

Studies of the Pathogenesis of Amebiasis, Including Characterization of a
Low Molecular Weight Protein Tyrosine Phosphatase

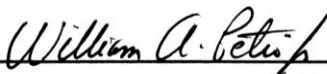
Alicia Selena Linfood
San Diego, California, USA


BS Biochemistry and Cell Biology, University of California San Diego, 1994
BA Biological Anthropology, University of California San Diego, 1994
MS Biology, California State University Long Beach, 1999

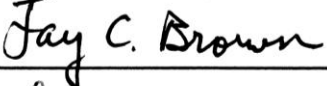
A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy


Department of Microbiology, Immunology, and Cancer Biology

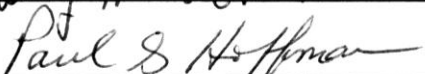
University of Virginia
December 2011














Abstract

Entamoeba histolytica is a protozoal eukaryotic intestinal parasite of humans, and is endemic in developing countries or under circumstances where hygiene is poor. In the Petri lab we are interested in amebic virulence: the how and why of the ability of this parasite to cause disease.

We chose RNAi to attempt to knock down gene expression in *E. histolytica*, and developed the short hairpin RNA interference system as a tool to reduce protein expression in order to study the function of virulence-involved genes. We used an episomal vector-based system utilizing the U6 promoter to drive expression of 29-basepair short hairpin RNAs (shRNAs) to knock down expression of four unrelated genes: PATMK, Igl, URE3-BP and EhC2A. PATMK levels were reduced by 99%, Igl levels by 72%, URE3-BP by 89%, and EhC2A by 97%. This shRNA system using the U6 promoter to drive expression of shRNAs is effective at knocking down gene expression in *E. histolytica*, and thus is a useful tool to study gene function in this organism, especially of virulence factors.

We have partially characterized an *E. histolytica* putative low molecular weight protein tyrosine phosphatase (LMW-PTP). The equivalent mammalian protein regulates cell proliferation, motility, and adhesion. These attributes are also needed for amebic virulence. The closest matches to this enzyme are found in plant or Bacteroides species, and are ~50% identical. Several different structures for this amebic tyrosine phosphatase have been solved, showing the ligand-induced conformational changes necessary for binding of substrate. In amebae, it is expressed at low but detectable levels, and can be

detected by immunoprecipitation followed by immunoblotting. When overexpressed as an epitope-tagged protein from an amebic expression vector, the LMW-PTP can be detected by immunoblotting without an immunoprecipitation step, and LMW-PTP-overexpressing transfectant amebae were used to visualize the cellular location of this protein. A mutant LMW-PTP protein in which the catalytic cysteine in the active site was replaced with a serine lacked phosphatase activity, and was used to identify a number of possible substrate proteins via mass spectrometry analysis. Seven of these putative substrate protein genes were cloned with an epitope tag for overexpression in amebae.

Acknowledgements

The work that makes up this dissertation was only possible because of the support from my mentor Dr. Bill Petri, who has advised me well, has helped polish my various writings and given me a lot of suggestions and support along my journey, as well as allowing me the opportunity to travel to places I likely would have not otherwise visited. I appreciate the members of my committee, Drs. Joanna Goldberg, Jay Brown, Amy Bouton, Bob Nakamoto and Paul Hoffman, for providing useful criticism and thoughtful suggestions throughout this process. I would like to thank the Department of Microbiology, Immunology, and Cancer Biology for having given me the opportunity to pursue this research. I also would like to thank Dr. Anindya Dutta for suggesting the use of short hairpin RNAs to knock down expression of amebic genes, and Dr. David Brautigam for his help and advice on my LMW-PTP project.

Many people in the Petri laboratory have helped me in this journey, including Girija Ramakrishan, Suzanne Stroup, and Lauren Lockhart, who were very patient in training me when I first started. Carol Gilchrist has been an invaluable resource for many things ameba. I appreciated bouncing ideas around with former and current fellow graduate students Doug Boettner, Eddie Moreno, Xiaoti Guo, Sarah Buss, and Nate Christy, and I am pleased that the shRNA tool that I developed was useful in Doug and Eddie being able to obtain needed data for their projects. I am indebted to Christina Bousquet and Ellyn Moore for their support work of growing stock amebae and other cultured cells, keeping the lab organized and running, and for ordering the reagents I need. My students Katie Good and Nona Jiang aided me with cloning, westerns,

transfections of amebae, and many, many plasmid minipreps. I would like to thank Carrie Cowardin for her confocal imaging work with my samples. I very much appreciate my family members who have been so supportive of me during this process: my sister Lara who encouraged me to move out here to pursue this work, my mother Virginia and brother Lawrence who chose to move out here with me while I pursued my PhD, and my grandmother Freda who always tells me how proud she is of me, the first person on her side of the family to earn a PhD. Also thanks to my father Gary: look, Dad, I finally did it. Now there's another Dr. Linford in the family.

Table of Contents

Abstract.....	ii
Acknowledgements	iv
Table of Contents	vi
List of Figures	viii
List of Tables.....	xiii
List of Abbreviations	xvi

Chapter 1. Introduction to the Parasite *Entamoeba histolytica*.

I. The Parasite <i>Entamoeba histolytica</i>	1
II. The Life Cycle of <i>Entamoeba histolytica</i>	2
III. The <i>Entamoeba histolytica</i> Genome	2
IV. Genome Reduplication in <i>Entamoeba histolytica</i>	3
V. <i>Entamoeba histolytica</i> Organelles	4
VI. <i>Entamoeba histolytica</i> Metabolism.....	4
VII. <i>Entamoeba histolytica</i> and Disease.....	5
VIII. Known Virulence Factors of <i>E. histolytica</i>	8
IX. Colonization of the Colonic Mucus by <i>E. histolytica</i>	8
X. Disease Following Colonization by <i>E. histolytica</i>	9
XI. The Gal/GalNAc Lectin is a Major Amebic Virulence Factor.....	10
XII. Structure of the Gal/GalNAc Lectin.....	11
XIII. Known <i>Entamoeba histolytica</i> Signal Transduction Molecules.....	13

XIV.	An <i>Entamoeba histolytica</i> Virulence-Related Transcription Factor.....	13
XV.	Goals of this Dissertation	14
Chapter 2. Short hairpin RNA-Mediated Knockdown of Protein Expression in <i>Entamoeba histolytica</i>.		
I.	Background on RNA Interference (RNAi).....	22
II.	Materials and Methods	32
III.	Results	45
IV.	Discussion.....	51
V.	Future Directions.....	54
Chapter 3. Partial Characterization of the <i>Entamoeba histolytica</i> Low Molecular Weight Protein Tyrosine Phosphatase.		
I.	Background on Protein Tyrosine Phosphatases	102
II.	Materials and Methods	118
III.	Results	141
IV.	Discussion.....	163
V.	Future Directions.....	174
	Literature Cited	280

List of Figures

Chapter 1.	Introduction to the Parasite <i>Entamoeba histolytica</i>	
1-1	The Life Cycle of <i>Entamoeba histolytica</i>	16
1-2	Diagram of the Gal/GalNAc Lectin in the <i>E. histolytica</i> Plasma Membrane	18
1-3	Invasive Amebiasis with Tissue Destruction	20
Chapter 2.	Short hairpin RNA-Mediated Knockdown of Protein Expression in <i>Entamoeba histolytica</i>	
2-1	Cleavage of RNA into siRNAs by Dicer	57
2-2	The RNA Interference (RNAi) Pathway	59
2-3	RNA-Dependent RNA Polymerase (RdRP) Function	61
2-4	RdRPs Can Transcribe RNA With or Without a Primer	63
2-5	shRNA System for <i>Entamoeba histolytica</i> : Generation of Short Hairpins by PCR	65
2-6	shRNA System for <i>Entamoeba histolytica</i> : Modification of the Amebic Expression Vector pGIR310 to Express shRNA, and the Expected Structures of the shRNA	67
2-7	Western Immunoblot for PATMK shRNA Transfectants	69
2-8	Western Immunoblot for Igl shRNA Transfectants	71
2-9	Western Immunoblot for URE3-BP shRNA Transfectants	73
2-10	Western Immunoblot for EhC2A Transfectants	75

2-11	Northern Blots of Small RNAs Extracted from Igl and PATMK Transfectants.....	77
------	---	----

**Chapter 3. Partial Characterization of the *Entamoeba histolytica*
Low Molecular Weight Protein Tyrosine Phosphatase**

3-1	The Protein Tyrosine Phosphatase Superfamily	179
3-2	Catalytic Mechanism of Protein Tyrosine Phosphatases.....	181
3-3	Protein Sequence Alignments Comparing the <i>E. histolytica</i> LMW-PTPs With Selected Plant LMW-PTPs.....	183
3-4	Protein Sequence Alignments Comparing the <i>E. histolytica</i> LMW-PTPs With Selected Bacterial LMW-PTPs.....	186
3-5	Protein Sequence Alignments Comparing the <i>E. histolytica</i> LMW-PTPs With Selected Unicellular Eukaryotic LMW-PTPs.....	189
3-6	Protein Sequence Alignments Comparing <i>E. histolytica</i> LMW-PTPs With Selected LMW-PTPs from Multicellular Animals.....	192
3-7	A Phylogenetic Tree for the Comparison Between the <i>Entamoeba histolytica</i> LMW-PTP protein XP_656359 and All Other Selected LMW-PTPs from Different Species.....	194
3-8	Structure of the <i>Entamoeba histolytica</i> LMW-PTP Apoenzyme with No Bound Ligand.....	196
3-9	Structure of the <i>Entamoeba histolytica</i> LMW-PTP with the Substrate Mimic HEPES Bound.....	198

3-10	Structure of the <i>Entamoeba histolytica</i> LMW-PTP with the Product Mimic Sulfate Bound	200
3-11	Triple Overlay of the Three <i>Entamoeba</i> <i>histolytica</i> LMW-PTP Structures	202
3-12	Crystal Shapes of the Crystallized LMW-PTP Proteins.....	204
3-13	Active Site Residues for the <i>E. histolytica</i> LMW-PTP with the Substrate Mimic HEPES Bound	206
3-14	Active Site Residues for the <i>E. histolytica</i> LMW-PTP with the Substrate Mimic HEPES Bound Using a Space-Filling Model.....	208
3-15	Space-Filling Model of Top and Side Views of the <i>E. histolytica</i> LMW-PTP Structure with HEPES Bound in the Active Site	210
3-16	Assay for Phosphatase Activity of the Purified Recombinant LMW-PTP Using a Generic Phosphatase Substrate	212
3-17	Test of Anti-LMW-PTP Sera from Rabbits Inoculated With Purified Recombinant LMW-PTP	214
3-18	Diagram of the Two-Vector Tetracycline-Inducible System for Expressing Wild-Type and Mutant Epitope-Tagged LMW-PTP	216
3-19	Assay for Phosphatase Activity of the Purified Recombinant Wild-Type and Substrate-Trapping Mutant LMW-PTP Proteins	218
3-20	Assay for Tetracycline Induction of Epitope-Tagged Wild-Type LMW-PTP Overexpression in Transfected Trophozoites	220

3-21	Test for Inhibition of Cellular Protein Tyrosine Phosphatases by the Addition of Sodium Orthovanadate and Iodoacetamide	222
3-22	Confocal Images of the Cellular Location of the C-Terminally Epitope-Tagged Wild-Type and Substrate-Trapping Mutant LMW-PTP Transfectants with Induced or Uninduced LMW-PTP Expression	224
3-23	A Representative Immunoblot of Pulldowns Using Substrate-Trapping Mutant LMW-PTP or BSA Protein Coupled to Affi-Gel 15 Beads	226
3-24	A Representative Silver Stain of Pulldowns Using Substrate-Trapping Mutant LMW-PTP or BSA Protein Coupled to Affi-Gel 15 Beads	228
3-25	Silver Stained Gels of Pulldowns Using Substrate-Trapping Mutant LMW-PTP or BSA Protein Coupled to Affi-Gel 15 Beads Submitted for Tryptic Digest and Mass Spectrometry Analysis.....	230
3-26	Sequences of Unique Peptides and their Locations in the Protein Sequences of Putative Substrate Proteins: 328t.00002 and Type A Flavoprotein	232
3-27	Sequences of Unique Peptides and their Locations in the Protein Sequences of Putative Substrate Proteins: Putative Protein Kinase and Conserved Hypothetical Protein GI56465028.....	234

3-28	Sequences of Unique Peptides and their Locations in the Protein Sequences of Putative Substrate Proteins: Putative ARP 2/3 Complex 34 kDa Subunit and Hypothetical Protein 503.t00001	236
3-29	Sequences of Unique Peptides and their Locations in the Protein Sequences of Putative Substrate Proteins: Putative Ribose- Phosphate Pyrophosphokinase and the Calcium-Binding and URE3-Sequence Specific DNA Binding Protein Transcription Factor URE3-BP.....	238
3-30	Protein Sequences and Predicted Tyrosine Phosphorylation Sites in the Proteins Selected from Mass Spectrometry Analysis for Cloning, Overexpression, and Further Analysis	240
3-31	Diagram of Epitope-Tagged Putative Substrate Protein Genes Subcloned into pGIR209.....	245

List of Tables

Chapter 2.	Short hairpin RNA-Mediated Knockdown of Protein Expression in <i>Entamoeba histolytica</i>	
2-1	Sequences Chosen to Generate shRNA Constructs That Were Successfully Transfected into Trophozoites	79
2-2	Oligos Used in Generating shRNA Constructs via PCR and Transfected into Trophozoites	81
2-3	Sequences of Oligos Used for Igl and URE3-BP Amplification in qRT-PCR.....	84
2-4	Proteins with Similar Sequences to PATMK Amplified in qRT-PCR, and Sequences of Oligos Used to Test for Off-Target Effects of PATMK Knockdown.....	86
2-5	Knockdown of PATMK mRNA and qRT-PCR Study To Evaluate if Expression of PATMK-Related Genes Is Affected by Transfectants Expressing shRNA Against PATMK (Off-Target Effects)	90
2-6	Summary of PATMK Protein Levels in PATMK shRNA Transfectants	92
2-7	Summary of Igl1 Protein Levels in Igl shRNA Transfectants	94
2-8	Summary of Igl mRNA Levels in Igl shRNA Transfectants.....	96

2-9	Summary of URE3-BP Protein Levels in URE3-BP shRNA Transfectants	98
2-10	Summary of mRNA Levels in GFP shRNA Control Transfectants, URE3-BP shRNA Transfectants, and Non-Transfected HM1:IMSS Trophozoites	100
Chapter 3. Partial Characterization of the <i>Entamoeba histolytica</i> Low Molecular Weight Protein Tyrosine Phosphatase		
3-1	Comparison of <i>E. histolytica</i> LMW-PTPs to Selected Representative LMW-PTPs Found in Plants	247
3-2	Comparison of <i>E. histolytica</i> LMW-PTPs to Selected Representative LMW-PTPs found in Bacteria.....	250
3-3	Comparison of <i>E. histolytica</i> LMW-PTPs to Selected Representative LMW-PTPs in Other Unicellular Eukaryotes	253
3-4	Comparison of <i>E. histolytica</i> LMW-PTPs to Selected Representative LMW-PTPs in Multicellular Organisms.....	256
3-5	Oligos Used to Clone the Wild-Type and Mutant Untagged and Epitope-Tagged <i>E. histolytica</i> LMW-PTP via PCR	259
3-6	Proteins Identified by Mass Spectrometry Analysis in the ~37 kDa Band from Silver-Stained Gels of the Pulldowns Using Mutant Substrate-Trapping LMW-PTP Attached to Affi-Gel 15 Beads Incubated with Amebic Lysate	261

3-7	Proteins Identified by Mass Spectrometry Analysis in the ~42 kDa Band from Silver-Stained Gels of the Pulldowns Using Mutant Substrate-Trapping LMW-PTP Attached to Affi-Gel 15 Beads Incubated with Amebic Lysate	267
3-8	Proteins Identified by Mass Spectrometry Analysis Analysis Selected for Epitope Tagging, Overexpression, and Further Characterization.....	273
3-9	Oligo Pairs used in PCR to Amplify the Genes for the Proteins Selected from the Mass Spectrometry Analysis Data in Table 3-8.....	275
3-10	Epitope Tags Added to LMW-PTP and Putative Substrate Clones.....	278

List of Abbreviations

aa	amino acid
AEBSF	4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride
Ala	alanine
Arg	arginine
Asp	aspartic acid
bp	base pair
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
Cys	cysteine
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylene diamine tetraacetic acid
EGTA	ethylene glycol tetraacetic acid
<i>E. dispar</i>	<i>Entamoeba dispar</i>
<i>E. histolytica</i>	<i>Entamoeba histolytica</i>
<i>E. invadens</i>	<i>Entamoeba invadens</i>
GalNAc	N-acetyl-D-galactosamine
Gal/GalNAc lectin	galactose- and N-acetyl-D-galactosamine-inhibitable lectin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	histidine
LMW-PTP	low molecular weight protein tyrosine phosphatase
µg	microgram
µl	microliter
µm	micrometer

μM	micromolar
mM	millimolar
MOPS	3-(N-morpholino)propanesulfonic acid
nt	nucleotide
oligo	oligonucleotide
PBS	phosphate-buffered saline solution at 1X concentration
PBS-Tween	PBS with 0.05% Tween 20 added
PCR	polymerase chain reaction
<i>p</i> NPP	<i>p</i> -nitrophenyl phosphate
PDGF	platelet-derived growth factor
PDGFr	platelet-derived growth factor receptor
PMSF	phenylmethanesulfonyl fluoride
PTP	protein tyrosine phosphatase
ROS	reactive oxygen species
qRT-PCR	quantitative real-time PCR
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser	serine
TBST	Tris-buffered saline with 0.1% Tween 20 added
TCEP	Tris(2-carboxyethyl)phosphine
Tris	2-Amino-2-(hydroxymethyl)-1,3-propanediol
Tyr	tyrosine

Chapter 1

Introduction to the Parasite *Entamoeba histolytica*

The Parasite *Entamoeba histolytica*

The human parasite *E. histolytica* is named for its ability to lyse human tissues [1]. The parasite recognizes a target cell and adheres to it via the Gal/GalNAc lectin (the major parasite adhesin), kills the target cell in a contact-dependent manner, and phagocytoses the apoptotic cell [1, 2, 3]. “Think of this protozoan parasite as a macrophage on steroids with pumped-up phagocytic, proteolytic, and cytolytic capabilities, invading human colonic mucosa, and occasionally penetrating through to the... circulation, reaching the liver, and causing fatal abscesses” [1]. There are three basic steps in the in vitro interaction of amebae with target cells: adherence, cytolysis, and phagocytosis [4, 5]. *E. histolytica* is an unusual unicellular protozoal parasite that phylogenetically is placed on one of the lowermost branches of the eukaryotic tree, closest to Dictostelium [6]. Infection with *E. histolytica* can cause amebic colitis (diarrhea with occult or gross blood in the stool) and liver abscesses in humans, while the related species *Entamoeba dispar*, genetically distinct but morphologically indistinguishable from *E. histolytica*, does not appear to cause disease in humans [7, 8, 9, 10].

The World Health Organization estimates that there are 50 million symptomatic clinical cases of amebic colitis annually worldwide, resulting in 100,000 deaths [7, 10, 12, 18]. Among protozoal infections, infection with *E. histolytica* is likely the second leading cause of death, after malaria [10, 11]. Humans are its only natural hosts, and there is no zoonotic reservoir [12, 13, 16] although other species, including several nonhuman primates, cats, dogs, and rats can also be infected [9] and there is a mouse model of amebic colitis [18, 19]. Parasite-induced diarrhea is

especially a concern in children, since persistent diarrhea in young children can result in long-term cognitive deficits [20, 21].

The Life Cycle of *Entamoeba histolytica*

E. histolytica organisms are spread by the fecal-oral route via contaminated food and water, and tend to endemically infect people in developing countries or under circumstances where hygiene is poor [11, 12]. *E. histolytica* has two stages in its life cycle: infective quadrinucleated spherical cysts 10-16 μm in diameter, and motile ameboid trophozoites, 20-40 μm in diameter (Figure 1-1) [9, 10]. The trophozoites are capable of locomotion via pseudopodia [9]. The cysts can survive for weeks in a moist environment [11]. After ingestion, the cysts pass through the stomach, where they excyst into four trophozoites in the small intestine [9, 13, 14] and the trophozoites pass into the colon. The trophozoites have a tropism for the colonic environment due to the large numbers of bacteria present in the colon (which are a food source for the trophozoites) and the mucosal surface to which they can bind [15]. Trophozoites eventually form tetra-nucleated cysts which then pass through the colon with the feces and complete the life cycle (Figure 1-1) [15]. 90% of people infected with *E. histolytica* are colonized asymptotically, with 10% of infections resulting in clinical disease [7, 10, 11, 12, 16, 17]. Infected carriers (cyst passers) can produce tens of millions of cysts daily, but do not suffer intestinal distress, and have no disease symptoms [13].

The *Entamoeba histolytica* Genome

E. histolytica has 9,938 predicted genes [22]. Its genes average 1.17 kb in size and comprise 49% of the genome [22]. The majority of *E. histolytica* genes lack introns; a quarter of *E. histolytica* genes contain a single intron, and only 6% contain two or more [22]. 96 of its genes

appear to have been laterally transferred from bacteria [22]. Slightly less than a third (31.8%) of its predicted proteins have no homologs in GenBank and are unique to this organism [22]. Its chromosome number is difficult to determine since (i) *E. histolytica* chromosomes do not condense at mitosis, (ii) there is extensive length variability between homologous chromosomes from different isolates, which may be due to expansion and contraction of sub-telomeric repeats [22], and (iii) *E. histolytica* can duplicate its genome more than once before cytokinesis [14, 23, 24]. There appear to be 14 linkage groups, [25] so *E. histolytica* likely has 14 chromosomes [26]. *E. histolytica* appears to be basically diploid like most other eukaryotes, but may be tetraploid for some chromosomes [25, 26].

Genome Reduplication in *Entamoeba histolytica*

E. histolytica appears to lack typical eukaryotic cell cycle checkpoints preventing reduplication of the genome until after cytokinesis [23], so the regulation of cell division in this organism differs from typical eukaryotes [14]. Homologous *Entamoeba* cell cycle-regulation genes are divergent from typical eukaryotic versions and thus may not have equivalent function [14]. *E. histolytica* cells in culture can duplicate their genomes several times before the cells divide, leading to polyploid cells, which are more prevalent in late log phase cultures [23, 24]. An *E. histolytica* nucleus may contain 1X to 10X genome copies, with some containing even more than this [23]. Multinucleated cells with up to four nuclei are also seen [9]. Since several genome copies are able to accumulate in each nucleus [23], this makes it difficult to perform genetic studies with this organism or to isolate recessive mutations [27]. There is no evidence of meiotic division in *E. histolytica*, and therefore no detectable homologous recombination of genes [14, 27, 28].

***Entamoeba histolytica* Organelles**

E. histolytica cells lack mitochondria; instead they have an atrophic remnant organelle that lacks a mitochondrial genome [6, 22, 29, 30]. There is no visible rough endoplasmic reticulum or Golgi apparatus [14] but *E. histolytica* encodes the basic elements of eukaryotic vesicle transport machinery. The cytoskeleton plays important roles in parasite motility, contact-dependent killing of other cells (bacterial and host) and phagocytosis of the killed cells [22].

***Entamoeba histolytica* Metabolism**

E. histolytica metabolism reflects its usual niche as a phagocytic resident of the human colon. These organisms have access to many bacterial and host-derived organic compounds, so they have lost many genes for mitochondrial metabolic pathways and *de novo* synthesis, and they have apparently acquired bacterial genes via lateral gene transfer, with 96 genes likely to be prokaryotic in origin [22]. *E. histolytica* lacks tricarboxylic acid (TCA) cycle and mitochondrial electron transport chain proteins, and so is an obligate fermenter, using bacterial-like fermentation enzymes [22]. Its glucose transporters are related to the prokaryote glucose/ribose transporter family [22]. The major donor group of the laterally transferred genes seems to be the *Cytophaga-Flavobacterium-Bacteroides* group of the phylum Bacteroidetes, with a majority (58%) of these genes encoding metabolic enzymes, while 41% have unknown functions [22]. The bacterially-related metabolic enzymes have increased the range of substrates available for energy generation and include tryptophanase, aspartase, and several glycosidases and sugar kinases [22]. *E. histolytica* may have acquired these genes from phagocytosed bacteria, since every engulfed bacterium could offer a potential gene replacement or new gene or metabolic pathway: that “you are what you eat” [31].

***Entamoeba histolytica* and Disease**

Amebiasis is a common parasitic infection caused by the protozoan *Entamoeba histolytica* [10, 32], and infected persons may or may not have apparent clinical symptoms [10]. Symptoms of *E. histolytica*-caused disease range from mild diarrhea to severe hemorrhagic colitis; liver abscesses can also result from *E. histolytica* infection [8]. The amebae can asymptotically persist in the human gut for months or even years [8]. The organisms are spread by the fecal-oral route via contaminated food and water, and tend to endemically infect people in developing countries or under circumstances where hygiene is poor [11, 12]. *E. histolytica* appears to have had a worldwide distribution before the advent of modern sanitation and hygiene. *E. histolytica* has been detected in latrine soil samples from thirteenth-century latrines used by Crusaders and pilgrims in the city of Acre (now in Israel), using the ELISA-based *E. histolytica* II assay kit (TechLab, Blacksburg, VA, USA), which detects the amebic Gal/GalNAc lectin (the major parasite adhesin); six of the eight samples tested were positive [33]. Another study tested human coprolites and cesspit soil samples of different ages collected from various locations around the world with the same assay kit, and found about 22% of the tested samples were positive for *E. histolytica*; the positive samples were from nineteenth-century Argentina, fourteenth- and eighteenth-century Belgium, third-century France, a 2500-year-old French sample, and pre-contact North American Anasazi samples dating to 1200-1300 A.D., with the oldest positive sample being a 5300-year-old Swiss sample [34].

