5-1-2008

YZGD pyridoxal phosphatase from P. thiaminolyticus; subcloning, expression, and purification for x-ray crystallography structure determination

Amanda M. Strassner

YZGD pyridoxal phosphatase from *P. thiaminolyticus*; subcloning, expression, and purification for x-ray crystallography structure determination

Approved: _____
Thesis Advisor

_____
Director of Bioinformatics

Submitted in partial fulfillment of the requirements for the Master of Science degree in Bioinformatics at Rochester Institute of Technology

Amanda M. Strassner

May 2008

Thesis/Dissertation Author Permission Statement

Title of thesis or dissertation: **YZGD pyridoxal phosphatase from *P. thiaminolyticus*; subcloning, expression, and purification for x-ray crystallography structure determination**

Name of author: **Amanda M. Strassner**
Degree: **Master of Science**
Program: **Bioinformatics**
College: **Science**

I understand that I must submit a print copy of my thesis or dissertation to the RIT Archives, per current RIT guidelines for the completion of my degree. I hereby grant to the Rochester Institute of Technology and its agents the non-exclusive license to archive and make accessible my thesis or dissertation in whole or in part in all forms of media in perpetuity. I retain all other ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Print Reproduction Permission Granted:

I, **Amanda M. Strassner**, hereby grant permission to the Rochester Institute of Technology to reproduce my print thesis or dissertation in whole or in part. Any reproduction will not be for commercial use or profit.

Signature of Author: _____ Date: _____

Inclusion in the RIT Digital Media Library Electronic Thesis & Dissertation (ETD) Archive:

I, **Amanda M. Strassner**, additionally grant to the Rochester Institute of Technology Digital Media Library (RIT DML) the non-exclusive license to archive and provide electronic access to my thesis or dissertation in whole or in part in all forms of media in perpetuity.

I understand that my work, in addition to its bibliographic record and abstract, will be available to the world-wide community of scholars and researchers through the RIT DML. I retain all other ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation. I am aware that the Rochester Institute of Technology does not require registration of copyright for ETDs.

I hereby certify that, if appropriate, I have obtained and attached written permission statements from the owners of each third party copyrighted matter to be included in my thesis or dissertation. I certify that the version I submitted is the same as that approved by my committee.

Signature of Author: _____ Date: _____

Table of Contents

Thesis Committee

Thesis advisor:
Dr. Suzanne O'Handley
Department of Chemistry
Rochester Institute of Technology

Committee Members:
Dr. Thomas Kim
Department of Chemistry
Rochester Institute of Technology

Dr. Michael Savka
Department of Biological Sciences
Rochester Institute of Technology

Abstract

The HAD superfamily contains enzymes that catalyze carbon or phosphate transfer reactions. One family within the HAD superfamily is the p-nitrophenyl phosphatase (PNPPase) family. Members of this family are all phosphatases that cleave the substrate analog p-nitrophenyl phosphate, in addition to their biologically significant substrate. Interestingly, among the small number of family members that have been characterized, there is a wide variety of biologically relevant activities. YZGD from *Paenibacillus thiaminolyticus* has pyridoxal phosphatase activity, catalyzed by a HAD superfamily domain, as well as a CDP-alcohol/sugar nucleotide hydrolase activity, catalyzed by a Nudix hydrolase domain. Determining the structure of this enzyme would result in valuable information to contribute to our knowledge and understanding of the HAD superfamily and the Nudix hydrolase superfamily. Sufficient quantities of sufficiently pure YZGD are necessary to perform x-ray crystal structure determination. A histidine tag was added to the enzyme so that it can be purified by nickel affinity chromatography. This was done by cloning the gene into pET19b, a plasmid containing a sequence encoding a string of ~10 histidines, and expressing the protein in *Escherichia coli BLR(DE3)*. The tagged YZGD expresses well, is soluble, and retains full activity. It appears to be completely purified by nickel affinity chromatography. This will now allow YZGD to be purified in large quantities, crystallized, and its structure determined.

Introduction

The HAD superfamily, named for haloacid dehalogenase (the first enzyme discovered for the superfamily), contains a remarkable variety of enzymes. Members are found in all three domains of life, and include dehalogenases, sugar phosphomutases, phosphonatases, ATPases, and phosphoesterases. There are 33 major enzyme families within the HAD superfamily, the majority of which are phosphoesterase families. At least five of the HAD superfamily proteins exist in all three domains of life, indicating that they were present in the last universal common ancestor of all life on earth (Burroughs, et al. 2006). HAD superfamily enzymes share a version of the Rossmanoid α/β fold, which has alternating β-α units that form a three-layered α/β "sandwich." One unique feature of the HAD superfamily fold is that it contains two sites in which additional domains, or "caps," can be inserted. These caps distinguish between different families within the HAD superfamily, and often provide important structural or functional residues. Members of the superfamily can contain caps in both sites, a cap in only one site, or no caps at all. (Figure 1). (Burroughs et al. 2006)

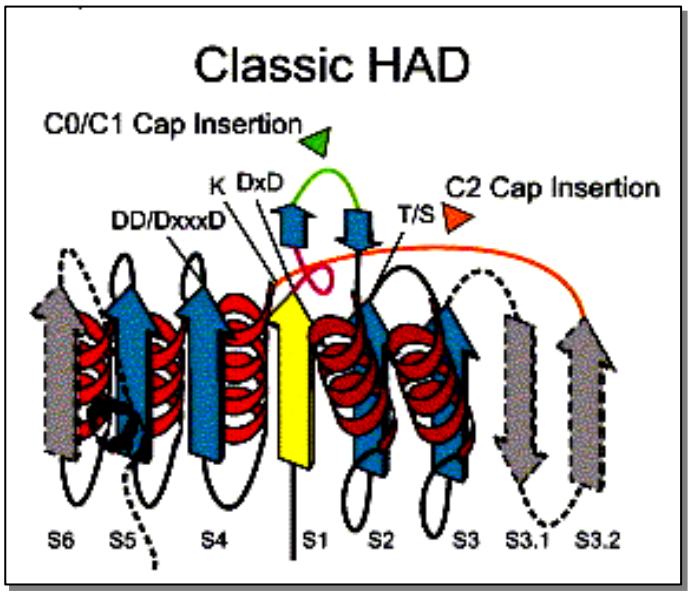

Figure 1: The classic HAD motif, showing the Rossmanoid fold and the positions in which caps can be inserted. (Burroughs et al. 2006)

The active site of the HAD superfamily members is surrounded by four loops. All superfamily members use an aspartate residue, located on loop 1, to catalyze a carbon or phosphoryl group transfer. The phosphotransferases within the superfamily all require a Mg^{2+} cofactor, which is coordinated by two carboxylate residues from loop 4. The phosphoryl group of the substrate interacts with either a serine or threonine on loop 2, and a lysine or arginine on loop 3. (Allen and Dunaway-Mariano 2004)

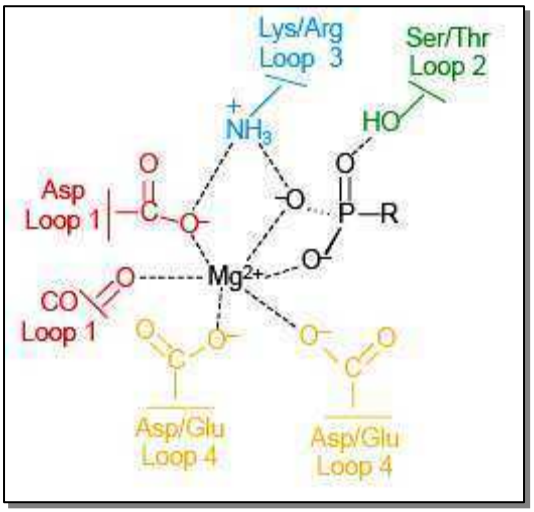

Figure 2: The important active site residues of HAD superfamily phosphotransferases (Allen and Dunaway-Mariano 2004).

One interesting family found within the HAD superfamily is the p-nitrophenyl phosphatase family (PNPPase) (Koonin & Tatusov 1994). The first members identified were enzymes from two species of yeast, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* (Attias & Bonnet 1972; Yang et al. 1991). These two enzymes were found to hydrolyze p-nitrophenyl phosphate, which is a substrate analog not found in biological systems, but thus far, no biologically relevant substrate has been found for these two enzymes. A large number of putative proteins have a high level of sequence similarity to the known family members, but only a few of these proteins have been characterized. Despite the small number of enzymes that have

been characterized, a variety of activities have been found within this family (Figure 2). NAGD from *Escherichia coli* is a family member that hydrolyzes uridine 5'-monophosphate (UMP) (Tremblay et al. 2006; Tirrell, I.M., Nguyen, A.T., Mentz, R.J., Slivka, E.J. and O'Handley, S.F. unpublished data). There are homologues of NAGD in other organisms as well, including *Shigella*, *Salmonella*, and *Yersinia pestis* (Figure 3). There are characterized pyridoxal phosphatases (PLPases) in the family: YZGD PLPase from *Paenibacillus thiaminolyticus*, which also contains a Nudix hydrolase domain (Tirrell et al. 2006), and PLPases found in humans and mice (Jang et al. 2007). Phosphoglycolate phosphatases (PGPases) are also found in this family; the two that have been characterized are from *Chlamydomonas reinhardtii* (a green algae) (Mamedov et al. 2001) and *Staphylococcus aureus* (Hill, J.K., Thomson, J.G., Strassner, A.M., and O'Handley, S.F., *in preparation*). The PGPase from *S. aureus* has been shown to be a virulence factor (Begun et al. 2005), indicating its significance in the cell.

