AMINO ACID SEQUENCE OF NUCLEOSIDE HYDROLASES FROM PLANTS

by

Kateryna V. Ogorodnik

A Thesis Submitted to the
Faculty of the College of Graduate Studies at
Middle Tennessee State University
in Partial Fulfillment of
the Requirements for the Degree of
Master of Science

Middle Tennessee State University
August 2014

Thesis Committee:

Dr. Paul C. Kline

Dr. Donald A. Burden

Dr. Mary Farone

ACKNOWLEDGEMENTS

Though only my name appears on the cover of this thesis, great people have contributed to its production. I owe my gratitude to all those people who have made this thesis possible and because of whom my graduate experience has been one that I will cherish forever.

I have been blessed to have an advisor, Dr. Paul Kline, to whom I would like to express my sincere gratitude for the continuous support of my study and research, for his patience, motivation, care, love, and immense knowledge. His guidance helped me in all the time of research and writing of thesis. I was amazingly fortunate having Dr. Kline as my advisor and mentor throughout my undergraduate and graduate study.

My gratitude and appreciation goes to the greatest instrument support engineer, Jessie Weatherly, who spent countless hours training me using various laboratory instruments and guiding me through getting good results. The smooth running of the equipment is much more a testament to his efforts than my own.

Also, I would like to thank my thesis committee: Dr. Burden, Dr. Mary Farone, and Dr. Kline for their encouragement, insightful comments, and hard questions. Another thanks goes to Dr. Melton for her valuable suggestions, recommendations, and job opportunities.

Many friends have helped me stay sane through this learning journey. Their support and care helped me overcome setbacks and stay focused on my graduate study. I greatly value their friendship and appreciate their belief in me.

None of my accomplishments would have been possible without the love and support of my family. My parents, brothers, and sister have been a constant source of love, concern, support, strength, and prayer.

Department of Chemistry has provided the support and equipment I needed to complete my research.

I would like to acknowledge financial, academic, and technical support of the Middle Tennessee State University and its wonderful staff.

Most important, thanks and glory goes to my God for blessing me with wonderful people on my path, for blessing me in health, wisdom, and strength, and for blessing me in research.

"Praise the Lord, all you nations;
extol him, all you people.
For great is his love toward us,
and the faithfulness of the
Lord endures forever.
Praise the Lord."
Psalm 117

ABSTRACT

Nucleoside hydrolases are key enzymes of the purine salvage pathways of various bacteria, yeast, parasitic protozoa, insects, fish, and plants. While the structures of nucleoside hydrolases from parasitic protozoans have been extensively studied, almost no structural information beyond subunit molecular weights is available for nucleoside hydrolases from plants.

Based on a nonlinear regression using the Michaelis-Menten equation, the Michaelis constants (K_m) of inosine for yellow lupin adenosine nucleosidase and inosine nucleosidase were determined. A K_m of 260 ± 75 μM for inosine nucleosidase and 9820 ± 13801 μM for adenosine nucleosidase was found. The high standard deviation of the K_m for adenosine nucleosidase was due to the solubility limits of inosine preventing a complete kinetic analysis.

Three nucleoside hydrolases from plants, adenosine and inosine nucleosidase from yellow lupin, and adenosine nucleosidase from soybean, along with rihC from *E. coli* have been partially sequenced by mass spectrometry. After a final polishing step based on isoelectric focusing was performed, the enzymes were subjected to tryptic digestion followed by separation of the resulting peptides using reverse phase HPLC. The peptides were subjected to MS/MS and the sequence of selected peptides determined using PEAKS software.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

FIGURE PAGE

FIGURE PAGE

CHAPTER I

INTRODUCTION

Nucleic acids, which allow organisms to transfer genetic information, are molecules essential for all forms of life. There are two types of nucleic acids: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). Nucleic acids are linear polymers composed of nucleotides linked by phosphodiester bonds. Nucleotides consist of three components: a nitrogenous base (a purine or pyrimidine nucleobase), a pentose sugar (ribose or 2-deoxyribose), and one or more phosphate groups. The most common nitrogenous bases are cytosine, guanine, adenine, thymine, and uracil. Nitrogenous bases consist of two groups, purines and pyrimidines. The substructure that consists of a nucleobase and sugar is termed a nucleoside. Nucleosides are composed of a cyclic furanoside-type sugar (β-D-ribose or β-D-2-deoxyribose) substituted at C1' by a heterocyclic purine or pyrimidine base attached by a β -glycosyl C1'-N linkage shown in Figure 1.[1] This linkage is in the β configuration with respect to the ribose sugar, in contrast to the α position of the hydrogen.[2] The base is relatively free to rotate around the glycosidic linkage although there are preferred conformations. The two extreme conformations of the base around the glycosidic linkage are *syn* and *anti.* In the illustration below the adenine ring in the *anti* conformation is nearly perpendicular to the furanose ring and projects away from the furanose ring. This is the favored conformation. In the *syn* conformation the adenine ring projects over the furanose ring.

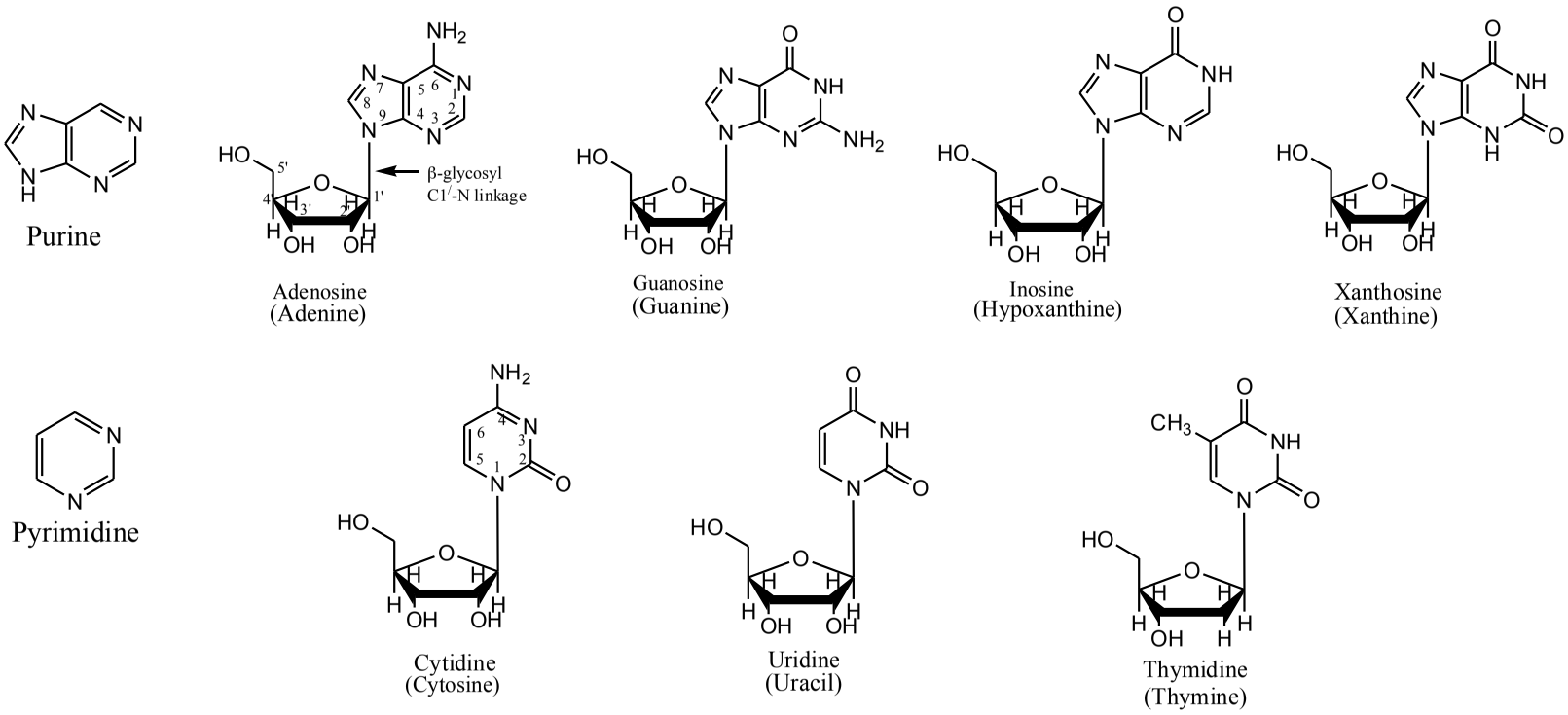


Figure 1. The common purine and pyrimidine nucleosides. Note thymidine contains 2'-deoxyribose, while ribose is the sugar found in the other nucleosides. The name of the nitrogenous base is in parentheses.

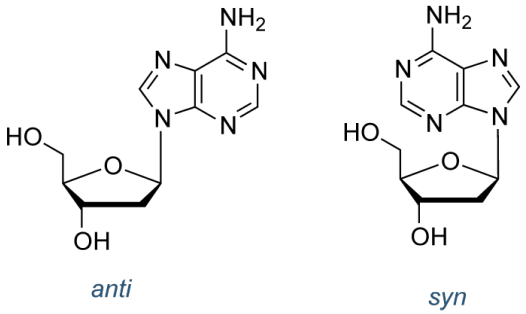


Figure 2. Orientation of base to sugar around glycosidic bond in nucleosides.

The *syn* conformation is found in purines in Z-form DNA, an unusual form of DNA. The *anti* conformation reflects the relative spatial orientation of the base and sugar found in most conformations of DNA.[2]

Nucleosides are synthesized by biochemical pathways that involve a series of chemical reactions in the cell. Cell growth depends on a supply of purine and pyrimidine nucleosides and nucleotides.

1. Purines and Pyrimidines

Purines are heterocyclic compounds with a six-membered ring fused with a five-membered imidazole ring (Figure 1). Adenine, guanine, and hypoxanthine are common purine bases.[3] These are synthesized in cells *de novo*, in multistep biochemical reactions with the base being built on a phosphorylated ribose sugar molecule. Adenine has an exocyclic amino group at the C6 position of the purine ring, while guanine has an exocyclic amino group at the C2 position and an exocyclic carbonyl group at the C6 position.[2] Hypoxanthine is found rarely as a constituent of nucleic acids. However hypoxanthine is found during the synthesis of purines,

Inosine monophosphate (IMP) is the common intermediate in the synthesis of AMP and GMP.[2]

Pyrimidines are six-membered rings containing two nitrogen atoms at the N1 and N3 positions. The common pyrimidine bases are cytosine, thymine, and uracil. The structures of thymine, cytosine, and uracil are shown in Figure 1. Cytosine contains an exocyclic amino group at C4 position. Thymine contains exocyclic carbonyl groups at the C4 and C2 positions with an exocyclic methyl group at the C5 position. Uracil is similar to thymine but lacks the methyl group at the C5 position. In addition, uracil is a component of ribonucleic acid (RNA), where it is used in place of thymine as one of the pyrimidines.[2]

2. Purine and Pyrimidine Metabolism

Nucleotides are crucial cellular components for plant growth, development, and metabolism. Purine and pyrimidine nucleotides play a fundamental role in information storage and retrieval in dividing and elongating tissues, and as components of transcripts.[4] The pathways for the synthesis of nucleotides in plant cells are similar to those found in animals and microorganisms. Nucleotides are synthesized from amino acids and small molecules via *de novo* pathways or recovered by salvage pathways where free nucleobases and nucleosides are preformed.[5] Salvage pathways recycle nucleosides and free bases. The *de novo* pathway enzymes build purine nucleotides using small molecules including the amino acids glutamine, glycine, and aspartate, the activated ribose precursor 5-phosphoribosyl-1-pyrophosphate (PRPP), N^{10}-formyl-THF, and carbon dioxide.[4, 6]

This is an energy-intensive pathway in which 6 ATPs are required in the purine biosynthetic pathway from ribose-5-phosphate to the branchpoints involving IMP (Figure 3). One ATP per step is consumed at steps 1, 3, 5, 6, 7, and 8. However, 7 high-energy phosphate bonds are consumed because α-PRPP formation in the first reaction followed by PP_i release in the next reaction represents the loss of 2 ATP equivalents. The *de novo* pathway route of purine nucleotide synthesis has a higher requirement for energy than the salvage pathway in terms of ATP.[6]

De novo synthesis of purines begins with the transfer of the amido group of glutamine to PRPP.[6] The stereochemistry of the anomeric carbon is α and must be inverted during the reaction sequence. The first committed reaction is catalyzed by PRPP amidotransferase (ATase) forming β −phosphoribosylamine (PRA). Glycine is attached to PRA with an amide bond by GAR synthetase catalyzing the formation of glycine amide ribonucleotide (GAR). This step requires the utilization of ATP. GAR is converted by GAR transformylase (GART) using N^{10}-formyl-THF to form formylglycinamide ribonucleotide (FGAR). Consumption of ATP and glutamine and catalysis by formylglycinamidine ribonucleotide (FGAM) synthetase leads to the FGAM formation. FGAM then undergoes ring closure by consuming an additional ATP molecule to form 5-aminoimidazole ribonucleotide (AIR) catalyzed by AIR-synthase. AIR is carboxylated by AIR carboxylase forming 4-carboxy aminoimidazole ribonucleotide (CAIR). Consumption of another ATP molecule and aspartate, catalyzed by SAICAR synthase, forms N-succinyl-5-aminoimidazole-4-carboxamide ribonucleotide (SAICAR).

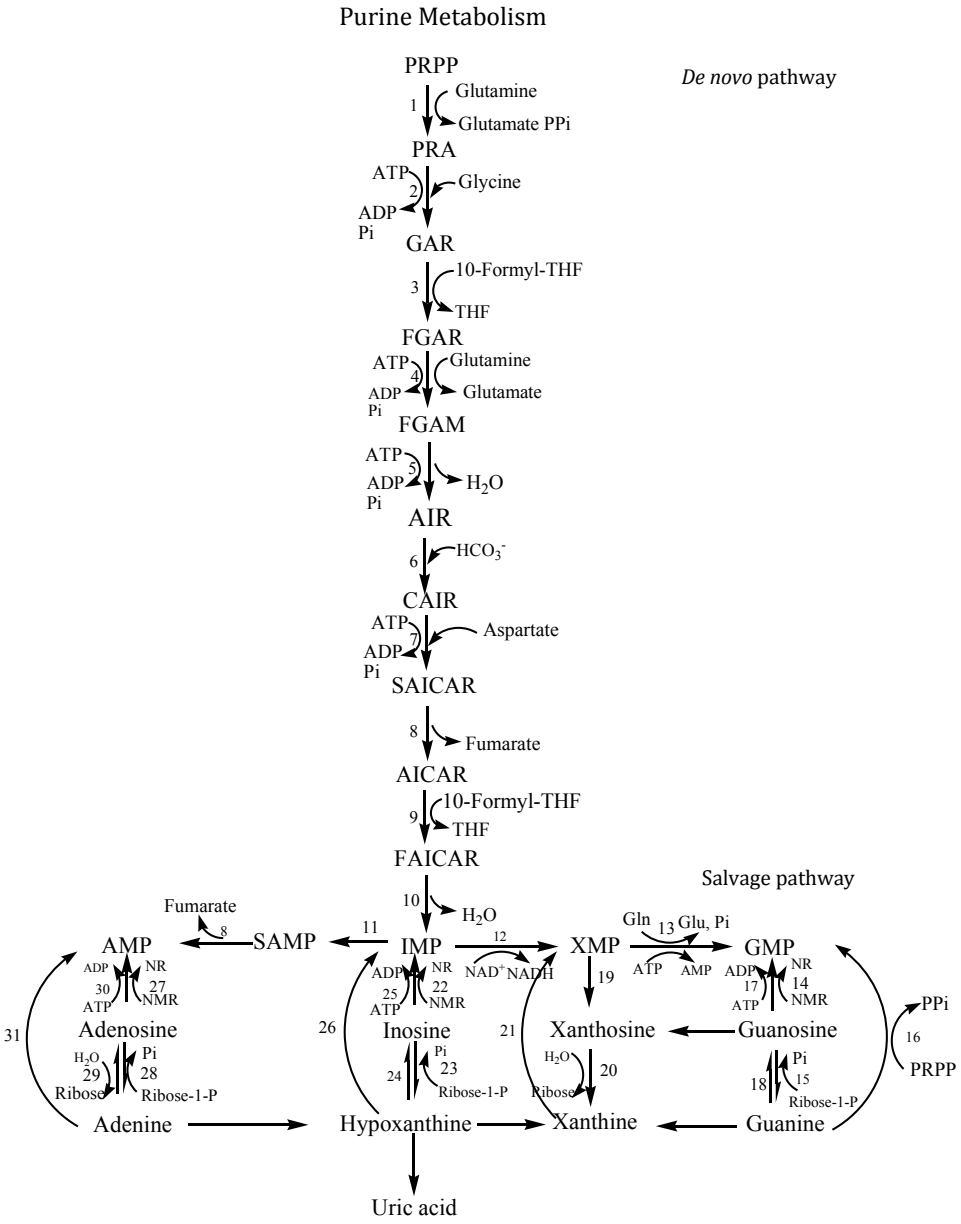


Figure 3. General schematic of purine metabolic pathways in plants. Metabolic components: PRPP- 5-phosphoribosyl-1-pyrophosphate, PRA-5-phosphoribosylamine, PPi- pyrophosphate, GAR-glycinamide ribonucleotide, THF-tetrahydrofolate, FGAR-formylglycinamide, FGAM-formylglycinamidine ribonucleotide, AIR-5-aminoimidazole ribonucleotide, CAIR-4-carboxy aminoimidazole ribonucleotide, SAICAR-N-succinyl-5-aminoimidazole-4-carboxamide ribonucleotide, AICAR-5-aminoimidazole-4-carboxamide ribonuleotide, FAICAR-5-formaminoimidazole-4-carboxamide ribonucleotide, IMP-inosine monophosphate, SAMP-adenylosuccinate, AMP-adenosine monophosphate, XMP-xanthosine monophosphate, GMP-guanosine monophosphate. Enzymes: 1. Amido phosphoribosyltransferase, 2. GAR synthetase, 3. GAR formyl transferase, 4. FGAM synthetase, 5. AIR synthetase, 6. AIR carboxylase, 7. SAICAR synthetas, 8. Adenylosuccinate lyase, 9. AICAR formyl transferase, 10. IMP cyclohydrolase, 11. SAMP synthetase, 12. IMP dehydrogenase, 13. GMP synthetase,14. non-specific nucleoside phosphoribosyltransferase, 15. Guanosine nucleosidase, 16. Guanine phosphoribosyltransferase, 17. Guanosine kinase, 18. Guanosine phosphorylase, 19. Xanthosine kinase, 20. Xanthosine nucleosidase, 21. Xanthine phosphoribosyltransferase, 22. non-specific nucleoside phosphoribosyltransferase, 23. Inosine nucleosidase, 24. Inosine phosphorylase, 25. Inosine kinase, 26. Hypoxanthine phosphoribosyltransferase, 27. non-specific nucleoside phosphoribosyltransferase, 28. Adenosine nucleosidase, 29. Adenosine phosphorylase, 30. Adenosine kinase, 31. Adenine phosphoribosyltransferase. Adapted with permission from Zrenner R, *et al.* **2006**. Annu. Rev. Plant Biol. *57*:805-36; Stasolla, C., *et al.* **2003**. J,Plant. Physiol. *160*: 1271-1295 and Moffatt B., *et al.* **2002**. The Arabidopsis Book. 1, e0018.