Outbreaks of amebiasis are rare in developed countries, but they can occur if municipal water supplies are underchlorinated, or become contaminated with human fecal matter [6, 35]. In the summer of 1998, there was an outbreak of amebiasis in the city of Tbilisi in the Republic of Georgia, resulting in 177 cases of suspected amebiasis [35]. It was hypothesized that the municipal drinking water was the source of infection, either because of inadequate water

treatment or contamination of municipal water in the distribution system due to breaks in sewer pipes: the pipe breaks as well as scheduled service interruptions to repair breaks and leaking pipes created major pressure drops that could have led to contaminated sewage being sucked into the water distribution system [35]. Consumption of fresh produce was also correlated with illness; the produce could have been tainted by exposure to contaminated municipal water sources either at markets or at home [35]. In the USA, there was an outbreak of amebiasis in Chicago in 1933 during the World's Fair, with more than 1400 people infected and close to 100 deaths. The source of the infection was traced back to a single hotel where water and sewer pipes had been cross-connected, and an overhead sewer pipe was leaking directly into the main tank for cooling the drinking water used throughout the hotel [36, 37, 38]. A small outbreak in the vicinity of San Francisco that occurred at the beginning of December 1908 was thought to have resulted from the consumption of fresh produce contaminated either by exposure to tainted water or by the use of human waste as fertilizer [39].

Entamoeba histolytica infection is endemic in many parts of the developing world. In a two-year study of a cohort of Bangladeshi preschool children from an urban slum in Mirpur, Dhaka, Bangladesh. 55% of the children were infected (colonized) by *E. histolytica* with 4% having colitis [40]. After eight years of observation, of the 162 children in the cohort, 90% of them had been infected with *E. histolytica* at least once, and 68% of them had at least one repeat infection, with *E. histolytica* accounting for almost 8% of all diarrheal disease [21].

Local environmental factors, diet, and gut flora likely contribute to the outcome of *E. histolytica* infection [8], as well as both parasite and host genetic factors. Infection with different strains of *E. histolytica* can cause consecutive infections, and it appears only a subset of *E. histolytica* parasites infecting the intestine also have the capacity to cause liver abscesses [7, 40]. Amebic liver abscess is rare in children and is more common in adult males than females [8, 18,

40]. A human genetic factor influencing amebic infection is a mucin gene, MUC2, which is polymorphic, having at least 12 alleles in Caucasian populations [12, 41]. A protective association was observed for individuals having the HLA class II allele DQB1*0601 and the heterozygous haplotype DQB1*0601/DRB1*1501: in a cohort study of Bangladeshi children who reside in an urban slum in the Mirpur district of Dhaka, Bangladesh, it was found that individuals with the heterozygous haplotype were about 10 times more likely to test negative for both *E. histolytica* and serum anti-lectin IgG [21, 42]. Acquired immunity against *E. histolytica* is associated with a mucosal IgA rather than a serum IgG antibody response against the carbohydrate-recognition domain (CRD) of the Gal/GalNAc lectin Hgl (heavy) subunit [18, 21, 40, 42, 43; 44]. In the cohort study, 81 children who had detectable anti-CRD IgA in the first year of the study had only a 6% incidence of infection with *E. histolytica* in the following year, compared to the 43% infection rate in 149 anti-CRD IgA-negative children [21, 40, 44]. A mucosal anti-lectin IgA response in mice vaccinated with native lectin or with the CRD of the Hgl subunit protected the mice against intestinal amebiasis [18]. The IgG response associated with susceptibility to amebic infection runs in families and thus has a genetic component [40].

Another human genetic factor influencing *E. histolytica* infection has been recently identified from the cohort study: a single amino acid residue change in the leptin receptor, from a glutamine to an arginine residue at position 223, was associated with increased susceptibility to amebic infection [45]. Homozygosity for the arginine residue in children was associated with a nearly fourfold increase in the likelihood of having an *E. histolytica* infection, and was associated in adult males with a nearly twofold greater susceptibility to amebic liver abscess [45]. Children homozygous for the glutamine allele had an average duration of about 2.5 years of no amebic infections from the time of entry into the study until the first amebic infection; heterozygous children had an average of about 1.5 years and children homozygous for the arginine allele had an

average of 1.0 years, indicating a single copy of the arginine allele increased susceptibility to amebic infection [45]. Mouse studies using mice homozygous or heterozygous for the arginine residue had significantly greater epithelial ulceration compared with mice homozygous for the glutamine residue, supporting the human findings [45]. The mechanism of action is hypothesized to be that the leptin receptor may help to promote repair of tissue damaged by *E. histolytica* infection, and the arginine allele may be less effective in this role [45].

Known Virulence Factors of *E. histolytica*

Virulence factors of *E. histolytica* include the Gal/GalNAc lectin, composed of three subunits, which is involved in adhesion to host cells by binding N-acetyl-D-galactosamine and D-galactose terminal residues on host cell glycoproteins (Figure 1-2) The heavy subunit of the Gal/GalNAc lectin has antigenic crossreactivity to CD59, a human leukocyte antigen that prevents C5b-9 complement membrane attack complex formation [12, 46, 47]. Cysteine proteinases degrade mucins and host extracellular matrix, as well as C3a and C5a of complement, secretory IgA, serum IgG, and pre-IL-18; amebapores form pores in host cell membranes and have homology to the pore-forming granulysin and NK-lysin produced by cytotoxic T cells and natural killer cells [10, 22, 47, 48, 49, 50, 51]. Some of the cysteine proteinases have predicted N-terminal transmembrane anchors, which may allow them to be tethered and localized to the surface of the amebic membrane [22]. Apoptotic killing of host cells by *E. histolytica* appears to require activation of host cell caspase 3-like caspases and is caspase 8- and 9-independent [49].

Colonization of the Colonic Mucus by *Entamoeba histolytica*

The first step in trophozoite colonization of the human colon is the adherence of trophozoites to the colonic mucus layer, mediated by the galactose and N-acetyl-D-galactosamine

binding lectin (Gal/GalNAc lectin) (Figure 1-2) on the trophozoite surface which binds to terminal galactose (Gal) and N-acetyl-D-galactosamine (GalNAc) residues in the mucin molecules, with 1000-fold greater affinity to GalNAc than to galactose residues [12, 13, 15, 32, 52]. Mucins are the major glycoproteins of the mucus layer [15, 52]. One of these major mucins in the human intestine is MUC2, which is extensively modified with O-linked N-acetyl-D-galactosamine (GalNAc) residues, and its gene is highly polymorphic and may influence susceptibility to infection [12]. The gel-like mucus layer protecting the underlying colon cells has an average thickness of 150 μm , about 7 times the diameter of a trophozoite [13]. Once bound to the mucins, the trophozoites can multiply; in most cases they do not appear to damage the host [13]. Trophozoites locally disrupt the mucin layer by producing degradative cysteine proteinases [15, 22, 50, 53], reducing the viscosity of the mucus and allowing bacteria to diffuse there [15]. One model for trophozoite encystment postulates that as the amoebae multiply and disrupt the mucus layer, each trophozoite is exposed to a decreasing amount of mucus, until the local mucin concentration around the trophozoites is so reduced that this causes clustering of Gal/GalNAc lectin molecules on the trophozoite surface [15]. This clustering could signal through the Gal/GalNAc lectin and stimulate differentiation of the trophozoites into cysts which then pass through the colon with the feces and complete the life cycle [15]. Tens of millions of cysts can be passed daily by an infected individual [13].

Disease Following Colonization by *E. histolytica*

Disruption of the mucus layer can expose the intestinal epithelium, allowing trophozoites to contact the epithelial cells [47]. Once through the mucus layer, trophozoites can attach to Gal and GalNAc residues on epithelial cell glycoproteins via the Gal/GalNAc lectin, dissolve the extracellular matrix of the colonic epithelium via their cysteine proteinases and invade into the

intestinal tissues [11, 12, 50] killing the cells, causing inflammation, and ultimately causing ulceration of the epithelium as they invade deeper into the tissue (Figure 1-3) [1]. *E. histolytica* trophozoites readily ingest red blood cells [9] (Figure 1-3), binding equally well to human red blood cells from all ABO human blood groups [5]. The presence of trophozoites with ingested red blood cells in fresh stool is strongly correlated with *E. histolytica* infection (rather than colonization with the morphologically identical but genetically distinct nonpathogenic *Entamoeba dispar*) and invasive disease (Figure 1-3) [9, 10]. If trophozoites reach the bloodstream, they can spread to the liver, lung, or brain, most commonly forming a liver abscess [11, 12]. Whether or not the amebae invade into tissue or remain in the mucus layer depends on factors such as the genetics of the amebae and the host, the host immune response, the makeup of the host's bacterial flora and the thickness of the intestinal mucus layer [12].

The Galactose- and N-acetyl-D-Galactosamine-Binding (Gal/GalNAc) Lectin is a Major Amebic Virulence Factor

The Gal/GalNAc lectin is one of the best characterized virulence factors of *E. histolytica* (Figure 1-2). Lectins are ubiquitous carbohydrate-binding proteins involved in a variety of cellular processes, including adhesion of cells to extracellular matrix and other cells [54, 55]. The Gal/GalNAc lectin is the major defined adhesin involved in adherence [12, 54] with other binding proteins involved to a lesser extent, likely after initial binding of the Gal/GalNAc lectin to the host cell [12]. *E. histolytica* kills target cells through direct contact [5] by adhering to galactose or N-acetyl-D-galactosamine (GalNAc) residues on glycoproteins on target cell surfaces via the Gal/GalNAc lectin in the first step of contact-dependent killing [1, 12, 56] Adherence and subsequent contact-dependent cell killing is inhibited by galactose or GalNAc but not other monosaccharides [12, 56, 57, 58]. In addition to colonic epithelial cells, *E. histolytica* has been

shown to kill human macrophages, monocytes, neutrophils, and T lymphocytes in a contact-dependent manner [57, 32, 58]. Virulent *E. histolytica* trophozoites can rapidly kill other cells: in one study, at a ratio of one amoeba to 100 neutrophils, 93% of the neutrophils were killed by the amoebae after 22 hours at 37°C; blocking adherence by the amoebic Gal/GalNAc lectin with 50 mM GalNAc prevented the amoebae from killing the neutrophils, and allowed the neutrophils to destroy the amoebae [57, 59]. After adhering to a target cell and killing it, the trophozoite phagocytoses the cell [5]. The Gal/GalNAc lectin also mediates resistance against the complement C5b-9 complexes which form the membrane attack complex [32] and thus are not killed by complement [5]. Lectin adhesion to host cells requires the presence of terminal Gal or GalNAc residues on glycoproteins of target cells; Chinese hamster ovary (CHO) glycosylation-deficient mutant cells with no terminal galactose or GalNAc residues are almost completely resistant to amoebic adherence and thus to being killed by amoebae [1, 12, 57]. The Gal/GalNAc lectin may be transferred to the membranes and also the cytoplasm of target epithelial cells, such as cultured hepatocytes, after trophozoite contact, and the lectin may also be secreted in the presence of target cells [4, 12, 60].

Structure of the Gal/GalNAc Lectin

The Gal/GalNAc lectin is composed of three subunits, the heavy subunit (Hgl), the light subunit (Lgl) and the intermediate subunit (Igl) (Figure 1-2). Hgl, the 170 kDa subunit, is the only one integral to the membrane, having an extracellular domain, a transmembrane segment, and a short section extending internally into the cytoplasm, and seems to be involved in intracellular signaling [12, 61, 62] since the cytoplasmic tail has homology to the cytoplasmic domains of $\beta 2$ and $\beta 7$ integrins, which bind the signaling molecules Shc and Grb2 [56, 62]. Hgl contains a known carbohydrate-recognition domain (CRD) in its extracellular portion that binds galactose

and GalNAc residues [16, 63]. Lgl is not an integral membrane protein, but is GPI (glycosylphosphatidyl-inositol)- anchored to the membrane at its C-terminal end, contains five conserved cysteine residues, and is disulfide-bonded to Hgl, forming a 260 kDa heterodimer, which is a novel configuration [3, 12, 32, 54, 64]. Lgl has major 35- and 31-kDa isoforms, differing in their post-translational modifications, and does not appear to be directly involved in adherence [12, 32, 54, 65] since it itself has no Gal or GalNAc binding activity [65]. Lgl plays a role in *E. histolytica* virulence: the avirulent Rahman strain has low transcription levels of *lgl*; an antisense transfectant of the virulent strain HM1:IMSS with a 60% reduction in gene expression of *lgl* had a significant decrease in virulence; and a HM1:IMSS transfectant overexpressing a truncated form of Lgl lacking the 55 N-terminal amino acids exhibited a dominant negative effect and had a significantly reduced ability to adhere to, kill, and phagocytose erythrocytes [3, 48]. Deletion of the Lgl GPI anchor prevents it from assembling into the Hgl-Lgl heterodimer [65]. The 150-kDa intermediate subunit Igl is also GPI-anchored and is noncovalently associated with the Hgl-Lgl heterodimer (Figure 1-2) [12, 66]. Igl appears to lack a carbohydrate-recognition motif, but has been shown to bind Gal and GalNAc and is thus identified as a galactose- and GalNAc binding lectin: when purified Igl was used to pre-treat CHO (Chinese hamster ovary) cells, adherence of *E. histolytica* trophozoites was significantly inhibited in a dose-dependent manner, but pretreatment of CHO cells with both purified Igl and Gal or GalNAc, but not glucose, allowed binding of trophozoites to the CHO cells [12, 56, 66]. Igl proteins contain CXXC motifs [56], and this motif is also found in a large family of putative transmembrane kinases in *E. histolytica* [67]. The Igl proteins also have limited sequence identity with *Giardia lamblia* variant surface glycoproteins, including some of the CXXC motifs, which in *G. lamblia* are implicated in protein-protein interactions [56].

Known *Entamoeba histolytica* Signal Transduction Molecules

E. histolytica utilizes a variety of different signal transduction systems to sense and interact with the environment, which includes about 270 putative protein kinases, including more than 90 receptor serine/threonine/tyrosine kinases [22, 67]. The essential kinase motifs are all conserved, but serine/threonine or tyrosine kinase specificity could not be determined based only on sequence analysis [67]. These receptor kinases fall into three groups based on their extracellular domains. The first group contains 50 receptor transmembrane kinases which have CXXC-rich repeats, which are also found in the Gal/GalNAc lectin intermediate subunit (Igl) and the *Giardia lamblia* variant-specific surface proteins [56, 67]; their most significant feature is a kinase domain that is distinct from other known kinases and has similarity to both serine/threonine and tyrosine kinases [67]. The second group of 32 has cysteine-rich CXC repeats, while the third group lacks cysteine-rich extracellular domains [22].

An *Entamoeba histolytica* Virulence-Related Transcription Factor

The Gal/GalNAc lectin is an important virulence factor, since it is required for a trophozoite to bind to mucins, or to target cell glycoproteins; its expression, and that of a number of other virulence-associated proteins, is regulated by URE3-BP, Upstream Regulatory Element-3 Binding Protein, a calcium-regulated transcription factor [68, 69]. URE3-BP binds to the Upstream Regulatory Element 3 (URE3) motif, TATTCTATT, found in the third of five major regulatory regions in the promoter of one of the genes encoding the heavy subunit of the Gal/GalNAc lectin, *hgl5* [70]. URE3-BP also regulates the expression of *fdx1*, another virulence-associated gene; URE3-BP downregulates transcription of *hgl5* and upregulates transcription of *fdx1* [68], so it can act an activator or repressor [69]. Transcriptome profiling of amebae inducibly expressing constitutively active URE3-BP was used to identify other genes regulated by URE3-

BP, and of the fifty modulated transcripts identified, more than half had the URE3 motif in their promoter regions [69]. Fifteen of these URE3-BP-regulated genes were annotated as potential membrane proteins, several of which were novel, and four other URE3-BP-regulated genes were metabolic enzymes that participate in the phospholipid degradation/ fatty acid assimilation pathway [69] suggesting that one function of URE3-BP is to remodel the surface of *E. histolytica* in response to a calcium signal [69]. Expression of the constitutively active URE3-BP in amoebae leads to an increase in their rate of transwell migration, which suggests a role in the regulation of cellular motility for URE3-BP [69]; increased motility of the parasite may play a role in virulence.

Goals of This Dissertation

In the Petri laboratory we are interested in studying virulence genes: the how and why of the ability of this parasite to cause disease: what stimulates a trophozoite to destroy and invade the intestinal epithelium or liver tissue, and how is that regulated? Why do the majority of infections in human hosts result in asymptomatic colonization, with only some resulting in tissue damage? What role does the internal environment of the host play?

Since the Gal/GalNAc heavy subunit does appear to have signaling capacity, and the other identified signal transduction molecules that are found in the membrane, such as the putative transmembrane receptor kinases, likely are involved in signaling, we were interested in developing tools such as gene knockdown that could be used to investigate the function of the individual proteins involved directly in virulence, or that are involved in signaling or propagating a signal that would induce an *Entamoeba histolytica* trophozoite to become more virulent.

In Chapter 2, the short hairpin RNA interference system is discussed, which was developed as a tool to knock down protein expression in *Entamoeba histolytica*, since there have

been no reports of gene knockouts generated via homologous recombination in this organism. We were able to knock down protein levels of four virulence-related genes with this tool: PATMK, which is involved in phagocytosis; URE3-BP, a calcium-regulated transcription factor; EhC2A, which translocates and sequesters URE3-BP to the plasma membrane; and Igl, the intermediate subunit of the Gal/GalNAc lectin. We were able to knock down protein levels by up to 99% using this system.

In Chapter 3, the partial characterization of a small phosphatase protein, the low molecular weight protein tyrosine phosphatase, is discussed. This protein has been characterized in mammals as regulating cell proliferation, motility, and adhesion through dephosphorylation of such substrates as growth factor receptors, cytoskeleton-associated proteins, and transcription factors. We were interested in the function of the LMW-PTP in *E. histolytica*, since it might play a role in virulence, as proliferation, motility, and adhesion are needed attributes for *E. histolytica* to cause disease. The protein structure has been solved, providing a picture of the ligand-induced conformational changes necessary for binding of substrate. The residues involved in catalysis are all conserved, and the recombinant protein was found to have phosphatase activity. Outside *Entamoeba*, the closest matches to this enzyme are found in selected plant or *Bacteroides* species. A mutant LMW-PTP protein in which the catalytic cysteine in the active site was replaced with a serine lacked phosphatase activity and was used to identify a number of possible substrate proteins via mass spectrometry.

Figure 1-1. The Life Cycle of *Entamoeba histolytica*. The life cycle of the parasite is simple, consisting of two stages: the quadrinucleated cysts, and the motile trophozoites. When cysts are ingested by a human host in contaminated food or water, the cysts pass through the stomach and into the small intestine, where four trophozoites excyst. The trophozoites are carried to the colon, where they colonize the mucus layer. 90% of humans colonized with this organism are infected with no symptoms, while 10% suffer invasive disease. The trophozoites replicate, ultimately encyst, and the cysts are passed out with the feces. The life cycle begins again when a new human host ingests the cysts.

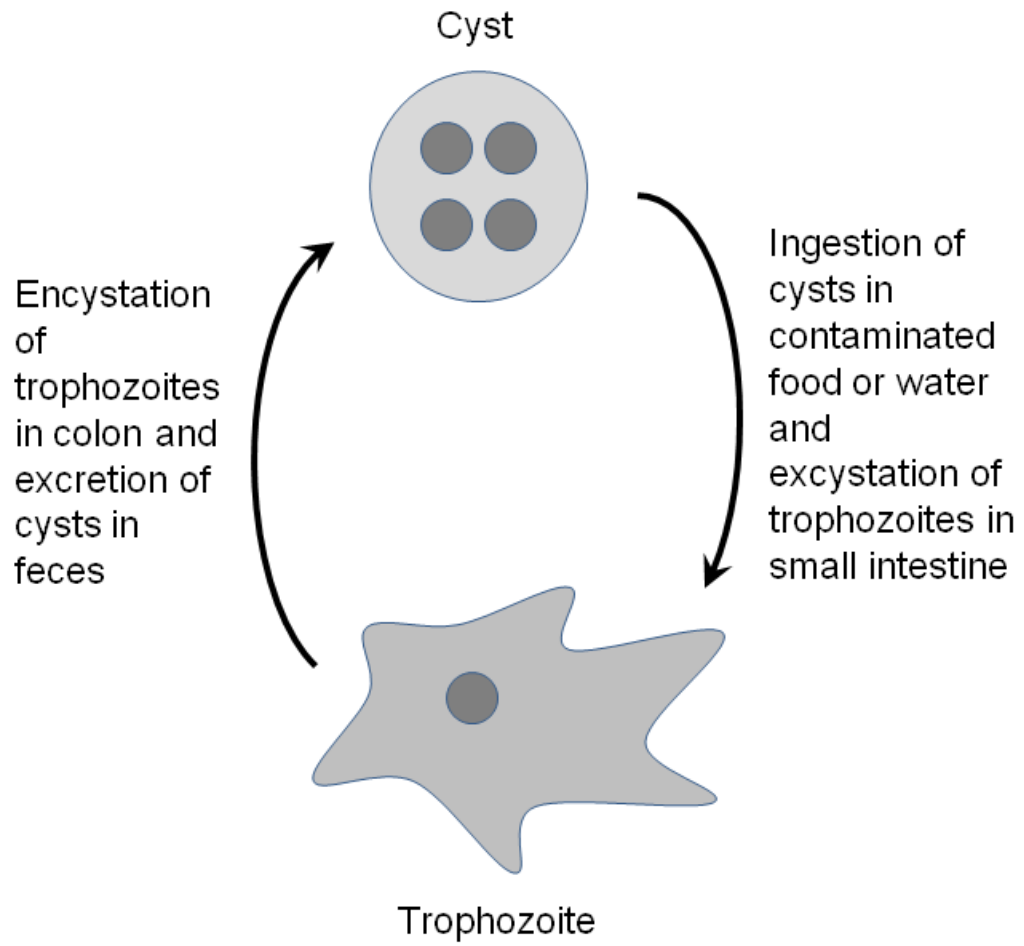


Figure 1-2. Diagram of the Gal/GalNAc Lectin in the *E. histolytica* Plasma Membrane. The galactose and N-acetyl-D-galactosamine binding lectin (Gal/GalNAc lectin) is found on the trophozoite surface. The Gal/GalNAc lectin binds to terminal galactose (Gal) and N-acetyl-D-galactosamine (GalNAc) residues in mucin molecules and on host cell glycoproteins. There are three subunits, Hgl, Igl, and Lgl, which are the heavy, intermediate, and light subunits respectively. The 170 kDa heavy subunit is an integral membrane protein containing the carbohydrate recognition domain that has been shown to bind galactose and GalNAc, and it has a short section extending internally into the cytoplasm, which seems to be involved in intracellular signaling. The GPI-anchored extracellular membrane protein Lgl is 31-35 kDa in size and forms a heterodimer with Hgl. The intermediate subunit Igl is 150 kDa in size. Igl, like Lgl, is an extracellular membrane protein GPI-anchored to the plasma membrane surface, and is noncovalently associated with the Hgl-Lgl heterodimer. Igl has also been shown to bind galactose and GalNAc. Hgl = heavy subunit; Lgl = light subunit; Igl = intermediate subunit; CRD = carbohydrate recognition domain in the heavy subunit. Figure originally published in Petri *et al* (2002) [12].