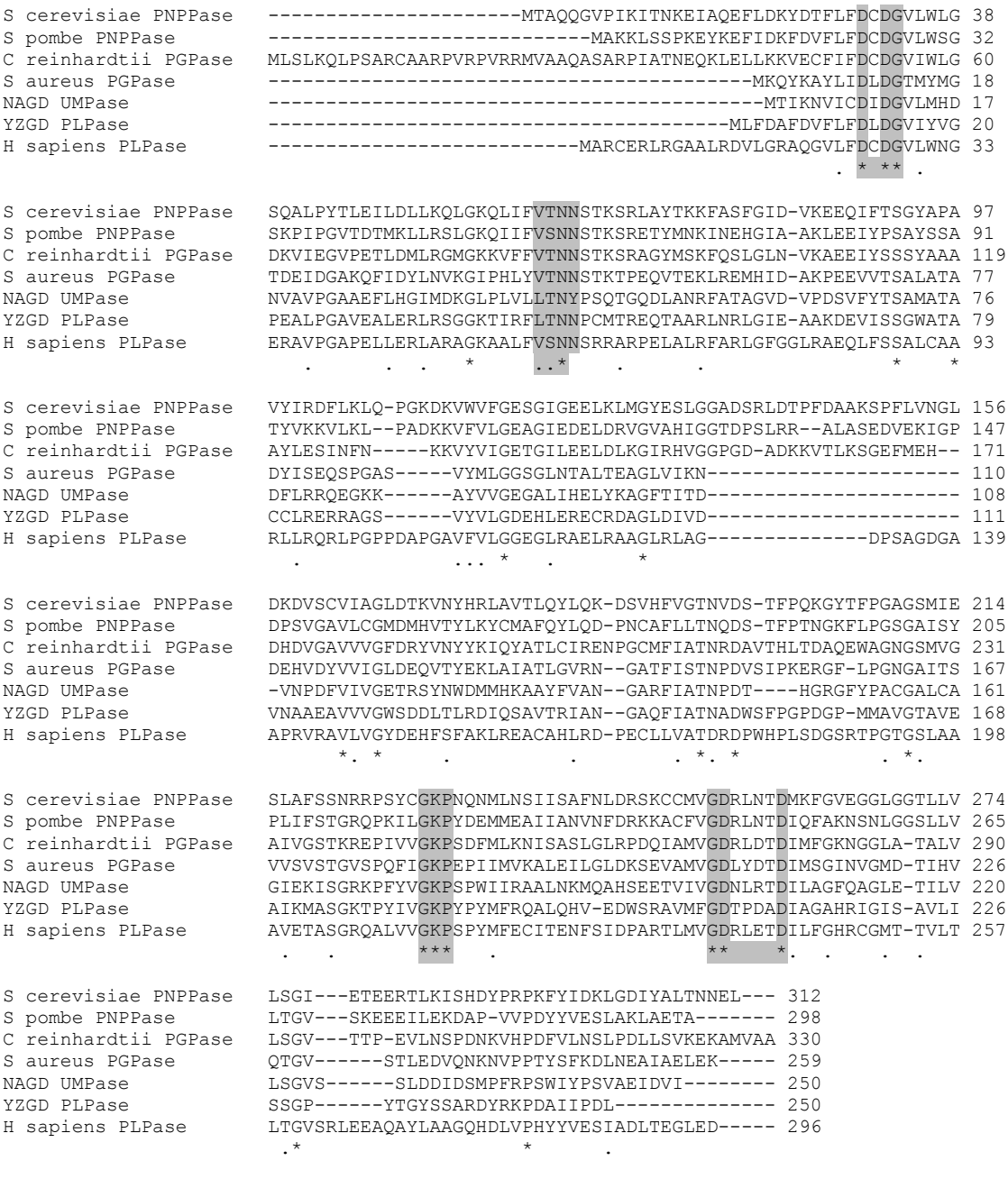

Figure 2: Multiple sequence alignment of known enzymes of the PNPPase family. Amino acids that are conserved are marked by characters below the sequences ('*' represents identical amino acids, and '.' represents similar amino acids). Residues conserved within all HAD superfamily members are highlighted in gray. The enzyme names are indicated to the left. YZGD has been truncated to show only its HAD superfamily domain (the first 250 amino acids), omitting the domain containing the Nudix signature sequence for clarity.

```
E_coli          MTIKNVICDIDGVLMHDNVAVPGAAEFLHGIMDKGLPLVLLTNYPSQTGQDLANRFATAGVDV
Shigella        MTIKNVICDIDGVLMHDNVAVPGAAEFLHGIMDKGLPLVLLTNYPSQTGQDLANRFATAGVDV
Salmonella      MTIKNVICDIDGVLMHDNVAVPGAAEFLTGILEKGLPLVLLTNYPSQTGQDLANRFATAGVNV
Y_pestis        MTIKSVICDIDGVLLHDNTAIKGANDFLARIQDAGMPLVILTNYPSQTAQDLTNRFITAGLDV
Shewanella      --MKNIICDIDGVLLHDNKLIPGSDKFIHRVLDQGNPLVILTNYPVQTGKDLQNRLSAAGIDV
                **.* .********.***  . *.  *.  . . * ***.***** **. ** **   **. *

E_coli          PDSVFYTSAMATADFLRRQEGKKAYVVGEGALIHELYKAGFTITDVNPDFVIVGETRSYNWDM
Shigella        PDSVFYTSAMATADFLRRQEGKKAYVVGEGALIHELYKAGFTITDVNPDFVIVGETRSYNWDM
Salmonella      PDSVFYTSAMATADFLRRQEGKKAYVVGEGALIHELYKAGFTITDVNPDFVIVGETRSYNWDM
Y_pestis        PESAFYTSAMATADFLRRQDGKKAYVIGEGALVHELYKAGFTITDINPDFVIVGETRSYNWDM
Shewanella      PEECFYTSAMATADFLKHQEGSKAFVIGEGALTHELYKAGFTITDINPDFVIVGETRSYNWDM
                *.  ************..*.* **.*.***** ************.*****************

E_coli          MHKAAYFVANGARFIATNPDTHGRGFYPACGALCAGIEKISGRKPFYVGKPSPWIIRAALNK
Shigella        MHKAAYFVANGARFIATNPDTHGRGFYPACGALCAGIEKISGRKPFYVGKPSPWIIRAALNK
Salmonella      MHKAAFFVANGARFIATNPDTHGRGFYPACGALCAGIEKISGRKPFYVGKPSPWIIRAALNK
Y_pestis        MHKAAYFVANGARFIATNPDSHGHGFAPACGALCAPIEKISGRKPFYVGKPSPWIIRAALNK
Shewanella      IHKAAGFVARGARFIATNPDTHGPAYSPACGALCSPIERITGKKPFYVGKPSSWIIRSALNH
                .**** *** **********.** .. *******. **.*.*.********* ****.***.

E_coli          MQAHSEETVIVGDNLRTDILAGFQAGLETILVLSGVSSLDDIDSMPFRPSWIYPSVAEIDVI
Shigella        MQAHSEETVIVGDNLRTDILAGFQAGLETILVLSGVSSLDDIDSMPFRPSWIYPSVAEIDVI
Salmonella      MQAHSEETVIVGDNLRTDILAGFQAGLETILVLSGVSTINDIDSMPFRPSWIYPSVAEIDVI
Y_pestis        MQAHSESTVIVGDNLRTDILAGFQAGLETILVLSGVSTLTDIDAMPFRPSYVYPSVADIDII
Shewanella      INGHSENTVIIGDNMRTDILAGFQAGLETILVTSGVSKLEDIDKEPFRPNHVFACAGDIDVV
                ...*** ***.***.***************** **** * ***  ****  .. ....**..
```

Figure 3: Alignment of homologues of NAGD in pathogenic bacteria. Identical amino acids are indicated by a '*' and similar amino acids are indicated by a '.'.

