Summer 6-19-2014

THE ROLE OF THE N-TERMINUS ON THE ENZYMATIC ACTIVITY OF dUTPase FROM *DICTYOSTELIUM DISCOIDEUM*

Kyle C. Varon
University of Nebraska-Lincoln, kvaron@huskers.unl.edu

THE ROLE OF THE N-TERMINUS ON THE ENZYMATIC ACTIVITY OF dUTPase FROM *DICTYOSTELIUM DISCOIDEUM*

by

Kyle Varon

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Master of Science

Major: Biological Sciences

Under the Supervision of Professor Catherine P. Chia

Lincoln, Nebraska

June 2014

THE ROLE OF THE N-TERMINUS ON THE ENZYMATIC ACTIVITY OF dUTPase FROM *DICTYOSTELIUM DISCOIDEUM*

Kyle Varon

University of Nebraska, 2014

Advisor: Catherine Chia

The enzyme deoxyuridine triphosphatase (dUTPase) (EC 3.6.1.23) converts dUTP to dUMP, thus shifting the dUTP to dUMP ratio in the cell. The molecule dUTP is subject to mis-incorporation into DNA due to lack of distinguishing by DNA polymerase. Uracil incorporation can be repaired with base excision repair mechanisms but may create overwhelming DNA strand breaks proving to be detrimental to the cell. Most dUTPases of eukaryotes are homotrimeric and contain five highly conserved motifs responsible for catalysis and substrate binding. Many dUTPases of eukaryotes possess a leading and core region in their sequence. The core region is composed of the five highly conserved motifs while the leading region is described as the N-terminal end of the enzyme with an uncharacterized role. Due to its similarity to mammalian cellular biology, the eukaryotic soil amoeba *Dictyostelium discoideum* was used as a model organism for understanding characteristics of dUTPase. The leading region of the *D. discoideum* dUTPase is 37 residues at the N-terminal end. Two separate sequences of dUTPase, a predicted native sequence (full-length version) and a sequence lacking the 37 N-terminal residues (core version) were cloned and expressed in the organism *Escherichia coli.* A comparative study format was utilized to understand the role of the N-terminal

leading region in relation to kinetic constants, pH and temperature optimum, metal dependence, and renaturation potential. Data revealed the core version to be superior to the full-length due to a lower K_m (0.44-fold lower) and greater k_{cat}/K_m (1.35-fold higher) than the full-length version.

Table of Contents

Chapter 1: Introduction

Maintenance of DNA integrity is essential for sustaining a functioning cellular environment. Lack of integrity has been found to lead to cellular complications and eventually apoptosis. Misincorporation of dUTP into replicating DNA is a potential recognition error of DNA polymerase (Kornberg, 2005). A cellular environment with high concentrations of dUTP can increase the number of misincorporations. Base and nucleotide excision repair mechanisms are capable of excising misincorporated uracil residues but require compromising the DNA for proper repair.

Large amounts of dUTP will statistically lead to an increase in misincorporation events and consequently to higher numbers of simultaneous DNA strand breaks that can prove to become detrimental to a cell. Conversion of dUTP to dUMP is a means of maintaining DNA integrity throughout replication by decreasing the amount of dUTP that can become misincorporated.

The enzyme deoxyuridine triphosphatase (dUTPase) converts deoxyuridine triphosphate (dUTP) to deoxyuridine monophosphate (dUMP) through the removal of the β and γ phosphate groups by hydrolytic cleavage (Figure 1.1). By converting dUTP to dUMP, dUTPase serves two roles for the cell. Firstly, decreasing dUTP levels reduces the amount of misincorporation into DNA. Secondly, dUMP is the reactant for the conversion of dUMP to deoxythymidine monophosphate (dTMP) by the enzyme thymidylate synthase (Mahagaokar, *et al.,* 1980). The molecule dTMP is subsequently phosphorylated by dTMP kinase and NDP kinase, creating dTTP, a molecule essential in

DNA replication to fulfill thymine demands of the cell. Cellular response to a lack of usable dTTP is apoptosis via thymine-less cell death (Cohen, 1971).

It is due to the dual role of dUTPase that it has been investigated in the context of knock-out studies. Investigation into the crucial nature of dUTPase showed a knock-out to be lethal in the case of *Saccharomyces cerevisiae, Escherichia coli,* and *Trypanosoma brucei* (Gadsden, *et al.,* 1993; Hochhauser and Weiss, 1978; Castillo-Acosta, et al., 2013). It should be noted that the dUTPase of *T. bruceii* also possesses dUDPase activity (conversion of dUDP to dUMP).

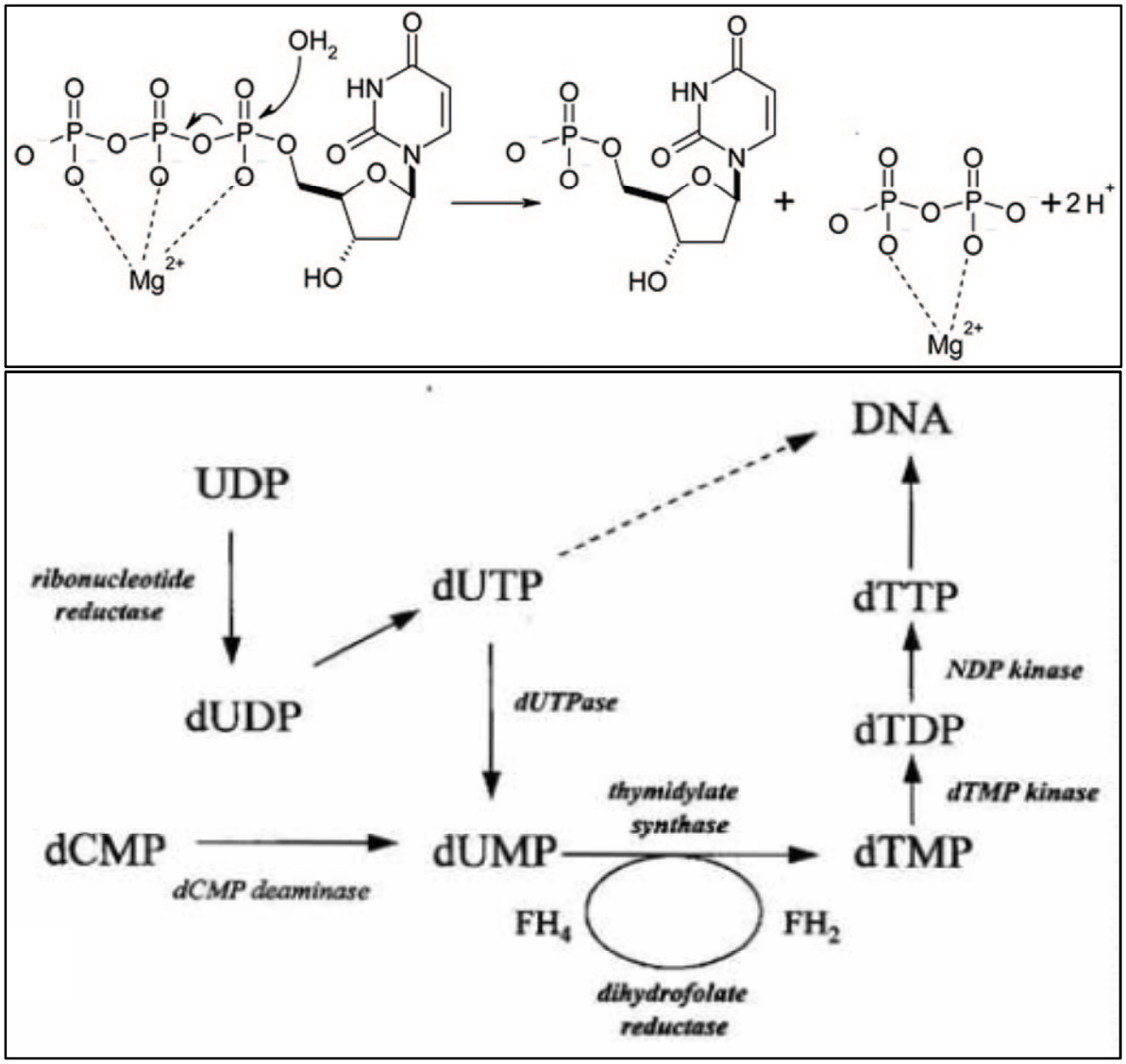

Figure 1.1: Role of dUTPase in nucleotide biosynthesis. A.) The reaction mechanism coordinated by dUTPase. A water molecule is coordinated in the active site of dUTPase and acts as the nucleophile in the attack on the α-phosphate group. The products of the reaction are pyrophosphate and $2H^+$. The present magnesium ion aids in coordinating the dUTP by bonding with an oxygen of the alpha, beta, and gamma phosphate groups. Figure from Vertessy and Toth (2009). B.) The role of dUTPase among other nucleotide metabolism enzymes. The role of dUTPase is to decrease cellular dUTP levels, accomplished through the direct conversion to dUMP. A second source of dUMP is through the deamination of dCMP by dCMP deaminase.Thymidylate synthase along with dihydrofolate reductase and tetrahydrofolate (FH_4) (reduced from dihydrofolate, FH_2) converts dUMP to dTMP which can then become phosphorylated by dTMP kinase and NDP kinase to form dTTP. Ribonucleotide reductase converts the ribose of UDP to deoxyribose, forming dUDP. Figure from McIntosh and Haynes (1997).

Studies that described the lethality of a dUTPase knockout further demonstrate the efficiency that a dUTPase inhibitor would have on eliminating cancer cell growth. Miyakoshi, *et al.,* 2012 investigated a novel class of dUTPase inhibitors for the future use in chemotherapy. The use of 1,2,3-triazole-containing uracil derivatives were found to be highly effective at inhibiting human dUTPase activity with a half maximal inhibitory concentration of approximately 0.029 µM. The lethality of the uracil derivatives holds promise as a more effective chemotherapy drug compared to 5-fluorouracil which has reported EC_{50} value of 0.27 µM in HeLa S3 cells (Miyakoshi, *et al.,* 2012).

Due to the crucial nature of dUTPase all organisms and many viruses, such as DNA, RNA, and retro-viruses contain the *dut* gene (Baldo AM and McClure MA, 1999). The tertiary structure of dUTPase varies throughout all organisms. Although some dUTPases are monomeric or dimeric, many dUTPases are homotrimeric, meaning they contain three identical subunits that collectively form an active enzyme (Castillo-Acosta, et al., 2012). As a eukaryote, *Dictyostelium discoideum* dUTPase is expected to be homotrimeric based on its sequence homology to characterized homotrimeric dUTPases.

Sequence characteristics of homotrimeric dUTPases are highly conserved with the presence of five conserved motifs (McGeoch DJ, 1990). The conserved motifs function collectively and in distinct ways to allow for proper binding of the substrate and alignment of the metal cofactor. Specifically, motifs I, II, and IV function in coordinating the phosphate chain of dUTP and the metal cofactor magnesium. In addition, the

conserved aspartate residue in motif I (see residue number 61 in sequence alignment below) is responsible for coordinating two water molecules which in turn allow for proper coordination of the metal ion required for catalysis. Motif V creates an environment favorable for catalysis by interacting with the active site. Finally, motif III creates steric hindrance that only allows the conformation of dUTP to enter into the active site. Information regarding the function of the conserved motifs was acquired from Vertessy and Toth, 2009. Alignment of the *D. discoideum* dUTPase homotrimer subunit with the dUTPase subunit sequences of *Dictyostelium purpureum, Escherichia coli, Arabidopsis thaliana, Saccharomyces cerevisiae,* and *Homo sapiens* can be seen in Figure 1.2.

```
                1                                                  50
D. discoideum   MPIEQKYFSLFSNLFKRLTTNNNNNNYLKMAPPNFETFKVKKLSDKAIIP
D. purpureum    ----------MSVLSRDEIVKAIEAGDIKITPYDEKAIGSASIDLTLGNE
E. coli         ------------------------------MKKIDVKILDPRVGKEFPLP
A. thaliana     -------MACVNEPSPKLQKLDRNGIHGDSSPS--PFFKVKKLSEKAVIP
S. cerevisiae   ----------------------------MTATSDKVLKIQLRSASATVP
H. sapiens      ------------MPCSEETPAISPSKRARPAEVGGMQLRFARLSEHATAP

                QRGSKGAAG-YDLSSAHELVVPAHG--KALAMTDLQIAIPDG-TYGRIAP
                FRVFKKNQGTIDVTEATNYQDFSEK--FVLAPGETFELEPGQMCLGITEE
                TYATSGSAG-LDLRACLNDAVELAPGDTTLVPTGLAIHIADPSLAAMMLP
                TRGSPLSAG-YDLSSAVDSKVPARG--KALIPTDLSIAVPEG-TYARIAP
                TKGSATAAG-YDIYASQDITIPAMG--QGMVSTDISFTVPVG-TYGRIAP
                TRGSARAAG-YDLYSAYDYTIPPME--KAVVKTDIQIALPSG-CYGRVAP

                RSGLAWKNFIDCG--AGVIDSDYRGNVGVVLFNHSDVDFKVAVGD-RVAQ
                TIDLSSKYCGLLEGRSRFARLGLFVHISAGFMNPGIKNRQVLEIFNASSN
                RSGLGHKHGIVLGNLVGLIDSDYQGQLMISVWNRGQDSFTIQPGE-RIAQ
                RSGLAWKHSIDVG--AGVIDADYRGPVGVILFNHSDADFEVKFGD-RIAQ
                RSGLAVKNGIQTG--AGVVDRDYTGEVKVVLFNHSQRDFAIKKGD-RVAQ
                RSGLAAKHFIDVG--AGVIDEDYRGNVGVVLFNFGKEKFEVKKGD-RIAQ

                LIFERIVTP-EPLEVDEIDETQRGAGGFGSTGVKVQN
                KLKLHPGTKFCQFVFMELKGNAVYHGRFANNTL----
```

```
MIFVPVVQA-EFNLVEDFDATDRGEGGFGHSGRQ---
LIIEKIVTP-DVVEVDDLDETVRGDGGFGSTGV----
LILEKIVDDAQIVVVDSLEESARGAGGFGSTGN----
LICERIFYP-EIEEVQALDDTERGSGGFGSTGKN---
```

Figure 1.2: dUTPase protein sequence alignment using ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/).
Accession codes are as follows: *D. discoideum:* Q54BW5.1; *D. purpureum:* EGC29029.1; *E. coli:* P06968.1; *A. thaliana:* Q9STG6.1; *S. cerevisiae:* P33317.2; *H.sapiens:* AAB71394. Sequence alignment shows the variation in the N- and C-terminal ends of dUTPase. Numbering refers to the *D. discoideum* dUTPase.

The five highly conserved motifs can be seen in alignment with dUTPase sequences of other species. The presented species were selected for the sequence alignment because they are representatives of the Kingdom Amoebozoa, Plantae, Eubacteria, Fungi, and Animalia. Although a diverse group of organisms, the sequence alignment data reveals the highly conserved nature of the typical dUTPase sequence. In addition, the N-termini or leading regions of the sequences are neither highly conserved in residue composition nor length.

Specifically, the N-terminus of the *D. discoideum* dUTPase is of interest for investigation due to its length and the composition which contains a portion of six asparagine residues in sequence. The polar patch created by the asparagine residues may have implications in protein-protein interactions (Ma, *et al*., 2003). Data have been presented showing evidence for mitochondrial and nuclear isoforms of dUTPase in humans (Ladner and Caradonna, 1997). In humans, organellar translocation of dUTPase may be mediated via the varying N-terminus. Unlike human dUTPase, *Dictyostelium discoideum* sequence information does not provide evidence for multiple isoforms of the enzyme at this time. Presently, there are no published data describing the role of

the N-terminus in enzymatic function or stability for any dUTPase. Sequence features for homotrimeric dUTPases have been annotated and the N-terminus has been shown to not contain any residues responsible for catalysis or subunit interfaces (Persson, et al., 2001). Presently, there are no published data on the *Dictyostelium discoideum* dUTPase. The importance of characterizing the *D. discoideum* dUTPase is to acquire data to understand kinetic efficiency relative to other species' versions of the enzyme. Specifically, objectives included obtaining enzymatic kinetic constants (K_m, V_{max}, k_{cat}, k_{cat}/K_m), the temperature and pH optima of the enzyme, as well as the metal dependence and thermal stability of the predicted homotrimers composed of either the predicted full-length dUTPase polypeptides or a core sequence lacking the 37 N-terminal residues. This information would allow for comparisons to other dUTPases and elucidate the role of the dUTPase N-terminus which does not contain residues that comprise any of the five conserved motifs.

Many species' dUTPase enzymes have characterized K_m values. Table 1.3 displays selected K_m data already published.

Organism	**K_m (µM)**
Allium cepa (onion) (Pardo & Gutierrez, 1990)	6.0
Solanum lycopersicum (tomato) (Pri-Hadash *et al.,* 1992)	0.50
Escherichia coli (Larsson *et al.,* 1996)	0.21
HeLa cells (Ladner *et al.,* 1996)	2.5
Trypanosoma cruzi (Bernier-Villamor *et al.,* 2002)	0.53
Chlorella virus PBCV-1 (Zhang *et al.*, 2005)	11.7
Arabidopsis thaliana (Chaiseeda 2009)	0.1±0.01
Saccharomyces cerevisiae (Tchigvintsev *et al.,* 2011)	13.2±0.6

Table 1.1: Reported K_m of various dUTPases. Table data is organized in chronological order. Table modified from Chaiseeda (2009). Recombinant technology was used to generate the *S. cerevisiae*, *A. thaliana*, PBCV-1, and *T. cruzi* dUTPase proteins.

As seen in Table 1.1, reported K_m values range from 0.1 µM (*Arabidopsis thaliana*) to as high as 13.2 µM (*Saccharomyces cerevisiae*). The *A. thaliana* and *S. cerevisiae* dUTPases both of which are eukaryotic, have reported K_m values that differ by 132-fold. However, BLAST results of the full-length *D. discoideum* dUTPase sequence show a high identity to *A. thaliana* dUTPase (66%) with the highest similarity being *D. purpureum* dUTPase (77%). The high identity between the full-length *D. discoideum* sequence and the *A. thaliana* dUTPase allows for the prediction that kinetic properties would be similar.

The dUTPase investigated in this study is from the unicellular soil amoeba *Dictyostelium discoideum.* Specifically, *D. discoideum* is considered a slime mold within the Kingdom Amoebozoa. Under environmental stress, such as nutrient deprivation, unicellular *Dictyostelium* cells aggregate into a multicellular structure and persist to form a fruiting body from which the next generation is created.

