C** and **D**) and the disordered alignment on the left in B shows the degree of structural variability of the dimer. Active site labels in **C** correspond to tbADAT2. Gray and red spheres are zinc and water, respectively. Alignments in **D** depict sequence and topology conservation between these homologues structures for ordered regions only. Stereo view available in appendix I.

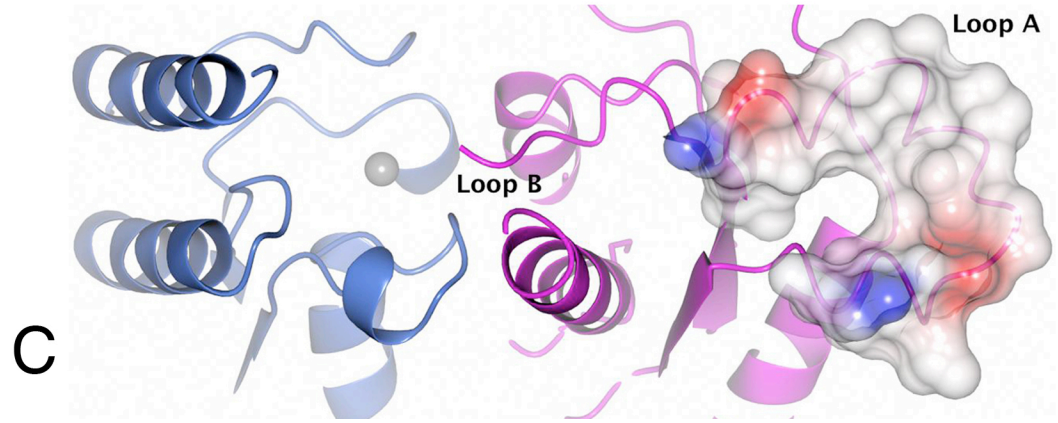
Figure 6.2 Sequence and structural analysis of non-homologous insertions in tbADAT2 amino acid sequence relative to homologous ADATs (**A**) Sequence alignment colored by conservation from lowest (blue) to highest (red). (**B**) Two views rotated by 180° about an axis in the plane of the paper (arrow) of the final tbADAT2 homodimer depicted as a ribbon diagram showing the location of the insertions identified in panel A. Loop A is seen near the axis of rotation. Loops B and C are not structured and are indicated by general location with green arrows. (**C**) Structures as in B showing a view from the top of the dimer looking into the binding site of blue monomer. Loop A is depicted as a space filling model colored for electrostatic potential. "Loop B" is indicated as such.


tbADAT2
hsADAT2
TadA_S_Aureus
TadA_A_tumafaci
Consistency
A
B
C
Unconserved 0 1 2 3 4 5 6 7 8 9 10 Conserved

A


B
A
A
B
C
C

B


Loop A
Loop B

C

Given the rotational symmetry of the dimer, loop A of one monomer is on the same face as the opening to the active site of the other monomer of the dimer. (Fig. 6.2 C and 6.3) A space filling model for loop A colored for electrostatic potential shows a relatively flat surface with regions of positive and negative potential that forms part of an extended face oriented toward the direction that substrates would approach the active site of ADAT2. Loop B of monomer A also forms a portion of this face and is positioned to form and or extend the substrate-binding surface adjacent to the active site; however, given the lack of density for this region, the orientation can only be inferred from the location and trajectory of the structured at the beginning and end of this loop. (Fig. 6.3) The same can be said regarding loop C, which covers the one side face the structure (if the substrate binding clefts are referred to as the top and bottom). The beginning and end of this loop suggests that loop C covers most of the side of the structure and likely interacts with loop C from the other monomer in the dimer. The function of these insertions can only by hypothesized at this point; however an interesting possibility will be discussed below that these insertions enable a significant domain swap upon hertero-dimerization with ADAT3. BLAST analysis shows that the A, B and C insertions are present in ADAT2 form other trypanosomes and possibly some *leishmania* species but are noticeably absent from ADATs from higher eukaryotes and prokaryotes.

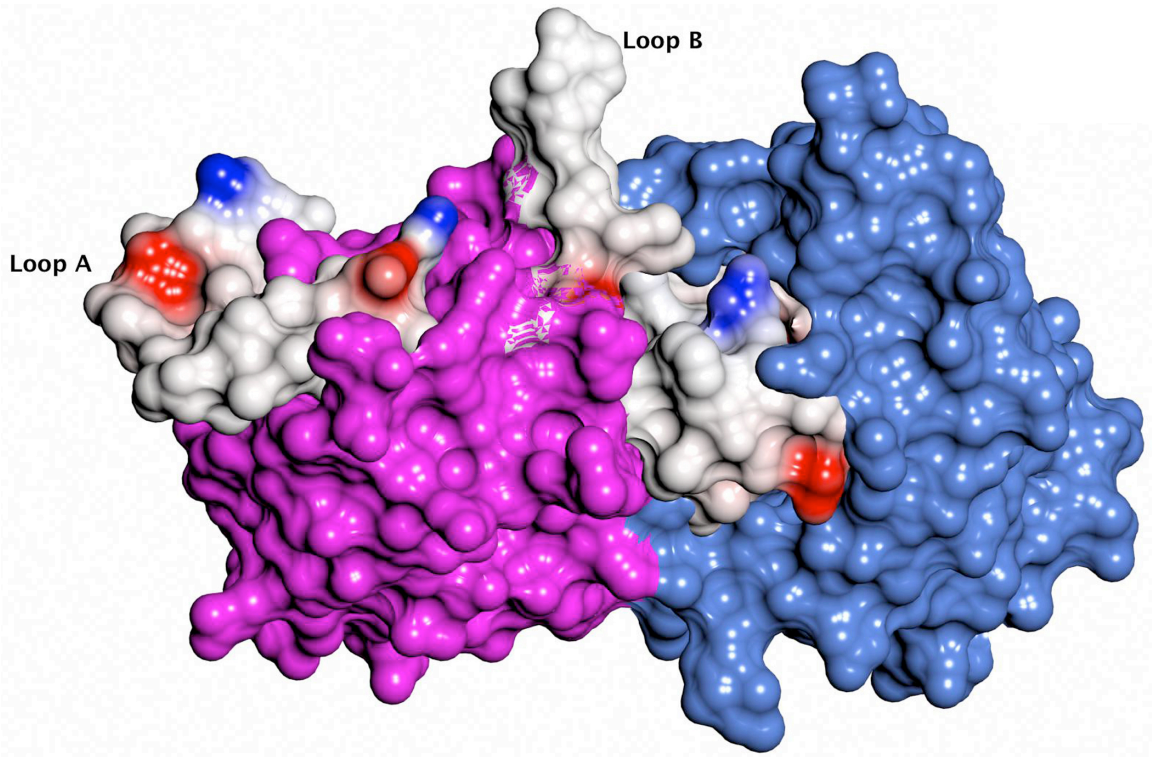


Figure 6.3 Space filling model of tbADAT2 showing Loops A and B in relation to the active site. Monomers are in magenta and blue. Loops are colored for electrostatic potential, as is the catalytic core of the blue monomer.

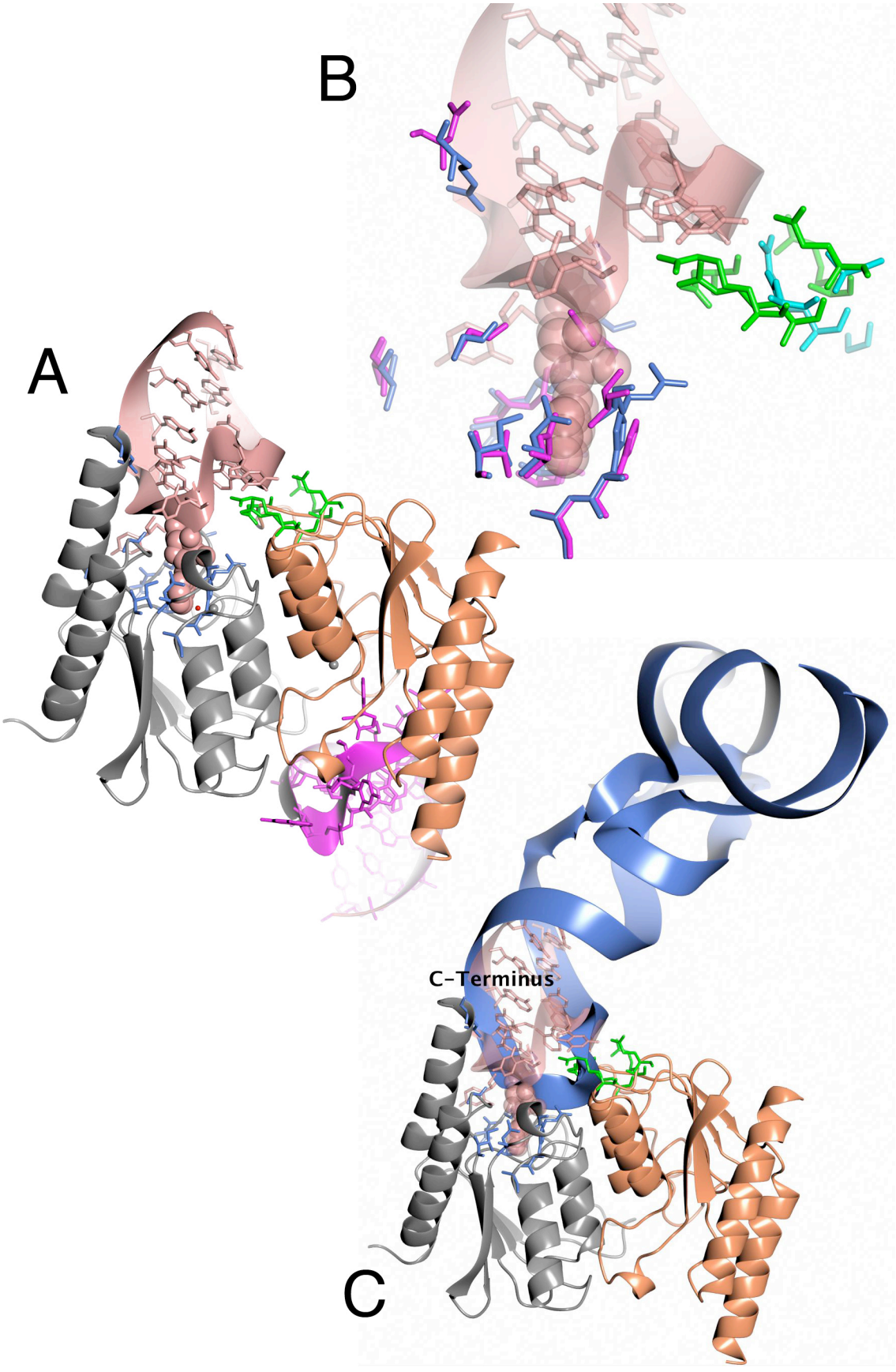
The active site: The binding of an RNAstem loop substrate has been characterized crystallographically for the prokaryotic saADAT which was shown to bind two RNA substrates per dimer, one each at two solvent exposed clefts on opposite sides of the dimer formed at the interface between the two monomers.[27] (Fig. 6.4A) The interactions between the dimer and each RNA stem loop were extensive with a total of 17 residues contacting each substrate. (Fig. 6.4B) At each binding site, the majority of the residues contacting the substrate arise from the same monomer containing the active site zinc associated with the respective active site, however, several significant contacts are made by residues of the adjacent monomer. A sequence alignment was used to identify the corresponding residues in tbADAT2. Of these 17 residues, 6 are identical between the

two structures. The remaining 11 equivalent positions in tbADAT2 show a few conservative mutations, but a significant portion have non-conservative mutations that appear on a structural alignment to alter or eliminate altogether the interaction those residues would make to the aligned RNA substrate. (Fig. 6.2B, Table 6.1) A model made by docking a tRNA structure onto the stem-loop of the saADAT structure does not indicate any additional regions of contact between the tRNA and the ADAT molecule. (Figure 6.3)

asADAT		tbADAT2	
A	B	A	B
G22	R70	**G42**	D123
I26	E72	V46	A125
N42	**R94**	**N79**	**R147**
R44	N123	T81	*L180*
H53	R125	**H90**	G185
A54		**A91**	
E81		**E134**	
D103		T156	
D104		N157	
K106		R159	
S138		V198	
R149		N209	

Table 6-1 Residues involved in RNA binding in saADAT vs. tbADAT2 based on sequence alignment. Bold residues are conserved.

Figure 6.4 Modified stem-loop bound to saADAT. A) Showing the dimer and bound stem loop side chains for residues that contact the RNA. Grey is monomer A with blue side-chains, Orange is monomer B with green side-chains. Pink and magenta show bound stem-loop molecules with the substrate analogue zebularine represented as spheres. B) Zoomed view of RNA-ADAT interaction. Colors as in A with tbADAT side-chains from equivalent monomers A and B in magenta and cyan. C) Equivalent view as in A with full length $tRNA^{ile\ 77}$ manually aligned to the stem-loop structure. Stereo view available in appendix I.


B
A
C-Terminus
C

Modeling of the ADAT2/3 homodimer

Spears *et al*. have generated two models of the hsADAT2/tbADAT3 homodimer with and without a proposed domain swap that involves the C-terminal 100 residues of ADAT2 in a symmetrical swap with an eqivalent C-terminal domain of ADAT3 [38]. This model was initially based on the hsADAT2 crystal structure and a computational model of tbADAT3. To generate a more suitable model of the tbADAT2/ADAT3 heterodimer, the models from Spears *et al*. were modified to replace the hsADAT2 portions with the tbADAT2 structure from the present study. No computational models were uses to make this substitution. When compared to the tbADAT2 structure, the two monomers of the hsADAT2 are rotated approximately 20° counterclockwise about an axis almost perpendicular to the dimer interface; consequently, the starting heterodimer model reflected this orientation. A new model was created by re-aligning the tbADAT3 portion of the original model with monomer B of the ADAT2 homodimer. This new orientation was then copied into a new model with ADAT2 monomer A to form the heterodimer. With the unswapped model of tbADAT2/3 as guide, the domain swap portions of each monomer were realigned to generate a representative domain swap model with the appropriate ADAT2 and ADAT3 amino acid sequence and residue numbering. (Fig. 6.5 A-D) The final domain swapped model comprises the native postions of residues 1-108 of ADAT2 and 1-272 of ADAT3 and swapped positions for residues 121-209 of ADAT2 and residues 276-365 of ADAT3. The cross-over of the domain-swapped region of ADAT2 corresponds with loop B discussed above, for which there is no density in the tbADAT2 homodimer model. For this reason, the actual cross-over loops between the native and swapped orientations of ADAT2 and ADAT3 were not modeled. This model

was used for structural comparisons and for the design of disulfide cross-linking experiments.

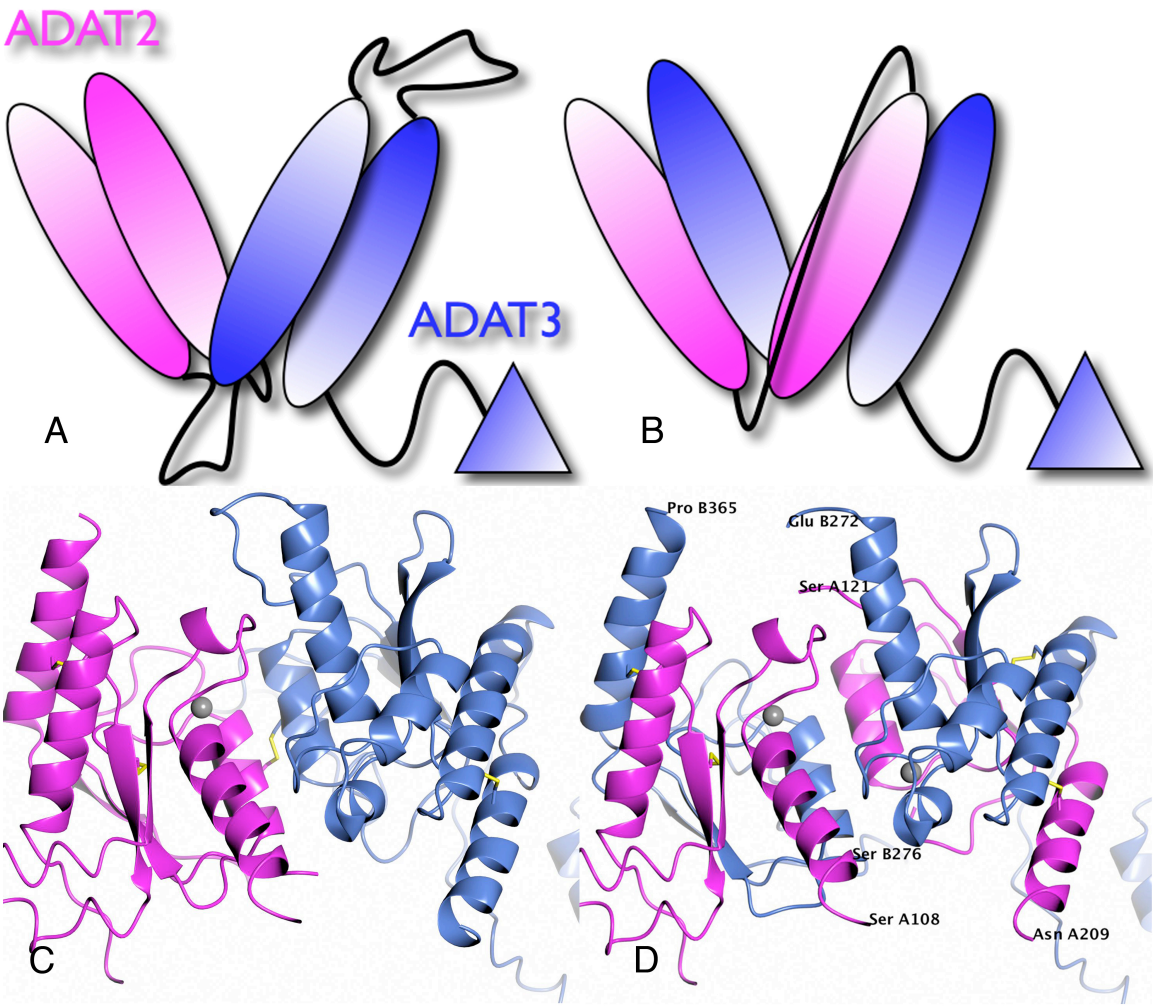


Figure 6.5 Purported domain swap models of tbADAT2/3 showing tbADAT2 (pink) and tbADAT3 (blue) as both cartoon and ribbon diagram depictions of the un-swapped dimers (**A** and **C**) and the domain swapped dimers (**B** and **D**). Proposed disulfide cross-links are shown in yellow (**B** and **D**). Stereo views available in Appendix I.

An analysis of the swapped and un-swapped tbADAT2/3 heterodimer models was conducted using the Disulphide by Design® web-server [61], which identified numerous sites amenable to favorable disulfide formation. These results were analyzed in the

context of the starting models to identify bond pairs that when cross-linked would yield a clear characteristic banding patterns on SDS-PAGE depending on the swapped state of the dimer. Candidate pairs were rejected if they were too close to the active site cysteines. This resulted in a total of eight pairs of mutations. (Table 6.1, Fig 6.6) Three additional cystein pairs were identified to generate several positive controls for this experiment. These included an intra-chain pair in ADAT2 (Cys34-Cys78), an inter-chain disulfide in the ADAT2 homodimer (Cys141), and an inter-chain disulfide in the ADAT2/3 heterodimer that would be the same in the either the swapped or non-swapped configurations (ADAT2 Cys141 – ADAT3 Cys 296) (Fig 6.6)

Table 6-2 Disulfide cross-linking mutagenesis

Pair #	Residue #1	Residue #2	Disulfide if domain swapped	Disulfide if no domain swap
1	ADAT2 P45C	ADAT3 H354C	A-yes	A-no
2	ADAT2 Cys48*	ADAT3 V268C	B-Yes	B- No
3	ADAT2 L200C	ADAT3 G201C	A' – Yes	A' – No
4	ADAT2 V131C	ADAT3 A204C	B' – Yes	B' – No
5	ADAT2 Y130C	ADAT3 F183C	C – Yes	N/A
6	ADAT2 P45C	ADAT2 L200C	A – No	A – Yes
7	ADAT2 Cys48	ADAT2 V131C	B – No	B – Yes
8	ADAT3 G201C	ADAT3H354C	A – No	A' – Yes

* Cysteine residues native to the sequence

These mutations were created by site directed mutagenesis of a his-tagged tbADAT3 construct in pet28^{Kan} and an untagged tbADAT2 construct in position two of pET-DUET^{Amp}. Co-transformations of these two pET vectors was very inefficient, frequently with only one or two double-resistant colonies per reaction compared to >100 for the same conditions with a single selection for one construct. The doubly transformed strains were successfully grown and induced, and the his-tagged ADAT2 and ADAT3 were purified in a manner similar to the method described above for ADAT2 with the major

difference being absence of reducing conditions to preserve disulfide formation; however, significant background, low yield, and considerable laddering complicated attempts at purification. It seems likely these difficulties were encountered due to the absence of reducing agent. Screening for optimal induction cell lines and induction condition identified the strongest induction in BL21-AI cells induced at 37°C for 2-3 hours with 0.5 mM IPTG and 0.1% arabinose; however purification attempts on material from these cells showed highly insoluble material that was not recoverable through either batch dialysis refolding nor on-column refolding. These experiments are ongoing with the focus on optimization of induction and of purification to identify the minimal concentration of reducing agent necessary to facilitate purification without compromising specific disulfide formation. Several additional options are available to pursue, including the creation of a minimal cysteine background sequence for both ADAT2 and ADAT3.

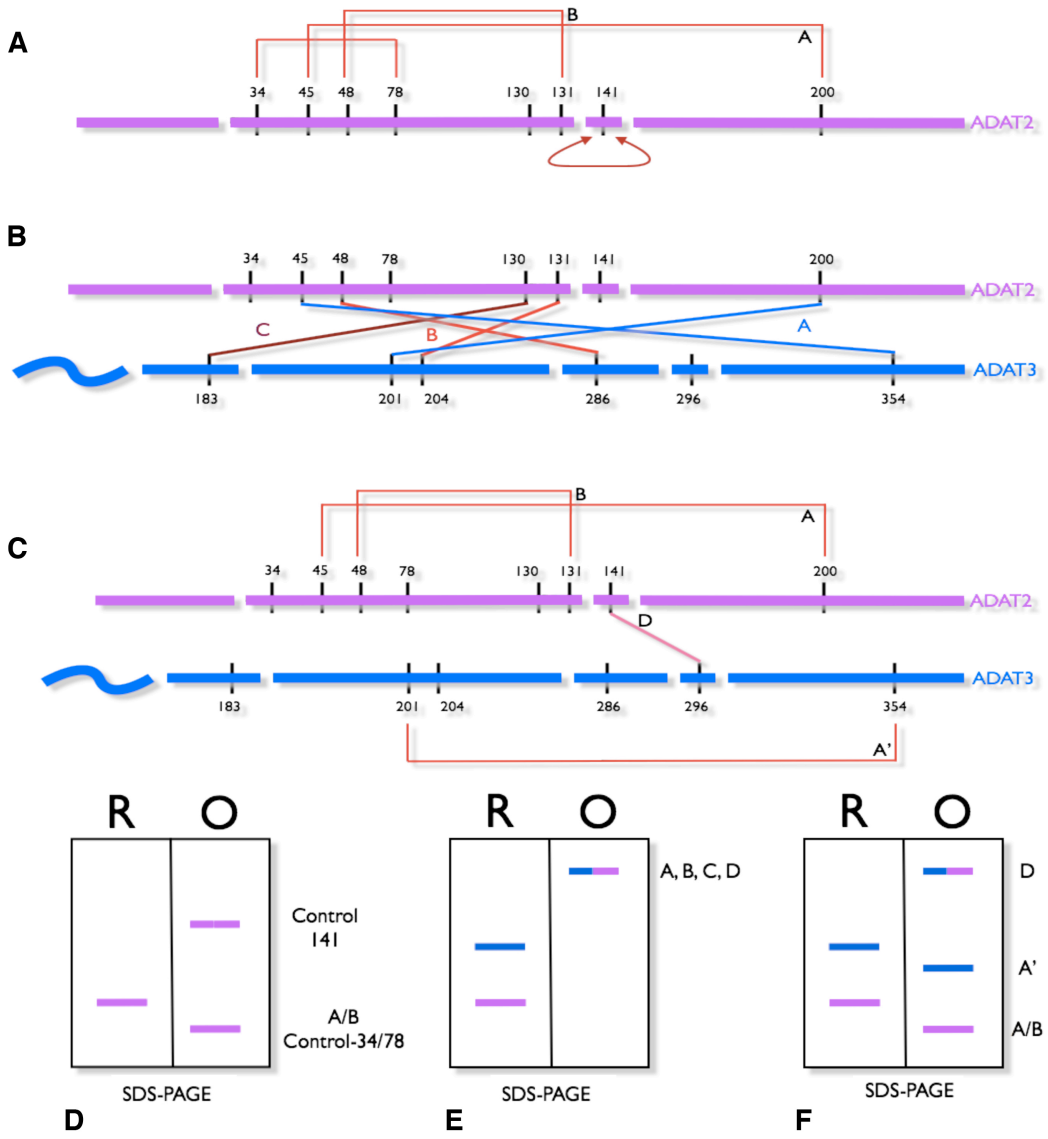


Figure 6.6. Schematics of disulfide cross-linking interrogation of ADAT2/3 heterodimer. Panels **A-C** depict predicted disulfide formation between introduced cysteine residues: **A** as a control in the ADAT2 homodimer, **B** in the domain swapped ADAT2/3 heterodimer, and **C** in the non-domain swapped ADAT2/3 heterodimer. Panels **D-E** depict predicted migration of purified material run on SDS PAGE in reducing (R) or oxidizing (O) conditions. Panels **D**, **E**, and **F** correspond to the schematics and disulfide pairs in **A**, **B**, and **C**, respectively, showing decreased migration due to inter-chain disulfide formation and increased migration due to intra-chain disulfide formation.

Chapter 7 Discussion

Comparisons between prokaryotic and eukaryotic tRNA editing of A34 in the anticodon wobble position highlight several distinct gaps in our understanding of how the eukaryotic enzymes function. Moving from prokarya to eukarya, the enzymes that deaminate A34 in tRNA exhibit a more permissive substrate specificity, a requirement for ADAT2/3 hetero-dimerization instead of ADAT2 homo-dimerization, and show a more diverse spectrum of enzymatic activity. Cells have evolved the ability and need to dynamically modify the genetic information they carry, and in many cases, including those performed by the ADAT family of enzymes, these modifications are essential. Understanding the regulation of these enzymes is an important step toward understanding the rules that govern RNA editing in higher eukarya and will likely pertain to a wider field of epigenetic gene regulation. The structure of tbADAT2 discussed below confirms previously known details about ADAT2 and provides several key insights about paths forward. This discussion will proceed in several parts to highlight the process of crystallizing the protein, determining the structure, and interpreting the structure in the context of pertinent extant literature. The discussion will conclude with a description of several experiments that are immediately relevant and propose a path forward.