YZGD is a particularly interesting member of the PNPPase family because, in addition to YZGD's HAD domain responsible for its PLPase activity, it has a second domain containing the signature sequence of the Nudix hydrolase superfamily responsible for its activity on CDP-alcohols, ADP-sugars, and to a limited extent GDP-sugars and UDP-sugars (Tirrell et al. 2006). The Nudix hydrolase superfamily contains diverse enzymes that hydrolyze nucleotide derivatives and are found in a variety of organisms (McLennan 2006). Therefore, the crystal structure of YZGD would be interesting to those who study the various Nudix hydrolases, in addition to those interested in the HAD superfamily.

In early 2006, the structure of the human PLPase family member was determined (PDB ID: 2CFT). The structure of YZGD could be compared to this structure to seek out similarities in structure that lead to specific PLPase activity. It would be the first structure of a prokaryotic PLPase in the HAD superfamily, as well as the first structure of an enzyme with dual phosphatase and Nudix hydrolase activity.

The *yzgd* gene has previously been subcloned into a pET plasmid and used to transform *E. coli HMS174(DE3)*. The YZGD protein has been shown to have good expression levels and solubility. Purification included sonication to break open the cells, streptomycin to remove the nucleic acids, ammonium sulfate fractionation, size exclusion chromatography, and anion exchange chromatography (Tirrell et al. 2006).

The goal of this project was to work out conditions that would increase the amount and purity of YZGD to be used in x-ray crystallography, since a high yield of pure protein is necessary for x-ray crystal structure determination. It was proposed to scale up the expression and purification to obtain milligram quantities of protein to work with. Expression was originally carried out at 37°C, however the expression temperature was lowered to 15°C to

increase both expression and solubility; this technique has been successful for a number of other proteins in the lab. Purification was modified, from using gravity chromatography to "FPLC" columns on the HPLC, for higher resolution in protein purification. The success of each purification step was measured using SDS-PAGE to evaluate purity, Bradford assays to quantify protein, and enzyme assays to determine specific activity. However, scaling up expression, lowering expression temperature, and modifying the chromatography procedures still did not yield pure enough YZGD, thus we created a His•tagged protein and carried out affinity chromatography purification. Here we show that we have successfully created a subclone that expresses soluble His•tagged YZGD, and that we can purify it to an essentially pure state via affinity chromatography.

Materials and Methods

YZGD expression and purification

Cell culture and expression

The YZGD protein was expressed in *Escherichia coli HMS174(DE3)* cells (Novagen) containing the pETYZGD plasmid. The cells were grown in Luria-Bertani media with 100μg/ml ampicillin at 37°C until they reached an A600 of 0.6. The cultures were induced to express the protein with 1mM isopropyl β-D-thiogalactoside. Expression was conducted under two conditions, to determine which would produce the greatest yield of soluble protein. In the first condition, the culture was incubated at 37°C, induced, then incubated for 3 more hours at 37°C, after which the cells were harvested. Alternatively, the culture was cooled from 37°C to 15°C, induced, and incubated for 24 hours at 15°C before harvesting.

Cell harvesting and sonication

Cells were harvested by centrifugation, washed with buffered saline, centrifuged again and stored at -80°C. Prior to sonication, cells were thawed and resuspended in two volumes of TED buffer (50mM Tris-HCl (pH 7.5), 1mM EDTA, 0.1mM DTT). Cells were broken open by sonication using a Branson sonifier, and the resulting solution was centrifuged to remove the insoluble proteins and cell debris from the soluble protein in the supernatant.

Streptomycin fractionation

Streptomycin sulfate was added to the sonicated supernatant to a concentration of 1% (w/v) and incubated, with stirring, for 15 minutes on ice. This was centrifuged to remove the precipitant containing nucleic acids.

Ammonium sulfate fractionation and concentration

Saturated ammonium sulfate was added to the resulting supernatant to a concentration of 55%, and incubated on ice for 15 minutes. The precipitate, containing YZGD, was obtained by centrifugation and resuspended in TED buffer.

Size exclusion chromatography

Size exclusion chromatography was conducted at room temperature, using the GE Healthcare HiPrep™ Sephacryl™ S-200 16/60 column, which has a volume of 120 ml of resin. All solvents were filtered, degassed, and sparged with helium to eliminate dissolved oxygen. The column buffer used contained 50mM Tris-HCl pH 7.5, 0.1mM DTT, 1mM EDTA, and 0.2M NaCl.

The column was equilibrated with 2 column volumes of column buffer. The protein sample (concentrated to less than 2ml) was loaded onto the column and the mobile phase was run at 0.5 ml/min. Fractions of 1ml were collected and monitored at 280nm and 290nm.

Dialysis

Prior to ion exchange chromatography, the protein sample was dialyzed to lower the salt concentration. Dialysis was performed with Pierce Slide-A-Lyzer dialysis cassettes with a 10kD molecular weight cutoff (MWCO) membrane. The dialysis buffer contained 50mM tris-HCl (pH 7.5) and 1mM EDTA. The sample was dialyzed for 6 hours, transferred to fresh buffer, then dialyzed for an additional 12 hours.

Ion exchange chromatography

FPLC Method

Buffer A contained Tris-HCl (pH 7.5) and 1mM EDTA. Buffer B contained Tris-HCl (pH 7.5), 1mM EDTA, and 1M NaCl. Buffers were filtered, degassed, and sparged with helium.

The GE Healthcare HiTrap™ Sepharose™ DEAE (Diethylaminoethyl) FF 1ml column was used. The column was washed with 5ml each of water, Buffer A, Buffer B, then Buffer A, at 1ml/min flow rate. The sample was loaded onto the column, and 1ml fractions were collected. The column was washed with 5ml of Buffer A.

YZGD was eluted with either a 20ml gradient from 0M NaCl (Buffer A) to 0.5M NaCl in Buffer A, or a stepwise elution was carried out where 5ml each of the following salt concentrations were run consecutively over the column: 0M, 0.1M, 0.2M, 0.3M, 0.4M, 0.5M NaCl.

Gravity Method

Buffer A contained 50mM Tris-HCl (pH 7.5) and 1mM EDTA. Additional buffers were made with the following NaCl concentrations: 0.1M, 0.2M, 0.25M, 0.3M, 0.4M, and 0.5M, in addition to 50mM Tris-HCl (pH 7.5) and 1mM EDTA.

A 5ml DEAE Sepharose column was equilibrated with ~120ml of buffer A. Dialyzed protein sample (4.5ml) was applied to the column, and fraction collection began, collecting ~2ml fractions. Sequentially, 8ml of each buffer was applied over the column, from the lowest to the highest NaCl concentration.

Protein Analysis

The Bradford protein assay (Bradford 1976) was used to determine protein concentration throughout the purification process. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using a stacked gel (Laemmli 1970) was used to examine the purity and quantity of YZGD.

Enzymatic activity assays were conducted to determine the levels of active YZGD protein. A standard reaction mixture contained 50mM Tris-Maleate (pH 5.0), 5mM $MgCl_2$, 1mM dithiothreitol, and 4mM pyridoxal phosphate (PLP). Between 0.5 and 4.0 units of YZGD were added to 50µl of the reaction mix. Reactions were incubated at 37°C for 30 minutes, and quenched by the addition of 50µl of 4 parts 20% packed Norit to 1 part 7% perchloric acid. After centrifugation, 50µl of the supernatant was analyzed for phosphate using the method of Ames and Dubin.

YZGD His•tag Subcloning, Expression and Purification

Plasmid Purification

The pETYZGD and pET19b (HisTag) plasmids were purified from overnight cultures of *E. coli BLR(DE3)* as described in Sambrook and Russell. The cells were disrupted by alkaline lysis using a NaOH/SDS solution, the chromosomal DNA was precipitated, and the cell extract was treated with RNase A and incubated at 37°C to degrade RNA. Phenol:chloroform was used to further purify the cell extract, and the plasmid DNA was precipitated with ethanol at -80°C. The sample was centrifuged, the ethanol discarded, and the DNA pellet was dried and resuspended in water. The DNA was further purified by a polyethylene glycol (PEG)-NaCl precipitation overnight on ice, using a 6.5% PEG and 0.8M NaCl solution to remove the tRNA.

DNA concentration was determined by measuring the absorbance at 260nm, and both concentration and purity were verified by gel electrophoresis.

Restriction Enzyme Digests

The *yzgd* gene was excised from the pETYZGD plasmid by digesting with NdeI and BlpI restriction enzymes (New England Biolabs). The pET19b plasmid was also digested with NdeI and BlpI to create matching "sticky" ends. After verification by gel electrophoresis that the digests were complete, the excised gene and the plasmid were gel-purified. This involved running each sample on a 0.7% agarose gel, staining with ethidium bromide, and physically cutting the appropriate band out of the gel. The DNA was extracted from the gel using the Gene-clean® kit (MPBiomedicals). The concentration of DNA was estimated using gel electrophoresis and DNA standards.