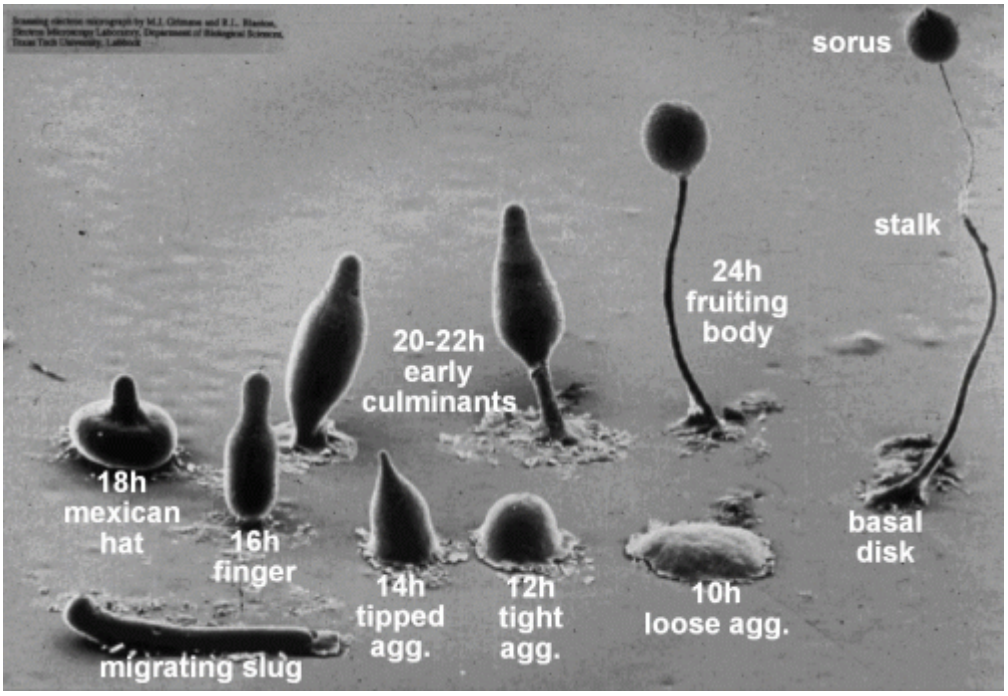

Figure 1.4: Lifecycle of the unicellular soil amoeba *Dictyostelium discoideum*. Environmental stimuli such as nutrient starvation triggers unicellular *D. discoideum* to begin aggregation by chemotaxis. The loose aggregate becomes more tight prior to forming into the tipped aggregate at 14 hours and eventually the finger at 16 hours. The finger morphology then becomes the migrating slug which is typically between 2 and 4 mm long (Cooper, 2000). The slug is able to move by the production of a cellulose sheath within the anterior portion (Tyler, 2000). The cells which compose the slug then differentiate into pre-stalk and pre-spore cells with differentiation-inducing factor (Tyler, 2000). Eventually, the anterior portion will become the stalk cells and the posterior portion the spore cells. The Mexican hat is formed when the posterior portion of the slug spreads out and the anterior portion protrudes upwards. The stalk and mature spores which form from the elongated Mexican hat structure are now ready to release spores, thus starting the lifecycle over again. SEM of Dictyostelium developmental stages. Copyright M.J. Grimson & R.L. Blanton; Biological Sciences Electron Microscopy Laboratory, Texas Tech University.

It is due to the unique nature of the *D. discoideum* life cycle that it has become an important eukaryotic model organism. Specifically, the phagocytic and chemotactic behavior exhibited is similar to that of mammalian macrophage behavior, making *Dictyostelium* an important model organism for motility and cellular biology

investigation. The *D. discoideum* dUTPase was selected to be characterized for several reasons. Firstly, the dUTPase enzyme in *D. discoideum* had not been purified or characterized. Secondly, characterization studies would lay the ground work for future projects investigating the expression patterns and *in vivo* activity of dUTPase within *D. discoideum.*

Within this study, three versions of the *D. discoideum* dUTPase were characterized with the monomeric unit being: a 19.8 kD hypothetical full length dUTPase (full-length version), a 15.5 kD version lacking the 37 N-terminal residues (core version), and an 9.4 kD version lacking the 118 N-terminal residues (truncated version). The dUTPase sequence is termed hypothetical since it is the product of gene annotation of genomic data compared to known dUTPase sequences. The objective of this comparative study was to obtain an enzymological characterization and understand the role of the N-terminus of dUTPase. Sequence information about dUTPases (Figure 1.2) has revealed a 'core' region of the protein sequence, containing residues responsible for catalysis and trimeric assembly (specific residues unknown) and a leading region at the N-terminus (Figure 1.2). The role of residues that comprise the leading region of *D. discoideum* dUTPase is unknown presently. However, the leading region of dUTPase has been cited in roles such as nuclear and mitochondrial localization within humans (Ladner *et al.,* 1996). Presently, there are no data to support or refute the importance of leading sequence residues in trimer assembly.

Since the residues responsible for enzyme activity were identified based on homology, the 37 N-terminal residues were excluded from the sequence to mimic the abbreviation applied for the investigation of *Arabidopsis* and *Chlorella* virus dUTPases (Bajaj and Moriyama, 2007; Homma and Moriyama, 2009). Furthermore, the prospect of crystallization is hypothesized to be more favorable without the N-terminal residues due to the unpredictable nature of the tail-end. The truncated *Arabidopsis* and *Chlorella* dUTPases were successfully crystallized, while their full-length sequences were unable to be used for kinetic or crystallographic analyses due to issues with stability (Hideaki Moriyama, personal communication).

Previously, a truncated dUTPase subunit lacking motifs I and II was generated by proteolytic cleavage of the core dUTPase (Nguyen 2010). The truncated dUTPase possessed activity and was characterized kinetically and the effect of temperature and pH on activity was recorded. The truncated dUTPase in this study is a recombinant version of the truncated protein previously investigated but possessed no measurable level of activity despite refolding attempts. Due to the absence of activity the hypotheses below were made solely between the full-length and core versions of dUTPase.

Before beginning experimentation, hypotheses were formulated for the prospective activity of the full-length and core version dUTPases with the different parameters to be investigated. The K_m of the core version was predicted to be lower than that of the full-length. The logic behind this prediction is that without the N-

terminal residues the enzyme would be more readily exposed to substrate molecules without steric hindrance. With less hindrance the substrate molecules would be exposed to the active site at a higher frequency and create a reduction in the $K_{m,}$ or the substrate concentration at which $1/2V_{max}$ is achieved. It is due to this difference in accessibility that the core version is expected to have greater specific activity than the full-length.

BLAST datum reveals the *A. thaliana* dUTPase to contain 66% identity to the *D. discoideum* dUTPase. The sequence identity, which includes the five conserved motifs responsible for the three active sites, is a measure of the similarity in characteristics between the two versions. Thus, it can be inferred that the *D. discoideum* full-length dUTPase would have a temperature optima similar to the *A. thaliana* dUTPase. As explained earlier, the role of the N-terminus is presently unknown. If the N-terminal residues absent in the core version are important in protecting the active site thermally, we would expect a decreased temperature optima compared to the full-length version.

The pH optima were investigated to determine optimal pH for activity for both versions of the enzyme. Understanding that extreme pH can denature a protein by extreme protonation or deprotonation of residues and thus cause secondary structure breakdown, it was expected that the pH optimum of the full-length would be close to neutral (pH 6.0-8.0). If the N-terminal residues are responsible for shrouding the active site and regulating a pH suitable for catalysis, we would expect the full-length version to possess more tolerance to varying pH and a greater sensitivity to pH in the core version.

Reliance of dUTPase activity on the metal cofactor Mg was tested directly with metal dependence trials. Several published studies on dUTPase have described the activity to be reduced greatly in the absence of Mg (Mustafi, *et al.,* 2003; Tchigvintsev, *et al.,* 2011). The role of Mg in the active site is highly specific, functioning in the coordination of the dUTP phosphate chain. If the active site lacks Mg, it would be expected that the phosphate chain could not properly be oriented and the rate of catalysis would decrease. It is due to the published data that activity for both versions of dUTPase is expected to decrease without the proper Mg metal cofactor. Specifically, activity of the full-length may be more hindered than the core as the N-terminal residues may prevent interactions with different metals.

Renaturation potential was measured to determine if the enzyme was capable of renaturing after exposure to high heat. Trials were predicted to reveal an inability for renaturation as the temperature treatment was higher than the predicted temperature optima. As denaturation is often times an irreversible process, neither of the versions was expected to regain any degree of activity lost through heat treatment.

Overall, the core version dUTPase was expected to have greater catalytic efficiency due to the potential of a more accessible active site. Temperature optimum of the core version was expected to be lower than the full-length if the N-terminal residues are important in protecting the active site thermally. The pH optimum of the core version was expected to be lower than the full-length if the N-terminal residues proved to be essential in maintaining a favorable active site pH. The level of activity with

cofactor substitutes was predicted to decrease in the absence of Mg as the cofactor. Finally, neither the full-length nor the core versions were expected to regain activity after a thermal denaturation.

Chapter 2: Materials and Methods

2.) Molecular Cloning

2.1.) Recombinant dUTPase

The coding sequence corresponding to the full-length dUTPase of *Dictyostelium discoideum* (DictyBase ID: DDB_G0293374; http://dictybase.org/gene/DDB_G0293374) was synthesized by GenScript (GenScript USA Inc., Piscataway, NJ) and inserted into vector pUC57 using restriction sites EcoRI and BamHI. K. Williams and C. Chia generated the construct to produce a His_6-tagged full-length protein. The *dut* gene was PCR-amplified using the sense primer and antisense primers (IDT, Coralville, IA) shown in Table 2.1 to generate NdeI and BamHI restriction sites that allowed its directional insertion into pET-15b.

C. Natarajan and H. Moriyama generated the construct to produce a His_6-tagged truncated dUTPase lacking residues 1-37 of the predicted N-terminus. To produce this core dUTPase protein, a PCR product with NdeI and BamHI restriction sites was generated using the primers indicated in Table 2.1, cloned into pGEMT (Promega, Madison, WI) and then released with NdeI and BamHI. The insert was subsequently ligated into pET-15b.

A truncated sequence of dUTPase (residues 98-179) was PCR-amplified from the full-length sequence by K. Varon using the primers indicated (Table 2.1). Overhang regions on the PCR product corresponding to the 9.4 kD segment and expression vector pET-15b (Novagen Inc., Madison, WI) were made by digesting with BamHI and NdeI.

Ligation was carried out overnight at 22°C with 1 µL T4 DNA ligase (Fermentas Inc., Pittsburgh, PA) with molar ratios of vector to insert of 1:1, 1:3, and 1:5. DNA was quantified using the Qubit dsDNA assay kit (Life Technologies, Carlsbad, CA). BL21 (DE3) competent cells (Life Technologies, Carlsbad, CA) were transformed using 5 µL of ligation mixtures per 50 µL of cells. Cells were heat shocked at 42°C for 30 seconds followed by addition of 250 µL pre-warmed SOC media (0.5% Yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose) and shaking at 225 rpm at 37°C. Bacteria were grown overnight on LB (1 L: 0.10% tryptone, 0.05% yeast extract, 0.10% NaCl, pH 7.5) agar (1.5%) with 0.1 mg/mL ampicillin (Thermo Fisher Scientific, Waltham, MA) at 32°C. Transformants were grown for 17 hours in LB liquid media (0.1 mg/mL ampicillin) at 32°C and plasmid DNA isolated using Miniprep and Maxiprep kits (Qiagen Inc., Valencia, CA) was sequenced (Sequetech. Corp., Mountain View, CA).

full-length	5’ GGCCCGGGATCCTTAATTTTGAACTTTAACACC 3’ 5’ CCAGTGAACATATGCCAATCGAACAAAAATATTTTTC 3’
core	5’ CATATGTTCAAAGTTAAAAAATTAT 3’ 5’ GGATCCTTA ATTTTGAACTTTAACA 3’
truncated	5’ ACCTATCATATGGGTGGTGGTGGTTCAGGTTTAGCATGGAAAAATTTCATCGATTGTGG 3’ 5’ GGCCCGGGATCCTTA ATTTTGAACTTTAACACC 3’

[1]The upper and lower sequences are the sense and antisense primers, respectively.

Table 2.1: PCR primers used to produce genes of different versions of D. discoideum dUTPases[1]

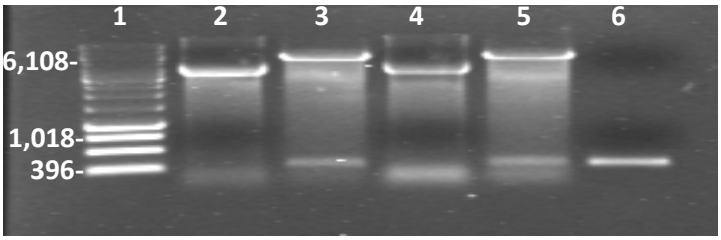

Figure 2.1: Restriction enzyme digest of expression plasmid for truncated (9.4 kD) dUTPase. Lane 1: DNA ladder (Fermentas Inc., Pittsburgh, PA); Lane 2: Undigested clone A; Lane 3: Digested (NdeI and BamHI) clone A; Lane 4: Undigested clone B; Lane 5: Digested (NdeI and BamHI) clone B; Lane 6: PCR product corresponding to the truncated polypeptide.

Plasmid DNA was digested with BamHI and NdeI and run on a 1.5% agarose gel for 30 minutes at 100 V to determine vector and insert sizes. Anticipated size of vector and insert was 5696 bp and 272 bp respectively. Lanes 3 and 4 contain insert and vector from digested plasmid. The insert in lane 6 lines up directly with the insert digested from the plasmid in lanes 3 and 5. Undigested plasmid in lanes 2 and 4 contain several bands that represent different conformations of the plasmid. Several species did not line up directly with expected DNA ladder size due to the short running time of the gels. Short running times were used to maintain the resolution of the short insert.

Plasmid DNA with the correct construct was then used to transform competent JM109 cells (Promega, Madison, WI). Expression of the truncated segment was induced with 1.0 mM isopropyl β-D-1-thiogalactopyranoside when cultures grown at 32°C in LB reached an OD600 of 0.5. Cells were harvested after 24 hours at 32°C and reached a final OD600 of ~1.4.

2.2.) Expression of Recombinant Proteins

For each version of dUTPase, a mini-culture (3 mL of LB, 0.1 mg/mL ampicillin) of a transformant containing the verified His_6-tagged construct was grown for approximately 12 hours at 32°C. The 3 mL culture was used to inoculate a 250 mL culture that was grown at 32°C until it reached an OD600 of 0.6 at which time protein expression was induced with a final concentration of 1.0 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cultures were grown for an additional 4 hours (full-length), 17 hours (core version), or 24 hours (truncated) at 32°C and then harvested by a Sorvall RC-5 Superspeed Refrigerated Centrifuge at 10,000 x G for 60 minutes with a GSA rotor cooled to 4°C (DuPont, Wilmington, DE).

2.3.) Protein Purification

2.3.1.) Purification of full-length and core dUTPase

Pelleted bacteria (from a 250 mL culture) were suspended in 12.5 mL NP1-10 (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) and sonicated at 4°C with a Sonicator 3000 (Misonix Inc., Farmingdale, NY) for 10 minutes (20 s rest/10 s burst) at a power setting of 3.5.

The bacterial lysate was clarified at 12,000 x G using a SS-34 rotor with a Sorvall RC-5 Superspeed Refrigerated Centrifuge for 60 minutes at 4°C and the clarified lysate then was loaded onto a HiTrap Chelating HP column(1 mL bed volume); GE Healthcare Life Sciences, Pittsburgh, PA) charged with $NiCl_2$ (HiTrap Chelating HP 1 mL- User Manual) and equilibrated with NP1-10. Purification was carried out with a Vision- Family Perfusion Chromatography Workstation (GMI Inc., Ramsey, MN) with a linear gradient of 0% to 100% NP1-250 (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0) at 1

mL/minute. The loading rate was adjusted to 0.5 mL/minute to allow for binding of the His_6-tagged protein. Clarified lysate was passed through the column five consecutive times to allow for total binding. The column was washed with NP1-20 (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0) for 20 minutes following binding. Unbound sample was analyzed by SDS-PAGE to determine the amount of target protein that remained unbound.

Protein was monitored by OD280 and typically a peak was observed 15 to 20 mL into the elution step (corresponding to 90 to 100%, NP1-250). Peak OD280 fractions were analyzed by SDS-PAGE and purity was assessed using ImageJ (Image Processing and Analysis in Java, http://rsbweb.nih.gov/ij/). Peak protein fractions were dialyzed into phosphate buffered saline (PBS: 150 mM NaCl, 1.9 mM $NaH_2PO_4*H_2O$, 8.1 mM Na_2HPO_4, pH 8.0).

The His_6-tagged proteins were treated with bovine alpha-thrombin (Haematologic Technologies Inc., Essex Junction, VT) (10 U /mg protein) at 22°C for 2h to remove the His_6-tag. Benzamidine (Sigma-Aldrich, St. Louis, MO) was added to a final concentration of 1 mM to inhibit thrombin activity. Loss of the His_6-tag was verified by SDS-PAGE. The His_6-core generally yielded 5 to 6 mg/250 mL cultures (eight preparations). The His_6-full-length generally yielded 4 to 5 mg/250 mL culture (three preparations). Protein preparations were dialyzed against PBS and stored at 4°C.

2.3.2.) Purification of truncated dUTPase

The clarified bacterial lysate was prepared with NP1-20 and 6 M Guanidine HCl (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 6 M Guanidine HCl, pH 8.0) for purification under denaturing conditions. After the lysate was prepared it was then mixed with a three mL bed volume of Ni-NTA agarose (Qiagen, Germantown, MD) at 4°C for 24 h. Before use, the matrix was washed with ten bed volumes of water, followed by ten bed volumes of NP1-10. The unbound protein was collected and the resin was washed with ten bed volumes of NP1-20 (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0) and the His_6-tagged protein was eluted using ten bed volumes of NP1-250 at 25°C. The OD280 of the elution was monitored to identify protein-containing fractions. Purity was determined by SDS-PAGE. Typically, the His_6-truncated construct yielded ~10 mg/250 mL culture (three preparations).

2.2.3.) Refolding trials of truncated dUTPase

Attempts were made to refold truncated dUTPase to obtain a measurable level of activity. A total of 50.0 µg truncated protein was diluted into 1 mL of refolding buffer. Refolding buffer was 10 mM bicine, 2 mM $MgCl_2$, pH 8.0 with 0.8 M arginine, 25% (v/v) ethanol, or 10 mM Triton X-100. Protein was incubated and allowed to refold for a total of 48 hours with samples taken for assay at 4 hours, 24 hours, and 48 hours (Clark, *et al.,* 1999).