The purification and crystallization of tbADAT2.

Recombinant expression of full-length tbADAT2 (wt) in *E. coli* was initially problematic with low expression, poor solubility, and aggregation, which on SDS-PAGE revealed a discrete laddering of the protein bands characteristic of inter-chain disulfide formation. An empiric approach of sequentially testing single Cys-to-Ser point mutants identified

Cys117Ser as a key mutation that solved all the initial problems of protein expression and purification. This residue is located in one of three notable insertions in the tbADAT2 sequence relative to the sequences of previously crystallized bacterial ADAT and hsADAT2 homologues. (Fig. 6.2 A-B) In retrospect both Cys117 in loop B and Cys178 in loop C suggest themselves as potential candidates for problematic cross-linking targets since both are located in regions of non-homology to other successfully crystallized proteins and both are in regions with no secondary structure prediction. Cys178 was never tested to determine if it too would alleviate the undesired cross-linking. Clearly empiric approaches are still necessary to assess difficulties in protein expression, but such an analysis could intelligently guide the process in general cases and could be useful in future crystallization attempts with the ADAT2/3 heterodimer.

Purification of tbADAT2 C117S (tbADAT2) for crystallization was very successful yielding a well-behaved, non-aggregating protein sample that did not show evidence of background disulfide formation. Gel filtration supported the conclusion that tbADAT2 exists as a dimer in solution *in vitro* (Fig. 3.1C-D), a result consistent with similar observations in the crystal structures of bacterial ADAT and hsADAT2 and with published experimental data [38]; however, oligomerization was never assessed at low/physiologic protein concentrations. Purified tbADAT2 was highly soluble and could be concentrated above 100mg/ml without evidence of aggregation on gel filtration, and crystallization screening of tbADAT2 was conducted at both 50 and 100 mg/ml. The initial conditions that were eventually optimized to produce data quality crystals were only successful at crystallizing tbADAT2 at the higher of these two concentrations, a

level that would rarely be considered for initial screening, which simply emphasizes the truly empiric nature of the technique.

Diffraction data from initial crystallization attempts was problematic in several respects. The resolution of the data was low at 3.2 Å, was highly mosaic, and was highly anisotropic. Additionally, attempts to identify an anomalous signal for SAD or MAD phase determination were unsuccessful. Without anomalous data, a molecular replacement solution was attempted to solve the structure of tbADAT2 using the 3.2 Å data followed by several rounds of automated model building. A solution was found, but the quality of initial maps did not suggest a well-refined model would be attainable. Structural anomalies were present in these initial models that were inconsistent with the core cytidine deaminase fold in general and specifically within the region of the zinc-coordinated active site. All the published structures of bacterial ADAT and hsADAT2 contain an active-site zinc coordinated by the PCIMCA...HAE motif that is very well ordered with continuous detailed density. Even with the low 3.2 Å resolution of this data in mind, the initial models based off the molecular replacement solution for tbADAT2 showed significantly distorted topology around the active site and did not show a clearly identifiable zinc ion that was separate and distinct from the peptide chain. Moreover, an inspection of the remaining regions showed the initial model harbored many regions of poorly built structure and regions that were obviously outside of clear density. While it is possible that manual model building and further rounds of refinement could have improved upon the initial molecular replacement results, it seemed clear that the final solution of such efforts would have yielded very little return and an alternate course was chosen. The values for the unit cell of these crystals, the contents of the asymmetric unit,

and the final space group were all significant data points, but cleaner data at higher resolution were required to move forward.

As a way to increase order in crystals, Lusty *et al*. have shown gentle glutaraldehyde cross-linking introduced into the crystallization experiment through vapor diffusion has been quite successful for several test cases.[73] To test whether the poor data quality for the first round of data collection was the result of a similar disordered crystal pathology, tbADAT2 crystals were cross-linked to varying levels; however, no improvement in the quality of diffraction was observed, and in many cases the quality of diffraction significantly decreased.

Another strategy to optimize diffraction was pursued through modification of the crystallization conditions and cryoprotection techniques. Na/K tartrate as a precipitant in the crystallization conditions opened the possibility to use high concentrations of organic-acid salts as cryoprotectants, which given their presence in the crystal initially, would putatively be less disruptive during cryo-cooling and thus improve the data quality over the first crystals protected in 20%glycerol.[72] Bujacz *et al*. have demonstrated how many salts of organic acids – including Na/K tartrate – can be used as cryoprotectants at higher concentrations; however the cryoprotective properties of Na/K tartrate are more difficult to utilize because they are present only concentrations that are close to the salt's maximum solubility at room temperature. In the context of cryoprotection in crystallography, an optimal cryoprotective agent is one that can be easily added to the crystallization conditions in addition to the necessary reagents for initial crystal formation. In the case of ADAT2 crystallized with Na/K tartrate at 4˚C, the crystallization conditions were such that there was not sufficient "room" in the mothor

liquor to achieve sufficiently high Na/K tartrate concentrations by adding a maximally concentrated Na/K tartrate stock solution. Even attempts to reach the desired concentrations using solid reagents proved unsuccessful. Attempts to transfer crystals into the highest possible tartrate conditions or into cryoprotectant conditions containing other combinations of organic salts failed to achieve a satisfactory result using data quality and diffraction resolution as endpoint criteria. One concern with these experiments was the possibility that the transfer into these alternate buffers or even into higher Na/K tartrate conditions was simply too rapid. A solution to address this concern was serial transfers between conditions of gradually increasing concentrations of cryoprotectant; however, this approach yielded similarly low quality results likely attributable to the destabilizing effects of the cryoprotectants in general or specifically to an excessive amount of manipulation of the crystals required for these protocols. After the failed attempts described above, an alternate approach was to perform the transfer to higher salt via a vapor diffusion method that is more analogous to dehydration protocols.

The cryoprotection technique of dehydration has been shown to improve and/or preserve the diffraction quality of crystals to a dramatic extent, and coincidentally has been shown to be beneficial in the recent structure of *S. mutans* deoxycytidylate deaminase, another member of the CD superfamily.[75, 78, 79] A modified protocol was developed that utilized Na/K tartrate as the dehydrant and allowed for minimal handling of the crystals while still using Na/K tartrate as the cryoprotectant. Crystals prepared in this way showed improved diffraction, lower mosaicity, and decreased anisotropy compared with the glycerol cryoprotection approach.

Data Collection and Processing.

Data were collected from more than 30 crystals of tbADAT2 cryoprotected in Na/K tartrate in order to identify the best crystals for data diffraction. Significant heterogeneity was seen across the samples both in terms of resolution and the in the quality of diffraction; however, even the weakest diffraction recorded amongst these crystals was consistently stronger than what had been observed previously, indicating that the better data quality observed in these samples was due to the dehydration/cryoprotection and not simply due to increased screening. Another indication of improved data quality was a detectable Zn anomalous signal. Several structures of ADAT2 homologues have been solved using Zn-MAD phasing methods [30, 31, 75], so the presence of zinc fluorescence in the crystal indicated that a *de-novo* structural solution for tbADAT2 might be possible.

For a data collection strategy, SAD was selected over MAD for several reasons. First was simply a practical consideration for maxiammly efficient use of beamline availability. Second was the observation that only 3-4 of the crystals yielded the best data, so the best strategy was to use the best two crystals to collect one dataset at the Zn-Peak and one dataset at the optimum wavelength for the beam (highest flux), thus measuring datasets at the maximum anomalous signal and the maximum diffraction, respectively. Finally, examples in the literature have recently shown success using Zn-SAD to solve structures of proteins with similar size and with similar Zn/amino acid ratios as found in tbADAT2, as discussed below.[80] Together these factors supported the choice of Zn-SAD as the best possible approach given the experimental scenario.

Structure Determination

Scaling the Zn-SAD dataset revealed a measurable anomalous signal extending to a resolution of 4.0 Å, and the initial phase solution identified 8 pairs of zinc atoms per unit cell (two per asymmetric unit) with a space group consistent with the initial scaled data and with the previous low-resolution molecular replacement solution. The initial phases from the Zn-SAD experiment combined with a higher resolution native data allowed the *de novo* structure of tbADAT2 to be built and refined with reasonable statistics (see below). The application of Zn-SAD in this instance highlights an emerging crystallophic method that deserves note.

Zinc phases in theory

Since the technical advancement of tuneable synchrotron x-ray sources, MAD and SAD phasing have yielded a trove of methods with which to solve crystallographic phases.[81, 82] Both for its ease of use and for its general applicability, the predominant application of these techniques has been through the use of seleno-methionine containing protein samples prepared from methionine auxotroph strains of bacteria. Such an approach can significantly increase the likelihood that an anomalously scattering heavy atom is in an ordered region of the structure. The allure of such "engineered" heavy atom sites is understandable, especially when compared to the relative uncertainty and limited practicality of soaking heavy metal ions into crystals. Nevertheless, difficulties with Se-Met derivatization do arise and can often be cumbersome and time-consuming to address if not altogether impossible. Given the relatively weak strength of the anomalous signal from selenium and the added flexibility of the methionine side-chain, the overall phasing

effectiveness of Se-Met is generally accepted to be no more than 1:100, heavy atoms:amino acids.[83] In proteins with low methionine content or in proteins in which the methionines are located in highly mobile regions of the structure, the MAD or SAD anomalous signal from Se-methionine can be insufficient. Finally, Se-methionoine substituted protein is frequently more difficult to purify and often requires an additional round of crystallization optimization. Of course the process of phasing data in crystallography is an empiric one, and if Se-met substitution doesn't work, researchers will continue with other techniques until a successful approach has been found; however, judging by the relative lack of structures solved by either MAD or SAD techniques using the anomalous signal of zinc, it would appear as though the use of zinc anomalous phasing has been overlooked when in many instances it may have proven quite powerful and perhaps much more expedient. This discrepancy begs the question: why use Se-MAD/SAD when a much more direct solution is available?

Zinc, despite having a similar-strength anomalous signal to selenium, and despite the native presence of structural and/or enzymatic zinc ions in a large number of proteins, is used infrequently with respect to selenium for MAD and even less so for SAD phase determination. There are many cases in the crystallography literature and several within the extant structures of the cytidine deaminase family where Se-methionine phasing has been used despite the presence of one or more structural ordered zinc(s). The reasons for this may be simply historic since it is a trend that appears to be changing as more structures solved by Zn-based phasing are reported in the literature. The crystallography field continues to explore the boundaries of anomalous phasing methods using many non-selenium heavy atoms for MAD and SAD experiments, many of which could be

considered first-line approaches, i.e. they take advantage of the native elements of the crystallized protein. Among these elements zinc is by not particularly unique for any physical property save for its relative abundance, and it is discussed in depth here simply as an example of an emerging trend and specifically because it was utilized in the structure determination of tbADAT2.

A more thorough experimental treatment of this idea has been provided by Dauter, Dauter, and Dodson, who in 2002, revisited fifteen previously solved structures and solved them anew using SAD methods phased off the anomalous signals of many different heavy atoms including two structures with native zinc and seven more with atoms natively incorporated into the crystallized proteins.[84] Dauter *et al*. discuss not only the broad applicability of the methods but also how the computational advancements over the last ten to twenty years – specifically solvent flattening and density modification algorithms – have improved the phasing power of anomalously scattering heavy atoms. Perhaps at the risk of belaboring the point, the results of Dauter *et al*. do not imply this progress in computational power has independently prompted the use of non-selenium heavy atoms, they simply demonstrate how all SAD and MAD techniques benefit from these computational techniques and are included in this discussion to highlight the potential for non-selenium SAD/MAD phasing.

The Fu group at Cornell have recently made careful studies using zinc-based phasing techniques to demonstrate the applicability of its use in *de novo* structure determination, the results of which show that in addition to being abundant zinc can also be quite powerful as a phasing tool relative to other elements. Meyer *et al* in 2006 and 2009 have demonstrated the use of zinc MAD[85] and SAD[86] phasing to improve the resolution of

RNA polymerase II to 4.1 and 3.8Å, respectively. These efforts demonstrate an impressive experimental phasing effectiveness of 1Zn:550 amino acids, which according to the authors' theoretical calculations could potentially extend to 1:1100, almost an order of magnitude beyond the power of Se-methionine. These authors further highlight several recent structures that have used Zn-MAD methods successfully with phasing effectiveness extending to 1:368 in other cases. A literature search identified several additional examples that have utilized Zn-SAD with phasing effectiveness ratios in the range of 1 zinc2:00-300 amino acids.[80, 87, 88] Again, these results are discussed alongside the Zn-SAD approach in the present tbADAT2 structure not to identify a new technological contribution, rather they are presented to demonstrate how this is an underutilized technique that will likely become much more common as the ease of use and overall phasing power available through zinc phasing is appreciated. Phasing with zinc is easily accessible at synchrotron-wavelength x-rays, and it is a technique that should be considered more frequently as a tool for *de novo* crystallography.

Zinc phases in practice, the structure of tbADAT2

For tbADAT2, the initial zinc phases yielded interpretable electron density maps that allowed 157 of the 225 residues of each monomer to be built through a combination of automated and manual model building. Density for the model was discontinuous with two significant breaks occurring at points that corresponded to insertions in the tbADAT2 sequence. (Fig. 5.4) Many attempts were made to build these regions; however, each attempt resulted in a model that 1) failed to improve the overall refinement stastics and 2) harbored problematic geometric statistics that were unresolvable. The final model

represents density that could unambiguously be interpreted using 2Fo-Fc maps contoured at 1σ

The phases of this initial model were refined against higher resolution data with the intention of improving the refinement of the model and extending the buildable portions of the model at the border regions of the discontinuities in electron density. These attempts succeeded at the former but not the latter of these two objectives. The maps resulting from these efforts showed small improvements in the sharpness of side-chain detail in some regions of the density, but overall the difference was limited. The main effect of the high-resolution data was to improve the overall refinement statistics of the structure. This process was not direct and involved several different approaches to data processing and refinement. The final combination of anomalous refinement on anisotropically scaled data with the use of TLS parameters was built up stepwise to produce for the final refinement strategy.

First among these approaches was the decision to treat the high-resolution data as an anomalous dataset. This data was collected at the optimum wavelength for peak x-ray flux for the X29 synchrotron beamline at NSLS, which at 1.070 Å lies in a region of the x-ray spectrum in which zinc atoms have a measurable anomalous signal. Typical scaling of this data using an I/σI cutoff of 2.0 resulted in an overall diffraction limit of 2.56 Å. Rescaling the data to include an anomalous signal decreased the overall resolution of the dataset from 2.56 to 2.69Å, but anomalous refinement with this data did show improvement in the overall refinement statistics. Visual inspection of the diffraction data had revealed diffraction to nearly 2.4Å, so the need to truncate this dataset at 2.56Å and then at 2.69Å for anomalous scaling prompted a more detailed

analysis in an effort to reclaim some of the higher resolution data that had been lost to truncation.

Anisotropic diffraction was one of the original problematic aspects of the initial tbADAT2 dataset at the initiation of this project. While the anisotropy in subsequent data was definitely improved by the cryoprotection techniques above, anisotropy was still evident in the high-resolution data and was significant enough to be visibly detectable, e.g. obvious diffraction extended to 2.4 Å in the strongest dimension and to about 2.7 Å in the weakest. Strong *et al*. have developed web-servers for the application of techniques they developed to address an extreme case of anisotropic diffraction they recently encountered.[48] Briefly the calculation applies an ellipsoidal truncation to the diffraction data and applies a negative thermal (B) factor to the remaining data to accommodate for global effect of the lost data, so the typical *sphere of reflection* with a radius of 1/resolution (Å) now has ellipsoidal dimensions of $1/\text{resolution}_{(a)}$ (Å) x $1/\text{resolution}_{(b)}$ (Å) x $1/\text{resolution}_{(c)}$(Å). (Fig. 7-1A) The difference in volume of these two shapes is the basis for an artificially low data completeness seen for the anisotropically truncated data used for the structure determination of tbADAT2 (Table 3.1) and is similar to statistics seen in the deposited data from Strong *et al*. A rigorous method for calculating data completeness within the ellipsoid was not available, however an approximation based strictly on the volume of the sphere of reflection for ideal data with a resolution of 2.5Å (r=1/2.5) and an ellipse with dimensions corresponding to a resolution of 2.8x2.8x2.5 (axbxc =1/2.8x1/2.8x1/2.5) yields a ratio of 1:0.77 V_{sphere}:$V_{ellipse}$, (Fig. 7-1B) which would indicate that a value of 80.1% calculated for this

data equivalent to a nearly complete dataset for the data within the limiting boundary of the ellipse.

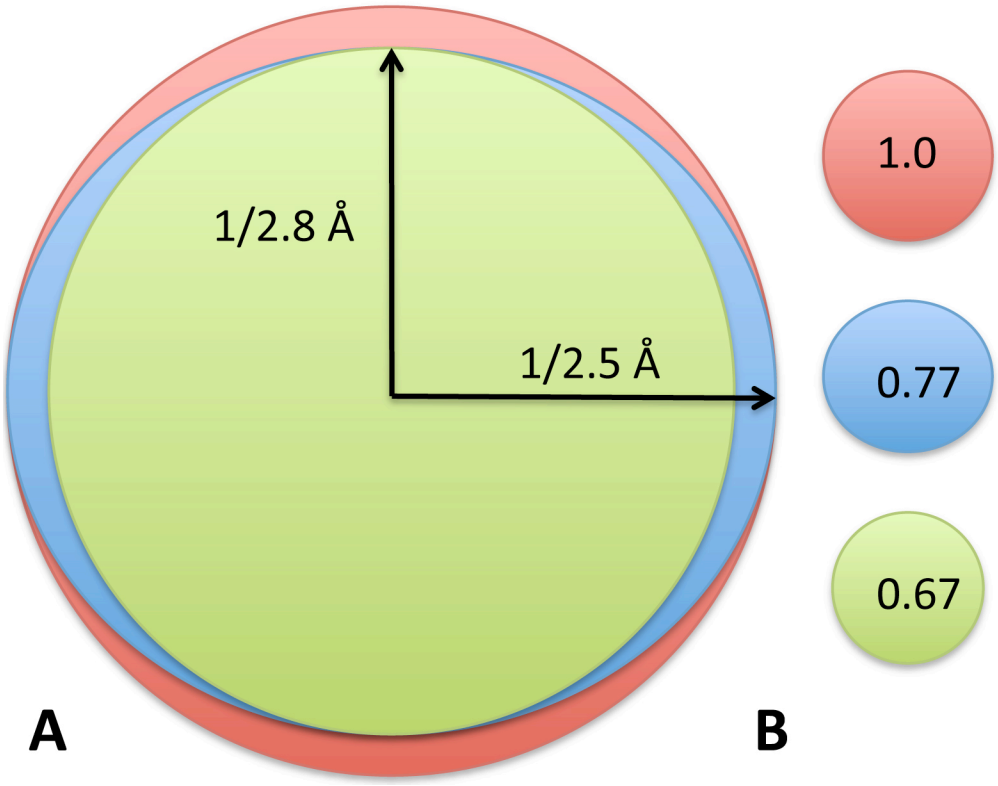


Figure 7-1. Schematic of anisotropic data truncation showing **A**) the differences in cross section between spheres of reflection with radii of 1/2.8Å (green) and 1/2.5Å (red) and an ellipsoid (blue) with dimensions 1/2.8Å x 1/2.8Å x 1/2.5Å cut along the longest dimension. **B**) The ratio of volumes between the three shapes is shown in B, with all volumes relative to the largest sphere.

Strengthening this conclusion is the calculation for completeness (>99%) for a normal spherical truncation of these data at 2.8Å. The anisotropic truncation can be performed on native or anomalously scaled data, so here the data were first scaled anomously to 2.5Å and then truncated ellipsoidally. This approach maintained the beneficial refinement effects of the anomalous data while also preserving as much of the collected data as possible. Combined, these two strategies yielded a 4-5% improvement in R_{free}.

The third parameter of refinement focused on optimizing the model used for refinement as opposed to the reflection data. Given the asymmetric unit comprised a single homodimer, initial refiment parameters were utilized that applied non-crystallographic symmetry to the model. This approach showed promising refinement statistics, so the additional technique of TLSMD analysis (torsion, libration, and screw motion determination) was utilized in which the starting model for refinement was analyzed for rigid body segments within the structure that have equivalent TLS motion, e.g. an α-helix, or a pair of β-sheets. The definition of these segments was then used for TLS refinement prior to secondary rounds of restrained refinement.

The final combination of these TLS refinement against the high-resolution dataset scaled anomously and truncated anisotropically was used for the final rounds of refinement, which resulted in a final model with an R_{free}/R_{work} of 29.6/25.8.

Structure analysis

The asymmetric unit of the tbADAT2 crystal structure was found to be a symmetric homodimer, a result in agreement with predictions based on gel-filtration data (above and [89]). As was expected from initial sequence homology, the tbADAT2 structure closely resembled the hsADAT2 structure and exhibited identical topology to the published bacterial ADAT structures in the region of the canonical CD fold. Structural alignments of the individual tbADAT2 monomers with bacterial ADAT and hs ADAT2 reveal r.m.s.d values of 1.3 and 1.9 $Å^2$, respectively. The dimer interface of tbADAT2 also shows significant similarities with the bacterial and human ADAT structures. As in the published ADAT structures, the tbADAT2 dimer interface is formed largely by a

symmetric interaction between the three α-helices in each monomer that contain the zinc-coordination motif; however, the specific orientation of the tbADAT2 interface differs from the known bacterial ADAT and hsADAT2 structures. It is important to note that some variability exists with regard to the dimer interface across all the known ADAT structures. While no specific attribute of ADAT function has been correlated with this observation, it is interesting to note that relative to one another, the tbADAT2 dimer interface is rotated to the largest degree in pairwise comparisons between the known structures: i.e. when a structural alignment of all the ADAT dimer structures is performed using just one monomer of each dimer, the specific monomers align very well, but the unaligned monomers of each of the dimers are rotated to a degree compared to the other structures with the ultimate effect being the separation of the two active sites in the monomer relative to the least rotated structure. The significance of this rotation awaits further study, but it could be part of a structural reason for tbADAT2 inactivity towards tRNA, or it could serve as the basis for a mechanism to accommodate non-tRNA substrates.

Like the hsADAT2 structure, the tbADAT2 structure shows a highly conserved and physiological zinc binding motif when compared to other bacterial ADAT structures. This would seem to negate initial speculation that the reason for inactivity is due to a distorted or somehow altered catalytic site. (Fig 6.1C) There is no structural reason evidenced in the catalytic core for either tbADAT2 nor hsADAT2 to be inactive. While the tbADAT does not show a catalytic water coordinated to the zinc, the lack of a refined catalytic water in the present structure is most likely attributable to the relatively low resolution of the diffraction compared to the structures to which it is being compared.

The only distinctive feature within the active site of tbADAT2 is the presence of Tyr145, which is contributed to the active site from the opposite monomer. The presence of a residue in this same location is typical in these structures (Fig. 7.2), but tyrosine in this case stands out as it is bulkier, less hydrophobic, and is oriented such that an interaction with the 2'-hydroxyl of the ribose could be possible. Again, the significance of this particular residue is moderated by the observation that the hsADAT2 contains a small, more hydrophic leucine in this position and is still inactive toward tRNA.

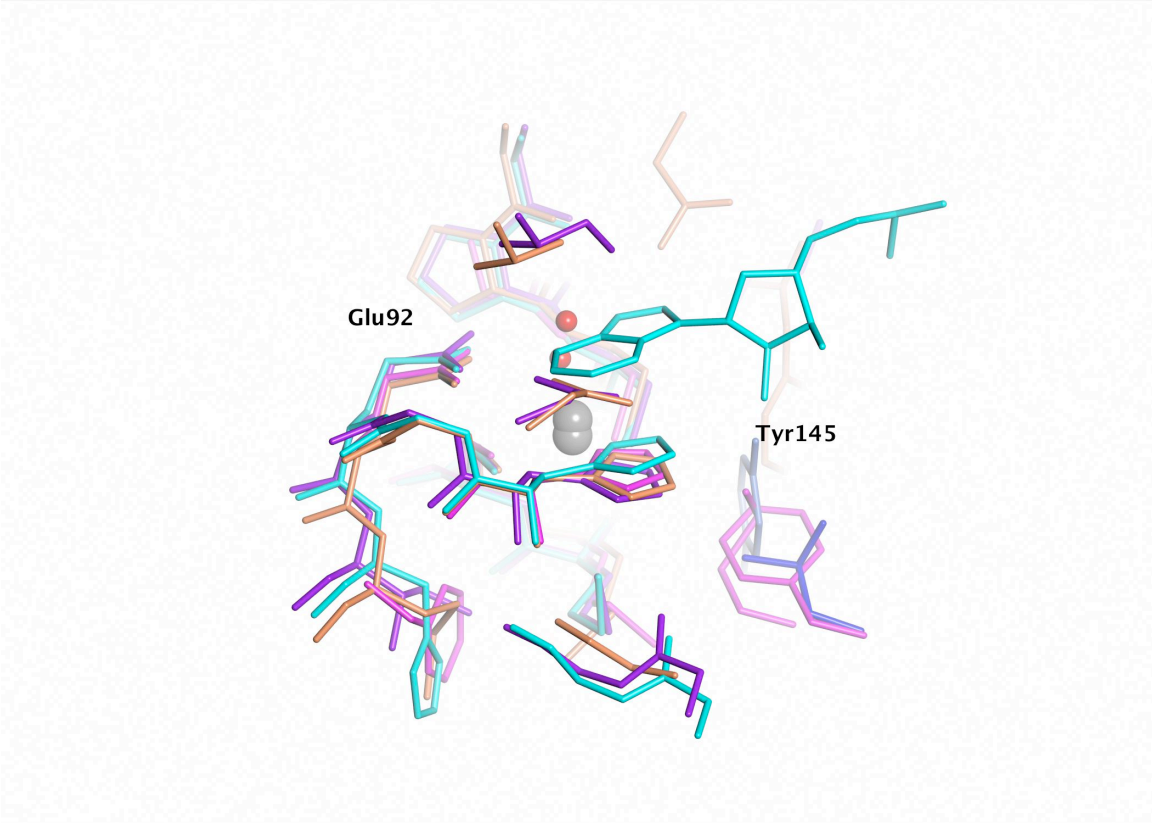


Figure 7.2 Active site of tbADAT2 showing presence of Tyr from the opposing monomer. tbADAT2 is in magenta and Tyr is light blue and labeled as such. Additional aligned structures are shown for comparison: saADAT2 with substrate (cyan), atADAT2 (orange) and hsADAT2 (purple). Stereo view in Appendix I.