Ligation

The *yzgd* gene was ligated into the pET19b plasmid using T4 DNA ligase (Invitrogen). Several ratios of plasmid:gene were used. The ligations were conducted at 15°C for 24 hours.

Transformation

E. coli DH5α competent cells (Invitrogen) were transformed with each of the ligation mixtures. Transformed cells were grown on LB plates containing 100 µg/ml ampicillin to select for cells containing the pET19YZGD plasmid (and therefore the ampicillin resistance gene).

Plasmids were purified according to the procedure in Sambrook and Russell, to determine whether they were successful recombinants. Overnight cultures were grown from colonies and

plasmids were purified from the cultures. The isolated plasmids were examined using gel electrophoresis to determine which colonies contained the recombinant plasmids (pET19YZGD(HisTag)).

Verification

As an additional verification step, the purified plasmids were digested using BamHI (New England Biolabs) and again run on a gel. There is a BamHI site in pET19YZGD but not in pET19b digested with NdeI and BlpI, so the recombinant plasmid will appear as primarily linear DNA on the gel, while the pET19b plasmid will still be a supercoiled plasmid or a smaller linear band if not digested fully prior to ligation.

Once successfully transformed cells were identified, several were used to create freezer stocks for future use. One freezer stock was used to create an overnight culture for a plasmid purification, and this plasmid DNA was used to transform *BLR(DE3)* competent cells (Novagen) to be used for expression. Freezer stocks were made of the expression cell line as well.

Expression

A 100ml culture of pET19YZGD(HT) in *E. coli BLR(DE3)* cells was prepared for expression. As a control, a culture of pETYZGD in *HMS174(DE3)* was grown under identical conditions. Expression, harvesting, and sonication were performed simultaneously in the two cultures, in the same manner as the experiments described earlier. After sonication, the cell extracts were tested for protein concentration and enzymatic activity. The samples were dialyzed in 20mM Tris-HCl (pH 8.0) to prepare them for the nickel affinity column chromatography.

Affinity Chromatography Purification

Nickel Affinity resin from Novagen was used to make two 1ml columns for affinity purification, one for the His•tagged protein and one for the untagged control. Each column was equilibrated with Binding Buffer, containing 20mM Tris-HCl (pH 8.0), 5M NaCl, and 5mM imidazole. After column equilibration, the protein sample was loaded and fractions were collected. The column was washed with 5ml Wash Buffer to elute all other proteins, which contained a higher imidazole concentration than Binding Buffer (60mM imidazole) but was otherwise the same. Next, 5ml Elution Buffer (250mM imidazole) was used to elute the bound His•tagged YZGD protein.

To determine the contents of each fraction, the protein concentration and enzymatic activity were measured, and fractions containing protein were run on a SDS polyacrylamide gel to examine each fraction and the purity of the eluted His•tagged YZGD protein.

Results and Discussion

YZGD expression and purification

The YZGD enzyme was first expressed and purified using the same methods as described in Tirrell et al. (2006). Table 1 and the protein gel in Figure 5 show the results of this initial purification. This protein came from a 2L batch of *E. coli* HMS(174) cells incubated at 37°C during expression of YZGD. The dark, wide band in lane 2 shows that the protein overexpressed successfully. However, throughout the purification steps, there were many other contaminating proteins remaining. This lead to the decision to express the protein at 15°C, with hope that it would express better and be more soluble. This has been the case with a number of other proteins studied in the lab.

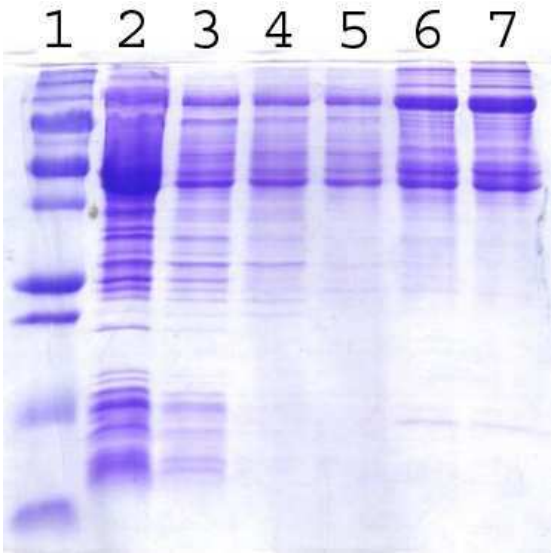

Figure 5: Purification of YZGD protein expressed at 37°C. Lane 1: molecular weight markers, Lane 2: crude extract, Lane 3: sonicated supernatant, Lane 4: streptomycin fractionation supernatant, Lane 5: 55% ammonium sulfate fraction, Lane 6: size exclusion pooled fractions, Lane 7: ion exchange pooled fractions. Some contaminating proteins were eliminated during each step of purification, but the final sample still contained many proteins other than YZGD.

Table 1: Purification of YZGD protein expressed at 37°C.

Step	Volume (ml)	Activity (units/ml)	Total units	[Protein] (mg/ml)	Specific activity (units/mg)	Total protein (mg)
Sonication	11.9	9.57	114	8.69	1.10	103
Ammonium sulfate 55%	1.28	56.7	72.3	23.9	2.37	30.5
Size Exclusion Chromatography	10.7	0.952	10.2	1.31	0.727	14.0
Ion Exchange Chromatography	7.00			1.50		10.5

When expressed at 15°C versus 37°C, the expression level of YZGD was comparable, or slightly better at 15°C as seen in Figure 6 (compare lane 3 to lane 7). As a negative control, samples of cells containing only the expression vector pET11b (and therefore not expressing YZGD) were included on the gel.

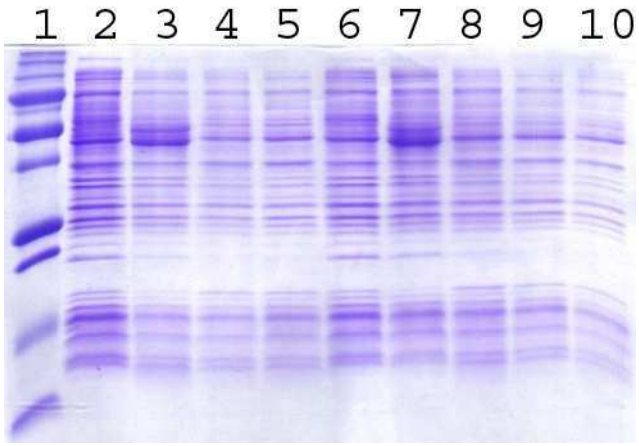

Figure 6: Comparison of expression of YZGD protein from pETYZGD in *E. coli HMS174(DE3)*, 15°C vs 37°C. Lane 1: molecular weight markers, Lane 2: 37°C pETYZGD prior to induction, Lane 3: 37°C pETYZGD at 3 hours after induction, Lane 4: 37°C pET11b prior to induction, Lane 5: 37°C pET11b at 3 hours after induction, Lane 6: 15°C pETYZGD prior to induction, Lane 7: 15°C pETYZGD at 24 hours after induction, Lane 8: 15°C pET11b prior to induction, Lane 9: 15°C pET11b at 24 hours after induction. YZGD protein was successfully expressed at both 37°C and 15°C.

While expressing at a lower temperature did not significantly increase expression, the change in expression temperature from 37°C to 15°C led to a significant increase in solubility. After sonicating the cells to break them open and centrifugation to remove insoluble proteins and other cellular debris, most of the YZGD protein was present in the supernatant.

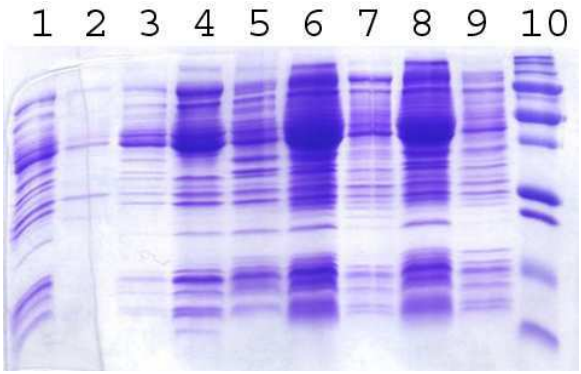

Figure 7: Comparison of solubility of YZGD protein, expressed at 15°C vs 37°C. Lane 1: 15°C crude extract, Lane 2: 15°C freeze thaw supernatant, Lane 3: 15°C freeze thaw pellet, Lane 4: 15°C sonicated supernatant, Lane 5: 15°C sonicated pellet, Lane 6: 37°C crude extract, Lane 7: 37°C sonicated supernatant, Lane 8: 37°C sonicated pellet, Lane 9: 15°C pET11b control sonicated supernatant, Lane 10: molecular weight markers. YZGD expressed at 15°C is mostly in the sonicated supernatant, indicating greater solubility compared to the protein expressed at 37°C.