2.4.) Protein Quantification

Protein was quantified using a Cary 50 Bio UV Vis Spectrophotometer (Varian Inc., Palo Alto, CA) simple reads function measuring at 280 nm. The machine was blanked with buffer just prior to acquiring absorbance measurements. The calculated

absorbances of a 0.1% solution (Absorbance 0.1%) (1 mg/mL) were 0.654 (full-length), 0.648 (core), and 0.752 (truncated) and were calculated for the cleaved versions (Protparam Tool, http://web.expasy.org/protparam/). The Absorbance 0.1%, calculated using the extinction coefficient, refers to the theoretical absorbance at 280 nm of a 1.0 mg /mL protein solution (in H_2O). A relationship of 1.0 mg/mL per 1.0 OD280 was used for this study and was reflective of data obtained by BCA assay (Pierce Biotechnology, Rockford, IL) data. According to the extinction coefficients calculated, a protein overestimation of 0.337, 0.345, and 0.192 mg/mL for the full-length, core, and truncated proteins is assumed by using the OD280 method of quantification.

2.5.) SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyze protein samples. Stacking (125 mM Tris, 0.1% SDS, pH 6.8) and resolving (150 mM Tris, 0.1% SDS, pH 8.8) gels were used when pouring SDS polyacrylamide gels.

Polyacrylamide gels (15% acrylamide (resolving layer); 37.5:1 acrylamide: bisacrylamide) were prepared using the Hoefer Dual GelCaster (Pharmacia Biotech, San Francisco, CA) and run for 2 hours at 100 V using Tris-glycine running buffer (25 mM Tris, 192.5 mM glycine, 0.1% SDS, pH 8.3). Samples were prepared by adding 10 µL SDS sample buffer (0.4 g/mL sucrose, 0.05 g/mL SDS, 312.5 mM Tris, 5 mM EDTA, pH 6.8) per 20 µL sample and treated for 10 minutes at 80°C. Gels were stained using Coomasie blue (0.25 g/L Coomassie R-250, 50% methanol, 10% acetic acid). PageRuler Prestained Protein Ladder (Thermo scientific, Walther, MA) was used to determine the relative

mobility of recombinant proteins. A plot of the log molecular weights was generated to approximate the mass of expressed proteins (Figure 2.2).

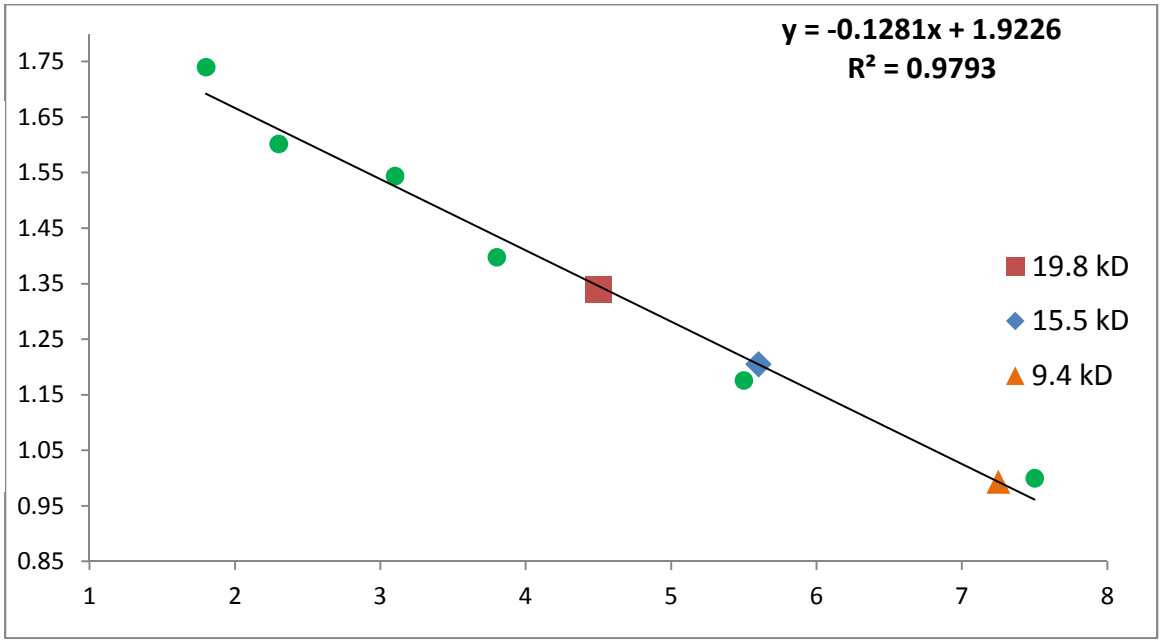

Figure 2.2: Size calculation by plotting migration distance of protein ladder. Vertical axis represents the log value of protein standards. Horizontal axis represents migration distance observed by each protein ladder constituent. The log was calculated for the molecular weights of the protein ladder components and plotted versus the individual migration distances. The log of the molecular mass was calculated using the migration distances of thrombin-cleaved recombinant dUTPases and the best fit line equation. With a log value of 1.205, the predicted 15.5 kD (core subunit) measures to 16.04 kD. The 19.8 kD (full-length subunit) with a calculated log value of 1.34, results in a predicted size of 21.8 kD. The 9.4 kD version (truncated subunit) measurements produced a log value of 0.994, predicting the size to be 9.9 kD. Gel measurements were consistent among multiple gel samples.

Sizes calculated through migration distance of SDS-PAGE provide an estimation of protein molecular weights. The core version sequence is expected to have a predicted size of 15.5 kD while measurements yield an estimation of 16.04 kD, a difference of 0.54 kD. Similarly, the full-length and truncated versions had estimated sizes that differed from calculated sizes by 2.0 kD and 0.5 kD respectively. The abbreviated number of

protein ladder measurements used in Figure 2.2 provides a more credible estimation of dUTPase mass than the entire data series (see Appendix, Figure 5.4). When using the entire data set, mass estimations differed from calculated sizes by 3.4 kD, 0.2 kD, and 1.3 kD for the full-length, core, and truncated versions respectively.

2.6.) Activity assays and data analysis

2.6.1.) Full-length and core kinetic data acquisition and analysis

Activity analyses of the core and full-length were carried out using a Cary 50 Bio UV Vis Spectrophotometer to measure changes in absorbance at 573 nm of the pH indicator cresol red.

To obtain dUTPase activity rates, recombinant protein (core: 10 µL at 0.58 µg/µL, 1.25×10^{-4} µmoles; full-length: 10 µL at 0.10 µg/µL, 1.68×10^{-5} µmoles) was placed in a cuvette and immediately upon addition of 990 µL of assay buffer (25 µM cresol red, 100 mM KCl, 250 µM bicine, pH 8.0) containing dUTP, absorbance changes were recorded. Buffer was prepared with varying concentrations of dUTP (0.1, 0.2, 0.4, 0.6, 0.8, 1.0, 2.0, 4.0, and 10.0 µM) (Roche Inc., Indianapolis, IN) and degassed for 30 minutes before use. Absorbance data were acquired every 0.0125 second and collected for a minimum of 1 minute. Slopes determined for the linear portion of the absorbance trace at each dUTP concentration were plotted and fitted to Michaelis-Menten parameters to determine the K_m using the analysis software program EnzFitter (Premier Biosoft Inc., Palo Alto, CA).

Kinetic constants were calculated using mathematical relationships presented by Larsson, *et al.,* 1996. Since a linear relationship results from the kinetic photometric trace, the absorbance values at different times can be used to calculate relative product produced at time *t*. Firstly, the concentration of product at time *t* was calculated using Equation 1. A_0, A_t, and A_∞ represent the initial absorbance, absorbance at time *t*, and final absorbance. $[S]_0$ represents the initial substrate concentration.

$$[P]_t = \frac{A_0 - A_t}{A_0 - A_\infty}[S]_0 \qquad \textbf{Equation 1}$$

The variables *V, K, t,* $[S]_0$*, and* $[P]_t$, represent respectively, the V_{max}, K_m, time, initial substrate concentration, and the concentration of product at time *t.*

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \xrightarrow{k_{cat}} E + P \qquad \textbf{Equation 2}$$

The Michaelis-Menten reaction (Equation 2) is characterized by the reversible formation of the enzyme substrate complex (***ES***) followed by the irreversible production of product and enzyme. The forward step responsible for formation of the enzyme substrate complex is referred to as k_1 while the reverse reaction responsible for the dissociation of the complex is termed k_{-1}.

$$\frac{[P]_t}{t} = V - \frac{k}{t} ln \frac{[S]_0}{[S]_0 - [P]_t} \qquad \textbf{Equation 3}$$

$$V = \frac{kln\left(\frac{[S]_0}{[S]_0 - [P]_t}\right) + [P]_t}{t} \qquad \textbf{Equation 4}$$

Since dUTPase is believed to follow the Michaelis-Menten equation, data can be interpreted using the rate equation above (Equation 4), derived from Equation 3 (Larsson, *et al.,* 1996). The calculated $[P]_t$ from Equation 1 was then used in Equation 4 to calculate the V_{max} based on properties of the assay. Within the equation *k, $[S]_0$, $[P]_t$, and t* represent the K_m, initial substrate concentration, product concentration at time *t,* and time for the reaction. In Equation 7, C_e is defined as enzyme concentration.

$$r_1 \approx k_1 C_e (1 + \frac{K_m}{C_e}) \qquad \textbf{Equation 5}$$

$$r_2 \approx \frac{K_{cat}}{1+\frac{K_m}{C_e}} \qquad \textbf{Equation 6}$$

$$K_{cat} \approx \frac{(K_m * r_2)+(C_e * r_2)}{C_e} \qquad \textbf{Equation 7}$$

For purposes of the k_{cat} expression, Equation 5 reduces to Equation 6 that is rearranged to determine k_{cat} (Equation 7). This reduction is under conditions in which enzyme concentration is greater than the enzyme K_m ($C_e >> K_m$) due to stoichiometric favorability for the reaction to drive forward from ***ES*** to ***product*** (r_2) rather than dissociate to ***E+S*** (r_1). This assumption allows for approximation of the reaction k_{cat} after obtaining the K_{obs}, or the slope of a first-order best fit line from a progress curve (Figure 2.2). Data used to approximate k_{cat} values were generated with 2.15×10^{-4} µmoles core (0.97 times K_m) and 1.68×10^{-5} µmoles full-length (0.034 times K_m). Since the k_{cat} kinetics in this study were not performed using a significantly higher concentration of enzyme to the observed K_m, the calculated k_{cat} values were not in accordance with the assumption required for measuring only the k_{cat} step of Equation 2 and thus were influenced by the slower rate of k_1.

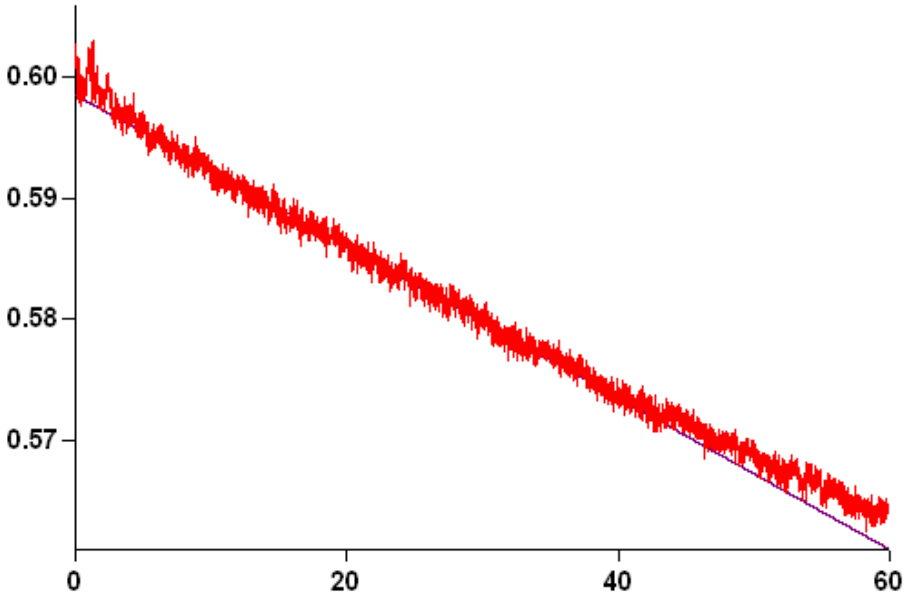

Figure 2.3: Example trace of the cresol red activity assay of core dUTPase with 10 µM dUTP. Reaction was carried out as described previously in Materials and Methods (Section 2.6.1).

The reaction trace shows a decrease in absorbance of the pH indicator cresol red. The reaction progresses and acidifies from the two protons released for each turnover. The best fit line represents the portion of the activity trace that remains linear.

2.7.) Measurement of dUMP product formed in end-point assays

Endpoint assays used a Vision- BioCAD Family Perfusion Chromatography Workstation (Version 3.01). An anion exchange protocol (developed by K. Chaiseeda and H. Moriyama; Figure 2.4) used three sequential HiTrap Q HP columns (1 mL bed volume each; GE life sciences, Piscataway, NJ)

A 20%-100% solution B gradient was used to separate dUTP from dUMP by relative hydrophilicity. The amount of dUMP produced at the end of a 10 minute reaction was obtained by measuring the area under the dUMP peak using the graphic display of chromatograms method editor. Two milliliters of the three milliliter reaction were analyzed per sample.

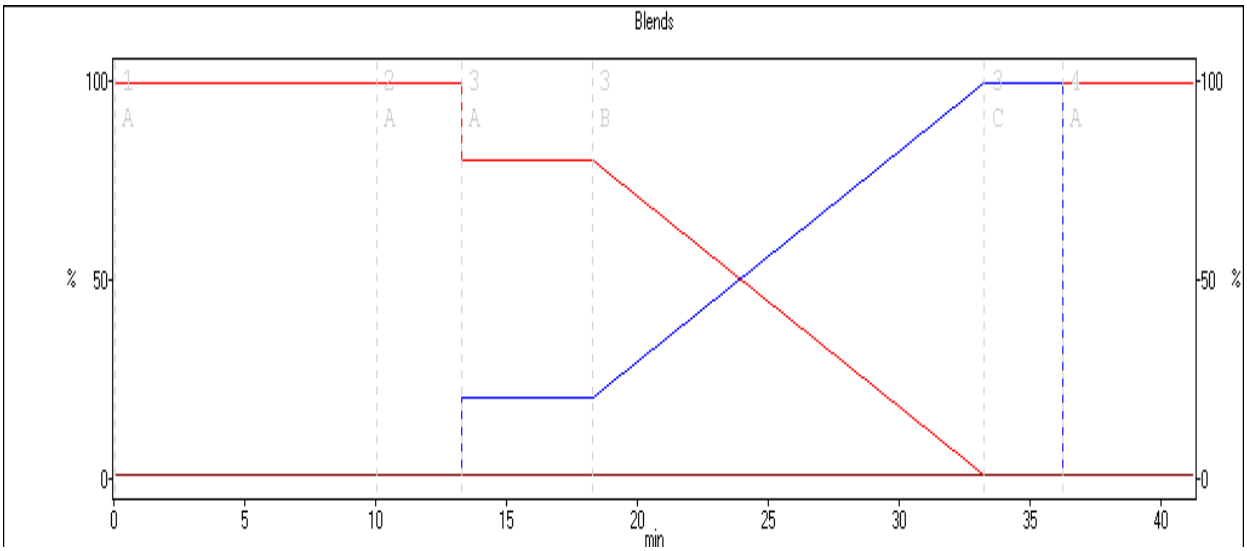

Figure 2.4: Diagram of automated buffer changes for anion exchange chromatography of dUMP and dUTP. Steps are denoted in grey lettering (e.g. 1A, 2A, 3A, etc.) Step 1A: 10 minutes of column equilibration with Solution A (5 mM KH_2PO_4, 10 mM KCl, pH 8.0). Step 2A: Sample injection (2 mL of 3 mL total reaction volume). Step 3A: Change in solution from 100% solution A to 80% solution A and 20% solution B (5 mM KH_2PO_4, 500 mM KCl, pH 8.0). Step 3B: Beginning of a linear gradient from 20% solution B to 100% solution B over a period of 15 minutes. Step 3C: 100% solution B for 5 minutes. Step 4A: Re-equilibration of the ion exchange column and plumbing to 100% solution A. X-axis shows time in minutes. Y-axis shows the percent of solution being used (red line, Solution A, blue line, Solution B). Source of the graphic featured is the HPLC method editor.

Trial runs were conducted using dUMP, or dUTP, or dUMP and dUTP to ensure the assay was functioning properly (Figure 2.4). Peaks were identified based on the solution injected. Peak locations and elution times of the two nucleotides was

consistent throughout all trials. The dUMP peak was consistently narrow with an A260 of approximately 0.075 (red peak at 5 minutes).

The dUTP peak (green peak between 14 and 15 minutes) has a smaller peak absorbance value consistently around 0.05 but is wider. While the height and width of peaks differed, the areas of the dUTP and dUMP peaks are comparable at the same concentration.

The different shapes of the peaks can be attributed to the point at which they are eluted. As dUMP requires lower salt concentrations to become eluted than the dUTP, the elution is more succinct, resulting in the narrow, sharp peak.

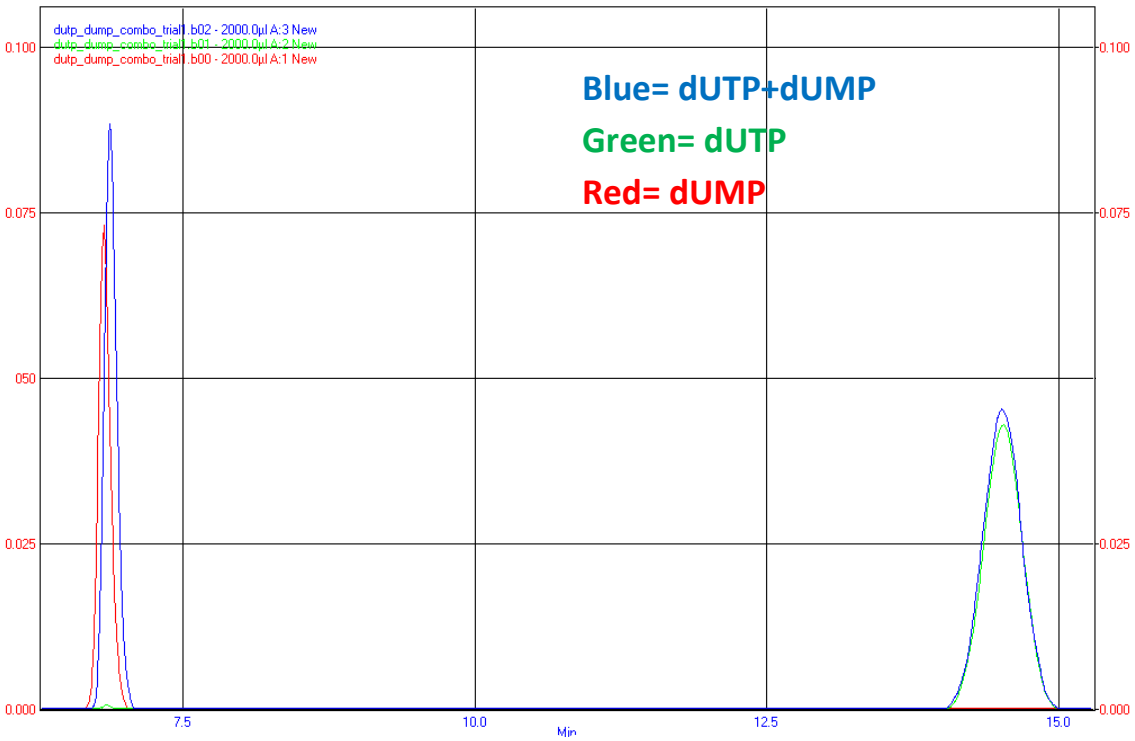

Figure 2.5: Elution profiles of dUMP and dUTP. Traces represent 2 mL injections of 10 µM (20 nanomoles) dUMP (red), dUTP (green) or both dUMP and dUTP (blue). Nucleotides were in 10 mM bicine, pH 8.0.