The second possible explanation for inactivity involves the affinity for the substrate. A structural comparison with the saADAT structure bound to stem-loop substrate analogue shows that several residues involved in RNA binding have undergone non-conservative mutation in the tbADAT2 (Table 6.1, Fig. 6.3B) The mutated residues in tbADAT2 correspond to residues in the saADAT structure that provide both specific and non-specific interactions with the RNA molecule. The importance of these residues in the saADAT structure is clear as they provide very sequence-specific interactions with the RNA substrate; however the importance of these contacts in ADAT2 is still an open question. It is clear that several key residues are not conserved in the tbADAT2 structure, and that the more extreme mutations – E-to-A and R-to-G – lie on the face of the opposing monomer to the one binding the catalytic zinc. (Fig. 6.3B on the right in cyan) Collectively these mutations support the conclusion that ADAT2 has evolved away from the substrate binding strategies of prokaryotic ADATs; however, it is important to note that it cannot be concluded that these mutations are the reason ADAT2 no longer has affinity for tRNA substrates. It has been shown recently that mutations in several of the conserved substrate binding residues, specifically E92 and K159, in the context of the ADAT2/3 heterodimer did not affect tRNA binding.[33] It is also worth noting that the region harboring the E-to-A and R-to-G mutations mentioned above would be replaced by ADAT3 in one of the two active sites in the heterodimer. From these results it is fair to conclude that bacterial ADATs and the ADAT2/3 of higher eukaryotes utilize strikingly different strategies to bind substrate, which would appear to have made the conservation of the residues involved in RNA binding in bacterial ADATs unnecessary for the purpose of binding tRNA. A consequence of this, however, is an ADAT2 enzyme

that no longer binds tRNA substrate due to an as-of-yet unknown reason. It seems plausible that it is due to the mutations alone, and an interesting proof-of-principle experiment would be to attempt to restore tRNA binding in tbADAT2 by restoring these mutations to the consensus residues in the ADAT structures. This would account for the same loss of tRNA binding in both the hsADAT2 and tbADAT2 homologues and would be a simple experiment to attempt in both cases. This explanation alone, however, does not leave much room to explain the novel functions that have been attributed to ADAT2 and the ADAT2/3 heterodimer.

To futher investigate these novel functions – specifically the *in vivo* C-to-U editing at position 32 of tRNA and *in vitro* C-to-U editing of DNA – the most likely candidates from a structural approach are the three distinct insertions in the tbADAT2 sequence. (Fig. 6.2A). These insertions are unique to the trypanosomatid enzymes and therefore stand as good candidates in which to identify a structural basis for these unique enzymatic features and may, additionally, suggest mechanisms for substrate recognition and the purported domain swap.

Two of these three loops are found on the face that is oriented toward the approaching substrate, and of these, only loop A is ordered. (Fig. 6.2B-C) The high thermal factors of loop A along with the lack of specific secondary structure and the lack of ordered electron density for loop B both indicate these two sub-domains are highly flexible, and given the length of both of these loops, have the capacity to sample a large range of conformation space near the catalytic site of the enzyme. With the proximity of loop B to the active site, this flexibility could easily be a mechanism by which substrate is

prevented from binding, and when open could create an extensive binding surface accommodating a wide variety of substrates.

The APOBEC cytidine deaminases are a group of enzymes that have been of intense focus due to their involvement with a number of important biological functions mediated by modification of DNA and RNA in higher mammals. Recently several structures of this family have been determined through x-ray crystallography and NMR, and while insightful, the mechanisms for substrate recognition and binding have yet to be completely elucidated.[90] APOBEC-3G is an interesting example of this family that binds ssDNA and RNA and provides innate viral immunity by interfering with viral replication within the cell. The structure of APOBEC-3G was recently solved by X-ray crystallography.[91] This structure revealed a series of loops on the surface adjacent to the active site zinc of APOBEC-3G, which have been proposed to create a DNA-binding cleft oriented towards the active site zinc. A structural alignment of APOBEC-3G and tbADAT2 clearly shows that these loops are oriented in much the same way as the loops of tbADAT2 (Fig. 7.3). APOBEC and ADAT families are not directly related and belong to different clades of the CD superfamily, so any similarity likely represents two independent evolutionary adaptations; however, the similarities are intriguing and certainly provide a basis for the possibility that the tbADAT is an active enzyme for which the substrate has yet to be identified. Moreover, the presence of APOBEC proteins in higher mammals could explain why hsADAT2 does not show the same modifications as the tbADAT2 of this study, since these enzymatic functions have been accommodated in the human cell by additional enzymes that are notably absent in lower eukaryotes.

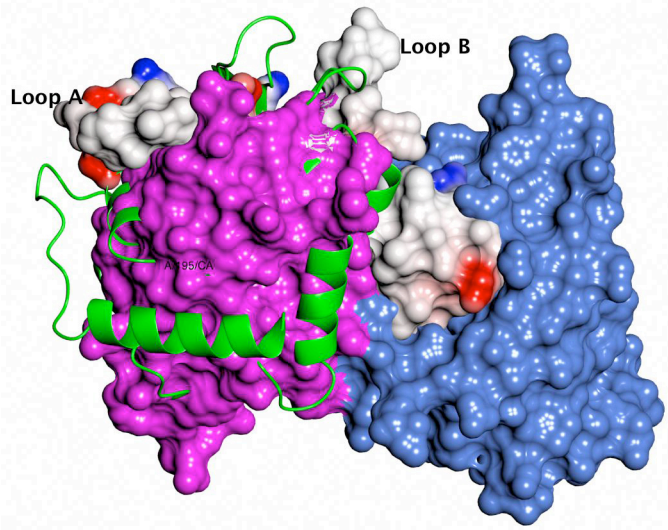


Figure 7.3 Alignment of APOBEC-3G and tbADAT2. APOBEC-3B is in green, tbADAT2 in magenta and blue. Loops A bnd B are colored for electrostatic potential as are the active site residues within 7Å of the active site zinc of the blue monomer. Stereo view in Appendix I.

Another interesting possibility for the function of these loops relates to the proposed domain swap model proposed for tbADAT2. The domain of tbADAT2 purported to swap with an analogous domain from ADAT3 begins in the middle of the disordered region of loop B. The original models for the domain swap were created using the hsADAT2 as a starting model, which specifically does not have the extra loop in this region. Consequently, these models show the stretch of amino acids connecting the swapped and un-swapped portions of hsADAT2 in an almost linear configuration with very little freedom of motion. (Fig. 1.3B) tbADAT2 has ten additional residues in this region, and therefore, it would seem to be able to more easily accommodate this domain swap with less strain on the intervening loop. Another interesting consequence of such a domain swap occurring at loop B would be the minimization of this loop as it is stretched from one monomer to the other resulting in a significant difference in the ADAT2

binding surfaces in these two conformations such that substrate binding by tbADAT2 in the homodimer would be significantly altered in the heterodimer configuration. Given that this loop is not structured in the present model, the possibility that a domain swap occurs in the tbADAT2 as well can't be definitively rejected. A series of disulfide cross-linking experiments have been designed and are presently being pursued, but the preliminary data from these experiments has yet to show any conclusive results.

Of the structural features identified in the tbADAT2 structure, loops A and B are of the greatest interest as they are both situated to potentially alter the substrate binding surface of the protein and/or participate in a domain swap. Several powerful tools are available to test the degree to which these loops are essential for activity and will be the future aim of this project. First would be to assess the importance of each loop independently both *in vitro* and *in vivo*. Many of the published experiments could easily be revisited using recombinantly expressed ADAT2 either alone or in the context of ADAT3 with varying levels of alteration to either of these loops. Since ADAT2 is essential in *T. brucei*, viability studies could be performed on strains that had a mutant ADAT2 allele with subsequent insertional knockouts or inducible RNAi downregulation of the wtADAT2 allele. These experiments would be further complimented by a more careful study of other examples of eukaryotic ADAT2 and ADAT3. Activity profiles of non-*T. brucei* enzymes are essential before a definitive role can be assigned to any particular domain. To date much of what is know about tbADAT2/3 has not been shown absent in other ADAT2/3 enzymes. The ability to clearly identify features of these enzymes will require a much more careful analysis of representative members of the eukaryotic ADAT family,

with the hsADAT2/3 enzymes being the most instructive specifically due to the absence of the loops A and B in the hsADAT2.

Another critical focus of continuing research will be the determination of eukaryotic ADAT2 homodimer function. The lack of a known function in eukaryotes in the presence of such a clear function in the prokaryotic homologues presents an interesting and complex question. The structures of hsADAT2 and the above tbADAT2 both demonstrate that eukaryotic ADAT2 clearly has the capacity to form a homodimer with an interface highly similar to bacterial ADAT and nearly identical in topology and buried surface area. The details of the dimer interface along with the demonstrated *in vitro* stability of the homodimer necessary to obtain the crystal structures of these proteins provide indirect evidence that a stable homodimer could exist *in vivo*, something for which direct evidence is still lacking. The function of such a homodimer remains to be clearly illucidated; however, the structure of the tbADAT2 homodimer presented above clearly exhibits features of the canonical cytidine deaminase active site with appropriate zinc-coordination and side-chain geometry. Save for occlusion of the active site by the A and B loops discussed above, there seems to be no structural reason to suggest that the tbADAT2 would be unable to catalyze substrate. Whether that substrate is a modified tRNA, or some other nucleic acid substrate remains to be determined. Several approaches to identifying such a substrate are currently underway, including attempts to identify possible DNA and RNA substrates.

The structure of tbADAT2 homodimer presented above reveals an enzyme harboring the canonical cytidine deaminase core fold and an otherwise normal functional appearance with no obvious indications as to why this enzyme is inactive. The structure suggests

several interesting possibilities as to why substrate binding is impaired in this structure including a rotated dimer interface and the presence of unstructured loops adjacent to the active site, but explanations for the inactivity of the enzyme remain elusive. The presence of these loops provides an intriguing possibility that structural features unique to tbADAT2 allow for a more diversified substrate profile, perhaps via mechanisms similar to those proposed for the AID and APOBEC family of proteins. The function of loop B may also be to facilitate a putative domain swap when hetero-dimerizing with ADAT3, allowing the constitution of a functional active site with the essential Glu92 proton shuttle of ADAT2 with the essential zinc binding residues of ADAT3. Together these results provide the guide for several interesting lines of investigation toward additional substrate identification in the homodimer as well as a helpful model with which to assess the function of the ADAT2/3 heterodimer. Significant work remains in order to determine the degree to which the results in tbADAT2/3 are unique to *T. brucei* vis-à-vis other eukaryotic ADAT2 and ADAT3 homologues. Specifically, are the alternate substrate profiles seen *in vitro* for the *T. brucei* enzymes reproducible with ADAT2/3 from other organisms? The hsADAT2 homologue does not have the same insertional loops as tbADAT2 but is otherwise highly homologous to the tbADAT2, so there is an interesting possibility that the results from *T. brucei* are indeed unique and attributable to the loops identified above. An important question will be to determine if the possibility of a domain swap is unique to *T. brucei* or a more widespread phenomenon in eukaryotes. The suggestion that loop B facilitates a domain swap due to its added flexibility could be strengthened by evidence that such a swap is not indicated in other organisms and could provide a clear path of experimentation to further understand the

mechanism of this fascinating possible mechanism. Finally, a full understanding of many of these questions awaits greater structural detail of ADAT3. No structural data is available to date for ADAT3 in any organism; furthermore, in addition to the domain of ADAT3 that is highly homologous to ADAT2, ADAT3 possesses an additional 160 residue N-terminal domain that has very low sequence similarity to other proteins. The structure and function of this additional domain is completely undefined and will be an important piece of data in the global understanding of the questions raised in the above discussion. Initial attempts to explore the solubility of this N-terminal domain have been successful and suggest that structural studies of this domain can be pursued in parallel with studies of ADAT2/3.

Appendix I

Stereo Views. The following are stereo views of figures from the results and discussion. They are numbered and arranged according to their appearance in the text. Figure legends are found in the body of the text.

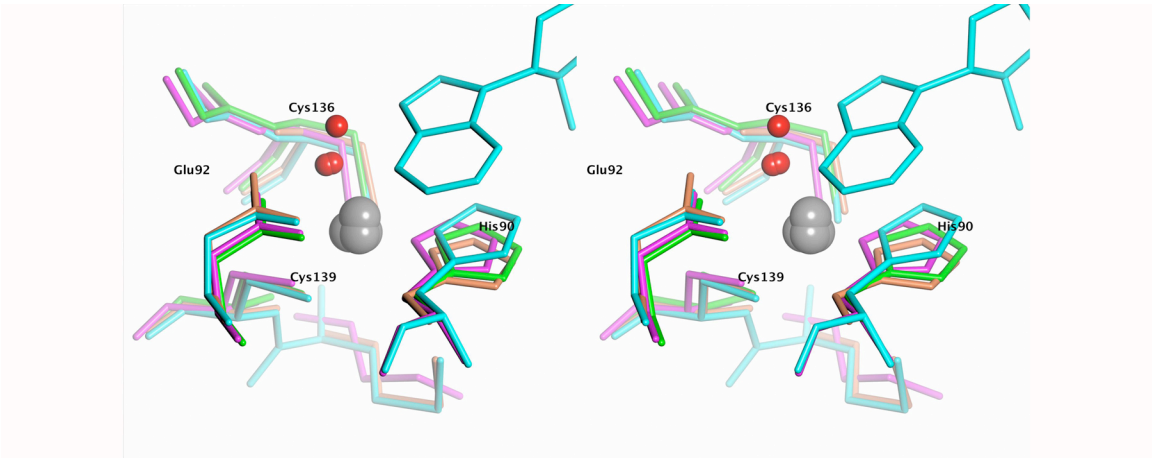


Figure 6.1C.

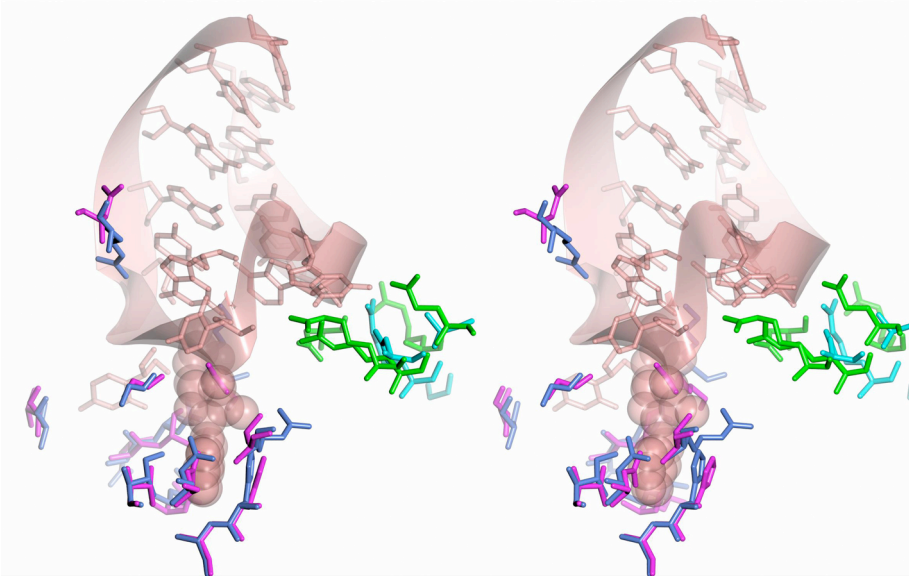


Figure 6.4

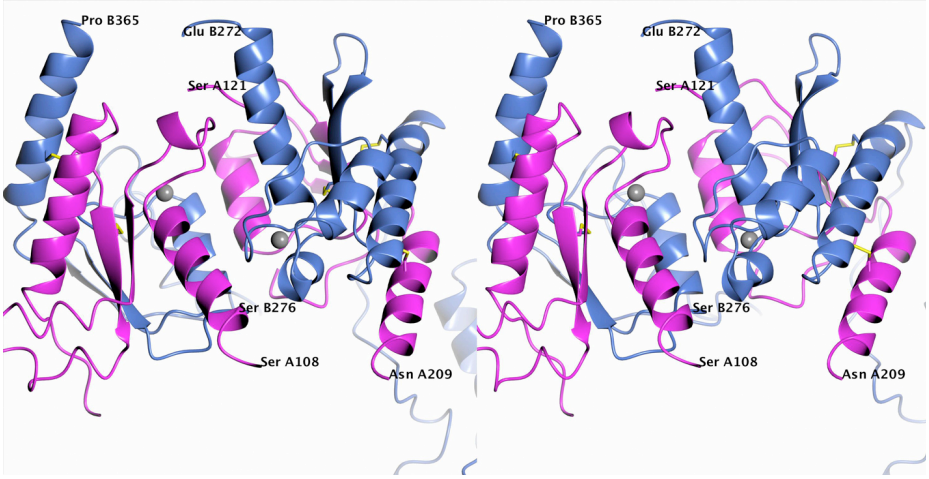


Fig 6.5 No-Swap

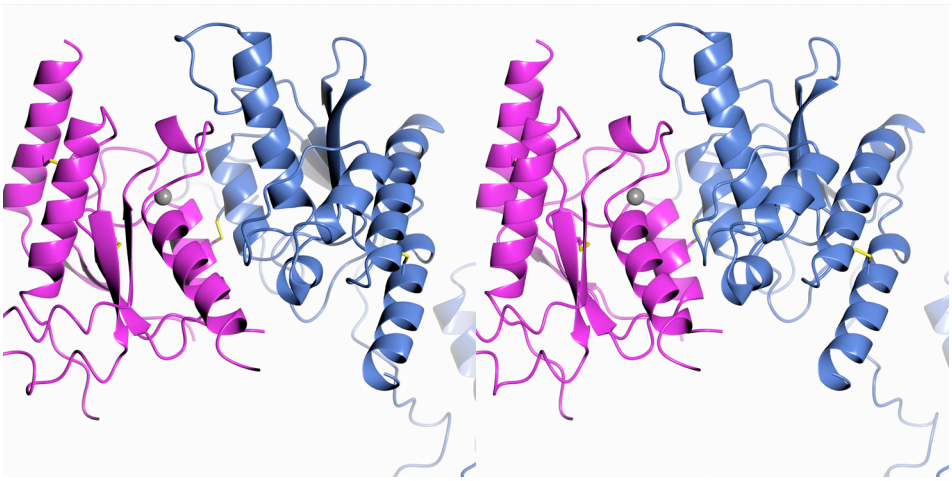


Fig 6.5 Domain Swapped

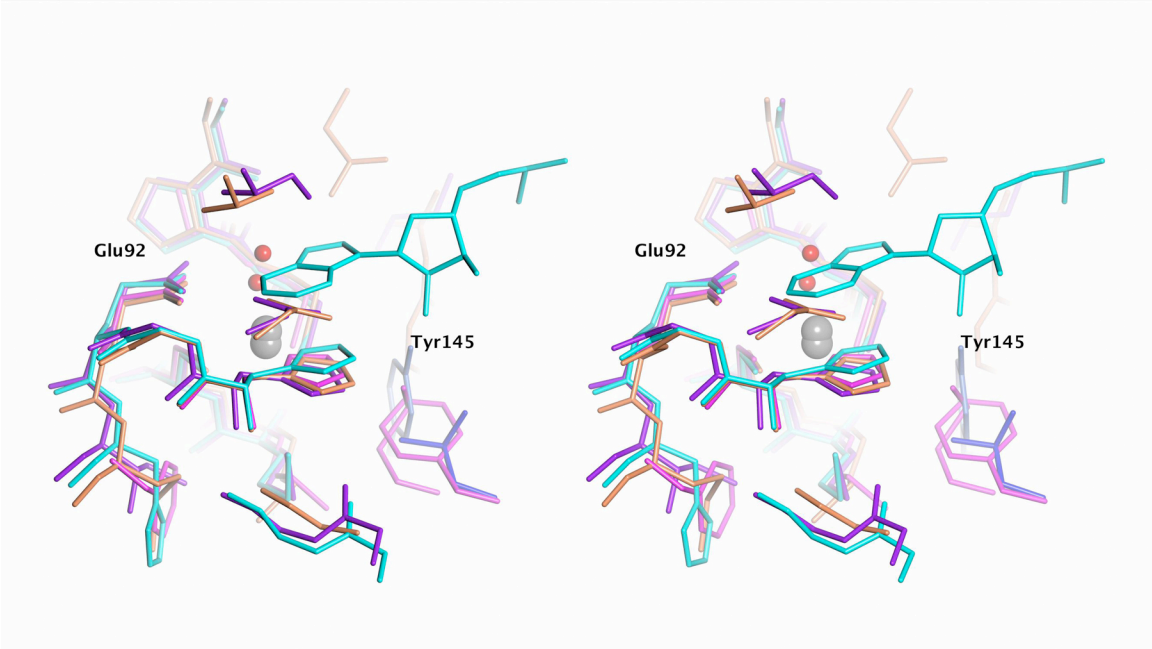


Fig. 7.2

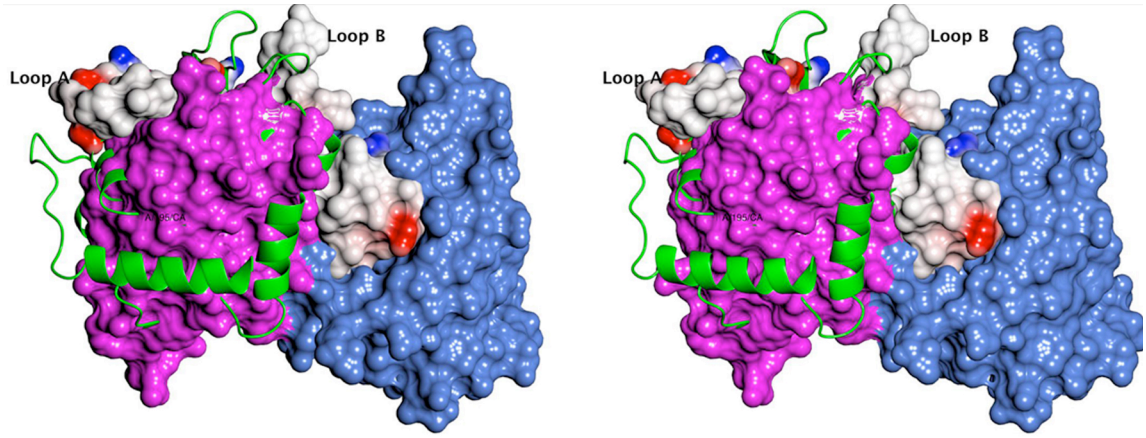


Fig 7.3

Appendix II:

The following is a summary of three additional projects that have formed the balance of my PhD research in the Stebbins lab. All three address issues of the structure and function of virulence factors from bacteria, exploring questions of how bacterial virulence factors function and how this function informs our understanding of the cells and organisms these bacteria infect. The first, work with Senior Research Associate Dragana Nesic in the Stebbins lab, resulted in the crystal structure of *Helicobacter pylori* virulence factor CagA bound to and inhibiting the human MarkII kinase.[92] For this project, my contributions will be presented in the results section and discussed in the context of the work Dr. Nesic and others contributed to this paper. The second and third projects involve the virulence factors SopB and YpkA from *Salmonella* and *Yersinia*, respectively. These projects did not result in publishable results to date. They will be presented with individual introductions, results, and discussions.

CagA from *H. pylori*

Introduction

Helicobacter pylori is a ubiquitous stomach pathogen found in humans at a prevalence ranging between 20-90% in regions across the globe, and in total it is believe to infect around 50% of the human population. *H. pylori* are unique in their ability to colonize the stomach where the resulting infection is typically asymptomatic. This observation was the basis for the contention against Warren's and Marshall's claim that these bacteria were a causative agent in peptic ulcer disease and a significant risk factor for the

development of gastric carcinoma, claims that proved to be correct and for which the pair were awarded the Nobel Prize in Physiology or Medicine in 2005.[93, 94]

H. pylori utilize a type IV secretion system to transport the 120-145 kDa cytotoxic associated gene A (CagA) protein into host cells [95], which increases the liklihood and severity of the gastrointestinal pathology of this pathogen.[96] Infection with $CagA^+$ *H. pylori* strains correlates with the incidence of premature myocardial infarction[97], atherosclerotic plaque build-up and stroke[98], and microalbuminuria in type II diabetes[99]. The direct link between the action of CagA and these pathologies is an area of active research focused on the many different cellular targets of CagA (for review, see Murata-Kamiya 2011).[100]

A major focus of CagA research has been to understand the function of the C-terminal region of the protein harboring the so-called EPIYA repeat domains (named repeat A-D), which vary in number and combination in a strain-specific manner.[101, 102] These repeats interact with the partition-defective and microtubule affinity-regulating kinase (PAR1-MARK) family of protein serine-threonine kinases.[103] PAR1 has a role in maintaining tight-junctions and cell polarity, so the disruption of normal PAR1 function has the potential to result in a significant disruption of normal endothelial function.

Results:

Dr. Nešic identified a 120 amino acid domain of CagA comprising EPIYA repeats A, B, and C, which formed a stable complex with PAR1 and crystallized at 10-25 mg/ml via vapor diffusion when reductively methylated and set up in 2:1 drops over 22.5-

25% polyethylene glycol (PEG) molecular weight 3350 Da, 300 mM Li2SO4, and 100 mM Bis-Tris pH 5.8-6.2, at 23ºC. Initial diffraction was recorded to ~2.8Å on an in-house x-ray source, and the diffraction data showed a very tightly spaced pattern suggesting a very long unit cell edge. This data proved very difficult to index and process, and closer inspection of the diffraction pattern revealed what appeared to be duplication of the pattern instead of a true long unit cell edge. The initial cryoprotection protocol of direct transfer into the crystallization buffer containing 5-10% PEG 400 and 7% glycerol was modified by introducing a step-wise increase, my introducing varying lengths of incubation at each step, changing the number of steps, and changing the method of cryo-cooling, e.g. LN2, cold-stream, liquid ethane. These various techniques did not significantly alter the diffraction pattern. Assuming a long unit cell, several data collection strategies were attempted with the aim of orienting the crystal in a plane perpendicular to rotation axis of the goniometer. These included the use of arc goniometers in-house and the use of a kappa offset goniometer on beamline X29 at the synchrotron. These results did not change the general diffraction pattern, which strengthened the conclusion that gross crystal morphology was the likely source of the poor data quality.