To build 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR), fumarate is released in a reaction catalyzed by adenylosuccinate lyase (ASL). To complete the *de novo* pathway, the final carbon of the purine ring is provided by N^{10}-formyl-THF to form 5-formaminoimidazole-4-carboxamide ribonucleotide (FAICAR). Lastly, FAICAR undergoes dehydration and ring closure to form IMP.[4]

IMP is the precursor to both AMP and GMP. IMP dehydrogenase (IMPDH) using NAD^+ as the hydrogen acceptor forms xanthosine monophosphate (XMP). Guanosine monophosphate is formed by the oxidation of IMP followed by amino group insertion provided by the amide nitrogen of glutamine. Formation of GMP is catalyzed by GMP synthetase (GMPAS). The second branch leads to adenosine monophosphate (AMP). It is synthesized by replacing the carboxyl group at C_6 with an amino group. Aspartate provides the amino group. GTP hydrolysis provides the energy to form adenylosuccinate (SAMP), which is catalyzed by SAMP synthetase. Adenylosuccinate lyase (ASL) catalyzes the removal of fumarate to form AMP.[4] Adenine, hypoxanthine, and guanine are converted into AMP, IMP, and GMP by the enzymes adenine, hypoxanthine, and guanine phosphoribosyltransferase respectively which transfer the phosphoribosyl moiety from PRPP.

The *de novo* pyrimidine nucleotide biosynthetic pathway, which is also referred to as the "orotate pathway," is usually defined as the formation of UMP from carbamoyl phosphate (CP), aspartate, and PRPP. The orotate pathway consists of six enzymatic reactions. The initial reaction catalyzed by CP synthetase (CPSase) produces CP by combination of carbonate, ATP, and an amino group from glutamine.

To form the pyrimidine ring from CP, three additional reactions are necessary: the phosphoribosyl group of PRPP is added to the pyrimidine base orotate, forming orotidine 5'-monophosphate (OMP) which is decarboxylated to make UMP, the first pyrimidine nucleotide. UMP is subsequently phosphorylated to UDP and UTP. The transfer of an amino group from glutamine to UTP by CTP synthetase leads to the synthesis of CTP. The molecular aspects of purine and pyrimidine biosynthesis and degradation are shown in Figure 3 and Figure 4 respectively.[4]

UMP synthesis leads to a decreased pyrimidine *de novo* synthesis accompanied by a stimulation of the less energy consuming salvage pathway.[4] The *de novo* pathway consists of six reactions (Figure 4). The formation of carbamoyl phosphate is the most important regulatory step in the pyrimidine pathway. Conversion of carbamoyl phosphate, catalyzed by aspartate transcarbamoylase, forms carbamoylaspartate by reaction with aspartate. Formation of the pyrimidine ring is catalyzed by dihydroorotase with addition of hydrogen ion forming dihydroorotate.

In the pyrimidine salvage pathway plant cells reutilize pyrimidine bases and nucleosides derived from the preformed nucleotides (Figure 4). Uridine, cytidine, deoxycytidine, and thymidine are converted to UMP, CMP, dCMP, dTMP nucleotides by their respective kinases.[6] Uracil is converted into UMP by uracil phosphoribosyltransferase in a reaction that transfers the phosphoribosyl moiety from PRPP to uracil.[5]

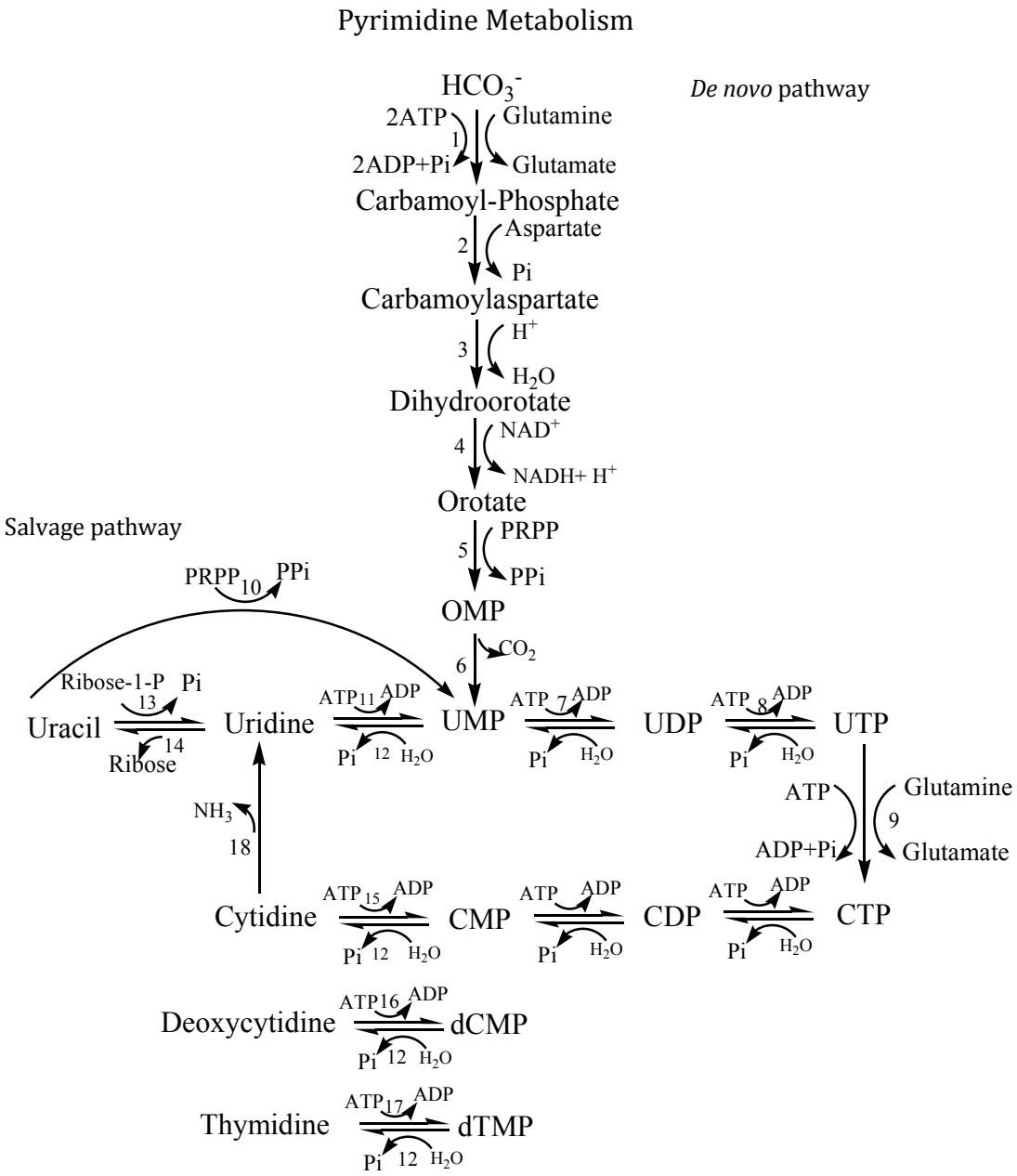


Figure 4. General schematic of pyrimidine metabolic pathways in plants. Metabolic components: ATP- adenosine triphosphate, ADP- adenosine diphosphate, P*i*- inorganic phosphate, PP*i*-pyrophosphate, OMP- orotidine 5'-monophosphate, UMP- uridine monophosphate, UDP- uridine diphosphate, UTP- uridine triphosphate, CTP- cytosine triphosphate, CDP- cytosine diphosphate, CMP- cytosine monophosphate. Enzymes: 1. Carbamoyl phosphate synthetase, 2. Aspartate transcarbamoylase, 3. Dihydroorotase, 4. Dihydroorotate dehydrogenase, 5-6. UMP synthase (orotate phosphoribosyltransferase plus orotidine-5'-phosphate decarboxylase), 7. UMP kinase, 8. Nucleoside diphosphate kinase, 9. CTP synthetase, 10. Uracil phosphoribosyltransferase, 11. Uridine kinase, 12. Non-specific nucleoside phosphotransferase, 13. Uridine phosphorylase, 14. Uridine nucleosidase, 15. Cytidine kinase, 16. Deoxycytidine kinase, 17. Thymidine kinase, 18. Cytidine deaminase. Adapted with permission from Zrenner R, *et al.* **2006**. Annu. Rev. Plant Biol. *57*:805-36; Stasolla, C., *et al.* **2003**. J,Plant. Physiol. *160*: 1271-1295 and Moffatt B., *et al.* **2002**. The Arabidopsis Book. *1*, e0018.

A central step in nucleoside and nucleobase salvage pathways is the hydrolysis of nucleosides to their nucleobases. In plants this is solely accomplished by nucleosidases, also known as nucleoside hydrolases.[7] Nucleoside hydrolases in parasitic protozoans have been studied extensively in recent years as potential targets for chemotherapeutic intervention.[7]

3. Nucleoside Hydrolases

Nucleoside hydrolases are a group of enzymes that catalyze the hydrolysis of the glycosidic bond between the nitrogenous base and pentose sugar of certain nucleosides. An example of this is adenosine nucleosidase (Figure 5). Various bacteria, yeast, parasitic protozoa, insects, fish, and plants contain homologues of nucleoside hydrolase, while they are absent in mammals.[5] In the nucleoside salvage pathways used by most of the protozoan parasites, a vital role is played by these enzymes, since the parasites are unable to synthesize purines *de novo.* In all organisms, the N-ribosidic bonds present in nucleosides and nucleotides should be hydrolysable because it is important for their regular metabolic function. In this pathway, the nucleoside hydrolases catalyzes the hydrolysis of the selected nucleosides, allowing recycling of the purine bases and ribose.[8]

Characterized nucleoside hydrolases are divided into four classes based on their substrate specificity. The first major class consists of nonspecific hydrolases acting on both purine and pyrimidine ribonucleosides. Examples include inosine-uridine preferring nucleoside hydrolase (IU-NHs) from *Crithidia fasciculata,*

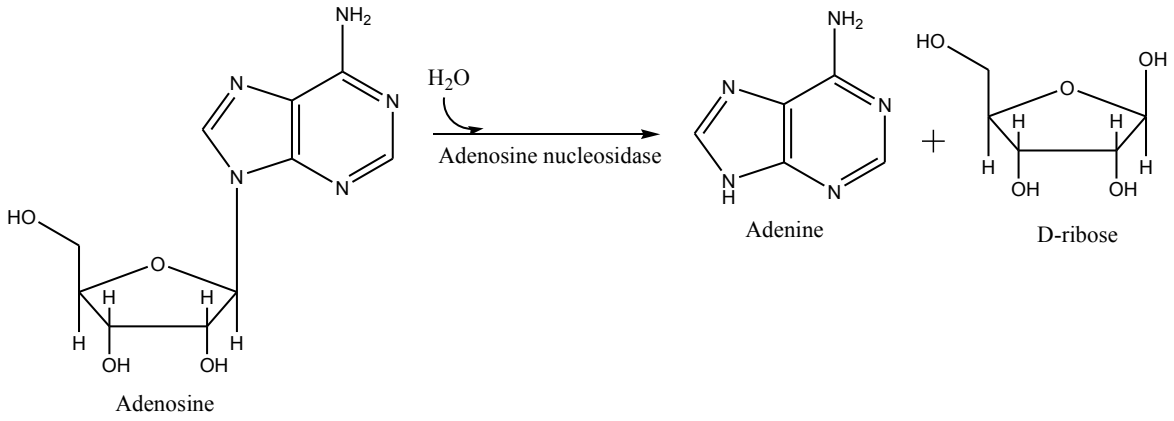


Figure 5. Reaction catalyzed by adenosine nucleosidase.

Leishmania major, Leishmania donovani, along with several types of bacteria.[2,8] The second class consists of enzymes that are specific for purine nucleosides, such as inosine-adenosine-guanosine preferring nucleoside hydrolases (IAG-NHs) found in *Trypanosoma brucei brucei, Trypanosoma vivax,* and *Ochrobactrum anthropic.*[2,8] The third class consists of the 6-oxo purine specific inosine-guanosine preferring nucleoside hydrolases (IG-NHs) found in *Crithidia fasciculata* and *Trypanosoma* species.[8] Another example is the calcium-stimulated guanosine-inosine nucleosidase from yellow lupin (*Lupinus luteus*) in which the ribosides of 6-oxopurines are the preferred substrates.[9] With respect to its substrate specificity, the lupin enzyme resembles the guanosine-inosine nucleosidase from *Crithidia fasciculata* for which these two nucleosides were the preferred substrates.[9] Also the recently purified nucleoside hydrolase from *Caenorhabditis elegans,* characterized by a low catalytic capacity compared to the parasitic enzymes, belongs to this class.[10] Finally, the fourth class is defined by the pyrimidine specific cytidine-

uridine preferring nucleoside hydrolases (CU-NHs) found in *Escherichia coli* and yeast. As more nucleoside hydrolases are characterized, the boundaries of the existing classification will probably fade.

4. Nucleoside Hydrolase from Yellow Lupin (*Lupinus luteus*)

Yellow lupin seeds were once a common food of the Mediterranean basin and Latin America. Today they are eaten as a snack food. Yellow lupin is an herb whose seeds and above ground parts are also used to make medicine. Yellow lupin is taken for urinary tract disorders, worm infections, fluid retention, and sometimes applied directly to the skin for ulcers. Yellow lupin seeds are a rich source of enzymes involved in the metabolism of nucleotides and nucleosides.[9] Inosine nucleosidase, an enzyme that hydrolyzes inosine to hypoxanthine and ribose, has been purified from yellow lupin.[11] The molecular weight of the enzyme is 62,000 Da and it exhibits optimum activity at pH 8.0.[11] Another nucleoside hydrolase, adenosine nucleosidase catalyzes the hydrolysis of adenosine to ribose and adenine.[12] It exists as a dimer with a native molecular weight of 72,000 Da and pH optimum of 7.5.[12] Inosine nucleosidase and adenosine nucleosidase complement each other with respect to their substrate specificity.[12] Adenosine is the best substrate for adenosine nucleosidase, followed by guanosine and inosine. For inosine nucleosidase inosine is the best substrate.[12] A third nucleoside hydrolase guanosine-inosine nucleosidase has recently been isolated from yellow lupin seeds.[9] It is a monomer with a molecular weight of 80,000 Da and pH optimum of 4.7-5.5.[9]

The K_m value determined for inosine is 65 μM.[9] This compares to a K_m value for adenosine of 4.8 μM for yellow lupin adenosine nucleosidase.[12]

5. Nucleoside Hydrolases from Other Plants

Nucleoside hydrolases have been isolated from other plant sources, such as barley leaves[13], coffee (*Coffea Arabica)* [14], artichoke shoots, wheat germ[15], spinach beet (*Beta vulgaris L.)*[17], tea leaves[16], tomato roots and leaves[18], and *Arabidopsis thaliana*[7]. Nucleoside hydrolases obtained from different sources have various molecular weights, subunit structures, and pH optima (Table 1).

6. Kinetic Analysis of Adenosine and Inosine Nucleosidase from Yellow Lupin

Enzyme kinetics is the study of the rates of enzyme-catalyzed reactions, which addresses the specificities and catalytic mechanisms of enzymes.[3] Kinetic analysis determines the maximum reaction velocity that the enzyme can attain and its binding affinities for substrates and inhibitors. Enzyme kinetics reveals how the rates of reactions are affected by variations in experimental conditions or in the concentration of enzyme or substrate. Chemical kinetics examines the relationship between the amount of product (P) formed in a unit of time (Δ [P]/Δt) and the conditions under which the reaction takes place.[3] The basis of most kinetic measurements is the observation that the rate, or velocity, of a reaction varies directly with concentration of each reactant, whether substrate or catalyst. At low concentrations of the substrate (S), velocity is proportional to [S], as expected for a first order reaction. However, velocity does not continually increase proportionally

Table 1. Sources, molecular weights, and pH optima of purine nucleosidases from various plant sources.

Source	Enzyme	Molecular Weight (Da)	pH Optimum
Barley leaves	Adenosine nucleosidase	66,000 dimer	4.7-5.4
Coffea Arabica young leaves	Adenosine nucleosidase	72,000 dimer	6.0
Wheat germ cells	Adenosine nucleosidase	59,000	4.7
Tea leaves	Adenosine nucleosidase	68,000	4.0-4.5
Tomato roots and leaves	Adenosine nucleosidase	68,000	5.0-6.0
Yellow lupin	Adenosine nucleosidase	72,000 dimer	7.5*
Yellow lupin	Inosine nucleosidase	62,000	8.0
Yellow lupin	Guanosine-inosine nucleosidase	80,000 monomer	4.7-5.5

*The adenosine nucleosidase isolated by Guranowski from the cotyledons of 5-day yellow lupin seedlings according to the procedure of Abusamhadneh and coworkers (2000) had an acidic pH optimum (between pH 4.7 and 5.5) rather than the optimum at pH 7.5.[9]

as [S] increases, but instead levels off for enzyme-catalyzed reactions. At high [S], velocity becomes independent of [S] and approaches a maximal limit, called V_{max}.[3]

The fundamental equation of enzyme kinetics used to calculate K_m and V_{max} is the Michaelis-Menten equation:

$$\text{Velocity, } \upsilon = \frac{V_{\max}[S]}{K_m + [S]}$$

The Michaelis constants, K_m, is the substrate concentration that gives a velocity equal to one-half the maximal velocity.[3] K_m is often a measure of the affinity of enzyme for substrate. The lower the K_m, the more tightly the substrate is bound. The kinetic parameters of an enzymatic reaction can provide valuable information about the specificity and mechanism of the reaction. Nucleoside hydrolases from various sources have various kinetic properties (Table 2).

Table 2. Comparison of the kinetic properties of various nucleoside hydrolases.