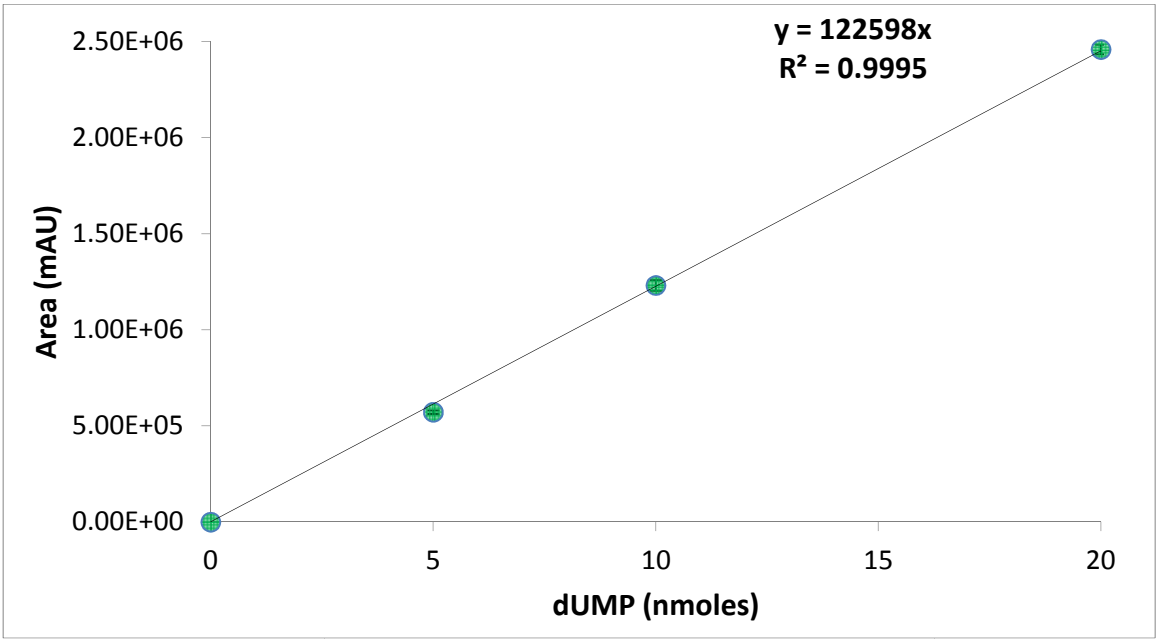

Figure 2.6: Standard curve of ion exchange chromatography of dUMP. The curve was created by injecting 5, 10, and 20 nmoles of dUMP and measuring the corresponding peak areas on the chromatogram. Y-axis unit, mAU, is defined as milliabsorbance units at 260 nm.

A standard curve was generated by chromatographing known amounts of dUMP (Sigma Aldrich, St. Louis, MO) and the corresponding area under the dUMP peak was measured (Figure 2.5). Chromatographed dUMP was prepared in 10 mM bicine buffer pH 8.0.

A	B	C	D	E	F
dUMP peak area	Reaction total calculated area	Convert to nmoles dUMP (B)/122598	µmoles dUTPase used	nmoles dUMP/minute/ µmoles dUTPase	µmoles dUMP/ µmoles dUTPase
	(A*3/2)	(B/122598)	(0.50 µg/46,563)	((C/10)/D)	(E/1000)
100,000	150,000	1.22	10.74×10^{-6}	3262.03	3.23

Table 2.2: End-point assay data analysis example.

Presented Table 2.2 displays the steps to convert raw dUMP peak areas provided by HPLC analysis to a value with units. Step A represents the peak area provided by chromatogram analyses. Step B normalizes the value (provided in A) to the total volume per reaction (3 mL) rather than the analyzed volume (2 mL). Step C utilizes the dUMP standard curve (Figure 2.6) best fit slope value to convert the raw data into nmoles dUMP. Step D is an independent calculation to determine the µmoles of dUTPase used per assay. In Step E the nmoles of dUMP were divided by µmoles of recombinant dUTPase (µg/molar mass; 46,563 kD for core version; 59,439 kD for full-length) used per assay to determine the nmoles of dUMP produced per mole of dUTPase. Step F is a conversion of nmoles dUMP to µmoles dUMP.

2.8.) Temperature and pH optima

For temperature optima trials, assay buffer (AB; 10 mM bicine, 2 mM $MgCl_2$, pH 8.0) was warmed to the experimental temperature (ranging from 20 to 70°C) for 10 minutes in a water bath. For pH optima trials, buffers were; 10 mM citrate, 2 mM $MgCl_2$ pH 4.0, 5.0, 6.0; 10 mM HEPES, 2 mM $MgCl_2$ pH 7.0, 8.0; or 10 mM borate, 2 mM $MgCl_2$ pH 9.0, 10.0), and assays were conducted at 25°C.

To start a reaction, protein (0.5 µg of the core or 0.25 µg of the full-length) was added to 3 mL of AB and the reaction was quenched after 10 minute with 60 µL 1 N HCl.

2.9.) Metal Dependence

The metal-dependence of dUTPase (0.5 µg of the core or 0.25 µg of the full-length) at 25°C was measured using 3 mL of AB containing 2 mM of: $MgCl_2$, $MnCl_2$, $CaCl_2$ or EDTA. Reactions were quenched after 10 minutes with 60 µL 1 N HCl.

3.0.) Renaturation Potential

Samples (0.5 µg of the core dUTPase or 0.25 µg of the full-length dUTPase) in AB were heated at 80°C for 10 minutes and either assayed immediately or placed at 4˚C for 20 minutes before assaying in AB at 25°C. Reactions were quenched after 10 minute with 60 µL 1 N HCl.

The activity of untreated dUTPase served as the control.

Chapter 3: Results

3.) Protein Sequences

The sequences of the three recombinant versions of dUTPase investigated are shown in Figure 3.1. Each version was engineered with a His_6-tag for purification and a thrombin cleavage sequence for removal of the affinity tag.

```
         10          20          30          40          50          60
MGSSHHHHHH SSGLVPRGSH MPIEQKYFSL FSNLFKRLTT NNNNNNYLKM APPNFETFKV
                                             MGSSHHHHHH SSGLVPRGSH MFKV

                  Motif I                                      Motif II
         70          80          90         100         110         120
KKLSDKAIIP QRGSKGAAGY DLSSAHELVV PAHGKALAMT DLQIAIPDGT YGRIAPRSGL
                                   MGSSHHHHHH SSGLVPRGSH MGGGGSGL

            Motif III                    Motif IV
        130         140         150         160         170         180
AWKNFIDCGA GVIDSDYRGN VGVVLFNHSD VDFKVAVGDR VAQLIFERIV TPEPLEVDEI

      Motif V
        190       199
DETQRGAGGF GSTGVKVQN
```

Figure 3.1: Amino acid sequence of full-length, core, and truncated recombinant dUTPase. The His_6-tag is annotated by red lettering. The thrombin cleavage consensus sequence is noted in blue lettering. Cleavage occurs at the peptide bond between arginine (R) and glycine (G). A secondary thrombin cleavage site is noted with yellow lettering (position 118). The five highly conserved motifs are highlighted in black. The core version lacks the 37 N-terminal residues of the full-length version. The truncated version lacks the 97 N-terminal residues of the full-length version. Four glycines were added to the N-terminus of the truncated sequence to provide flexibility.

3.1) Sequence Analysis

Protein BLAST of the full-length dUTPase sequence revealed similarity to other dUTPases. Results showed a sequence identity consistently around 70% with the majority of identity being from the five conserved motifs of dUTPase. Although high

sequence identity was observed, the leading region of dUTPase was consistently unique from other compared sequences.

Lack of published information on the role of the N-terminus in dUTPase is a driving factor in the investigation of the 37 N-terminal residues of the *D. discoideum* dUTPase. Sequence information was analyzed using bioinformatics tools prior to biochemical characterization as a means to place the *D. discoideum* sequence in context of similarity to other dUTPases. Protein BLAST (National Institute of Biotechnology Information, Bethesda, MD) was used to comprehensively investigate the novelty of the specified sequence. The 37 N-terminal residues: **MPIEQKYFSLFSNLFKRLTTNNNNNNYLKMAPPNFET**, were entered into protein BLAST to identify potential sequence identity across known protein sequences. The returned data revealed there to be no protein sequence matches related to the *D. discoideum* N-terminal residues. The lack of hits on the 37 N-terminal residues reveals the absence of conservation between the N-terminus of the *D. discoideum* dUTPase and other dUTPases.

The enzyme dUTPase belonging to *D. discoideum* is among the superfamily of homotrimeric dUTPases. Sequence alignment analysis reveals complete conservation of the five conserved motifs (composed of catalytic residues) among homotrimeric dUTPases (Persson *et al.,* 2001). However, an interesting aspect of the *D. discoideum* dUTPase sequence is the N-terminus that does not contain residues annotated as being crucial for enzymatic functioning (trimer interface or catalytic residues). The N-terminal

residues of dUTPase differ among species in length and residue composition. Overall, the top fifty-one BLAST hits were analyzed (Appendix 5.5) for N-terminal length quantification. The net residue difference was measured with the *D. discoideum* dUTPase as a template. Calculations revealed of the fifty-one measured, twenty-three sequences had longer N-termini (1 to 54 residues longer) and twenty-six sequences had shorter N-termini (1 to 50 residues).

The majority of the top fifty-one BLAST hits belonged to the Kingdom Plantae (60%) with the next greatest representation from Fungi (16%). These data characterize the *D. discoideum* dUTPase N-terminus as longer than 53% of entries analyzed.

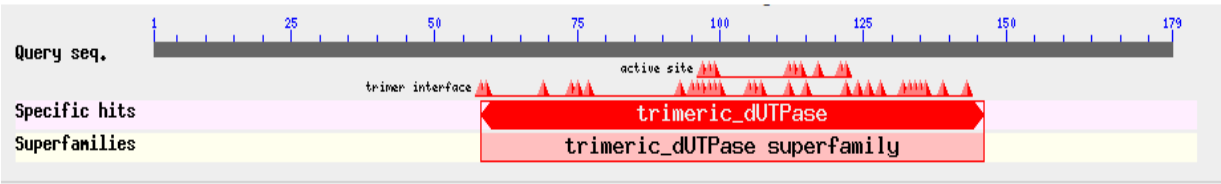

Figure 3.2: Basic Local Alignment Search Tool (BLAST) of full-length *D. discoideum* dUTPase. This figure represents the protein sequence annotation of dUTPase, revealing the location of trimer interface and active site residues. BLAST results can be located in the Appendix Section 5.5.

3.2) Full-length, core, and truncated dUTPase purification

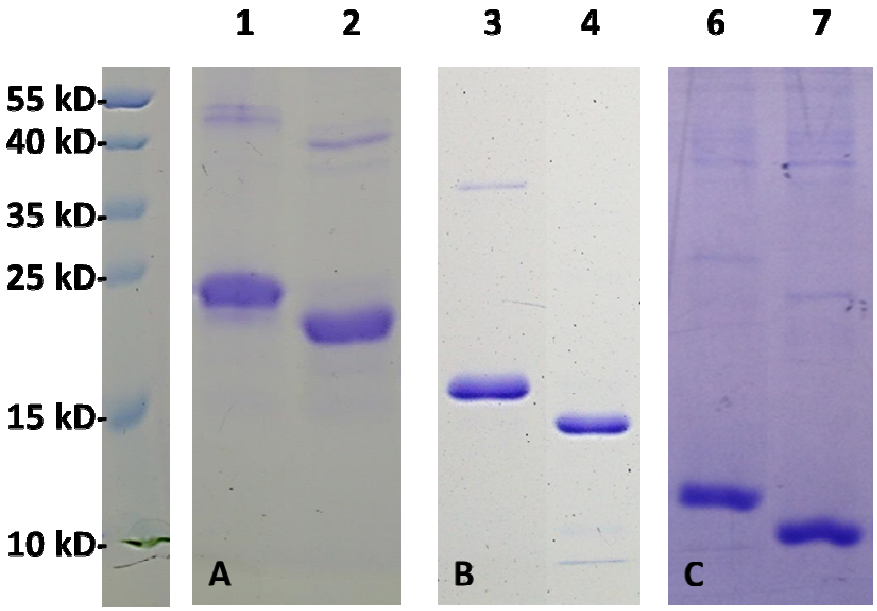

Figure 3.3: SDS-PAGE (15% acrylamide) analysis of purified recombinant dUTPase. Recombinant versions of full-length, core, and truncated dUTPase were cloned protein expression. Each protein was expressed with a His_6-tag for purification using HiTrap Chelating HP (1 mL bed volume); GE Healthcare Life Sciences, Pittsburgh, PA). A.) Purified and processed full-length polypeptide. Lanes 1 and 2 contain 8.0 µg His_6-tagged full-length and 8.0 µg post-cleavage full-length dUTPase. B.) Purified and processed core polypeptide. Lane 3 contains 5 µg His_6-tagged core dUTPase. Lane 4 contains 5.0 µg post-cleavage core dUTPase. C.) Purified and processed truncated polypeptide. Lanes 6 and 7 contain, respectively, 5.0 µg His_6-tagged and 5.0 µg post-cleavage truncated dUTPase. Protein ladder markers and sizes are displayed on the left.

This SDS-PAGE shows the processing of the His_6-tagged full-length, core, and truncated dUTPase polypeptides into a post-cleavage form. Protein was quantified using OD280 absorption readings. All three versions of dUTPase were isolated with high purity (>95%) and sizes were close to expected when analyzed using SDS-PAGE. Specifically, each dUTPase version ran higher than expected, appearing to have a greater molecular mass than predicted sizes. Each protein sample contained higher molecular weight species when analyzed by SDS-PAGE. Although no immunoblot analyses were utilized, the higher molecular weight species are believed to be trimeric versions of dUTPase.

The higher molecular weight species, which are present before and after cleavage decrease in mass after cleavage consistent with the removal of the His_6-tag. The predicted mass of high molecular weight species based on the protein ladder was accurate in the truncated version but lower than expected in the full-length and core versions respectively. Within the protein ladder, lower molecular weight species were found to have linear migration distances. While higher molecular weight species also migrated linearly, the slope was different than the regression of the lower molecular weight species. Due to the change in slope the R^2 for the entire data set was much lower than either the low or high molecular weight end individually (see Appendix, Figure 5.4). For purposes of size estimation, the data for the lower molecular weight species was used since it covers the range of expected protein sizes (see Materials and Methods, Figure 2.2).

3.3) Michaelis-Menten Kinetics

3.3.1) Truncated dUTPase

Truncated dUTPase was assayed but no measurable activity was detected. Refolding attempts included exposing recombinant protein to triton X-100, arginine, and ethanol for an extended period of time for regaining activity (see Materials and Methods Section 2.2.3). The rationale for the refolding trials was to provide the protein an optimal environment for proper folding. The methodology was originally published by Clark *et al.,* 1999.

3.3.2) Core and full-length dUTPase

Kinetic constants were found using the cresol red method (see Materials and Methods Section 2.6.1). For each dUTP hydrolyzed, two protons are released (see introduction Figure 1.1 for reaction mechanism). The pH indicator cresol red undergoes a colorimetric change as the pH changes. The change in color is detected at 573 nm allowing dUTPase activity to be monitored. Kinetic constants were calculated using a plot of the slope value from the linear portion of each reaction trace.

Varying amounts of core dUTPase was measured at 10 µM dUTP to determine linearity of the reaction. The amount of dUTPase used for the reaction displayed different rates. The concentration of 5.85 µg/mL (1.25×10^{-4} µmoles dUTPase) fell in the middle of the linear portion and was used for all of the analyses used to calculate K_m and V_{max} of the core version.

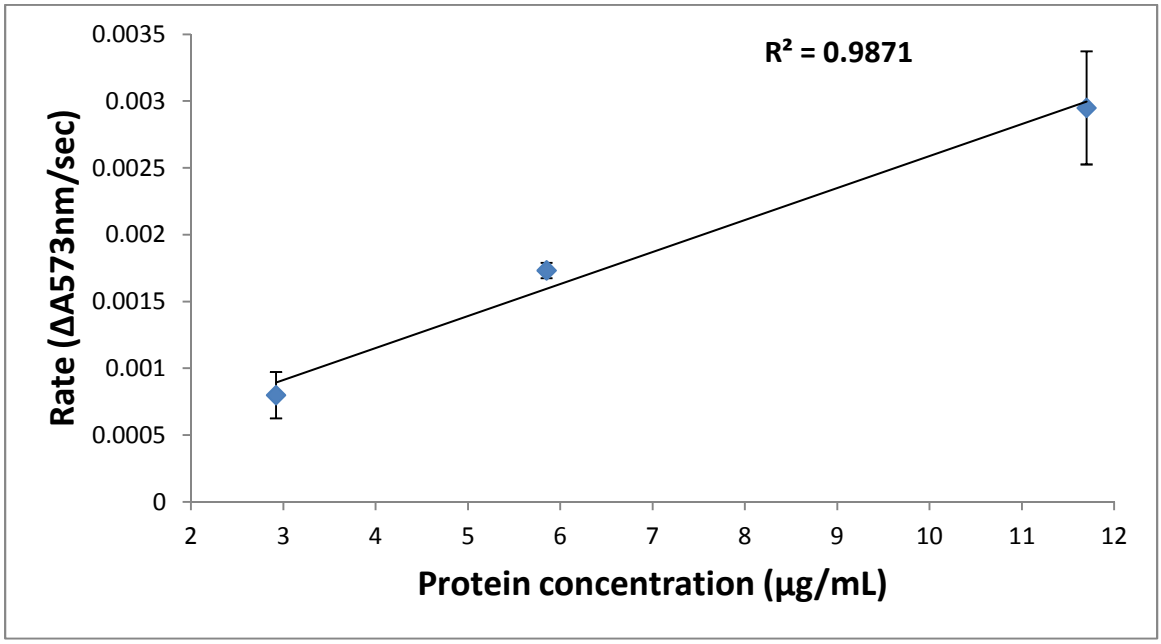

Figure 3.4: Range of linearity for the activity of the core dUTPase. Protein concentrations of 2.92 µg/mL, 5.85 µg/mL, and 11.7 µg/mL were used to determine the linear portion of the activity curve at 10 µM dUTP.