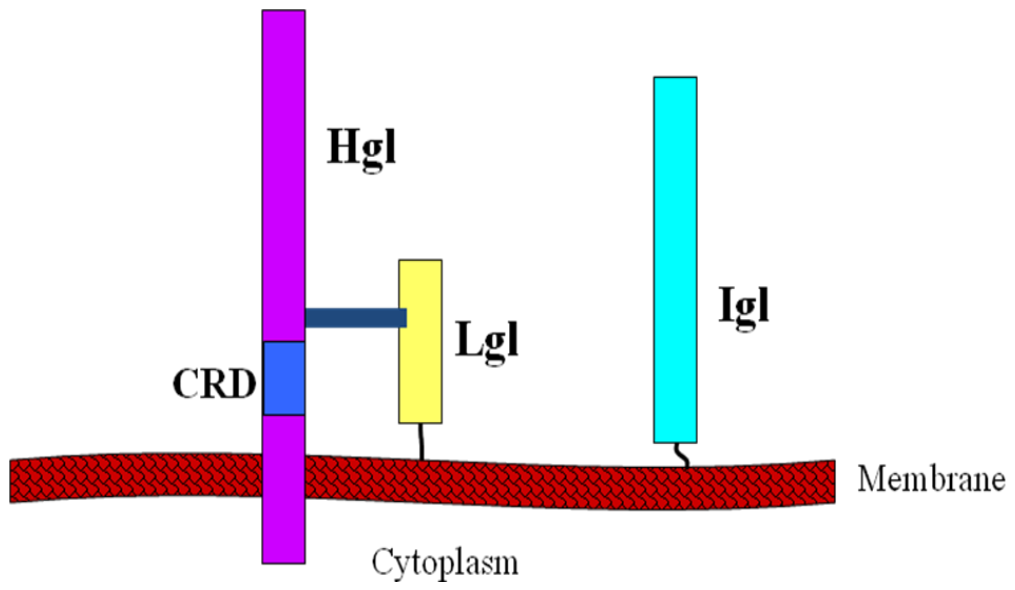
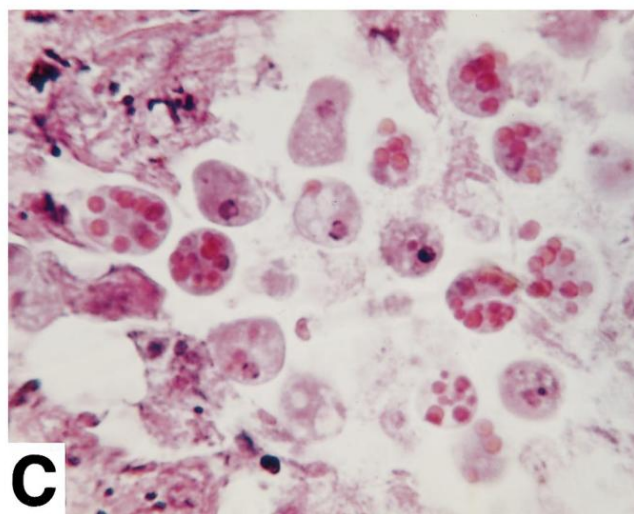
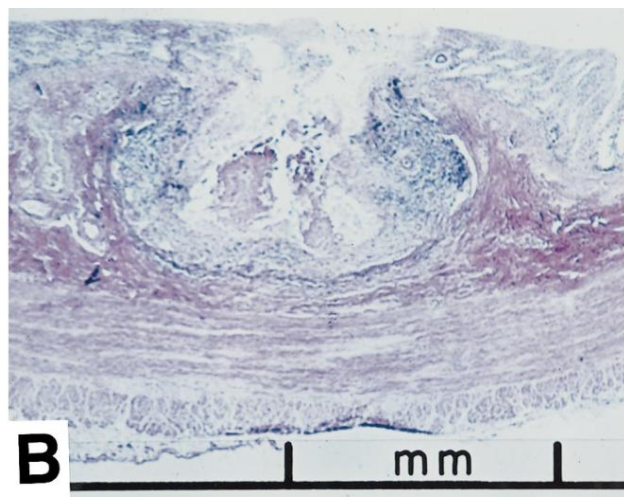
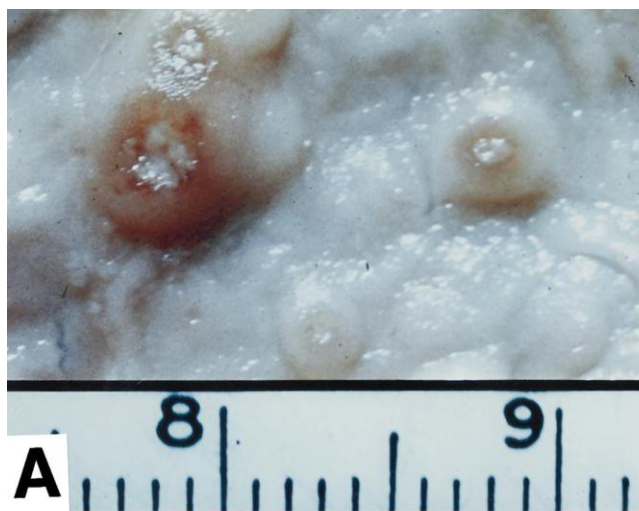


Figure 1-3: Invasive Amebiasis with Tissue Destruction. (A). Colonic ulcers of the intestinal epithelium, of about 1 mm diameter. (B). Cross-section of one of the colonic ulcers. The ulcer has the characteristic flask shape. Note the tissue debris in the center. (C). *Entamoeba histolytica* trophozoites in the ulcer. They contain ingested red blood cells, and are surrounded by the tissue debris resulting from their destruction of the sub-mucosa. Figure originally published in Petri and Singh (1999) [71].



Chapter 2

Short Hairpin RNA-Mediated Knockdown of Protein Expression in *Entamoeba histolytica*

Note: The shRNA data on Igl, URE3-BP, and EhC2A has been previously published in Linford *et al* (2009) [72]. Some of the shRNA data on PATMK was included in Boettner *et al* (2008) [73]. Characterization of the role of PATMK in *Entamoeba histolytica* was performed by Douglas Boettner [73] and that of EhC2A was performed by Heriberto Moreno [74]. Studies characterizing URE3-BP function, except for those involving shRNA knockdowns, were performed by Carol Gilchrist [68, 69, 75, 76, 77].

Background on RNA Interference (RNAi)

The discovery that double-stranded RNA (dsRNA) introduced into eukaryotic cells can initiate post-transcriptional sequence-specific gene silencing of cellular genes [78] via translational repression or degradation of messenger RNA (mRNA) has become an important tool in assessing or manipulating gene function in most eukaryotic cells. This mechanism of RNA interference (RNAi) may have evolved as a defense against viruses and transposable elements with dsRNA intermediates [79, 80, 81, 82, 83, 84, 85, 86, 87]. Unexpected gene silencing was observed first in petunia plants [88, 89], followed by reports from researchers working with worms (*Caenorhabditis elegans*), a fungus (*Neurospora crassa*), fruit flies (*Drosophila melanogaster*), and mice; the common factor was the introduction of DNA or RNA homologous to the sequence of an endogenous gene into cells, resulting in a decrease in the mRNA level of that endogenous gene [89]. The causative agent of the knockdown was discovered by Fire *et al* in 1998; injection of dsRNA caused “potent and specific genetic interference” in *C. elegans* [78].

Hamilton and Baulcombe (1999) first identified the small RNA intermediates in this process (~25 basepair (bp) in size) in plants which were undergoing silencing [90]. These small RNA intermediates, named short interfering RNAs (siRNAs), result from the dsRNA being cleaved at 21 to 23 bp intervals [91] and are involved in the degradation of complementary mRNA [92, 93, 94, 95]. The dsRNA is cleaved into siRNAs by an RNase III-type protein, Dicer [96] resulting in dsRNA duplexes with symmetrical two or three nucleotide (nt) 3' overhangs (Figure 2-1) [96, 97, 98]. A protein with tandem dsRNA-binding domains, R2D2 in *Drosophila* [99] and RDE-4 in *C. elegans* [100], is found to be associated with Dicer (Figure 2-2) [97, 99, 100]. This Dicer-dsRNA-binding protein heterodimer remains associated with the siRNA [97, 99, 101], and the siRNA-protein complex is incorporated into a larger complex, the RNA-induced silencing complex (RISC) (Figure 2-2) [97, 99, 102] which includes Argonaute "Slicer" protein [91, 93, 97, 98, 103, 104]. The incorporated siRNA is unwound, which requires an RNA helicase (Armitage in *Drosophila*) [97, 102] and its antisense strand is used as a guide for the RISC to find its target mRNA match, with the other strand, the passenger strand, being lost, and the mRNA is then cleaved by Argonaute [89, 98, 103, 105, 106, 107, 108, 109], or the equivalent protein, Aubergine, in *Drosophila* [110] (Figure 2-2).

Synthetic siRNAs, when introduced into mammalian cells, can mediate gene silencing, so the "dicing" step of dsRNA into siRNAs could be bypassed [92]. This was an important discovery, allowing RNAi to be used effectively in mammalian cells, since the interferon response is activated in response to dsRNA longer than 30 bp, ultimately causing sequence-nonspecific mRNA degradation [92, 98, 105, 107, 111, 112]. The antisense strand of the siRNA shows greater silencing efficiency if it has, relative to the mRNA target, a relatively thermodynamically unstable 5' end, because this favors incorporation of the antisense rather than

the sense strand of the siRNA into the RISC effector complex, where it is used as a guide to find the cognate mRNA [113, 114].

MicroRNAs

Most eukaryotic cells have an endogenous RNA interference pathway that (usually) results not in mRNA cleavage and subsequent degradation, but prevents transcription of mRNAs. The effectors, called microRNAs (miRNAs), are small noncoding RNAs that direct the posttranscriptional silencing of gene expression [98, 115, 116]. miRNAs are similar to siRNAs, use much of the same machinery as RNA interference, and also function in the form of ribonucleoprotein complexes, miRISCs (miRNA-induced silencing complexes) [116]. miRNAs are derived from longer primary transcripts that are expressed from RNA polymerase II promoters [115, 117, 118]. miRNAs extensively regulate gene expression in animals, plants, and protozoa [116]. The majority of miRNA genes are located in intergenic regions, or in antisense orientation to annotated genes [87, 115, 118, 119], with most of the remaining miRNA genes being located in the introns of the transcribed genes that they regulate [118]. These primary transcripts, pri-miRNAs, can be up to thousands of nucleotides in length, forming a long RNA hairpin which contains frequent mismatches and bulges [115, 117]. These long hairpins are cleaved in the nucleus into shorter ~70 nt RNA hairpins called pre-miRNAs, by Drosha [115, 120] in conjunction with a double-stranded RNA recognition protein, which in *Drosophila* and *C. elegans* is named Pasha and in mammalian cells DGCR8 [115, 120, 121, 122, 123]. The pre-miRNAs are exported into the cytoplasm from the nucleus by Exportin 5 [115, 120, 124, 125] where they are recognized and cleaved into the 22 nt miRNAs by Dicer in conjunction with another dsRNA-binding protein, called Loquacious in *Drosophila* and TRBP in mammals [115, 125, 126]. The miRNA is taken up by the effector complex, miRISC, the equivalent of the

siRNA-RISC, where the miRNA is unwound, the passenger strand is lost, and the miRNA guides the recognition of the miRNA-binding sites on the target mRNA, leading to translational repression and therefore inhibition of protein synthesis [97, 98, 115, 116]. In general, miRNA-mRNA duplexes contain mismatches and bulges in the center of the miRNA which prevents the cleavage of the mRNA [116]. It is the amount of complementarity that the guide siRNA or miRNA has with the mRNA target that appears to determine whether the mRNA will be cleaved. Even though siRNAs and miRNAs are processed from different dsRNA precursors, they can be functionally interchangeable, since a partially complementary siRNA can direct translational repression of a target site, while a fully complementary miRNA can direct cleavage of a target mRNA [97, 115, 128, 129]. Most of the miRNA sites that have been characterized are found in the 3' untranslated regions (UTRs) of the mRNA [116, 130, 131]. In animals, miRNAs have been found to also target 5'UTRs and coding regions of mRNAs [116, 130].

RNA-Dependent RNA Polymerases and RNA Interference

RNAi effects can be amplified by the action of RNA-dependent RNA polymerases (RdRPs). The primary siRNAs resulting from cleavage of the original introduced “trigger” dsRNA [132] can act as guides for RdRPs which form new dsRNAs using the cognate mRNA as a template, with the siRNAs acting as primers to copy and amplify the mRNAs [86, 133, 134, 135]. The resulting dsRNA is then cleaved into secondary siRNAs, resulting in amplification of the RNAi signal, and thus increasing the potency of the original introduced siRNA or dsRNA [133], since the new dsRNAs are subsequently degraded into more siRNAs [133, 136]. These secondary siRNAs have sequences corresponding to target mRNAs but differing from the original “trigger” dsRNA [86, 132, 133, 135]. These secondary siRNAs are more effective at inducing mRNA destruction than the primary siRNAs [137]. Genes for RdRPs have now been identified in

most organisms, and were first identified in such species as tomato [138, 139], *Arabidopsis thaliana* [140, 141], tobacco (*Nicotiana benthamiana*) [141], *Neurospora crassa* [142] and *Dictyostelium discoideum* [143]. One model for amplified gene silencing by RNAi is that primary siRNAs, formed from the original dsRNA, act as guides (presumably in an Argonaute-dependent manner) to recruit RdRP to targeted transcripts (Figure 2-3) [132]. It was found when short oligos are used as primers in vitro, tomato RdRP can synthesize complementary RNA from both RNA and DNA templates (Figure 2-4) [138]. The tomato RdRP is also capable of unprimed synthesis of RNA: without a primer, the RdRP initiates synthesis at or near the 3' terminal nucleotide of a single-stranded template: the nucleotide where the RdRP begins transcription needs to have phosphate groups both 5' and 3', so if the very first 3' terminal nucleotide has a hydroxyl group, then the transcript will begin at the penultimate nucleotide (Figure 2-4) [138].

RNA interference-associated proteins in *Entamoeba histolytica*

E. histolytica possesses the basic molecular machinery for RNAi. It has a gene [GenBank: XM_645408] [22, 144, 145] encoding a protein which has a single RNase III domain and possesses RNase III activity, and could perform the Dicer role as a dimer (Figure 2-1, Figure 2-2). Most Dicers have two RNase III domains [146, 147]; however, the RNAi-competent budding yeast *Saccharomyces castellii* Dicer also contains a single RNase III domain, which appears to function as a dimer [146, 148], so there is evidence for other “atypical” Dicers that function in RNAi. *E. histolytica* also has two Argonaute homologs [GenBank: XM_651344, XM_651422] [22, 50, 144, 145], and an RNA-dependent RNA polymerase gene (Figure 2-3, Figure 2-4) [GenBank: XM_646217] [22, 144]. What appear to be secondary siRNAs resulting from the action of the RNA-dependent RNA polymerase protein have been detected in *E.*

histolytica [149], and seventeen putative miRNAs have been identified using a bioinformatic approach [144].

Why use RNA interference for knockdown of gene expression in *Entamoeba histolytica*?

Exploitation of RNAi for gene expression knockdown is an attractive approach for *E. histolytica*, as there is no evidence for meiotic division or detectable homologous recombination of genes [14, 27, 28]. An additional challenge is that in *E. histolytica*, multiple copies of the genome, and even multiple nuclei, can accumulate before cell division [23, 24]. This is due to an apparent lack of usual eukaryotic cell cycle regulatory checkpoints [23, 24], since *Entamoeba* cell-cycle regulation genes are divergent from typical eukaryotic versions and may not have equivalent function [14]. This presents difficulties in studying gene function, as it has not been possible to generate gene knockouts using standard techniques, and difficult to isolate recessive mutations [27, 150]. The study of the function of individual genes in the past has therefore been limited to other techniques, such as over-expression of wild-type or mutant genes, or by other methods of gene inactivation such as antisense [150, 151, 152]. Methods of RNAi used in *E. histolytica* to knock down protein expression have included the use of long dsRNA being expressed by an *E. histolytica* RNA polymerase II promoter, which was used successfully to knock down expression of the *E. histolytica* proteins Diaphanous, Klp5, and ESTIRP [27, 153, 154] or by soaking the cells in artificial siRNAs to knock down expression of γ -tubulin [28]. These reports of RNAi use in *E. histolytica* showed knockdown of a single gene or of a gene family of related proteins. Here, we report in this study the success of the method of expression of short hairpin RNAs driven by the *E. histolytica* U6 promoter to knock down protein expression in *E. histolytica* of four unrelated genes.

Expression of short hairpin RNAs

Short hairpin RNAs (shRNAs) have a similar structure to siRNAs except the sense and antisense strands are connected at one end by a short loop, function like siRNAs to knock down gene expression (Figure 2-1) [155], and like artificial siRNAs, do not induce interferon [156]. The shRNAs can be produced from an expression vector as a single transcript from an RNA polymerase III promoter. The eukaryotic U6 promoter offers two advantages over other RNA polymerase III promoters for shRNA expression: first, the promoter region immediately upstream of the transcribed sequence for the U6 small nuclear RNA gene includes all the required regulatory elements, in contrast to most RNA genes transcribed by RNA polymerase III, which usually include regulatory elements within the gene sequence; second, the termination sequence is very short, consisting of 4 to 6 thymidine residues [155, 157, 158, 159, 160, 161, 162]. The first nucleotide in the shRNA to ensure effective transcription should be a G [155, 156, 158]. A variety of shRNA stem and loop lengths have been tested, including a 19-bp stem with a 3-nt loop [155]; a 21-bp stem with a 9-nt loop [156]; 29-bp and 19-bp stems with a 4-nt loop [163], and 29- and 19-bp stems with both 4-nt and 9-nt loops [164]; 19-base stems with 5-nt, 7-nt, and 9-nt loops were tested by Brummelkamp *et al* (2002) in a system using the H1 promoter, which like the U6 promoter, normally drives transcription of a small RNA, finding that only the 9-nt loop produced siRNAs [162]. This loop sequence, UUCAAGAGA [162], has become commonly used in shRNA constructs, and is also used in the constructs in this study. Longer hairpins with 29-bp stems appear to be better inhibitors of gene expression than ones with shorter 19-21 bp stems; it may be that longer shRNAs are more effective since they better resemble endogenous microRNAs (miRNAs), which regulate gene expression [165, 166]. This increased effectiveness has also been seen for similarly-sized longer artificial siRNAs, with only one siRNA generated per longer shRNA or siRNA [113, 165, 167].

Genes selected for knockdown

The four genes selected for knockdown, PATMK, Igl, URE3-BP, and EhC2A, are genes involved in amebic virulence. They were selected since as known virulence-related genes, they could be used to test if shRNA knockdown was possible in *E. histolytica*, and we were seeking to create an additional tool for further studying the function and role of these genes, and others, in amebic virulence.

PATMK, Phagosome-Associated Transmembrane Kinase 96 [GenBank: XM_650501], is the protein Douglas Boettner characterized for his PhD dissertation [73]. PATMK is a 146 kDa protein [73]. It is a member of a transmembrane kinase family first identified by its similarity to Igl in its CXXC motifs; more than 90 putative transmembrane kinases (TMKs) have been identified in the *E. histolytica* genome [56, 67]. PATMK is found in the early phagosome [73]. It contains an intracellular putative kinase domain, a short membrane-spanning region, and has an ectodomain containing 25 CXXC repeats [67, 73]. Its sequence contains residues necessary for both serine/threonine and tyrosine kinase activity, so its kinase specificity could not be predicted and it may actually be a pseudokinase, since it lacks the usual conserved glycine-rich ATP-orienting motif expected to be found in an active kinase [73].

Igl, a 150 kDa protein, is the intermediate subunit of the galactose- and N-acetyl-D-galactosamine (Gal/GalNAc) -binding lectin [56, 66]. The Gal/GalNAc lectin, the major defined amebic adhesin, is a virulence factor mediating adherence to target cells in the first step of contact-dependent cell killing [12, 168, 169]. It is composed of three subunits: the heavy subunit Hgl, which has a defined carbohydrate-recognition domain, the intermediate subunit Igl, and the light subunit Lgl (Figure 1-2) [12, 56]. Igl and Lgl are both glycosylphosphatidylinositol- (GPI-) anchored to the membrane surface, and while Lgl is disulfide-bonded to Hgl, Igl is non-covalently associated with the Hgl-Lgl heterodimer [12]. The lectin binds to terminal galactose or

GalNAc residues in glycoproteins on the surfaces of target cells [1, 12, 54, 56]. Igl was identified in a study to screen murine antibodies that would recognize surface proteins of *E. histolytica* and not those of the related non-pathogen *Entamoeba dispar*, using *E. histolytica* strain HM1:IMSS trophozoites [66, 170]. Its two genes, Igl1 (GenBank: AF337950) and Igl2 (GenBank: XM_647302) code for ~1100 aa proteins that are 81% identical and contain 32 CXXC repeats [56]. CXXC repeats are also found in the family of transmembrane kinases and the *Giardia lamblia* variant-specific surface proteins [8, 67, 171]. Igl, along with Hgl, is also found in the phagosome [73, 172]. Igl has the ability to bind to galactose and GalNAc [66, 169]. Adherence of *E. histolytica* trophozoites to CHO cells pretreated with purified Igl was significantly inhibited; pretreatment of CHO cells with both Igl and galactose or GalNAc, but not Igl and glucose, allowed binding of trophozoites to the CHO cells [66]. The specific role of Igl in virulence is still undefined.

URE3-BP, Upstream Regulatory Element 3-Binding Protein [GenBank: AF291721], is an unusual 22.6 kDa calcium-regulated transcription factor having two canonical EF hand motifs [68, 69, 76] and one non-canonical EF hand motif [74]. The human neuronal protein DREAM (Downstream Regulatory Element Antagonist Modulator) is the only other known calcium-responsive eukaryotic transcription factor with EF hands [68, 74, 76, 173]. Neither URE3-BP nor DREAM contains a canonical DNA-binding domain [68]. URE3-BP binds to a specific DNA sequence, the Upstream Regulatory Element 3 (URE3) motif, which has the consensus sequence TATTCTATT [68, 69, 75, 77]. The URE3 motif is one of five regulatory regions in the promoter for *hgl5* (which encodes a heavy subunit of the Gal/GalNAc lectin), and is also present in the ferredoxin (*fdx1*) promoter [68, 70, 75, 77]. URE3-BP can function to either activate or repress gene expression [69, 75, 77]. URE3-BP action is regulated by calcium: when URE3-BP binds calcium, it dissociates from the URE3 motif [69, 74, 76, 77]. URE3 motifs have been found in

other promoters, and are enriched in promoters of genes with modulated expression in the mouse model of amebic infection [69, 77]. URE3-BP localizes to both the cytoplasm and the nucleus in trophozoites [69, 74] and can also be found localized to the inner plasma membrane [68, 74]. There is an additional level of calcium-activated regulation of URE3-BP: it interacts with a calcium-responsive binding partner [74, 77]. Increased levels of intracellular calcium inhibit URE3-BP binding to the URE3 motif, promoting the interaction between URE3-BP and the *Entamoeba histolytica* C2A protein (EhC2A) with EhC2A sequestering URE3-BP at the plasma membrane [74, 77].

EhC2A protein. *Entamoeba histolytica* C2A protein (EhC2A) [GenBank: XM_650207] was identified and characterized by Heriberto Moreno for his PhD dissertation [74]. This protein is a 22 kDa calcium-binding membrane protein which has a conserved N-terminal calcium-binding C2 domain and a proline-rich C-terminal tail [72, 74]. C2 domains are Ca²⁺-binding motifs that allow calcium-dependent anchoring to, or interaction with, phospholipids in membranes, and can target proteins to specific membranes in eukaryotic cells [74]. Other proteins containing C2 domains include the tumor suppressor PTEN and copine proteins [174]. EhC2A is localized to the cytoplasm and plasma membrane [74]. In the presence of increased calcium, EhC2A and URE3-BP are strongly associated with the plasma membrane [74]. This translocation and sequestering of URE3-BP by EhC2A to the plasma membrane appears to be an additional enhanced regulatory mechanism, since the sequestered URE3-BP is no longer in the nucleus, and is not available to bind to the promoters of the genes it regulates [74].

Materials and Methods

Culture of Trophozoites

E. histolytica strain HM1:IMSS trophozoites were grown axenically in TYI-S-33 (Trypticase-yeast extract-iron-serum) (TYI) medium supplemented with 1X Diamond's vitamins (SAFC Biosciences, Lenexa, KS, USA), 15% heat-inactivated bovine serum (Gemini Bio-Products, West Sacramento, CA), 100 U of penicillin/ml and 100 mg streptomycin sulfate/ml (Gibco/Invitrogen, Carlsbad, CA, USA), at 37°C in T-25 tissue culture flasks [175] in a volume of 50 ml.

Transfection of Amebae

Plasmid DNA for each construct was prepared using the HiSpeed Qiagen Maxi Kit (Qiagen, Valencia, CA, USA). The plasmid DNA was ethanol-precipitated overnight at -20°C, and was resuspended to a concentration between 1 and 5 µg/µl in sterile 10 mM Tris pH 7.5. Medium 199 (M199) (Gibco BRL/Invitrogen, Carlsbad, CA, USA) was supplemented with 5.7 mM cysteine, 25 mM HEPES, and 0.6 mM ascorbic acid [176], adjusted to between pH 6.8 and 7.0 and filter-sterilized. Twenty (or thirty) µg plasmid DNA was diluted in 100 (or 200) µl supplemented M199 medium (M199S) in sterile 2 ml microcentrifuge tubes was mixed with 15 (or 30) µl of SuperFect or Attractene transfection reagent (Qiagen, Valencia, CA, USA), and incubated at room temperature to allow transfection-complex formation as per the manufacturer's instructions. Heat-inactivated bovine serum was added to the remaining M199S to a 15% concentration. Amebae were harvested by tapping the tissue culture flasks on the benchtop or by icing the flasks for 15 minutes, and flask contents were transferred in the hood to 50 ml conical tubes. Conical tubes were centrifuged at 200 × g for 5 min at 4°C, and pelleted amebae were

suspended in M199S with serum to 2.5×10^5 amebae/ml. Each tube containing transfection complexes was filled with the suspended trophozoites (1.8 to 1.9 ml), the contents were gently mixed by inverting the tubes, and the tubes were incubated horizontally for 3 hours at 37°C. Tube contents were added to warm TYI in 25 cm² tissue culture flasks, and incubated overnight at 37°C. 15 µg/ml hygromycin (Invitrogen, Carlsbad, CA, USA) was added for selection after the overnight incubation [177]. After 3-5 days, 25 ml of the TYI was removed to a new 25 cm² tissue culture flask, and 25 ml fresh TYI with hygromycin was added to each of the flasks. Transfectants were usually apparent 1–2 weeks after transfection.

***E. histolytica* shRNA Constructs**

All short hairpin RNAs (shRNAs) used in this study were driven by the RNA polymerase III U6 promoter (GenBank: U43841), which normally promotes the expression of the endogenous single-copy spliceosomal small nuclear U6 RNA (Figure 2-5) [178] and were cloned into the amebic expression vector pGIR310 (Figure 2-6) [72, 73], a modification of the expression vector pGIR308 [177, 179]. pGIR310 was modified by the addition of a short polylinker containing *Hind*III, *Sal*I, and *Not*I restriction sites (Figure 2-6) [72, 73]. Modified pGIR310 conferred resistance to hygromycin B in *E. histolytica* and to ampicillin in *Escherichia coli* (Figure 2-6). This approach utilizing the U6 promoter has been successfully used in mammalian cell lines [180]. All shRNA constructs used in these studies have the same structure: they form a short hairpin, consisting of a 29-base sense strand, followed by a 9 base loop, followed by the 29-base antisense strand (Figure 2-5, Figure 2-6). 29-base pair hairpins were chosen rather than the standard 21-23mers, since artificial siRNAs may be more effective in knocking down protein expression [167].