	IU-NH *Lm*[21] K_m (μM)	IU-NH *Cf*[20] K_m (μM)	IG-NH *Cf*[22] K_m (μM)	IAG-NH *Tv*[1] K_m (μM)	IAG-NH *Tbb*[19] K_m (μM)
Adenosine	185	460	106	8.5	15
Inosine	445	380	16	5.4	18
Guanosine	140	420	77	3.8	46
Cytidine	422	4700	3700	925	106
Uridine	234	1220		586	102
p-NPR[a]	185	110		257	562

[a] *p*-Nitrophenyl-β-D-ribofuranoside

7. Structure

Amino acid sequence alignments of nucleoside hydrolases from various organisms highlight a recurring N-terminal DXDXXXDD motif as a fingerprint of nucleoside hydrolases.[23] On the basis of their sequence identity and the conservation of key amino acid residues, the nucleoside hydrolase sequences in the various databases appear to fall into three distinct groups (Figure 6).[23] The conserved histidine of Group I NH appears to fulfill the function of proton donor, postulated to be essential for catalysis.[24] The C-terminal regions of enzymes belonging to group I contain the conserved histidine in their nucleobase-binding pocket. In group II enzymes, the histidine is replaced by a tryptophan. In group III, various amino acids such as a cysteine, proline, or an asparagine can occur in this position. The functional data suggest that the group I enzymes include the IU- and CU-NHs, group II comprises the IAG-NHs, and group III includes the IG-NHs hydrolases.[23] The N-terminal region, which contains the aspartate cluster mentioned above is involved in binding of the active site Ca^{2+} ion and activation of the water nucleophile, and is conserved in all nucleoside hydrolase proteins.[23] In both regions, the protein residues involved in interactions with the calcium ion or ribose ring are highlighted (Figure 6).[23]

Structural and functional similarities and differences between two representatives of the superfamily of nucleosidase hydrolases from *Crithidia fasciculata* and *Trypanosoma vivax* will be discussed. The crystal structures of the trypanosome parasite, *Crithidia fasciculata,* inosine-uridine nucleoside hydrolase

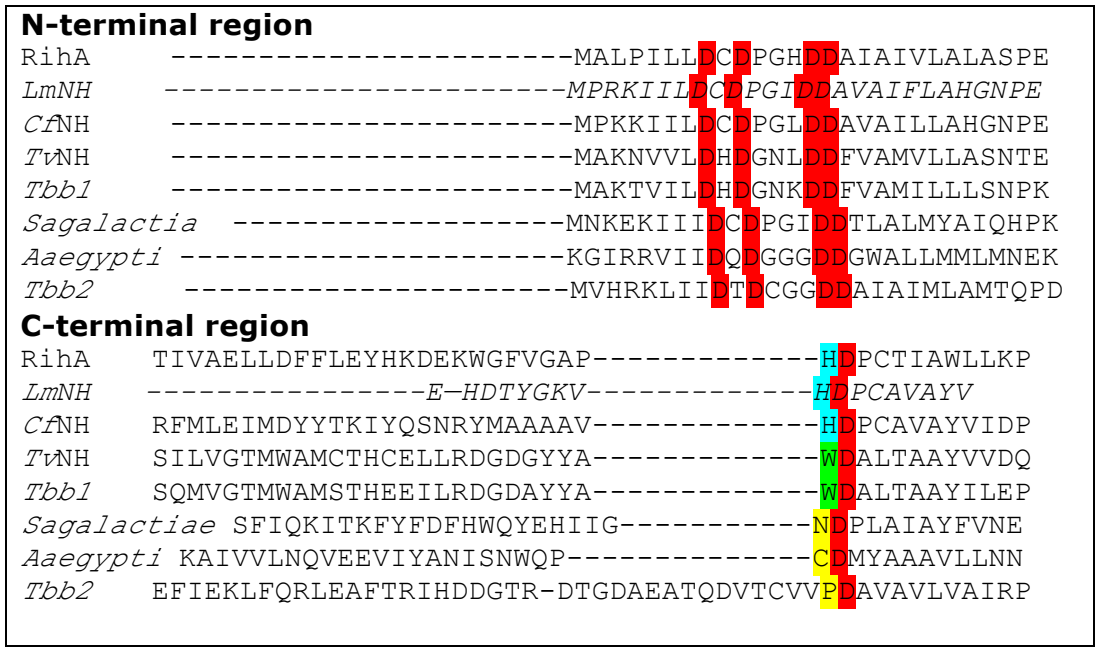


Figure 6. Amino acid sequence alignment for 8 selected putative NH proteins: RihA, the CU-specific NH from *E.coli*; *Cf*NH, the IU-specific NH from *Crithidia fasciculata*; *Tv*NH, the IAG-specific NH from *Trypanosoma vivax*; *Tbb1,* the IAG- specific NH from *Trypanosoma brucei brucei*; *Sagalactiae*, the NH from *S. Agalactiae*; *Aaegypti*, the NH from *Aedes aegypti* and *Tbb2* the second NH from *Trypanosoma brucei brucei*. Top panel, the N-terminal region with the aspartate cluster highlighted in red. Bottom, the C-terminal region of the nucleobase-binding pocket. The conserved histidine of group I NHs is highlighted in blue, the conserved tryptophan of group II in green and the residue at this position for group III in yellow. The conserved aspartate residue is highlighted in red. Reprinted with permission from Giabbai, Barbara; Degano, Massimo. *Structure.* **2004**, *12,* 739-749.

(IU-NH) was solved along with the inosine-adenosine-guanosine-nucleoside hydrolase (IAG-specific) enzyme from *Trypanosoma vivax.*

IU-NH crystallized as homotetramers and IAG-NH as a homodimer (Figure 7 a,b). Both subgroups have similar monomeric subunits, which are common in architecture, topology, and have a single globular domain (Figure 7 c, d).[8] At the C-terminal end of the central sheet is the location of the active site. Two loops (loop I and loop II in Figure 7 c, d) are positioned on each side of the active site. The subunits of the IU-NH and IAG-NH are arranged in different quaternary architectures having different subunit-subunit interfaces.[8] Each subunit consists of ten β-strands, 12 α-helices, and three short 3_{10} helices. The first six strands of the β sheet arrangement resemble a dinucleotide binding motif, also called Rossmann fold. As is demonstrated in Figure 7 (e, f), the α/β core of the monomer is composed of an eight-stranded β sheet (seven parallel and one antiparallel strand) and a few surrounding α helices.[8]

In each subunit of nucleosidase hydrolase, one deep narrow active site is present.[8] Ca^{2+} ion is tightly bound at the bottom of the active site. In *Crithidia fasciculata* the nucleobase-binding pocket is composed of Phe167, Ile81, Tyr229, Tyr225, and His82 (Figure 8 a, c). However, there are no crystal structures of IU-NH in complex with a genuine heterocyclic nucleobase.[8] In *Trypanosoma vivax,* an octacoordinated calcium metal is chelated through a conserved network of interactions connecting the main chain carbonyl oxygen of Thr137, the side chain oxygens of Asp261, Asp15, and Asp10, and three water molecules (Figure 8 b, d).[8]

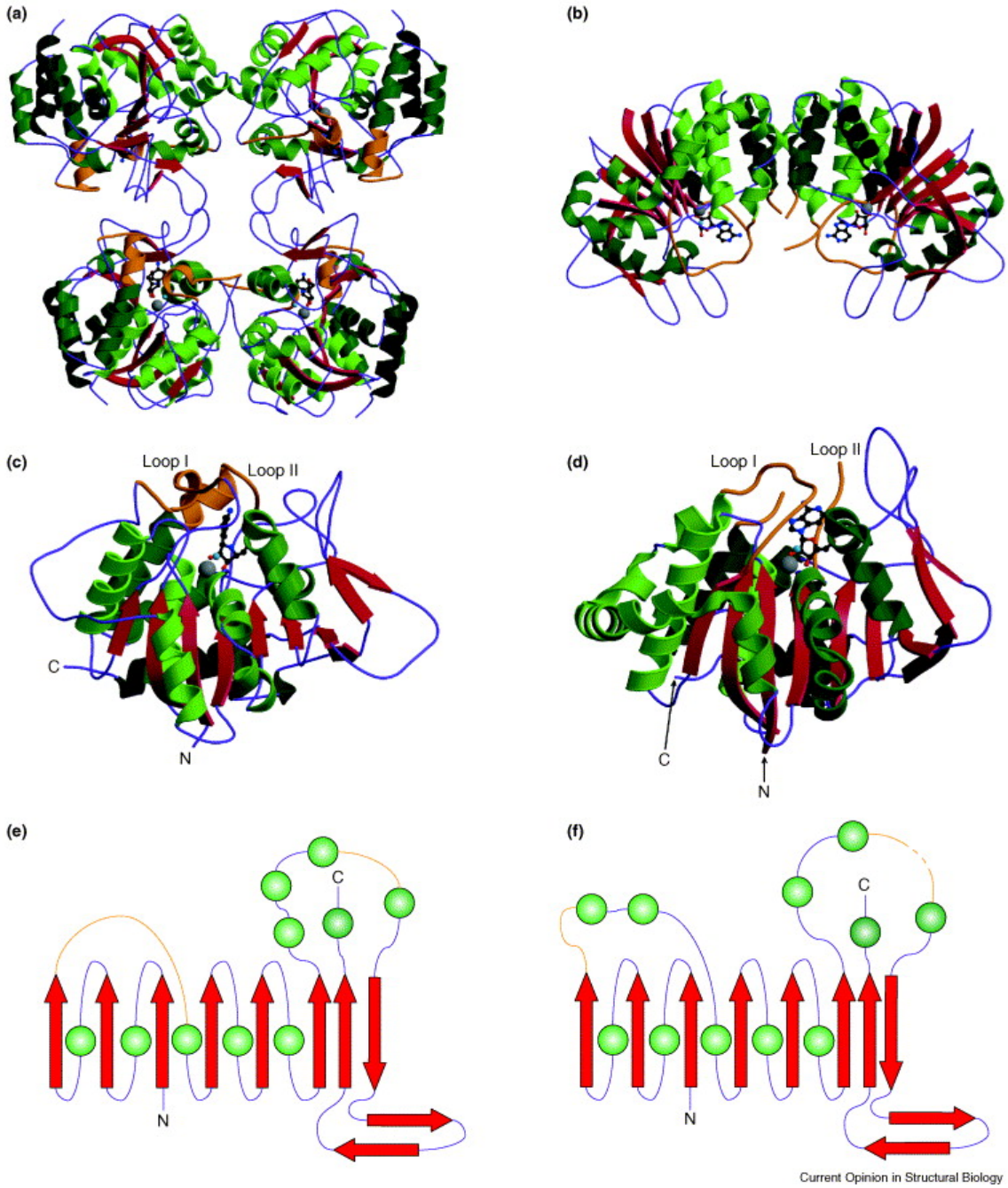


Figure 7. Comparison of the quaternary structures, tertiary structures, and topologies of *Crithidia fasciculata* IU-NH (a,c,e) and *Trypanosoma vivax* IAG-NH(b,d,f). α Helices are represented in green and β sheet strands are in red. The Ca^{2+} ion and the nucleophilic water molecule are in the active site pockets and represented as grey and blue spheres. In orange color are two flexible loops in the vicinity of the active site. Reprinted with permission from Versees, W; Steyaert, J. *Current Opinion in Structural Biology.* **2003**, *13*, 731-738.

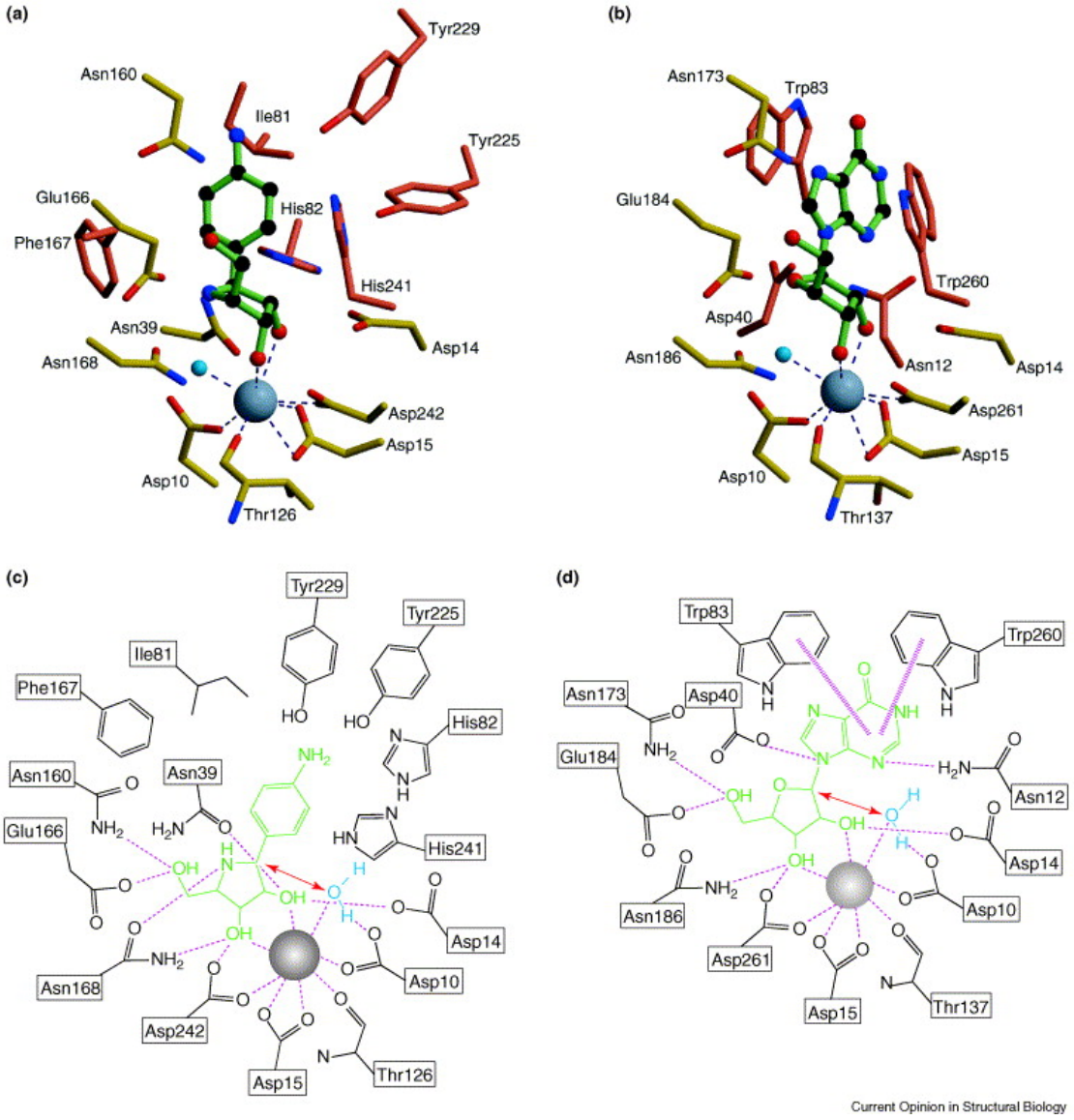


Figure 8. Comparison of active sites of IU-NH *Crithidia fasciculata* (a, c) and IAG-NH *Trypanosoma vivax* (b, d). The enzyme bound ligands are represented in green. The nucleophilic water molecule and Ca^{2+} ion are represented as blue and grey spheres, respectively. Schematic representation of the intermolecular interactions is in (c, d) part. Magenta dotted lines indicate possible interactions. The pink arrows show the nucleophile and the electrophilic center of the substitution reaction. In the *Trypanosoma vivax* NH-inosine complex, the nucleobase is stacked between the side chains of Trp83 and Trp260. Catalytically relevant interactions in the nucleobase-binding pocket of *Crithidia fasciculata* does not allow a straightforward interpretation because of the poor analogy of the *p*-aminophenyl group with nucleobases. Reprinted with permission from Versees, W; Steyaert, J. *Current Opinion in Structural Biology.* **2003**, *13*, 731-738.

Two calcium-bound water molecules are replaced by the 2'-and 3'-hydroxyl groups of the sugar in the complex of active site. The remaining calcium-chelated water molecule interacts with an Asp10 and is positioned to attack the anomeric carbon. The network of interactions involving the Ca^{2+} ion, the 2'-, 3'- and 5' -hydroxyls of the sugar, and conserved residues Asn173, Glu184, Asn186, Asp261, and Asp14 establishes the specificity of nucleosidase hydrolases for ribose (Figure 8 b, d).[8] A mutagenesis analysis demonstrated that π-π stacking interaction between a purine base and two tryptophans (Trp83 and Trp260) is responsible for the purine specificity of the IAG-NH.[8] Few additional interactions are formed between the enzyme and the purine base where only Asn12 and Asp40 are located within interaction distance.[8] It can be concluded that the aromatic stacking complex is responsible for the purine specificity in IAG-NH and is not present in the IU-NH.

8. Bio-analytical Techniques

Proteomics is defined as the large-scale study of proteins, in particular their structures and functions.[25] One area of proteomics is the identification of newly purified proteins by comparing their sequences to previously identified proteins.[25] A critical step in the identification of new proteins by mass spectrometry is the conversion of proteins to peptides by trypsin. Mass spectrometry of whole proteins is less sensitive than peptide MS and the mass of the intact protein by itself is insufficient for identification.[25] Trypsin is a serine protease that cleaves peptide bonds at the carbonyl side of lysine and arginine residues, creating peptides both in the preferred mass range for MS sequencing and with a basic residue at the carboxyl

terminus of the peptide, producing an interpretable peptide fragmentation mass spectra pattern.[25] After a chromatographic separation of the complex mixture of peptides derived from a tryptic digestion, mass spectra are collected and software is used to find proteins containing the peptides identified. In complex proteomic samples, protein identification is performed by searching databases with search engines such as BLAST, Mascot, Sequest, or Phenyx.[25]

9. Tryptic Digestion

Protein identification follows two different workflows depending on the approach (Figure 9). In-gel digestion coupled with mass spectrometric analysis is a powerful tool for the identification and characterization of proteins.[26] During gel electrophoresis, proteins are separated in one or two dimensions (1-D/ 2-D) and an enzymatic digestion is performed in-gel, which is a time-consuming process.[27] The proteins are degraded enzymatically to peptides by trypsin or other proteases. Excised and destained protein separated bands (spots) are in-gel reduced, alkylated, and degraded by trypsin. Denaturation and reduction is carried out by a combination of heat and 1,4-dithiothreitol (DTT) reagent. DTT is a strong reducing agent that reduces the disulfide bonds and prevents inter- and intra-molecular disulfide formation between cysteines in the protein to avoid renaturation of the protein.[25] Alkylation of cysteine is necessary to prevent the renaturation of the protein, with the most common alkylating reagent, being iodoacetamide (IAM).[25]

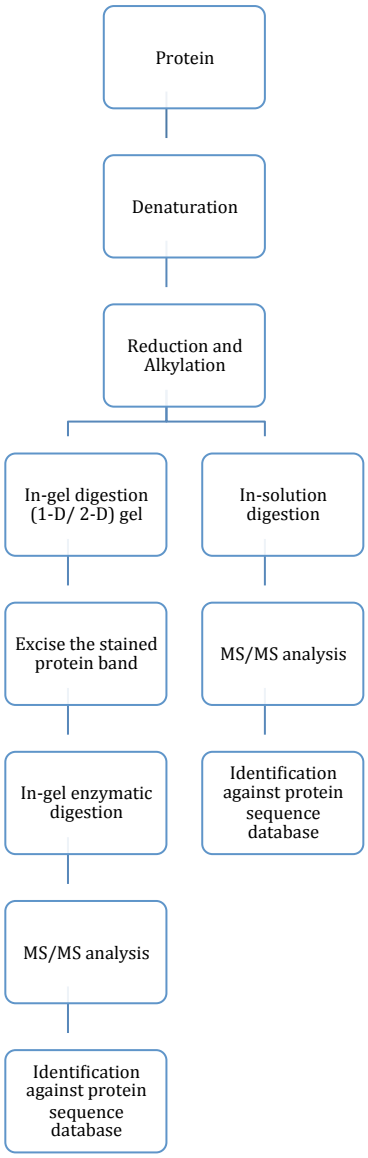


Figure 9. Workflow of in-gel (left) and in-solution (right) digestion and subsequent MS/MS analysis of a protein sample. Redrawn from Hustoft, Hanne, *et al.* **2012**, *Integrative Proteomics.* 73-89.