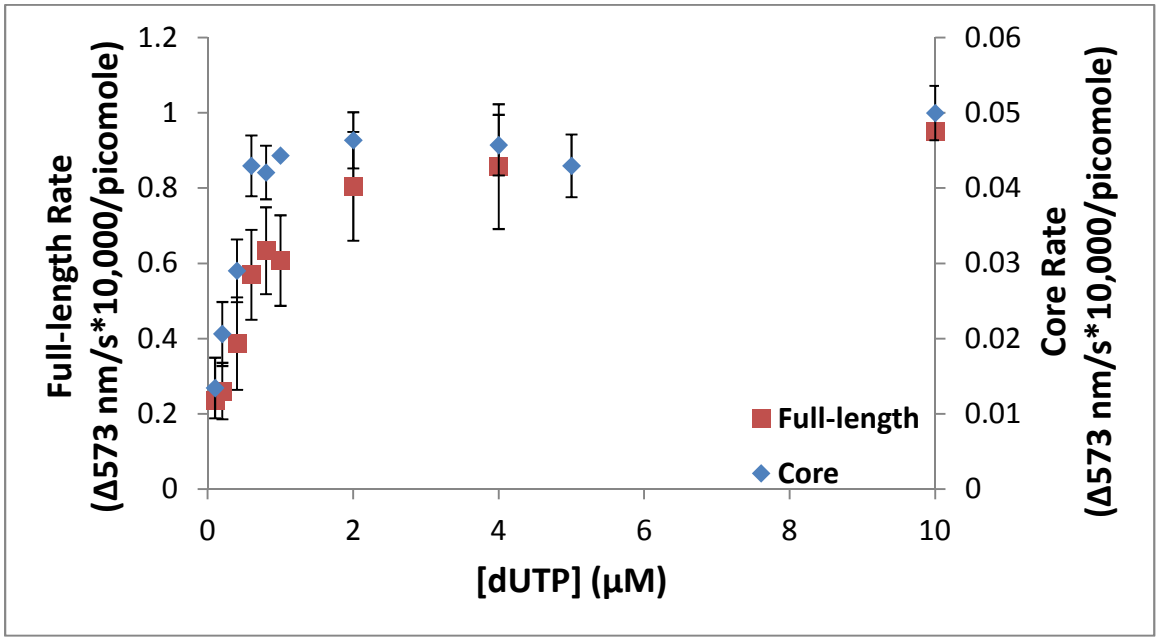

Figure 3.5: Michaelis-Menten kinetics of core and full-length dUTPase used to determine K_m and V_{max}. Rates were determined by measuring the decrease in OD573 of the linear portion of the absorbance trace, and normalized to protein mass. Error bars represent the standard deviation within each substrate concentration. Ten replicates were analyzed per substrate concentration. KV collected data for the core dUTPase. Data for the full-length dUTPase were collected by Kathleen Williams and Catherine Chia.

Michaelis-Menten data acquired (Figure 3.5) was replicated and used directly to generate kinetic constants of the full-length and core dUTPase. K_m values were calculated by Enzfitter kinetic software and Lineweaver-Burke plot analyses. Enzfitter kinetic analyses utilized non-linear regression data (Figure 3.5) fitted to the Michaelis-Menten equation:

$$v = \frac{V_{max} \times [S]}{K_m + [S]}$$

Where V_{max} is the maximal velocity of the enzyme, *[S]* is the substrate concentration, K_m is the Michaelis-Menten constant, and *V* is the initial velocity of the enzyme.

The V_{max} and k_{cat} constants were calculated using mathematical expressions characterized by Larsson, *et al.* (1996). V_{max} was calculated by first finding the concentration of the product at time *t* and then using it to solve directly with equations described in Materials and Methods Section 2.6.1. k_{cat} was similarly calculated using equations described in Materials and Methods Section 2.6.1. Examples of calculations can be seen in Appendix Sections 5.1 and 5.2.

Version of dUTPase	**K_m (µM)**	**V_{max} (µM s^{-1})**	**k_{cat} (s^{-1})**	**k_{cat}/K_m (1/µM*s)**
Core	0.22±0.05	0.097	0.23±0.08	1.05
Full-length	0.50±0.06	0.153	0.37±0.08	0.74
Fold difference (Core/Full-length)	0.44	0.63	0.62	1.42

Table 3.1: Kinetic constants of core and full-length dUTPase.

The degree of variation between both versions is expressed in the 'fold difference' row (Table 3.1) and was calculated by dividing the 'core value' by the 'full-length value'. Kinetic constants acquired reveal the core version to be catalytically superior to the full-length dUTPase based on k_{cat}/K_m values. Specifically, the kinetic constants that are superior in the core version are the K_m and k_{cat}/K_m with a 0.44-fold decrease and 1.42-fold increase respectively. The full-length version has superior V_{max} and k_{cat} constants with 1.58-fold and 1.68-fold higher values than the core.

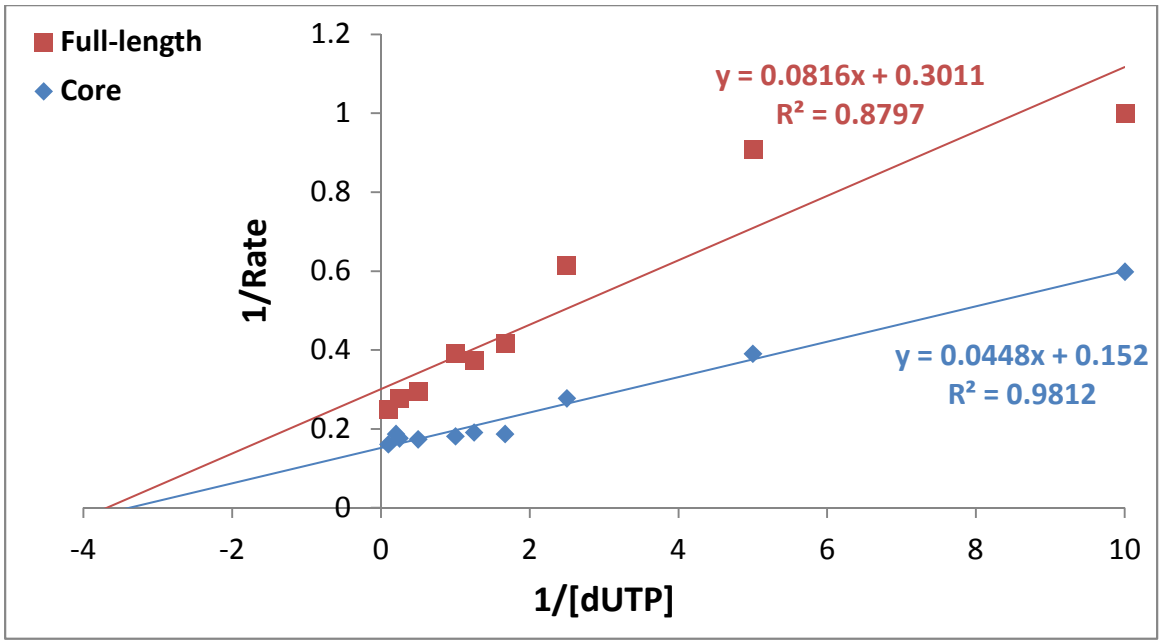

Figure 3.6: Lineweaver-Burke plot of full-length and core activity data. Calculated V_{max} and K_m from this plot are presented in Table 3.2.

	Y-intercept	X-intercept	K_m (µM)	V_{max}
Core	0.152	-3.39	0.295	6.60
Full-length	0.301	-3.69	0.271	3.32
Fold difference (Core/Full-length)	-	-	1.09	1.99

Table 3.2: Lineweaver-Burke plot calculated kinetic constants. Y-intercept and x-intercept values were calculated from regression equations in Figure 3.6. The fold difference values were calculated relative to the full-length version. Units for V_{max} values are ΔA573nm*10,000/s.

K_m and V_{max} values approximated using the Lineweaver-Burke plot (Figure 3.6) can be seen in Table 3.2. The double reciprocal representation of the kinetic data is a means for extrapolating the V_{max} and K_m by plotting the reciprocal of generated kinetic data. Kinetic constants calculated using the Enzfitter software (Table 3.1) are used below for comparison with data from Figure 3.6. Based on the Lineweaver-Burke plot, the V_{max} (ΔA573nm*10,000/s) for the full-length is 3.32 while the V_{max} of the core is 6.60. The

calculated K_m values were 0.295 µM for the core and 0.271 µM for the full-length. The data from the Lineweaver-Burke plot (Table 3.2) shows a 1.99-fold difference in the V_{max} values and a 1.09-fold difference in the K_m values compared to the 0.63- and 0.44-fold differences of the software-calculated constants. The discrepancy in comparative data may be attributable to the R^2 value, or the measure of how well data points fit a specific model. The full length best fit line has an R^2 of 0.8797 compared to the core version best fit value of 0.9812. The lower R^2 of the full-length linear best fit reduces the credibility of the double reciprocal plot data. The lower credibility of the full-length data series may account for the difference in V_{max} and K_m values of the full-length dUTPase when compared to values in Table 3.1.

3.4) End-point assays

3.4.1) Temperature Optima

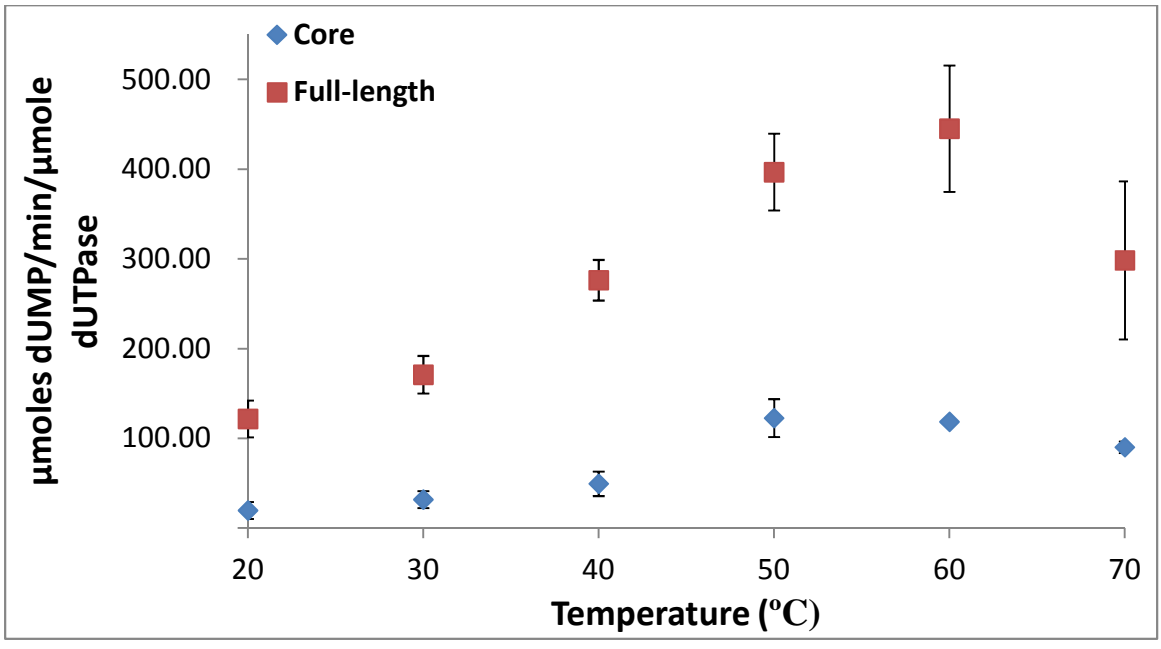

Figure 3.7: The activity of the core and full-length dUTPases as a function of temperature. Data and error bars are representative of three trials.

The core dUTPase displayed a shallow peak with a maximum between 50°C and 60°C with the highest specific activity at 50°C of 122.6 µmoles dUMP/minute/µmole core. The full-length dUTPase showed a more distinct temperature optimum of 60°C with a specific activity of 445.1 µmoles dUMP/minute/µmole full-length.

3.4.2) pH Optima

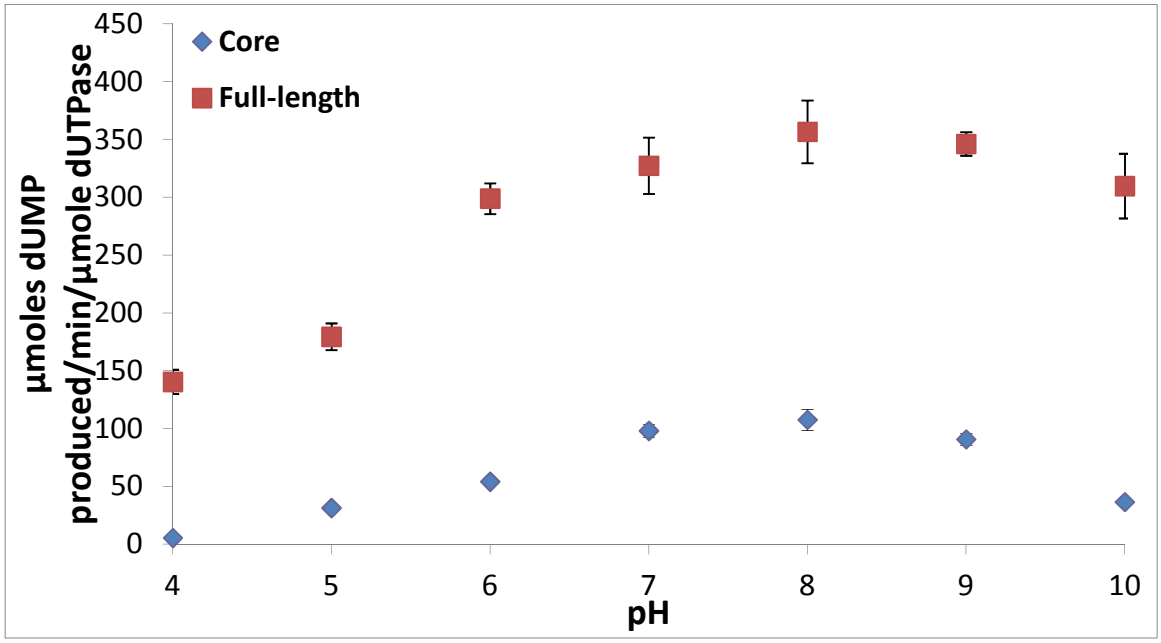

Figure 3.8: The activity of the core and full-length dUTPases as a function of pH. Reactions at each pH were performed in triplicate within each set. Data and error bars are representative of three trials.

The full-length dUTPase showed a pH optimum of 8.0 with a specific activity of 356.6 µmole dUMP/min/µmole full-length dUTPase. The core displayed optimal activity at pH 8.0 with a specific activity of 107.6 µmole dUMP/min/µmole core.

The end-point assay measuring amount of dUMP per reaction was used to generate the pH and temperature optima data (see Materials and Methods Section 2.7). Temperature and pH optimum datum was obtained with 10.74 x 10^{-6} µmoles core and

4.21 x 10^{-6} µmoles full-length. The V_{max} was calculated to be 0.63-fold lower in the core (Table 3.1). If the difference in V_{max} of core and full-length dUTPases was consistent in the temperature trials, the maximal rate of the core at its temperature optimum would be 280.4 µmoles dUMP/minute/µmole core. The measured specific activity for the core version was 122.6 µmoles dUMP/minute/µmole core, making the maximum rate 0.73-fold lower than the full-length. A 0.63-fold lower V_{max} would predict the specific activity of the pH optimum to be 224.7 µmole dUMP/min/µmole core dUTPase. The observed rate of 107.6 µmole dUMP/min/µmole core places the core dUTPase rate at 0.70-fold lower than the observed full-length value.

3.4.3) Metal dependence

The core version displayed varying levels of activity depending on the compound. The highest level of activity detected was in the presence of Mg (100±4.0%) followed by Mn (83.5±11%), Ca (43.2±27%), and EDTA (0.3±25%). Since activity for the core enzyme was highest with Mg, it was considered 100% activity with the other activity levels represented as a relative percentage.

Similarly, the full-length version displayed varying levels of activity. The highest level of activity was with Mg (100±0.2%) followed by Mn (74.1±3.9%), Ca (18.24±17.8%), and EDTA (2.27±48.8%).

Comparing the effect various compounds have on activity between the two versions places the presented data into context. The relative activity measured with Mn was 9.4% higher in the core version than the full-length. The effect Ca had on activity

was greater in the full-length where it reduced activity by 81.76% whereas the core version showed a 56.8% decrease. Activity of both versions with EDTA was low with a 99.7% and 97.7% reduction in activity for the core and full-length respectively.

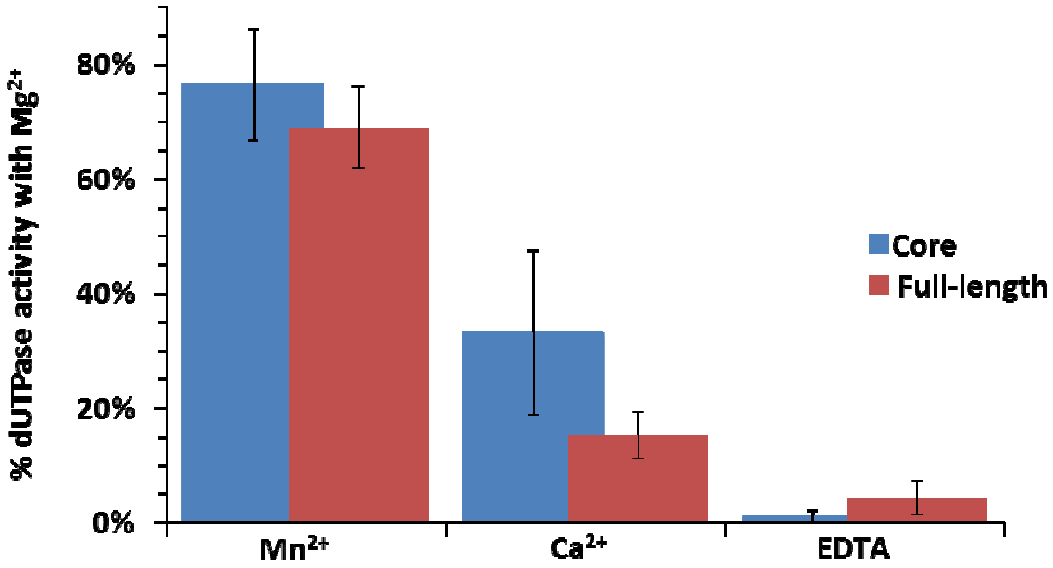

Figure 3.9: Metal dependence of core and full-length dUTPase. Percentages represent percent activity relative to $MgCl_2$ which was assumed to have optimal activity (Vértessy and Tóth, 2009). A representative trial is presented for each version. See Appendix Section 5.5 for data from other trials and table presentation of the data.