Sequence Selection for shRNA Constructs

The Ambion siRNA Finder web site [181] was used to select possible siRNA sequences of 21-mers beginning with two adenosine residues. To select sequences that would target Igl1 and Igl2 both separately and simultaneously, those portions of their coding sequences which were identical or divergent were input separately. For EhC2A the portion of the gene sequence selected for targeting was the poly-proline region (bases 301–567) since this region is least similar to the other gene family members. The entire coding sequence of PATMK and of URE3-BP was used to select siRNA sequences. From the pool of selected 21-mer sequences, those with runs of more than 4 As or Ts were eliminated, and those with GC content between 30% and 50% were lengthened to 29 bp by adding the next eight bases in the genomic sequence. To check that a given 29-bp sequence would only match its own gene sequence (open reading frame), the *E. histolytica* Genome Project database at the J. Craig Venter Institute (JCVI, formerly TIGR) web site [182] was used to check each 29-bp sequence by BLAST, using the sequence search function, with the CDS (all contigs) database selected with the expect value set to 100. At least four sequences were selected per gene, with non-unique ones eliminated. To create a scrambled control sequence, one of the selected sequences per gene was chosen, and the bases were scrambled up (each still began with the AA dinucleotide like the others); these sequences were checked against the database, this time to make certain they matched nothing in the *E. histolytica* genome. In addition, a sequence targeted to the green fluorescent protein (GFP) was included as a control [180]. The chosen sequences, those ultimately transfected into *E. histolytica* HM1:IMSS trophozoites, are shown in Table 1-1. Sequences for constructs that did not successfully transfect are not shown.

shRNA Primer Design

Primers were designed based on the method of Gou *et al* (2003) [180] to yield PCR-generated shRNA constructs containing in the final product the *E. histolytica* U6 promoter [178] followed by the sense strand of the hairpin, the 9 bp loop (TTCAAGAGA) [162], the antisense strand of the hairpin, and the U6 terminator sequence [180]. The two-step PCR process is diagrammed in Figure 2-5. An *ApaI* restriction site (GGGCCC) was included between the 3' end of the U6 promoter and the beginning of the shRNA sequence [180]. To facilitate cloning of the PCR product into the expression vector, a *HindIII* site was added to the 5' end of the U6 promoter sequence, and a *NotI* site was added following the terminator sequence. The selected siRNA sequences, shown in Table 2-1, were used to design oligos to create shRNAs via PCR. Two rounds of PCR were employed to generate the final shRNA constructs, using one forward primer and two reverse primers, whose sequences are listed in Table 2-2. In the first round of PCR, the *E. histolytica* U6 promoter followed by the sense strand and the loop were generated using a forward primer amplifying the 5' end of the U6 promoter and a first reverse primer containing the sequence of the sense strand of the shRNA and the loop (Figure 2-5, Table 2-2). A second round of PCR created the completed shRNA construct: the product from the first round was used as a template, using the same forward primer as in the first round, and a second reverse primer containing the sequence of the loop, the antisense strand sequence, and the U6 termination sequence (Figure 2-5, Table 2-2). The control GFP sequence [180] was used to design oligos for making a shRNA control construct. Sense strand sequences chosen to make the PATMK, Igl, URE3-BP and EhC2A shRNA constructs successfully transfected into trophozoites are shown in Table 2-1, and PCR oligos used to amplify these sequences to generate shRNAs via PCR are shown in Table 2-2.

PCR Conditions for Generating shRNAs

Initially, *E. histolytica* genomic DNA was used as a template for the first round of PATMK and Igl shRNA PCRs. For the URE3-BP and EhC2A shRNA PCRs, the cloned U6 promoter in the PATMK or Igl constructs was used as the PCR template for the first round of PCR: either PATMK or Igl shRNA plasmids were digested with *HindIII* and *ApaI* and the U6 promoter was gel-purified using the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA, USA) and used as the template for the first round of PCR. The first PCR round generated the sense strand of the hairpin and the loop. Reaction volumes of 40 μ l were set up, each consisting of 0.6 μ l SAHARA™ DNA polymerase (Bioline USA Inc., Tauton, MA, USA), 4 μ l 10X SAHARA™ PCR buffer, 3.2 μ l 50 mM MgCl₂, 2 μ l dNTP mix (stock 10 mM each), 0.4 μ l U6 *HindIII* forward oligo (100 μ M stock), 0.4 μ l R1 oligo (100 μ M stock), 1 μ l template DNA (200 ng of *E. histolytica* genomic DNA or 25 ng gel-purified digest of *HindIII/ApaI* U6 promoter), and 28.4 μ l sterile water. Cycling conditions were as follows: 95°C for 8 minutes, 10 cycles of 95°C 45 sec, 40°C 1 min, 68°C 1 min 30 sec; 25 cycles of 95°C 45 seconds, 52°C 1 min, 68°C 1 min 30 sec, and a 5 min final extension at 68°C. 5 μ l of each PCR product was subjected to agarose gel electrophoresis to check that the products were ~380 bp. In the second PCR round, the first round PCR product was used as a template to add the antisense strand of the hairpin, the terminator sequence and the *NotI* site. Each 100 μ l-volume reaction contained 2 μ l SAHARA™ DNA Polymerase (Bioline USA Inc., Tauton, MA, USA), 10 μ l 10X SAHARA™ PCR buffer, 8 μ l 50 mM MgCl₂, 5 μ l dNTP mix (10 mM each), 0.8 μ l U6 *HindIII* forward oligo (100 μ M), 0.8 μ l R2 oligo (100 μ M), 2 μ l PCR product from the first PCR round, and 71.4 μ l sterile water. Cycling conditions were: 95°C for 8 minutes, 10 cycles of 95°C 45 sec, 18.5°C 1 min 30 sec, 68°C 1 min 30 sec; 30 cycles of 95°C 45 seconds, 55°C 1 min, 68°C 1 min 30 sec, and a 5 min final extension at 68°C. The low annealing temperature in the early cycles of the second PCR was used since the

loop is the only overlap between the first round product and the second round reverse oligo. The second round PCR products were checked by agarose gel electrophoresis for products of the correct size (~420 bp). Sometimes a smaller product was present in addition to the correct size product in the final PCR product; this was ignored since it had no effect on the subsequent cloning steps.

The final PCR products were ethanol-precipitated, then they and modified pGIR310 were digested with *HindIII* and *NotI*. The digested expression vector was gel-purified using the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA, USA). The PCR products were ethanol-precipitated at -20°C overnight, resuspended, then ligated into modified pGIR 310 using High Concentration T4 DNA Ligase (New England Biolabs, Ipswich, MA, USA), incubated overnight at 16°C, ligated into modified pGIR310, transformed into *E. coli*, and colonies screened. Plasmids with inserts were sequenced using the Sp6 promoter present in the vector, and those with perfect U6 promoter and hairpin sequences were cultured, plasmids were isolated using the Qiagen HiSpeed Maxiprep Kit (Qiagen, Valencia, CA, USA), and transformed into HM1:IMSS strain trophozoites as described above.

Western Immunoblotting

The PATMK, Igl, URE3-BP, or EhC2A shRNA transfectants were cultivated in 25 cm² tissue culture flasks and selected beginning with 15 µg/ml of hygromycin, with the hygromycin level increased every 24 hours until the final level of selection was reached, and this level was maintained for 48 hours before harvesting. PATMK transfectants [73] were selected with 30 µg/ml of hygromycin. The GFP control transfectants, all three Igl transfectant lines, and the URE3-BP (350–378) transfectants were selected with 100 µg/ml, the URE3-BP (580–608) shRNA transfectants with 75 µg/ml, and the EhC2A samples with 90 µg/ml hygromycin. The

final concentration of hygromycin selection differs since the selection was increased until the desired level of knockdown was achieved. There were three biological replicates per shRNA transfectant, and one for the HM1:IMSS nontransfected trophozoites. Trophozoites were harvested as described above for transfection, counted, resuspended in ice cold Lysis Buffer (150 mM NaCl, 50 mM Tris, 5X Sigma protease inhibitor cocktail (P2714)) (Sigma-Aldrich, St. Louis, MO, USA), and 25 µg/ml E-64 (Sigma-Aldrich, St. Louis, MO, USA) at an initial concentration of 2×10^6 – 5×10^6 amebae/ml, and lysed by sonication by pulsing twice for 10 seconds each with a 10 second rest on ice between pulses, using a Branson 150D sonifier set to 5 (Branson Ultrasonics Corporation, Danbury, CT, USA). For PATMK gels [73], lysate from 5×10^4 or 1×10^5 amebae was loaded per lane. For samples other than PATMK, the amount of protein in the lysate was measured using a bicinchoninic acid (BCA) protein assay kit (Pierce/Thermo Scientific, Rockford, IL, USA) or with the Bio-Rad protein assay kit (Bradford) (Bio-Rad, Hercules, CA, USA). Sample lysates were diluted to the same protein concentration, with up to 60 µg total protein loaded per lane, and samples were subjected to SDS-PAGE on 8% or 10% (PATMK), 12% (Igl) or 15% (URE-BP and EhC2A) Tris-glycine gels in running buffer (25 mM Tris, 192 mM glycine, 1% SDS, pH 8.3). Gels were run under reducing conditions for all samples except for the Igl samples, which were briefly heated, not boiled. After SDS-PAGE, samples were transferred to Immobilon-P PVDF transfer membrane with pore size 0.45 µm (Millipore, Billerica, MA, USA) with either the Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad, Hercules, CA, USA) at 20 volts for 1 hour 30 minutes (URE3-BP and EhC2A) or the Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad, Hercules, CA, USA) for 100 volts for 1 hour 45 minutes (PATMK and Igl). The same transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol) was used in each case. All sample lysates and lysate dilutions were done in triplicate (technical replicates) or at least three separate times.

After protein transfer, sample PVDF membranes were cut in half so each half could be probed separately, were blocked in 5% milk in TBST (43 mM Tris-HCl, 7 mM Tris base, 200 mM NaCl, 0.1% Tween 20, pH 7.35), and incubated with either antibodies or sera against PATMK, Igl1, URE3-BP, EhC2A, or control antibodies against the light subunit (Lgl) of the Gal/GalNAc lectin, or against actin. All antibodies or sera against amebic proteins were non-commercial, raised by immunizing rabbits or mice with recombinant amebic protein or synthesized peptide. Douglas Boettner was responsible for the production of the PATMK antibody. PATMK antiserum was raised by immunizing New Zealand White rabbits (Covance, Princeton, NJ, USA) with the synthesized peptide EIQQQNPISTSLKISKISSD (residues 130–150 of PATMK) conjugated to KLH [73]. Heriberto (Eddie) Moreno was responsible for the production of the EhC2A antibody. EhC2A polyclonal antiserum was raised by immunizing New Zealand White rabbits with purified recombinant GST-tagged EhC2A fusion protein (Cocalico Biologicals, Inc., Reamstown, PA, USA) [74]. Carol Gilchrist was responsible for the production of the URE3-BP antibody. For preparation of URE3-BP antibody, mice were immunized with recombinant URE3-BP, hybridoma cell lines producing anti-URE3-BP monoclonal antibody were prepared, and monoclonal antibodies against URE3-BP were produced [76]. James McCoy was responsible for the production of the Lgl antibody; recombinant purified Lgl1-GST fusion protein was used in conjunction with complete Freund's adjuvant to immunize mice [54]. David Beck was responsible for the production of the Igl antibody. For preparation of Igl antibody, mice were immunized with recombinant Igl1, hybridoma cell lines producing anti-Igl1 monoclonal antibody were prepared, monoclonal antibodies against Igl were produced, and these were tested by Lauren Lockhart for reactivity against recombinant Igl1. Commercially available anti-actin antibodies (Sigma-Aldrich, St. Louis, MO, USA or Santa Cruz Biotechnology, Santa Cruz, CA, USA) were used to detect amebic actin as a loading control.

The ECL kit from Roche (Roche Applied Science, Indianapolis, IN, USA) was used to treat membranes after secondary antibody incubation, bands were visualized on film, film images were electronically scanned, and Scion Image Beta 4.0.3 software (Scion Corporation, Frederick, MD, USA) was used to quantify band intensity. The background value was subtracted from each band value, the ratio of Igl, EhC2A, or URE3-BP band value to the control actin (or Lgl) band value was taken for normalization, and then the sample shRNA transfectant lines were compared to the control GFP shRNA or scrambled shRNA transfectant line. HM1:IMSS nontransfected samples were also included. Values for each shRNA transfectant were averaged, and the SD and/or SE for each average was calculated using the total number of biological replicates multiplied by the number of technical replicates. Statistical analysis was performed using Student's t test (two-tailed) or ANOVA. The GraphPad QuickCalcs P-value calculator was used to calculate the P-values [183].

Isolation of Total RNA and cDNA Generation

Igl, URE3-BP, and control GFP transfectant shRNA lines were selected with hygromycin as described above for western blotting, and samples were collected and frozen in TRIzol reagent (Invitrogen, Carlsbad, CA, USA) at -80°C for RNA isolation at the same time as those harvested for crude lysate for protein analysis. Total RNA was isolated from each shRNA transfectant and nontransfected HM1:IMSS sample using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and the RNA was quantified before DNase I treatment by using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific/Thermo Fisher Scientific, Waltham, MA, USA). 100 µg total RNA was treated with RNase-free recombinant DNase I (Roche, Indianapolis, IN, USA) for 30 minutes at 37°C, was treated with RNase-free recombinant DNase I (Roche, Indianapolis, IN, USA) for 30 minutes at 37°C, and purified on RNeasy columns using the RNeasy Mini kit as per the

manufacturer's instructions (Qiagen, Valencia, CA, USA). The RNA was re-quantified using the NanoDrop, and 5 µg RNA per sample was reverse-transcribed using SuperScriptII reverse transcriptase (Invitrogen, Carlsbad, CA, USA) with anchored oligo dT as the primer, including sample controls with no reverse transcriptase added (no-RT controls). Each cDNA sample was quantified using the NanoDrop and sample cDNAs were diluted to 100 ng/µl. To check samples for residual DNA contamination in the no-RT controls, each was screened with primers specific for the Jacob cyst-specific gene [67]. If residual DNA contamination was observed, the RNA was treated again with DNase I as above, purified on RNeasy columns and re-screened for residual DNA contamination.

Quantitative Reverse-Transcription Real-Time PCR (qRT-PCR)

After the screen for residual DNA contamination was completed, the cDNA concentrations were quantified, and sample cDNAs were diluted to 100 ng/µl. HM1:IMSS cDNA was also serially-diluted for generating a standard curve. Primers used for qRT-PCR in this study were selected to amplify <300 bp sections of mRNA and designed such that they would amplify sections of the gene not targeted by shRNAs. Amplification of actin and/or Jacob [67] was performed for use as a normalization control. The oligos amplifying the 5' end of URE3-BP were designed by Carol Gilchrist; the other URE3-BP oligos, the PATMK oligos and the Igl oligos were designed by Alicia Linford [72]. Oligo sequences used in qRT-PCR are shown in Table 2-3. Each oligo pair was checked using the *E. histolytica* genomic database [TIGR database *Entamoeba histolytica* Genome Project web site [182] to validate that only the gene intended would be amplified, except for actin and Jacob, which were designed to detect all family members [67]. An MJ Research Opticon2 DNA Engine (Bio-Rad, Hercules, CA, USA) was utilized for all qRT-PCR runs. 200 ng (2 µl) of each sample or control cDNA, or serially-diluted

HM1:IMSS cDNA for standard curves, was added to each sample well in a 96-well plate for each set of amplifications. cDNA from each biological replicate was run in quadruplicate (technical replicates), and there were three biological replicates per transfectant line, except for HM1:IMSS nontransfected samples, which had one biological replicate. No-RT controls were also included for each set of samples. Each well contained in addition to the cDNA: 1.25 U HotStarTaq (Qiagen, Valencia, CA, USA), 1X HotStarTaq PCR Buffer, 0.5 mM MgCl₂, 200 μM each dNTP, 1 μM each oligo, 1:10,000 dilution of SybrGreen (Invitrogen, Carlsbad, CA, USA), 0.15% Triton X-100, and water to a volume to 25 μl per well. qRT-PCR cycling conditions were 95°C for 15 minutes, followed by 40 cycles of 95°C 30 s; 54°C 30 s; 72°C 45 s, followed by one cycle of 72°C for 3 min. At the end of amplification, a melt curve was performed from 70°C to 95°C, increasing 0.2°C every cycle with a 5-second hold. The C(t) values were averaged for each oligo pair for each set of technical replicates, and sample values were normalized to the housekeeping gene actin. The GFP shRNA transfectant line was used as a baseline control for comparison to the URE3-BP and Igl shRNA transfectant lines; for PATMK samples, the scrambled PATMK shRNA transfectant line was used as a control. HM1:IMSS samples were included as a secondary control. The differences in gene expression for the URE3-BP and Igl transfectant lines as compared to the GFP transfectant line were calculated by using both the relative standard curve [184] and the comparative C(t) method ($\Delta\Delta C(t)$ method) [185, 186, 187]. Statistical analysis was performed using Student's t test (two-tailed), groups were also compared using ANOVA, and the GraphPad QuickCalcs P-value calculator [183] was used to calculate P-values.

qRT-PCR Study for Transfectants Expressing shRNAs Against PATMK To Assay if Expression of Related Genes Are Being Affected (Off-Target Effects)

Three biological replicates each of the *E. histolytica* transfectant lines stably transfected with the shRNA construct that yielded the best protein knockdown, PATMK (3532-3560), or with the control PATMK (3532-3560 scrambled), were selected with 30 µg of hygromycin B per ml of TYI medium before harvesting. Total RNA was isolated as described in the Isolation of Total RNA and cDNA Generation section under Materials and Methods. Nine genes with sequence identity ~80% to PATMK (3532-3560) were chosen for this RT-PCR study and were originally identified by BLAST using the TIGR *Entamoeba histolytica* database web site [182]. Six of the nine were annotated as protein kinases or putative protein kinases, and oligo pairs were designed to amplify short sections of each of these genes (Table 2-4). Oligo pairs were checked using the TIGR *E. histolytica* genomic database at the web site [182] to validate that only the selected gene transcript would be amplified, except for actin and Jacob, which detect all family members (Table 2-3, Table 2-4) [67]. qRT-PCR was performed as described in the Quantitative Reverse-Transcription Real-Time PCR (qRT-PCR) section above. qRT-PCR cycling conditions were 95°C for 15 minutes, followed by 45 cycles of 95°C 30 s; 47°C 30 s; 72°C 30 s, followed by one cycle of 72°C for 3 min followed by a melt curve from 45°C to 90°C increasing 0.2°C every cycle with a 2-second hold. For real-time PCR quantitation, all samples were run in triplicate, data from multiple runs were combined, the C(t) values were averaged and calculations were performed as described in the Quantitative Reverse-Transcription Real-Time PCR (qRT-PCR) section.

Isolation of Small RNAs

Three of the Igl shRNA transfectant lines, Igl (1198–1226), Igl (2412–2440), and Igl (2777–2805) [72] as well as the two PATMK knockdown shRNA lines, PATMK (2273–2301) and PATMK (3532–3560), and the PATMK scrambled control [73] were grown in 25 cm² tissue culture flasks. Igl shRNA transfectants were selected with 100 µg/ml hygromycin and PATMK transfectants with 30 µg/ml hygromycin. A minimum of 50 µg small RNAs were isolated from each sample as well as from control nontransfected HM1:IMSS trophozoites using the mirVana™ miRNA Isolation Kit (Applied Biosystems/Ambion, Austin, TX, USA) as per the manufacturer's instructions.

Northern Blotting of Small RNAs

Oligo probes were designed to match the sense or antisense strands of each hairpin. Fifty µg of small RNAs were loaded per lane on a 12% denaturing acrylamide gel, subjected to electrophoresis, and transferred to Hybond™-N+ nylon membrane (Amersham Biosciences/GE Healthcare Biosciences Corp, Piscataway, NJ, USA) as per the manufacturer's instructions. rRNA bands were analyzed to insure equal RNA loading. Oligo probes matching to the sense or antisense strands of the hairpins were end-labelled with ³²P and were hybridized with each corresponding sample blot strip overnight at 37°C overnight, washed with low and medium stringency conditions, and exposed overnight to film.

Results

Construction of shRNA constructs

The RNA polymerase III promoter of the *E. histolytica* U6 gene (GenBank: U43841) [178] was amplified beginning at -333 from the transcription start site of the U6 small nuclear RNA gene, and the shRNA-encoding DNA was added by PCR at the transcription start site (Figure 2-5) [73, 180]. The resulting U6 promoter-shRNA constructs were cloned into pGIR310 modified to contain a short polylinker (Figure 2-6). The shRNAs were designed to have a 29-nucleotide complementary stem with a 9-nucleotide loop (Figure 2-6). The sense strand sequences of the shRNA constructs transfected into HM1:IMSS trophozoites, the oligonucleotide (oligo) sequences used to create them by PCR, and the oligo sequences used in quantitative reverse-transcription real-time PCR (qRT-PCR) amplification to assess mRNA knockdown are shown in Tables 2-1, 2-2, and 2-3.

shRNA Transfectants

Transfectants were maintained at 15 µg/ml hygromycin. For knockdown studies, the hygromycin concentration was increased every 24 hours until the final level of selection was achieved, and was maintained for 48 hours, in order to increase the copy number of the episomal shRNA vector [62, 188, 189]. The level of hygromycin selection was increased until the desired knockdown was attained, up to 100 µg/ml. Transfected trophozoites selected with 100 µg/ml hygromycin were able to grow and divide for at least two weeks under continuous selection at this level. A shRNA directed against green fluorescent protein (GFP) [180], with a sequence matching nothing in the *E. histolytica* genome, was utilized as a control for transfection and hygromycin selection for the Igl and URE3-BP transfectants. GFP shRNA transfectants were

selected with the same level of hygromycin as other shRNA transfectants. For PATMK and EhC2A, a scrambled control for each which matched nothing in the *E. histolytica* genome was created, containing the same nucleotides as either the PATMK (3532-3560) or the EhC2A (363–391) shRNA, but in a different order. Note: PATMK (3532-3560) is referred to as PATMK (3552-3581) in Boettner *et al* (2008) [73], which is due to an earlier typographical error; “3532” was accidentally written as “3552”. Sequences of the shRNA sense strands are shown in Table 2-1. Non-transfected HM1-IMSS amebae were also included, with the results for western blotting and qRT-PCR being statistically the same as the GFP controls. Three biological replicates were tested per shRNA transfectant. All sample trophozoites were cultured in 25 cm² tissue culture flasks, and were harvested for crude lysate and for RNA isolation on the same day from the same flask. For protein and mRNA comparison, actin was used as the "housekeeping" control gene, as a loading and normalization control, except for the PATMK shRNA immunoblots, where the light subunit of the Gal/GalNAc lectin was used as the loading and normalization control [73].

Knockdown of PATMK protein

Four shRNA constructs targeted PATMK, targeting different regions of the mRNA: PATMK (325-353), PATMK (1326-1354) PATMK (2273-2301), and PATMK (3532-3560). A scrambled control, PATMK (3532-3560) scrambled, contained the same bases as PATMK (3532-3560), but in a different order, and matched nothing in the *E. histolytica* genome. Transfected trophozoites were selected with 30 µg/ml hygromycin for 48 hours before harvesting. Blots were probed with anti-PATMK antibody, and with anti-Lgl antibody (against the light subunit of the Gal/GalNAc lectin) for use as a loading and normalization control. The level of PATMK in the PATMK (3532-3560) scrambled shRNA transfectants was defined to be 100% (Figure 2-7, Table 2-6). All PATMK shRNA constructs were able to knock down PATMK protein expression to

some extent. Interestingly, the closer to the 3' end the shRNA target sequence was, the more effective the shRNA was in knocking down PATMK protein expression (Figure 2-7, Table 2-6). PATMK (325-353) had $53.6 \pm 4.5\%$ and PATMK (1326-1354) had $45.2 \pm 5.2\%$ of the PATMK (3532-3560) scrambled control PATMK level. PATMK (2273-2301) had $7.4 \pm 3.5\%$ and PATMK (3532-3560) had $0.6 \pm 1.1\%$ of the PATMK (3532-3560) scrambled control PATMK level (Figure 2-7, Table 2-6). Note: PATMK (3532-3560) is referred to as PATMK (3552-3581) in Boettner *et al* (2008) [73], which is due to an earlier typographical error; “3532” was accidentally written as “3552”.

Knockdown of PATMK mRNA and qRT-PCR study to evaluate if expression of PATMK-related genes is affected by transfectants expressing shRNA against PATMK (off-target effects).

Genes with sequence identity to the sequences for the most effective shRNA construct against PATMK, PATMK (3532-3560), were originally identified by BLAST using the TIGR *Entamoeba histolytica* database web site [182]. Nine genes with sequence identity ~80% to PATMK (3532-3560) were chosen for this RT-PCR study [73]. Six of these nine are known or putative protein kinases. RNA was isolated from PATMK (3532-3560) and control PATMK (3532-3560) scrambled shRNA samples as described above in the Isolation of Total RNA and cDNA Generation section under Materials and Methods, and primers for qRT-PCR amplifying these nine genes were designed (Table 2-4) and checked using the TIGR *Entamoeba histolytica* database web site [182] to make sure they amplified only their target gene. qRT-PCR was performed as in the Quantitative Reverse-Transcription Real-Time PCR (qRT-PCR) section above. There was no statistical difference between the PATMK (3532-3560) and control PATMK (3532-3560) scrambled shRNA transfectants in expression levels of any of these nine genes, so

there were no measureable off-target effects (Table 2-5) [73]. PATMK mRNA levels were reduced in PATMK (3532-3560) to undetectable levels (Table 2-5) [73].