Purification of the protein expressed at 15°C included the steps shown in Figure 8. YZGD did not elute from the cells in a freeze-thaw step, and thus the pellet from this step was processed, as it contained YZGD. A portion of YZGD was found to be soluble upon sonication, and the sonicated supernatant was processed further. Streptomycin removed nucleic acid (DNA) from the sample. Two rounds of anion exchange chromatography using a DEAE FPLC column from GE Healthcare removed some contaminating proteins from YZGD, but not enough to render YZGD pure. A gravity DEAE anion exchange column further purified YZGD, but still not enough for the enzyme to be pure enough to crystallize for x-ray crystal structure determination. Thus we decided to tackle the problem from a new perspective, to create a His•tagged protein and purify this tagged protein by nickel affinity chromatography.

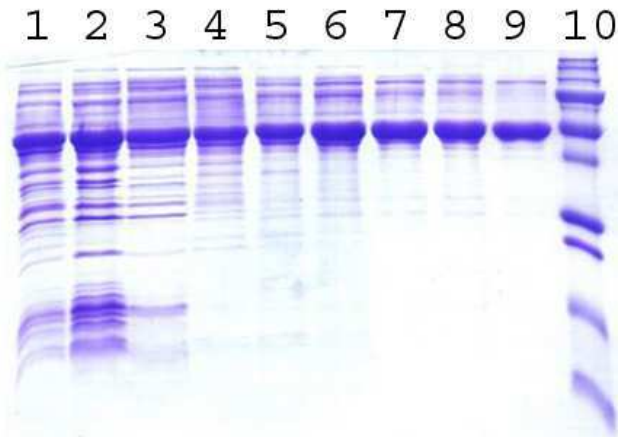

Figure 8: Purification of YZGD protein expressed at 15°C. Lane 1: crude extract, Lane 2: freeze-thaw pellet, Lane 3: sonicated supernatant, Lane 4: streptomycin supernatant, Lane 5: FPLC ion exchange fraction, Lane 6: 55% ammonium sulfate fraction, Lane 7: FPLC ion exchange fraction, Lane 8: dialyzed fraction, Lane 9: gravity ion exchange fraction, Lane 10: molecular weight markers. The final protein sample (lane 9) was not sufficiently pure for x-ray crystal structure determination.

Table 2: Purification of YZGD protein expressed at 15°C.

Step	Volume (ml)	Activity* (units/ml)	Total activity* (units)	[Protein] (mg/ml)	Specific activity* (units/mg)	Total protein (mg)
Sonication	2.45	31.2	76.4	17.8	1.75	43.6
Streptomycin fraction	4.00			6.70		26.8
Ion Exchange Chromatography	6.00			1.77		10.6
Ammonium Sulfate Fractionation and Dialysis	2.40			2.98		7.15
Ion Exchange Chromatography	4.50					
Gravity Ion Exchange Chromatography				0.350		

*** Detailed activity information was not determined for many steps of this procedure, because the purity of YZGD was not good enough to support using this procedure to prepare samples for crystallography.**

YZGD His•tag subcloning and purification

Since the purification of YZGD proved ineffective, it was decided that a His•tag would be added to the protein. This sequence of histidines, attached to one end of the protein, would allow it to bind to Ni^{2+} ions. Therefore, the cell extract could be run over a nickel affinity column, and when the bound protein is eluted, it should be pure YZGD. The pET19b His•tag expression vector was used. The map of this vector is shown in Figure 9, with the chosen restriction enzyme target sites marked by red rectangles. Computational analysis showed that the *yzgd* gene did not contain either of these sites, so the same restriction enzymes could be used both to excise *yzgd* out of pETYZGD and to digest the pET19b plasmid, giving them "sticky" ends that can be ligated together easily.

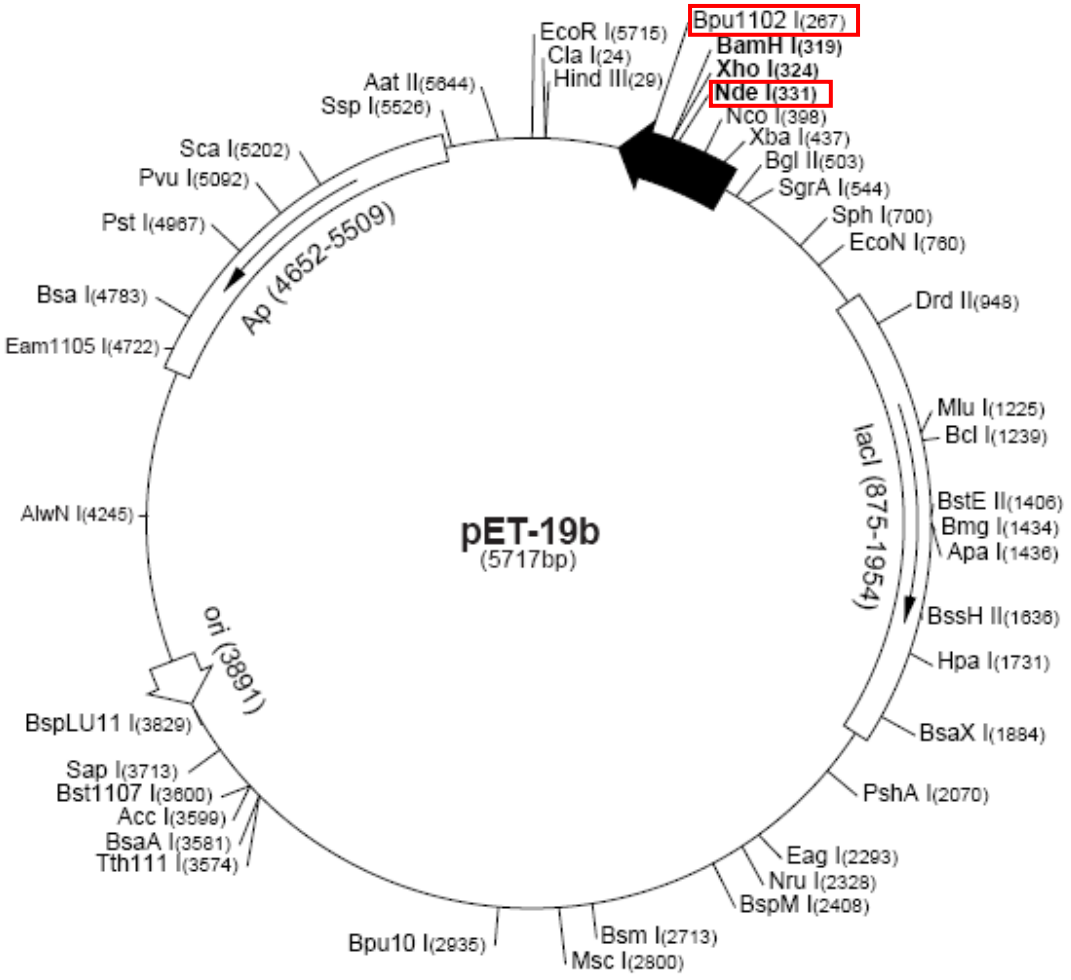

Figure 9: pET19b plasmid map. Red rectangles indicate the digestion sites for the selected restriction enzymes.
(http://www.emdbiosciences.com/docs/NDIS/TB049-000.pdf)

Plasmid Purification

The pET19b and pETYZGD plasmids were each purified from *E. coli DH5α* cells. The amount of plasmid DNA obtained was calculated from the absorbance of the solution at 260nm, using the following equation:

DNA Concentration (ng/µl) = A260(OD) · 50(ng · µl^{-1} · OD^{-1}) · dilution factor

where the dilution factor was 100. The total DNA was calculated by multiplying the concentration by 120 µl since that is the volume of plasmid obtained. The concentration and purity were verified qualitatively by gel electrophoresis (Fig. 10). The intensity of the band of obtained plasmid matches the control band.

Table 3: Quantification of isolated pET19b and pETYZGD.