3.4.4) Renaturation potential

Heat treated core showed an initial 82% decrease in activity. Heat treated core placed at 4°C for 10 minutes had 60.8% more activity than the 80˚C treated sample. The decrease in activity of heat treated core shows thermal denaturation occurred by the end of the 10 minute treatment. The 60.8% increase in activity following 10 minutes at 4°C is indicative of renaturation.

Heat treated full-length dUTPase showed a 97.59-99.7% decrease in activity. Heat treated full-length placed at 4°C for 10 minutes had 0.07-17.1% more activity than the original heat treated sample. The 97.59-99.7% decrease in activity following heat treatment reflects near complete thermal denaturation. Denaturation of full-length dUTPase is largely irreversible as indicated by the meager activity recovered following the 4°C treatment.

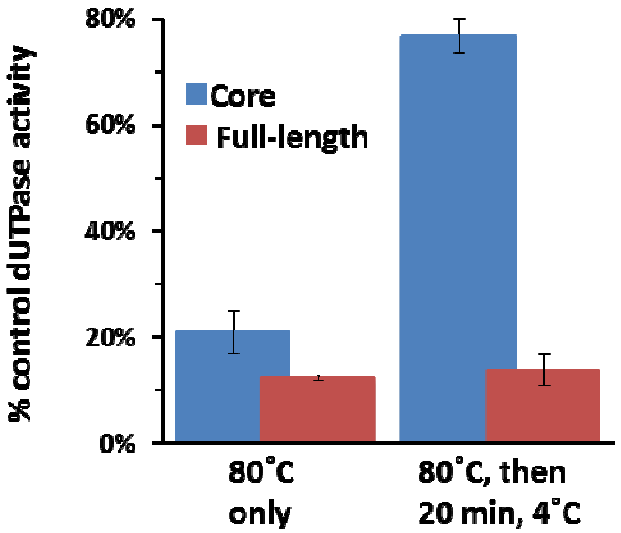

Figure 3.10: Renaturation potential of core and full-length dUTPases when heated to 80°C. Heat treatment was performed as described in Materials and Methods. Data presented are representative of two trials.

Chapter 4: Discussion

4.1) Background

The role of dUTPase is to catalyze the hydrolysis of dUTP to dUMP. This enzymatic action allows for dTTP biosynthesis by providing the precursor molecule dUMP and decreases the dUTP pool. An increased dUTP pool could prove detrimental to the cell through an increase in uracil incorporation into DNA during synthesis. Increased base excision repair can be used to remove uracilated DNA but in doing so compromises the overall stability of the molecule. In addition, dUMP is required as the precursor to the molecule dTTP, required for DNA synthesis.

The two versions of *Dictyostelium discoideum* dUTPase investigated in this study differ only by 37 N-terminal residues, referred to as the leading region. The leading region contains no residues noted in the literature as participating in catalysis or trimer interface and therefore has a questionable contribution to the overall activity of the enzyme. This study focuses on understanding the role of the N-terminal residues as it pertains to the overall activity of the enzyme. Specifically, kinetic constants, temperature optima, pH optima, metal dependence, and renaturation potential data were obtained and will be analyzed in Section 4.3 and 4.4.

There is a lack of enzymatic data on the truncated dUTPase due to the inability of obtaining active enzyme through purification and refolding trials. Truncated dUTPase data generated previously used a glutathione-s-transferase (GST) fusion protein with the

core dUTPase sequence (Nguyen 2010) rather than the His_6-tagged construct used in this study. Cleavage of the fusion protein occurred after purification and spontaneous folding. It is believed that the conformation of the fusion protein allowed for the truncated peptide to remain in an active state after cleavage. Early experimentation aimed at replicating truncated dUTPase data with the same construct was unsuccessful due to the inability to separate cleavage products. Additional trials aimed at obtaining truncated data utilized the core version construct and a unique cleavage protocol to cleave the core version into a truncated and remnant peptide but was ultimately unsuccessful in separation of the truncated and remnant peptides.

4.2) Protein Purification

Purification of recombinant proteins was carried out as described in Materials and Methods Section 2.3. His_6-tagged protein was cleaved to remove the purification tag using the protease thrombin. Cleavage efficiency was greater than 95% among all versions of dUTPase investigated. Purity of sample preparations was high (>95%), any other protein species present were dUTPase in origin and were a result of cleavage at the secondary thrombin cleavage site (see Results, Figure 3.3). All statements regarding purity of sample preparations were based on data from SDS-PAGE and Coomassie blue staining.

4.3) Enzyme Kinetics and Analyses

The Michaelis constant represents the concentration of substrate required for an enzyme to operate at $1/2V_{max}$. In addition, the K_m may be an inverse relationship of substrate affinity for an enzyme active site (Lehninger, *et al*, 2005). In kinetic analyses the K_m inversely represents the catalytic efficiency of an enzyme. The K_m of dUTPases varies greatly among compared species with the largest reported value of 13.2±0.06 µM (*S. cerevisiae*) and the smallest of 0.1±0.01 µM (*A. thaliana*) (see Introduction, Table 1.1). This study established the K_m of the full-length and core dUTPases to be 0.50±0.06 µM and 0.22±0.06 µM, respectively (see Results, Table 3.1). The two-fold decrease in K_m indicates that the absence of the 37 N-terminal residues on the core version increased active site affinity for the substrate.

The maximal velocity (V_{max}) of an enzyme refers to the maximum rate at which the enzyme can function under saturating conditions of substrate. The calculated V_{max} observed for the versions was 0.097 µM s^{-1} for the core and 0.153 µM s^{-1} for the full-length, a 0.63-fold decrease (see Results, Table 3.1). The turnover number (k_{cat}) refers to the number of molecules processed for each active site per unit time. The k_{cat} was 0.23 s^{-1} for the core version and 0.37 s^{-1} for the full-length, a 0.62-fold decrease. Data analyses reveals the residues which compose the N-terminus of dUTPase although not annotated as participating in catalysis, affect the rate of activity (see Results, Figure 3.2). Given the data presented, a role of the specific residues may be stability of the active site. A level of stability surrounding the active site of enzymes would be necessary to perform catalysis at the maximal rate. By removing the residues that provide the added stability to the active site, the enzyme has a smaller maximal velocity.

The core version, with a K_m 0.44-fold smaller than the full-length version is evidence for a role of the N-terminus in substrate binding efficiency. The calculated K_m values support the hypothesis stated that if the 37 N-terminal residues are absent the K_m will be reduced. The change in K_m prediction was made with the assumption that the leading region of the full-length acts as a covering over the active site. Figure 4.1 was generated using Chimera molecular modeling program with the published *S. cerevisiae* dUTPase structure (Tchigvintsev *et al.,* 2011; PDB ID 3P48) and a modeled *D. discoideum* dUTPase monomer. The modeled *D. discoideum* dUTPase monomer was generated using Phyre2 (Protein Fold Recognition Server). Figure 4.1 shows the predicted locations of the N-termini of the *S. cerevisiae* trimer and a single subunit of the *D. discoideum* dUTPase. The *D. discoideum* dUTPase N-terminus is 29 residues longer than the *S. cerevisiae* dUTPase. Published crystal structure for *S. cerevisiae* (Tchigvintsev *et al.,* 2011) lacks the structural data for only its first four residues, and the modeled *D. discoideum* dUTPase, due to its longer N-terminus, lacks structural data for the first thirty-five residues. Since the number of residues lacking from the *S. cerevisiae* structure is far less than the *D. discoideum* it is assumed to be a more representative location for the N-terminus. The first seven highlighted *D. discoideum* residues represent the first seven that have structural data modeled. This is a result of a sequence alignment with sequences that have structural data. The highlighted residues that represent the N-terminal region lie along the bottom of the molecule (Figure 4.1 Panel A). Although the N-terminus appears to be located away from the active site both N and C-termini are predicted to be in close proximity to the active sites as seen in Figure 4.1 Panel B.

Similar to other dUTPase structures, the active sites of the *D. discoideum* dUTPase (Panel A) can be characterized as external, providing evidence that the N-termini of the *D. discoideum* dUTPase could in theory provide coverage.

Without the N-termini covering the active site, substrate molecules would be able to more readily interact with the active site. Further experimentation would be required to visualize the active site of the core and the full-length dUTPases. X-ray crystallographic analysis could be employed to determine the N-terminal structure and exact mechanism responsible for the reduction in the K_m of the core version. This could be done by generating a crystal structure for the full-length and core versions of the *D. discoideum* enzyme. Structural overlay analyses could then be performed with programs such as Pymol, Coot, or Chimera.

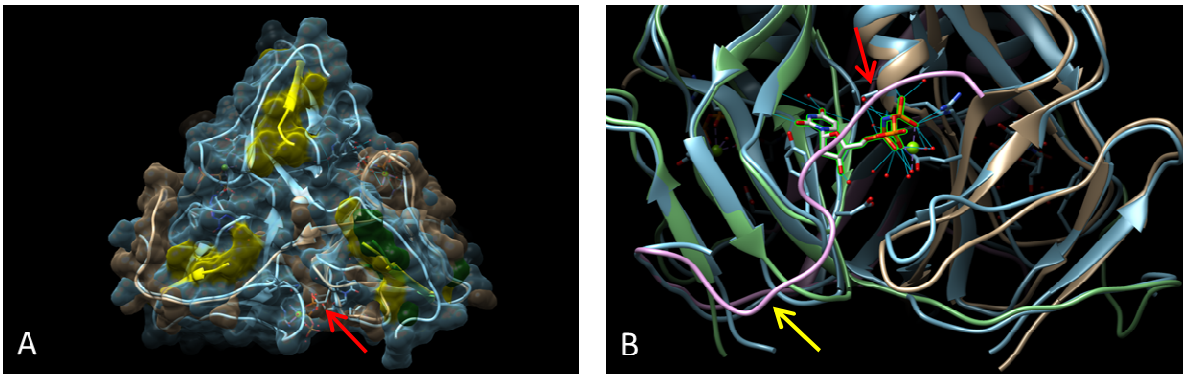

Figure 4.1: *D. discoideum* dUTPase modeled using the known structure of dUTPase from S. *cerevisiae*. *S. cerevisiae* dUTPase (trimer) is colored light blue with the first seven N-terminal residues of each subunit colored yellow. Hypothetical *D. discoideum* dUTPase (monomer) is tan with the first seven N-terminal residues of the monomer colored green. Panel A represents a bottom view of overlaid *S. cerevisiae* trimeric and *D. discoideum* monomeric dUTPase structures (looking bottom up). The red arrow in Panel A indicates the location of a single active site. The active site featured is in complex with the nucleotide analogue dUMPnPP. Panel B represents a zoomed in side view of trimeric *S. cerevisiae* (blue) and *D. discoideum* (Chain A- green, Chain B- tan, Chain C- purple)

dUTPase in complex with dUMPnPP. The red arrow indicates the C-terminus of Chain C covering the featured active site. The yellow arrow indicates the location of the N-terminus of Chain A. The *D. discoideum* dUTPase structure was generated using Phyre 2 (Protein Fold Recognition Server). Structural comparison was carried out using Chimera molecular modeling software.

4.4) Temperature and pH Optima Analyses

The pH and temperature effects on activity reveal differences in the optimal operating temperature and pH between the two versions of dUTPase. The temperature optima data of the core version displays a plateau indicative of a broad optima range. The full-length version displayed a temperature optimum of 60°C, a relatively high value since *D. discoideum* grows optimally at 22°C. I propose that the high optimal temperature is not a result of evolving due to necessity but rather a testament to the overall stability of the enzyme. Since sequence homology to other trimeric dUTPases makes it likely *D. discoideum* dUTPase is also a trimer, stability and assembly of the functional protein must be a favorable process. The favorability of the trimer assembly is an evolutionary characteristic that would contribute to stability and resistance to temperature extremes.

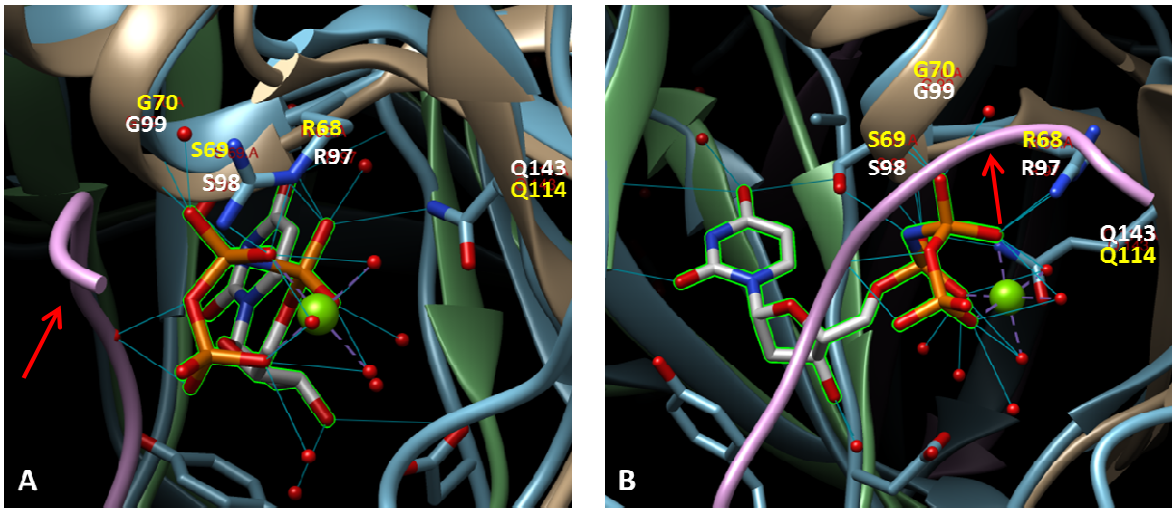

Figure 4.2: Active site comparison of *S. cerevisiae* and hypothetical *D. discoideum* dUTPase. Crystal structure of *S. cerevisiae* dUTPase bound with dUMPnPP nucleotide analogue (Tchigvintsev, *et al.*, 2011) was accessed using protein data base (PDB) ID 3P48. The full-length *D. discoideum* dUTPase sequence was used to generate a hypothetical structure based on homology to sequences with known structures using the online tool Phyre2. The hypothetical *D. discoideum* structure (Chain A- green, Chain B- tan, Chain C- purple) and the known yeast dUTPase (blue peptide) structures were then paired to understand the similarities among the residues responsible for substrate binding in the active site. Labeled residues (white, *D. discoideum*; yellow, *S. cerevisiae*) are part of conserved motif II with the exception of Q114 and Q143 which are part of motif IV. The residues Q114 and Q143 are labeled due to their conservation as active site residues. The thin blue lines linking separate atoms represent predicted hydrogen bonding between residues in the active site and the analogue in the yeast enzyme. The red spheres represent water molecules. Dashed purple lines represent Mg^{2+} coordination bonds. Random coiled regions are narrow tubes while alpha helices and beta sheets are broad and flat. The atoms of molecules are identified by color with dark red representing oxygen, orange representing phosphorous, dark blue representing nitrogen, and green identifying magnesium. Red arrows in Panels A and B indicate the location of the C-terminus. Both panels A and B are presented to provide added perspective on the *D. discoideum* dUTPase active site.

Structural alignment of the published dUTPase structures and the hypothetical *D. discoideum* dUTPase can be used for analyses, specifically yielding insight into substrate binding and the catalytic mechanism. The hypothetical *D. discoideum* dUTPase structure was matched with the *S. cerevisiae* dUTPase (Figure 4.2) because it is the

closest related organism that has a published dUTPase structure with a substrate-bound active site. The *S. cerevisiae* dUTPase has a reported K_m value of 13.2±.06 µM (Tchigvintsev *et al.,* 2011), or 26.4-fold higher than the full-length *D. discoideum* dUTPase. The difference between the two K_m values implies either an additional level of modulation or structural differences in the active site.

Although it appears there is a difference in the C-termini of the *D. discoideum* and *S. cerevisiae* dUTPase structural comparison yields a high similarity between the two structures (Figure 4.2). Evidence suggests the C-termini of the *D. discoideum* and *S. cerevisiae* dUTPase cover active sites. The modeled *D. discoideum* dUTPase C-termini can be visualized in Figure 4.2 and clearly covers the active site. The *S. cerevisiae* dUTPase C-terminus structure featured in Figure 4.1 (Panel B) lacks fourteen additional residues present in sequence data. The additional C-terminal residues were visualized by X-ray crystallographic analyses (Tchigvintsev *et al.,* 2011) and have been revealed to cover the active site in the presence of substrate. The close proximity of the C-termini to the active site could have implications in the kinetic characteristics of the enzyme; particularly the K_m due to increased steric hindrance. As stated earlier, the *Km* of the *S. cerevisiae* dUTPase is 26.4-fold higher than the full-length *D. discoideum* dUTPase. However, it should be stated that the relatively high K_m of *S. cerevisiae* may be attributable to other presently unknown structural characteristics rather than the position of the C-termini.

Although based on hypothetical modeling, Figure 4.2 provides strong evidence for the proximity of the *D. discoideum* dUTPase C-terminus and the active site. Due to the N-terminal (Figure 4.1) and C-terminal (Figure 4.2) modeling above, it is a possibility that both the N- and C-termini exist in proximity to the active site.