The crystallization drops for CagA-PAR1 were heterogeneous with the supposed "best" crystals representing only 10-20% of the total number of crystals in each drop. These best crystals were used for the above experiments and were selected for size (~0.2-0.3mm in all dimensions), regularity, and thickness. The conclusion above regarding crysal morphology prompted the screening of many different crystal sizes and forms in the initial crystallization drops. The same cryoprotection techniques as above were again

used in various combinations and high quality diffraction data without the tight splitting was obtained from crystals that were square thin plates with edges ~0.15mm. Several crystals of this form were cryoprotected by serial transfer into the above cryo-buffer in 3-4 step-wise increments and flash-cooled in LN2. Diffraction data from these crystals showed diffraction to 2.2Å that was regular and without the splitting morphology seen previously.

***Discussion*:**

The diffraction data collected at the synchrotron allowed the structure of CagA-PAR1 to by solved by molecular replacement using the previous PAR1(PDB ID 2HAK)[104] structure as a search model. The structure revealed that all but 14 residues of the CagA sub-domain were ordered in the structure and formed an extended coil that bound the active site of PAR1 as a substrate mimic. These results were surprising for several reasons, first of which was the lack of the EPIYA motif in the bound sequence of interpretable density and for lack of density for the remainder of the CagA peptide. Analysis of the crystallized protein by SDS-PAGE confirmed the full CagA construct was present in the crystals, but whether it interacts in a specific way with the kinase remains unknown. Structurally, the results were quite interesting as CagA was shown to trap PAR1 in a pseudo-active state in which the lobes of the kinase hinge toward one another and the activation loop of the kinase adopts an activated conformation as though Mg^{2+} and ATP are present, despite the lack thereof. Together these results show how a very small domain within CagA is sufficient to effectively inhibit PAR1

Type III Virulence Factors SopB of Salmonella and YpkA of Yersinia

Introduction to Type III Secretion Systems

Numerous species of Gram-negative bacteria exhibit a syringe-like organ likely evolved from the flagellar apparatus that facilitates complex interactions between the bacterium and its specific host and is essential for virulence, pathogenicity, and in some cases even symbiosis.[105] Termed type III secretion (TTS), this mechanism of virulence was first identified 21 years ago when evidence for secretion of virulence determinants in *Yersinia spp*. was shown that was independent of either N- or C-terminal signal-sequences and lacked *sec*-dependent signaling. This secretion activity was found to rely on the expression of a large operon consisting of at least 13 genes – a list that would later be expanded to more than 20 – and be found to comprise the proteins necessary for the formation of the syringe-like needle complex.[106-108] In the first years following the initial discovery, TTS was quickly identified in many well-known Gram-negative bacterial species that cause significant disease in humans, domesticated animals, and essential food crops.[109] Much about TTS has been learned in the intervening years about both the function of the secretion apparatus and about the virulence factors that are transported via this mechanism. A web-based database has been established to, among other things, simply track and normalize the nomenclature of this field of research as it continues to grow. A quick inspection on the database of organisms that exhibit TTS and the number of TTS virulence factors in each of these species highlights the impressive complexity and diversity of host-microbe interactions that are mediated by this mechanism:

Salmonella spp., for example, utilize two different TTS systems, each for a unique purpose and each with its own set of virulence determinants; *Chlymadia spp*. transport more than 25 different virulence factors into the host cell via TTS, the pan-genome of Pseudamonas syringae contains 57 families of TTS virulence factors, and interestingly, *Rhizobium spp*. utilize TTS systems to facilitate the well-known symbiosis between these species of bacteria and leguminous plants.[105, 110]

TTS is purported to have arisen from a distant genetic duplication event of genes comprising the flagellar apparatus.[111] Significant similarities have been observed at the genetic, structural, and organizational level between TTS and the flagellar apparatus. These similarities were the main basis for the early modeling and purported functional mechanisms of TTS,[112] and evidence of significant cross-talk between the flagellar and TTS systems strengthens such a comparison.[113] Both structures exhibit a basal complex anchored in the bacterial inner-membrane that contains within it a second structure that spans the inner and outer bacterial cell membranes and is itself traversed by a narrow central pore only 2-3 nm wide. In the flagellar apparatus, flagellin subunits purportedly transit through this pore and emerge at the surface of the growing flagellum leading to elongation of the flagellar tail.[114] In the case of TTS, instead of a flagellar tail, the basal body in the bacterial inner membrane connects to the needle apparatus, which unlike the flagellum, is straight, narrow, and typically much shorter. The needle component of TTS is fairly consistent between species varying between 45-60 nm and is capped at the needle tip by a putative homo-pentamer that facilitates docking with host cell membranes.[115, 116] (Fig. Appendix II-1) Much like the case of the flagellar apparatus, the needle appendage purportedly grows by transporting the needle subunit proteins through

its central pore to the lengthening tip of the appendage. Once the appropriate length has been reached, the needle complex is capped and the specificity for transported substrates switches from subunits of the needle itself to the TTS virulence factors. (see IpaB-D in Fig. Appendix II-1) It has been shown that specific TTS components are involved in both the control of needle length and the control of substrate specificity, although the strength of the interaction between these two mechanisms and the specific mechanism of signal transduction back and forth between the needle tip and the basal body remains to be determined.[117] Such a mechanism purports that a ruler-like protein somehow measures the growing needle (see Spa32 in Fig. Appendix II-1), and once the appropriate length is achieved, a substrate-specificity switch occurs causing the growing needle to be capped by a host-cell docking complex and then for virulence factor transport to begin once successful docking with the host cell membrane is established. Such events would seemingly require signaling events between the needle tip and the basal body to trigger the switch in substrate specificity to stop needle growth and initiate cap formation and then again to initiate virulence factor transport upon host-cell contact. How these events are regulated is presently unknown. They raise fascinating questions about what types of signal transduction could be possible in this context as well as what structural mechanisms are utilized to control the flow of not just proteins but of ions and water. Electron micrographs of bacteria actively expressing TTS needle complexes show multiple needles studding the surface of the bacterium. It is interesting to consider mechanisms that could prevent these needles from serving as a conduit for ions and small molecules between the intra- and extracellular milieu while still maintaining the ability to

transit large proteins; however, such mechanisms remain strictly speculative as-of-yet as no structural data with adequate resolution to address these questions has been published.

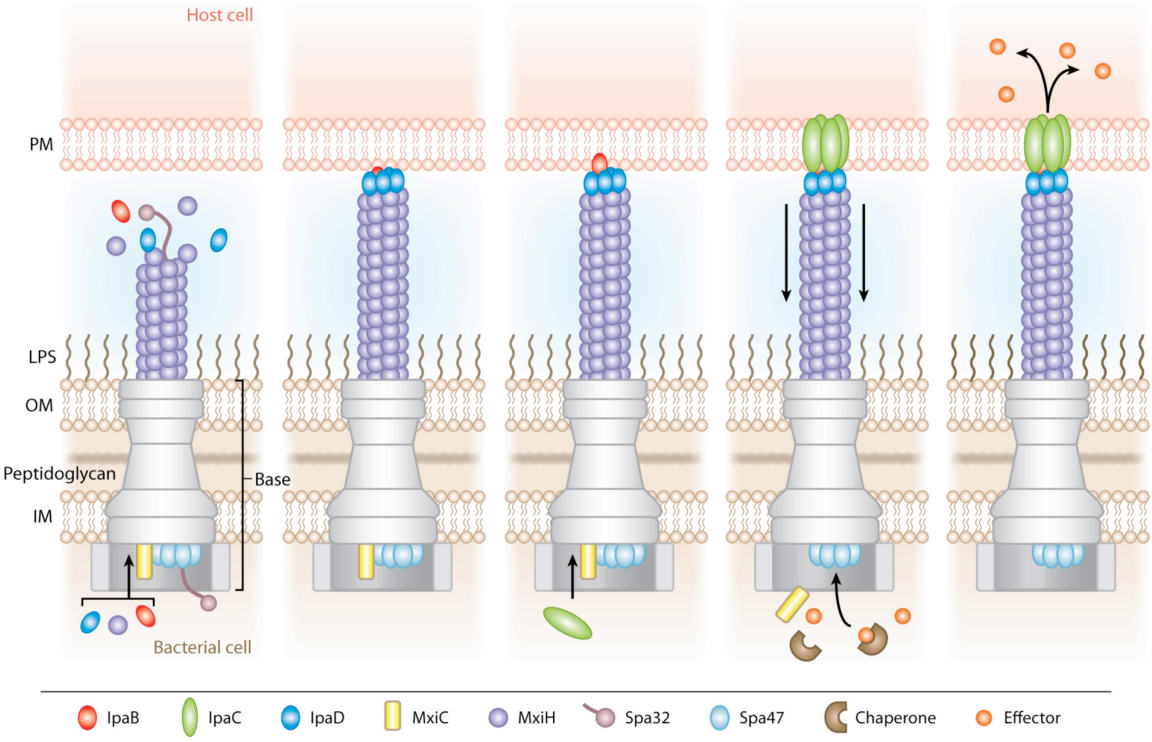
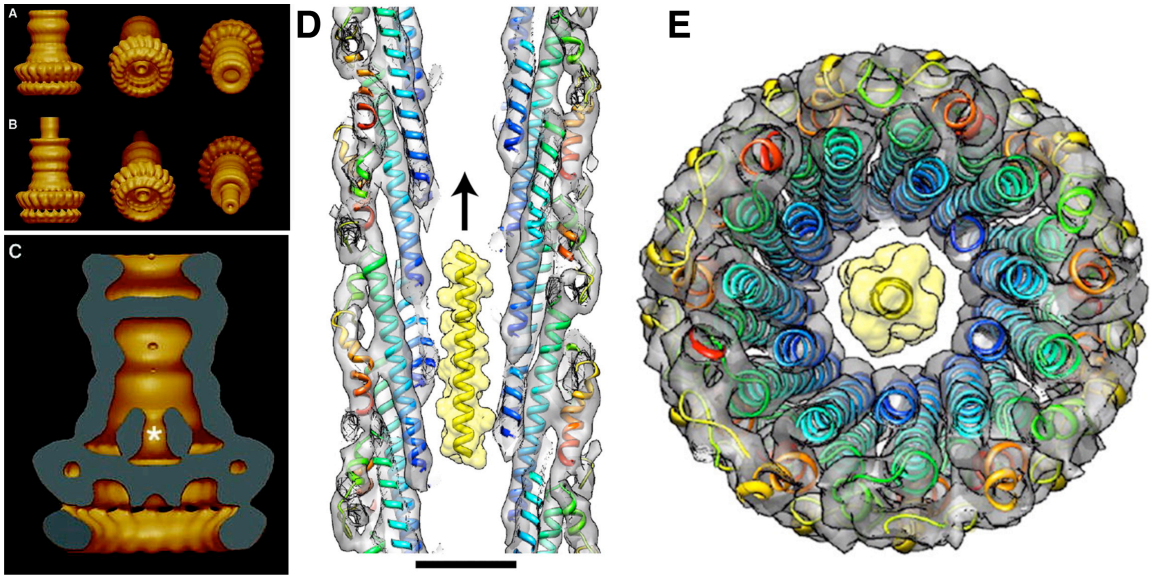


Figure Appendix II-1 Diagram of Type III secretion needle complex assembly and virulence factor transport. From Hayes et al. 2010.[118] Protein names are from *Shigella spp*. Schematic shows the assembly of of needle subunit MxiH in the presence of IpaB, C, and D which all are critical elements for docking with and pore formation in the host cell membrane. Spa32 has been shown to regulate needle length and substrate-specificity switching. MxiC has been assigned a gatekeeper role on the bacterial face of the basal body, and Spa47 is an essential ATPase that provides energy for transport.

Similar to the rotation of the flagellar apparatus, transport via TTS is an ATP-dependent process. Mutation in *Yersinia* of the TTS-associated ATPase YscN not only renders strains of *Y. pestis* avirulent but also allows the successful use of these strains as live-attenuated vaccines for bubonic plague in mice.[119] As mentioned above, it is unclear if there is a rotational aspect in the function of the ATPase in TTS, and if so how this would

be linked to protein transport. The size of the central pore of the needle complex is putatively the size of an α-helix (fig. Appendix-II D-F), so transit of TTS virulence factors is believed to involve either an unfolding of already-translated bacterial virulence factors or a process whereby translation and transit are coupled.[120] Possible evidence for the latter of these scenarios is found in reports that the mRNA of TTS virulence factors contains within it sequences that target it to the secretion apparatus.[121] How such a model is consistent with the role of the TTS virulence chaperones (see below) is not clear. Neither is it clear how the mechanism by which an ATPase-driven proton-motive force potentiates the possible unfolding and subsequent transit of virulence factors through the TTS apparatus. Several groups have attempted a structural approach to these questions through different combinations of electron microscopy, crystallography, and molecular modeling, and the "view" of these structures now extends to a resolution of sub-10 Å (Fig. Appendix II-2); however, many of the above questions await an atomic level description of the basal body and needle complex structures.[120, 122, 123]

Figure Appendix II-2 Structure of Type III Secretion system by 3D Cryo-electronmicroscopy. The *S. typhimurium* basal body at 17Å resolution is shown in **A**, in cross-section in **C**, and with the needle complex **B** (from Marlovits *et al*. 2004)[124] The structure of the needle from *Shigella flexneri* is shown at 7Å resolution in longitudinal (**D**), and horizontal (**E**) cross-section. Ribbon diagrams show a model of the needle complex subunit fit to the experimental cryo-electron density in gray. Density depicted in yellow is to scale for a linear α-helix, and the scale bar in D is 20Å. (from Fujii *et al*. 2012)[120]



As exemplified by the mutation to the *Yersinia* TTS-ATPase mentioned above, functional TTS is often essential for successful pathogenesis. Being unique to bacteria, many critical pieces of the TTS machinery have stimulated intense interest as non-traditional drug targets.[125, 126] Equally interesting have been the initial successes in manipulating TTS in order to utilize it as a drug/antigen delivery system within host organisms, particularly higher mammals. Exploiting both the ability of bacteria to target specific cell types and the ability of TTS to deliver large proteins directly into the host cell cytoplasm, these experiments could provide new and unique therapies for otherwise intractable diseases by delivering antibodies directly into host cells as a means of inhibiting abnormal cellular pathways or potentiating senescence/apoptosis[127], or by delivering disease specific antigens to the appropriate cellular components of the immune system.[128, 129]

It is important to note that there is not universal acceptance of the syringe-like mechanism of TTS translocation of effectors into host cells. Edgren *et al*. have recently

argued that direct evidence of effector translocation through the needle complex has yet to be shown. As an alternate model these authors offer a model that would involve the needle complex serving as a docking/sensing appendage that determines proximity to susceptible host cells. TTS effectors would then diffuse across the space between the two cells and be transported into host cells through a mechanism similar to AB toxin transport.[130] (Fig. Appendix II-3)

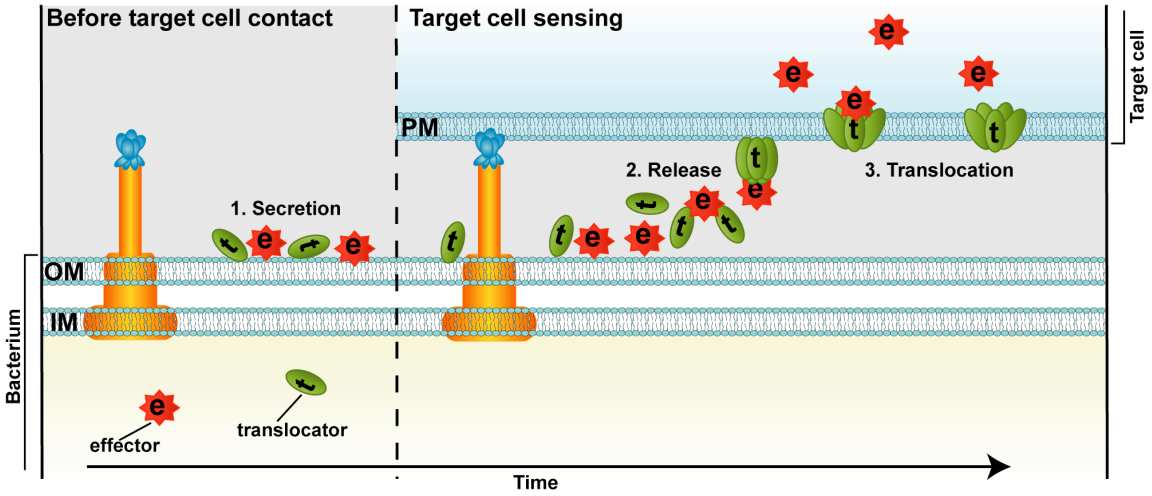


Figure Appendix II-3. Model for non needle-like injection mechanism for Type III secretion proposed by Edgren *et al*. in which the needle-complex is a sensing/docking organelle that then triggers demarginilization of effector/translocator complexes that diffuse onto the surface of the host cell and form pores that can then transport effectors into the host cell.[130]

Proteins secreted via TTS and the flagellar apparatus all contain an essential N-terminal secretion signal and putatively an additional secretion signal in their 5' region of their respective mRNA.[121, 131] Each secreted TTS protein – including subunits required for needle assembly – interacts specifically with one of a family of highly conserved chaperone molecules that target the effector to the secretion apparatus and stabilize the

effector and/or mask its activity prior to transport. Button *et al*. have shown how the translation of TTS virulence factors is tightly coupled to the co-translation of the appropriate TTS chaperone such that if chaperone translation is not present then translation of the paired virulence factor is inhibited by an RNA-mediated mechanism, suggesting that the genomic organization of TTS is an essential component to their regulation.[132] The nature of the interaction between TTS chaperones and their effectors is only generally understood. There are three groups of TTS secretion chaperones classified as such depending on whether the chaperone interacts with a single or with multiple effectors, or specifically with the two proteins involved in formed the docking pore at the tip of the needle complex.[133] (Fig. Appendix II-4) Interestingly many of these chaperones do not have directly homologous sequences across species but they still exhibit a highly conserved tertiary structure.[134] Structures of several TTS chaperones and effector-chaperone complexes have been determined using x-ray crystallography. The striking feature of these structural studies is the high degree of conservation of the overall shape of the fold as well as the identification of a generalized structural motif for effector-chaperone interaction that will help guide future research in this field.[135-138] Questions remain, however, regarding the interaction of full-length effector molecules with their respective chaperones, as well as the nature of the TTSS secretion signal in general. The results by Button *et al*. mentioned above as well as the general genetic organization of the TTS chaperones all highlight the importance they play in the overall success of this mode of virulence. In fact many of the questions that are actively being pursued in regard to TTS systems – the global regulation of the apparatus, how the order of effector transport is established and maintained, how is substrate specificity selected for – will almost

certainly involve the role of the chaperone proteins as part of the answer.

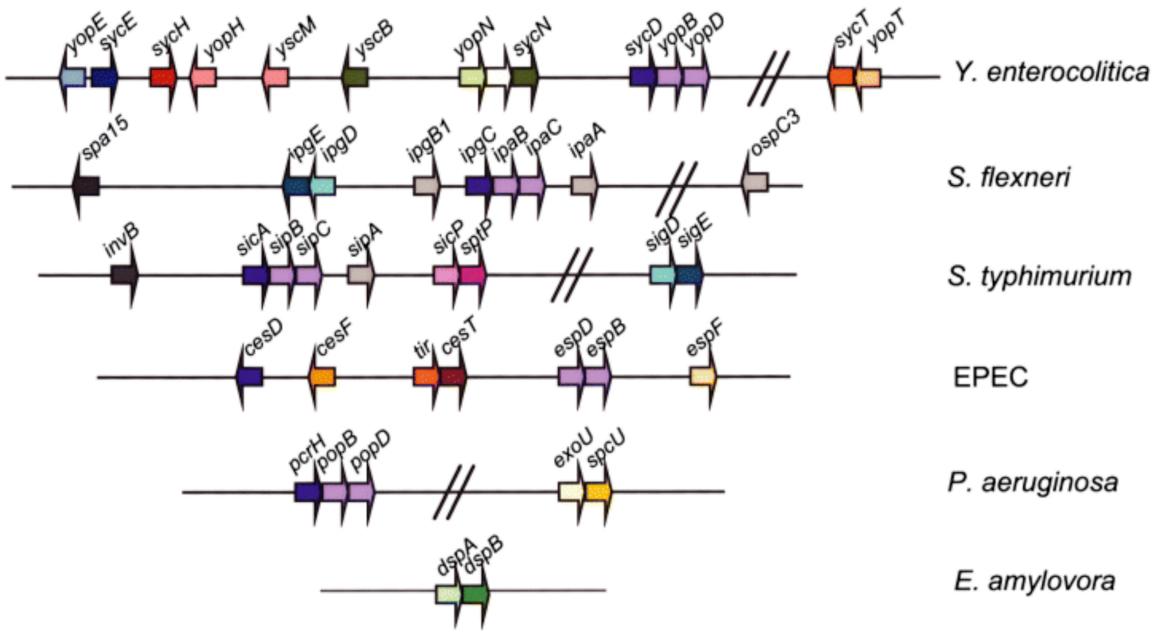


Figure Appendix II-4. Depiction as published by Page and Parsot of genes for chaperones (light colors) and the secreted virulence factors they bind (corresponding dark color) from several well-known TTS systems.[133]

What is known about TTS outlines a system that is amazingly complex and highly evolved. Attempts to understand such a complex system at a mechanistic level, including structural studies of the entire needle complex, highlight the degree of progress that has been made in the state of the art of structural biology. More importantly, however, is the degree to which TTS is essential to virulence and how an understanding of the effectors of a particular bacterial species can reveal a tremendous amount of essential insight into the pathogenesis of that organism. While the basal body, needle complex, and chaperone delivery system are all highly conserved across bacterial species, the gamut of TTS effectors and the effects they mediate as they are deployed in unique combinations by each genus and species of bacteria reveals a great deal about mechanisms of bacterial

virulence and adaptation. TTS allows bacteria to co-opt native host-cell pathways in order to facilitate bacterial survival to the greatest degree possible. Studying how this TTS-mediated reprogramming is accomplished not only elucidates mechanisms of bacterial virulence but also highlights host-cell biochemical pathways in exquisite detail. Quite amazingly many pathogenic bacteria express TTS effectors that are convergently evolved to positively or negatively regulate eukaryotic signaling pathways. In addition, the abundance of TTS in pathogens that cause significant human and agricultural disease along with the above discussion about the very real prospect of exploiting TTS as a drugable bacteriostatic target and, in a beautifully poetic manner, with the prospect of utilizing TTS as a drug delivery mechanism all argue for the importance of a greater understanding of this system. Below is a discussion of two projects aimed at understanding TTS effectors SopB and YpkA from *Salmonella spp*. and *Yersinia spp*., resceptively.

Salmonella spp.

Salmonellae comprise a genus of gram-negative facultative anaerobic bacteria that are intracellular pathogens in vertebrates. As of 2004, there are 2541 characterized serovars distributed between two species, *S. enterica* (2419) and *S. bongori* (22).[139] The host specificity of salmonellae is highly variable with some serovars infecting various hosts, while others are highly host specific. Furthermore, a given serovar highly pathogenic in one host may be mild to asymptomatic in another. *S. enterica* serovar Typhimurium (*S. typhimurium*), for example, is a common cause of enteric disease in humans and agricultural animals, and is lethal in mice, while *S. enterica* serovar Typhi (*S. typhi*) infects humans exclusively.