Knockdown of Igl protein

Four Igl shRNA constructs targeted Igl. One construct, Igl1 (272–300), specifically targeted Igl1. Three constructs, Igl (1198–1226), Igl (2412–2440), and Igl (2777–2805), were targeted to sequences conserved in both Igl1 and 2 (Table 2-1). The GFP shRNA transfectants were used as controls. Transfected trophozoites were selected with 100 $\mu\text{g/ml}$ hygromycin for 48 hours before harvesting. Blots were probed with anti-Igl1 antibody, and with anti-actin antibody as a loading and normalization control. The level of Igl1 in the GFP shRNA transfectants was defined to be 100% (Figure 2-8, Table 2-11). The Igl1-specific (272–300) shRNA transfectant had a decreased amount of Igl1 protein, $27.8 \pm 3.9\%$, as compared to the GFP shRNA control (Figure 2-8, Table 2-7). Igl (1198–1226) had $42.3 \pm 6.2\%$ and Igl (2777–2805) had $38.1 \pm 9.4\%$ of the GFP control Igl1 level. The Igl (2412–2440) shRNA construct had no effect on Igl1 levels ($95.3 \pm 9.7\%$ of the level in the GFP shRNA transfectants) (Figure 2-8, Table 2-7). HM1:IMSS nontransfected amebae were not statistically different from the GFP shRNA control (Table 2-7). The Igl (1198–1226) and Igl (2777–2805) transfectants, when selected with 30 $\mu\text{g/ml}$ hygromycin rather than 100 $\mu\text{g/ml}$, yielded less knockdown, having $\sim 70\%$ and $\sim 65\%$ of the control level of Igl1 (data not shown).

Knockdown of Igl mRNA

Short sections of Igl were amplified via qRT-PCR using template cDNAs synthesized from the Igl and control GFP shRNA transfectant mRNAs. Four oligo pairs were used to amplify Igl. Two sets of oligos targeted both Igl1 and Igl2 simultaneously, with one pair amplifying a 5'

section and the other a 3' section conserved in both Igl1 and Igl2. The two others were specific for Igl1 or Igl2, targeting the least-conserved region between the proteins. The oligo sequences and regions of Igl transcript amplification are shown in Table 2-3, and summarized qRT-PCR data for Igl is shown in Table 2-8. All Igl shRNA samples were compared to the GFP control shRNA transfectants. Three of the four Igl shRNA transfectants showed knockdown of Igl transcripts for all sets of oligo pairs, ranging between ~60 and ~80% of the Igl level in the GFP shRNA control (Table 2-8). The Igl (2412–2440) shRNA transfectants did not show any knockdown, and the HM1:IMSS nontransfected trophozoites were not statistically different from the GFP shRNA control (Table 2-8).

Knockdown of URE3-BP protein

Two shRNA constructs were used to target URE3-BP: URE3-BP (350–378) and URE3-BP (580–608). Transfected trophozoites were selected with 100 µg/ml hygromycin (GFP control or URE3-BP (350–378) shRNA) or 75 µg/ml hygromycin (URE3-BP (580–608) shRNA) for 48 hours before harvesting. Actin was used as a normalization and loading control. There was significant reduction of URE3-BP protein in both URE3-BP shRNA transfectants: for URE3-BP (350–378) it was $10.8 \pm 1.0\%$ and $13.8 \pm 2.6\%$ for URE3-BP (580–608) as compared to the GFP shRNA control (Figure 2-9, Table 2-9). HM1:IMSS samples were also included, but were not statistically different from the GFP shRNA control (Table 2-9).

Knockdown of URE3-BP mRNA

Three different oligo pairs, one amplifying the 5' end of URE3-BP, one the 3' end, and one a section in the middle, were used in qRT-PCR to amplify URE3-BP in cDNA from GFP shRNA control transfectants, URE3-BP (350–378) and URE3-BP (580–608) shRNA

transfectants, and HM1:IMSS nontransfected trophozoites. Oligo sequences are shown in Table 2-3. Actin was used as the normalization control. The URE3-BP (350–378) shRNA transfectant had an average of about 69% of the GFP control URE3-BP transcript level, and the URE3-BP (580–608) shRNA transfectant had about 13% of the GFP shRNA control URE3-BP level (Table 2-10).

Knockdown of EhC2A protein

Two shRNA constructs targeted EhC2A, EhC2A (363–391) and EhC2A (502–530). Transfectants were selected with 90 µg/ml hygromycin for 48 hours before harvesting. The scrambled control EhC2A (363–391 scrambled) shRNA transfectant was used as a control for EhC2A protein levels. HM1:IMSS nontransfected amoebae were not included. The level of EhC2A protein in the EhC2A (363–391 scrambled) control shRNA transfectant was defined as $100 \pm 5.0\%$ (\pm SE). The EhC2A (363–391) shRNA transfectant yielded a knockdown of EhC2A protein to a level of $3.0 \pm 0.4\%$ ($P < 0.0001$). The EhC2A (502–530) shRNA transfectant had no knockdown effect on EhC2A levels ($106.1 \pm 7.3\%$) and was statistically the same ($P = 0.3141$) as the EhC2A (363–391 scrambled) shRNA control transfectant (Figure 2-10). Student's t test was used for statistical analysis. qRT-PCR was not performed for these samples.

Northern blots of small RNAs

We tested if the shRNAs were actually being expressed by the U6 promoter. The PATMK samples were included because they had been shown to have significant knockdown of PATMK protein levels as compared to the scrambled PATMK shRNA control transfectant [73] and therefore would be good candidates for expressing the shRNAs. Northern blotting of the PATMK and Igl shRNA transfectant small RNAs was performed. Transfected trophozoites were

selected with 30 µg/ml hygromycin for 48 hours before harvesting, since we had seen protein knockdown previously at that level of selection [73]. Non-transfected HM1:IMSS amoebae were included as a negative control. Fifty µg of small RNAs from PATMK shRNA transfectants [73] and the Igl shRNA transfectants were probed with oligo probes targeting the respective sense and antisense strands of the shRNAs (Figure 2-11). The PATMK (3532–3560) [73] and Igl (2777–2805) shRNA samples had substantial expression of ~70 and ~30 nucleotide small RNAs, the expected sizes for the unprocessed hairpin and the processed siRNA respectively. There was not a correlation of the small RNA abundance with the degree of protein knockdown, as the small RNAs were more abundant in Igl (2777–2805) than in Igl (1198–1226), yet both had similar degrees of knockdown.

Discussion

We have utilized the U6 promoter to drive expression of shRNAs with a 29-bp stem and a 9-nt loop to knock down protein expression of four unrelated genes: the phagocytosis-associated transmembrane protein PATMK [73], the extracellularly-anchored Gal/GalNAc lectin intermediate subunit Igl; the calcium-regulated transcription factor, URE3-BP (upstream regulatory element 3-binding protein); and EhC2A, a membrane-binding protein that interacts with URE3-BP. The highest level of protein knockdown for Igl was 72%, for URE3-BP 89%, for EhC2A 97%, and more than 99% for PATMK. We concluded that this was an effective system for gene knockdown in *E. histolytica*. This method has advantages over other methods used for gene silencing: the U6-shRNA expression cassettes are small (~420 bp), appear to be active against different types of genes, yield significant knockdown, and the expression vector, once

transfected, allows continuous expression of shRNAs, thus avoiding performing multiple transfections, and the shRNAs can be easily synthesized via PCR.

Not every transfected shRNA construct was equally effective in silencing gene expression. The PATMK (325-353) shRNA construct knocked down PATMK protein by approximately half, while the PATMK (3532-3560) shRNA construct yielded almost complete knockdown of PATMK protein and no detectable mRNA as measured by qRT-PCR. Other shRNAs were ineffective at producing protein knockdown. For example, neither the EhC2A (502–530) nor the Igl (2412–2440) shRNA construct reduced gene expression. In the case of Igl (2412–2440), the run of four thymidines at positions 19–23 in the shRNA sense strand could possibly cause RNA polymerase III to terminate the transcript prematurely. In the case of the EhC2A (502–530) shRNA construct, the shRNA could bind to two locations in the mRNA, both the originally targeted region (502–530), but also about 100 nt upstream (406–434), and perhaps this had an inhibitory effect on the ability of this shRNA to knock down protein expression.

Factors other than the shRNA sequence affect the ability of a shRNA to down-regulate gene expression. The secondary structure of the transcript affects the ability of the RISC to bind to its target site [190, 191], and the relative abundance and stability of an mRNA may play a significant role in determining whether a given shRNA will effectively lead to the degradation of its target message. In addition, the stability and abundance of a protein product could also be a determinant in the detection of a knockdown phenotype. The protein with the least knockdown in these studies, Igl, was the most abundant; PATMK was the least abundant and had the most knockdown [192]. The level of hygromycin utilized to select for transfectants was also an important determinant of the extent of protein knockdown. Igl knockdown was twice as effective with 100 $\mu\text{g}/\text{ml}$ as with 30 $\mu\text{g}/\text{ml}$ of hygromycin selection (data not shown); however, there is a limit to how much selective antibiotic the shRNA transfectants can tolerate. The individual gene

whose product is being knocked down also plays a role in how much antibiotic can be used; Heriberto Moreno found that the EhC2A (363–391) shRNA transfectants yielded the best knockdown when selected with 90 $\mu\text{g/ml}$ of hygromycin, but increasing selection to 100 $\mu\text{g/ml}$ hygromycin killed the amoebae; Igl and PATMK shRNA transfectants could tolerate a concentration of at least 125 $\mu\text{g/ml}$ hygromycin selection.

The qRT-PCR data was not correlated directly with the level of protein knockdown. For the Igl transfectants, the mRNA knockdown level was not as high as the protein knockdown level, indicating the possibility that the protein could have a high turnover rate or be somewhat unstable. For URE3-BP, the URE3-BP (350–378) and (580–608) transfectants had similar levels of protein knockdown; however, the mRNA levels in the URE3-BP (350–378) transfectants were higher (67% of the control level), as compared with the URE3-BP (580–608) transfectants (13.5% of the control level). This difference is probably not due to partial mRNA decay, since the qRT-PCR data showed consistent URE3-BP levels among the three oligo pairs amplifying the 5', middle, and 3' sections of the transcript. One possible explanation could be that the secondary structure of the URE3-BP mRNA at the location of the URE3-BP (350–378) shRNA could be interfering with the RISC cleaving the mRNA but still allowing RISC binding, allowing for a degree of translational inhibition in addition to some mRNA destruction.

The *E. histolytica* U6 promoter appears to be functional and producing shRNAs: the northern blots of the small RNAs detected two sizes of small RNAs when probed with oligos that were complementary to the individual sense and antisense strands of the shRNAs. These may represent the unprocessed hairpin and the resulting siRNAs after Dicer processing. Surprisingly, the abundance of the small RNA was not proportional to the level of silencing. Northern blots may not be sensitive enough to identify low-level small RNA production, with low-level production adequate for protein knockdown. It is also possible that some of the expressed

hairpins could have been more quickly processed and incorporated into the RISC than others, and were therefore not detected.

In conclusion, we report the knockdown of four genes: PATMK, phagocytosis-associated transmembrane kinase 96; Igl, the intermediate subunit of the Gal/GalNAc lectin; the calcium-responsive transcription factor URE3-BP; and the membrane-binding protein EhC2A, by transfecting *E. histolytica* with expression vectors using the *E. histolytica* U6 promoter to drive expression of shRNAs targeting endogenous genes. These genes come from different families, with different functions, so this shRNA knockdown method appears robust and not specific to only one gene or gene family.

Future Directions

There is always room for improvement of any given system. The shRNA constructs could be expressed using a tetracycline-inducible expression vector, as there may be situations in which having constant expression of shRNAs is undesirable. One of the first reports of using a tetracycline-inducible knockdown system is by Ohkawa and Taira (2000) [194]. The addition of tetracycline operator sequences (DNA sequences where the tetracycline repressor binds) to the human U6 promoter was used to inducibly express antisense RNA to knock down expression of a reporter gene in HeLa cells; the modified U6 promoter was 80% as active as the wild-type unmodified promoter [194]. Tetracycline-inducible expression of shRNAs from this modified U6 promoter was able to knock down protein levels of DNA methyltransferase 1 (DNMT1) in the HCT116 human colorectal cancer cell line [195]. This system could be adapted to express shRNAs inducibly in amoebae: a tetracycline-regulatable vector has successfully been used to express antisense RNA in *E. histolytica* using an RNA polymerase II promoter [196, 197].

Since microRNAs do not have perfectly matched stems, a number of other researchers have tried synthesizing imperfectly matched hairpins such that the sense (passenger) strand contains mismatches to the antisense (guide) strand. Miyagishi *et al* 2004 [193] compared a number of shRNA stem lengths, a number of mutations introduced into the sense strand sequence, and different loops in shRNAs expressed from the U6 promoter in their ability to knock down luciferase expression in HeLa S3 cells, and found that introducing three or four C to T or A to G mutations in the sense strand using 21 and 25 nt stem lengths allowed for more effective knockdown, and an 11 nt loop with the sequence GUGUGCUGUCC (which is derived from microRNA miR-26b) [193] was more effective than shRNAs with the commonly used loop sequence, UUCAAGAGA [162]. It may be useful to experiment with *E. histolytica* shRNAs designed with imperfect matching of the sense strand and with different loops. Having an imperfectly matched sense (passenger) strand may help insure that the antisense (guide) strand would be incorporated into the RISC, and the shRNA would more closely resemble a miRNA.

Multiple cassettes of U6 or H1 promoters, each one driving expression of its own shRNA, can be cloned into the same expression vector, such that multiple genes are targeted for knockdown simultaneously [198]. This might be a useful approach for knocking down more than one gene, or targeting several sites with shRNAs on one or two genes. There are also shRNAs that are based on microRNA structures (artificial microRNAs) which have been used in human cultured cells, which can be expressed from RNA polymerase II promoters [199] and can be very effective, especially those that have been based on miR-155 or miR-30 [199, 200, 201]; however, regular shRNAs may be sometimes more effective [199]. Finally, constructs have been made using a combination of these features: two separate shRNAs with imperfectly matched stems targeting the same gene that were part of an artificial miR-30 backbone were expressed inducibly from a tetracycline-operator-containing U6 promoter for successful knockdown of expression

[202]. Currently, microRNA structures are not well characterized in *E. histolytica*, but seventeen putative miRNAs have been identified using a bioinformatic approach [144]. The use of microRNA-based shRNA expression vectors which can express multiple shRNAs inducibly may be a useful option in the future.

Figure 2-1. Cleavage of RNA into siRNAs by Dicer. Dicer cleaves long double-stranded or hairpin RNAs into short interfering RNAs (siRNAs). A typical siRNAs has 5' phosphorylated ends, a 19-nucleotide duplexed region and 2- or 3-nucleotide unpaired and unphosphorylated 3' ends. The 5'-PO₄ group is important for recognition of the siRNA by Dicer [114]. Figure adapted from Walton *et al* 2010 [114] and Dykxhoorn *et al* 2003 [98].

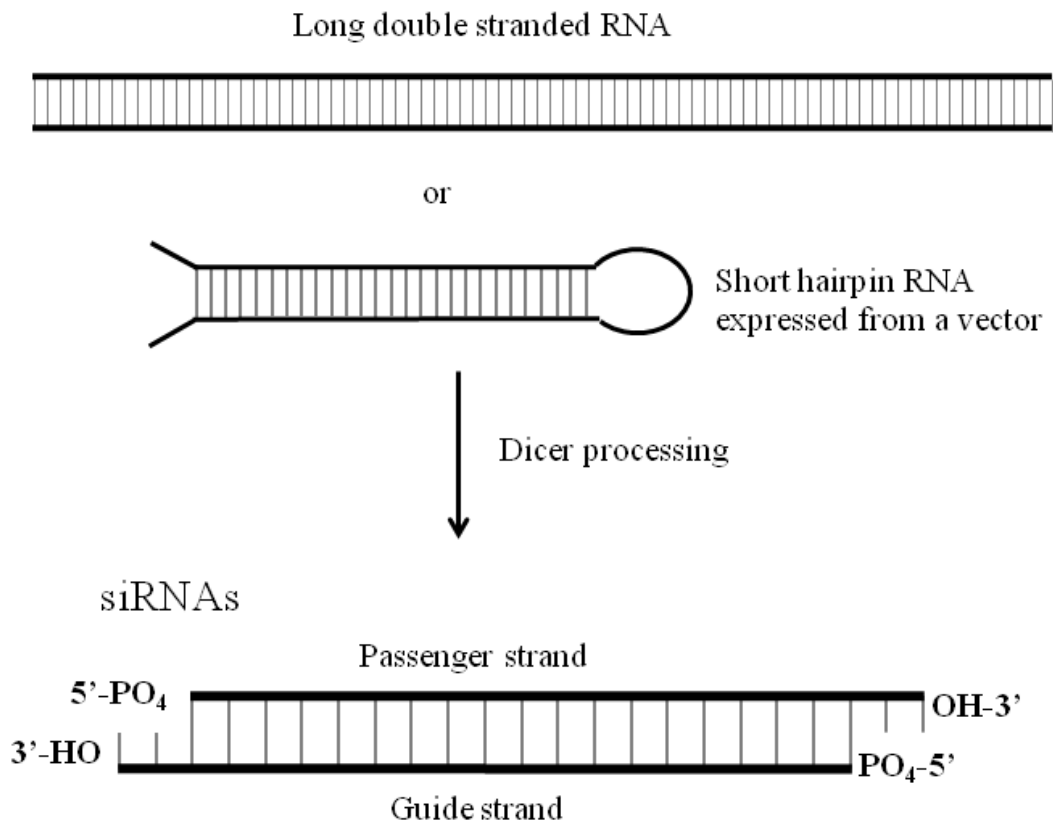


Figure 2-2. The RNA interference (RNAi) pathway. This pathway has been well characterized for *Drosophila* and should be similar in *Entamoeba*. Double-stranded RNAs or short hairpin RNAs (shRNAs) (Figure 2-1) are processed by Dicer/dsRNA binding protein heterodimer (R2D2 is the helicase in *Drosophila*) into small interfering RNAs (siRNAs). The duplexed siRNA is unwound in an ATP-dependent manner starting at the 5' terminus that has the lowest relative free energy of base pairing. This siRNA strand, the guide strand, is also then preferentially taken up by the RNA-induced silencing complex (RISC) which includes an RNA helicase (Armitage protein in *Drosophila*) and the Argonaute "Slicer" protein (Aubergine protein in *Drosophila*). In *Drosophila*, RISC also contains other proteins: the Vasa intronic gene (VIG), the Tudor staphylococcal nuclease (TSN), and the *Drosophila* homolog of the Fragile-X related protein (dFXR) [97]. Similar proteins (not yet identified) may be components of the *E. histolytica* RISC. The single-stranded siRNA guides the activated RISC to the homologous site on the cognate mRNA, cleaving the mRNA. Figure adapted from Dykxhoorn & Lieberman (2005) [97] and Dykxhoorn *et al* (2003) [98].

Double-stranded RNA

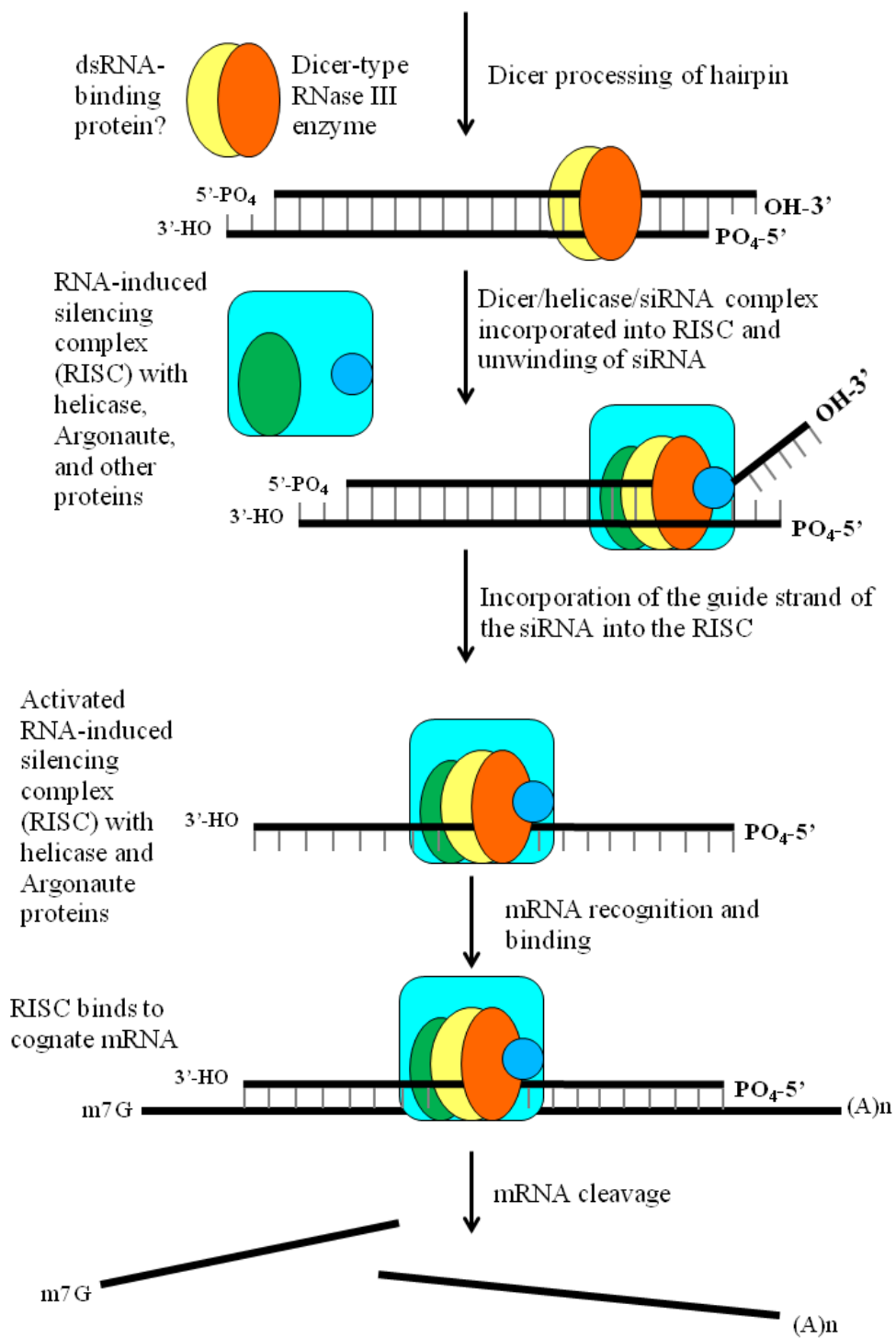


Figure 2-3. RNA-dependent RNA polymerase (RdRP) function. One model for amplification of the RNAi signal is that primary siRNAs, formed from a dsRNA trigger (the original introduced dsRNA) act as guides (presumably in an Argonaute-dependent manner) to recruit RdRPs to targeted transcripts (mRNA). This recruitment leads to de novo synthesis of short antisense RNAs that must be processed into siRNAs and incorporated into RISCs that are capable of finding additional silencing targets [132].

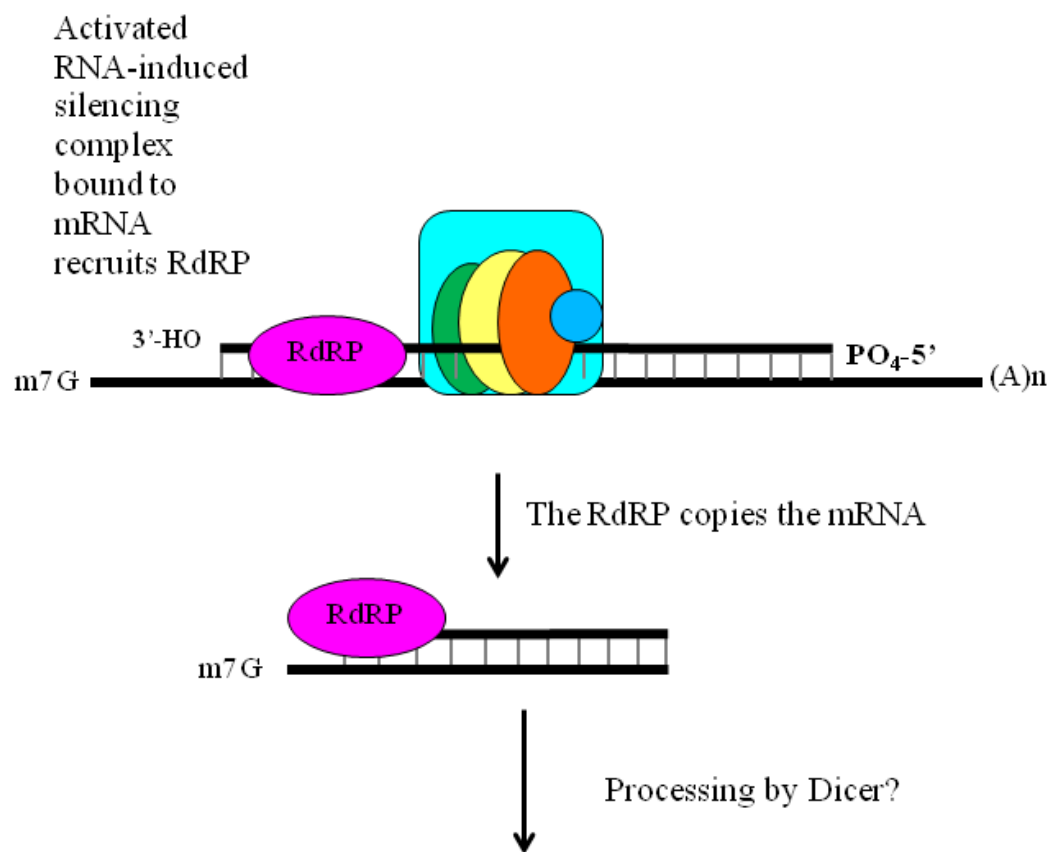


Figure 2-4. RdRPs can transcribe RNA with or without a primer. Plant RdRPs have been well characterized and the *E. histolytica* RdRP may function by a similar mechanism. It was found that when short oligos are used as primers in vitro, tomato RdRPs can synthesize complementary RNA from both RNA and DNA templates. The RdRP is also capable of unprimed synthesis of RNA: without a primer, the RdRP initiates synthesis at or near the 3' terminal nucleotide of a single-stranded template: the nucleotide where the RdRP begins transcription needs to have phosphate groups both 5' and 3' (if the very first 3' terminal nucleotide has a hydroxyl group then the transcript will begin at the penultimate nucleotide) [138]. The RISC may or may not be needed as a guide.