Structurally this characteristic may increase the steric hindrance for dUTP and would explain the higher K_m value in the *D. discoideum* full-length dUTPase when compared to the core version. The majority of the hydrogen bonds formed between active site residues in motif II and the substrate in the yeast protein likely have counterparts in the *D. discoideum* enzyme due to the identical residues. Since the dUTPase active site is external, it relies on water for many hydrogen bonds used for stabilizing dUTP. For further understanding of the mechanism behind the high temperature optimum in dUTPase, a structural and kinetic study of a dUTPase with a low temperature optimum would be required to draw conclusions about the precise mechanism that allows for different properties. The dUTPase of the archaeon *Thermococcus pacificus* would be an ideal candidate for a dUTPase with an extreme temperature optimum once a crystal structure is published to accompany the sequence and enzyme characterization published by Cho *et al.* (2012). The mechanism that allows a high temperature optimum (85°C) of *T. pacificus* compared to the moderate temperature optima of the full-length dUTPase (60°C) may be elucidated with accompanying crystal structure data. Figure 4.3 displays a sequence alignment of several dUTPases with temperature optima ranging from 50°C (PBCV-1) to 85°C (*T. pacificus*) with the five highly conserved motifs highlighted.

```
D. discoideum  MPIEQKYFSLFSNLFKRLTTNNNNNNYLKMAPPNFETFKVKKLSDKAIIP 50
PBCV-1         ---------------------------------MSSLLVKKLVESATTP 16
T. pacificus   --------------------MMLPDWKIRKEILIEPFNEKSLQPAGYDL 29

               QRGSKGAAG---YDLSSAHELVVPAHGKALAMTDLQIAIPDGTYGRIAPR 97
               MRGSEGAAG---YDISSVEDVVVPAMGRIAVSTGISIRVPNGTYGRIAPR 63
               RVGKEAYINGRLINVEKEGKVVIPPKTHALILTLERVKLPDDVMGDMKLR 79

               SGLAWKNFIDCGAGVIDSDYRGNVGVVLFNHSDVDFKVAVGDRVAQLIFE 147
               SGLAYKYGIDVLAGVIDSDYRGELKAILYNTTERDYIIKKGDRIAQLILE 113
               SSLAREGLLGSFA-WVDPGWDGNLTLMLFNASEEPVELNYGERFVQIAFI 128

               RIVTPEPLEVDEIDETQRGAGGFGSTGVKVQN 179
               QIVTPDVAVVLELEDTARGGGGFGSTGI---- 141
               RLEGP---AKNPYRGNYQGSQHLVLSKRRK-- 155
```

Figure 4.3: Sequence comparison among dUTPases with a range of temperature optima. The five highly conserved motifs of homotrimeric dUTPases are highlighted in black. See Section 4.5 for details about Serine 81 (red) and Arginine 84 (blue) of the dUTPase of Chorella virus PBCV-1. The multiple sequence alignment was performed by the European Bioinformatics Institute (http://www.ebi.ac.uk/Tools/msa/clustalw2/). Protein accession numbers were *D. discoideum* (Q54BW5.1), PBCV-1 (NP_048907.1), and *T. pacificus* (AFJ04533). Motif I of *T. pacificus* is a greater distance from motif II than the other two sequences, resulting in the observed misalignment.

The sequence comparison in Figure 4.3 is a preliminary step in understanding potential mechanisms that allow for varying degrees of temperature optima. Through analysis of published work on the above sequences it is possible to draw conclusions as to the mechanisms that allow for variation in characteristics.

4.5) Structural Analyses

Sequence information for *Thermococcus pacificus* was used to generate a hypothetical structural model with the modeling engine Phyre2. The structure was paired with the hypothetical structure of *D. discoideum* dUTPase to visualize similarities and differences between the structures. As seen in panel A of Figure 4.3, the two

hypothetical structures are similar with both containing the characteristic beta sheet and alpha helix present in many dUTPase structures.

Chlorella virus PBCV-1 has a dUTPase with a published temperature optimum of 50°C (Zhang *et al.,* 2005). The PBCV-1 investigation involved site-directed mutagenesis of serine 81 and arginine 84 of the PBCV-1 sequence (Figure 4.2) to glutamate 81 and threonine 84. The resulting trimer displayed a depressed temperature optimum of 37°C, implying that interactions Ser 81 and Arg 84 participate in are important in the temperature optimum. The corresponding or homologous residues in the *D. discoideum* dUTPase sequence are Ser 115 and Arg 118. It is logical to assume these residues are involved in the same interactions as Ser 81 and Arg 84, which may partially account for the temperature optimum of the PBCV-1 dUTPase. Although information regarding Ser 115 and Arg 118 is important, it does not lend perspective into the difference in temperature optima between the two versions of *D. discoideum* dUTPase. Since both Ser 115 and Arg 118 are located in core and full-length sequences it does not explain the lower temperature optimum displayed by the core version. It should be noted that in the alignment the equivalent residues to the *D. discoideum* Ser 115 and Arg 118 are Pro 97 and Asp 100 in the *T. pacificus* dUTPase. The chemical properties of these two residues differ from the conserved Ser 115 and Arg 118 (seen in *D. discoideum* and PBCV-1). Proline does not contain a free amino group and thus would be a favorable residue for secondary structure turns. Aspartate holds a negative charge in a deprotonated state which is the likely state due to the characteristically low pKa value (~4.0) whereas arginine holds a net positive charge. The different chemical

properties of Ser and Arg may allow for more and closer interactions with other residues such as increased hydrogen bonding or shorter length hydrogen bonding allowing for a more stable molecule. Panel B of Figure 4.4 shows one unique hydrogen bond from Asp 99 to Gly 144 in the *T. pacificus* dUTPase. This candidate hydrogen bond represents one of many potential unique bonds found in the *T. pacificus* dUTPase. Collectively, it is unique hydrogen bonds that could provide additional stability for the tertiary structure, allowing for a greater temperature optimum.

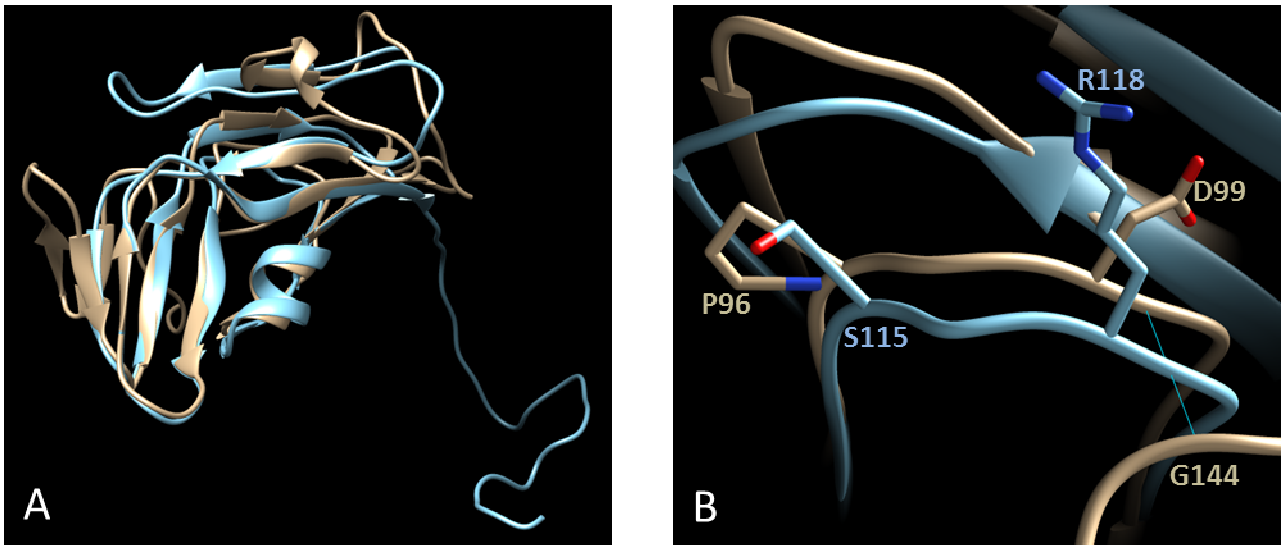

Figure 4.4: Structural comparison of hypothetical dUTPases belonging to *T. pacificus* and *D. discoideum*. Computational modeling of hypothetical dUTPase structures was carried out with the modeling engine Phyre2. The *D. discoideum* structure is presented in blue while the *T. pacificus* is tan. Panel A shows a whole peptide comparison between *D. discoideum* and *T. pacificus* dUTPase monomers. Panel B shows conserved residues Ser 115 and Arg 118 with the positionally equivalent residues Pro 96 and Asp 99 in *T. pacificus* dUTPase.

The pH optima data collected on both versions of dUTPase reveals catalysis can occur throughout a broad pH range. The gradual peak observed for both versions (see Results, Figure 3.9) indicates the residues that comprise the N-terminal leading region

are not specifically functional in buffering of the active site. The active site of enzymes may be sensitive to variation in pH, in which case formation of a highly specific microenvironment may be necessary to preserve activity. Active site residues of motif II are featured in Figure 4.2 and are essential for understanding the pH temperature optima of dUTPase. Specifically, the *D. discoideum* residues of importance are Arg 97 (pKa 12.48), Ser 98 (pKa 9.15), Gly 99, and Gln 143 (pKa 9.13). All of these featured residues (of motif II and IV) and several conserved in motif III (Asp 114 and Asp 116, pKa 3.86) possess extreme pKa values for the side chain. By having an extreme pKa value, the pH of the environment would need to drastically change for deprotonation or protonation to occur. It is probable that the broad pH operating range reported for both versions of dUTPase can be attributed to the high proportion of residues with extreme pKa values. The role of the N-terminus on pH optimum of dUTPase is still uncertain however. It can be assumed that if the N-terminus was responsible for buffering the active site, one would expect the pH optima curves to be vastly different between the two versions. Due to the lack of data on the position of the *D. discoideum* dUTPase N-terminus it cannot be confirmed whether or not there is in fact a buffering effect. The N-terminus is 16.2 % (6/37) composed of residues with charged R groups. The low proportion of residues with ionizable R groups is evidence that the N-terminus may not be functional in buffering active site pH.

It is widely known that dUTPase performs catalysis most effectively when in complex with the metal ion magnesium. Metal dependence trials were conducted to test how dUTPase activity is affected when magnesium is replaced with other metal

cofactors. Experimental data from metal dependence trials display similar conclusions for both versions of dUTPase. Aside from magnesium, all other metals used as a substitute cofactor were detrimental to enzyme activity. The substances with the greatest to least activity were: magnesium, manganese, calcium, and EDTA. The results were expected since residues responsible for coordination of the magnesium ion (motifs I, II, and IV) are present in both the core and full-length versions. Conserved Asp 61 in motif I of many dUTPases (see Introduction, Figure 1.2) is largely responsible for coordinating water molecules which in turn coordinate the magnesium ion required for catalysis (Vértessy and Tóth, 2009). The equivalent aspartate residue is conserved in the *D. discoideum* sequence and is present in the full-length and core sequences at positions 61 and 24 respectively (see Introduction, Figure 1.2). Metal dependence trials conducted on the *S. cerevisiae* dUTPase displayed a similar trend observed in *D. discoideum* dUTPase with the greatest activity exhibited by magnesium followed by manganese, calcium, and EDTA.

Renaturation potential data were acquired to determine how readily dUTPase was able to withstand extreme heat and renature (see Results, Figure 3.11). Presently, there are no published data regarding renaturation of dUTPase following thermal treatment. Data acquired shows the core version was able to renature and regain activity whereas the full-length was unable to do so significantly. It can be concluded that renaturation as robust as the core version relies on the absence of the N-terminus. The absence of an N-terminus allows for renaturation as a favorable process in the core

version whereas the full-length cannot refold apparently due to the presence of the N-terminus.

Overall, the role of the N-terminus of the *D. discoideum* dUTPase appears to be modulation of activity. This is shown in the temperature and pH optima data where the core version dUTPase possesses overall less activity and is less responsive to changes in pH and temperature (see Results, Figures 3.8 and 3.9). The N-terminus also appears to contain residues essential for the enzyme to operate at maximal velocity. The catalytic efficiency of the core is greater than the full-length with a 1.35-fold greater k_{cat}/K_m. Contributing to the increase in catalytic efficiency is the elevated k_{cat} (0.59-fold greater) and lower K_m (0.44-fold lower) of the core dUTPase. As noted in the Materials and Methods (Section 2.6), k_{cat} approximations are not credible since the condition of enzyme concentration was not approximately 10-fold higher than the K_m.

Appendix

5.) Kinetic constants calculations

5.1) Core:

Calculation of concentration of product at time *t* (Equation 1, Materials and Methods Section 2.6.1)

$$[P]_t = \frac{A_0 - A_t}{A_0 - A_\infty}[S]_0$$

$$[P]_t = \frac{(0.5974 - 0.5888)}{(0.5974 - 0.5250)}10\ \mu\text{M}$$

$$[P]_t = 1.19\ \mu\text{M}$$

Calculation of V_{max} (Equation 4, Materials and Methods Section 2.6.1)

$$V = \frac{kln\left(\frac{[S]_0}{[S]_0 - [P]_t}\right) + [P]_t}{t}$$

$$V = \frac{0.22ln\left(\frac{10\ \mu\text{M}}{10\ \mu M - 1.19\ \mu M}\right) + 1.19\ \mu\text{M}}{13\ s}$$

$$V = 0.097\ \mu\text{M}\ s^{-1}$$

Approximation of k_{cat} (Equation 7, Materials and Methods Section 2.6.1)

$$k_{cat} \approx \frac{(K_m * r_2) + (C_e * r_2)}{C_e}$$

$$k_{cat} \approx \frac{(0.22\ \mu\text{M} * 0.1704) + (0.215\ \mu\text{M} * 0.1704)}{0.215\ \mu\text{M}}$$

$$k_{cat} \approx 0.345\ s^{-1}$$

5.2) Full-length:

Calculation of concentration of product at time *t* (Equation 1, Materials and Methods Section 2.6.1)

$$[P]_t = \frac{A_0 - A_t}{A_0 - A_\infty}[S]_0$$

$$[P]_t = \frac{0.7614 - 0.7550}{0.7614 - 0.7319}(10\ \mu M)$$

$$[P]_t = 2.17\ \mu M$$

Calculation of V_{max} (Equation 4, Materials and Methods Section 2.6.1)

$$V = \frac{kln\left(\frac{[S]_0}{[S]_0 - [P]_t}\right) + [P]_t}{t}$$

$$V = \frac{0.5ln\left(\frac{10\ \mu M}{10\ \mu M - 2.17\ \mu M}\right) + 2.17\ \mu M}{15\ s}$$

$$V = 0.153\ \mu M\ s^{-1}$$

Approximation of k_{cat} (Equation 7, Materials and Methods Section 2.6.1)

$$k_{cat} \approx \frac{(K_m * r_2) + (C_e * r_2)}{C_e}$$

$$k_{cat} \approx \frac{(0.5\ \mu M * 0.0142) + (0.017\ \mu M * 0.0142)}{0.017\ \mu M}$$

$$k_{cat} \approx 0.3444\ s^{-1}$$

Calculations used to determine V_{max} and k_{cat} were carried out using Equations 1, 4, and 7 presented in the Materials and Methods (Section 2.6.1). The order of the calculations is presented specifically since the product of Equation 1 is required to solve for V_{max} (Equation 4). Within Equation 1, the variables A_0, A_∞, and A_t stand for initial, final, and absorbance at time *t* which were obtained through analysis of specific kinetic traces. k_{cat} calculations were carried out using Equation 7 and a calculated slope constant (*K*) produced from an exponential best fit line applied to specific kinetic traces. The above calculations represent just one calculation for each equation per version of dUTPase. Additional k_{cat} calculations and corresponding data can be seen below.

5.3) k_{cat} Calculations (Core and full-length):

	***K* values (s^{-1})**	**k_{cat} (μM s^{-1})**	**Average k_{cat}**	**Standard deviation**	**[Enzyme](μM)**
Core	0.1704	0.3448	0.2335	0.0788	0.2150
	0.1701	0.3442			
	0.0935	0.1892			
	0.0935	0.2157			
	0.1066	0.1497			
	0.0740	0.2112			
	0.0890	0.1801			
Full-length	0.0142	0.4318	0.3741	0.0817	0.0170
	0.0104	0.3163			

Table 5.1: k_{cat} values of full-length and core dUTPase. The data presented above represents the respective slope constants (*K* values) from kinetic traces of the full-length and core dUTPase.

Accurate determination of the k_{cat} requires enzyme concentration to exceed the K_m. Since these requirements were understood after kinetic data acquisition, data from

existing traces was utilized for k_{cat} approximation. The core data that came closest to meeting the requirement of [Enzyme]>>K_m were trials that utilized 10 µg protein (0.2150 µM) and 10 µM dUTP. Full-length data used for k_{cat} approximation were traces that utilized 1 µg protein (0.0170 µM) and 10 µM dUTP.

5.4) dUTPase activity from cresol red assays (Materials and Methods Section 3.3)

Full-length dUTPase activity

[dUTP] (µM)	Average ΔA573nm *10,000	Standard Deviation
0.1	1.00	0
0.2	1.10	0.32
0.4	1.63	0.52
0.6	2.40	0.50
0.8	2.67	0.49
1.0	2.56	0.51
2.0	3.39	0.61
4.0	3.61	0.70
10	4.00	0

Core dUTPase activity

[dUTP] (µM)	Average ΔA573nm *10,000	Standard Deviation
0.1	1.67	0.50
0.2	2.56	0.53
0.4	3.6	0.52
0.6	5.33	0.50
0.8	5.22	0.44
1.0	5.5	0
2.0	5.75	0.46
4.0	5.67	0.50
5.0	5.33	0.52
10.0	6.2	0.45

Table 5.2: Data used to calculate K_m and V_{max} of full-length and core dUTPase. Average slope values were calculated by averaging the linear portion of ten replicates per substrate concentration. Standard deviation was calculated for each substrate concentration. Table A represents the core kinetics data set. Table B represents the full-length data set generated by Kathleen Williams and Catherine Chia.