Salmonellae spread between hosts via fecal-oral contamination, cause four types of diseases – enteric fever (typhoid), enterocolitis/diarrhea, disseminated bacteremia, and chronic asymptomatic carriage – and are a significant health concern for humans and livestock.[140] Globally an estimated 200 million to 1.3 billion cases of non-typhoidal *Salmonella* infections occur annually resulting in 3 million deaths. Infections with *S. enterica* serovar Typhi contribute an additional 17 million cases with 600,000 fatalities annually, in spite of the presence of an effective vaccine. In the United States, the CDC estimates that 1.6 million non-Typhoidal *Salmonella* infections occur annually resulting in nearly 16,000 hospitalizations and 600 fatalities. *Salmonella* is the leading cause of morbidity and mortality among bacterial food-born diseases in the U.S., with a combined health and agricultural economic impact in 2007 of $2.5 billion.[141] While enteric cases of *Salmonella* are self-limiting, and bacteremia and typhoid are typically treatable with antibiotics, the emergence of antibiotic resistant salmonellae continues to be an increasing worldwide health risk. Cases of multi-drug resistance (MDR) have risen significantly in the last 10 years due in part to the broad-scale prophylactic use of antibiotics in animal husbandry and to the presence of a transferable genetic element in salmonellae containing several drug resistance genes.[142, 143] MDR salmonellae tolerate ampicillin, chloramphenicol/florfenicol, streptomycin, sulfonamides, tetracycline, and 3rd generation cephalosporins. Such isolates exist in at least 43 countries, are increasingly associated with nosocomial outbreaks, and have been implicated in increased morbidity and mortality.[142-146]

Pathogenesis

Salmonellae exhibit two essential virulence phenotypes necessary to establish local

infection and promote disseminated disease: host-cell invasion and intracellular pathogenesis. Salmonellae cause disease by invading otherwise non-phagocytic cells in the host intestine, inducing inflammation, subsequently infecting attracted phagocytic cells, and spreading systemically via lymphatic trafficking in phagocytic/antigen-presenting cells. Once introduced into the gastrointestinal tract, salmonellae adhere to and induce phagocytosis in cells of the small intestine, specifically enterocytes, Peyer's patches, mesenteric lymph nodes, and newly identified cryptopatches or solitary intestinal lymphoid tissue.[147] Intracellular bacteria persist in a spacious phagosome that matures into a structure known as the Salmonella containing vacuole (SCV). This structure exhibits membrane markers of an early endosome and is able to evade trafficking to lysosomal compartments. From within the SVC, salmonellae induce host-cell secretion of IL-8 and pathogen-elicited epithelial chemo-attractant (PEEC), eliciting local inflammation and attracting phagocytes. By an unknown mechanism, the SVC allows salmonellae to traffic to and across the basolateral surface of infected cells where the bacteria repeat the process of host-cell invasion in the newly-attracted phagocytic cells, a stage that is essential for virulence.[148] Once within phagocytic cells, salmonellae induce secretion of mature IL1-β and IL-18, amplifying inflammation and inducing pyroptosis of infected phagocytes by an intracellular caspase-1 mediated mechanism. While seemingly paradoxical, the active proliferation of a proinflammatory environment appears to benefit salmonellae by attracting migratory phagocytic cells, which salmonellae readily infect and utilize for systemic spread, and by inhibiting competing and/or protective intestinal fauna.[149]

Salmonella Pathogenicity Islands

The majority of the genes necessary for the pathogenesis of salmonellae are located on pathogenicity islands (SPI) distributed throughout the *Salmonella* genome. Twelve SPIs are distributed differentially across *Salmonella spp*. and are perhaps most notable for the presence of two type three secretion systems (TTSS) in SPI-1 and SPI-2, which are critical for host-cell invasion and intracellular pathogenesis, respectively.[150] Each TTSS translocates a unique set of SPI-encoded effectors that are spread across the highly conserved SPI-1, -2, and -5, which are found in all but one *Salmonella spp*.[151, 152] The temporal regulation of the SPI-1 and SPI-2 TTSS and their respective effectors allows for a complex and highly orchestrated manipulation of host cell biology. These effectors possess a variety of different functions that mimic or modify the host cell molecular machinery such that a large number of cellular pathways are coordinately affected.[153]

Salmonella outer protein B (SopB, previously SigD)

SopB is one of five SPI-1 TTSS effectors necessary for the invasion phenotype in Salmonellae and is found in all but one known subspecies, which lacks the SPI-1 TTSS altogether.[151-153] SopB is a phosphoinositol (PtdIns) phosphatase active throughout invasion and intracellular phases of salmonellae infections.[154, 155] SopB comprises 560 residues, with N-terminal chaperone binding and secretion signal domains, a ubiquitinated domain (amino acids 29-116), a membrane association domain (a.a.117-167), and three C-terminal PtdIns phosphatase homology domains.[156, 157] (Fig. Appendix II-5A) PtdIns are a major class of signaling molecules within the host cell and are involved in the regulation of many cellular processes.[158] The enzymatic activity of SopB generates fluxes in the concentration at least four signaling PtdIns and has been shown in

a phosphoproteomic analysis to lead to be essential for nearly 50% of all cellular phosphorylation events that occur within host cells follwing Salmonella invasion.[159] These fluxes are the basis for at least five characterized effects within the host cell:[160-162]

i. *AKT activation and anti-apoptosis*: the synaptojanin (5-phosphatase) homology domain at the C-terminus of SopB is required to activate AKT for 6-8 hours post infection by generating PtdIns(5)P from PtdIns(4,5)P_2.[157, 163, 164]

ii. *Chloride secretion*: the hydrolysis of pentakiphosphate to yield PI(1,4,5,6)P_4 antagonizes the binding of Cl^- channel inhibitor PI(3,4,5)P_3, the effects of which causes 50% of the fluid secretion in bovine ligated ilial loop models of salmonellae infection.[155, 165, 166]

iii. *Creation and maintenance of Salmonella Containing Vesicle (SCV)*: elimination of PtdIns(4,5)P_2 by SopB promotes membrane fission to create the spacious phagosome and subsequent SCV, in which the high PtdIns(3)P content is maintained by SopB thus preventing degradative lysosomal trafficking.[167, 168]

iv. *Actin remodeling*: SopB alters the actin cytoskeleton by activating RhoG via SH3-containing guanine nucleotide exchange factor (SGEF), which is activated by a flux in PtdIns of unknown identity. Additionally, SopB, SopE, SopE2, and SipA disrupt tight junctions via Rho-family GTPase activation.[169-171] Interestingly, SopB lacking C-terminal phosphatase domains has been reported to disrupt actin polarization.[172]

v. *CDC42 activation*: SopB activates CDC42 in a phosphatase-dependent manner, leading to actin-mediated cytoskeletal rearrangements.[173] The interaction of

SopB and CDC42 has been localized to SopB residues 1-142. SopB constructs mutated to lack ability to interact with CDC42 do not traffic to the SVC, which correlates with decreased rates of intracellular *Salmonella* replication within the host cell.[174] Curiously, phosphate-dead mutant (R468A) *inactivated* CDC42 when heterologously expressed in yeast, giving further potential evidence for a phosphatase independent activity as mentioned above.[175] The recent structure of SopB 29-181 in complex with CDC42 gives further evidence for phosphatase independent activity as it shows that SopB contains a guanine nucleotide dissociation inhibitor (GDI) that recognizes the CDC42 GTPase binding domain, thereby inhibiting CDC42 and preventing downstream signaling events mediated though this pathway.[176]

Salmonella invasion gene E (SigE): SopB type III secretion chaperone

The structure of SigE has been solved crystallographically to 1.9Å resolution, and site-directed mutagenesis of essential hydrophobic and electrostatic interactions has demonstrated the diversity of interactions present in this particular chaperone-effector pair.[135, 137]

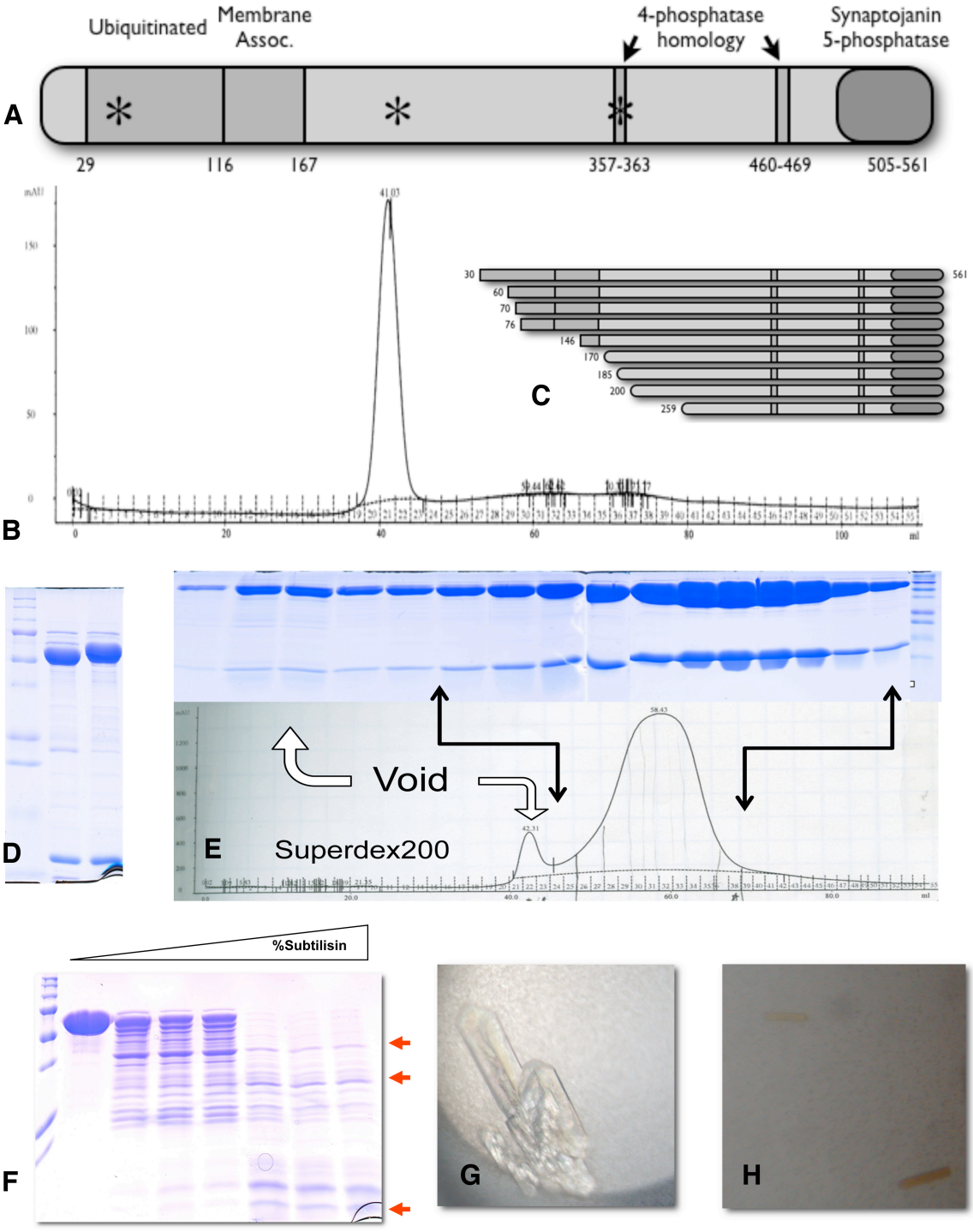
Results

SopB (30-561) was successfully cloned from genomic *Salmonella enterica* serovar LT2 and expressed recombinantly in *E. coli* as a construct containing an N-terminal 6xHis-tag cleavable with 3C protease. Various strategies were used to attempt to clone the full length sequence including optimization of PCR cycles with gradient annealing, melting, and extension protocols as well as with different polymerases; however none of the

techniques were successful. Levels of SopB30-561 expression were unsatisfactory (<1 mg/ml) and yielded aggregated samples upon purification. A series of N-Terminal truncations were created in an attempt to identify a more-soluble construct: 60-, 70-, 146-, 170-, 185-, 200, and 259-561. (Fig. Appendix II-5 B) All constructs were created successfully, but none exhibited increased solubility. Secondary and tertiary structure prediction using PHYRE identified neither an obvious domain structure nor a structural homologue that could assist the rational design of a more soluble construct.[177]

Co-expression of SopB 30-561 with SigE yielded a dramatic increase in SopB expression and solubility (100-120 mg/L of cultured cells and >80 mg/ml for purified protein, respectively), which allowed for purification of SopB in complex with SigE to high purity for crystallization screening. The SopB:SigE complex was very stable such that attempts to separate the two proteins were not successful using native buffer conditions, including both anion and cation exchange chromatography. Sparse matrix crystallization screening of this SopB30-561:SigE complex at both 15 and 31 mg/mL at both room temperature and at 4°C yielded two hits in 96-well format at 4°C (50% MPD, 200 mM Am. Phosphate, and 100 mM Tris pH 8.5; and 25% Tert-butanol, 100 mM Tris pH 8.5). (Figure SoPB 1 G-H) Multiple attempts to reproduce these crystals in 96- and 24-well format in both sitting and hanging drop failed to reproduce the initial crystal hits.

Figure Appendix II-5. Initial purification and crystallization of SopB. **A**. Schematic of SopB showing the key features of the protein including a region of shown to be ubiquitinated, a membrane association region (also CDC42-interaction domain), and phosphatase domains. * denote sites of proteolysis (see panel F) B. Representative Superdex 200 gel-filtration elution profile for SopB 30-561 and all other constructs shown in panel **C**. Note that nearly all protein elutes at 40-42 mL, which is the void volume for the column. **C** shows schematic of all constructs tested for SopB overexpression. None improved stability. **D**. 6xHis(3C)SopB 30-561 before (right) and after (left) 3C proteolysis to remove 6x-His purification tag. **E**. Final gel-filtration purification step of SopB30-561/SigE-fl showing that a large majority of the material remains in complex with compliment and that few contimanants were seen. **F**. Limited proteolysis of SopB30-561 w. SigE. Triangular shape shows the concentration of Subtilisin. Red arrows indicate bands sent for N-terminal sequencing, and the confirmed sites of proteolytic lesions are indicated by (*) in panel **A**. G and H show initial crystal hits from primary round of screening. **G**. Initial crystal hit condition 50% MPD, 200mM Am. Phosphate, 100 mM Tris pH 8.5 **H** Initial crystal hit condition. 25% Tert-butanol, 100mM Tris pH 8.5.


A
Ubiquitinated
Membrane Assoc.
4-phosphatase homology
Synaptojanin 5-phosphatase
29
116
167
357-363
460-469
505-561
B
C
30
60
70
76
146
170
185
200
259
561
D
E
Void
Superdex200
%Subtilisin
F
G
H

Other purification options were attempted for SopB in complex with SigE to yield the best possible crystallization material, but nothing yielded results that were objectively better than the results described above. The one exception to this statement was the use of ultracentrifucation prior to final gel-filtration purifiction. (figure UUIOP). While gel filtration offered a clean means by which to separate soluble aggregates from mono-disperse soluble material, ultra-centrifugation (80,000g for 60 minutes at 4°C) accomplished the same level of purification without gel filtration, and could therefore also allow a cleaner starting material before the final gel-filtration polishing steps. It seems that this step could easily apply as a general purification step for crystallization samples and should be considered for future projects. In this project it did not provide any different results and crystallization screens of ultra-centrifuged material yielded similar results to the previous trials.

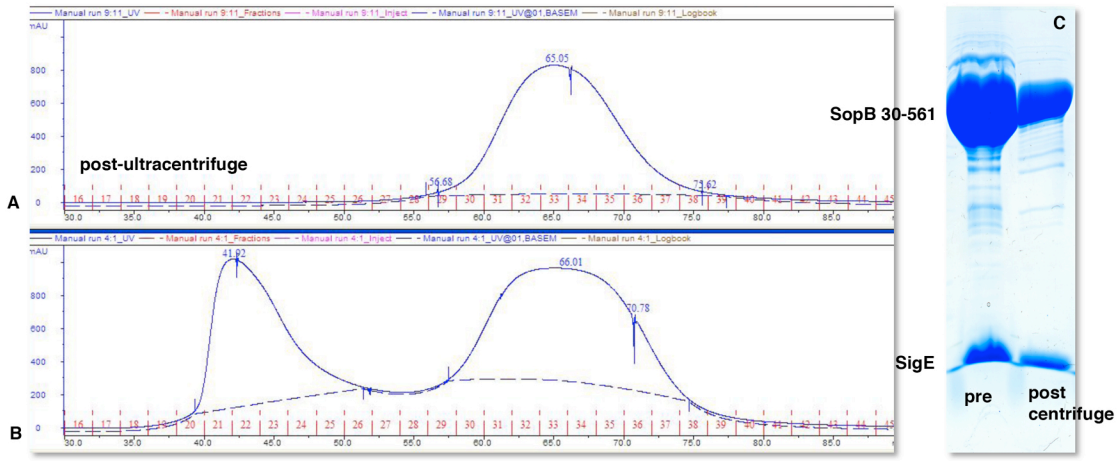


Figure Appendix II-6 Purification of SopB30-561/SigE complex via ultra-centrifucation (80,000 xg). Gel filtration of the pre-(**B**) and post-centrifucation (**A**) samples. Note the large void peak in **B** that is completely removed by centrifugation, as shown in **A**. **C** SDS-PAGE of samples from the middle of the elution curve from the pre- and post-ultra-centrifuge.

Having failed to obtain reproducible crystals with SopB30-561 in complex with SigE, several residues of SopB that were predicted to be surface exposed and to contribute significantly to surface entropy have been mutated to alanine to determine if these regions were inhibiting crystal formation.[178] The resulting constructs SopB118-121AAAA, 256-257AA, and 312-313AA were characterized, but also failed to yield positive crystal hits.

Further attempts were made to identify a stable SopB construct that could be expressed in the absence of SigE. Limited proteolysis of the SopB30-561:SigE complex with increasing concentrations of subtilisin did not reveal definitive results as would be characteristic from a multi-domain protein, i.e. the banding pattern of the degradation products did not reveal fragments that were clearly stable over a range of subtilisin concentrations. Edman sequencing of the degradative products identified cleavage at residues 53, 229, and 360. (Fgure sopB-1 A and F) As several constructs from the initial round of N-terminal truncations had already characterized constructs that were similar to these results, the proteolysis results did not offer significant progress. Cleavage at residue 360 was in the middle of the putative PtdIns(4)P phosphatase domain, so this residue was not used for construct design. Instead, residue 229 was used as a basis for two additional constructs, 229-561, and 229-505, both of which yielded negative results.

Reductive alkylation of the SopB30-561:SigE complex was used to further expand possible crystallization conditions. Interestingly, gel-filtration and anion-exchange purification steps following reductive alkylation of the complex resulted in a clean separation of the two proteins. (Figure 2 SopB A-D) Purified thusly, reductively alkylated SopB 30-561 (RASopB 30-561) was screened by spare-matrix crystallization

screening at 10-12 mg/mL at both room temp and 4°C with no identified hits. A sample of reductively alkylated RASopB was also sent to the Hauptman-Woodward Medical Research Institute for high-throughput crystallization screening using a 1536-well crystallization screen designed for membrane proteins.[179] These screens yielded several crystallization hits that were optimized in 96-well format in-house around the following conditions: 1.1 and 2.2 M CaCl2 in 100 mM Na-Acetate (pH 5.9), HEPES (pH 6.8), and Tris (pH 7.4). (figure sopB 2 F) UV microscopy of these hits suggested they were protein containing crystals and optimization was extensively pursued including multiple formats, micro and macro seeding, and additive screening, which identified T-butanol as an additive that promoted cleaner crystal edges, decreased nucleation, and increased crystal growth.[180] After multiple failed attempts to optimize these crystals to diffraction-sized samples (the largest crystals remained well under 25 μm), it was noted that older trays had begun to grow crystals in the well solution, i.e. not in the experimental drop, and thus could only be salt crystals. The crystal morphology of the well crystals matched the morphology in the drop with the almost certain conclusion that the crystals being optimized were salt. This conclusion was suspected originally due to the presence of Ca^{2+} ions as the sole precipitant at two very different concentrations with almost identical crystallization results. Calcium, being a divalent cation, has been found empirically to readily form salt crystals in a variety of crystallization conditions. The chemical basis for this behavior likely stems from the low solubility of many salts formed by divalent cations, which, interestingly is also the basis for the presence of divalent cations in the most common forms of kidney stones. This limited solubility has prompted many commercial crystallization screens to limit the inclusion of divalent cation reagents due

the high false-positive rate that such conditions give in crystallization screens.[181]

Given that the RA-SopB30-561 was able to be purified away from SigE, another round of limited proteolysis was attempted to determine if the lack of SigE would present a different digestion pattern and possibly suggest a more consildated domain architecture or at least additional proteolysis sites that could be useful for future expression trials. Unfortunately, the digestion patters were similarly problematic to those seen for SopB30-561/SigE. The banding pattern upon degradation did not yield a small number of stable domains and instead yielded many bands that seemed fairly unstable even over a narrow range of subtilisin concentration. (Fig. Appendix II-5F)

Due to the large number of lysine residues in SopB (50), it was possible the results of reductive alkylation yielded a highly heterogenous population of SopB molecules that would not be favorable for crystallization. Consequently separation of the SopB30-561/SigE complex was attempted using affinity chromatography under denaturing conditions followed by on-column refolding. This protocol yielded purified SopB30-561(Figure UIOUPO); however, the refolded SopB was not stable and quickly aggregated, suggesting that SigE does stabilize a domain that is inherently unstable in the absence of a chaperone or modification by reductive alkylation.

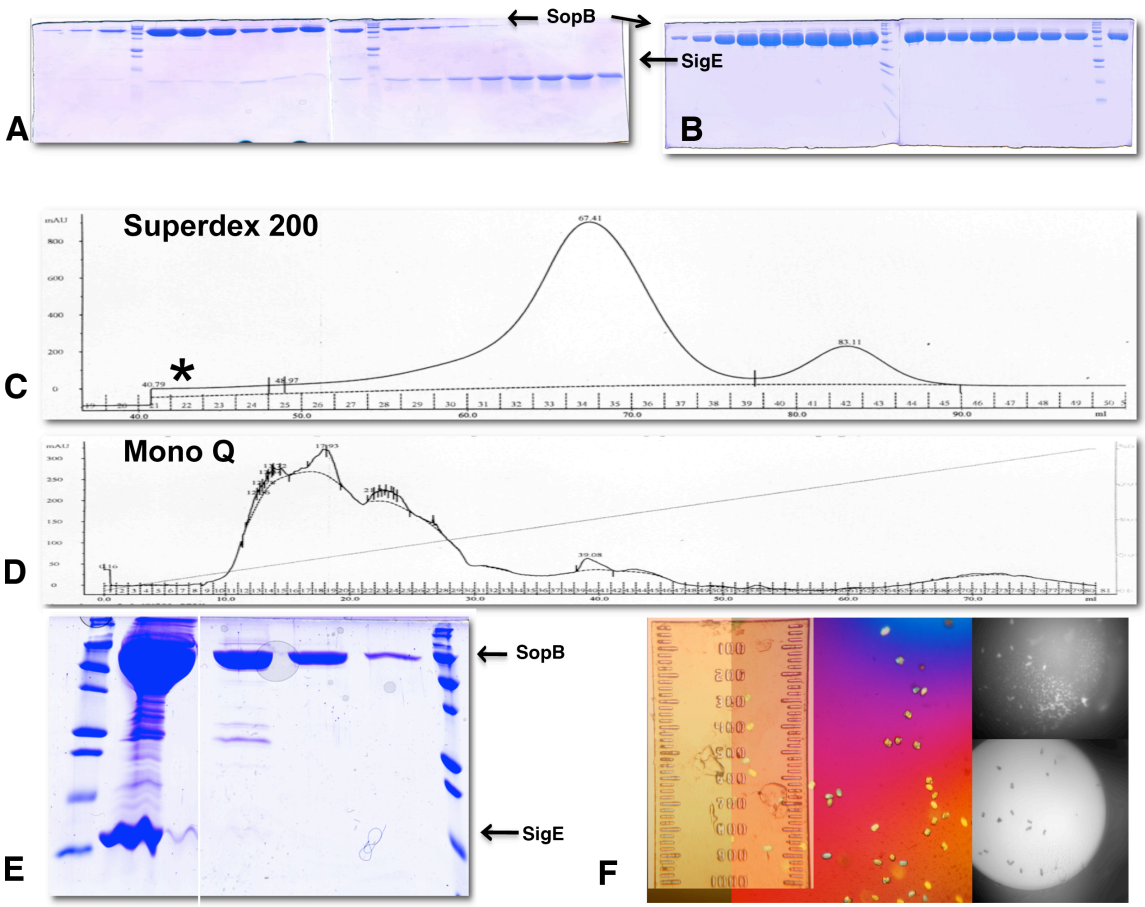


Figure Appendix II-7 Purification and crystallization of reductively alkylated SopB 30-561. **A**. SDS-PAGe analysis of Superdex 200 gel-filtration purification (**C**) of reductively alkylated SopB 30-561 from SigE. **B**. SDS-PAGE analysis of SourceQ anion exchange purification (**D**) of reductively alkylated SopB30-561. **E**. SDS-PAGE of eluted fractions from Ni-NTA of on-column refolded SopB30-561 showing separation from SigE during refolding. **F**. Initial crystal hits (far right, bright field on bottom, UV in top) of high throughput crystallization screen for reductively alkylated SopB30-561. Optimized crystals are shown in polarized brighfield with ruler marked in micro meters.

The crude extent of the domain in SopB that binds to the SigE chaperone was investigated by co-expressing SigE with the series of N-terminal truncations mentioned above. Favorable interaction of SigE with SopB was observed indirectly through increases in SopB expression and directly through co-purification of SigE with SopB:

constructs 70-561, 146-561, and 170-561 co-purified with SigE but constructs 185-561, 200-561 or 229-561 did not. This experiment confirmed that SigE interacts with the N-terminus of SopB and that the C-terminal limit of the SigE-association domain is somewhere between residues 170-185. (Fig. Appendix II-8) The following C-terminal truncations of this domain were co-epressed with SigE in an attempt to identify the N-terminal limit of the region that was critical for successful co-expression of SigE and SopB: SopB 30-229, 30-185, 30-170, 30-145, 30-110, 53-229, 53-185, 53-170, and 53-145. All of these constructs co-expressed well with SigE (data not shown). Together these data suggested that a SigE-association domain resides in the N-terminus of SopB and can is completely contained in the SopB30-185 construct.

The assertion that the above experiments identify the minimal SigE-association domain in SopB assumes that the condition of successful co-expression offers a read-out forassociation. Given that SigE is known to function as a SopB chaperone, this does not seem like an unreasonable assumption, but it is certainly possible that additional interaction between SigE and SopB could exist outside of this domain and that such an interaction could have significant functional significance. Following the belief that the most "crystallizeable" protein constructs are often those that contain the minimal extent of the domain of interest, structural studies were pursued for SopB 30-185 in complex with SigE.

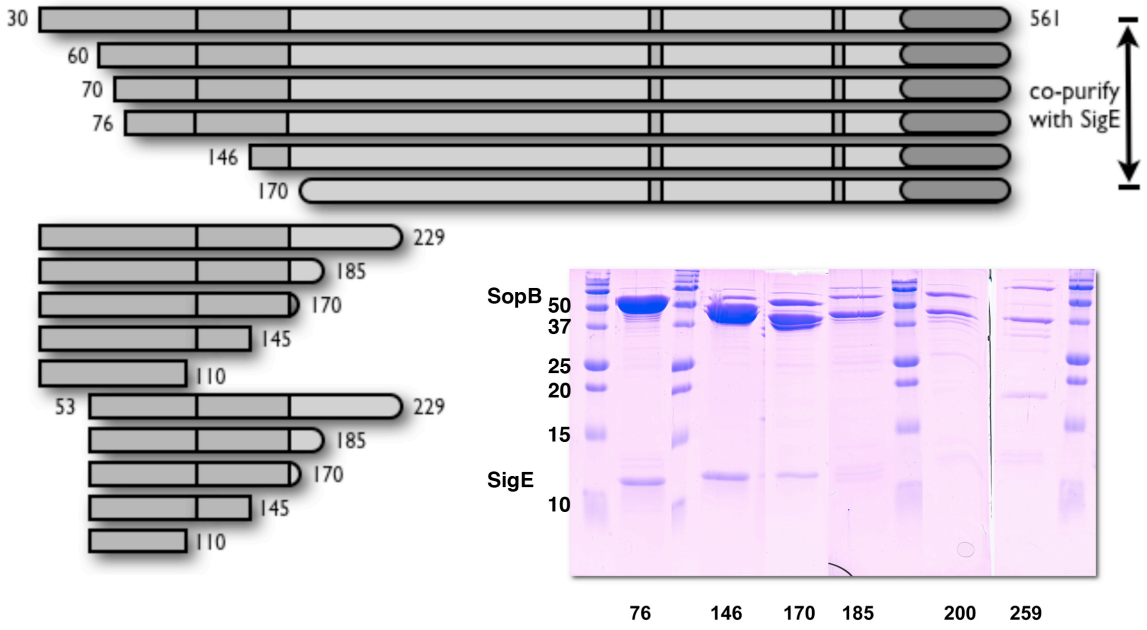


Figure Appendix II-8. Identification of SigE binding domain in SopB. Schematics represent constructs that were characterized for coppurification with SopB. The top set of schematics show N-terminal truncations that identify the C-terminal limit of the SigE binding domain, for which the SDS-PAGE analysis Ni-NTA purifications of His-tagged SopB constructs co-expressed with untagged full-length SigE. The numbers along the bottom of the gel correspond to the N-terminal residue of the construct. The bottom two groups of schematics show C-terminal truncations of SopB, which were designed to identify the N-terminal boundary of the SigE binding comain. All of these constructs purified with SigE (data not shown), so the N-terminal border of the SigE binding domain is not absolutely clear, but is assumed to be somewhere between 30 and 110.