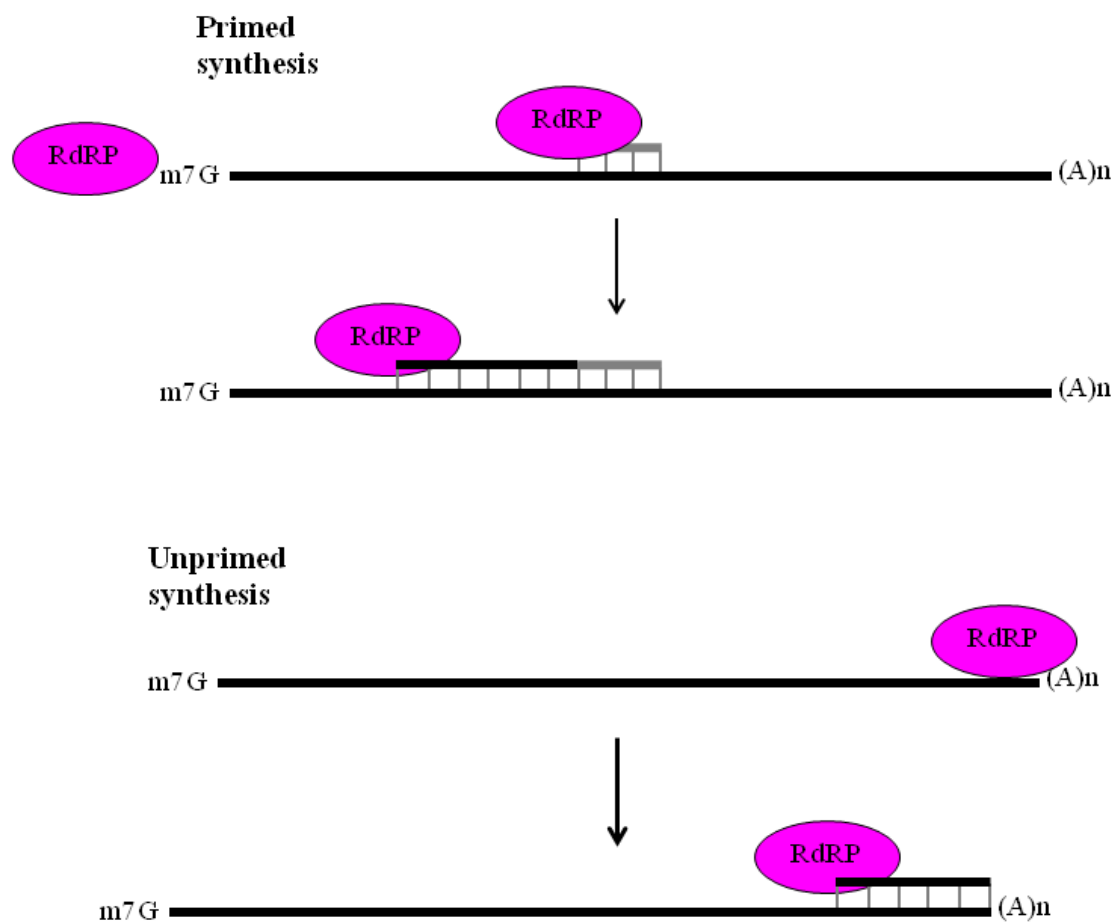
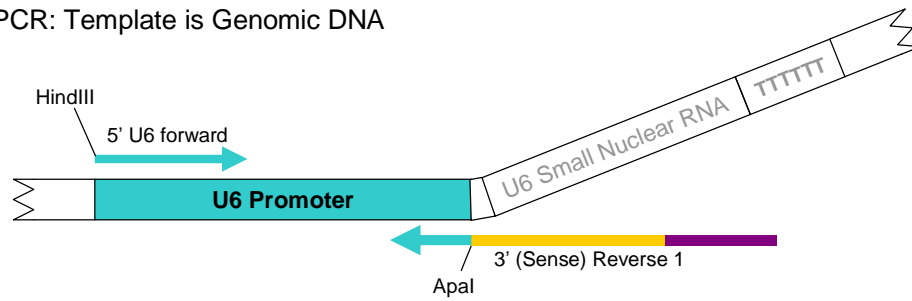


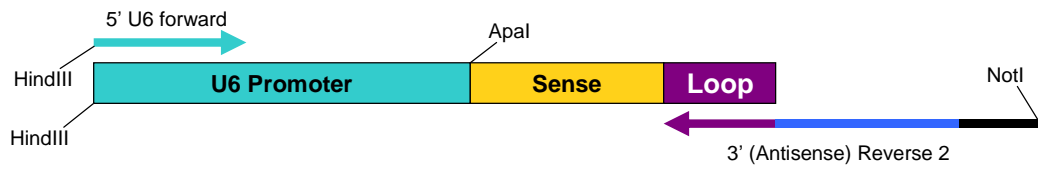
Figure 2-5. shRNA system for *Entamoeba histolytica*: Generation of short hairpins by PCR.

Diagram of the two-step PCR process for generating short hairpins shRNA constructs were made using the method of Gou *et al* (2003) [180]. Genomic DNA (or subsequently, the cloned U6 promoter) was used as a template to amplify the *E. histolytica* U6 promoter and to add the hairpins. The primers in the first PCR round were the forward primer, containing a *Hind*III site and 5' end of the U6 promoter, and a first reverse primer, containing the U6 promoter 3' end, the shRNA sense strand sequence, and the 9-nucleotide loop. To yield the final product, in the second PCR round, the same forward primer was used, with a second reverse primer containing the loop sequence, the antisense strand sequence, the termination sequence, and a *Not*I recognition site, using the first round product as a template. The primers used to generate the PCR products are listed in Table 2-2.

1st PCR: Template is Genomic DNA



2nd PCR: Template is 1st PCR product



Final PCR Product

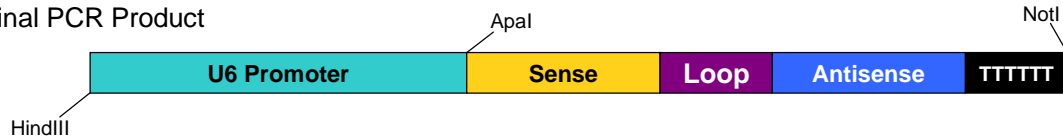
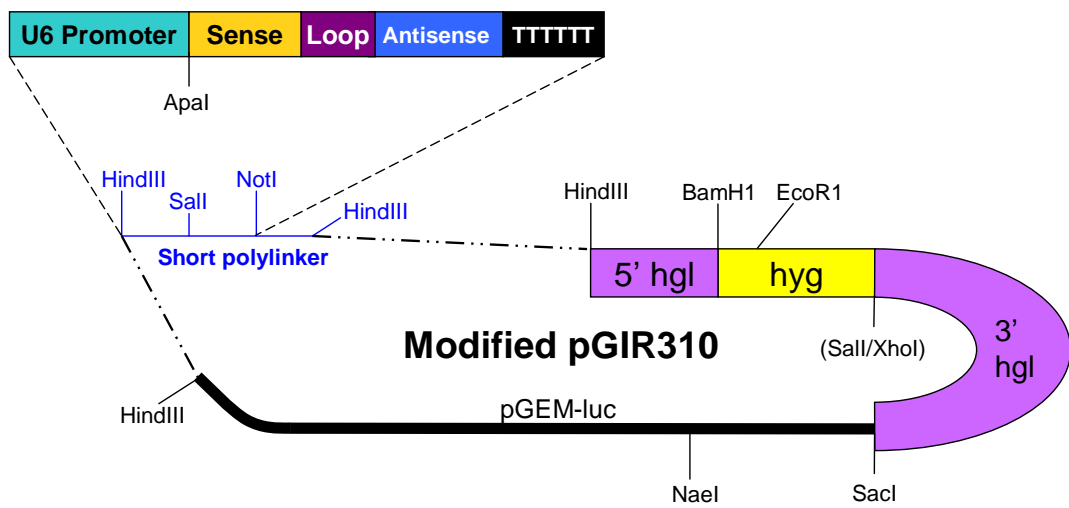
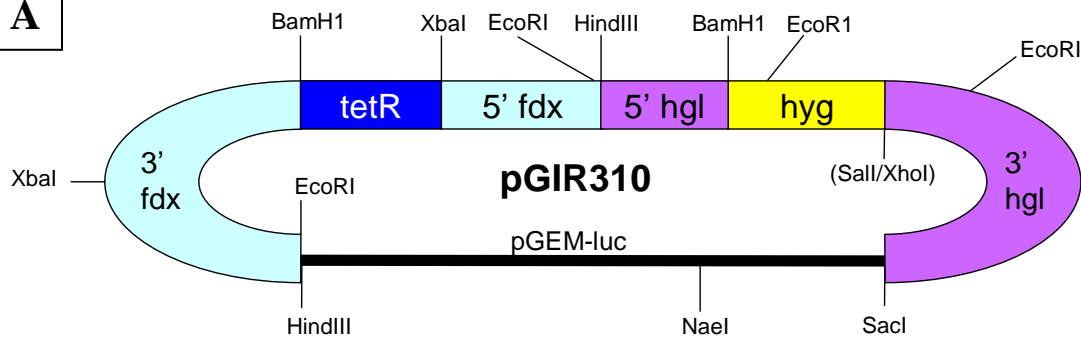


Figure 2-6. shRNA system for *Entamoeba histolytica*: Modification of the amebic expression vector pGIR310 to express shRNA, and the structure of the expected shRNA. (A). The tetracycline repressor cassette in expression vector pGIR310, a modification of pGIR308 [177, 179], was replaced with a polylinker containing a *SalI* and *NotI* site, flanked by *HindIII* sites. PCR products were cloned into the *HindIII* and *NotI* sites. pGIR310 confers hygromycin resistance in amebae and ampicillin resistance in *E. coli* bacteria. **(B).** Expected structure of 29-basepair shRNA before and a possible structure after processing by Dicer (Figure 2-1, Figure 2-2). The 29-basepair stem and 9-nucleotide loop are shown. The siRNAs generated by Dicer and the RdRP in *E. histolytica* appear to be larger, ~27 nt, rather than the usual mammalian 21-23 nt [149].

A



B

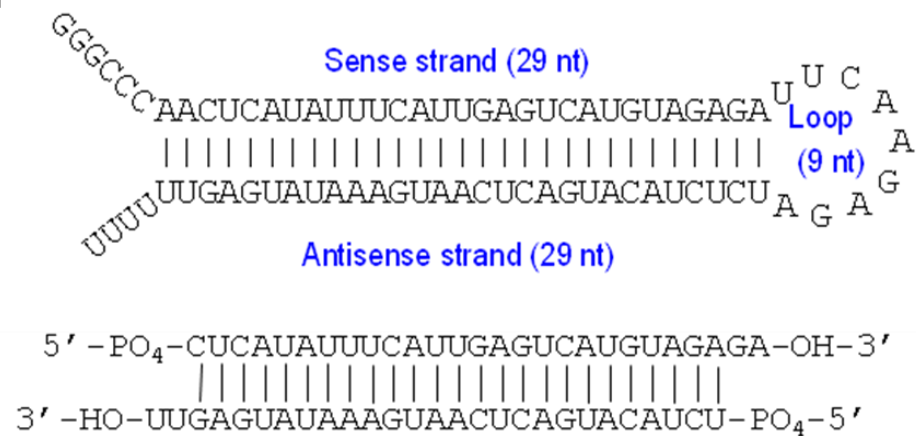
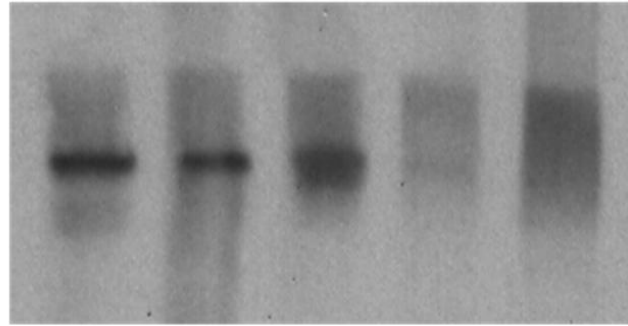


Figure 2-7. Western blot for PATMK shRNA transfectants. A representative western blot is shown with one biological replicate each for the PATMK (3532-3560) scrambled transfectant and four transfectant lines, with each shRNA targeting a different section of the PATMK gene: PATMK (325-353), PATMK (1326-1354), PATMK (2273-2301), and PATMK (3532-3560). Results shown are representative of three biological replicates per shRNA transfectant. Each membrane was probed with either anti-Lgl (light subunit of the Gal/GalNAc lectin) antibody as a loading control, or with anti-PATMK antibody. PATMK protein levels for the PATMK shRNA transfectant amoebae are summarized in Table 2-6.

PATMK



PATMK (3532-3560) scrambled shRNA

PATMK (325-353) shRNA

PATMK (1326-1354) shRNA

PATMK (2273-2301) shRNA

PATMK (3532-3560) shRNA

Lgl

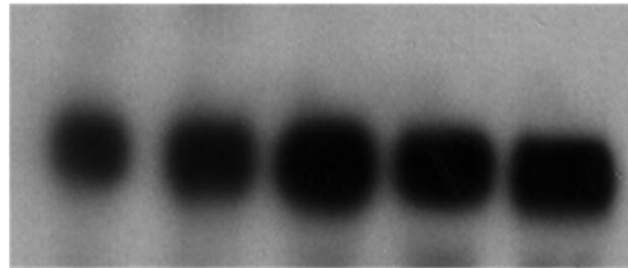


Figure 2-8. Western blot for Igl shRNA transfectants. A representative western blot is shown with one biological replicate each for the GFP control shRNA transfectant, the Igl1-specific (272–300), the Igl (1198–1226), the Igl (2412–2440), and the Igl (2777–2805) shRNA transfectants. HM1:IMSS samples are not shown. Results shown are representative of three biological replicates per shRNA transfectant with each sample run in triplicate. Serial dilutions of the crude lysates (1:2, 1:4, and 1:8) were also performed for each sample. Each membrane was probed with anti-actin antibody as a loading control, or with anti-Igl1 antibody. Igl1 protein levels for the Igl shRNA and GFP shRNA transfectants and HM1:IMSS nontransfected amoebae are summarized in Table 2-7.

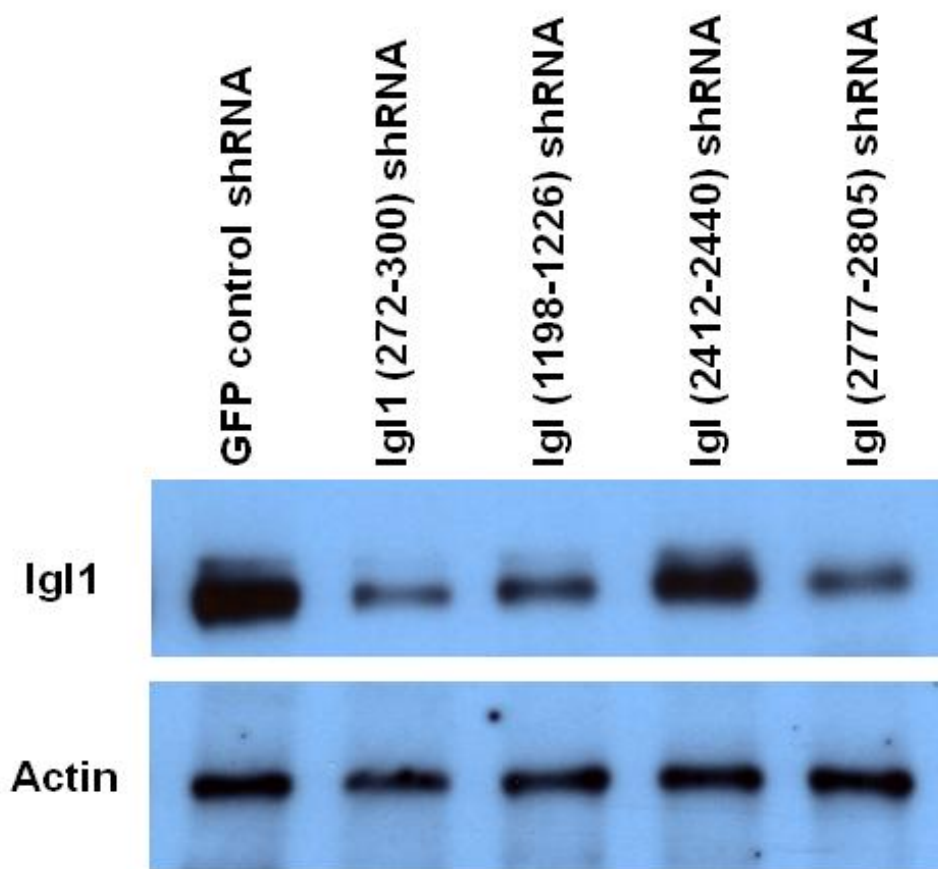


Figure 2-9. Western blot for URE3-BP shRNA transfectants. A representative western blot is shown with three biological replicates each (one dilution shown) for GFP control, URE3-BP (350–378), and URE3-BP (580–608) shRNA transfectants. HM1:IMSS samples are not shown. Results are representative of three biological replicates per shRNA transfectant with each sample run in triplicate. Serial dilutions of the crude lysates (1:2, 1:4, and 1:8) were also done for each sample. Each membrane was probed with anti-actin antibody as a loading control, or with anti-URE3-BP antibody. URE3-BP protein levels are summarized in Table 2-9.

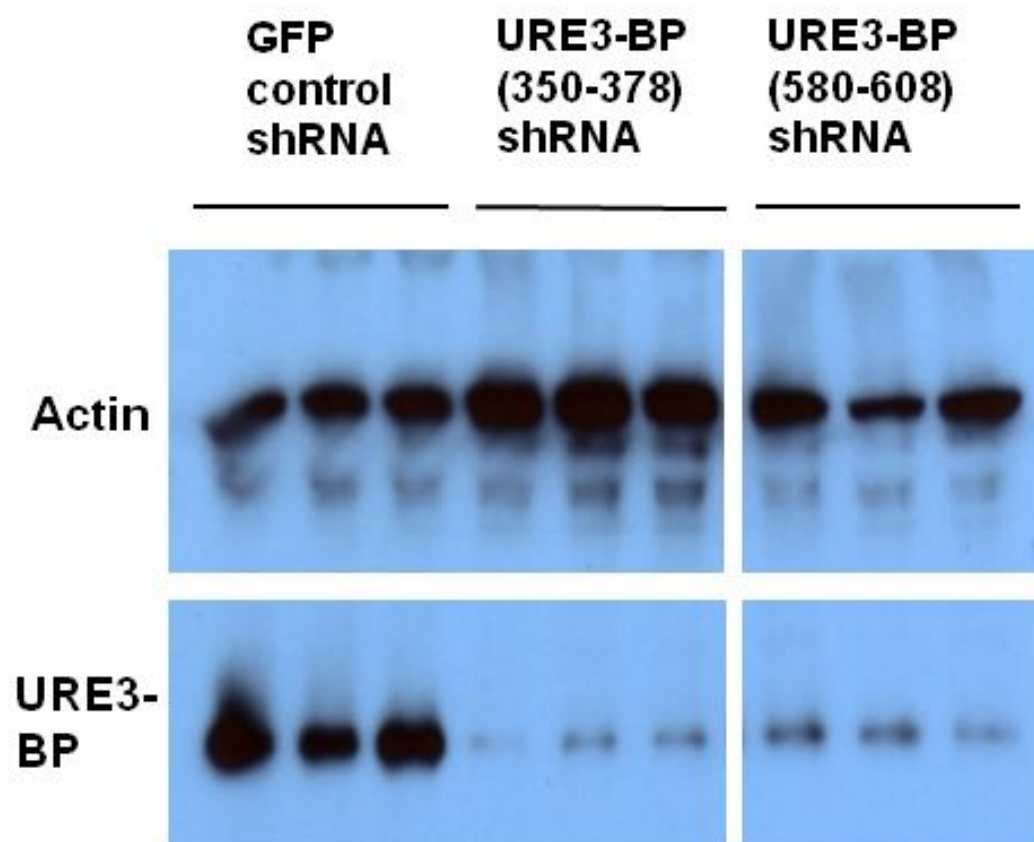


Figure 2-10. Western blot for EhC2A transfectants. A representative western blot is shown with three biological replicates each for EhC2A (363–391), EhC2A (502–530), and EhC2A (363–391 scrambled control) shRNA transfectants. Results are representative of three biological replicates per shRNA transfectant with each sample run in triplicate. Each sample was also serially diluted 1:2, 1:4, and 1:8. Each membrane was probed with anti-EhC2A antibody and with anti-actin antibody as a loading control. The level of EhC2A protein in the scrambled control transfectant was defined as 100% ($\pm 5\%$). The EhC2A (363–391) shRNA transfectant had strongly reduced levels of EhC2A protein: it was only $3.0 \pm 0.4\%$ of the scrambled control ($P = <0.0001$). The EhC2A (502–530) shRNA transfectant had no knockdown effect on EhC2A levels ($106.1 \pm 7.3\%$) ($P = 0.3141$). The two-tailed Student's t test was used for comparing sample averages.

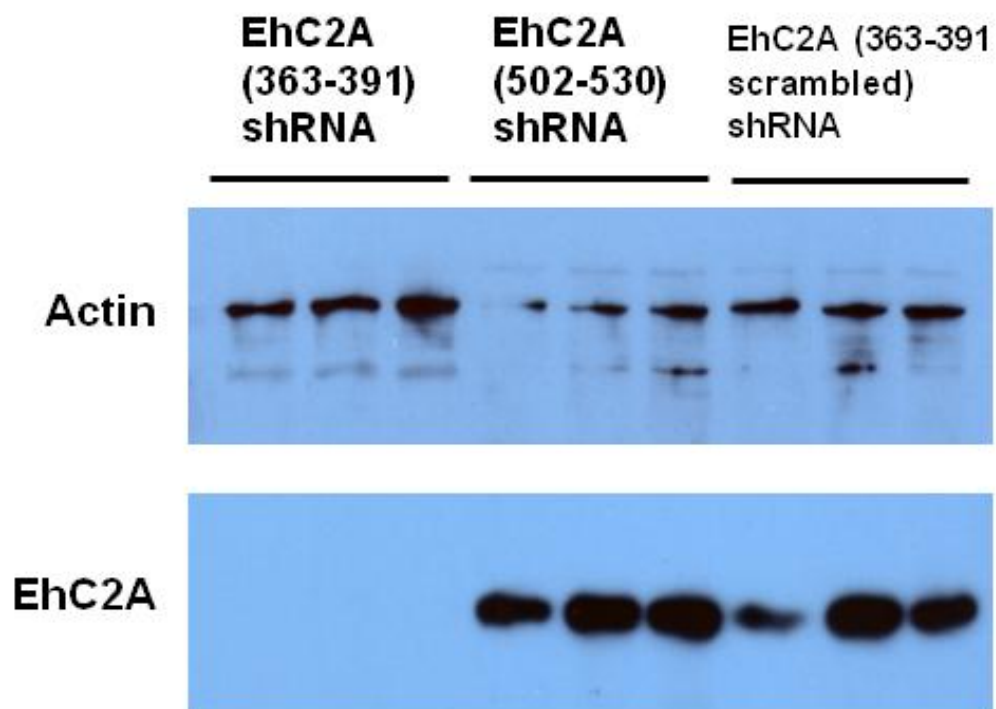


Figure 2-11. Northern blots of small RNAs extracted from Igl and PATMK transfectants.

To test if the U6 promoter was driving hairpin expression, shRNA transfectants (PATMK (3532-3560), PATMK (2273-2301), PATMK (3532-3560 scrambled) [73] Igl (1198-1226), Igl (2412-2440), and Igl (2777-2805) were selected with 30 µg/ml hygromycin for 48 hours before harvesting. HM1:IMSS non-transfected amoebae were included as negative controls. Small RNAs were extracted using the mirVana™ miRNA Isolation Kit (Ambion) (Applied Biosystems/Ambion, Austin, TX, USA). Fifty µg small RNA were loaded per lane on a 12% denaturing acrylamide gel, subjected to electrophoresis, and transferred to membrane. Oligo probes matching to the antisense strands of the hairpins were end-labeled with ³²P and were hybridized with each corresponding sample blot overnight at 37°C overnight, washed with low and medium stringency conditions, and exposed overnight to film. Note the two product sizes, which may correspond to the unprocessed hairpin (~60-70 nucleotides) (blue arrows) and the processed siRNA products (~30 nucleotides) (red arrows).