Sample	A260 of 1:100 dilution	[DNA] (ng/µl)	Total DNA (µg)
pET19b	0.047	235	28.2
pETYZGD	0.041	205	24.6

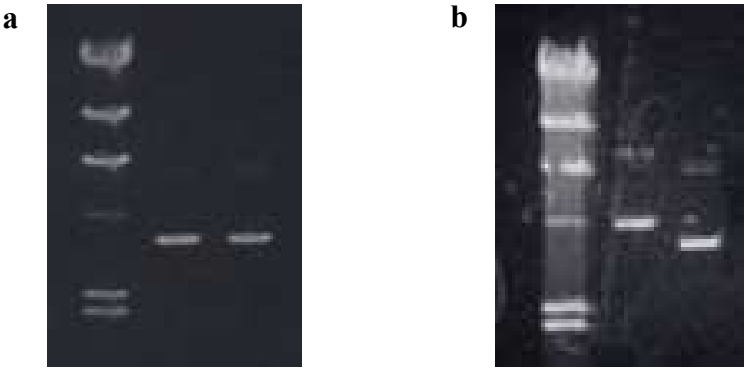

Figure 10: Agarose gel electrophoresis of purified pET19b and pETYZGD.

a) **pET19b purification. Lane 1: λ HindIII molecular weight markers, Lane 2: 100ng pET19b, Lane 3: 100ng pET11b control.**
b) **pETYZGD purification. Lane 1: λ HindIII molecular weight markers, Lane 2: 100ng pETYZGD, Lane 3: 100ng pET11b control.**

Digestion and purification

The pET19b plasmid was digested with BlpI and NdeI restriction enzymes. Simultaneously, single digests were prepared with each of the enzymes. After 1 hour of digestion, these single digests were run on a 0.7% agarose gel alongside the undigested plasmid, for comparison (Fig. 11a). Since the BlpI sample still contained some undigested plasmid, the digestion was allowed to proceed. After 2.5 hours (Fig. 11b), both samples appeared to be completely digested, and it was safe to assume that the double digest had also been successful.

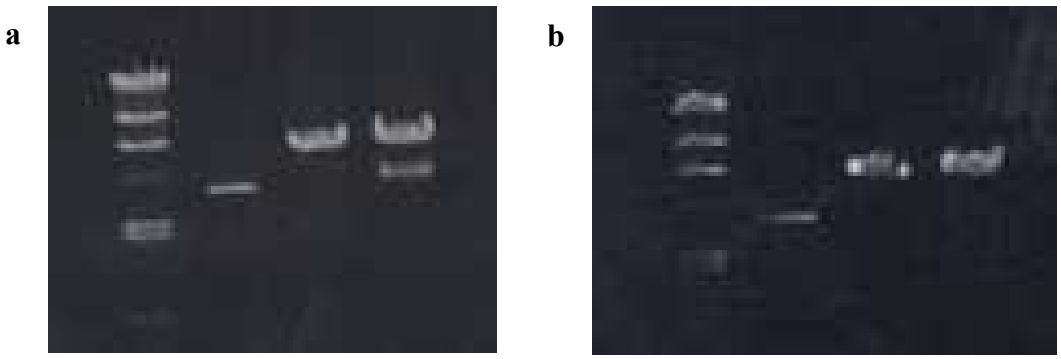

Figure 11: Digestion of pET19b by BlpI and NdeI.
a) Results after 1 hour. b) Results after 2.5 hours. Lane 1: λ HindIII molecular weight markers, Lane 2: undigested pET19b, Lane 3: pET19b digested by NdeI, Lane 4: pET19b digested by BlpI. While the BlpI digestion was not complete at 1 hour, both reactions appeared to be complete by 2.5 hours.

The same procedure was conducted with the pETYZGD plasmid, to excise the *yzgd* gene from the plasmid. After 1 hour, the two single digests of pETYZGD were analyzed by gel electrophoresis (Fig. 12). The majority of the DNA was linear, digested plasmid, but there was still a small fraction of remaining undigested plasmid. The digestions were allowed to proceed for a total of 2.5 hours to complete digestion (Fig. 14).

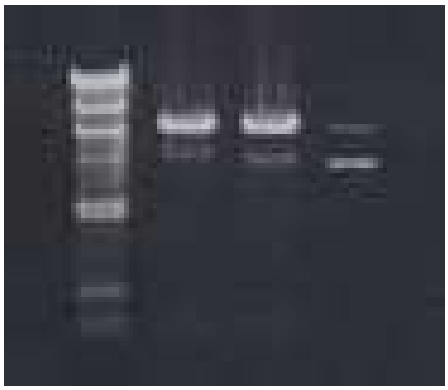

Figure 12: Digestion of pETYZGD by BlpI and NdeI, after 1 hour. Lane 1: λ HindIII molecular weight markers, Lane 2: pETYZGD digested by NdeI, Lane 3: pETYZGD digested by BlpI, Lane 4: undigested pETYZGD. There was still a small amount of undigested plasmid after 1 hour which was completely digested after 2.5 hours.

The digested pET19b plasmid was run in one large lane on a gel to be cut out for purification, as shown in Figure 13. Only one band is visible because the fragment excised between the two restriction sites is so small (~60 bases) that it would have quickly run off of this gel. Unfortunately, this means that any plasmids that were only digested once will be similar in size to the plasmids that were completely digested, and will not be separated in the gel, hence the importance of the single digest controls shown above. The band is consistent with the 5.7kb size of the pET19b plasmid.

Figure 13: Gel purification of digested pET19b plasmid DNA. Lane 1: λ HindIII molecular weight markers, Lane 2: 1µl of digested pET19b, Lane 3: rest of digested pET19b. The band was cut from the gel for purification.

Digested pETYZGD was also run on a gel for purification (Fig. 14). The bottom band is the *yzgd* gene, to be cut out of the gel for purification. This is consistent with the known size of the *yzgd* gene, approximately 1.2kb The YZGD gene is well separated from the plasmid DNA for ease in purification.

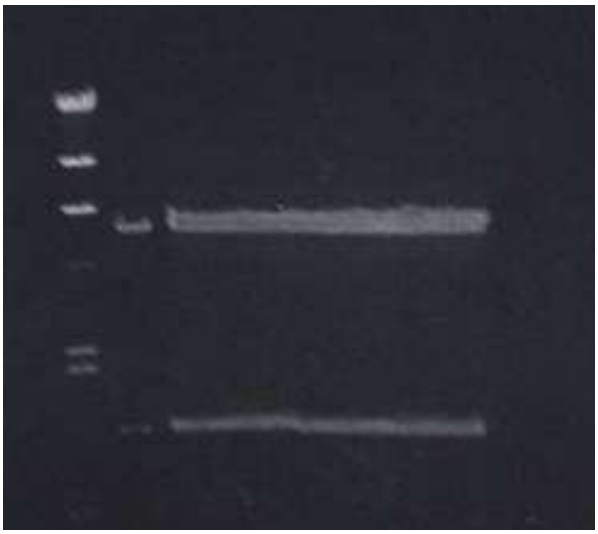

Figure 14: Gel purification of *yzgd* gene from digested pETYZGD plasmid DNA. Lane 1: λ HindIII molecular weight markers, Lane 2: 1µl of digested pETYZGD, Lane 3: rest of digested pETYZGD. The bottom band, which is the yzgd gene, was cut from the gel for purification.

The concentration of purified plasmid was determined by gel electrophoresis. Two different volumes of plasmid, 1µl and 2µl, were run alongside 100ng of pET11b plasmid. The bands were visually compared to the control to estimate the concentration. For pET19b, shown in Figure 15, the 2µl band was comparable to the 100ng control. Therefore, the concentration was determined to be approximately 50ng/µl.

Figure 15: Quantitation of purified digested pET19b plasmid. Lane 1: λ HindIII molecular weight markers, Lane 2: 1µl pET19b, Lane 3: 2µl pET19b, Lane 4: 100ng pET11b for comparison. By comparison to the band of known concentration, the pET19b solution appears to be about 50ng/µl.

The same procedure was followed to quantitate the *yzgd* DNA (Fig. 16). In this case, the 2µl band was about half the intensity of the 100ng control band. Therefore, there were approximately 50ng of DNA in the 2µl band, and the concentration was estimated at 25ng/µl.

Figure 16: Quantitation of purified excised *yzgd* gene. Lane 1: λ HindIII molecular weight markers, Lane 2: 1µl *yzgd*, Lane 3: 2µl *yzgd*, Lane 4: 100ng pET11b for comparison. By comparison to the band of known concentration, the yzgd solution appears to be about 25ng/µl.

Ligation and Transformation

As shown in Table 4, varying amounts of the purified *yzgd* were added to the digested pET19b plasmid, and ligated together to form a new recombinant plasmid. The resulting plasmids were transformed into *E. coli DH5α* cells and grown on selective LB plates containing 100 μg/ml ampicillin. The number of colonies that grew is also shown in Table 4. The colonies resulting from ligation E, the negative control to which no *yzgd* gene was added, indicate that some pET19b plasmids were not completely doubly digested, and re-ligated to themselves. However, the other ligations yielded 4-10 times more colonies, an initial indication that the subcloning was probably successful.

Table 4: Ligations of pET19b with *yzgd*, and results of transformation into *E. coli DH5α* cells. pET19b = 5500bp, yzgd = 1239bp

Ligation	pET19b (ng)	*yzgd* (ng)	pET19b:*yzgd* ratio	# colonies
A	200	6.25	2.4	13
B	200	12.5	1.2	16
C	200	25	0.6	37
D	200	50	0.3	15
E	200	0	---	4
pUC control	---	---	---	~160

Twelve colonies were chosen from ligations A-D, grown as overnight cultures in LB media, and plasmids were isolated from these cultures. The isolated plasmids were analyzed by gel electrophoresis to determine if they were the new recombinant pET19bYZGD plasmid. This is shown in Figure 17. Plasmid DNA from colonies A2 and B2 was consistent in size with the pET19b plasmid, which reinforces the earlier assumption that some fraction of the pET19b plasmids were only digested by one of the two restriction enzymes and were ligated back together. Colony A1 contains a smaller unknown DNA. All other colonies appeared to be of the correct size to contain the new pET19YZGD plasmid.