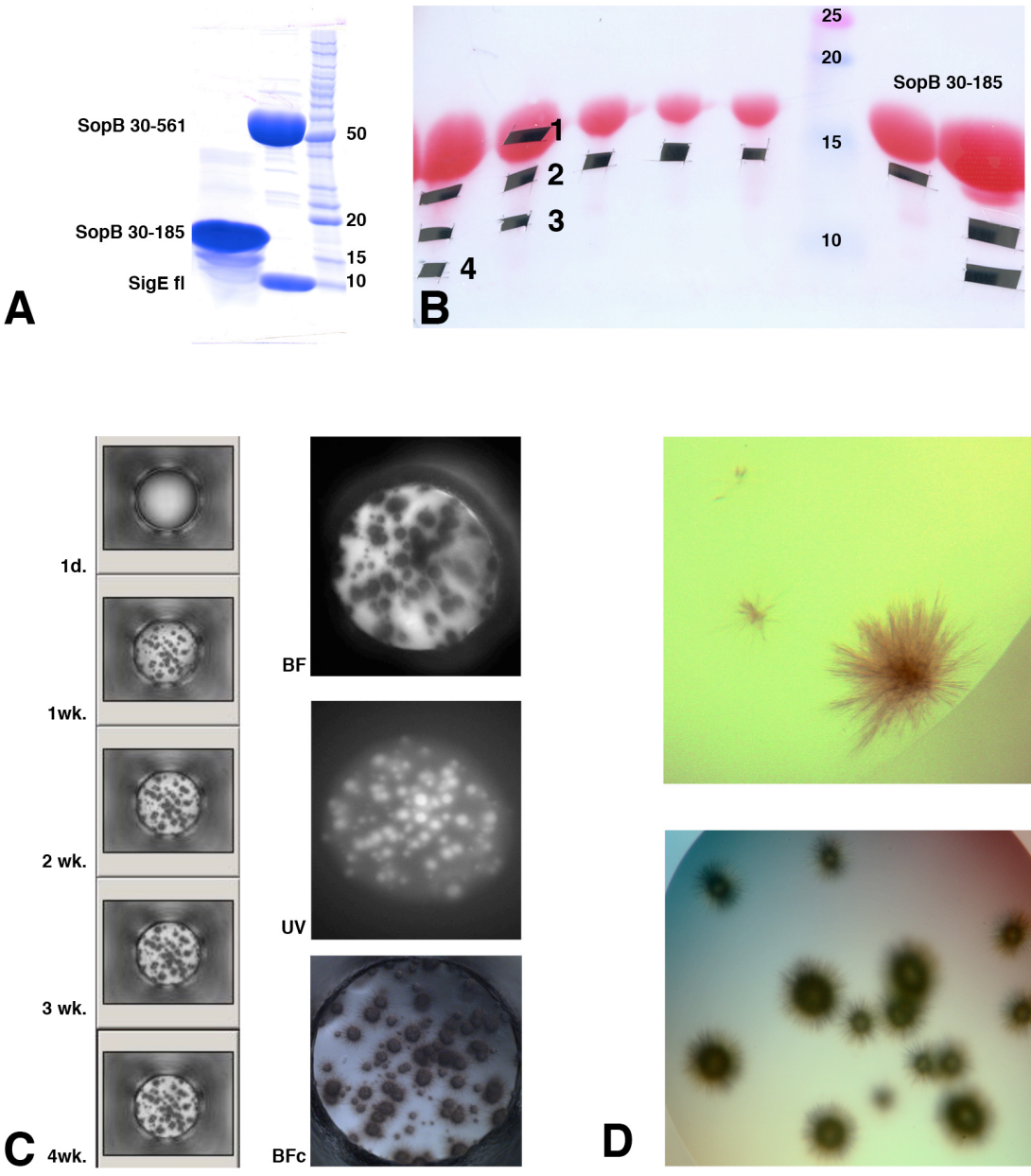
SopB 30-185 was co-expressed and purified successfully in a manner similar to SopB30-561 utilizing a 3C-protease cleavable 6xHis tag on the SopB construct with rounds of Ni-NTA purification pre- and then post 3C digest, with subsequent gel-filtration purification. (figure panel A) The yield of 12L of culture was concentrated to 13.6 mg/mL for the purposes of crystallization screening. SDS-PAGE analysis of the concentrated sample indicated that SigE had progressively been lost during the purification leaving behind

what appeard to be SopB30-185 alone. Trace bands of lower molecular weight than SopB30-185 and of appropriate molecular weight for SigE were noted and Edmun sequencing was performed to determine whether SigE was still present in a decreased ratio or whether the smaller molecular weight bands represented breakdown/proteolysis of SopB30-185. (figure UIOPSUD panel B). A total of four separte sequencing samples were sent and all were found to have the N-terminal sequence of SopB, indicating that SigE had been lost during purification and the banding on the SDS-PAGE of the concentrated sample represented C-terminal truncations of the SopB30-185 construc. This result was interesting since it suggested that SopB30-185 was only soluble when co-expressed with SigE but was stable in the absence of SigE thereafter. Moreover, unlike SopB30-561, the affinity between SopB and SigE seems to be diminished when only the minimal SigE-association domain is expressed since the full-length constructs could never be separated by gel-filtration or anion exchange wherease the SopB30-185 construct was separated with gel-filtration alone. This last observation somewhat calls into question the deliniation of SopB residues 30-185 as the only SigE interaction domain described above.

The material obtained from the steps above was sent for high-throughput crystallization screening at the Hauptman-Woodward Medical Research Institute mentioned above for RA-SopB30-561. Several conditions yielded positive hits, with all showing a similar sea-urchin morphology of clumps of fine needles with a dense core. Bright-field and UV images gave strong evidence that the crystallized material was proteinaceous. (figure SHJDF panel C) The conditions were reproducible in-house in 96-well format and several methods of optimization were employed, including micro-seeding and additive

screening, which identified several conditions that maximized the fine needle growth and limited the central clumps. The best of these efforts is shown in figure OSDFSDFIOP panel D; however, these images also show the limit of optimization that was achieved with these conditions. Ultimately the optimized crystals were still only 25-50 μm in length and immeasurably thin, which precluded any assessment for diffraction on the in-house x-ray source.

Figure Appendix II-9. Crystallization trials of SigE binding domain of SopB in complex with SigE. **A**. SDS-PAGE of final purified sample of SopB30-185 used for crystal screening. A sample of SopB30-561 with SigE is run alongside for comparision and to demonstrate a notebable lack of SigE band in the SopB30-185 sample. **B**. PVDF membrane showing several different concentration so of SopB30-185 used for crystal screening analyzed for Edmen sequencing. Bands labeled 1-4 were excised and sequenced and all four had the N-terminal sequence of SopB30-185, confirming the absence of SigE from these samples. **C**. Crystal hits from high throughput screening of SopB30-185 sample shown in A. Left strip shows progression of crystal growth and three imaged on right show, from top, higher quality bright field (BF), uv (UV), and color bright field (BFc) images of this condition after 4 weeks: 400 mM Magnesium Chloride Hexahydrate, 100 mM MES pH 6.0, 30% PEG 3350 (w/v). **D**. Two images of optimized crystallization conditions form high-throughput screen in C that contain the additive benzamidine HCl. The longest needles in D are approximately 25 micrometers long.


A
SopB 30-561
SopB 30-185
SigE fl
50
20
15
10
B
25
20
SopB 30-185
15
10
1
2
3
4
C
1d.
1wk.
2 wk.
3 wk.
4wk.
BF
UV
BFc
D

Discussion

The ability to explore the structural biology of SopB was consistently hampered by the inability to obtain high-quality, reproducible crystals of SopB either alone or in comples with its TTS chaperone SigE. Initial difficulties of low yield and aggregative behavior were solved successfully through co-expression with SigE and through the use of reductive alkylation. Both approaches yielded high-quality, non-aggregating material that seemed optimal for crystallization screening. The behavior of SopB when co-expressed with SigE showed an extremely stable complex that remained in all but denaturing conditions. Empiric construct design in the context of SigE co-expression allowed the identification of a minimal region comprising residues 30-185 of SopB that was necessary and sufficient for soluble expression with SigE, but further rounds of purification resulted in a pure sample of SopB30-185. The depature of this result from the behavior of the SopB30-561/SigE complex was interesting as it suggested that SigE plays an important role in stabilizing and solubilizing the growing translated SopB, and it also suggested that additional contacts outside of the 30-185 could contribute to the interaction between SopB and SigE thus reinforcing this interaction in the full-length co-expression. These results are further interested as recent articles have been published identifying a minimal CDC42-interacting domain. The interaction of SopB with CDC42 has been shown to contribute to the overall successful invasion of Salmonella into the host cell but the SopB CDC42 interaction seems independent of the phosphatase activity of SopB. The identified CDC-42 interaction domain of SopB maps very closely to the minimal SigE-interaction domain described above. The interaction of the N-terminus of SopB and CDC42 has also recently been shown crystallographically.[176] Again, these

results strengthen the conclusion from the above experiments that SopB30-185 represents a sub-domain of the protein that exhibits an innate level of stability and also functions via protein-protein interaction to help affect bacterial invasion. Unfortunately none of the crystallization screening hits yielded material that was productive in generating diffraction data. There are still many questions regarding the structural biology of SopB that remain to be answered and further research into this TTS virulence factor will definitely be an interesting to topic to follow for many upcoming years.

Yersinia spp. and YpkA

Yersinia are genus of Gram negative rod-shaped bacteria that contain three important human pathogens: *Y. enterocolitica*, *Y. pseudotuberculosis*, and *Y. pestis*. The first two of these species cause acquired food-borne illness that results in a spectrum of disease ranging from acute to chronic gastro-enteric infection with septicemia in advanced cases. These infections are typically self-limiting in healthy individuals but can result in hospitilzation and even death in the young, elderly, and immune compromised.[182] *Y. enterocolitica* is by far the most common of the two food-borne species and is typically associated with undercooked pork and unpasteurized milk products.

The far greater interest in this genus stems from the significant influence on human history exerted by *Y. pestis*. Transferred between mammalian hosts via a flea vector and spread from host to host via flea bite, *Y. pestis* is the causative agent of pneumonic, septicemic, and most famously bubonic plague. The significant prevalence of this bacterium throughout the world was first appreciated in the late 1890s by Alexandre Yersin, a disciple of Koch and Roux, who isolated the bacterium during an outbreak of

the plague in Hong Kong. Yersin went on to study the bacterium extensively throughout Asia and developed an antisera to *Y. pestis* that dramatically increased survival, which prompted in the mid 20th century the renaming of the genus in his name.[183] While some controversy has surrounded the attribution of the great plagues of human history to *Y. pestis* alone, genetic testing has confirmed evidence of the organism in isolates from two significant instances of pandemic plague that took the lives of millions of people throughout Europe and Central Asia in 1347 and 1353.[184] *Y. pestis* has without a doubt had one of the most profound influences on human history of any infectious organism; moreover, the specter of *Y. pestis* infection has gone from bad to worse during many historical periods as this pathogen has time and again been utilized as a biological weapon in almost every corner of the globe.[185] This latter threat combined with the persistent endemic nature of the disease in all continents except Australia and Antartica continues to prompt extensive research into this organism as new means of rapidly detecting and effectively treating infections caused by this organism are in constant need of attention.

Unlike *Salmonella spp*. the TTS virulence system of *Yersinia spp*. is located on a virulence plasmid that, although different between the three *Yersinia* species, is highly conserved with regard to the virulence factors encoded therein. TTS is not the only mechanism of virulence in *Yersinia*, but the combined effect of the six main TTS virulence factors in *Yersinia*, named *Yersinia* outer proteins (Yop), has a significant influence on the pathogenicity of disease.[186] The concerted effect of the various Yops is to suppress and reorient the host immune response, effecting critical immune surveillance mechanisms, particularly the phagocytic behavior of macrophages. The suppression of

monocytic phagocytosis of *Pasteurella pestis* (now known as *Y. pestis* after Yersin) was first documented in 1969 by Cavanaugh and Randall.[187] Further evidence of extracellular multiplication of *Y. pseudotuberculosis* was shown by electron microscopy by Simonet *et al.* in 1990 only in bacteria harboring the *Yersinia* virulence plasmid, which has subsequently been shown to contain the complete TTS and attendant virulence effectors.[186, 188] The net result of the Yop family is to suppress bacterial internalization by phagocytic cells and to promote extracellular replication in an environment free of host-immune surveillance.

Yersinia protein Kinase A: YpkA

YpkA is 730 residues long and was identified as one of the first Yop (also known as YopO) proteins nearly 20 years ago when it was first recognized as a secreted Ser/Thr kinase indispensible to virulence.[189] Initially a substrate for YpkA was not known, but while that search continued, YpkA was subsequently found to interact with the Rho family of GTPases,[190-192] an activity that was found to correspond to an inhibition of actin function within the host cell that was independent of the Ser/Thr kinase activity of YpkA.[193] Interestingly, YpkA has also been shown to interact with G-actin (non-muscle actin) strongly enough to inhibit actin polymerization at both the extreme C- and N-terminal ends of the YpkA protein and that actin binding is necessary and sufficient for kinase activation in YpkA, which in the absence of actin is otherwise kinase-inactive.[69] The interaction between actin and YpkA is interesting as the essential domains for G-actin binding are removed by 200-300 residues from the actual kinase domain of the protein. No structural biology of the entire protein has been published to date, but an exceptional characterization of the inhibitory activity of YpkA on the Rho family of

GTPases was the product of a prior graduate student in the Stebbins lab, Gerd Prehna.[194, 195] In his graduate work Dr. Prehna demonstrated that a 200 residue C-terminal domain of YpkA constitutes a GDI domain that inhibits members of the Rho family of GTPases. It has also recently been shown that a target for the kinase activity of YpkA is $G\alpha_q$,[196] which together with Rho GDI activity both explain how YpkA activity could disrupt cytoskeletal function in host cells. Secondary effects from YpkA-mediated dysfunction of cellular pathways include actin polymerization inhibition leading to decreased Fc-γ recepter internalization and inhibition of $G\alpha_q$, which has been purported to contribute to the profound bleeding phenotype of *Y. pestis* plague.[197, 198] While a significant amounts of structural characterization for both the GDI and kinase activites of YpkA have been published, there remains a large gap in the understanding of the kinase domain function in YpkA. How actin interacts with and modulates the kinase activity is a very interesting question as well as a very unique biochemical activity that has not been observed in the literature outside of YpkA.

Prior work by Gerd Prehna to identify the sub-domains of YpkA had determined that a domain containing the kinase motif of YpkA was contained within residues 115-428. Once identified this domain was initially screened for crystallization without success. Additionally two compounds had been identified using a high-throughput screening protocol to determine molecules that increased the stability of YpkA in solution. The basis behind the screen for these small molecules empirically tested the T_m of YpkA in the presence of each of a large number of small molecules. Positive interactions that increased the stability of the YpkA construct were identified by an increase in the T_m of the YpkA-ligand complex above the baseline T_m for YpkA alone. This screen identified

two compounds that were then used for later crystallization trials described below: demnacanthal and quercetin dihydrate. Studies have demonstrated a strong correlation between increased stability of protein constructs (as measured by T_m) and successful crystallization by vapor diffusion, and it was believed that use of the molecules identified in the melting-temperature screen would do the same for YpkA.[199]

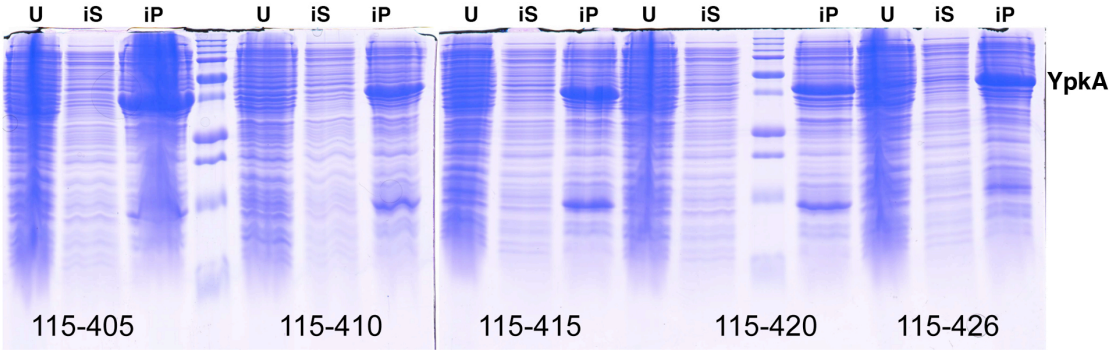


Figure Appendix II-10. Induction and solubility trials of C-terminal truncations of YpkA kinase domain. For each construct three lanes are shown indicating un-induced cells (**U**), the induced soluble fraction (**iS**) and the induced insoluble/pellet fraction (**iP**).

Initial attempts to identify the minimal soluble kinase domain of YpkA were not terribly enlightening as they confirmed the 115-428 construct that had already been partially characterized previously by Dr. Prehna in his graduate studies was the most likely minimal domain (Fig. Appendix II-10). N-terminal truncations from 115 were not attempted as the secondary structure prediction of full-length YpkA showed that residue 115 was at the beginning of a stretch of amino acids with a very high probability to form an α-helix. Instead empiric C-terminal truncations were made to determine whether a shorter construct could be identified. In total 5 constructs were tested for soluble expression (115-405, -410, -415, -420, and -426). All five over-expressed well but were found exclusively in the insoluble fraction following cell lysis. (Fig. Appendix II-10)

Once the minimal domain had been established, crystallization studies were pursued on YpkA 115-428 as the optimal kinase domain construct.

Purification of YpkA 115-428 was very successful, with results similar to the prior work of Dr. Prehna mentioned above; however, given the initial inability to crystallize this domain, several different approaches were utilized as described below. Briefly YpkA was overexpressed recombinantly in *E. coli* with a 3C-protease cleavable 6x-His purification tag. Upon initial purification over Ni-NTA, the 6x-His tag was cleaved and purified away from the 115-428 constructs by a second round of Ni-NTA purification. Subsequent purification steps included FPLC gel-filtration followed by ion-exchange chromatography. Interestingly, when this purification protocol was followed, the homogeneous material obtained from gel-filtration separated into two discrete peaks on anion exchange chromatography. (Fig. Appendix II-11A) Likely a result of a post-translational modification of at least two different sites, the specific chemistry behind the separation of these peaks was not explored much further. The two peaks were designated "early" and "late" and repeat purification steps showed that the nature of the chemical difference between these two populations of YpkA 115-428 was stable, i.e. following the first round of anion exchange, the "late" peak could be passed over gel filtration again, re-concentrated, purified over anion exchange a second time, and would still elute as a single "late" peak. Reductive alkylation of YpkA 115-428 prior to anion exchange eliminated the division of the protein into two populations (not shown) and would instead yield a single peak. This result simply demonstrates that whatever the modification causing the different behavior of the native YpkA 115-428 on anion exchange, it could be reversed by the chemistry involved with reductive alkylation. Experiments such as mass-

spec analysis or empiric treatment with phosphatases and glycosylases to attempt to reverse the modification were considered but not pursued. With respect to the discussion of T_m above, it could be highly productive to revisit these results to assess the T_m of the early vs. late vs. reductively alkylated populations to assess which, if any, would be most likely to crystallize. SDS-PAGE of these three samples – RA-YpkA, 115-428, "early" YpkA 115-428, and "late" YpkA 115-428 – did not show an appreciable difference in the constitution of these samples with regard to purity or degradation kinetics, so any differences between these samples would be more chemical in nature and could have a great deal of influence on cryatallizeability. (Figure Appendix II-11B)

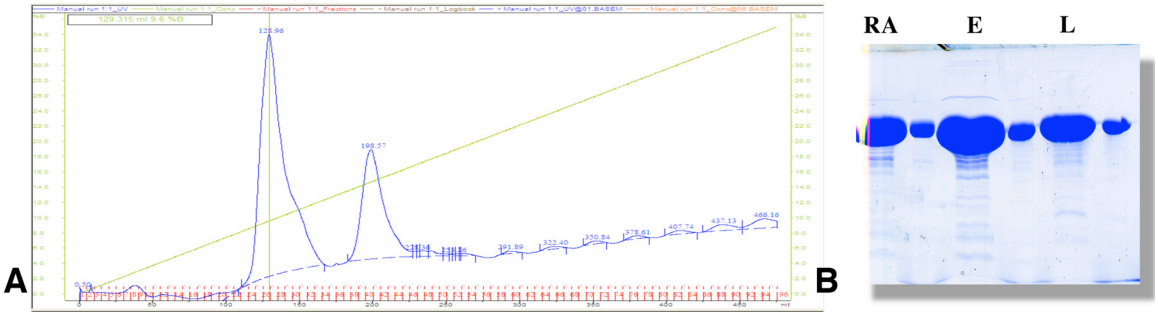


Figure Appendix II-11. Anion exchange chromatography of YpkA 115-428. **A**. Elution profile of YpkA 115-428 from SourceQ (GE Scientific) using FPCL. **B**. SDS-PAGE of final purification of three samples of YpkA 115-428 used for crystallization screening. RA is reductive alkylation (left two lanes), E ("early") middle two lanes, and L ("late") right two lanes. The left lane of each of the three pairs was loaded with 10 µl of protein sample, the right lane of each pair loaded with 2 µl of sample.

Reductively alkylated, "early", and "late" YpkA 115-428 samples were screened for crystallization using sparse matrix screening reagents as above for SopB; however, no significant hits were observed for these samples in-house. Also as with SopB above, samples were submitted for 1536-condition high throughput screening at the Hauptman-

Woodward Medical Research Institute. YpkA is predicted to be a cytosolic protein and specifically not membrane-assocaited, so the screens used were a non-membrane protein crystalliztion formulation. RA YpkA 115-428 (20 mg/ml) sent for high-throughput screening yielded only negative results. Similarly negative results were seen for "early" YpkA 115-428 (20 mg/ml) with 1 mM of either of the two stabilizing compounds mentioned above. Finally, attempts to co-crystallize YpkA 115-428 AMP-PNP (a non-hydrolyzable ATP analogue) were also pursued at the high throughput facility. Magnesium was included in the initial crystallization mix with AMP-PNP as ATP binding is often $Mg2^{+}$-dependent. This screen yielded a variety of crystallization hits (Fig. Appendix II-12 A), but close inspection revealed identical crystal morphology across a variety of crystallization conditions. Moreover, all the crystallization conditions of the screen that gave positive hits for YpkA 115-428 with Mg^{2+} and ATP had some form of phosphate. Like the examples above in SopB with divalent calcium, these results raised a very high suspicion that all the positive results were false-negative salt crystals likely of magnesium phosphate, which has a solubility constant that is six orders of magnitude lower than the original $MgCl_2$-hexahydrate that was originally used to introduce $Mg2^{+}$ ions into the crystallization mix. These suspicions were clearly confirmed by UV microscopy, which showed crystals completely transparent to UV light. (Fig Appendix II-12 B and C)

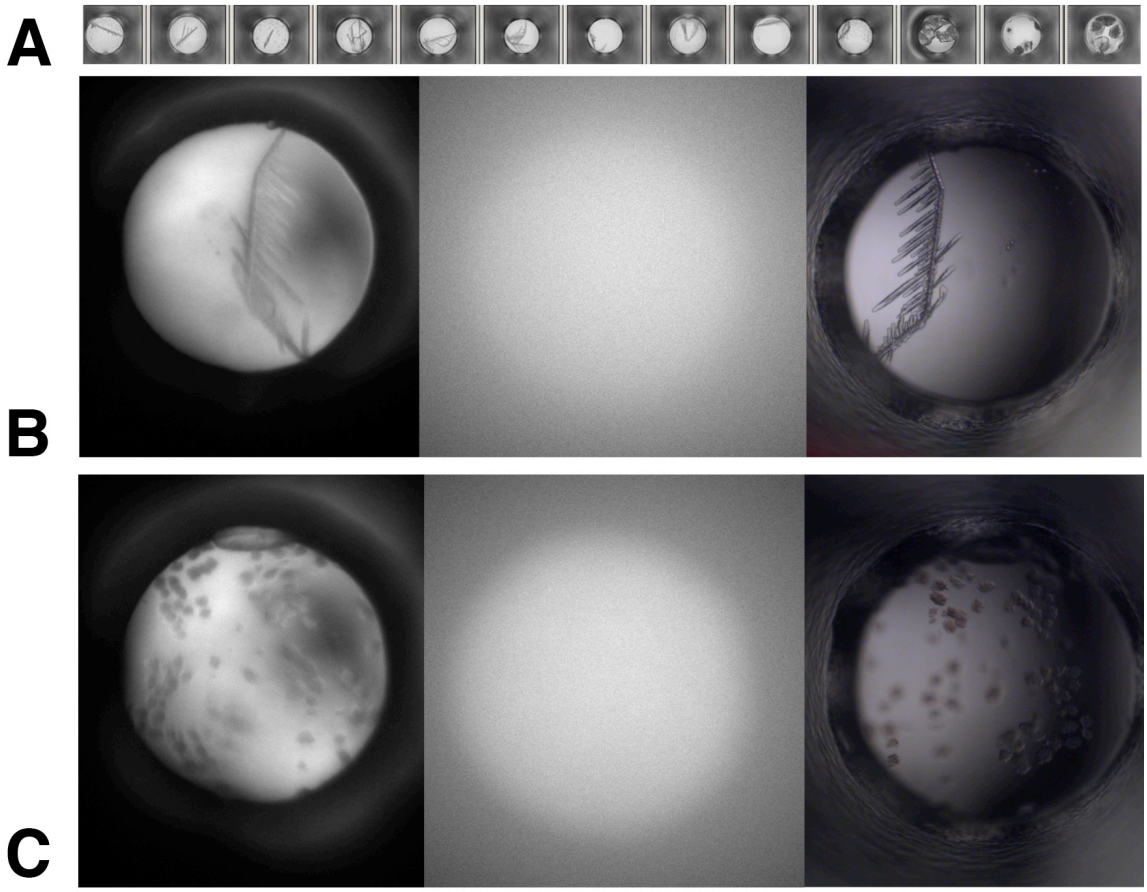


Figure Appendix II-12. High-throughput crystallization screening of YpkA 115-428 with Mg^{2+} and AMP-PNP. Characteristic positive hits for this screen are shown in **A**. Panels B and C both show three views of the same hit showing, from left to right, gray-scale bright-field, ultraviolet, and color bright-field views of the well. Note UV microscopy detects proteinaceous crystalline material. Clear images as shown in B and C indicate salt crystals, in this case likely due to the presence of the Mg^{2+}. **B** shows a condition containing 50% PEG/Tascimate pH 6.8 (Hampton) and 50% of the following mixture: 0.25% w/v 1,2-Diaminocyclohexane sulfate, 0.25% w/v 1,4-Cyclohexanedicarboxylic acid, 0.25% w/v Methylenediphosphonic acid, 0.25% w/v Sulfanilic acid, and 0.02 M HEPES sodium pH 6.8. **C** shows a condition containing Sodium dihydrogen phosphate, Potassium dihydrogen phosphate, Na- HEPES pH 7.5, and 25% (v/v) Glycerol-anhydrous.