In the in-solution approach the proteins or peptides are separated chromatographically typically by reverse phase HPLC. The proteins are reduced, alkylated, and digested in-solution using a serine protease.[25] Acquired MS/MS spectrum is stored for matching against protein sequence databases.[28]

10. Peptide Fragmentation and Analysis by MS/MS

Mass spectrometry is an analytical technique that measures the molecular weight of molecules based upon the motion of a charged particle in an electric or magnetic field.[29] The sample molecules are converted into ions in the gas phase and separated according to their mass to charge ratio (*m/z*). Positive and negatively charged ions can be formed. Mass spectrometry can be used in two different ways for identification: it can be used to indicate the composition of a particular analyte or the structure of a particular analyte.[29] In the first example, the instrument is run in MS mode, where the actual mass of the analyte is measured. This mass is indicative of the composition of the sample. For example, peptide mass fingerprinting uses the mass of each of the tryptic peptides in the spectrum for a database search. In the second example, the instrument is used in MS/MS mode to obtain structural information relating to the analyte. In this mode, the mass of the analyte is determined and the analyte fragmented within the mass spectrometer to yield structural information. This method is common for the determination of small molecule structure and peptide sequence.[29]

Fragmentation is generated by collision-induced dissociation (CID) within the mass spectrometer, which is called tandem mass spectrometry (MS/MS). Three

types of MS/MS experiments are performed routinely within proteomics. The analyzer region can be regarded in three parts: MS1-product ion scan, collision cell-parent ion scan, and MS2-neutral loss scan.[29] In a product ion scan, the first part of an analyzer, MS1, is used to specifically select the ion of interest, the precursor ion (i.e. peptide). The precursor ion is allowed into the collision cell where it will undergo CID, where the peptide precursor ions collide with molecules of the collision gas (argon or nitrogen) causing the precursor ion to fragment yielding a distribution of fragment ions, or product ions. The resultant product ions are resolved by the third part of the analyzer, MS2, before detection at the detector producing a product ion spectrum.[29]

As the result of the low energy collision fragmentation, the peptide precursor ion fragments yield a distribution of product ions in a complimentary ion series forming a ladder, which indicates peptide sequence.[29] The accepted nomenclature for fragment ions was first proposed by Roepstorff and Fohlmann and later modified (Figure 10).[30] Fragments will only be detected if they carry at least one charge. If this charge is retained on the N-terminal fragment, the ion is classified as either *a, b,* or *c.* If the charge is retained on the C-terminal, the ion type is either *x, y,* or *z.*

The product ion spectrum can be used to identify a protein by manual determination of the peptide sequence. The determined sequence can then be used to search the sequence databases or by automatic correlation of the native product ion spectra with the sequence database. After the mass spectrometer analysis,

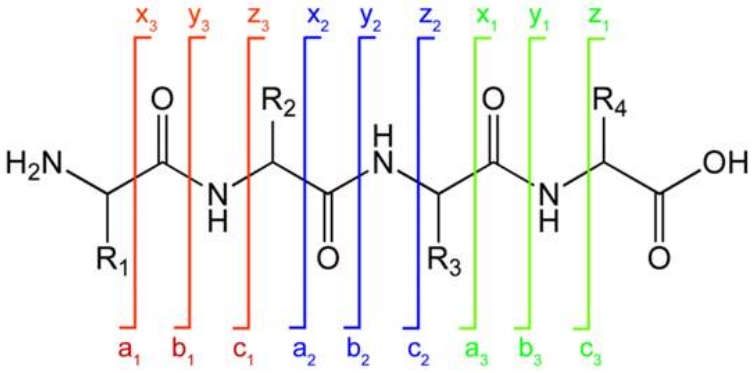


Figure 10. Schematic diagram of daughter ion nomenclature, adapted from Roepstoff and Fohlmann[30]. A subscript indicates the number of residues in the fragment.

data are analyzed in order to obtain information regarding the amino acid sequence of the peptide and further used for protein identification and characterization.[31]

11. Research Goal

The purpose of this study is purification and determination of amino acid sequences of nucleoside hydrolases from various seeds using a combination of tryptic digestion with MS/MS characterization and comparison to the known nucleoside hydrolases. To accomplish this goal, methods to digest proteins must be developed. In addition, parameters for MS/MS analysis using a mass spectrometer must be determined.

CHAPTER II

MATERIALS AND METHODS

1. Equipment and Instrumentation

Protein purification was carried out by column chromatography on a GE Healthcare AKTA Fast Protein Liquid Chromatography (FPLC) system. The AKTA FPLC system was equipped with a variety of columns including a Mono Q (26/10) ion exchange column, Sephacryl S100 (26/60) size exclusion column, and Bio-Rad CHT hydroxyapatite column. UV-Vis spectroscopy was carried out on a Hitachi U-2900 spectrophotometer. Polyacrylamide gel electrophoresis was run on a Hoeffer SE250 vertical electrophoresis system. Enzyme activities were determined on a Dionex Ultimate 3000 HPLC equipped with a thermostatted autosampler, column oven, and UV detector. An Agilent 3100 OFFGel Fractionator separated proteins according to their isoelectric points. Peptide sequencing was carried out on a Waters MS/MS system using BLAST software. PEAKS software for analysis of MS/MS data was obtained from Bioinformatics Solutions Inc.

2. Materials and Reagents

Yellow lupin (*Lupinus luteus)* seeds were obtained from B and T World Seeds Sarl, France. Tris (hydroxymethyl) aminomethane (Tris buffer), Millipore centrifugal concentrators, and Lonza 15% SDS-PAGE gels were purchased from Fisher Scientific. Protamine sulfate, polyvinylpolypyrrolidone (PVPP), dithiothreitol (DTT), ω −aminohexyl-agarose for column chromatography, protease inhibitor cocktail for plant cell and tissue extracts, purine nucleosides, and purine bases were

obtained from Sigma-Aldrich. Laemmli sample buffer, 10X Tris/Glycine buffer, Precision Plus molecular weight marker, protein assay dye reagent and protein standard II were obtained from Bio-Rad. GelCode Blue Safe Protein Stain for SDS-PAGE gels was obtained from Thermo Scientific. Mass spectrometry grade trypsin was purchased from G-Biosciences. All other chemicals were of reagent grade.

3. Yellow Lupin Seeds Germination

Yellow lupin seeds (100.0 g) were surface sterilized with 10% bleach for 8 minutes. Seeds were washed several times with tap water and placed on moistened paper towels to germinate in a closed container at room temperature. Four days after germination, the seed coat was removed and the cotyledons were homogenized in 500 mL of 50 mM Tris buffer pH 7.2 containing 0.15 g of DTT, 5% (w/v) polyvinylpolypyrrolidone, 500 μL of protease inhibitor cocktail, and 2% (w/v) protamine sulfate to precipitate nucleic acids to prepare the initial extract.

4. Purification of Adenosine Nucleosidase

Adenosine nucleosidase was purified by a modification of the method of Abusamhadneh[4]. The initial extract was centrifuged at 20,000*x g* for 30 min. The supernatant was removed and heated to 55 °C for 5 min, cooled on ice to 4 °C, and centrifuged for 30 min at 20,000*x g*.

After centrifugation, proteins in the supernatant (470 mL) were precipitated by addition of ammonium sulfate. Solid ammonium sulfate (77.08 g) was added to the supernatant to give 30% saturation at 4°C, stirred for 2 hours and centrifuged at 4°C at 20,000*x g* for 30 min. The precipitate was discarded and the supernatant

saturation was increased to 60% by adding 90.5 g of ammonium sulfate. The solution was stirred for 2 hours and then centrifuged at 20,000*x g* for 30 min and the supernatant removed. The pellet was resuspended in 50 mM Tris buffer at pH 7.2 and dialyzed overnight against 500 mL of 10 mM Tris pH 7.2 buffer. The buffer was changed 3 times.

The dialyzed sample was applied to a Mono Q (26/10) FPLC ion-exchange column equilibrated against 10 mM Tris buffer pH 7.2. The column was washed with 150 mL of 10 mM Tris pH 7.2 buffer and the enzyme was eluted with a 500 mL linear gradient of 0-1 M NaCl in the same buffer. Collected fractions were monitored for absorbance at 280 nm for the presence of protein and assayed for activity by reducing sugar assay using adenosine as the substrate. Activity containing fractions were pooled, concentrated, and dialyzed against 10 mM potassium phosphate pH 6.8.

The sample was next applied to Bio-Rad CHT hydroxyapatite column equilibrated against the same buffer used for dialysis. The column was washed with 30 mL of 10 mM potassium phosphate pH 6.8 and eluted with a linear gradient 110 mL of 10- 400 mM potassium phosphate pH 6.8. Fractions were assayed for protein and activity as described above. Activity containing fractions were pooled and concentrated for the next step.

The concentrated enzyme solution was applied to a size exclusion FPLC column (Sephacryl S-100 26/60). The column was eluted with 1L of 100 mM

potassium phosphate with 300 mM NaCl pH 7.0. The fractions were assayed for activity and fractions containing active enzyme were pooled and concentrated.

5. Reducing Sugar Assay

One mL of adenosine reaction mixture (1 mM adenosine in 50 mM Tris buffer pH 7.2) was added to a clean test tube and 100 µL of enzyme-containing solution was added. The sample was incubated for 4 hours at 32 ^{0}C in a water bath. Then, 300 µL of copper reagent (4% Na_2CO_3, 1.6% glycine, 0.045% $CuSO_4 \cdot H_2O$ in 400 mL H_2O) and 300 µL of neocuproine solution (0.12% neocuproine dissolved in 400 mL water, pH 3.0) were added to stop the reaction. Samples were incubated for 7 minutes at 95 ^{0}C. After cooling, the absorbance was measured at 450 nm. The ribose amount was determined by comparison to a standard curve (Figure 11).

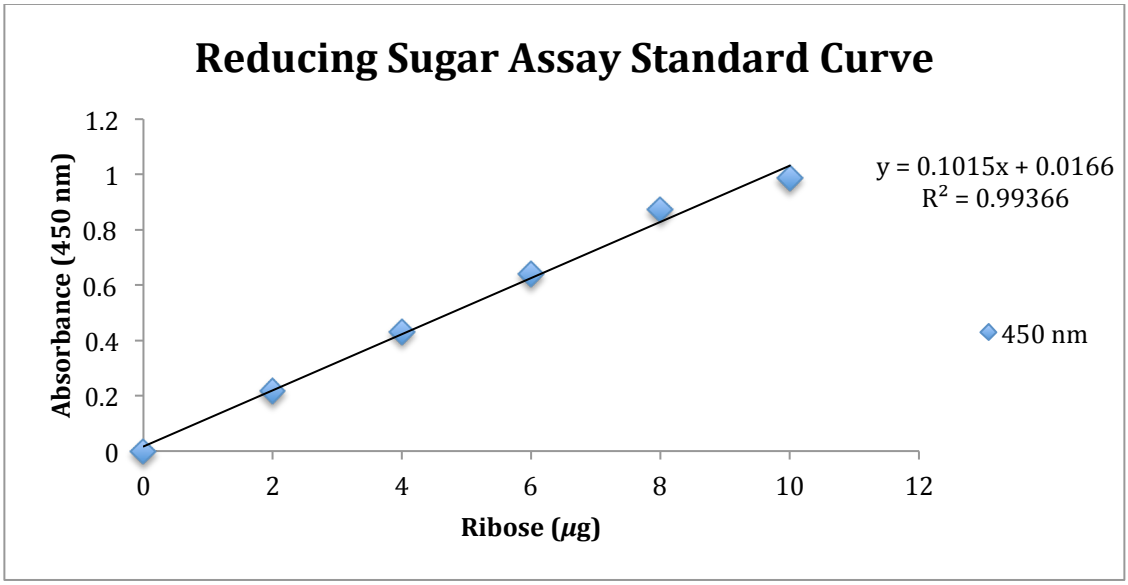


Figure 11. Standard curve for reducing sugar assay relating absorbance at 450 nm to amount of ribose.

6. HPLC Analysis of Enzyme Activity

One mL of the reaction mixture (1 mM adenosine or inosine in 50 mM Tris buffer pH 7.2) was placed in an HPLC vial and 100 µL of enzyme sample was added. The solution was incubated for 2 hours at 32 ^{0}C in a water bath. The amount of nucleoside and base present was determined by measuring the absorbance at 254 nm of the effluent from a Phenomenex Hypersil C_{18} column eluted with 98% 10 mM ammonium acetate pH 5.2 and 2% methanol. The difference in extinction coefficients between the various nucleosides and bases was calculated from a series of stock solutions. Activity was calculated by dividing previously calculated amount of base produced by the reaction time.

Additional enzymes were analyzed representing nucleoside hydrolases from *E. coli,* soybean, and Alaska pea.

7. Determination of Enzyme Purity by 1-D Gel Electrophoresis

Electrophoresis buffer was prepared by diluting 10x Tris/glycine/SDS solution (30 mL) to 300 mL with water. Sample buffer was prepared by adding 50 µL of β −mercaptoethanol to 950 µL of Bio-Rad Laemmli sample buffer. Protein sample (10 µL) was added to 10 µL of Laemmli buffer and was centrifuged for 10 sec. The sample was heated at 95 ^{0}C for 5 min and centrifuged for 10 sec. The sample was loaded onto a 15% SDS-PAGE gel along with Precision Plus Protein Unstained Markers. The gel was electrophoresed at a constant current of 30 mA for about 45 minutes. After completion the gel was carefully removed from its plastic holders, washed with distilled water 3 times for 5 min each, and then covered with

GelCode Blue Safe Protein Stain. After 4 hours the dye was removed and the gel was washed with distilled water. A calibration curve was constructed using the Precision Plus Protein molecular weight markers. The molecular weight of the sample subunit(s) was determined by comparison to the calibration curve. Figure 12.

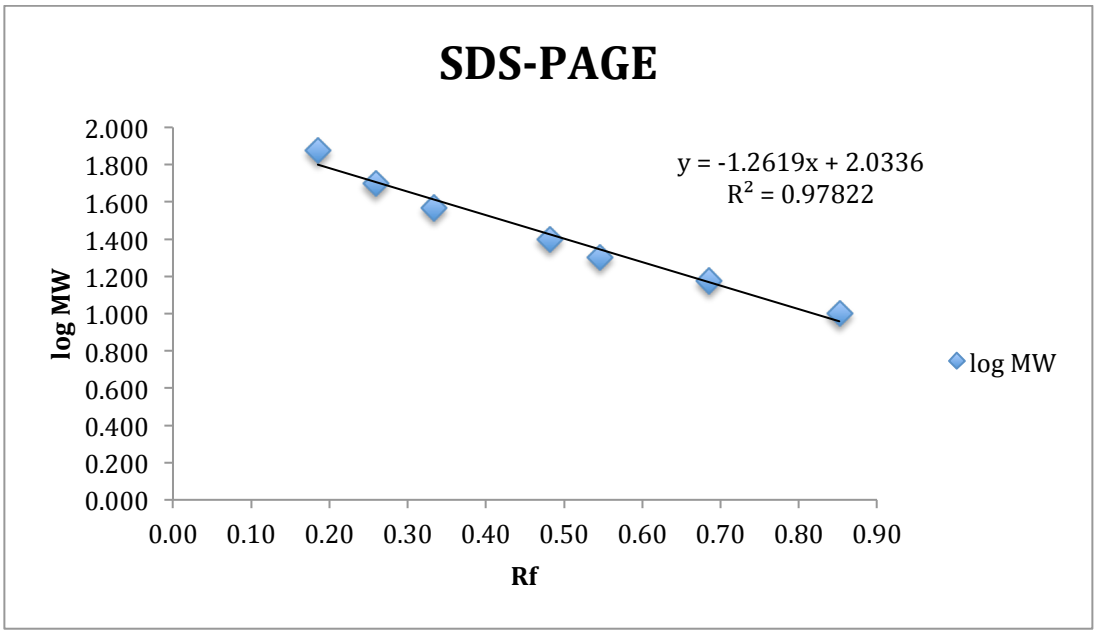


Figure 12. SDS-PAGE Precision Plus Protein standard curve.

8. Determination of Protein Concentration by Bio-Rad Protein Assay

The assay mixture consisted of standard, water, and dye (200 µL) combined to a total volume of 1000 µL. The total volume of standard or sample and water was 800 µL. After addition of the dye, the sample was left to react for 5 min at room temperature. The absorbance was measured at 595 nm and the amount of protein

in the sample determined by comparison to a standard curve. The standard used was bovine albumin (1.53 mg/mL), Figure 13.

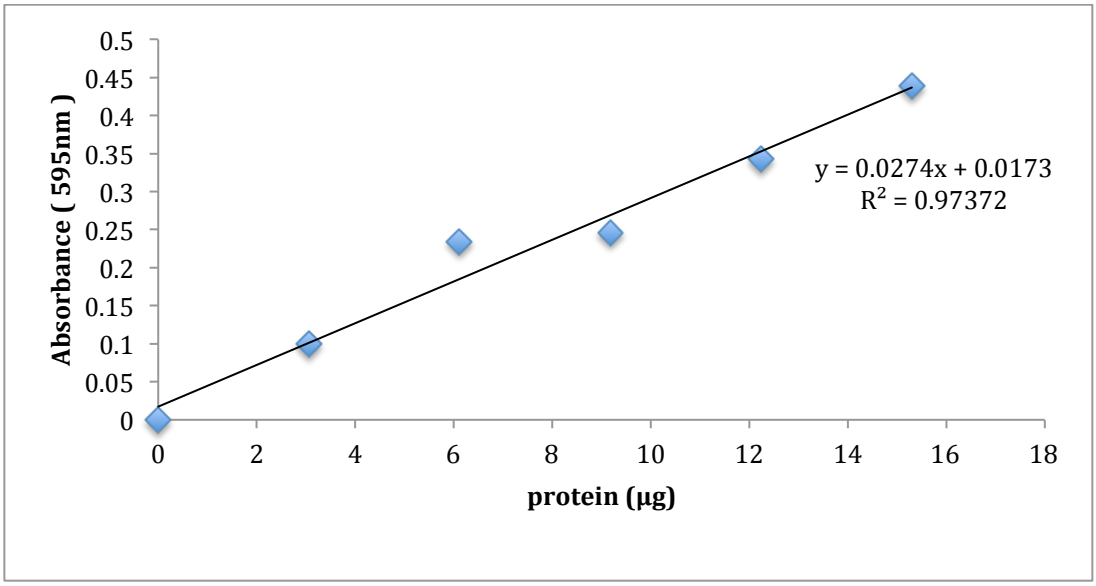


Figure 13. Standard curve of protein absorbance at 595 nm.

9. OFFGEL Fractionation

Protein OFFGEL stock solution was prepared by mixing the thiourea, DTT, glycerol solution (6 mL), and OFFGEL buffer (600 µL). The total volume was brought to the 50 mL mark on the bottle with deionized water.