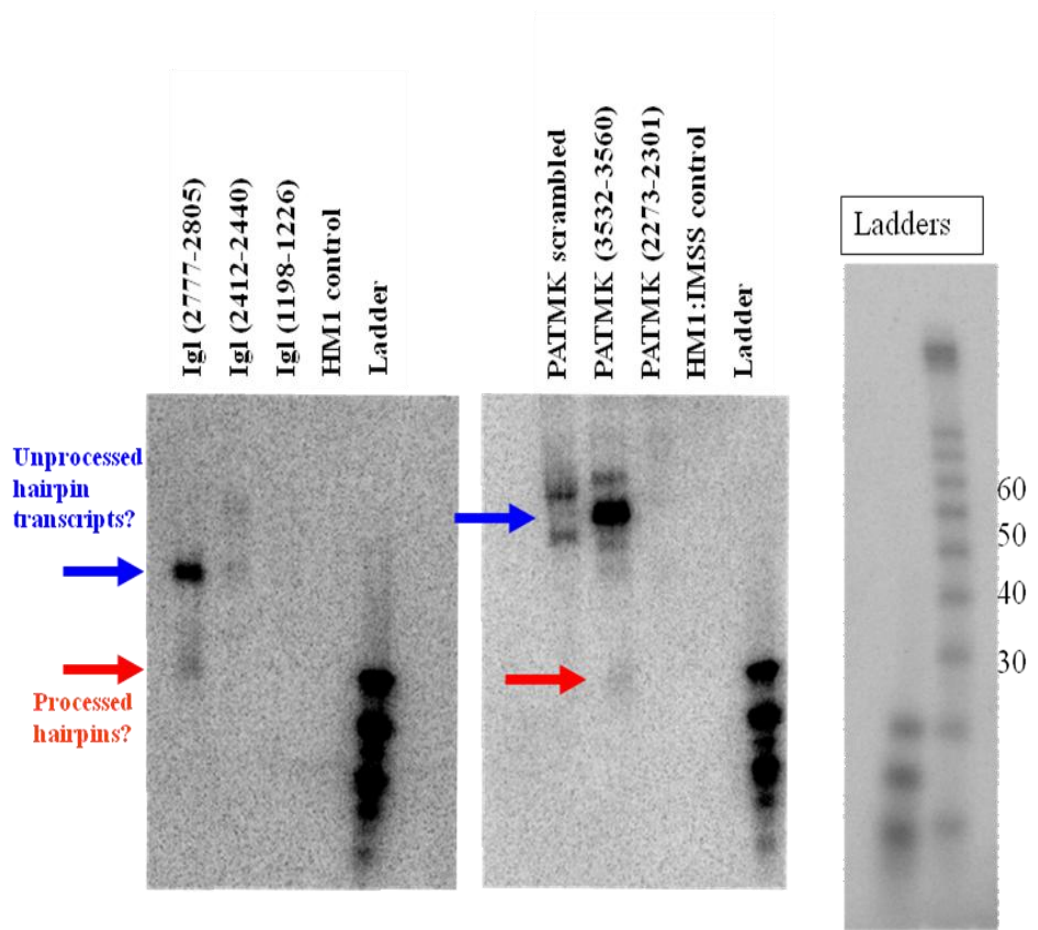


Table 2-1. Sequences chosen to generate shRNA constructs that were successfully transfected into amebae. The Ambion/Applied Biosystems siRNA Target Finder web site [181] was used to select 21-mers from the entire coding sequence of URE3-BP, the poly-proline region of EhC2A, or the identical or divergent regions of Igl1 and Igl2, which were then checked for sufficient GC content, lengthened to 29 nucleotides, and tested for sufficient sequence uniqueness by checking each 29-mer against the *E. histolytica* draft genome, originally using the BLAST function of the *E. histolytica* Genome Project database (TIGR Database *Entamoeba histolytica* Genome Project) [182]. A scrambled sequence was created as a control for EhC2A. A sequence directed against GFP [180] was included as a control for the Igl and URE3-BP selections. The constructs are named such that the numbers in parentheses following the gene name indicated the location of the shRNA sense strand within that gene sequence.

*This shRNA is unlikely to bind to Igl2 mRNA due to the poor match: there is only 69% identity (20 of 29 bases), which includes five gaps.

Gene (Construct)	Sequence	Location in mRNA (bp from ATG)	Total length of target mRNA (bp)
PATMK (325-353)	AATGGAGACTTATCAGTTG ATAAGATTCT	325-353	3840
PATMK (1326-1354)	AAACTCATATTTTCATTGAG TCATGTAGAG	1326-1354	
PATMK (2273-2301)	AAGGGTATTATTTAGAACA ATATGAGTGT	2273-2301	
PATMK (3532-3560)	AAGGGAATAGGAAGTCCAA TTTATATGGC	3532-3560	
PATMK (3532-3560 scrambled)	AATGCAAGAAGTATGGAGA CATCTATGTG	Does not match any <i>E. histolytica</i> mRNA	None
GFP	AAGGTGATGCAACATACGG AAAAC	Does not match any <i>E. histolytica</i> mRNA	None
Igl1 (272-300)	AAGTAAATACATCATCACA CTCTGGAAAT	272-300 (Igl1); 282-306 (Igl2)*	Igl1 3306; Igl2 3318
Igl (1198-1226)	AATGGACTTACATTGAATG GAACTCATTG	1198-1226 (both Igl1 and Igl2)	
Igl (2412-2440)	AACAGAATGTTTCAGATGGT TTTAGTGGAC	2412-2440 (Igl1); 2424-2452 (Igl2)	
Igl (2777-2805)	AAGGAACATGTATACCATG CACATCACCA	2777-2805 (Igl1); 2789-2817 (Igl2)	
URE3-BP (350-378)	AACTTGCATACAATCTCTT CGTTATGAAC	350-378	663
URE3-BP (580-608)	AATCCATACTATGGTCCAA TGAAACCATT	580-608	
EhC2A (363-391)	AATGGTTCCACCAATGCAA CCAGGCATGA	363-391	567
EhC2A (502-530)	AAGCTTACCCACCACCTGG ATATCCACCA	502-530; also 406-434	
EhC2A (363-391 scrambled)	AAGGCTAGACAATCCAGAC CGTTCCAGAT	Does not match any <i>E. histolytica</i> mRNA	None

Table 2-2. Oligos used for generating shRNA constructs by PCR and transfected into amoebae. The sequences shown in Table 2-1 were used to design primers for two-step PCR, based on the method used by Gou *et al* (2003) [180] and diagrammed in Figure 2-5. The final PCR product contained the *E. histolytica* U6 promoter with a *Hind*III site on the 5' end, an *Apa*I site at the 3' end of the U6 promoter, the 29-nt sense strand of the hairpin, the 9 bp loop TTCAAGAGA, the antisense strand of the hairpin, and the U6 terminator sequence followed by a *Not*I restriction site. The forward primer, "U6 *Hind*III forward", contained the *Hind*III recognition site and the 5' end of the U6 promoter, the first reverse primer (R1) contained the sequence of the sense strand of the shRNA and the future loop, and the second reverse primer (R2) contained the loop sequence, the antisense strand sequence, and the U6 termination sequence. A control GFP sequence [180] was used to design oligos for creating a shRNA construct as a transfection control.

Oligo Name	Oligo Sequence
U6 <i>Hind</i> III forward [72, 73]	CTACTGAAGCTTGTTTTTATGAAAAAGTGTATTTGC
PATMK (325-353) R1 [73]	TCTCTTGAAAGAATCTTATCAACTGATAAGTCTCCAG GGCCC AATTTTATTTTTCTTTTTATCC
PATMK (325-353) R2 [73]	GAATGCGGCCGCAAAAAATGGAGACTTATCAGTTGAT AAGATTCT TCTCTTGAA
PATMK (1326-1354) R1	TCTCTTGAACCTCTACATGACTCAATGAAATATGAGTT TGGGCC AATTTTATTTTTCTTTTTATCC
PATMK (1326-1354) R2	GAATGCGGCCGCAAAAACTCATATTTTCATTGAGTCA TG TAGAG TCTCTTGAA
PATMK (2273-2301) R1 [73]	TCTCTTGAAACACTCATATTTGTTCTAAATAATACCCT TGGGCC AATTTTATTTTTCTTTTTATCC
PATMK (2273-2301) R2 [73]	GAATGCGGCCGCAAAAAAGGGTATTATTTAGAACAAAT ATGAGTGT TCTCTTGAA
PATMK (3532-3560) R1 [73]	TCTCTTGAAGCCATATAAATTGGACTTCCTATTCCCT TGGGCC AATTTTATTTTTCTTTTTATCC
PATMK (3532-3560) R2 [73]	GAATGCGGCCGCAAAAAAGGGAATAGGAAGTCCAATT TATATGGC TCTCTTGAA
PATMK (3532-3560) scrambled R1 [73]	TCTCTTGAATCTGGAACGGTCTGGATTGTCTAG CCTTGGGCC AATTTTATTTTTCTTTTTATCC
PATMK (3532-3560) scrambled R2 [73]	GAATGCGGCCGCAAAAAAGGCTAGACAATCCAGACCG TTCCAGAT TCTCTTGAA
GFP R1 [72, 180]	TCTCTTGAA GTTTTCCGTATGTTGCATCACCTT GGGCC AATTTTATTTTTCTTTTTATCC
GFP R2 [72, 180]	TCGATCGCGGCCGCAAAAAAGGTGATGCAACATACGG AAAAC TCTCTTGAA

Table 2-2 is continued on page 83.

Table 2-2, continued from page 82.

Igl1 (272-300) R1 [72]	TCTCTTGAAATTTCCAGAGTGTGATGATGATGATTTTAC TT GGGCCCAATTTTATTTTTCTTTTTATCC
Igl1 (272-300) R2 [72]	TCGATCGCGGCCGCAAAAAAGTAAATACATCATCAC ACTCTGGAAAT TCTCTTGAA
Igl (1198-1226) R1 [72]	TCTCTTGAACAATGAGTTCCATTCAATGTAAGTCCA TT GGGCCCAATTTTATTTTTCTTTTTATCC
Igl (1198-1226) R2 [72]	TCGATCGCGGCCGCAAAAAATGGACTTACATTGAAT GGAACTCATTG TCTCTTGAA
Igl (2412-2440) R1 [72]	TCTCTTGAAGTCCACTAAAACCATCTGAACATTCTG TT GGGCCCAATTTTATTTTTCTTTTTATCC
Igl (2412-2440) R2 [72]	TCGATCGCGGCCGCAAAAAACAGAATGTTTCCAGATGG TTTTAGTGGAC TCTCTTGAA
Igl (2777-2805) R1 [72]	TCTCTTGAATGGTGATGTGCATGGTATACATGTTCC TT GGGCCCAATTTTATTTTTCTTTTTATCC
Igl (2777-2805) R2 [72]	TCGATCGAGCTCGCGGCCGCAAAAAAGGAACATGTA TACCATGCACATCACCA TCTCTTGAA
URE3-BP (337) R1 [72]	TCTCTTGAAGTTCATAACGAAGAGATTGTATGCAAG TT GGGCCCAATTTTATTTTTCTT
URE3-BP (337) R2 [72]	TCGATCGAGCTCGCGGCCGCAAAAAACTTGCATACA ATCTCTTCGTTATGAAC TCTCTTGAA
URE3-BP (567) R1 [72]	TCTCTTGAATAATGGTTTTTATTGGACCATAGTATGGA TT GGGCCCAATTTTATTTTTCTTTTTATCC
URE3-BP(567) R2 [72]	TCGATCGAGCTCGCGGCCGCAAAAAATCCATACTAT GGTCCAATGAAACCATT TCTCTTGAA
EhC2A (363-391) R1 [72, 74]	TCTCTTGAATCATGCCTGGTTGCATTGGTGGAACCA TT GGGCCCAATTTTATTTTTCTTTTTATCC
EhC2A (363-391) R2 [72, 74]	TCGATCGCGGCCGCAAAAAATGGTTCCACCAATGCA ACCAGGCATGA TCTCTTGAA
EhC2A (502-530) R1 [72]	TCTCTTGAATGGTGGATATCCAGGTGGTGGGTAAGC TT GGGCCCAATTTTATTTTTCTTTTTATCC
EhC2A (502-530) R2 [72]	TCGATCGCGGCCGCAAAAAATGGTTCCACCAATGCA ACCAGGCATGA TCTCTTGAA
EhC2A (363-391 scrambled) R1 [72, 74]	TCTCTTGAATCTGGAACGGTCTGGATTGTCTAGCC TT GGGCCCAATTTTATTTTTCTTTTTATCC
EhC2A (363-391 scrambled) R2 [72, 74]	TCGATCGCGGCCGCAAAAAAGGCTAGACAATCCAGA CCGTTCCAGAT TCTCTTGAA

Table 2-3: Sequences of oligos used for Igl and URE3-BP amplification in qRT-PCR. Oligo pairs were designed to amplify short sections of Igl or URE3-BP. For Igl, four pairs of oligos were used. Two pairs were designed to amplify both Igl1 and Igl2 simultaneously: one pair amplified the 5' end of the gene (Igl 5' oligo pair) and one the 3' end (Igl 3' oligo pair). To amplify Igl1 and Igl2 separately, oligo pairs were designed to amplify a short section unique to Igl1 or Igl2 (Igl1 or Igl2 oligo pair, which have separate forward primers, but the same reverse primer in common) near the 5' end of the mRNA. Three oligo pairs were used to amplify short sections of URE3-BP: one pair the 5' end, one pair near the middle, and one pair near the 3' end. The actin and Jacob primers were designed to amplify all family members or alleles [67].

Oligo Name	Oligo Sequence	mRNA/ cDNA section amplified (bp from ATG)	Total length of mRNA (bp)
Igl 5' F [72]	GCTGTTCCACATTGTG CATCAGTTTCAAATG	85–450 (Igl1)	3306 (Igl1), 3318 (Igl2)
Igl 5' R [72]	TTCTGCATGATCTTCT GTAGTTGCATTATCAC ATAAC	85–459 (Igl2)	
Igl 3' F [72]	TGAAGGCACTTCTACA GAAGATAATAAAAT	2967–3166 (Igl1)	
Igl 3' R [72]	TATGTCTTGAACATGG AATACATGATC	2979–3178 (Igl2)	
Igl1 F [72]	TCTTGTAATAAGTTCC CGGAGCA	634–841 (Igl1)	
Igl1 R [72]	CATCAGAAACAGTACA TCTTTTATTACATG		
Igl2 F [72]	GTACTAAATACCCAGA TCATTGTTCAA	643–841 (Igl2)	
Igl2 R [72]	CATCAGAAACAGTACA TCTTTTATTACATG		
URE3-BP 5' F [72]	CCTGTAGCTAATTTCT GTTTATGGAATC	10–155	663
URE3-BP 5' R [72]	CTTGTATATTGATCTA ATGGGATAGTGTTAAG		
URE3-BP Middle F [72]	GATGAGAATTTTTGAT ACTGATTTTAAATGGAC	276–454	
URE3-BP Middle R [72]	GATTAATATAGAATCC AAGTTGTTGAAGAG		
URE3-BP 3' F [72]	CTGTGATCTTAATTGT TGGATTG	504–658	
URE3-BP 3' R [72]	CCAAGAGGGAAGTAAC AACGT		
Actin F [67, 72]	GCACTTGTTGTAGATA ATGGATCAGGAATG	125 bp, variable location (detects all family members/ alleles)	variable
Actin R [67, 72]	ACCCATACCAGCCATA ACTGAAACG		
Jacob F [67, 72]	CAAAGGAGTTCAAATG GGATGTGTTAG	156-287 (detects all family members/ alleles)	435
Jacob R [67, 72]	TTATTTGGTGTAGGAG TTGGTAATGGG		

Table 2-4. Proteins with similar sequences to PATMK amplified in qRT-PCR, and sequences of oligos used to amplify them to test for off-target effects in PATMK

knockdown. To test that only PATMK and not any other related gene was being knocked down, nine genes, six of them protein kinases or putative protein kinases, were chosen with sequence identity ~80% to PATMK (3532-3560), the PATMK shRNA that gave the best knockdown. Oligo pairs were designed that would amplify short sections <200 bp of each tested gene or of PATMK.

Protein	GenBank Accession Number	Description
PATMK (control)	XM_650501	Putative tyrosine kinase
115m.00124	XM_647447	Hypothetical protein
115m.00129	XM_647427	Hypothetical protein
93m.00137	XM_648024	Hypothetical protein
99m.00167	XM_647846	Putative tyrosine kinase
TMK 32 (72m.00194)	XM_648689	Putative tyrosine kinase
243.m00079	XM_651923	Protein tyrosine kinase domain-containing protein
TMK 11 (15m.00355)	XM_651172	Putative tyrosine kinase
TMK 56 (5m.00482)	XM_651923	Protein tyrosine kinase domain-containing protein
TMK 79 (71m.00130)	XM_645052	Protein tyrosine kinase domain-containing protein
Actin (control)	XM_001914362, XM_650398, XM_650064, XM_001913751, XM_651646, XM_642962, XM_643736	actin or putative actin
Jacob (control)	XM_650162, XM_645825, XM_642844	Cyst wall-specific glycoprotein Jacob

Table 2-4 is continued on page 88.

Table 2-4, continued from page 87.

Oligo Name	Oligo Sequence	mRNA/ cDNA section amplified (bp from ATG)	Total length of mRNA (bp)
115m.00124 F	CGTACCAATCGAAGTGGTTG	984-1117	1329
115m.00124 R	GACTATTTGAGTTTGGTTTAGTTTC TTC		
115m.00129 F	GACTACCTGTTTCATATCGTCC	1145-1262	3420
115m.00129 R	GACATTCTTACCATTAAAATTGGTT GAG		
93m.00137 F	CTTCTTCAACCCCCAGTGTTA	971-1130	3570
93m.00137 R	CACATTCCATATAATGTTTCAGATT CATAAC		
99m.00167 F	AGAAGAAAAAGTGTTAAAGCAGAAT CTACT	1834-1973	3375
99m.00167 R	GCCAAATTAATGTTAATTCCCATT TATCTG		
TMK 32 (72m.00194) F	GATTCTGGAAGTTGTTATCAATGTG AA	841-1007	4047
TMK 32 (72m.00194) R	GAATAATACCCATCTTCACATATTG AACA		
243.m00079 F	CGTGCTGAATGGAGAAGAGTT	2677-2813	3591
243.m00079 R	TTTACAATATATGGACAACGAATAC GTTC		
TMK 11 (15m.00355) F	CCAGGTGATGGTCAATTTTCTG	1348-1507	5451
TMK 11 (15m.00355) R	CTAATGTTGATACATCTAATACTCC TGAA		
TMK 56 (5m.00482) F	GATCCGAGTCGTCCATATACA	550-703	3591
TMK 56 (5m.00482) R	CAAATTCAAGGTCTGTTCCAATTTT AC		
TMK 79 (71m.00130) F	CCTATTGGTGGTGGTACATTTG	2647-2801	3615
TMK 79 (71m.00130) R	GGACAACGAATACGTTCCATC		

Table 2-4 is continued on page 89.

Table 2-4, continued from page 88.

Oligo Name	Oligo Sequence	mRNA/ cDNA section amplified (bp from ATG)	Total length of mRNA (bp)
PATMK (Phak1) F	GGAAATGGGTGTGCTGTTTGTCAAAA GGATATTA	2056-2225	3840
PATMK (Phak1) R	CCAATAGTTTCATTTAAAGCAACACAC TTCGCG		
Actin F [67, 72]	GCACTTGTTGTAGATAATGGATCAGGA ATG	(22-147), (22-147), (87-212), (22-147), (58-183), (22-147), (78-203)	229, 1131, 1230, 1131, 1687, 1158, 1214
Actin R [67, 72]	ACCCATACCAGCCATAACTGAAACG		
Jacob F [67, 72]	CAAAGGAGTTCAAATGGGATGTGTTAG	156-287	435
Jacob R [67, 72]	TTATTTGGTGTAGGAGTTGGTAATGGG		

Table 2-5. Knockdown of PATMK mRNA, and qRT-PCR study to evaluate if expression of PATMK-related genes is affected by transfectants expressing shRNA against PATMK (off-target effects). To test that only PATMK, and not any other related gene, was being knocked down, nine genes, six of them protein kinases or putative protein kinases, were chosen with sequence identity ~80% to PATMK (3532-3560), the PATMK shRNA that gave the best protein knockdown. Three biological replicates each of the PATMK shRNA transfectant lines PATMK (3532-3560) or the control PATMK (3532-3560 scrambled) were selected with 30 micrograms of hygromycin B per ml of TYI medium before harvesting for RNA extraction and cDNA synthesis. Samples were amplified with primer sets for each of the nine chosen genes or with actin or Jacob primers in qRT-PCR. qRT-PCR was performed as described in the qRT-PCR section. qRT-PCR cycling conditions were 95°C for 15 minutes, followed by 45 cycles of 95°C 30 s; 47°C 30 s; 72°C 30 s, followed by one cycle of 72°C for 3 min, followed by a melt curve from 45°C to 90°C increasing 0.2°C every cycle with a 2-second hold. For quantitation, all samples were run in triplicate and data was combined from multiple runs. C(t) values were averaged and calculations performed as described in the Quantitative Reverse-Transcription Real-Time PCR (qRT-PCR) section.

Gene Amplified	$\Delta\Delta C(t)$ = Difference between the test gene $\Delta C(t)$ for PATMK (3532-3560) scrambled and PATMK (3532-3560)	t value for $\Delta\Delta C(t)$	P value	Statistically significant?
115m.00124 (XM_647447)	0.869 ± 3.379	0.772	0.4623	No
115m.00129 (XM_647427)	0.604 ± 3.784	0.479	0.6448	No
93m.00137 (XM_648024)	0.332 ± 2.388	0.417	0.6876	No
99m.00167 (XM_647846)	-0.528 ± 2.353	-0.673	0.5199	No
TMK32 (72m.00194) (XM_648689)	-0.679 ± 2.802	-0.727	0.4880	No
243.m00179 (XM_651923)	1.910 ± 3.408	1.681	0.1313	No
TMK11 (15m.00355) (XM_651172)	-0.894 ± 2.549	-1.052	0.3235	No
TMK56 (5m.00482) (XM_651923)	-0.359 ± 3.406	-0.316	0.7601	No
TMK79 (71m.00130) (XM_645052)	0.845 ± 3.158	0.803	0.4452	No
PATMK (26.m00300) (XM_650501)	$>14.848 \pm 2.839$	<-15.690	<0.0001	Extremely

Table 2-6: Summary of PATMK protein levels in PATMK shRNA transfectants. The average level of PATMK protein in the PATMK (3532-3560) scrambled control shRNA transfectants was defined as 100% expression of PATMK protein for computational purposes. Protein levels of PATMK, using Lgl (the light subunit of the Gal/GalNAc lectin) as a standard for comparison, were quantified from western blotting. Values are expressed as the percentage of the PATMK (3532-3560) scrambled level \pm SE, with the P-value following each, calculated using Student's t test. All PATMK shRNAs allowed for statistically significant knockdown of PATMK protein ($P < 0.05$). For western blotting, there were three biological replicates. A representative western blot is shown in Figure 2-7.

shRNA Transfectant or Control Sample	% of Control Protein Level (\pm SE)	P-value
PATMK (3532-3560) scrambled	100.0 \pm 5.7%	--
PATMK (325-353)	53.6 \pm 4.5%	0.0436
PATMK (1326-1354)	45.2 \pm 5.2%	0.0426
PATMK (2273-2301)	7.4 \pm 3.5%	0.016
PATMK (3532-3560)	0.6 \pm 1.1%	0.0051

Table 2-7. Summary of Igl1 protein levels in Igl shRNA transfectants. The average level of Igl1 protein in the GFP control shRNA transfectants was defined as 100% expression of Igl1 protein for computational purposes. Protein levels of Igl1, using actin as a standard for comparison, were quantified from western blotting. Values are expressed as the percentage of the GFP level \pm SE, with the P-value following each, calculated using Student's t test. For western blotting, there were three biological replicates, each run in triplicate sets of serial dilutions (1:2, 1:4, and 1:8), with the exception of the HM1:IMSS nontransfected samples having one biological replicate rather than three. Protein levels were not statistically different between the Igl1 (272–300), Igl (1198–1226), and Igl (2777–2805) samples (tested with ANOVA, one-tailed, $\alpha = 0.05$, $0.1 < P < 0.25$) or the GFP, HM1:IMSS, and Igl (2412–2440) samples (tested with ANOVA, one-tailed, $\alpha = 0.05$, $P > 0.25$). A representative western blot is shown in Figure 2-8.

shRNA Transfectant or Control Sample	% of Igl1 Protein Level (\pm SE)	P-value
GFP	100.0 \pm 3.6	
HM1:IMSS	115.5 \pm 11.8	0.1449
Igl (2412–2440)	95.3 \pm 3.2	0.2078
Igl1 (272–300)	27.8 \pm 1.3	< 0.0001
Igl (1198–1226)	42.3 \pm 2.1	< 0.0001
Igl (2777–2805)	38.1 \pm 3.1	< 0.0001

Table 2-8: Summary of Igl mRNA levels in Igl shRNA transfectants. For qRT-PCR, samples were amplified with the actin oligo pair as a control, or with four pairs of Igl oligos: Igl 5', amplifying the 5' end of both Igl1 and Igl2, Igl 3', amplifying both Igl1 and Igl2 at the 3' end, and oligos specific for Igl1 and Igl2 individually, amplifying Igl1- or Igl2-specific sequences near the 5' end. Oligo sequences are shown in Table 2-3. Three biological replicates were each assayed in quadruplicate sets with each oligo pair, with the exception of the HM1:IMSS samples, which had one biological replicate. Igl and actin levels were calculated by using both the relative standard curve and the $\Delta\Delta C(t)$ method [185, 187] and actin was used as the normalization control. The average level of Igl in the GFP control shRNA transfectants was defined as 100% expression of Igl mRNA for computational purposes. Igl levels in the Igl transfectant samples and nontransfected HM1:IMSS were compared to the GFP control, and are shown as the percentage of Igl mRNA relative to the GFP control (\pm SE). Statistical analysis was performed using Student's t test (two-tailed), groups were compared using ANOVA, and the GraphPad QuickCalcs P-value calculator [183] was used to calculate P-values.