Figure 17: Purified plasmids from transformations of *E. coli DH5α* with pET19b-YZGD ligation mixtures. Lane 1: λ HindIII molecular weight markers, Lane 2: pET19b control, Lanes 3-14: purified plasmid from colonies A1, A2, B1, B2, C1, C2, C3, C4, C5, C6, D1, D2. Colonies B1, C1, C2, C3, C4, C5, C6, D1, D2 contain plasmids larger than pET19b, indicating that they probably contain pET19YZGD.

To verify the success of subcloning, the isolated plasmids were digested with BamHI. The pET19b plasmid contains a single BamHI site, which is cut out when cloning the *yzgd* gene into the plasmid. Since the *yzgd* gene itself contains a BamHI restriction site, the pET19YZGD(HT) plasmid should be digested by BamHI to produce a linear plasmid. Figure 18 confirms the results shown in Figure 17. A1 contains a plasmid too small to be pET19YZGD(HT). A2 and B2 contain undigested pET19b (a small amount is relaxed or digested plasmid (upper band) but a majority is supercoiled plasmid (lower band)). The rest of the colonies appear to contain pET19YZGD(HT) which can be digested by BamHI. The overnight culture made from colony C6 was streaked to single colonies, and a single colony from this plate was cultured and used to make a freezer stock.

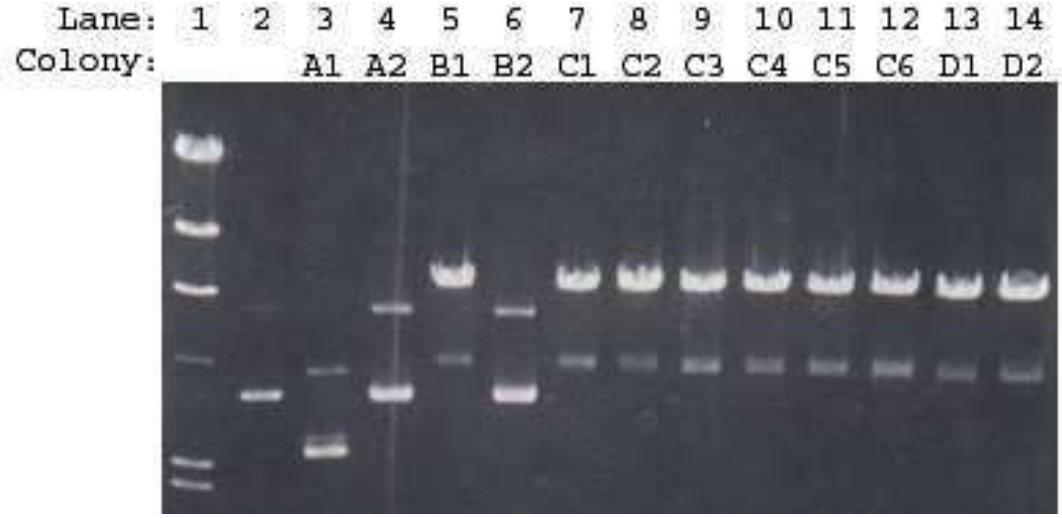

Figure 18: Purified plasmids from transformations of *E. coli DH5α* with pET19b-YZGD ligation mixtures, digested with BamH1. Lane 1: λ HindIII molecular weight markers, Lane 2: pET19b control, Lanes 3-14: digested plasmids from colonies A1, A2, B1, B2, C1, C2, C3, C4, C5, C6, D1, D2. Colonies B1, C1, C2, C3, C4, C5, C6, D1, D2 appear to contain mostly digested, linear plasmid of the predicted correct size, indicating that they contain pET19YZGD.

The recombinant pET19YZGD(HT) plasmid was isolated from the transformed *E. coli*. As with the plasmid purifications described earlier, the concentration of isolated DNA was determined by measuring its absorbance at 260nm (Table 5) and verified by gel electrophoresis (Fig. 19).

Table 5: Amount of isolated pET19YZGD(HT) DNA.

Sample	A260 of 1:100 dilution	[DNA] (ng/µl)	Total DNA (µg)
pET19YZGD(HT)	0.048	240	28.8

Figure 19: Purified pET19YZGD(HT) plasmid. Lane 1: λ HindIII molecular weight markers, Lane 2: 100ng pET19YZGD, Lane 3: 100ng pET19b, Lane 3: 100ng pET11b as a control. All of the samples appear pure, and of the same quantity.

Expression

The isolated pET19YZGD(HT) was used to transform *E. coli* cells for expression, and freezer stocks were made. The new tagged protein and the original untagged protein (as a control) were overexpressed simultaneously. Figure 20 shows the results of this expression. The expression level of the His•tagged protein was comparable to that of the untagged protein, possibly higher. An increase in expression level could be because different strains of *E. coli* were used to express the protein. The BLR(DE3) cells used to express the tagged protein may be more effective than the HMS(174) strain used for the pETYZGD.

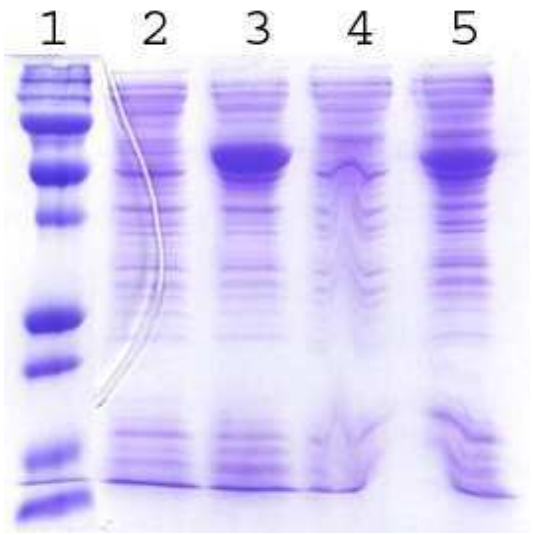

Figure 20: Comparison of expression: YZGD protein from pETYZGD plasmid in *E. coli HMS174(DE3)* vs YZGD(HT) protein from pET19YZGD(HisTag) plasmid *in E. coli*. Lane 1: molecular weight markers, Lane 2: expression of pET19YZGD(HT) prior to induction, Lane 3: expression of pET19YZGD(HT) at 24 hours after induction, Lane 4: expression of pETYZGD prior to induction, Lane 5: expression of pETYZGD at 24 hours after induction. The His•tagged YZGD expressed just as well as the untagged YZGD, indicating that the tag does not interfere with expression.

The solubility of the tagged and untagged YZGD was ascertained by analyzing the protein composition of the sonicated supernatants and pellets by SDS-PAGE (Figure 21). The solubility level of the His•tagged YZGD protein was satisfactory, with ~50% of the tagged protein being present in the supernatant after sonication and centrifugation (lane 8).

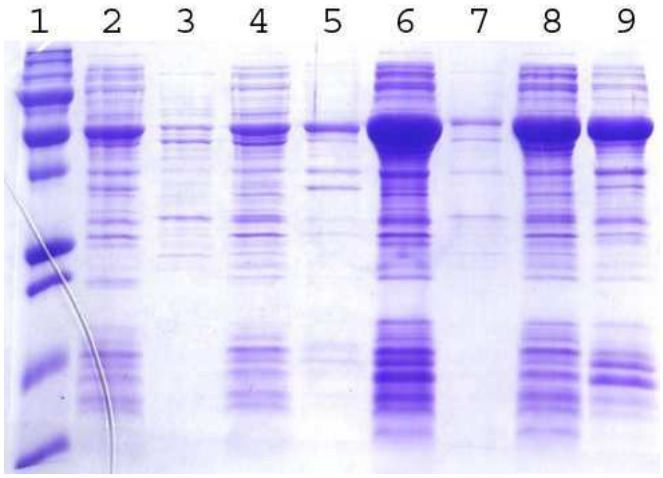

Figure 21: Comparison of solubility: YZGD protein from pETYZGD plasmid in *E. coli HMS174(DE3)* vs YZGD(HT) protein from pET19YZGD(HT) plasmid *in E. coli*. Lane 1: molecular weight markers, Lane 2: pETYZGD freeze thaw pellet, Lane 3: pETYZGD freeze thaw supernatant, Lane 4: pETYZGD sonicated supernatant, Lane 5: pETYZGD sonicated pellet, Lane 6: pET19YZGD(HT) freeze thaw pellet, Lane 7: pET19YZGD(HT) freeze thaw supernatant, Lane 8: pET19YZGD(HT) sonicated supernatant, Lane 9: pET19YZGD(HT) sonicated pellet. Both the untagged and tagged protein show good solubility, with a significant portion of the YZGD protein located in the supernatant after sonication.