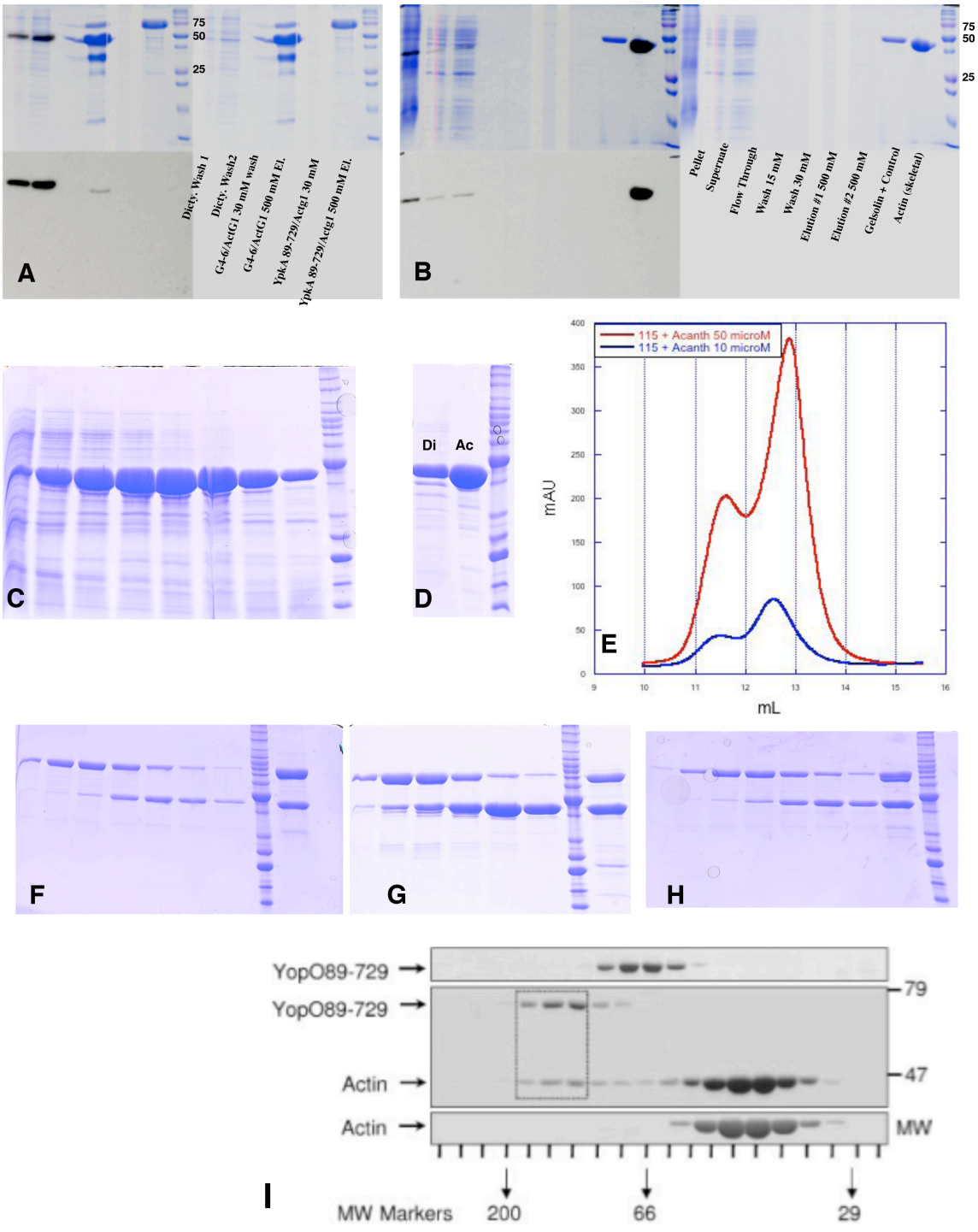
Several additional strategies were chosen to pursue structural information of the YpkA kinase domain. The first of which was to consider crystallization of a much larger construct, the second and third involved crystallizing YpkA in the presence of additional proteins known to associate with YpkA with the hypothesis that a protein complex could be more stable and thus more crystallizeable. Dr. Prehna had successfully crystallized the C-terminus of YpkA in complex with Rac1, so the possibility was explored that the inclusion of the C-terminal domain could stabilize the whole construct and results in favorable crystallization. Furthermore, since the pervious structure of YpkA had been accomplished as a complex with Rac1 crystallization screens were attempted with YpkA 115-731 that was purified with or without the co-expression of Rac1. These trials were conducted using the Crystal Phoenix (Art Robbins Instruments) robotic crystallization tool allowing a wide array of conditions to be samples in multiple drop:well ratios. No significant hits were identified from these attempts. The potential results of a crystal structure of the full 115-731 YpkA would provide interesting insight into the state of TTS virulence factors as whole proteins as all the structures to date of this family of proteins have been limited to single domains. It would also be of great significance to understand how domains interact, particularly when the substrate for one domain – e.g. Rac1 in this case – is bound. In this scenerio what does the other domain do/look like in the presence of Rac1 binding. (Note, for all experiments listed above, the original cDNA used for subcloning was the sequence originally cloned by Gerd Prehna for his thesis research and comprised a YpkA gene numbered 1-731. For all experiments listed below, the YpkA construct used for the actin binding experiments by Trasak *et al*. was kindly provided by said authors).[69]

As a means of pursuing these questions and in a manner similar to the above experiments described for Rac1, the decision was made to attempt to co-crystallize YpkA in the presence of actin. Stable YpkA-actin complexes that could be purified on gel-filtration had been shown in the literature,[69] and studies had shown that regions of YpkA interact with actin.[200] (see Fig Appendix II-13I) These experiments would require a source of actin, and specifically non-muscle actin in the case of YpkA. The initial experiment to produce a YpkA-actin complex involved co-expression of the human form of non-muscle gamma-1 actin (hACTG1). A synthetic gene was produced using the cDNA sequence of ACTG1 and codon-optimized for expression in *E. coli*. This construct was then co-expressed using a Duet two site expression plasmid with YpkA 89-729, the exact construct that had been shown in the literature to form a complex with actin. Unfortunately no co-expression of hACTG- was observed in these experiments as detected by western blot analysis. (Fig. Appendix II-13A) Co-expression of hACTG1 with gelsolin, another actin binding protein, also failed to show significant actin production. Actin has several unique post-translational modifications including a 3-methyl-histidine at position 73 that is conserved across almost all known actin homologues with the exception being some strains of yeast.[201, 202] Perhaps for this reason, for its polymer qualities, or for other unknown reasons it has historically been difficult to recombinately express actin in any typical protein expression systems. Since there are no simple sources of non-muscle actin, the first and most obvious step was the above experiment despite the slim chance of success.

Non-muscle actin is more difficult to obtain in large quantities because, unlike muscle actin, it is present at much lower levels in cells that express these actin iso-forms than the

levels of muscle actin in myocytes. Cell types that utilize non-muscle actin do so to dynamically alter cell shape and morphology. Mobile cell types such as macrophages and neutrophils require non-muscle actin, and phagocytosis/endocytosis in endothelial and neuronal cells similarly involve the focal polymerization of non-muscle actins. Mobile unicellular organisms are also a prolific source of non-muscle actins, so model systems such as *Dictyostelium* and *Acanthamoeba* are often used to study actin dynamics. Various protocols exist for purifying non-muscle actin from these organisms, and an interesting pull-down method of non-muscle actin purification from fibroblasts using recombinant his-tagged Gelsolin has also been described.[70, 203] Specifically, the YpkA-actin binding studies by Trasak *et al*. mentioned above used *Dictyostelium discoidium* actin. In the face of an inability to express non-muscle actin recombinantly, both of these approaches – native expression from *D. discoidium* and *Acanthamoeba* as well as pull-down purification using 6x-his-Gelsolin and 6x-his-YpkA89-729 – were pursued as a means of acquiring crystallization amounts of non-muscle actin.

Figure Appendix II-13 Purification and complex formation of non-muscle actin with YpkA. **A**. From left to right: initial purifications of non-muscle actin from *D. discoidium*(two lanes), 30mM inidazole wash and 500 mM imidazole elution from Ni-NTA resin of ACTG1/Gelsolin coexpression, 30 mM imidazole wash and 500 mM elution of ACTG1/YpkA89-729 coexpression. Panel is divided as follows: anti-Acin Western (lower left), SDS-PAGE (upper-right), and overlay (upper left) of actin purification. **B**. Pull-down purification of *D. discoidium* actin using his-tagged Gelsolin. Panel divisions as in A. **C**. SDS-PAGE of eluted fractions from initial DE52 ion-exchange purification of *D. discoidium* actin showing several actin-rich fractions (prominent band). **D**. SDS-PAGE of final purified *D.discoidium* (Di) and *Acanthamoeba* (Ac) actin. **E**. Representative SD-200 gel-filtration elution profile for YpkA89-729 mixed with *Acanthamoeba* actin showing two elution peaks and no void peak, which would be at ~8.0 mL for this column. **F, G, and H**. SDS-PAGE of eluted gel-filtration fractions from mixtures of YpkA89-729 with D.discoidium (F), Acanthamoeba (G), and Rabbit skeletal (H) actin. Notably no right-shift in the YpkA or actin profiles was seen, and no complex formation was observed. **I**. Complex formation between YpkA 89-729 and D. discoidium actin as published by Trasek *et al*.[69] A clear left-shift is seen in the mixed sample, and fractions that contain both actin and YpkA are clearly seen.


A
75
50
25
Dicty. Wash 1
Dicty. Wash2
G4-6/ActG1 30 mM wash
G4-6/ActG1 500 mM El.
YpkA 89-729/Actg1 30 mM
YpkA 89-729/Actg1 500 mM El.
B
75
50
25
Pellet
Supernate
Flow Through
Wash 15 mM
Wash 30 mM
Elution #1 500 mM
Elution #2 500 mM
Gelsolin + Control
Actin (skeletal)
C
D
Di
Ac
E
115 + Acanth 50 microM
115 + Acanth 10 microM
mAU
mL
F
G
H
I
YopO89-729
YopO89-729
Actin
Actin
79
47
MW
MW Markers
200
66
29

D. discoidium stock cultures were provided by Laura Macro of the Simon Lab at Rockefeller and *Acanthamoeba* stock cultures were obtained from the ATCC and the Mullins lab at UCSF. A plasmid containing the gelsolin construct described above was kindly provided by Ohki *et al*. and used as an actin-binding pull-down reagent with actin-rich lysates from *Dictyostelium* and *Acanthamoeba* with very limited success. Purified gelsolin was added to cleared lysates of *D. discoidium* and Ni-NTA pull-down assays were attempted; however, little to none of the original gelsolin was recovered (Fig. Appendix II-13 A-B). The most likely explanation for these results was rapid degradation due to high levels of protease in the cellular lysates. Similar results were seen when using YpkA as a pull-down reagent (not-shown). Attempts to increase protease inhibition were unsuccessful and ultimately this approach was abandoned for plans to purify non-muscle actin biochemically from *D. discoidium* and *Acanthamoeba* lysates.

Large volume cultures of *D. discoidium* and *Acanthamoeba* were grown without significant difficulty and large-scale purification of actin was attempted using the published protocol of Zuchero *et al*. for *Acanthamoeba* and also the unpublished protocol used by Trasak *et al*. for their work with YpkA (personal communication, see methods). Essentially both protocols use bulk anion exchange resin as a first pass purification step for cleared cell lysates followed by rounds of polymerization, ultracentrifugation, and depolymerization to yield purified actin monomers. While it was certainly the case that a significant amount of actin could be obtained from these lysates and that fractions containing enriched concentrations of actin were obtained from the initial anion exchange purification steps, the subsequent purification steps proved difficult to execute

successfully. (Fig. Appendix II-13 C). Whereas both protocols that were used stated that polymerization post-anion exchange should yield readable visible polymerization of actin, this was never observed in multiple purification attempts. Some material did seem to polymerize, and an even small percentage of the pelleted polymerized actin would de-polymerize, so the final yield from these attempts was frustratingly small but sufficient for a few preliminary gel-filtration studies. (Fig. Appendix II-13D) Additional technical difficulties included clogged ion-exchange columns and bacterial/fungal contamination in dialysis/storage steps of the protocol. Alex Kelly of the Funabiki lab at Rockefeller (formerly of the Mullins lab at UCSF) kindly obtained a sample of purified *Acanthamoeba* actin from the Mullins lab to use as a standard and initial quality control. (Fig. Appendix II-13D) Once monomeric forms of the two non-muscle actins were prepared, YpkA89-729 was dialyzed into actin G-buffer and mixed with monomeric actin. Complex formation was assessed by gel-filtration in a manner as consistent with the protocol of Trasak et al. as technically possible. The one difference between these studies and those published by Trasak et al. regard the size/volume of the gel-filtration column. The published studies of YpkA-Actin complex formation used analytical scale gel-filtration columns (3.2 x 300 mm SD200), which were not found to be available in any lab that was contacted at Rockefeller, Memorial Sloan Kettering, and Cornell. Instead a small preparative scale gel-filtration column was used that had significantly higher bed volumes (30 x 300 mm SD200) and thus required a much higher sample volume to achieve equivalent concentrations of proteins on the column. In personal communication Trasak *et al*. reported they had successfully reproduced the experiment with preparative scale columns, but direct data from these experiments was not available

for comparison. Ultimately, attempts to reproduce the published binding results between YpkA 89-729 and *Dictyostelium* non-muscle actin were unsuccessful, despite the use of the exact same construct from the published results, the same source of non-muscle actin, and nearly identical protocols. (Fig Appendix II-13E and F vs I). No left-shift in the elution volume of either YpkA nor actin was observed, and co-elution of YpkA and actin was not observed. The elution volumes of actin and YpkA were both appropriate for a monodisperse species in solution suggesting that the buffer conditions were appropriate to limit polymerization of actin and also were not unfavorable to YpkA solubility. These negative results – i.e. no YpkA-actin complex formation – were again seen with *Acanthamoeba* actin (an experiment predicted to be a second positive control) and with rabbit skeletal muscle actin (an experiment designed to be a negative control). (Fig Appendix II-13 G, and H respectively) There are many variables that remain to be optimized in these experiments. First among them is the purification of actin from both *Dictyostelium* and *Acanthamoeba*. The lack of clear polymerization of actin during the purification process raises suspicions about the activity of the actin that was purified originally. Strict adherance to the protocol was maintained at all times, but this is not to say that some significant step is being overlooked either in execution of the steps of purification or in preparation of reagents. Future collaborations with a lab that performs these types of experiments regularly will be necessary for future work in this area. Second is the significance of the size of gel-filtration column. It is clear from the published data regarding YpkA-actin complex formation that only a fraction of the material loaded onto the column forms a complex. This suggests that the affinity of the complex between YpkA and actin might be relatively low and needs the high local

concentration provided by the smaller diameter analytic gel-filtration to maintain a population of the two proteins in a complexed state. It would be very interesting to know whether the collected fractions that had both YpkA and actin would stay in complex when ran again over the same column, or whether the dilution factor involved in this experiment would ultimately cause dissociation of the complex. These are just a few of the important questions that can be answered in a more robust and objective way once an experimental protocol is established that is reproducible. Finally, having these reagents prepared, i.e. monomeric actin and YpkA89-729, it would be possible to conduct an entirely different set of experiments that were alluded to above regarding the interaction of between the substrates of these multidomain virulence factors. Clear binding domains for Rac1 and actin have been established for YpkA, but whether these two proteins can bind YpkA at the same time is a question for which there is virtually no data. The consequences of different binding affinities for these two substrates is very interesting to consider as it could suggest that there is a "hard-wired" set of effects that YpkA will have on the cell depending on the state of the cell at any given time. It could switch between Rac1 inhibition if actin concentrations are low, but switch to actin-induced kinase activity of $G\alpha_q$ if Rac1 is for some reason low or otherwise inhibited by other YpkA molecules. How different affinities for different substrates would orchestrate the multitude of downstream effects that are mediated by these molecules can only be speculated, but having these tools in place could prove a valuable first step towards answering these questions. Finally, the issue of attempting to purify and/or generate a complex between YpkA and the substrate of its kinase activity remains a completely unexplored field. Gαq has been clearly identified as a target of actin-activated YpkA kinase activity. Having

purified actin, Rac1, and hopefully Gαq could add yet another level of complexity to the above discussion that would shed valuable light on the integration of these various signaling mechanisms mediated by YpkA inside host cells.

Summary:

Structural studies of the YpkA kinase domain remain in their infancy with attempts to study the structural biology of the kinase domain experiencing very limited success to date. The reasons for the lack of success are still unclear at this point. Multiple approaches have been successful at generating soluble, well-behaved YpkA protein both of a limited kinase domain and also of the multi-domain constructs. The success of the YpkA GDI-domain structure in complex with Rac1 suggest that such approaches can succeed provided the correct empiric combination of variables. There are many questions that remain unanswered regarding the activity of YpkA as a whole – i.e. both the GDI domain and the kinase domain together – as well as several important questions about the mechanism of the kinase domain. Of particular interest is how recognition and binding of actin at sites very remote by residue number to the kinase domain could activate the activity of YpkA towards Gαq. The above studies outline several approaches that will be pursued in the future.

References

1. World-Health-Orginization. in World Health Orginization Fact Sheets (2010).
2. World-Health-Orginization. (2011).
3. Maudlin, I. African trypanosomiasis. *Ann. Trop. Med. Parasitol.* **100**, 679-701 (2006).
4. Malvy, D. & Chappuis, F. Sleeping sickness. *Clin Microbiol Infect* **17**, 986-95 (2011).
5. Dias, J.C., Silveira, A.C. & Schofield, C.J. The impact of Chagas disease control in Latin America: a review. *Mem Inst Oswaldo Cruz* **97**, 603-12 (2002).
6. Barrett, M.P. & Gilbert, I.H. Perspectives for new drugs against trypanosomiasis and leishmaniasis. *Curr Top Med Chem* **2**, 471-82 (2002).
7. Paulnock, D., Freeman, B. & Mansfield, J. Modulation of innate immunity by African Trypanosomes. *Parasitology* **137**, 2051-2063 (2010).
8. Figueiredo, L.M., Cross, G.A. & Janzen, C.J. Epigenetic regulation in African trypanosomes: a new kid on the block. *Nat Rev Microbiol* **7**, 504-13 (2009).
9. Knoop, V. When you can't trust the DNA: RNA editing changes transcript sequences. *Cell Mol Life Sci* **68**, 567-86 (2011).
10. Berriman, M. et al. The genome of the African trypanosome Trypanosoma brucei. *Science* **309**, 416-22 (2005).
11. Benne, R. et al. Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* **46**, 819-26 (1986).
12. Sturm, N.R. & Simpson, L. Kinetoplast DNA minicircles encode guide RNAs for editing of cytochrome oxidase subunit III mRNA. *Cell* **61**, 879-84 (1990).
13. Borst, P. & Sabatini, R. in Annu Rev Microbiol 235-51 (2008).
14. van Leeuwen, F. et al. beta-D-glucosyl-hydroxymethyluracil is a conserved DNA modification in kinetoplastid protozoans and is abundant in their telomeres. *Proc Natl Acad Sci USA* **95**, 2366-71 (1998).
15. Cross, M. et al. in Mol Microbiol 37-47 (2002).

16. Genest, P.-A., Ter Riet, B., Cijsouw, T., van Luenen, H.G.A.M. & Borst, P. Telomeric localization of the modified DNA base J in the genome of the protozoan parasite Leishmania. *Nucleic Acids Res* **35**, 2116-24 (2007).
17. Genest, P.-A. et al. Formation of linear inverted repeat amplicons following targeting of an essential gene in Leishmania. *Nucleic Acids Res* **33**, 1699-709 (2005).
18. Palenchar, J. & Bellofatto, V. Gene transcription in trypanosomes. *Molecular and Biochemical Parasitology* **146**, 135-41 (2006).
19. CRICK, F.H. Codon--anticodon pairing: the wobble hypothesis. *Journal of Molecular Biology* **19**, 548-55 (1966).
20. Agris, P., Vendeix, F.A. & Graham, W. tRNA's wobble decoding of the genome: 40 years of modification. *Journal of Molecular Biology* **366**, 1-13 (2007).
21. Luo, M. & Schramm, V.L. Transition state structure of E. coli tRNA-specific adenosine deaminase. *J Am Chem Soc* **130**, 2649-55 (2008).
22. Schaub, M. & Keller, W. RNA editing by adenosine deaminases generates RNA and protein diversity. *Biochimie* **84**, 791-803 (2002).
23. Gerber, A.P. & Keller, W. An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science* **286**, 1146-9 (1999).
24. Rubio, M.A. et al. An adenosine-to-inosine tRNA-editing enzyme that can perform C-to-U deamination of DNA. *Proc Natl Acad Sci USA* **104**, 7821-6 (2007).
25. Auxilien, S., Crain, P.F., Trewyn, R.W. & Grosjean, H. Mechanism, specificity and general properties of the yeast enzyme catalysing the formation of inosine 34 in the anticodon of transfer RNA. *Journal of Molecular Biology* **262**, 437-58 (1996).
26. Conticello, S.G., Langlois, M.-A. & Neuberger, M.S. Insights into DNA deaminases. *Nature Structural & Molecular Biology* **14**, 7-9 (2007).
27. Losey, H., Ruthenburg, A. & Verdine, G. Crystal structure of Staphylococcus aureus tRNA adenosine deaminase TadA in complex with RNA. *Nat Struct Mol Biol* **13**, 153-159 (2006).
28. Kim, J. et al. in Biochemistry 6407-6416 (2006).

29. Kuratani, M. et al. in Journal of Biological Chemistry 16002-16008 (2005).
30. Lee, W. et al. in Proteins 1016-1019 (2007).
31. Elias, Y. & Huang, R.H. Biochemical and structural studies of A-to-I editing by tRNA:A34 deaminases at the wobble position of transfer RNA. *Biochemistry* **44**, 12057-65 (2005).
32. Welin, M., Tresaugues, L., Andersson, J., Arrowsmith, C.H., Berglund, H., Collins, R., Dahlgren, L.G., Edwards, A.M., Flodin, S., Flores, A., Graslund, S., Hammarstrom, M., Johansson, A., Johansson, I., Karlberg, T., Kotenyova, T., Lehtio, L., Moche, M., Nilsson, M.E., Nyman, T., Olesen, K., Persson, C., Sagemark, J., Schueler, H., Thorsell, A.G., van der Berg, S., Wisniewska, M., Wikstrom, M., Nordlund, P., . pdb 3DH1: Crystal structure of human tRNA-specific adenosine-34 deaminase subunit ADAT2. *Structural Genomics Consortium* (2011).
33. Ragone, F.L. et al. The C-terminal end of the Trypanosoma brucei editing deaminase plays a critical role in tRNA binding. *RNA* **17**, 1296-306 (2011).
34. Rubio, M.A., Ragone, F.L., Gaston, K.W., Ibba, M. & Alfonzo, J.D. C to U editing stimulates A to I editing in the anticodon loop of a cytoplasmic threonyl tRNA in Trypanosoma brucei. *J Biol Chem* **281**, 115-20 (2006).
35. Fritz, E.L. & Papavasiliou, F.N. Cytidine deaminases: AIDing DNA demethylation? *Genes & Development* **24**, 2107-14 (2010).
36. Conticello, S.G., Thomas, C.J., Petersen-Mahrt, S.K. & Neuberger, M.S. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Molecular Biology and Evolution* **22**, 367-77 (2005).
37. Keegan, L.P. et al. The properties of a tRNA-specific adenosine deaminase from Drosophila melanogaster support an evolutionary link between pre-mRNA editing and tRNA modification. *Mol Cell Biol* **20**, 825-33 (2000).
38. Spears, J. et al. A Single Zinc Ion Is Sufficient for an Active Trypanosoma brucei tRNA Editing Deaminase. *Journal of Biological Chemistry* **286**, 20366-20374 (2011).

39. Liaw, S. Crystal Structure of Bacillus subtilis Guanine Deaminase: THE FIRST DOMAIN-SWAPPED STRUCTURE IN THE CYTIDINE DEAMINASE SUPERFAMILY. *Journal of Biological Chemistry* **279**, 35479-35485 (2004).
40. Spears, J.L. et al. A Single Zinc Ion Is Sufficient for an Active Trypanosoma brucei tRNA Editing Deaminase. *Journal of Biological Chemistry* **286**, 20366-20374 (2011).
41. Evans, G. & Pettifer, R.F. in J Appl Crystallogr 82-86 (2001).
42. Z. Otwinowski, W.M. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology* **276**, 19 (1997).
43. Kantardjieff, K.A. & Rupp, B. in Protein Science 1865-1871 (2003).
44. MATTHEWS, B.W. in Journal of Molecular Biology 491-& (1968).
45. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
46. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
47. Vagin, A. et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* **60**, 2184-2195 (2004).
48. Strong, M. et al. Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc Natl Acad Sci USA* **103**, 8060-5 (2006).
49. Painter, J. & Merritt, E.A. TLSMD web server for the generation of multi-group TLS models. *Journal of Applied Crystallography* **39**, 109-111 (2006).
50. Howlin, B., Butler, S.A., Moss, D.S., Harris, G.W. & Driessen, H.P.C. TLSANL: TLS parameter-analysis program for segmented anisotropic refinement of macromolecular structures. *Journal of Applied Crystallography* **26**, 622-624 (1993).
51. Stein, N. in J Appl Crystallogr 641-643 (2008).
52. BAILEY, S. in Acta Crystallogr D 760-763 (1994).