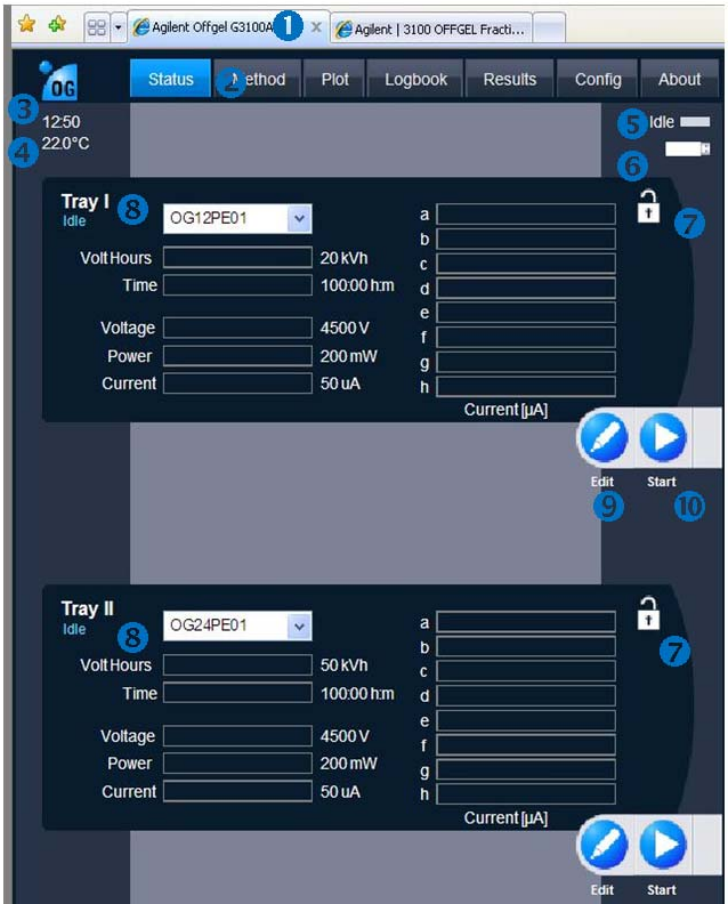
The following procedure is described for a single sample use in a low-resolution 12 cm strip/ 12 well frame. Salts were avoided because they interfere with isoelectric focusing (IEF) separation. Total salt concentration should not exceed 10 mM for optimal fractionation. Protein IPG strip rehydration solution was prepared by mixing 560 µL of protein OFFGEL stock solution and 140 µL of

deionized water. The OFFGEL protein sample was prepared by adding 360 µL of sample to 144 µL of protein OFFGEL stock solution.

A tray was placed in the OFFGel Fractionator with an orientation such that the fixed electrode side (low pH/ anode side) was on the left and the tray handle was on the right. The protective backing from the IPG strip gel was removed. The IPG strip was placed in the tray with the gel side up. The strip was pulled as far left as possible until the strip touched the left edge of the tray. The well frame was placed against the mechanical stop on the tray and the frame pressed down until it snapped into place. IPG strip rehydration solution (40 µL) was pipetted into each of the wells. After adding the sample, the tray was gently tapped on the surface of the bench to ensure the solution was in contact with the gel. Using tweezers, four electrode pads were wetted with IPG strip rehydration solution. One wetted electrode pad was placed on each protruding end of the IPG strip, making sure there was no gap between the pad and the frame. A second wetted electrode pad was placed on top of the first electrode pads. The IPG gel was allowed to swell for 15 min. Into each well, 150 µL of prepared protein sample was loaded. The cover seal was placed over the frame and pressed down gently on each well to secure a proper fit. Purified water (10 µL) was applied to the electrode pads at each end of the IPG gel. The tray was placed on the instrument platform. Cover fluid (mineral oil) was pipetted onto the gel strip ends. Plastic fixed electrode (anode) was selected and the two tabs of the electrode were placed into the slots on the left side of the tray. The fixed electrode was then rotated into position over the electrode pad. The tray

was slid, without lifting it, into the anode connector. The moveable electrode was inserted and pushed down until it clicked in place. The lid was closed and the fractionation started.

The Agilent 3100 OFFGEL Fractionator was controlled via LAN from an Internet browser interface. The communication of the browser application and the instrument is based on the display of webpage content retrieved from the instrument. The user interface provides several tabs to control the instrument: status, method, plot, logbook, result, config, and about (Figure 14). The Status tab is the main screen that was used to select and load methods, to start and stop them. It was also used to process and modify run parameters interactively of a method that was running. In the pull down menu for Method Selection a supplied default method was chosen, such as OG12PR01 (for use with 3100 OFFGel low resolution, 12 well frames). The fractionation was started after assembly of the tray with the start button. The method screen was used to view, modify, and store methods. It shows all method parameters and a graphical view of the timetable settings. Plot screen allowed the viewing over time in a XY plot. Up to four signals could be selected. Blue light below the anode connector of the respective tray indicated that the instrument was running. A blinking blue light indicated that the fractionation was finished. Upon completion, the fractionation run was stopped; the fractions were retrieved, and analyzed by SDS-PAGE electrophoresis.



1 Web Browser Tab
2 OFFGEL interface tabs
3 Instrument clock
4 Platform temperature
5 Status indicator
6 USB indicator
7 Application Lock
8 Tray method
9 Method edit
10 Fractionation start

Figure 14. Overview of Agilent 3100 OFFGel Fractionator interface. Reprinted from Agilent 3100 OFFGEL Fractionator user manual.

10. In Solution Tryptic Digestion

A digestion buffer was prepared by dissolving 10 mg of ammonium bicarbonate in 2.5 mL of ultrapure water. A reducing buffer was prepared by mixing 8 mg of DTT and 500 µL of ultrapure water. Digestion buffer (15 µL), reducing buffer (1.5 µL), and protein solution (10 µL) were added to a microcentrifuge tube and the final volume adjusted to 27 µL with ultrapure water. The procedure below

has been found to work well with protein concentrations in the range of 0.1-1.0 mg/mL. The sample was centrifuged for 10 sec and incubated at 95 ^{0}C for 5 min.

The alkylation buffer was prepared by dissolving 9 mg of iodoacetamide in 500 µL of ultrapure water. The alkylation buffer (3 µL) was added to the reaction mixture and incubated in the dark at room temperature for 20 min.

A trypsin stock solution was prepared with a concentration of 0.1 µg/µL by adding 200 µL of trypsin resuspension buffer to 20 µg of trypsin. Trypsin stock solution (1 µL) was added and the reaction mixture was centrifuged for 10 sec. The reaction mixture was incubated at 37 ^{0}C for 3 hours. Formic acid (1 µL) was added to the reaction mixture to stop digestion. The resulting peptides were analyzed by tandem mass spectrometry.

11. Ultra High Performance Liquid Chromatography and Tandem Mass Spectrometry

Tryptically digested nucleoside hydrolases were separated into peptides by reverse phase Ultra High Performance Liquid Chromatography (UHPLC). The UHPLC conditions are presented in Table 3a and the gradient method is shown in Table 3b. The eluent was then passed into a Waters Synapt q-ToF tandem mass spectrometer. A total ion chromatograph (TIC) was collected as well as the collision induced disassociated (CID) MS/MS spectra of the 5 largest multiply charged ions observed. The MS/MS conditions are presented in Table 4. Collected data were then processed with Maxent3 software by Waters or PEAKS Studio by Bioinformatics Solutions Inc. The fragment ion spectra were assigned a peptide

sequence and a BLAST search predicted the protein sequences that produced the peptides.

Table 3a. UHPLC instrument conditions for UHPLC analysis.

Sample size:	10 mL
Mobile Phase Flow Rate:	0.200 ml/min
Solvent A:	5 % Formic acid in water
Solvent B:	5 % Formic acid in acetonitrile
Initial Acetonitrile concentration:	10%
Run time:	25.00 min
Column Temperature:	30°C

Table 3b. UHPLC gradient program.

Time (min)	Flow (mL /min)	Water (%)	Acetonitrile (%)
0.00	0.200	90.0	10.0
5.00	0.200	90.0	10.0
20.00	0.200	34.0	66.0
23.00	0.200	34.0	66.0
23.10	0.200	100.0	0.0
25.00	0.200	100.0	0.0

Table 4. Waters q-ToF MS/MS operating conditions.

Acquisition Time:	23 min
Ion Source	Electro Spray
Polarity:	Positive
Analyser Mode:	V Mode
Capillary Voltage	3000 Volts
Source Temperature	90°C
Desolvation Gas Temperature	320°C
Desolvation Gas Flow Rate	800 L/h
MS TIC Scan Range	100-1500 Da
MS Collision Energy	6.0 Volts
MS Survey Threshold	>50 counts/second
Scan Time:	0.5 seconds
MS/MS Scan Range	50-2000 Da
Number of Simultaneous ions for MS/MS	5
Scans Per MS/MS Ion	3
Peak Selection for MS/MS	Charge State
Charge States of Interest	+2, +3, +4
Exclude background Ion From MS/MS	400.8977
Low Mass Collision Energy Ramp	10-30 Volts
High Mass Collision Energy Ramp	20-40 Volts

12. Kinetic Analysis

The kinetic constant K_m was determined at pH 7.2 by steady-state analysis in which the velocity was measured in the presence of varying nucleoside concentration. Reaction mixtures consisted of varying concentrations of the nucleoside under study in 10 mM Tris pH 7.2 (990 µL). The reaction was initiated by the addition of 10 µL of nucleoside hydrolase. The inosine substrate concentrations were 100 µM, 250 µM, 500 µM, 1000 µM, 1500 µM, and 2000 µM for adenosine nucleosidase. The inosine concentrations were 10 µM, 25 µM, 50 µM, 100 µM, 200 µM, 300 µM, and 400 µM for inosine nucleosidase. The amount of hypoxanthine produced was determined by HPLC as described earlier. The reaction velocity was calculated by dividing the amount of hypoxanthine produced by the reaction time. The kinetic constant K_m for inosine nucleosidase was determined by nonlinear regression of the kinetic data to the Michaelis-Menten equation using Kaleidagraph software.

CHAPTER III

RESULTS AND DISCUSSION

The enzyme nucleoside hydrolase plays an important role in the nucleoside salvage pathways in a variety of organisms including parasitic protozoans and plants. Adenosine nucleosidase and inosine nucleosidase, enzymes involved in hydrolysis of selected purine nucleosides to their bases and ribose sugar, have been isolated from a wide variety of plants. The enzymes from different species though similar in some properties, have different substrate specificities, molecular weights, kinetic constants, pH optima, and subunit structures.

Germination of yellow lupin seeds, preparation of an initial extract, heat treatment, ammonium sulfate precipitation (60% for adenosine nucleosidase and 70% for inosine nucleosidase) ion exchange chromatography, size exclusion chromatography, and affinity chromatography were steps used in the purification of adenosine and inosine nucleosidase from yellow lupin.

The most important parameter used to distinguish adenosine nucleosidase and inosine nucleosidase are the difference in substrate specificity and the accompanying kinetic parameters.[11,12] The Michaelis constant, K_m, is the concentration of substrate required by the enzyme to achieve half of maximum velocity, V_{max}, of the reaction. For many enzyme-substrate combinations, K_m represents the affinity of the enzyme toward the substrate. The kinetic constant, K_m for inosine was obtained by steady state kinetic analysis at pH 7.2 using adenosine nucleosidase and inosine nucleosidase as the enzymes. The substrate

concentrations used were 100 μM, 250 μM, 500 μM, 1000 μM, 1500 μM, and 2000 μM for adenosine nucleosidase and 10 μM, 25 μM, 50 μM, 100 μM, 200 μM, 300 μM, and 400 μM for inosine nucleosidase. Based on a direct nonlinear regression of the kinetic data using the Michaelis-Menten equation, the K_m of inosine nucleosidase at pH 7.2 was found to be 260 $\pm$ 75 μM using inosine as the substrate. A similar nonlinear regression analysis leads to a K_m of 9820 $\pm$ 13801 μM for adenosine nucleosidase using inosine as the substrate (Figures 15 and 16, respectively). The K_m for adenosine nucleosidase has a large standard deviation, because a complete kinetic analysis could not be done due to the solubility limits of inosine. However, the very different K_m values for the same substrate, inosine, does show that the two proteins are kinetically different from one another. This was expected since inosine is a best substrate for the 70% ammonium sulfate fraction of yellow lupin (inosine nucleosidase) and inosine is a poor substrate for the 60% ammonium sulfate fraction (adenosine nucleosidase).

The differing kinetic properties indicate that inosine nucleosidase and adenosine nucleosidase represent two different enzymes. This is consistent with the situation in a number of organisms in which multiple nucleoside hydrolases are present. For example, three different nucleoside hydrolases have been isolated from the insect parasite *Crithidia fasiculata.*[20, 22] Another organism in which multiple nucleoside hydrolases have been isolated is *E. coli.*[33]

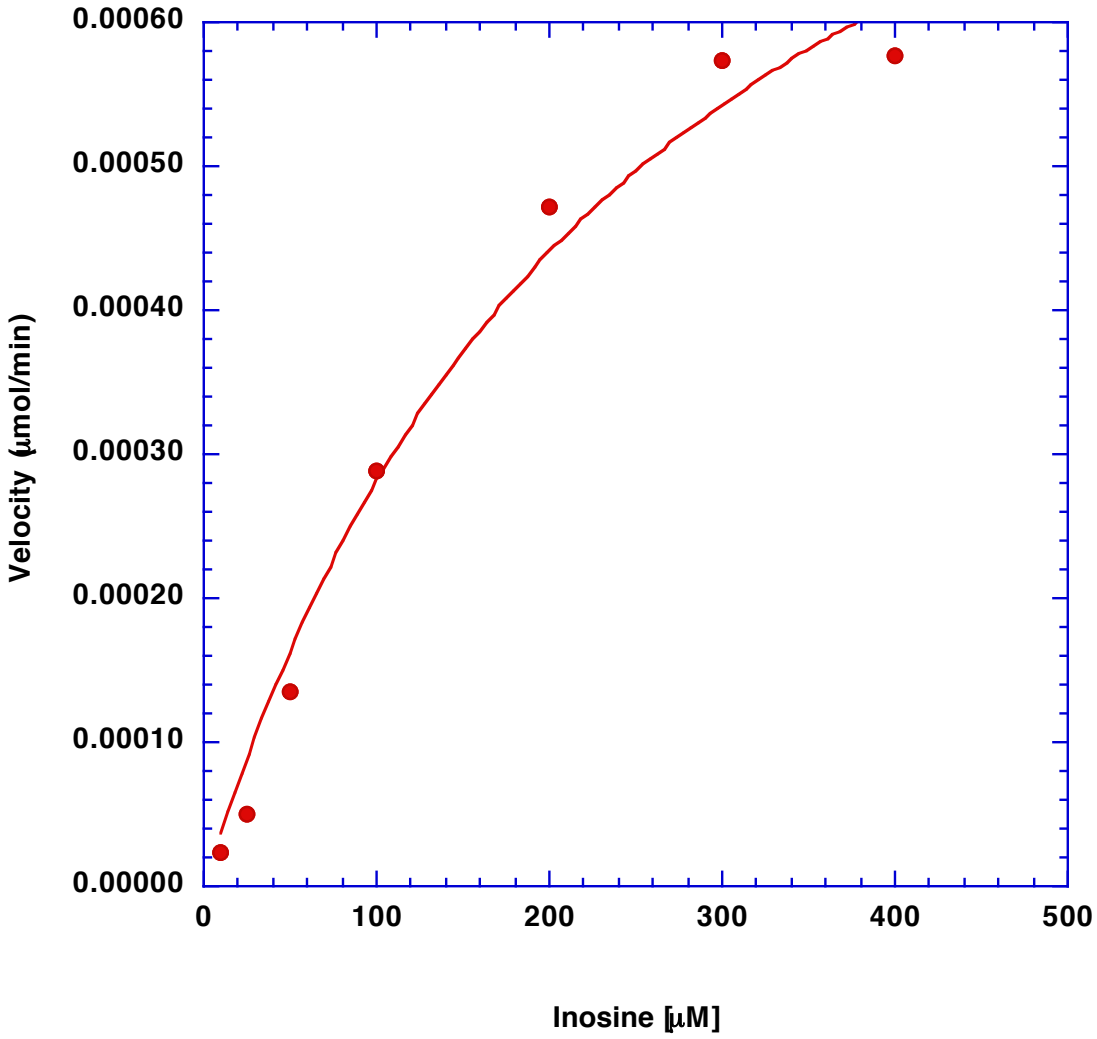


Figure 15. Kinetic analysis of inosine nucleosidase at pH 7.2 of yellow lupin (70% ammonium sulfate). The substrate (inosine) concentrations used were 10 µM, 25 µM, 50 µM, 100 µM, 200 µM, 300 µM, and 400 µM. Direct fitting of the data to the Michaelis-Menten equation yield a K_m of 260 ± 75 µM.

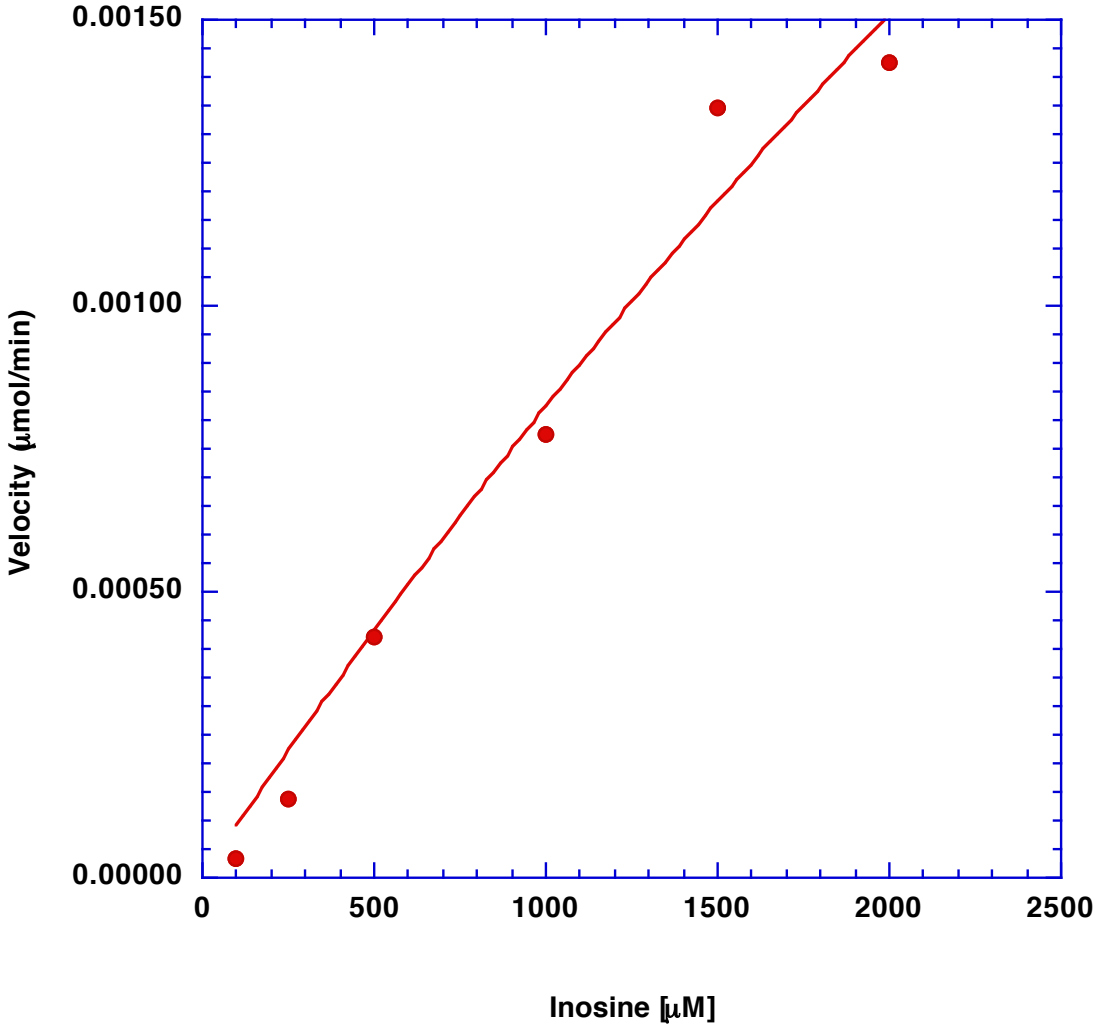


Figure 16. Kinetic analysis of adenosine nucleosidase at pH 7.2 of yellow lupin (60% ammonium sulfate). The substrate (inosine) concentrations used were 100 µM, 250 µM, 500 µM, 1000 µM, 1500 µM, and 2000 µM. Direct fitting of the data to the Michaelis-Menten equation yield a K_m of 9820 ± 13801 µM.