5.5.) Metal Dependence- supplemental data

(Results Section 3.4.3)

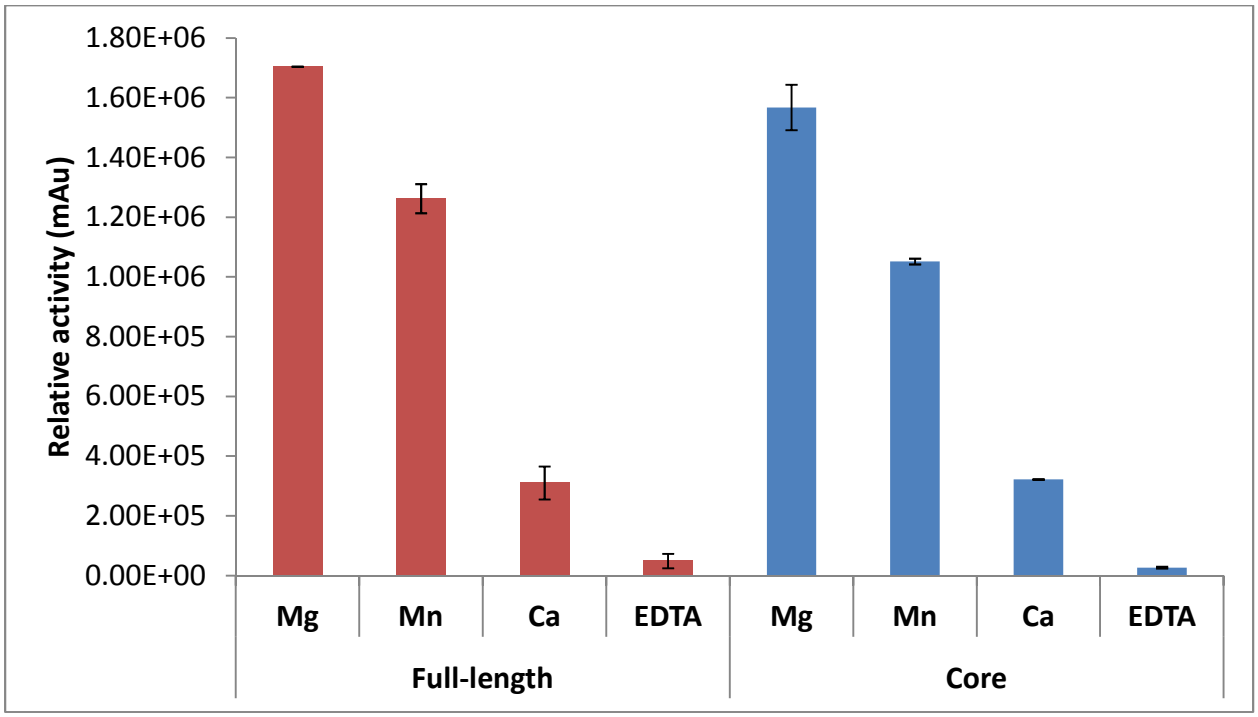

Figure 5.3: Supplemental metal dependence data. The supplemental trials presented for the full-length and core version were generated from trials carried out identically to the presented trials in the results section (Figure 3.9).

	Metal or EDTA (2 mM)			
dUTPase	**$MgCl_2$**	**$MnCl_2$**	**$CaCl_2$**	**EDTA**
Core	**100.0±4.0%**	**83.5±11%**	**43.2±27%**	**0.3±25%**
Full-length	**100.0±0.2%**	**74.1±3.9%**	**18.24±17.8%**	**2.27±48.8%**

Table 5.3: Metal dependence of core and full-length dUTPase. Percentages represent percent activity relative to $MgCl_2$ which was assumed to have optimal activity (Vértessy and Tóth, 2009). Data presented is from Figure 5.3.

5.6) Supplemental BLAST data

Organism	Identity (%)	Query cover (%)	E value	Total score	Query	Sbjct	Longer NT	Shorter NT	Kingdom	Common name
D. discoideum	100	100	1.00E-128	370	1	1	-	-	Amoebozoa	-
D. purpureum	77	84	5.00E-77	238	27	1		26	Amoebozoa	-
O. lucimarinus	72	77	6.00E-66	210	38	6	-	32	Viridiplantae	Algae
V. carteri	71	78	7.00E-66	211	36	62	26	-	Viridiplantae	Algae
T. cacao	72	77	5.00E-65	209	38	62	24	-	Viridiplantae	Cacao tree
E. salsugineum	72	77	6.00E-65	208	38	37	-	1	Viridiplantae	-
M. notabilis	72	77	9.00E-65	207	38	36	-	2	Animalia	Moth
C. clementina	73	77	1.00E-64	209	38	82	44	-	Viridiplantae	Clementine
P. vulgaris	67	85	2.00E-64	209	23	73	-	50	Bacteria	-
S. tuberosum	71	77	2.00E-64	206	38	13	25	-	Viridiplantae	Potato
A. thaliana	66	88	3.00E-64	206	17	10	-	7	Viridiplantae	-
V. vinifera	63	85	3.00E-64	206	23	24	1	-	Viridiplantae	Common Grape
C. sinensis	73	77	4.00E-64	207	38	82	44	-	Animalia	Liver Fluke
O. tauri	70	77	5.00E-64	205	38	5	-	33	Viridiplantae	-
S. moellendorffii	69	79	1.00E-63	204	37	2	-	35	Viridiplantae	-
G. aurea	69	77	2.00E-63	203	38	3	-	35	Viridiplantae	-
S. lycopersicum	70	79	2.00E-63	204	34	28	-	6	Viridiplantae	Tomato
P. sitchensis	72	77	3.00E-63	204	38	49	11	-	Viridiplantae	Sitka Spruce
P. trichocarpa	71	77	3.00E-63	203	38	33	-	5	Viridiplantae	Black Cottonwood
P. persica	70	77	4.00E-63	203	38	37	-	5	Viridiplantae	Peach
S. italica	70	82	5.00E-63	204	29	66	37	1	Viridiplantae	-
C. thermophilum	70	75	6.00E-63	204	40	58	18	-	Fungi	-
C. arietinum	71	77	7.00E-63	202	38	13	-	25	Viridiplantae	-
A. lyrata	72	77	8.00E-63	203	38	41	3	-	Viridiplantae	-
P. patens	68	79	8.00E-63	202	36	21	-	15	Viridiplantae	-
C. reinhardtii	68	82	1.00E-62	201	29	9	-	20	Viridiplantae	-
O. brachyantha	58	96	2.00E-62	203	3	43	40	-	Viridiplantae	-
C. rubella	70	77	2.00E-62	201	38	39	1	-	Viridiplantae	-
M. guttatus	70	77	2.00E-62	201	38	38	0	0	Viridiplantae	Spotted Monkey Flower
S. bicolor	72	77	3.00E-62	201	38	37	-	1	Viridiplantae	-
Z. mays	71	77	2.00E-61	201	38	85	47	-	Viridiplantae	Corn
C. biroi	64	85	3.00E-61	198	24	2	-	22	Metazoa	-
C. higginsianum	61	90	3.00E-61	201	16	23	7	-	Fungi	-
P. infestans	69	78	3.00E-61	199	38	12	-	26	Stramenopiles	-
C. subellipsoidea	67	78	3.00E-61	199	35	37	2	-	Viridiplantae	-
C. gloeosporioides	60	91	5.00E-61	198	16	18	2	-	Fungi	-
C. sativus	70	77	6.00E-61	199	38	86	48	-	Viridiplantae	-
C. graminicola	61	91	7.00E-61	198	16	18	2	-	Fungi	-
S. diclina	67	78	1.00E-60	198	36	4	-	32	Stramenopiles	-

Table 5.4: BLAST data results. The top fifty hits (based on overall relatedness) are featured. The column 'Identity' displays the percentage of the residues that are identical between the hit and the query sequence. 'Query' and 'Subject' refer to the first residue of the alignment for both the query sequence and the subject sequence. 'Longer NT' and 'Shorter NT' refer to the number of N-termini residues that are either longer or shorter than the *D. discoideum* dUTPase sequence.

5.7) Extended migration distance plot

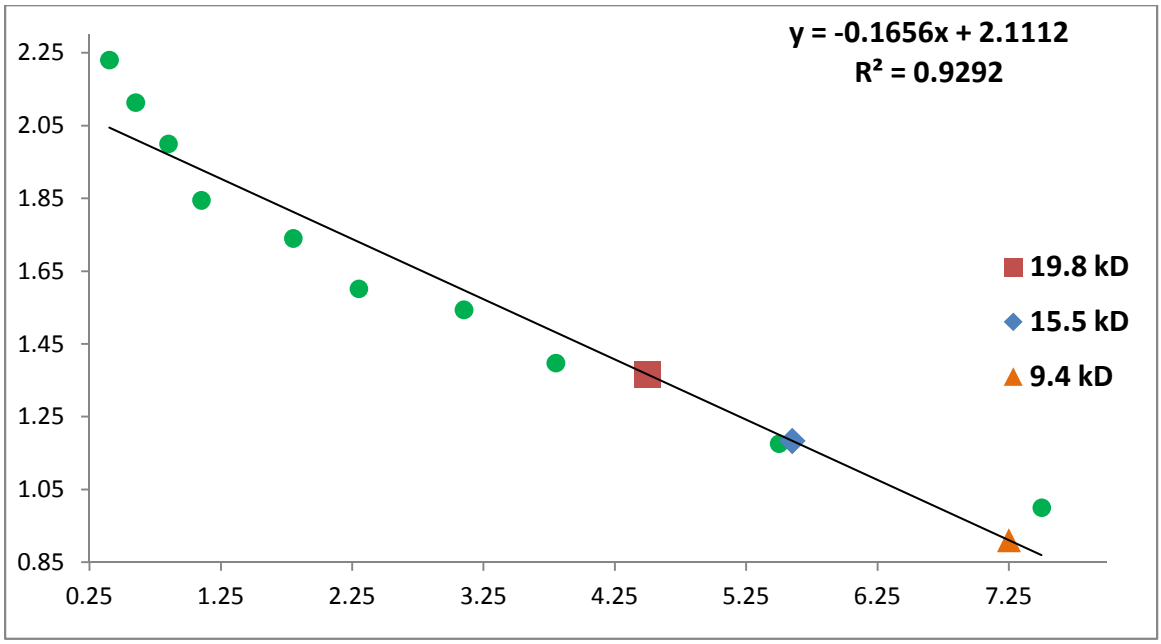

Figure 5.4: Size calculation by plotting migration distance of protein ladder. Vertical axis represents the log value of protein standards. Horizontal axis represents migration distance observed by each protein ladder constituent. The log was calculated for the molecular weights of the protein ladder components and plotted versus the individual migration distances. The log of the molecular mass was calculated using the migration distances of thrombin-cleaved recombinant dUTPases and the best fit line equation. With a log value of 1.184, the predicted 15.5 kD (core subunit) measures to 15.3 kD. The 19.8 kD (full-length subunit) with a calculated log value of 1.366, results in a predicted size of 23.2 kD. The 9.4 kD version (truncated subunit) measurements produced a log value of 0.9106, predicting the size to be 8.1 kD. Gel measurements were consistent among multiple gel samples. Inclusion of additional high molecular weight marker measurements was used to demonstrate the decrease in credibility.

Works Cited

Bajaj, M., & Moriyama, H. (2007). Purification, crystallization and preliminary crystallographic analysis of deoxyuridine triphosphate nucleotidohydrolase from *Arabidopsis thaliana*. *Acta Crystallographica.Section F, Structural Biology and Crystallization Communications, 63*(Pt 5), 409-411. doi:10.1107/S1744309107016004

Baldo, A. M., & McClure, M. A. (1999). Evolution and horizontal transfer of dUTPase-encoding genes in viruses and their hosts. *Journal of Virology, 73*(9), 7710-7721.

Bernier-Villamor, V., Camacho, A., Hidalgo-Zarco, F., Perez, J., Ruiz-Perez, L. M., & Gonzalez-Pacanowska, D. (2002). Characterization of deoxyuridine 5'-triphosphate nucleotidohydrolase from *Trypanosoma cruzi*. *FEBS Letters, 526*(1-3), 147-150.

Castillo-Acosta, V. M., Aguilar-Pereyra, F., Bart, J. M., Navarro, M., Ruiz-Perez, L. M., Vidal, A. E., & Gonzalez-Pacanowska, D. (2012). Increased uracil insertion in DNA is cytotoxic and increases the frequency of mutation, double strand break formation and VSG switching in *Trypanosoma brucei*. *DNA Repair, 11*(12), 986-995. doi:10.1016/j.dnarep.2012.09.007; 10.1016/j.dnarep.2012.09.007

Castillo-Acosta, V. M., Aguilar-Pereyra, F., Garcia-Caballero, D., Vidal, A. E., Ruiz-Perez, L. M., & Gonzalez-Pacanowska, D. (2013). Pyrimidine requirements in deoxyuridine triphosphate nucleotidohydrolase deficient *Trypanosoma brucei* mutants.

Molecular and Biochemical Parasitology, 187(1), 9-13. doi:10.1016/j.molbiopara.2012.11.003; 10.1016/j.molbiopara.2012.11.003

Chaiseeda, K. (2009). Structure and kinetic studies on deoxyuridine triphosphatase from *Arabidopsis thaliana.* (Unpublished M.S.). University of Nebraska- Lincoln,

Cho, S. S., Sun, Y., Yu, M., Kwon, S. H., & Kwon, S. T. (2012). Characterization and PCR applications of dUTPase from the hyperthermophilic euryarchaeon *Thermococcus pacificus*. *Enzyme and Microbial Technology, 51*(6-7), 342-347. doi:10.1016/j.enzmictec.2012.08.001; 10.1016/j.enzmictec.2012.08.001

Cohen, S. S. (1971). On the nature of thymineless death. *Annals of the New York Academy of Sciences, 186*, 292-301.

Cooper, G. M. (2000). Chapter 1. An overview of cells and cell research. *The cell: A molecular approach* (2nd Ed.). Sunderland, Massachusetts: Sinauer Associates.

De Bernardez Clark, E., Schwarz, E., & Rudolph, R. (1999). Inhibition of aggregation side reactions during in vitro protein folding. *Methods in Enzymology, 309*, 217-236.

Gadsden, M. H., McIntosh, E. M., Game, J. C., Wilson, P. J., & Haynes, R. H. (1993). dUTP pyrophosphatase is an essential enzyme in *Saccharomyces cerevisiae*. *The EMBO Journal, 12*(11), 4425-4431.

Hochhauser, S. J., & Weiss, B. (1978). *Escherichia coli* mutants deficient in deoxyuridine triphosphatase. *Journal of Bacteriology, 134*(1), 157-166.

Homma, K., & Moriyama, H. (2009). Crystallization and crystal-packing studies of Chlorella virus deoxyuridine triphosphatase. *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications, 65*(Pt 10), 1030-1034. doi: 10.1107/S1744309109034459; 10.1107/S1744309109034459

Kornberg, A. (2005). Uracil incorporation in DNA. *DNA replication* (Second ed., pp. 500) University Science Books.

Ladner, R. D., & Caradonna, S. J. (1997). The human dUTPase gene encodes both nuclear and mitochondrial isoforms. Differential expression of the isoforms and characterization of a cDNA encoding the mitochondrial species. *The Journal of Biological Chemistry, 272*(30), 19072-19080.

Ladner, R. D., McNulty, D. E., Carr, S. A., Roberts, G. D., & Caradonna, S. J. (1996). Characterization of distinct nuclear and mitochondrial forms of human deoxyuridine triphosphate nucleotidohydrolase. *The Journal of Biological Chemistry, 271*(13), 7745-7751.

Larsson, G., Nyman, P. O., & Kvassman, J. O. (1996). Kinetic characterization of dUTPase from *Escherichia coli*. *The Journal of Biological Chemistry, 271*(39), 24010-24016.

Ma, B., Elkayam, T., Wolfson, H., & Nussinov, R. (2003). Protein-protein interactions: Structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proceedings of the National Academy of Sciences of the United States of America, 100*(10), 5772-5777. doi:10.1073/pnas.1030237100

Mahagaokar, S., Rao, P. N., & Orengo, A. (1980). Deoxyuridine triphosphate nucleotidohydrolase of HeLa cells. *The International Journal of Biochemistry, 11*(5), 415-421.

McGeoch, D. J. (1990). Protein sequence comparisons show that the 'pseudoproteases' encoded by poxviruses and certain retroviruses belong to the deoxyuridine triphosphatase family. *Nucleic Acids Research, 18*(14), 4105-4110.

McIntosh, E. M., & Haynes, R. H. (1997). dUTP pyrophosphatase as a potential target for chemotherapeutic drug development. *Acta Biochimica Polonica, 44*(2), 159-171.

Miyakoshi, H., Miyahara, S., Yokogawa, T., Endoh, K., Muto, T., Yano, W., Wakasa, T., Ueno, H., Tee Chong, K., Taguchi, J., Nomura, M., Takao, Y., Fujioka, A., Hashimoto, A., Itou, K., Yamamura, K., Shuto, S., Nagasawa, H., Fukuoka, M. (2012). 1,2,3-triazole-containing uracil derivatives with excellent pharmacokinetics as a novel class of potent human deoxyuridine triphosphatase inhibitors. *Journal of Medicinal Chemistry, 55*(14), 6427-6437. doi:10.1021/jm3004174; 10.1021/jm3004174

Mustafi, D., Bekesi, A., Vertessy, B. G., & Makinen, M. W. (2003). Catalytic and structural role of the metal ion in dUTP pyrophosphatase. *Proceedings of the National Academy of Sciences of the United States of America, 100*(10), 5670-5675. doi:10.1073/pnas.1031504100

Nguyen, P. (2011). Truncated dUTPase of *Dictyostelium discoideum* is active and likely to form trimers. (Unpublished M.S.). University of Nebraska- Lincoln,

Pardo, E. G., & Gutierrez, C. (1990). Cell cycle- and differentiation stage-dependent variation of dUTPase activity in higher plant cells. *Experimental Cell Research, 186*(1), 90-98.

Persson, R., Cedergren-Zeppezauer, E. S., & Wilson, K. S. (2001). Homotrimeric dUTPases; structural solutions for specific recognition and hydrolysis of dUTP. *Current Protein & Peptide Science, 2*(4), 287-300.

Pri-Hadash, A., Hareven, D., & Lifschitz, E. (1992). A meristem-related gene from tomato encodes a dUTPase: Analysis of expression in vegetative and floral meristems. *The Plant Cell, 4*(2), 149-159. doi:10.1105/tpc.4.2.149

SEM of Dictyostelium developmental stages. Copyright M.J. Grimson & R.L. Blanton; Biological Sciences Electron Microscopy Laboratory, Texas Tech University

Tchigvintsev, A., Singer, A. U., Flick, R., Petit, P., Brown, G., Evdokimova, E., Savchenko, A., Yakunin, A. F. (2011). Structure and activity of the *Saccharomyces cerevisiae* dUTP pyrophosphatase DUT1, an essential housekeeping enzyme. *The Biochemical Journal, 437*(2), 243-253. doi:10.1042/BJ20110304; 10.1042/BJ20110304

Tyler, M. S. (2000). *Developmental biology: A guide for experimental study* (2nd Ed.). Sunderland, Massachusetts: Sinauer Associates.

Vertessy, B. G., & Toth, J. (2009). Keeping uracil out of DNA: Physiological role, structure and catalytic mechanism of dUTPases. *Accounts of Chemical Research, 42*(1), 97-106. doi:10.1021/ar800114w; 10.1021/ar800114w

Zhang, Y., Moriyama, H., Homma, K., & Van Etten, J. L. (2005). Chlorella virus-encoded deoxyuridine triphosphatases exhibit different temperature optima. *Journal of Virology, 79*(15), 9945-9953. doi:10.1128/JVI.79.15.9945-9953.2005