53. Langer, G., Cohen, S.X., Lamzin, V.S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* **3**, 1171-9 (2008).
54. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).
55. Colovos, C. & Yeates, T.O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* **2**, 1511-9 (1993).
56. Guex N, P.M. SWISS_MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 10 (1997).
57. McNicholas, S., Potterton, E., Wilson, K.S. & Noble, M.E. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr* **67**, 386-94 (2011).
58. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545-9 (2010).
59. Kelley, L. & Sternberg, M. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363-71 (2009).
60. Costantini, S., Colonna, G. & Facchiano, A.M. ESBRI: a web server for evaluating salt bridges in proteins. *Bioinformation* **3**, 137-8 (2008).
61. Dombkowski, A.A. Disulfide by Design™ : a computational method for the rational design of disulfide bonds in proteins. *Bioinformatics* **19**, 1852-1853 (2003).
62. Sorenson, M.K. & Darst, S.A. Disulfide cross-linking indicates that FlgM-bound and free sigma28 adopt similar conformations. *Proc Natl Acad Sci USA* **103**, 16722-7 (2006).
63. Carlo Lapid, Y.G. (2006).
64. Liu, H. & Naismith, J.H. in Bmc Biotechnol 91 (2008).
65. Simossis, V.A. & Heringa, J. PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res* **33**, W289-94 (2005).

66. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876-82 (1997).
67. Di Tommaso, P. et al. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res* **39**, W13-7 (2011).
68. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-10 (1990).
69. Trasak, C. et al. Yersinia Protein Kinase YopO Is Activated by A Novel G-actin Binding Process. *Journal of Biological Chemistry* **282**, 2268 (2007).
70. Zuchero, J.B. In vitro actin assembly assays and purification from Acanthamoeba. *Methods Mol Biol* **370**, 213-26 (2007).
71. Ashworth, J.M. & Watts, D.J. Metabolism of the cellular slime mould Dictyostelium discoideum grown in axenic culture. *Biochem J* **119**, 175-82 (1970).
72. Bujacz, G., Wrzesniewska, B. & Bujacz, A. Cryoprotection properties of salts of organic acids: a case study for a tetragonal crystal of HEW lysozyme. *Acta Crystallogr D Biol Crystallogr* **66**, 789-96 (2010).
73. Lusty, C.J. in J Appl Crystallogr 106-112 (1999).
74. Teh, A.H. et al. The 1.48 A resolution crystal structure of the homotetrameric cytidine deaminase from mouse. *Biochemistry* **45**, 7825-33 (2006).
75. Hou, H.F., Liang, Y.H., Li, L.F., Su, X.D. & Dong, Y.H. Crystal structures of Streptococcus mutans 2'-deoxycytidylate deaminase and its complex with substrate analog and allosteric regulator dCTP x Mg2+. *Journal of Molecular Biology* **377**, 220-31 (2008).
76. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* **372**, 774-97 (2007).
77. Silvian, L.F., Wang, J. & Steitz, T.A. Insights into editing from an ile-tRNA synthetase structure with tRNAile and mupirocin. *Science* **285**, 1074-7 (1999).
78. Heras, B. et al. in Structure 139-145 (2003).

79. Sam, M., Abbani, M., Cascio, D., Johnson, R. & Clubb, R. Crystallization, dehydration and preliminary X-ray analysis of excisionase (Xis) proteins cooperatively bound to DNA. *Acta Crystallogr F Struct Biol Cryst Commun* **62**, 825-8 (2006).
80. Wenzel, S., Martins, B.M., Rösch, P. & Wöhrl, B.M. Crystal structure of the human transcription elongation factor DSIF hSpt4 subunit in complex with the hSpt5 dimerization interface. *Biochem J* **425**, 373-80 (2010).
81. Hendrickson, W.A., Horton, J.R. & LeMaster, D.M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *The EMBO Journal* **9**, 1665-72 (1990).
82. Dodson, E. in Acta Crystallogr D 1958-1965 (2003).
83. Boggon, T.J. & Shapiro, L. Screening for phasing atoms in protein crystallography. *Structure* **8**, R143-9 (2000).
84. Dauter, Z., Dauter, M. & Dodson, E. in Acta Crystallogr D 494-506 (2002).
85. Meyer, P.A., Ye, P., Zhang, M., Suh, M.H. & Fu, J. Phasing RNA polymerase II using intrinsically bound Zn atoms: an updated structural model. *Structure* **14**, 973-82 (2006).
86. Meyer, P.A., Ye, P., Suh, M.H., Zhang, M. & Fu, J. Structure of the 12-subunit RNA polymerase II refined with the aid of anomalous diffraction data. *J Biol Chem* **284**, 12933-9 (2009).
87. Tumbale, P. et al. Structure of an aprataxin-DNA complex with insights into AOA1 neurodegenerative disease. *Nature Structural & Molecular Biology* **18**, 1189-95 (2011).
88. Wen, H. et al. Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation. *J Biol Chem* **285**, 9322-6 (2010).
89. Spears, J.L. Trypanosoma brucei tRNA Editing Deaminase: Conserved Deaminase Core, Unique Deaminase Features. *Thesis* (2011).
90. Bransteitter, R., Prochnow, C. & Chen, X.S. The current structural and functional understanding of APOBEC deaminases. *Cell Mol Life Sci* **66**, 3137-47 (2009).

91. Holden, L.G. et al. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature* **456**, 121-4 (2008).
92. Nesić, D. et al. Helicobacter pylori CagA inhibits PAR1-MARK family kinases by mimicking host substrates. *Nature Structural & Molecular Biology* **17**, 130-2 (2010).
93. Marshall, B. in Chemmedchem 783-802 (2006).
94. Warren, J.R. in Chemmedchem 672-+ (2006).
95. Odenbreit, S. et al. Translocation of Helicobacter pylori CagA into gastric epithelial cells by type IV secretion. *Science* **287**, 1497-500 (2000).
96. Blaser, M.J. et al. Infection with Helicobacter pylori strains possessing cagA is associated with an increased risk of developing adenocarcinoma of the stomach. *Cancer Res* **55**, 2111-5 (1995).
97. Gunn, M., Stephens, J.C., Thompson, J.R., Rathbone, B.J. & Samani, N.J. Significant association of cagA positive Helicobacter pylori strains with risk of premature myocardial infarction. *Heart* **84**, 267-71 (2000).
98. Gabrielli, M. et al. CagA-positive cytotoxic H. pylori strains as a link between plaque instability and atherosclerotic stroke. *European Heart Journal* **25**, 64-8 (2004).
99. Pietroiusti, A., Giuliano, M., Magrini, A., Bergamaschi, A. & Galante, A. Cytotoxin-associated gene A strains of Helicobacter pylori represent a risk factor for the development of microalbuminuria in type 2 diabetes. *Diabetes Care* **29**, 1399-401 (2006).
100. Murata-Kamiya, N. Pathophysiological functions of the CagA oncoprotein during infection by Helicobacter pylori. *Microbes Infect* **13**, 799-807 (2011).
101. Covacci, A. et al. Molecular characterization of the 128-kDa immunodominant antigen of Helicobacter pylori associated with cytotoxicity and duodenal ulcer. *Proc Natl Acad Sci USA* **90**, 5791-5 (1993).
102. Naito, M. et al. Influence of EPIYA-repeat polymorphism on the phosphorylation-dependent biological activity of Helicobacter pylori CagA. *Gastroenterology* **130**, 1181-90 (2006).

103. Saadat, I. et al. Helicobacter pylori CagA targets PAR1/MARK kinase to disrupt epithelial cell polarity. *Nature* **447**, 330-3 (2007).
104. Marx, A. et al. Structural variations in the catalytic and ubiquitin-associated domains of microtubule-associated protein/microtubule affinity regulating kinase (MARK) 1 and MARK2. *J Biol Chem* **281**, 27586-99 (2006).
105. Wang, Y., Huang, H., Sun, M.a., Zhang, Q. & Guo, D. T3DB: an integrated database for bacterial Type III Secretion System. *BMC bioinformatics* **13**, 66 (2012).
106. Michiels, T. & Cornelis, G.R. Secretion of hybrid proteins by the Yersinia Yop export system. *J Bacteriol* **173**, 1677-85 (1991).
107. Salmond, G.P. & Reeves, P.J. Membrane traffic wardens and protein secretion in gram-negative bacteria. *Trends in Biochemical Sciences* **18**, 7-12 (1993).
108. Rosqvist, R., Forsberg, A. & Wolf-Watz, H. Intracellular targeting of the Yersinia YopE cytotoxin in mammalian cells induces actin microfilament disruption. *Infect Immun* **59**, 4562-9 (1991).
109. Hueck, C.J. Type III Protein Secretion Systems in Bacterial Pathogens of Animals and Plants. *Microbiology and Molecular Biology Reviews* **62**, 379 (1998).
110. Lindeberg, M., Cunnac, S. & Collmer, A. Pseudomonas syringae type III effector repertoires: last words in endless arguments. *Trends in Microbiology* **20**, 199-208 (2012).
111. Nguyen, L., Paulsen, I.T., Tchieu, J., Hueck, C.J. & Saier, M.H., Jr. Phylogenetic analyses of the constituents of Type III protein secretion systems. *J Mol Microbiol Biotechnol* **2**, 125-44 (2000).
112. Blocker, A., Komoriya, K. & Aizawa, S.-I. Type III secretion systems and bacterial flagella: insights into their function from structural similarities. *Proc Natl Acad Sci USA* **100**, 3027-30 (2003).
113. Soscia, C., Hachani, A., Bernadac, A., Filloux, A. & Bleves, S. Cross talk between type III secretion and flagellar assembly systems in Pseudomonas aeruginosa. *J Bacteriol* **189**, 3124-32 (2007).
114. Namba, K., Yamashita, I. & Vonderviszt, F. Structure of the core and central channel of bacterial flagella. *Nature* **342**, 648-54 (1989).

115. Botteaux, A., Sani, M., Kayath, C.A., Boekema, E.J. & Allaoui, A. Spa32 interaction with the inner-membrane Spa40 component of the type III secretion system of Shigella flexneri is required for the control of the needle length by a molecular tape measure mechanism. *Mol Microbiol* **70**, 1515-28 (2008).
116. Epler, C.R., Dickenson, N.E., Bullitt, E. & Picking, W.L. Ultrastructural Analysis of IpaD at the Tip of the Nascent MxiH Type III Secretion Apparatus of Shigella flexneri. *J Mol Biol*.
117. Shen, D.K., Moriya, N., Martinez-Argudo, I. & Blocker, A. Needle length control and the secretion substrate specificity switch are only loosely coupled in the T3S apparatus of Shigella. *Microbiology*.
118. Hayes, C.S., Aoki, S.K. & Low, D.A. Bacterial contact-dependent delivery systems. *Annu Rev Genet* **44**, 71-90 (2010).
119. Bozue, J. et al. A Yersinia pestis YscN ATPase mutant functions as a live attenuated vaccine against bubonic plague in mice. *FEMS Microbiol Lett*.
120. Fujii, T. et al. Structure of a type III secretion needle at 7-Å resolution provides insights into its assembly and signaling mechanisms. *Proc Natl Acad Sci USA* **109**, 4461-6 (2012).
121. Aldridge, P. & Hughes, K.T. How and when are substrates selected for type III secretion? *Trends Microbiol* (2001).
122. Marlovits, T.C. & Stebbins, C.E. Type III secretion systems shape up as they ship out. *Current Opinion in Microbiology* (2009).
123. Sanowar, S. et al. Interactions of the transmembrane polymeric rings of the Salmonella enterica serovar Typhimurium type III secretion system. *MBio* **1** (2010).
124. Marlovits, T.C. et al. Structural insights into the assembly of the type III secretion needle complex. *Science* **306**, 1040-2 (2004).
125. Kline, T. et al. Tethered thiazolidinone dimers as inhibitors of the bacterial type III secretion system. *Bioorg Med Chem Lett* **19**, 1340-3 (2009).
126. Kline, T., Felise, H.B., Sanowar, S. & Miller, S.I. The type III secretion system as a source of novel antibacterial drug targets. *Curr Drug Targets* **13**, 338-51.

127. Blanco-Toribio, A., Muyldermans, S., Frankel, G. & Fernández, L.Á. Direct injection of functional single-domain antibodies from E. coli into human cells. *PLoS ONE* **5**, e15227 (2010).
128. Hegazy, W.A., Xu, X., Metelitsa, L. & Hensel, M. Evaluation of Salmonella enterica type III secretion system effector proteins as carriers for heterologous vaccine antigens. *Infect Immun* **80**, 1193-202.
129. Hegazy, W.A. & Hensel, M. Salmonella enterica as a vaccine carrier. *Future Microbiol* **7**, 111-27.
130. Edgren, T., Forsberg, A., Rosqvist, R. & Wolf-Watz, H. Type III Secretion in Yersinia: Injectisome or Not? *PLoS Pathog* **8**, e1002669 (2012).
131. Evdokimov, A.G. et al. Similar modes of polypeptide recognition by export chaperones in flagellar biosynthesis and type III secretion. *Nat Struct Biol* **10**, 789-93 (2003).
132. Button, J.E. & Galán, J.E. Regulation of chaperone/effector complex synthesis in a bacterial type III secretion system. *Molecular Microbiology* **81**, 1474-83 (2011).
133. Page, A.-L. & Parsot, C. Chaperones of the type III secretion pathway: jacks of all trades. *Molecular Microbiology* **46**, 1-11 (2002).
134. Fattori, J., Prando, A., Martins, A.M., Rodrigues, F.H. & Tasic, L. Bacterial secretion chaperones. *Protein Pept Lett* **18**, 158-66.
135. Knodler, L.A., Bertero, M., Yip, C., Strynadka, N.C. & Steele-Mortimer, O. Structure-based mutagenesis of SigE verifies the importance of hydrophobic and electrostatic residues in type III chaperone function. *Mol Microbiol* **62**, 928-40 (2006).
136. Lilic, M., Vujanac, M. & Stebbins, C. A Common Structural Motif in the Binding of Virulence Factors to Bacterial Secretion Chaperones. *Molecular Cell* **21**, 653-664 (2006).
137. Luo, Y. et al. Structural and biochemical characterization of the type III secretion chaperones CesT and SigE. *Nat Struct Biol* **8**, 1031-6 (2001).
138. Yip, C.K., Finlay, B. & Strynadka, N.C.J. Structural characterization of a type III secretion system filament protein in complex with its …. *Nature Structural & Molecular Biology* (2004).

139. Popoff, M.Y., Bockemühl, J. & Gheesling, L.L. Supplement 2002 (no. 46) to the Kauffmann-White scheme. *Research in Microbiology* **155**, 568-70 (2004).
140. Coburn, B., Grassl, G. & Finlay, B. Salmonella, the host and disease: a brief review. *Immunol Cell Biol* **85**, 112-118 (2007).
141. Mead, P.S. et al. Food-related illness and death in the United States. *Emerging Infect Dis* **5**, 607-25 (1999).
142. McEwen, S.A. & Fedorka-Cray, P.J. Antimicrobial use and resistance in animals. *Clin Infect Dis* (2002).
143. Mulvey, M., Boyd, D., Olson, A., Doublet, B. & Cloeckaert, A. The genetics of Salmonella genomic island 1. *Microbes and Infection* **8**, 1915-1922 (2006).
144. Arlet, G. et al. Salmonella resistant to extended-spectrum cephalosporins: prevalence and epidemiology. *Microbes and Infection* **8**, 1945-1954 (2006).
145. Butaye, P. et al. The clonal spread of multidrug-resistant non-typhi Salmonella serotypes. *Microbes and Infection* **8**, 1891-1897 (2006).
146. Molbak, K. Human health consequences of antimicrobial drug-resistant Salmonella and other foodborne pathogens. *Clin Infect Dis* (2005).
147. Grassl, G.A. & Finlay, B.B. Pathogenesis of enteric Salmonella infections. *Current Opinion in Gastroenterology* (2008).
148. Wijburg, O.L.C., Simmons, C.P. & van Rooijen, N. Dual role for macrophages in vivo in pathogenesis and control of murine Salmonella enterica var. …. *European Journal of Immunology* (2000).
149. Stecher, B., Robbiani, R., Walker, A.W., Westendorf, A.M. & al., e. Salmonella enterica Serovar Typhimurium Exploits Inflammation to Compete with the Intestinal …. *PLoS Biol* (2007).
150. Hensel, M. Evolution of pathogenicity islands of Salmonella enterica. *International Journal of Medical Microbiology* (2004).
151. Hu, Q. et al. Salmonella serovar Senftenberg human clinical isolates lacking SPI-1. *J Clin Microbiol* (2008).
152. Prager, R. et al. Molecular properties of Salmonella enterica serotype paratyphi B distinguish between its systemic and its enteric pathovars. *J Clin Microbiol* **41**, 4270-8 (2003).

153. Haraga, A., Ohlson, M.B. & Miller…, S.I. Salmonellae interplay with host cells. *Nat Rev Microbiol* (2008).

154. Galyov, E.E. et al. A secreted effector protein of Salmonella dublin is translocated into eukaryotic cells and mediates inflammation and fluid secretion in infected ileal mucosa. *Mol Microbiol* **25**, 903-12 (1997).

155. Norris, F.A., Wilson, M.P., Wallis, T.S., Galyov, E.E. & Majerus, P.W. SopB, a protein required for virulence of Salmonella dublin, is an inositol phosphate phosphatase. *Proc Natl Acad Sci USA* **95**, 14057-9 (1998).

156. Marcus, S.L., Knodler, L.A. & Finlay, B.B. Salmonella enterica serovar Typhimurium effector SigD/SopB is membrane-associated and ubiquitinated inside host cells. *Cell Microbiol* **4**, 435-46 (2002).

157. Marcus, S.L., Wenk, M.R., Steele-Mortimer, O. & Finlay, B.B. A synaptojanin-homologous region of Salmonella typhimurium SigD is essential for inositol phosphatase activity and Akt activation. *FEBS Lett* **494**, 201-7 (2001).

158. Di Paolo, G. & De Camilli, P. Phosphoinositides in cell regulation and membrane dynamics. *Nature* **443**, 651-7 (2006).

159. Rogers, L.D., Brown, N.F., Fang, Y., Pelech, S. & Foster, L.J. Phosphoproteomic Analysis of Salmonella-Infected Cells Identifies Key Kinase Regulators and SopB-Dependent Host Phosphorylation Events. *Science Signaling* **4**, rs9-rs9 (2011).

160. Brumell, J.H. & Grinstein, S. Role of lipid-mediated signal transduction in bacterial internalization. *Cellular Microbiology* (2003).

161. Drecktrah, D., Knodler, L.A. & Steele-Mortimer, O. Modulation and utilization of host cell phosphoinositides by Salmonella spp. *Infect Immun* **72**, 4331-5 (2004).

162. Hilbi, H. Modulation of phosphoinositide metabolism by pathogenic bacteria. *Cell Microbiol* **8**, 1697-706 (2006).

163. Knodler, L.A., Finlay, B.B. & Steele-Mortimer, O. The Salmonella effector protein SopB protects epithelial cells from apoptosis by sustained activation of Akt. *J Biol Chem* **280**, 9058-64 (2005).

164. Pendaries, C., Tronchère, H., Arbibe, L., Mounier, J. & al., e. PtdIns (5) P activates the host cell PI3-kinase/Akt pathway during Shigella flexneri infection. *The EMBO Journal* (2006).

165. Uzzau, S. & Fasano, A. Cross-talk between enteric pathogens and the intestine. *Cell Microbiol* **2**, 83-9 (2000).

166. Zhang, S. et al. in Infect Immun 3843-55 (2002).

167. Hernandez, L.D., Hueffer, K., Wenk, M.R. & Galán, J.E. Salmonella modulates vesicular traffic by altering phosphoinositide metabolism. *Science* **304**, 1805-7 (2004).

168. Terebiznik, M.R. et al. Elimination of host cell PtdIns(4,5)P(2) by bacterial SigD promotes membrane fission during invasion by Salmonella. *Nat Cell Biol* **4**, 766-73 (2002).

169. Boyle, E.C. & Finlay, B.B. Leaky guts and lipid rafts. *Trends Microbiol* **13**, 560-3 (2005).

170. Patel, J.C. & Galán, J.E. Differential activation and function of Rho GTPases during Salmonella-host cell interactions. *J Cell Biol* **175**, 453-63 (2006).

171. Zhou, D., Chen, L.M., Hernandez, L., Shears, S.B. & Galán, J.E. A Salmonella inositol polyphosphatase acts in conjunction with other bacterial effectors to promote host cell actin cytoskeleton rearrangements and bacterial internalization. *Mol Microbiol* **39**, 248-59 (2001).

172. Alemán, A. et al. The amino-terminal non-catalytic region of Salmonella typhimurium SigD affects actin organization in yeast and mammalian cells. *Cell Microbiol* **7**, 1432-46 (2005).

173. Zhou, D. & Galán, J. Salmonella entry into host cells: the work in concert of type III secreted effector proteins. *Microbes Infect* **3**, 1293-8 (2001).

174. Rodríguez-Escudero, I., Ferrer, N.L., Rotger, R., Cid, V.J. & Molina, M. Interaction of the Salmonella Typhimurium effector protein SopB with host cell Cdc42 is involved in intracellular replication. *Mol Microbiol* **80**, 1220-40 (2011).

175. Rodríguez-Escudero, I., Rotger, R., Cid, V.J. & Molina, M. Inhibition of Cdc42-dependent signalling in Saccharomyces cerevisiae by phosphatase-dead

SigD/SopB from Salmonella typhimurium. *Microbiology (Reading, Engl)* **152**, 3437-52 (2006).

176. Burkinshaw, B.J., Prehna, G., Worrall, L.J. & Strynadka, N.C.J. Structure of Salmonella effector protein SopB N-terminal domain in complex with host Rho GTPase Cdc42. *J Biol Chem* **287**, 13348-55 (2012).

177. Bennett-Lovsey, R.M., Herbert, A.D. & Sternberg, M.J. Exploring the extremes of sequence/structure space with ensemble fold recognition in the program …. *Proteins* (2007).

178. Goldschmidt, L., Cooper, D.R. & Derewenda, Z.S. Toward rational protein crystallization: A Web server for the design of crystallizable protein …. *Protein Science* (2007).

179. Koszelak-Rosenblum, M. et al. Determination and application of empirically derived detergent phase boundaries to effectively crystallize membrane proteins. *Protein Science* **18**, 1828-39 (2009).

180. Gill, H.S. Evaluating the efficacy of tryptophan fluorescence and absorbance as a selection tool for identifying protein crystals. *Acta Crystallogr F Struct Biol Cryst Commun* **66**, 364-72 (2010).

181. Gorrec, F. The MORPHEUS protein crystallization screen. *Journal of Applied Crystallography* **42**, 1035-1042 (2009).

182. Long, C. et al. Yersinia pseudotuberculosis and Y. enterocolitica infections, FoodNet, 1996-2007. *Emerging Infect Dis* **16**, 566-7 (2010).

183. Alexandre Yersin, B.W. Etymologia: yersinia. *Emerging Infect Dis* **16**, 496 (2010).

184. Haensch, S. et al. Distinct clones of Yersinia pestis caused the black death. *PLoS Pathog* **6**, e1001134 (2010).

185. Doyle, R.J. & Lee, N.C. Microbes, warfare, religion, and human institutions. *Can J Microbiol* **32**, 193-200 (1986).

186. Aepfelbacher, M., Trasak, C. & Ruckdeschel, K. Effector functions of pathogenic Yersinia species. *THROMBOSIS AND HAEMOSTASIS-STUTTGART-* (2007).

187. CAVANAUGH, D.C. & RANDALL, R. The role of multiplication of Pasteurella pestis in mononuclear phagocytes in the pathogenesis of flea-borne plague. *J Immunol* **83**, 348-63 (1959).

188. Simonet, M., Richard, S. & Berche, P. Electron microscopic evidence for in vivo extracellular localization of Yersinia pseudotuberculosis harboring the pYV plasmid. *Infect Immun* **58**, 841-5 (1990).

189. Galyov, E.E., Hakansson, S., Forsberg, A. & Wolf-Watz, H. A secreted protein kinase of Yersinia pseudotuberculosis is an indispensable virulence determinant. *Nature* (1993).

190. Juris, S.J., Rudolph, A.E., Huddler, D., Orth, K. & Dixon, J.E. A distinctive role for the Yersinia protein kinase: Actin binding, kinase activation, and cytoskeleton disruption. *Proceedings of the National Academy of Sciences* **97**, 9431 (2000).

191. Dukuzumuremyi, J. et al. The Yersinia Protein Kinase A Is a Host Factor Inducible RhoA/Rac-binding Virulence Factor. *Journal of Biological Chemistry* **275**, 35281 (2000).

192. Barz, C., Abahji, T.N., Trülzsch, K. & Heesemann, J. The Yersinia Ser/Thr protein kinase YpkA/YopO directly interacts with the small GTPases RhoA and Rac-1. *FEBS Letters* **482**, 139-43 (2000).

193. Nejedlik, L., Pierfelice, T. & Geiser, J.R. Actin distribution is disrupted upon expression of Yersinia YopO/YpkA in yeast. *Yeast* **21**, 759-68 (2004).

194. Hu, X., Prehna, G. & Stebbins, C.E. Targeting plague virulence factors: a combined machine learning method and multiple conformational virtual screening for the discovery of Yersinia protein kinase A inhibitors. *J Med Chem* **50**, 3980-3 (2007).

195. Prehna, G., Ivanov, M.I., Bliska, J.B. & Stebbins, C.E. Yersinia virulence depends on mimicry of host Rho-family nucleotide dissociation inhibitors. *Cell* **126**, 869-80 (2006).

196. Navarro, L. et al. Identification of a molecular target for the Yersinia protein kinase A. *Mol Cell* **26**, 465-77 (2007).

197. Groves, E. et al. Sequestering of Rac by the Yersinia effector YopO blocks Fcgamma receptor-mediated phagocytosis. *J Biol Chem* **285**, 4087-98 (2010).

198. Laskowski-Arce, M.A. & Orth, K. The elusive activity of the Yersinia protein kinase A kinase domain is revealed. *Trends Microbiol* **15**, 437-40 (2007).
199. Dupeux, F., Rower, M., Seroul, G., Blot, D. & Marquez, J.A. A thermal stability assay can help to estimate the crystallization likelihood of biological samples. *Acta Crystallogr D Biol Crystallogr* **67**, 915-9.
200. Juris, S., Rudolph, A., Huddler, D., Orth, K. & Dixon, J. A distinctive role for the Yersinia protein kinase: Actin binding, kinase activation, and cytoskeleton disruption. *Proceedings of the National Academy of Sciences* **97**, 9431 (2000).
201. Johnson, P., Harris, C.I. & Perry, S.V. 3-methylhistidine in actin and other muscle proteins. *Biochem J* **105**, 361-70 (1967).
202. Kalhor, H.R. et al. A highly conserved 3-methylhistidine modification is absent in yeast actin. *Arch Biochem Biophys* **370**, 105-11 (1999).
203. Ohki, T., Ohno, C., Oyama, K., Mikhailenko, S. & Ishiwata, S. Purification of cytoplasmic actin by affinity chromatography using the C-terminal half of gelsolin. *Biochem Biophys Res Commun* **383**, 146-150 (2009).