The Michaelis constant for inosine from other nucleoside hydrolases reveals a wide variety of values. For nucleoside hydrolases isolated from the parasitic protozoans, *Crithidia fasciculata*, *Trypanosoma brucei brucei*, *Trypanosoma cruzi*, and *Leishmania donovani*, the inosine K_m values were 380 ± 30 μM, 18 ± 1 μM, 18 ± 0.1 μM and 349 ± 143 μM respectively.[20, 19, 34, 35] The Michaelis constant for inosine for nucleoside hydrolase from Jerusalem artichoke shoots was 2.5 μM.[36] Several nucleoside hydrolases have been isolated from yellow lupin. The K_m values for calcium-stimulated guanosine-inosine nucleoside hydrolase and inosine nucleosidase were 2.7 ± 0.3 μM and 65 μM respectively.[9,11] The value of 65 μM reported by Guranoswki is relatively close to the value of 260 ± 75 μM reported in this thesis. The recently published K_m value for rihC from *E.coli*, a bacterial enzyme, was 422±225 μM for inosine.[37]

The structures of nucleoside hydrolases from parasitic protozoans and bacteria have been extensively studied. Amino acid sequences are known for nucleoside hydrolases from *Crithidia fasciculata*[8, 20], *Leishmania major*[21], *Trypanosoma vivax*[1], *Escherichia coli*[23], and *Trypanosoma brucei brucei*[19]. In addition crystal structures for a number of these enzymes are known.[23, 21] However while a number of nucleoside hydrolases have been isolated from plants almost no structural information beyond subunit molecular weights from SDS-PAGE is available.

The basic principle of sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is to separate proteins based on their ability to move

within a gel under the influence of an electric current. The distance a protein travels is a function of length of their polypeptide chains or of their molecular weight. SDS-PAGE of several purified nucleoside hydrolases and molecular weight standards are shown in Figure 17. The molecular weight of purified adenosine nucleosidase from yellow lupin was 36 kDa, which is similar to the result reported by Abusamhadneh[4].

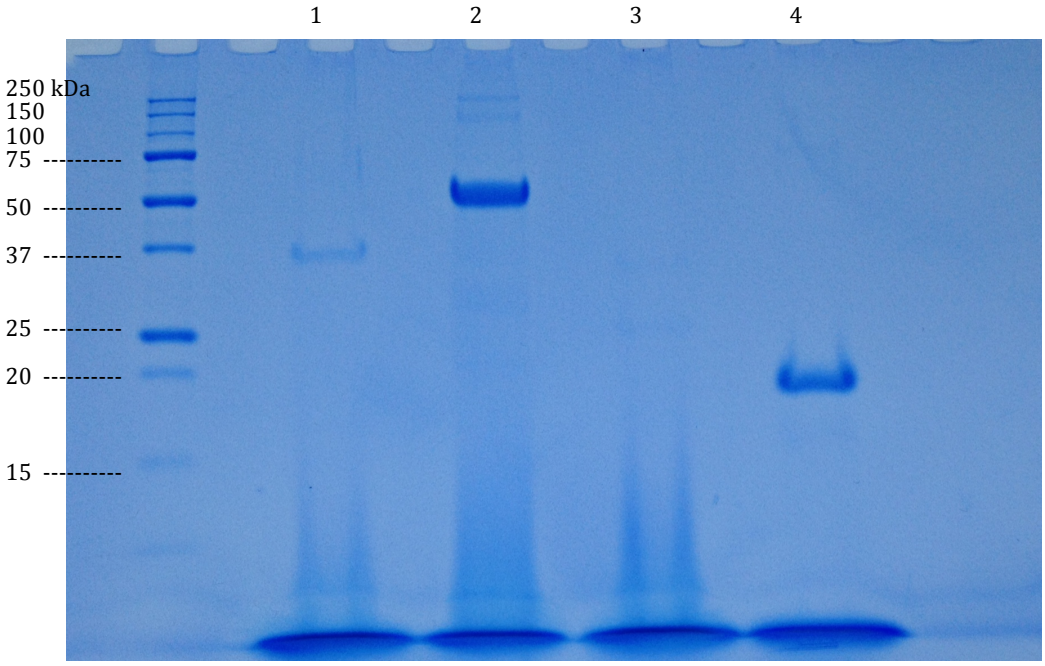


Figure 17. SDS-PAGE of purified nucleoside hydrolases after OFFGEL fractionator. Lane 1 is yellow lupin 60% (adenosine nucleosidase). Lane 2 is yellow lupin 70% (inosine nucleosidase). Lane 3 is *E.coli* rihC (nucleoside hydrolase). Lane 4 is soybean (nucleoside hydrolase).

The denatured inosine nucleosidase yielded a band at approximately 62 kDa. Guranowski reported inosine nucleosidase has a molecular weight of 62 kDa[11]. In comparison to other purine nucleosidases, the yellow lupin inosine nucleosidase

was relatively large. Two additional nucleoside hydrolases were also analyzed. The first was a recombinant form of rihC, a nonspecific nucleoside hydrolase from *E. coli* similar to IU-NH from *Crithidia fasciculata.* This enzyme has a subunit molecular weight of 36 kDa.[34] The other one was nonspecific nucleoside hydrolase, isolated from soybean. The subunit molecular weight of this enzyme was 19 kDa (Figure 17).

OFFGEL electrophoresis is a novel technique that separates proteins or peptides according to their isoelectric points (pI), the pH at which a protein has no overall charge. It provides a high-resolution separation, and experimentally derived pI information. The basis of the technique is a preformed gel strip, which contains a pH gradient. When a voltage is applied to the ends of the gel strip, the proteins or peptides move through the gel until the molecules reach a well covering a portion of the gel where the pH equals the pI of the molecule. At this point the protein does not move any further along the gel. Multiple standards such as β-lactoglobulin, cytochrome c, and carbonic anhydrase of known pI were run to ensure that the instrument functioned correctly.

A recombinant form of *E. coli* rihC, designated *E. coli* B6, was analyzed on the OFFGEL fractionator. Two bands were apparent for the *E. coli* B6 sample after analysis by SDS-PAGE of the results from the OFFGEL fractionator. The bands from wells 4-5 on the OFFGEL fractionator corresponded to a molecular weight of 36 kDa and pI values of 5.17 and 5.32. Based on the amino acid sequence the calculated molecular weight was 31893 daltons. The calculated pI was 5.92. Both calculations

and experimental values are within the expected uncertainty of the methods. This further supports the proper functioning of the method.

Yellow lupin adenosine nucleosidase and inosine nucleosidase were also fractionated by isoelectric focusing. The major band appeared in well 4 for yellow lupin adenosine nucleosidase and in well 9 for yellow lupin inosine nucleosidase, corresponding to pIs of 5.17 and 6.15, respectively.

Adenosine nucleosidase from soybean was subjected to isoelectric focusing and a major band in well 5 was present. This corresponds to a pI value of 5.28 and a molecular weight of 19 kDa (Figure 17). Samples from wells were collected separately and further desalted for sequence analysis.

The procedure to determine the sequence and ultimately the identity of a protein by mass spectrometry is well established (Figure 18). Mass spectrometry of whole proteins is less sensitive than peptide mass spectrometry and the mass of the intact protein by itself is insufficient for identification. Therefore, proteins were degraded enzymatically to peptides by trypsin, leading to peptides with C-terminally protonated amino acids, which provides an advantage in subsequent peptide sequencing.[32] The peptides were separated by high-pressure liquid chromatography and eluted into an electrospray ion source where they are nebulized into small, highly charged droplets. After evaporation of solvent, multiply protonated peptides enter the collision chamber of the mass spectrometer and a mass spectrum of the peptides eluting at this time point was collected. The

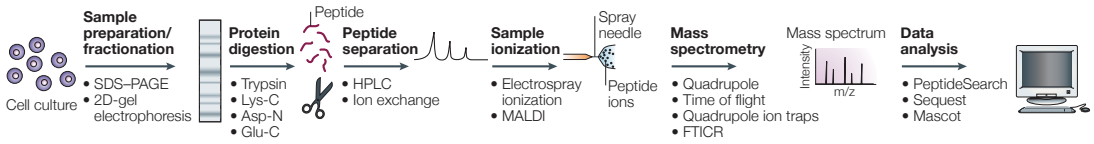


Figure 18. Workflow for protein identification by mass spectrometry. Reprinted with permission from Steen, H.; Mann, M. *Nature Reviews* **2004**, *5,* 699-711.

computer generates a prioritized list of possible sequences for these peptides, which were stored for matching against protein sequence databases.

The vast majority of proteins are too large to be sequenced directly and as stated earlier the mass of the intact protein is insufficient to identify a protein. Therefore the protein must be fragmented and the individual peptides sequenced. The most widely used protease to generate fragments of target proteins is trypsin, which cleaves the peptide bond on the carboxyl side of arginine or lysine, except when they are followed by a proline. In this study four nucleoside hydrolases were subjected to digestion by trypsin and the resulting peptides separated by reverse phase chromatography. Selected peptides were then sequenced by mass spectrometry. The four proteins analyzed in this study were rihC from *E. coli*, adenosine nucleosidase from yellow lupin, inosine nucleosidease from yellow lupin, and adenosine nucleosidase from soybean. The resulting chromatograms are shown in Figures 19-22. As can be seen from the chromatograms in Figures 20 and 21, adenosine nucleosidase and inosine nucleosidase had different tryptic digest peptide patterns again confirming these two proteins are different from one another.

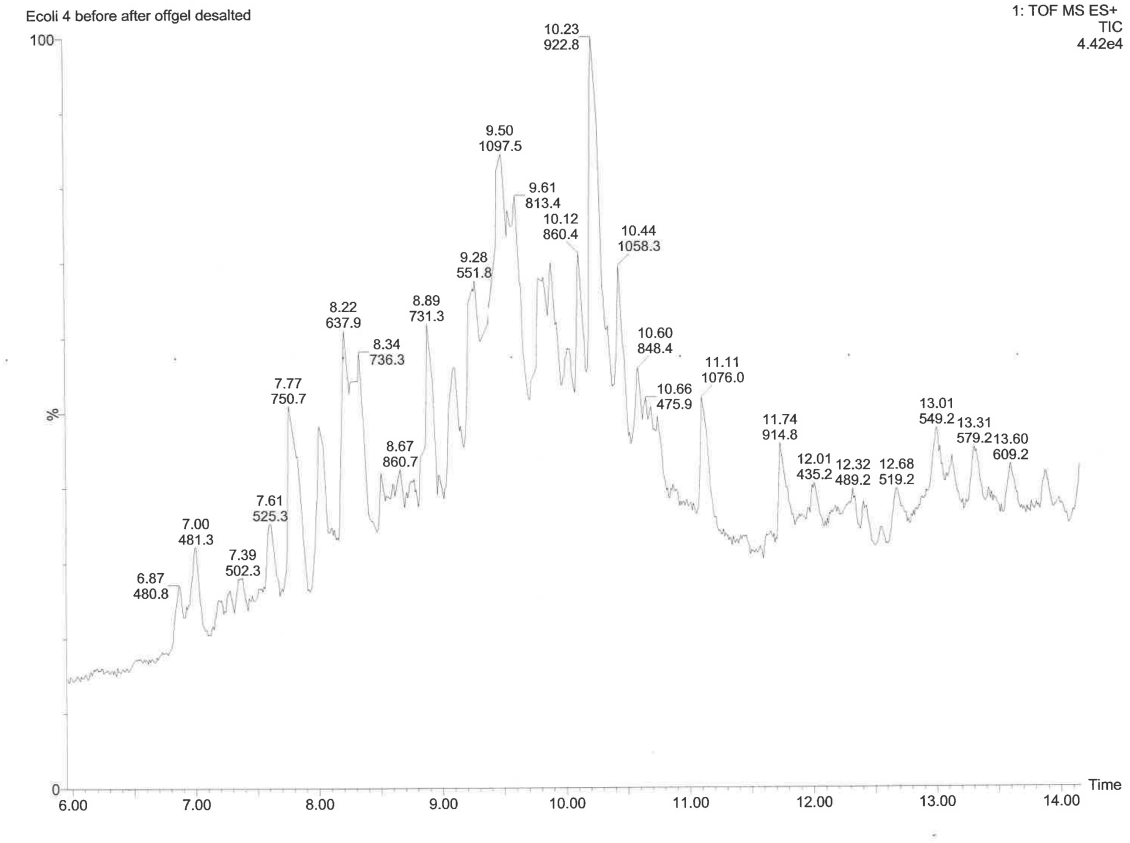


Figure 19. Partial chromatogram of peptides from tryptic digest of *E. coli* B6 (*E. coli* rihC).

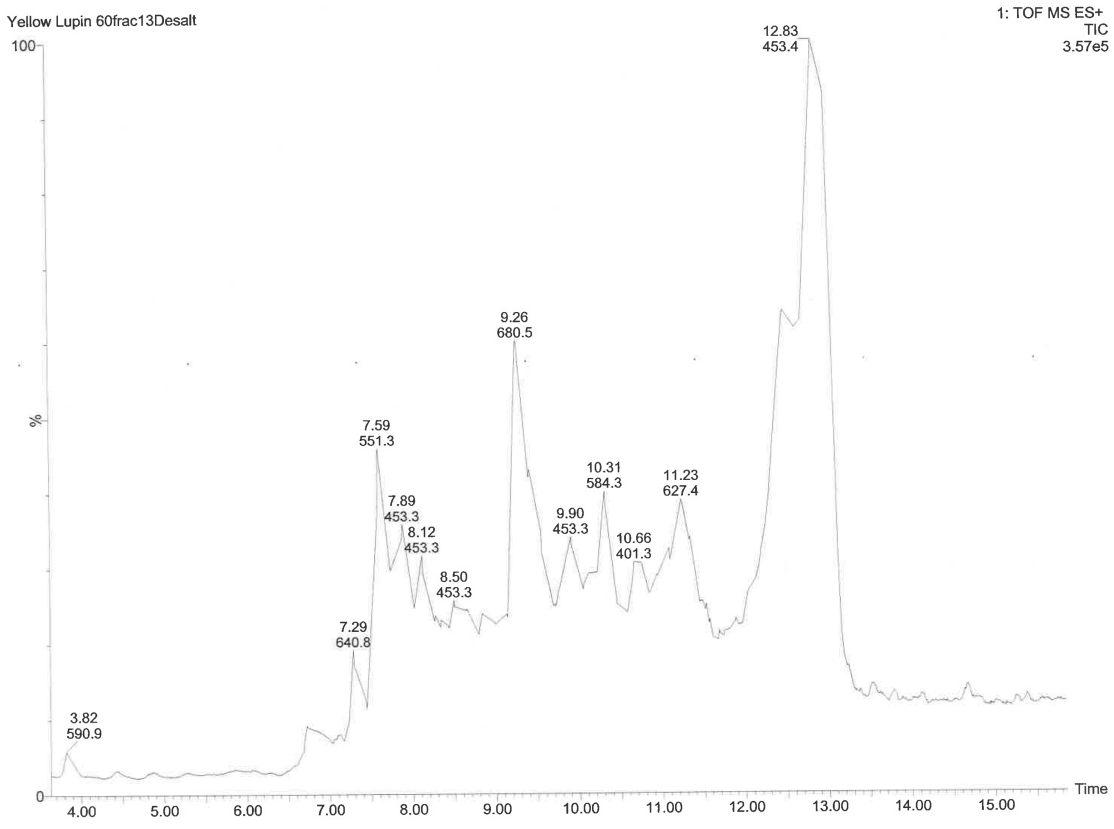


Figure 20. Partial chromatogram of peptides from tryptic digest of yellow lupin adenosine nucleosidase.

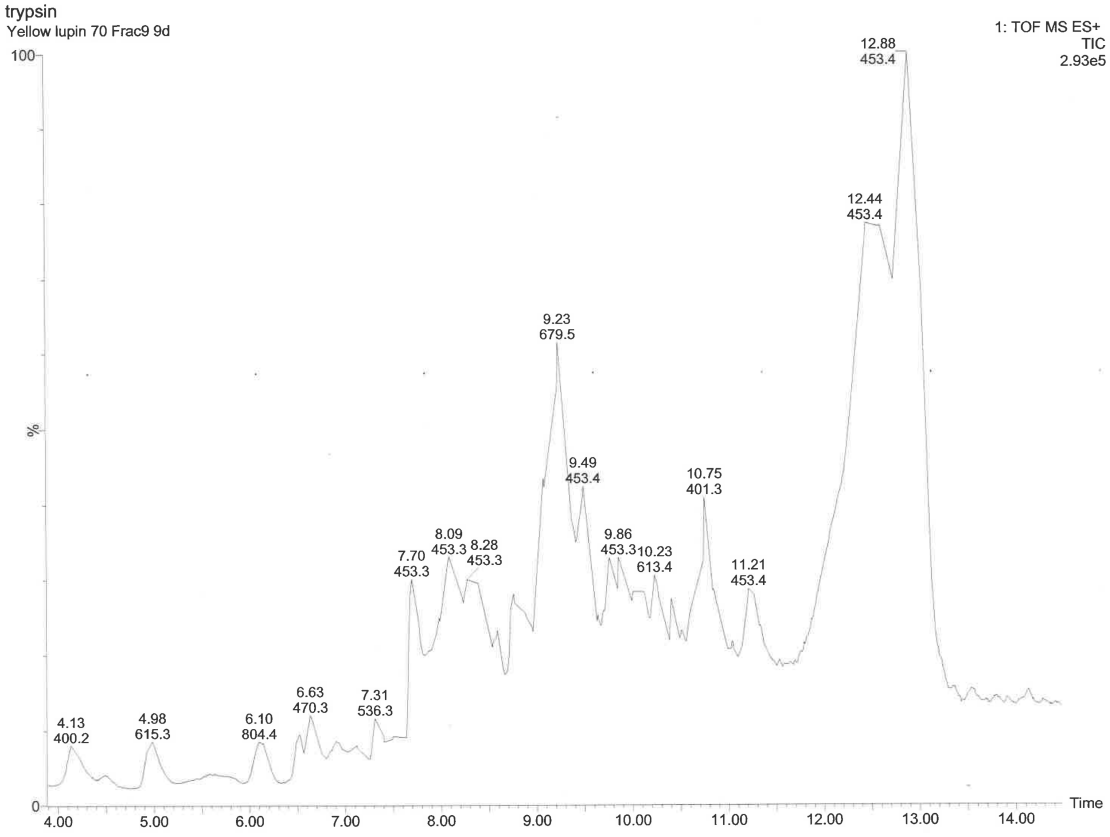


Figure 21. Partial chromatogram of peptides from tryptic digest of yellow lupin inosine nucleosidase.