The protein concentration of the sonicated supernatant of each culture was measured, and is recorded in Table 6. The total protein in pET19YZGD(HT) in BLR(DE3) was nearly double that of the total protein in pETYZGD in *HMS174(DE3)*. This could be due to the use of different *E. coli* strains or to better expression or solubility of YZGD.

Likewise, the enzymatic activity was analyzed for each culture. The pET19ZGD(HT) in BLR(DE3) culture had more than double the activity as compared to pETYZGD in *HMS174(DE3)*, measured in units per mg of total protein. Each unit is one nanomole of substrate cleaved per minute. The increase in activity is most likely due to the increased solubility of the protein described above.

Table 6: Concentration and enzymatic activity of tagged and untagged enzymes.

Sample	Volume (ml)	Activity (units/ml)	Total units	[Protein] (mg/ml)	Specific activity (units/mg)	Total protein (mg)
YZGD	1.06	8.26	8.76	6.5	1.27	6.9
YZGD(HT)	1.038	35.0	36.3	11.5	3.05	11.9

Affinity Purification

A pilot affinity chromatography purification experiment was performed to verify that this method would effectively purify the His•tagged protein. Simultaneous experiments were performed using the tagged and untagged protein, and the results are shown in Figure 22. When extract containing untagged YZGD was loaded onto the nickel column, proteins washed through the column. The band of overexpressed YZGD is clearly visible in these fractions. When the elution buffer was used, no other proteins were eluted from the columnm indicating that no proteins had bound to the nickel-containing resin. In the case of the His•tagged YZGD (lanes 1-5), many proteins eluted in the wash fractions, but little, if any, YZGD(HT) eluted in this stage. When elution buffer was applied, YZGD(HT) was eluted from the column, and appears to be pure as seen in lanes 4 and 5 of Figure 22.

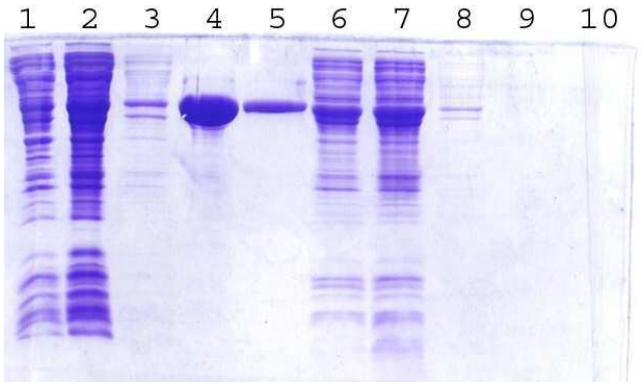

Figure 22: Ni-NTA affinity chromatography of tagged and untagged proteins. Lane 1: YZGD(HT) wash fraction 1, Lane 2: YZGD(HT) wash fraction 2, Lane 3: YZGD(HT) wash fraction 3, Lane 4: YZGD(HT) elution fraction 1, Lane 5: YZGD(HT) elution fraction 2, Lane 6: YZGD wash fraction 1, Lane 7: YZGD wash fraction 2, Lane 8: YZGD wash fraction 3, Lane 9: YZGD elution fraction 1, Lane 10: YZGD elution fraction 2. The untagged YZGD protein washes through the column immediately, while the tagged YZGD(HT) binds to the column until eluted, essentially devoid of other proteins.

Conclusions

The first purification methods for YZGD, involving size exclusion and ion exchange chromatography, were insufficient for obtaining protein pure enough for x-ray crystallography. While these methods had previously been used to purify the enzyme for determination of its substrate and quantitation of activity, the methods did not yield enough pure protein for crystallography. Adding a histidine tag to YZGD before evaluating its activity would have been unwise, since this modification might have altered the enzyme's activity in some way that we would have been unaware of. However, since the activity of YZGD has now been thoroughly evaluated (Tirrell et al. 2006), any change in activity could be detected.

The tagged YZGD(HT) protein has been shown to have good expression and solubility. It also appears to purify very effectively over a nickel affinity chromatography column, and retains its enzymatic activity throughout all of these steps, with a final specific activity comparable to the purified untagged YZGD. Since this work, Julian Ramos has scaled up expression and purification to obtain a larger amount of the YZGD protein. He also obtained excellent purification and activity. The next step in the project is the crystallization of the protein with the help of our collaborator, Dr. Joseph Wedekind from the University of Rochester. It is possible that the tag will need to be removed from the protein before it will successfully crystallize into its native structure. The crystallization process will be considerably easier now that there is a simple and effective way to purify large amounts of YZGD protein.

The eventual determination of the three-dimensional structure of YZGD will be important to scientists who study HAD superfamily enzymes, as well as those who study Nudix hydrolase enzymes. The structure of YZGD can be compared to known PLPase structures within the nitrophenyl phosphatase (PNPPase) family of the HAD superfamily, PNPPase family members with other substrates, HAD superfamily members not in the PNPPase family, and

Nudix hydrolases. As more enzymes are characterized, in terms of both activity and structure, it becomes easier to identify features that determine substrate specificity. For these reasons, the structure of YZGD will be of great value.

References

Allen, K.N. and Dunaway-Mariano, D. (2004) Phosphoryl group transfer: evolution of a catalytic scaffold. *Trends Biochem. Sci*. **29**, 495-503.

Ames, B.N. and Dubin, D.T. (1960) The role of polyamines in the neutralization of bacteriophage deoxyribonucleic acid. *J. Biol. Chem*. **235**, 769–775.

Attias, J. and Bonnet, J.L. (1972) A Specific Alkaline p-Nitrophenyl Phosphatase Activity from Baker's Yeast. *Biochim. Biophys. Acta* **268**, 422–430.

Begun, J., Sifri, C.D., Goldman, S., Calderwood, S.B., and Ausubel, F.M. (2005) Staphylococcus aureus Virulence Factors Identified by Using a High-Throughput Caenorhabditis elegans-Killing Model. *Infection and Immunity* **73**, 872-877.

Bradford, M. (1976) A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* **72**, 248-254.

Burroughs, A.M., Allen, K.N., Dunaway-Mariano, D., and Aravind, L. (2006) Evolutionary Genomics of the HAD Superfamily: Understanding the Structural Adaptations and Catalytic Diversity in a Superfamily of Phosphoesterases and Allied Enzymes. *J. Mol. Biol*. **361**, 1003-1034.

Jang, Y.M., Kim, D.W., Kang, T.C., Won, M.H., Baek, N.I., Moon, B. J., Choi, S.Y., and Kwon, O.S. (2003) Human Pyridoxal Phosphatase: Molecular Cloning, Functional Expression and Tissue Distribution. *J. Biol. Chem.* **278**, 50040-50046.

Koonin, E.V., and Tatusov, R. L. (1994) Computer Analysis of Bacterial Haloacid Dehalogenases Defines a Large Superfamily of Hydrolases with Diverse Specificity: Application of an Iterative Approach to Database Search. *J. Mol. Biol.* **244**, 125-132.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Mamedov, T.G., Suzuki, K., Miura, K., Kucho, K., and Fukuzawa, H. (2001) Characteristics and Sequence of Phosphoglycolate Phosphatase from a Eukaryotic Green Alga *Chlamydomonas reinhardtii*. *J. Biol. Chem.* **276**, 45573-45579.

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

Sambrook, J., and Russell, D.W. (2001) *Molecular Cloning: A Laboratory Manual*. 3rd ed. NY: Cold Spring Harbor Press.

Tirrell, I.M., Wall, J.L., Daley, C.J., Denial, S.J., Tennis, F.G., Galens, K.G. and O'Handley, S.F. (2006) YZGD from *Paenibacillus thiaminolyticus*, a Pyridoxal Phosphatase of the HAD (Haloacid Dehalogenase) Superfamily and a Versatile Member of the Nudix (Nucleoside Diphosphate x) Hydrolase Superfamily. *Biochem. J.* **394**, 665–674.

Tremblay, L.W., Dunaway-Mariano, D., Allen, K. (2006) Structure and Activity Analyses of *Escherichia coli* K-12 NAGD Provide Insight into the Evolution of Biochemical Function in the Haloalkanoic Acid Dehalogenase Superfamily. *Biochemistry* **45**, 1183-1193.

Yang, J., Dhamija, S.S., and Schweingruber, M.E. (1991) Characterization of the Specific p-Nitrophenyl Phosphatase Gene and Protein of *Schizosaccharomyces pombe*. *Eur. J. Biochem.* **198**, 493-497.