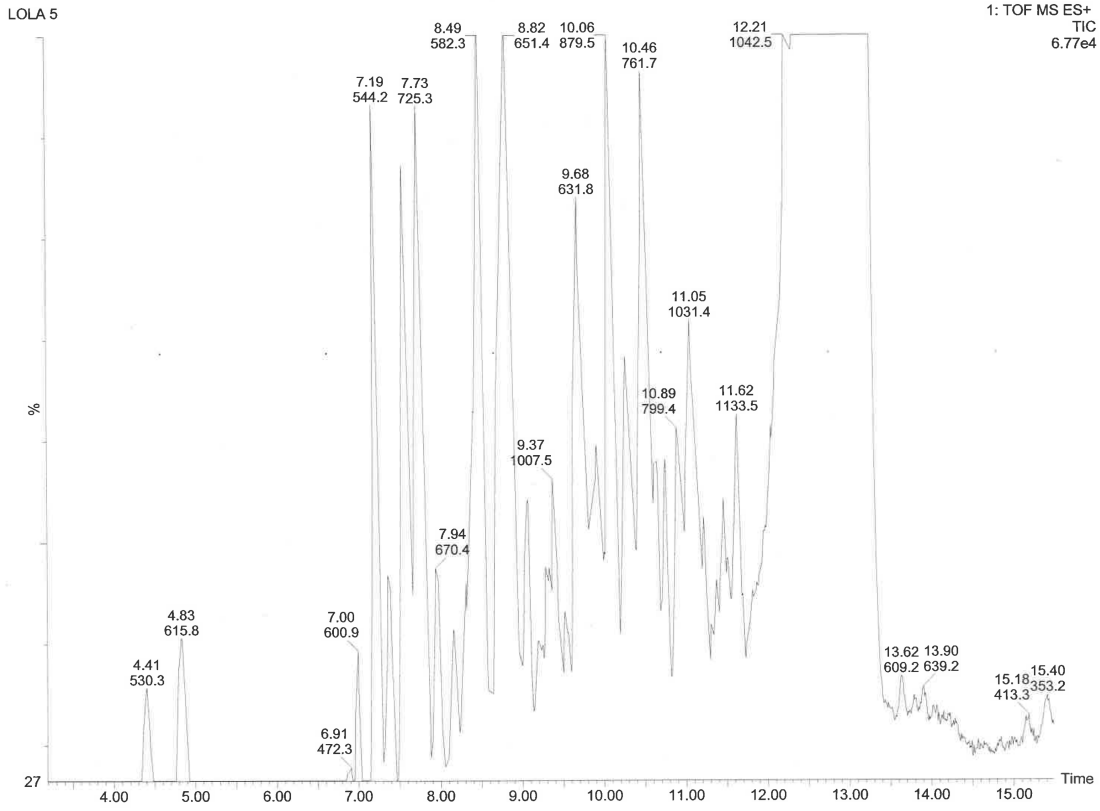


Figure 22. Partial chromatogram of peptides from tryptic digest of soybean adenosine nucleosidase.

Analysis of *E. coli* B6 sample was used to confirm the procedures developed to determine the amino acid sequence of a peptide by mass spectrometry and identify the protein based on its sequence. One tryptic digestion was required to generate Figure 23. The sequence of *E. coli* rihC has been determined and was compared to the sequence information generated by mass spectrometry (Figure 23). Of the 296 amino acids of *E. coli* rihC, mass spectrometry was able to determine 136 or 46%. Of the 136 amino acids identified there were no mismatches between the known sequences and the MS-generated sequence.

Peptide mapping is a technique used to identify unknown proteins based upon the mass of peptides or the sequence of peptides generated by cleavage of the original protein. The peptide masses or sequences are compared to known protein sequences in a database. Using a computer program such as Swissprot or BLAST (basic local alignment search tool), the MS-generated peptides are analyzed to determine the best match to known proteins.

Using the peptides generated from *E. coli* B6 a BLAST search was carried out. A partial list of the results is shown in Figure 24. The score of each alignment is indicated by one of five different colors, which divides the range of scores into five groups. Alignment is color-coded; pink is in the score range of 80-200. The description table provides a summary of the database sequences identified by BLAST to be similar to those of the input query. At the top of the list is non-specific ribonucleoside hydrolase rihC (*Escherichia coli*) with the highest percent identity of 52%. As can be seen from Figure 24, the top 20 results were various nucleoside

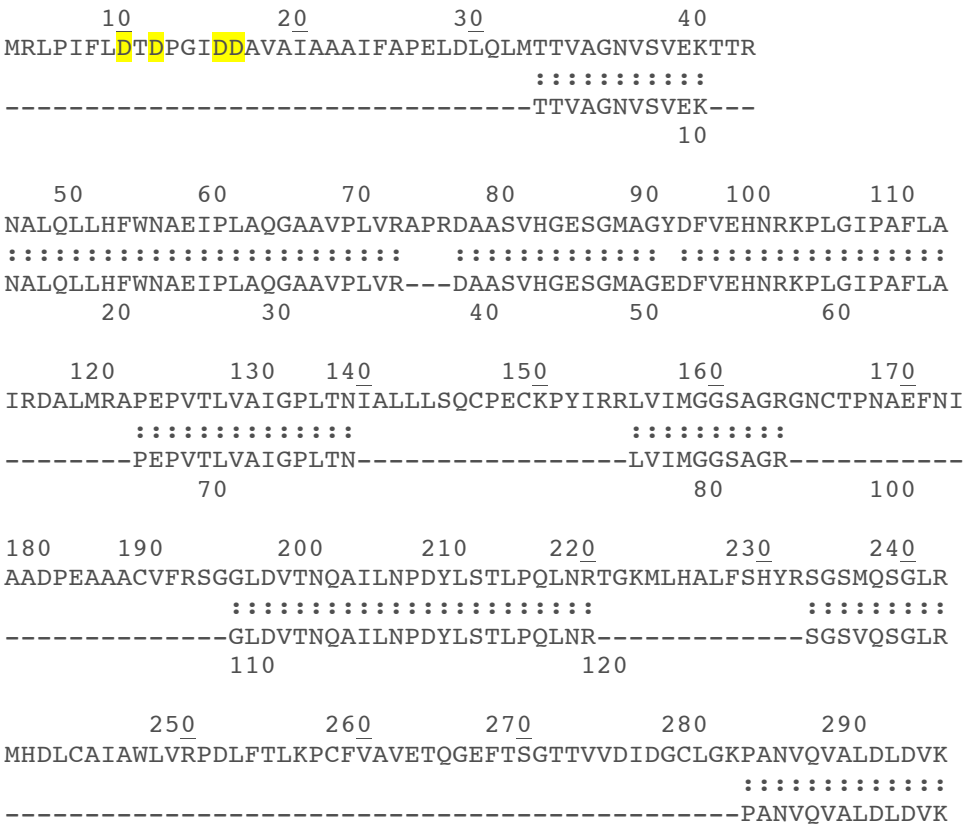
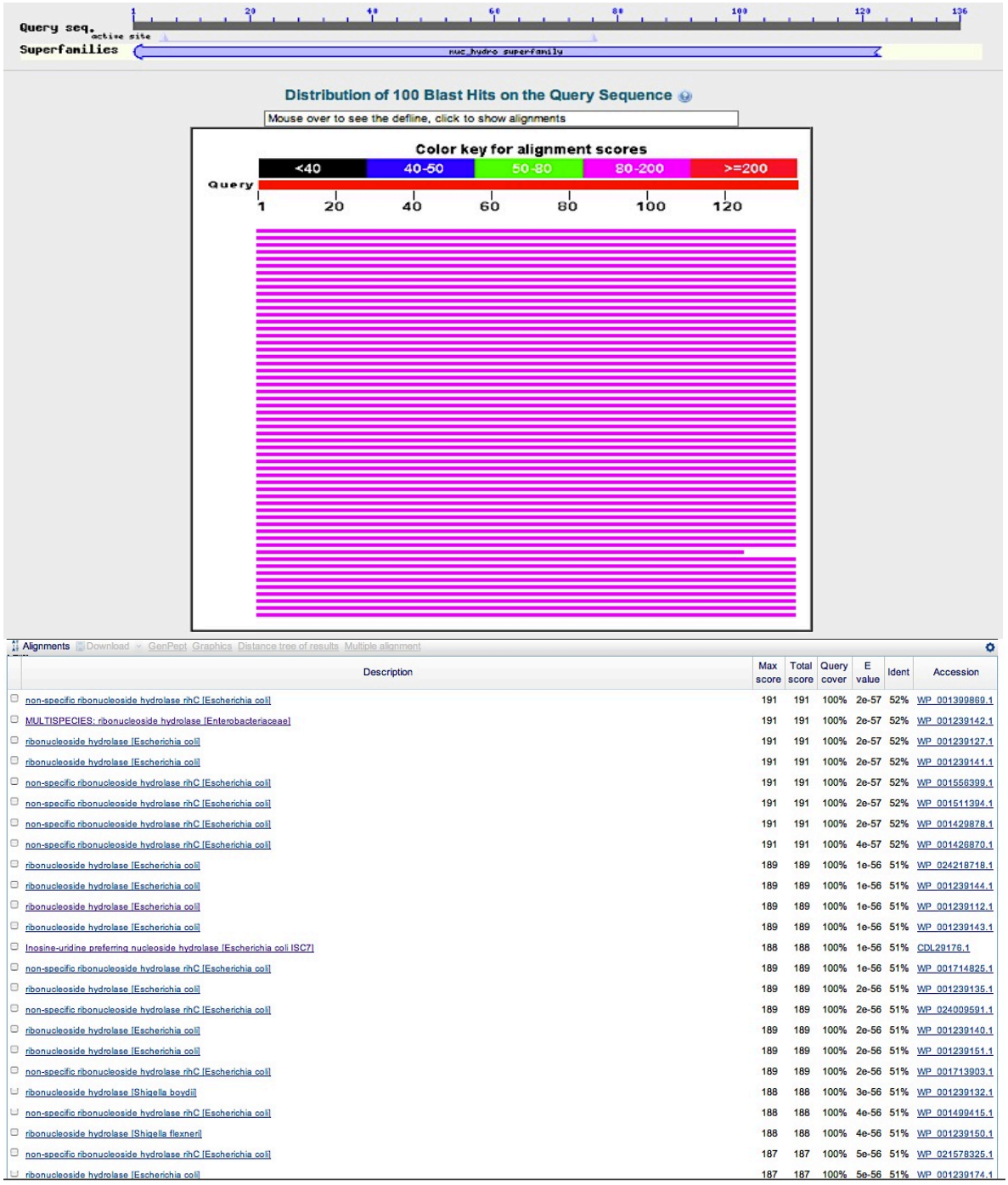


Figure 23. Sequence alignment of known *E. coli* rihC (top line) and *E. coli* B6 nucleoside hydrolase (bottom line). The alignment was made using LALING web sequence analysis software. Dotted line indicates the alignment of two identical amino acids and dashes indicates missing sequence. Known sequence is adapted from Hunt, C.; Gillani, N.; Farone, A.; Rezaei, M.; Kline, P.C. *Biochim. Biophys. Acta* **2005**, *1751*, 140-149.



Alignments | Download | GenPept | Graphics | Distance tree of results | Multiple alignment

Description	Max score	Total score	Query cover	E value	Ident	Accession
non-specific ribonucleoside hydrolase rihC [Escherichia coli]	191	191	100%	2e-57	52%	WP_001399869.1
MULTISPECIES: ribonucleoside hydrolase [Enterobacteriaceae]	191	191	100%	2e-57	52%	WP_001239142.1
ribonucleoside hydrolase [Escherichia coli]	191	191	100%	2e-57	52%	WP_001239127.1
ribonucleoside hydrolase [Escherichia coli]	191	191	100%	2e-57	52%	WP_001239141.1
non-specific ribonucleoside hydrolase rihC [Escherichia coli]	191	191	100%	2e-57	52%	WP_001556399.1
non-specific ribonucleoside hydrolase rihC [Escherichia coli]	191	191	100%	2e-57	52%	WP_001511394.1
non-specific ribonucleoside hydrolase rihC [Escherichia coli]	191	191	100%	2e-57	52%	WP_001429878.1
non-specific ribonucleoside hydrolase rihC [Escherichia coli]	191	191	100%	4e-57	52%	WP_001426870.1
ribonucleoside hydrolase [Escherichia coli]	189	189	100%	1e-56	51%	WP_024218718.1
ribonucleoside hydrolase [Escherichia coli]	189	189	100%	1e-56	51%	WP_001239144.1
ribonucleoside hydrolase [Escherichia coli]	189	189	100%	1e-56	51%	WP_001239112.1
ribonucleoside hydrolase [Escherichia coli]	189	189	100%	1e-56	51%	WP_001239143.1
Inosine-uridine preferring nucleoside hydrolase [Escherichia coli ISC7]	188	188	100%	1e-56	51%	CDL29176.1
non-specific ribonucleoside hydrolase rihC [Escherichia coli]	189	189	100%	1e-56	51%	WP_001714825.1
ribonucleoside hydrolase [Escherichia coli]	189	189	100%	2e-56	51%	WP_001239135.1
non-specific ribonucleoside hydrolase rihC [Escherichia coli]	189	189	100%	2e-56	51%	WP_024009591.1
ribonucleoside hydrolase [Escherichia coli]	189	189	100%	2e-56	51%	WP_001239140.1
ribonucleoside hydrolase [Escherichia coli]	189	189	100%	2e-56	51%	WP_001239151.1
non-specific ribonucleoside hydrolase rihC [Escherichia coli]	189	189	100%	2e-56	51%	WP_001713903.1
ribonucleoside hydrolase [Shigella boydii]	188	188	100%	3e-56	51%	WP_001239132.1
non-specific ribonucleoside hydrolase rihC [Escherichia coli]	188	188	100%	4e-56	51%	WP_001499415.1
ribonucleoside hydrolase [Shigella flexneri]	188	188	100%	4e-56	51%	WP_001239150.1
non-specific ribonucleoside hydrolase rihC [Escherichia coli]	187	187	100%	5e-56	51%	WP_021578325.1
ribonucleoside hydrolase [Escherichia coli]	187	187	100%	5e-56	51%	WP_001239174.1

Figure 24. Top results from BLAST analysis of peptides generated by tryptic digest of rihC *E.coli*. (sample *E. coli* B6)

hydrolases from *E. coli*. The sequence from *E. coli* B6 was also aligned against the *Arabidopsis thaliana* genome and was found to have a 37% identity to a 336 amino acid protein that has been identified as uridine nucleosidase. Another alignment was done against known proteins from *Crithidia fasciculata* resulting in a 31% identity with a protein identified as inosine-uridine nucleoside hydrolase.

A similar analysis was carried out on three nucleoside hydrolases from plants, adenosine nucleosidase and inosine nucleosidase from yellow lupin, and a protein tentatively identified on the basis of substrate specificity as adenosine nucleosidase from soybean.

The sequences of the peptides for adenosine nucleosidase and inosine nucleosidase from yellow lupin and adenosine nucleosidase from soybean were determined using PEAKS software.[38] Based on MS/MS data, the PEAKS software looks at all possible amino acid combinations and calculates the best match to the data. Because of this it does not rely upon protein databases. The sequences of 8 peptides from adenosine nucleosidase from yellow lupin are shown in Figure 25.

```
       10         20         30         40         50         60
MELVDAARPL LALGQISERL NVQVSDVKEL VADDAWLNSE FIATVQQRIV QGLPIDEFSR

        70
LDLTAQELSE EK
```

Figure 25. Sequence of peptides determined by mass spectrometry from adenosine nucleosidase from yellow lupin (72 amino acids).

Based on the SDS-PAGE identified molecular weight of adenosine nucleosidase, the percent coverage was approximately 25%. It must be pointed out,

that at this time it is not possible to place the sequenced peptides within the protein. In other words while the sequence within the peptides runs from the N-terminal to C-terminal end, it is not possible to tell the order of the peptides within the protein. A similar analysis was carried out for inosine nucleosidase from yellow lupin and adenosine nucleosidase from soybean. The sequences of those protein peptides are shown in Figures 26 and 27 respectively. In the case of inosine nucleosidase, the coverage of amino acids in the proteins was 25%, while for soybean adenosine nucleosidase it was 35%. The same limitation that was discussed for yellow lupin adenosine nucleosidase on the order of the peptides within the protein applies for inosine nucleosidase from yellow lupin and soybean adenosine nucleosidase.

```
10          20          30          40          50          60
LENLQNYRIV EFQSKPNTLI LPKHSDADYI LVVLNGRATI TIVNPDKRIL NPDDNQNLRL

         70          80          90         100         110         120
AIPINNPGNF YDFYPSSTKN TLEATFNTRY EEIQRVSKEQ IQELRHAQSS SGRGKPSESG

        130
PFFGNFYEIT PDR
```

Figure 26. Sequence of peptides determined by mass spectrometry from inosine nucleosidase from yellow lupin (133 amino acids).

```
         10          20          30          40          50
SRGIGTIISS PYRFIAEGHP LCVGIPTEWS VVEDLPEGPA VKDAMDGWFR VSDDEFNNY
```

Figure 27. Sequence of peptides determined by mass spectrometry from adenosine nucleosidase from soybean (59 amino acids).

Inspection of the peptide sequences determined did not reveal the recurring N-terminal DXDXXXDD motif that has been found in other nucleoside hydrolases. In

addition the highly conserved histidine residue seen in nucleoside hydrolases from other sources was not seen.

Using the peptides generated from a tryptic digestion of yellow lupin of adenosine nucleosidase and a tryptic digestion of inosine nucleosidase, a BLAST search was carried out, similar to the analysis that was done for *E. coli* B6. The lists of possible proteins are shown in Figures 28 and 29. While the results did not specifically identify a nucleosides hydrolase, it did indicate that the sequences belongs to *Lupinus* genus.

A BLAST sequence was also carried out on the peptides that resulted from a tryptic digest of adenosine nucleosidase from soybean. Again the results did not identify a nucleoside hydrolase, although a number of proteins from soybeans have a score of at least 99 (Figure 30).

Additional sequence information will need to be gathered for the plant nucleoside hydrolases to ultimately determine their structure and mechanism.

Description	Max score	Total score	Query cover	E value	Ident	Accession
cytosolic malate dehydrogenase [Lupinus angustifolius]	60.8	103	84%	3e-09	54%	AEB60994.1
cytosolic malate dehydrogenase [Lupinus angustifolius]	60.5	108	84%	3e-09	54%	AEB60995.1
malate dehydrogenase [Lupinus albus]	60.5	103	84%	4e-09	54%	AAO15574.1
hypothetical protein Osl_34036 [Oryza sativa Indica Group]	58.9	58.9	84%	1e-08	35%	EEC67171.1
PREDICTED: malate dehydrogenase, cytoplasmic-like [Fragaria vesca subsp. vesca]	58.5	100	84%	2e-08	53%	XP_004296370.1
NAD-dependent malate dehydrogenase [Cucumis sativus]	55.5	55.5	51%	4e-08	49%	ABK55707.1
malate dehydrogenase, cytoplasmic-like [Glycine max]	57.4	102	84%	5e-08	52%	NP_001243291.1
RecName: Full=Malate dehydrogenase; Short=NtRed-2 [Nicotiana tabacum]	57.0	99.4	84%	6e-08	51%	Q9FSF0.1
RecName: Full=Malate dehydrogenase, cytoplasmic [Mesembryanthemum crystallinum]	56.6	89.7	84%	8e-08	51%	Q24047.1
cytosolic malate dehydrogenase [Malus domestica]	56.2	95.9	84%	1e-07	50%	ABB36659.1
unknown [Medicago truncatula]	56.2	98.6	84%	1e-07	50%	AFK34402.1
Malate dehydrogenase [Morus notabilis]	56.2	94.7	84%	1e-07	51%	EXC01795.1
malate dehydrogenase [Capsicum chinense]	54.3	99.0	84%	1e-07	49%	CAJ13707.1
Malate dehydrogenase [Medicago truncatula]	55.8	98.2	84%	1e-07	50%	XP_003590025.1
hypothetical protein MIMGU_mgv1a009751mg [Mimulus guttatus]	55.8	99.7	84%	2e-07	49%	EYU32552.1
PREDICTED: malate dehydrogenase, cytoplasmic-like [Cucumis sativus]	55.5	95.5	84%	2e-07	49%	XP_004147146.1
hypothetical protein POPTR_0010s08180g [Populus trichocarpa]	55.1	96.3	84%	3e-07	50%	XP_006378335.1
malate dehydrogenase [Phaseolus vulgaris]	55.5	96.3	84%	3e-07	50%	AGZ15381.1
RecName: Full=Malate dehydrogenase, cytoplasmic [Beta vulgaris]	55.1	90.1	84%	3e-07	50%	Q9SML8.1
PREDICTED: malate dehydrogenase, cytoplasmic-like [Solanum lycopersicum]	55.1	99.7	84%	3e-07	50%	XP_004247734.1
hypothetical protein CISIN_1g020022mg [Citrus sinensis]	55.1	55.1	51%	3e-07	49%	KDO57590.1
PREDICTED: malate dehydrogenase, cytoplasmic-like [Setaria italica]	55.1	99.0	84%	4e-07	50%	XP_004982864.1
malate dehydrogenase-like protein [Solanum tuberosum]	54.7	96.3	84%	4e-07	50%	NP_001275487.1
PREDICTED: malate dehydrogenase, cytoplasmic-like [Cicer arietinum]	54.7	93.6	84%	4e-07	49%	XP_004495378.1
hypothetical protein PHAVU_007G273500g [Phaseolus vulgaris]	54.7	96.7	84%	4e-07	50%	XP_007145851.1

Figure 28. Top results from BLAST analysis of peptides generated by tryptic digest of adenosine nucleosidase from yellow lupin.

Description	Max score	Total score	Query cover	E value	Ident	Accession
conglutin beta 2 [Lupinus angustifolius]	217	217	100%	1e-64	63%	AEB33713.1
conglutin beta [Lupinus angustifolius]	208	208	100%	6e-62	61%	ABR21771.1
conglutin beta 7 [Lupinus angustifolius]	208	208	100%	3e-61	61%	AEB33718.1
conglutin beta [Lupinus angustifolius]	204	204	99%	8e-61	59%	ABR21772.1
conglutin beta 5 [Lupinus angustifolius]	206	206	100%	3e-60	60%	AEB33716.1
conglutin beta 3 [Lupinus angustifolius]	204	204	100%	5e-60	60%	AEB33714.1
conglutin beta [Lupinus angustifolius]	204	204	99%	1e-59	59%	ACB05815.1
conglutin beta 1 [Lupinus angustifolius]	204	204	99%	1e-59	59%	AEB33712.1
conglutin beta 4 [Lupinus angustifolius]	203	203	100%	3e-59	60%	AEB33715.1
conglutin beta 6 [Lupinus angustifolius]	202	202	100%	3e-59	60%	AEB33717.1
beta-conglutin precursor [Lupinus albus]	201	201	100%	8e-59	58%	AAS97865.1
vicilin-like protein [Lupinus albus]	200	200	100%	1e-58	58%	CAI84850.2
BLAD [Lupinus albus]	163	163	72%	5e-48	70%	ABB13526.1
7S globulin-2 [Vigna angularis]	140	140	100%	8e-37	42%	BAF56571.1
Chain A, Crystal Structure Of Adzuki Bean 7s Globulin-3 [Vigna angularis]	140	140	100%	9e-37	42%	2EAA_A
seed storage protein B [Vigna luteola]	139	139	100%	1e-36	43%	AAZ06681.1
seed storage protein A [Vigna luteola]	139	139	100%	1e-36	43%	AAZ06680.1
Chain A, Crystal Structure Of Adzuki Bean 7s Globulin-1 [Vigna angularis]	138	138	100%	4e-36	41%	2EA7_A
8S globulin beta isoform precursor [Vigna radiata]	133	133	100%	2e-34	41%	ABG02282.1
Vicilin protein [Vigna unguiculata]	132	132	100%	4e-34	40%	CAP19802.1
8S globulin alpha' isoform precursor [Vigna radiata]	128	128	100%	2e-32	42%	ABG02281.1
RecName: Full=Beta-conglycinin, beta chain; Flags: Precursor [Glycine max]	126	126	100%	7e-32	41%	P25974.1
PREDICTED: beta-conglycinin, beta chain-like [Glycine max]	126	126	100%	7e-32	41%	XP_003556052.1
beta-conglycinin beta subunit [Glycine max]	126	126	100%	7e-32	41%	BAD98463.1
Chain A, Crystal Structures Of Recombinant And Native Soybean Beta- Conglycinin Beta Homotrimers [Glycine max]	126	126	100%	8e-32	41%	1IPK_A
Chain A, Crystal Structures Of Recombinant And Native Soybean Beta-Conglycinin Beta Homotrimers Complexes With N-Acetyl-D-Glucosamine [Glycine max]	126	126	100%	9e-32	41%	1IPJ_A

Figure 29. Top results from BLAST analysis of peptides generated by tryptic digest of inosine nucleosidase from yellow lupin..

Description	Max score	Total score	Query cover	E value	Ident	Accession
Kunitz trypsin inhibitor [Glycine max]	100	100	98%	2e-24	71%	BAB63285.1
Kunitz trypsin inhibitor [Glycine soja]	100	100	98%	2e-24	71%	BAF95191.1
trypsin inhibitor subtype A precursor [Glycine max]	100	100	98%	2e-24	71%	NP_001236752.1
trypsin inhibitor A precursor [Glycine max]	100	100	98%	2e-24	71%	NP_001238611.1
Chain A, Soybean Trypsin Inhibitor [Glycine max]	99.0	99.0	98%	3e-24	71%	1BA7_A
Chain B, Complex Porcine Pancreatic TrypsinSOYBEAN TRYPSIN Inhibitor, Orthorhombic Crystal Form [Glycine max]	99.0	99.0	98%	3e-24	71%	1AVW_B
kunitz trypsin inhibitor 3 [Glycine max]	99.8	99.8	98%	3e-24	71%	AAF65315.1
uncharacterized protein LOC100527408 precursor [Glycine max]	99.0	99.0	98%	7e-24	70%	NP_001235651.1
Kunitz trypsin inhibitor [Glycine soja]	99.0	99.0	98%	7e-24	70%	BAF95192.1
Kunitz trypsin inhibitor [Glycine soja]	98.2	98.2	98%	1e-23	70%	BAF95193.1
Kunitz trypsin inhibitor [Glycine max]	97.4	97.4	98%	2e-23	70%	BAD04941.1
trypsin inhibitor C (Kunitz) - soybean [Glycine max]	96.3	96.3	98%	3e-23	70%	TISYC
Kunitz trypsin inhibitor [Glycine soja]	97.1	97.1	98%	3e-23	69%	BAE94326.1
soybean kunitz trypsin inhibitor [Glycine soja]	96.3	96.3	98%	6e-23	68%	AGC92015.1
trypsin inhibitor subtype B precursor [Glycine max]	94.4	94.4	98%	4e-22	68%	NP_001238220.1
Kunitz trypsin inhibitor [Glycine soja]	94.4	94.4	98%	4e-22	68%	BAD04942.1
RecName: Full=Trypsin inhibitor B; AltName: Full=Kunitz-type trypsin inhibitor B [Glycine max]	93.6	93.6	98%	4e-22	68%	P01071.1
Kunitz trypsin inhibitor [Glycine soja]	92.4	92.4	98%	2e-21	66%	BAF95190.1
Kunitz trypsin inhibitor [Glycine max]	89.4	89.4	98%	2e-20	67%	AAK20289.1
truncated Kunitz trypsin inhibitor [Glycine max]	86.7	86.7	84%	8e-20	70%	AAK20290.1
Kunitz trypsin inhibitor [Glycine tomentella]	78.6	78.6	98%	2e-16	58%	BAG68489.1
RecName: Full=Turmerin [Curcuma longa]	69.7	69.7	100%	6e-14	64%	P85278.2

Figure 30. Top results from BLAST analysis of peptides generated by tryptic digest of adenosine nucleosidase from soybean.

CHAPTER IV

CONCLUSION

Nucleoside hydrolases catalyze the hydrolysis of the N-glycosidic bond in purine and pyrimidine nucleosides resulting in ribose and the corresponding base. For example adenosine nucleosidase catalyzes the hydrolysis of the glycosidic bond in adenosine resulting in ribose and adenine. The enzyme is part of the nucleoside salvage pathway allowing the reuse of the base without the energy expenditure necessary to synthesize the base *de novo*. While the enzymes from the parasitic protozoans and bacteria have been studied extensively little information is available on the structure of the enzyme from plants.

Inosine nucleosidase and adenosine nucleosidase were shown to be two different proteins based on a number of parameters. The Michaelis constant using inosine as the substrate was found to be 260 ± 75 μM for inosine nucleosidase and 9820 ± 13801 μM for adenosine nucleosidase. The large standard deviation for adenosine nucleosidase is due to the solubility limit of inosine preventing a complete kinetic analysis. Additional evidence that these are two separate proteins was found in their isoelectric points. Using isoelectric focusing of yellow lupin adenosine nucleosidase and inosine nucleosidase the isoelectric points were 5.17 and 6.15, respectively. The isoelectric points were nearly a full pH unit different indicating significantly different overall amino acid composition.

Techniques for sequencing proteins by tandem mass spectrometry were developed. Sample of rihC from *E. coli* was used as a standard. An overall coverage

of 46% of the amino acids in the protein was determined and aligned with the known sequence of the protein. Using similar techniques the sequence of three nucleoside hydrolases from plants were determined. Based on the molecular weight of yellow lupin adenosine nucleosidase and inosine nucleosidase and soybean adenosine nucleosidase, approximately 25%, 25%, and 35% of the amino acids in each protein were determined respectively.

The four proteins in this study were also subjected to a BLAST analysis. RihC from *E. coli* was identified as a nucleoside hydrolase. However, when the proteins from the plant sources were subjected to the same analysis they were not identified as a nucleoside hydrolase. These proteins were correctly identified as coming from yellow lupin and soybean.

While initial work has been done in sequencing nucleoside hydrolases from plants, additional work remains to be done. A recurring motif of aspartate residues in the active sites of nucleoside hydrolases from a number of organisms has been identified. This recurring motif has yet to be found in the sequence information determined so far. Finally a BLAST analysis of these proteins did not identify them as nucleoside hydrolases. Therefore additional sequence information must be determined for these enzymes to properly classify them.

REFERENCES

1. Versees, W.; Decanniere, K.; Pelle, R.; Depoorter, J.; Brosens, E.; Parkin, D.W.; Steyaert, J. Structure and Function of a Novel Purine Specific Nucleoside Hydrolase from *Trypanosoma vivax*. *J. Mol. Biol.* **2001**, *307*,1363-1379.

2. Sinden, Richard. *DNA Structure and Function*. London: Gulf Professional Publishing, **1994**.

3. Grisham, Charles; Reginald Garrett. *Biochemistry*. Brooks Cole: Belmont, CA. **2005**.

4. Zrenner, Rita; Stitt, Mark; Sonnewald, Uwe; Boldt, Ralf. Pyrimidine and Purine Biosynthesis and Degradation in Plants. *The Annual Review of Plant Biology.* **2006**, *57*, 805-836.

5. Stasolla, Claudio; Katahira, Riko; Thorpe, Trevor; Ashihara, Hiroshi. Purine and Pyrimidine Nucleotide Metabolism in Higher Plants. *J. Plant Physiol.* **2003,** *160*, 1271-1295.

6. Moffatt, Barbara; Hiroshi, Ashihara. Purine and Pyrimidine Nucleotide Synthesis and Metabolism. *The Arabidopsis Book* **2002**, *1*,e0018.

7. Jung, Benjamin; Hoffmann, Christiane; Mohlmann, Torsten. Arabidopsis Nucleoside Hydrolase Involved in Intracellular and Extracellular Degradation of Purines. *The Plant Journal*. **2011**, *65*, 703-711.

8. Versees, Wim; Steyaert, Jan. Catalysis by Nucleoside Hydrolases. *Current Opinion in Structural Biology.* **2003**, *13*, 731-738.

9. Szuwart, Maciej; Starzynska, Elzbieta; Pietrowska-Borek, Malgorzata; Guranowski, Andrzej. Calcium-Stimulated Guanosine-Inosine Nucleosidase from Yellow Lupin (*Lupinus luteus*). *Phytochemistry* **2006**, *67*, 1476-1485.

10. Hansen, Michael Riis; Dandanell, Gert. Purification and Characterization of RihC, a Xanthosine-Inosine-Uridine-Adenosine-Preferring Hydrolase from *Salmonella enterica* Serovar Typhimurium. *Biochim. et Biophys Acta* **2005**, *1723*, 55-62.

11. Guranowski, Andrzej. Purine Catabolism in Plants. *Plant Physiol.* **1982**. *70*, 344-349.

12. Abusamhadneh, Ekram; McDonald, Nancy; Kline, Paul. Isolation and Characterization of Adenosine Nucleoside from Yellow Lupin (*Lupinus luteus*). *Plant Sci.* **2000**, *153*, 25-32.

13. Guranowski, A.; Schneider, Z. Purification and Characterization of Adenosine Nucleosidase from Barley Leaves. *Biochim. Biophys. Acta.* **1977**, *482*, 145-158.

14. Campos, Alexandre; Rijo-Johansen, Maria; Carneiro, Maria; Fevereiro, Pedro. Purification and Characterization of Adenosine Nucleosidase from *Coffea Arabica* Young Leaves. *Phytochemistry* **2005**, *66*, 147-151.

15. Chen, Chong-maw; Kristopeit, Susan. Metabolism of Cytokinin: Deribosylation of Cytokinin Ribonucleoside by Adenosine Nucleosidase from Wheat Germ Cells. *Plant Physiol.* **1981**, *68,* 1020-1023.

16. Imagawa, Hiroshi; Yamano, Haruko; Inoue, Kazuhiko; Takino, Yoshinori. Purification and Properties of Adenosine Nucleosidases from Tea Leaves. *Agric. Biol. Chem.* **1979,** *43*, 2337-2342.

17. Poulton, Jonathan; Butt, Vernon. Partial Purification and Properties of Adenosine Nucleosidase from Leaves of Spinach Beet (*Beta vulgaris L.*). *Planta* **1976**, *131*, 179-185.

18. Burch, L.R.; Stuchbury, T. Purification and Properties of Adenosine Nucleosidase from Tomato (*Lycopersicon esculentum*) Roots and Leaves. *Plant Physiol.* **1986**, *125*, 267-273.

19. Parkin, D.W. Purine-Specific Nucleoside N-ribohydrolase from *Trypanosoma brucei brucei*. Purification, Specificity, and Kinetic Mechanism. *J. Biol. Chem.* **1996**, *271*, 21713-21719.

20. Parkin, D.W.; Horenstein, B. A.; Abdulah, D.R.; Estupinan, B.; Schramm, V. L. Nucleoside Hydrolase from *Crithidia fasciculata.* Metabolic Role, Purification, Specificity, and Kinetic Mechanism. *J. Biol. Chem.* **1991**, *266*, 20658-20665.

21. Shi, W.; Schramm, V. L.; Almo, S. C. Nucleoside Hydrolase from *Leishmania major.* Cloning, Expression, Catalytic Properties, Transition State Inhibitors, and the 2.5-A Crystal Structure. *J. Biol. Chem.* **1999**, *274*, 21114-21120.

22. Estupinan, B.; Schramm, V. L. Guanosine-Inosine-Preferring Nucleoside N-glycohydrolase from *Crithidia fasciculata. J. Biol. Chem.* **1994**, *269,* 23068-23073.

23. Giabbai, Barbara; Degano, Massimo. Crystal Structure to 1.7 *A* of the *Escherichia coli* Pyrimidine Nucleoside Hydrolase YeiK, a Novel Candidate for Cancer Gene Therapy. *Structure* **2004**, *12*, 739-749.

24. Versees, W.; Decanniere, K.; Holsbeke, E.V.; Devroede, N.; Steyaert, J. Enzyme-Substrate Interactions in the Purine-specific Nucleoside Hydrolase from *Trypanosoma vivax*. *J. Biol. Chem.* **2002**, *277*, 15938-15946.

25. Hustoft, Hanne; Malerod, Helle; Wilson, Steven; Reubsaet, Leon; Lundanes, Elsa; Greibrokk, Tyge. A Critical Review of Trypsin Digestion for LC-MS Based Proteomics. *Integrative Proteomics.* **2012**, 73-89.

26. Lahm, H.W.;Langen, H. Mass spectrometry: A Tool for the Identification of Proteins Separated by Gels. *Electrophoresis.* **2000**, *21*, 2105-2114.

27. Lopez-Ferrer, D. Sample Treatment for Protein Identification by Mass Spectrometry-based Techniques. *TrAC Trends in Analytical Chemistry* **2006** *10*, 996-1005.

28. Aebersold, Ruedi; Mann, Matthias. Mass Spectrometry-based Proteomics. *Nature* **2003**, *422*, 198-207.

29. Westermeier, R.; Naven, T. *Proteomics in Practice*. Germany: Wiley-VCH, **2002**.

30. Roepstorff, P.; Fohlman J. Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed. Mass Spectrom*. **1984**, *11*, 601.

31. Prasain, Jeevan. *Tandem Mass Spectrometry-Applications and Principles*. Croatia: InTech, **2012**.

32. Steen, H.; Mann, M. The ABC's (and XYZ's) of Peptide Sequencing. *Nature Reviews* **2004**, *5,* 699-711.

33. Petersen, C.; Moller, L.B. The RihA, RihB, and RihC Ribonucleoside Hydrolases of *Escherichia coli*: Substrate Specificity, Gene Expression, and Regulation. *J. Biol. Chem.* **2001**, *276*, 884-894.

34. Miller, Richard; Sabouring, Carol; Krenitsky, Thomas; Berens, Randolph; Marr, Joseph. Nucleoside Hydrolases from Trypanosoma cruzi. *J. Biol. Chem*. **1984**, *259*, 5073-5077.

35. Cui, Liwang; Rajasekariah, G.R.; Martin, Samuel. A Nonspecific Nucleoside Hydrolase from *Leishmania donovani*: Implications for Purine Salvage. *Gene*. **2001**, *280*, 153-162.

36. Le Floc'h, F.; LeFleuriel, J. The Purine Nucleosidases of Jerusalem Artichoke Shoots. *Phytochemistry.* **1981**, *20*, 2127-2129.

37. Arivett, B.; Farone, M.; Masiragani, R.; Burden, A.; Judge, S.; Osinloye, A.; Minici, C.; Degano, M.; Robinson, M.; Kline P. Characterization of Inosine-Uridine Nucleoside Hydrolase (RihC) from *Escherichia coli*. *Biochim. et Biophys. Acta* **2014**, *1844*, 656-662.

38. Ma, Bin; Zhang, Kaizhong; Hendrie, Christopher; Liang, Chengzhi; Doherty-Kirby, Amanda; Lajoie, Gilles. PEAKS: Powerful Software for Peptide *De novo* Sequencing by Tandem Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2337-2342.