shRNA Transfectant or Control Sample	Igl 5' Oligo Pair	P-value	Igl 3' Oligo Pair	P-value	Igl1 Oligo Pair	P-value	Igl2 Oligo Pair	P-value
GFP6	100.0 ± 4.1	--	100.0 ± 4.9	--	100.0 ± 3.0	--	100.0 ± 4.0	--
HM1:IMSS	101.4 ± 4.3	0.7741	96.1 ± 3.5	0.3239	105.5 ± 3.1	0.1382	103.9 ± 6.1	0.5713
Igl (2412–2440)	100.6 ± 5.0	0.9172	103.4 ± 9.1	0.7717	91.1 ± 6.9	0.2426	106.0 ± 5.2	0.2919
Igl1 (272–300)	71.3 ± 2.9	<0.0001	67.1 ± 3.0	<0.0001	61.1 ± 3.2	<0.0001	70.2 ± 2.7	<0.0001
Igl (1198–1226)	70.9 ± 2.7	<0.0001	62.1 ± 1.6	<0.0001	68.3 ± 2.5	<0.0001	76.8 ± 1.6	<0.0001
Igl (2777–2805)	68.1 ± 3.3	<0.0001	62.3 ± 2.9	<0.0001	74.1 ± 3.3	<0.0001	77.8 ± 3.0	<0.0001

Table 2-9: Summary of URE3-BP protein levels in URE3-BP shRNA transfectants. The average level of URE3-BP protein was defined as being 100% in the GFP shRNA control transfectants. The levels of URE3-BP and the actin standard were quantified from western immunoblots. Values are expressed as the percentage of URE3-BP protein of the GFP control shRNA transfectant level \pm SE, with the P-value following each. There were three biological replicates for all samples, each blotted in triplicate plus serial dilutions (1:2, 1:4, and 1:8) except for HM1:IMSS nontransfectants, which had one biological replicate run in triplicate plus serial dilutions. All samples were normalized to actin and compared to the GFP control shRNA transfectants using Student's t test. URE3-BP protein levels were not statistically different between the URE3-BP (350-378) and (560-608) samples (two-tailed Student's t test for comparing two sample averages, $P = 0.3262$), or between the GFP control shRNA transfectants and HM1:IMSS nontransfected samples (two-tailed Student's t test for comparing two sample averages, $P = 0.2346$). A representative western blot is shown in Figure 2-9.

shRNA Transfectant or Control Sample	% of Control Protein Level (\pm SE)	P-value
GFP	100.0 \pm 9.9	--
HM1:IMSS	111.3 \pm 15.8	0.6189
URE3-BP (350–378)	10.8 \pm 1.0	<0.0001
URE3-BP (580–608)	13.8 \pm 2.6	<0.0001

Table 2-10: Summary of mRNA levels in GFP shRNA control transfectants, URE3-BP shRNA transfectants, and nontransfected HM1:IMSS trophozoites. The average URE3-BP transcript level as measured by qRT-PCR and normalized to actin was defined as being 100% in the GFP shRNA control transfectants. HM1:IMSS nontransfected amoebae were also included. Three different oligo pairs amplifying the 5', middle, and 3' sections of URE3-BP were used (sequences and locations are shown in Table 3-3). Student's t test was used for statistical analysis. Three biological replicates were each assayed in quadruplicate with each oligo pair, with the exception of the HM1:IMSS samples, which had one biological replicate. Values are expressed as the percentage of URE3-BP mRNA of the GFP control shRNA transfectant level \pm SE, with the P-value following each.

shRNA Transfectant or Control Sample	URE3-BP 5' Oligo Pair	P-value	URE3-BP Middle Oligo Pair	P-value	URE3-BP 3' Oligo Pair	P-value
GFP	100.0 ± 2.9	--	100 ± 2.8	--	100 ± 4.3	--
HM1:IMSS	106.4 ± 5.8	0.2928	108.9 ± 5.6	0.1008	102.8 ± 5.0	0.5792
URE3-BP (350–378)	67.0 ± 2.5	<0.0001	67.4 ± 2.0	<0.0001	72.2 ± 2.8	<0.0001
URE3-BP (580–608)	12.4 ± 0.8	<0.0001	13.5 ± 3.3	<0.0001	12.5 ± 3.8	<0.0001

Chapter 3

Partial Characterization of the *Entamoeba histolytica* Low Molecular Weight Protein Tyrosine Phosphatase

Background on Protein Tyrosine Phosphatases

Tyrosine Phosphorylation of Proteins

The phosphorylation of proteins on tyrosine residues plays important roles in the regulation of cellular processes including cell migration, differentiation, proliferation, growth, metabolism, gene transcription, cell cycle, cell-cell communication, response to environmental changes, and cell survival via signal transduction pathways [203, 204]. Modification of proteins by phosphorylation of tyrosine residues by tyrosine-specific kinases was not discovered until 1979, when researchers working with the tumor virus proteins middle T antigen from polyomavirus [205], *v*-Src protein from Rous sarcoma virus [206], and Abl protein from the Abelson leukemia virus [207], found these proteins had the ability to phosphorylate tyrosine residues *in vitro* [205, 208]. These findings suggested that tyrosine phosphorylation of substrate proteins could be involved in morphological changes and unregulated growth of tumor cells [208]. Tyrosine phosphatases were discovered somewhat later: one of the first protein tyrosine phosphatases found, PTP1B, was first purified and sequenced in 1988 [203, 209, 210].

Tyrosine phosphorylation of proteins is mediated by the balanced action of protein tyrosine kinases and protein tyrosine phosphatases; this reversible process of protein phosphorylation is highly regulated and requires both a protein kinase and a protein phosphatase to work as a team [203, 204, 211]. Resting cells typically have low levels of protein phosphotyrosines [212, 213]. The pattern of cellular proteins with phosphorylated tyrosine

residues results from the combined action of these kinases and phosphatases [203, 204].

Theoretically, phosphorylation at a particular tyrosine residue can be regulated by changing the activity of its kinase or phosphatase, or both [204]. The dysregulation of phosphorylation control can disrupt cellular processes, resulting in pathological situations, such as cancer, if the kinase or phosphatase is mutated or anomalously expressed, resulting in abnormal phosphorylation patterns of their substrates [203, 211, 214]. Many dominant oncogenes are mutated versions of protein tyrosine kinases or phosphatases [203]. One of the best known of these is the proto-oncogene kinase Src, originally identified in its oncogenic form, *v*-Src, which could form tumors in birds infected with the Rous sarcoma virus [215, 216]. *v*-Src is missing its C-terminal segment, which contains a negatively regulatory tyrosine residue that when phosphorylated, puts the enzyme in an inactive state [215, 216]. The *v*-Src can no longer be switched off, so its action can no longer be regulated [215, 216]. An example of a dysregulated phosphatase causing disease is the overexpression of PRL3 in metastatic colon cancer [217, 218].

The Protein Tyrosine Phosphatase Superfamily

Protein tyrosine phosphatases make up a large family of enzymes that are vital in regulating other proteins in signal transduction pathways by catalyzing the removal of a phosphate group (dephosphorylation) from a tyrosine residue. All protein tyrosine phosphatases (PTPs) contain the conserved motif CX₅R, comprising the enzyme active site [203, 212, 219, 220, 221, 222, 223, 224, 225]. The active site, also called the PTP loop or P-loop, binds the phosphate group of the phosphotyrosine [212, 219, 222, 223]. The PTP superfamily is divided into three class types based on the substrate specificities and structures of the members (Figure 3-1); they all share the same catalytic mechanism that depends upon the formation of a thiol-phosphate intermediate and general acid-general base catalysis, which requires there be a conserved cysteine

residue in the active site [212, 220, 227, 221, 222, 223, 225, 226]. However, beyond the conservation of the active site, these three class types are only distantly related and have little sequence similarity [203, 220, 222, 223, 225, 227]. Type I PTPs includes the dual-specificity phosphatases, which dephosphorylate phosphoserine and phosphothreonine residues in addition to phosphotyrosine residues, and also includes the classical PTPs, which are subdivided into a non-receptor cytosolic group and a receptor-like group whose members have a transmembrane domain (Figure 3-1) [203, 212, 217, 220, 221, 222, 225]. Each of these is further divided into subgroups. Subgroups in the dual-specificity protein phosphatase group include MAPK phosphatases, Slingshots, Cdc14s, PTENs, myotubularins, and atypical dual-specificity phosphatases [212, 222]. Subgroups in the classical protein tyrosine phosphatase non-receptor group include PTP1B, SHPs, TC-PTP, PTP-PEST, and PTPL1 [212, 219, 222]. Subgroups in the classical receptor-like group (few ligands have been yet identified) include CD45, LAR, PTP α , RPTP σ , and RPTP γ [212, 219, 222]; most members of this group contain one or two intracellular catalytic domains, a single transmembrane domain, and a variable extracellular ligand-binding domain [212, 220, 221, 222]. The Type III PTP class contains only the Cdc25 protein tyrosine phosphatases: Cdc25A, Cdc25B, and Cdc25C [212, 221, 222, 225]. The Type II PTP class also contains only one member: the low molecular weight protein tyrosine phosphatases (LMW-PTPs) [212, 221, 222, 223, 225] (Figure 3-1).

Substrate Catalysis by Protein Tyrosine Phosphatases

All the three-dimensional structures that have been solved for the PTP catalytic domains share the same features, even though the phosphatase signature motif is localized at different positions in PTP proteins [219, 225]. This is because all protein tyrosine phosphatases share a common mechanism to catalyze the removal of the phosphoryl group from their phosphotyrosine

substrates. Three key structural features are conserved: the catalytic cysteine and arginine in the active site, along with the general acid/base containing surface loop, which contains the conserved aspartic acid residue (Figure 3-2) [224, 225, 228]. The common catalytic strategy involves a two-step mechanism: first, the thiol side chain of the invariant active site catalytic cysteine serves as a nucleophile and attacks the phosphoryl group from the substrate, forming a thiol-phosphate intermediate (Figure 3-2) [203, 224, 225, 229, 230, 231, 232]. The invariant active site arginine residue forms hydrogen bonds with the substrate phosphoryl group via its guanidinium group, and stabilizes the transition state during hydrolysis [224, 225, 232]. Binding of substrate promotes conformational change of the general acid/base containing surface loop, which brings the conserved aspartic acid residue into the active site, positioning its carboxyl group close to the leaving group oxygen of the substrate, allowing for proton transfer [203, 224, 229, 232]. The aspartic acid first acts as a general acid by providing a proton to the oxygen of the tyrosine residue leaving group, enhancing formation of the enzyme-product complex (Figure 3-2) [203, 212, 219, 224, 225]. In the second step, the aspartic acid acts as a general base by accepting a proton from water, and allows the hydrolysis of the enzyme-product complex, releasing inorganic phosphate and returning the enzyme to its resting state (Figure 3-2) [203, 212, 219, 224, 225, 229]. It is unusual to have a phosphocysteine intermediate in a phosphatase-catalyzed hydrolysis reaction [203, 225]; in the same reaction catalyzed by serine/threonine phosphatases, an activated water molecule attacks the phosphorus substrate atom [225, 233]. The enzyme may have evolved to form a cysteine phosphate intermediate in catalysis because the bond energy, 45-50 kcal/mol, of the P-S bond is less than the 95-100 kcal/mol of the P-O bond, making the P-S bond easier to cleave [225].

Oxidation of the Active Site Catalytic Cysteine of Protein Tyrosine Phosphatases

Oxidation of the catalytic cysteine residue is important in regulating the activity of PTPs [234] since the side chain of the conserved cysteine at physiological pH is that of a thiolate anion ($-S^-$) rather than that of a thiol group ($-SH$) and has a low pKa value [225, 227]. Whether or not the catalytic cysteine residue can be oxidized or not is strongly dependent on pH, indicating its microenvironment plays an important role [234]. The thiolate anion of the catalytic cysteine is highly susceptible to being oxidized, and an oxidized cysteine can no longer function in the dephosphorylation reaction [227, 235]. Reactive oxygen species (ROS) such as H_2O_2 or O_2^- are generated transiently under certain conditions such as cell stimulation with cytokines and growth factors or oxidative bursts [211]. This can result in oxidation of the catalytic cysteine, forming sulfenic acid from the thiolate anion ($-S^-$ to $-S-OH$), and rendering the enzyme inactive [212, 235, 236] since it can no longer form a thiol-phosphate intermediate [211]. Highly oxidizing conditions lead to further oxidation, allowing the formation of sulfinic ($-S-O_2H$) and sulfonic acid ($-S-O_3H$) formation [212, 235]. The oxidation of the thiolate anion of the catalytic cysteine to sulfenic acid is generally reversible, while the further conversion to sulfinic and sulfonic acid is irreversible [212, 235, 236].

PTPs have protective mechanisms to prevent irreversible oxidation of their catalytic cysteines. In the classical PTPs, the sulfenic acid is converted to cyclic sulfonamide [212, 235, 237, 238] which is a five-membered ring that forms when the cysteine sulfur binds to the backbone nitrogen of the neighboring serine residue; the cyclic sulfenamide is not readily further oxidized and can be reduced by thiols, including reduced glutathione [212, 235] or reduced in vitro by reducing agents such as DTT or β -mercaptoethanol [235, 236, 239, 240]. The non-classical PTPs CDC25, PTEN, and LMW-PTPs use a different protective mechanism to prevent

their catalytic cysteines from being irreversibly oxidized: they can form intra-molecular disulfide bonds between the catalytic cysteine and an adjacent cysteine [212, 235, 241, 240]. The LMW-PTPs have the same protective mechanism of forming a disulfide bond between cysteine residues, but are unique in that the disulfide bond in the LMW-PTP is formed between the catalytic cysteine residue and an additional cysteine residue located within the active site [242], which in the human LMW-PTP are residues Cys12 and Cys17. These two cysteine residues in the active site are conserved in both prokaryotic and eukaryotic LMW-PTPs [242]. A mutant with a Cys17 to Ala17 change could not form an intramolecular disulfide bond and was unable to recover its activity after treatment with H_2O_2 followed by treatment with cellular reduced glutathione [242]. In the cell, the oxidation of LMW-PTP can be reversed by the glutaredoxin/glutathione/glutathione reductase/NADPH system [211, 228, 243]. In vitro the inactivation of the LMW-PTP by oxidation is reversible by treatment with reducing agents such as DTT [243]. This disulfide bond between the two cysteine residues in the active site protects the active site cysteine residue from irreversible oxidation and also allows for efficiency and speed in the redox regulation of the LMW-PTP activity [243, 244].

The Role of Reactive Oxygen Species and Redox Regulation of Protein Tyrosine Phosphatases

Reactive oxygen species (ROS) such as H_2O_2 or O_2^- are produced in cells following stimuli such as the binding of ligands such as cytokines and growth factors cellular receptors or exposure to ultraviolet light [211, 234, 235, 239]. For example, platelet-derived growth factor (PDGF) stimulation induced ROS levels in cells comparable to adding up to 1.0 mM H_2O_2 exogenously [234, 235, 239]; these PDGF-induced ROS levels are more than sufficient to oxidize active site cysteines of protein tyrosine phosphatases (PTPs) [235]. Growth factor signaling

depends on ROS production, since treatment of cells with catalase (an enzyme that converts H_2O_2 to O_2 and H_2O) abolishes growth factor signaling [235, 239, 245]. The inactivation of PTPs by growth factor-induced oxidation would explain why ROS are essential for growth factor signaling [212, 235]; they would help regulate the activity of PTPs via oxidation of the catalytic cysteines, which would render the PTPs inactive for a period until the catalytic cysteines were re-reduced, returning activity to the enzymes [211]. For example, the PDGF receptor has been found to be a substrate for the mammalian LMW-PTP (discussed below in the Identified Substrates of LMW-PTPs section). When cells are stimulated with PDGF, after only ten minutes about 80% of the LMW-PTP molecules become oxidized and their activity inhibited, but after 45 minutes, about 70% of the LMW-PTP molecules are reduced and regain their catalytic activity [228, 243]. Caselli *et al* (1995) [246] found that protein tyrosine phosphatases can also be inactivated by nitric oxide, which is produced from L-arginine in macrophages, endothelial cells and other cell types, and would have a similar function to ROS.

Modifications of the Conserved Cysteine and Arginine Residues Affect Catalysis

Experiments in which the catalytic cysteine is irreversibly modified with the alkylating agents iodoacetate or iodoacetamide show that the invariant cysteine residue in the phosphatase active is absolutely required for phosphatase activity; this residue performs the nucleophilic attack on the phosphorus atom in the phosphate group of the substrate, producing a covalent phosphoenzyme intermediate [225, 226, 247, 248, 249, 250, 251]. Replacement of the catalytic cysteine residue with a serine or alanine residue eliminates the ability of the PTP to form a phospho-enzyme intermediate [225, 231, 247] and results in an enzyme with no detectable catalytic activity when tested with *p*-nitrophenyl phosphate [251, 252]. Mutant PTPs in which the conserved arginine has been changed to an alanine [251], a lysine, or a methionine residue [250]

also have undetectable catalytic activity. Since cysteine to serine (Cys to Ser) mutants retain the ability to bind substrates, this allows their use in isolating and identifying PTP substrates in vivo [225]. A Cys to Ser mutation is replacing the negatively charged thiolate anion with a hydroxyl group with a neutral charge at physiological pH [225]. While the active site Cys to Ser mutant PTP binds with the same affinity as the wild-type PTP to ligands and substrate analogs [225, 229, 253], the lack of catalytic activity in this mutant may be enhanced due to structural or dynamic changes in the active site, and the active site residues and other associated residues may be more flexible in this mutant [225].

Low Molecular Weight Protein Tyrosine Phosphatases

LMW-PTP proteins are found in most organisms including Archaea, bacteria, and eukaryotes, suggesting they have an important role in cellular functions [217, 254]. In general, organisms seem to have one or two LMW-PTP genes. *Entamoeba histolytica* has two LMW-PTP genes (GenBank: XM_651267 and XM_648265), each which codes for a 157 aa protein (GenBank: XP_656359 and XP_653357). These two proteins are identical except for a single residue change at position 85 from an alanine to a valine. The nonpathogenic related species *Entamoeba dispar* and the reptile parasite *Entamoeba invadens* both have a single LMW-PTP gene. The plant *Arabidopsis thaliana* has only one LMW-PTP gene [255, 256], as does the rice plant *Oryza sativa* and the green alga *Chlamydomonas reinhardtii*; the black cottonwood tree *Populus trichocarpa* has two [256]. Two LMW-PTP proteins expressed from two different genes in the *primo* locus that are 50% identical to each other have been identified in the developing *Drosophila* pupal retina [257]. All mammals, including human beings, seem to have two active isoforms resulting from an alternately spliced transcript of a single gene, and these isoforms are usually coexpressed; their precise role, in terms of substrate specificity or functional regulation, is

not yet known [258]. Like other mammals, human beings have a single LMW-PTP gene [211, 217, 222], which gives rise to two active proteins, with either exon 3 or exon 4 being incorporated [211, 227, 259]. There are also two inactive forms: one in which both exons 3 and 4 are deleted [211, 227, 260] and one in which alternate splicing generates a frameshift [211, 261]. LMW-PTPs have no particular tissue-specific expression in mammals [211, 262].

LMW-PTPs have their active site close to the N-terminus of the protein; in all the other PTPs the active site is towards the C-terminus [219, 222]. The active site, or P loop, of LMW-PTPs has the characteristic conserved sequence CLGNICR, which conforms to the general phosphatase active site sequence CX₅R [203, 228]. The conserved aspartic acid is found in the DPYY loop which also contains two conserved tyrosines [212, 219]. The mammalian LMW-PTPs are 18 kDa enzymes that are widely expressed in various tissues [228, 244]. They were studied initially as polymorphic low molecular weight soluble acid phosphatases extracted from human red blood cells [263, 264] which had activity on the generic tyrosine phosphatase substrate *p*-nitrophenyl phosphate (*p*NPP) [203, 264, 265]. In humans LMW-PTP is genetically polymorphic [262, 263, 264]. In 1969 a LMW-PTP from mammalian tissue, bovine liver, was first purified to homogeneity [203, 266]. Since then, mammalian LMW-PTPs have been studied quite extensively, having been isolated as acid phosphatases from many tissues including human placenta [230, 262, 267], human liver [262], bovine liver [230, 262, 266], and rat brain [230, 268]. The first LMW-PTP cloned gene was cloned from a bovine heart cDNA library [230]. The two active human isoforms were cloned from a placental cDNA library using the cloned bovine LMW-PTP gene as a probe [262]. The *Saccharomyces cerevisiae* LMW-PTP was cloned by using an oligonucleotide probe derived from the sequence of the mammalian LMW-PTP active site [269, 270].

The Structure of Low Molecular Weight Protein Tyrosine Phosphatases

There have been many LMW-PTP proteins from diverse organisms crystallized and their structures determined, including those from *Homo sapiens* [223, 262, 271], *Bos taurus* [272, 273, 274], *Saccharomyces cerevisiae* [270], *Mycobacterium tuberculosis* [275], *Bacillus subtilis* [276], *Campylobacter jejuni* [277], and the bovine parasite *Tritrichomonas foetus* [278]. The LMW-PTP structures from different organisms all have the same basic structure: the enzyme is characterized by a central twisted four-stranded β -sheet surrounded by five α -helices [270, 275, 278]. The two β - α - β motifs form a Rossmann fold, and the active site (P loop) is in the first β - α - β motif, at the C-terminus of β 1 and the N-terminus of α 1 [270, 275]. The P-loop conformation is stabilized by a network of hydrogen bonding and with substrate, substrate analogs or inhibitors bound, is the most rigid, and ordered, part of the enzyme [270, 272, 273, 278]. The active site pocket is quite deep in all phosphotyrosine-specific phosphatase enzyme structures, about 9 Å, allowing specificity for phosphotyrosine substrates [224, 279, 280], but tyrosine-specific PTPs can not catalyze the removal of phosphate groups from phosphoserine and phosphothreonine, which have shorter side chains [203]. In contrast, the dual-specificity phosphatases have shallower active site pockets of about 6 Å deep that can accommodate phosphotyrosine, -serine, and -threonine [224, 280].

Substrate-Trapping Mutants

The Cys12 to Ser12 mutation in the mammalian LMW-PTP abolishes enzymatic activity, but still allows substrate binding [228, 250, 252, 281, 282], and acts as a dominant negative in vivo, since the catalytically inactive enzyme competes for substrate binding with the active endogenous LMW-PTP [228, 281, 282]. These mutants can be used in studies to trap possible substrates, either in vivo or in vitro, since they still bind substrate but can not carry out the

catalytic reaction. Different substrate-trapping mutants have been tried (reviewed in Blanchetot *et al* 2005) [283]: the mutations that have been generated in LMW-PTPs include: (i) mutation of the catalytic active site cysteine residue to a serine, (ii) mutation of the general acid/base aspartic acid residue to an alanine, and (iii) double mutants with both the Cys to Ser and Asp to Ala mutations [283].

Identified Substrates of LMW-PTPs in Mammalian Cells

The functions of LMW-PTPs have been most studied in mammalian cells. In mammalian cells, LMW-PTPs primarily regulate growth factor receptors, mainly by counteracting their signaling [228, 242] as well as influencing adhesion and cytoskeletal organization [219, 228]. Studies using NIH 3T3 murine fibroblast cells overexpressing the Cys12 to Ser12 substrate-trapping mutant show that activated platelet-derived growth factor receptor (PDGFr) is an LMW-PTP substrate [228, 282, 284], as is the insulin receptor [228, 281] fibroblast growth factor receptor [212, 228, 285], the Rho regulator p190RhoGAP [212, 286], β -catenin (which links cadherins to the actin cytoskeleton) [287], the Ephrin A2 receptor tyrosine kinase [288] and focal adhesion kinase p125FAK [212, 222, 228, 289]. Caveolin-1 was identified as a substrate in murine endothelial H-end cells; it is also tyrosine-phosphorylated *in vivo* by Src kinases [213, 258]. Ephrin B1 receptor tyrosine kinase was identified as a substrate using human renal microvascular endothelial cells (HRMEC) [212, 228, 290], and the transcription factor STAT5 (signal transducer and activator of transcription-5) was identified as a LMW-PTP substrate in human DAMI megakaryocytic cells [212, 222, 291, 292].

Identified Substrates of LMW-PTPs in Non-Mammalian Cells

Less work has been done with non-mammalian LMW-PTPs in identifying putative substrates. For example, in the photosynthetic cyanobacterium *Synechocystis* sp. PCC 6803, the hypothetical LMW-PTP protein Slr0328 was expressed recombinantly as a double substrate-trapping mutant with mutations to both the catalytic cysteine (Cys to Ser) and the general acid/base aspartic acid (Asp to Ala) attached to *N*-hydroxysuccin-imide (NHS)-activated Sepharose beads and incubated with clarified lysate [293]. Tryptic peptides were analyzed by mass spectrometry to identify putative substrate proteins, which included PsaD (photosystem I subunit II), CpcD (phycocyanin rod linker protein) and phycocyanin- α and - β subunits, thus the *Synechocystis* LMW-PTP appears to be involved in regulating photosynthesis-associated proteins [293]. In another example, the pathogen *Erwinia amylovora* (the causative agent of fireblight on rosaceous plants such as apple and pear trees) synthesizes the capsular exopolysaccharide amylovoran, which is necessary for pathogenicity [294]. Its LMW-PTP protein AmsI was found to have phosphatase activity, and overexpression of AmsI resulted in a strong reduction in the level of exopolysaccharide synthesis, so this LMW-PTP appears to be involved in exopolysaccharide biosynthesis in this organism [294]. In general, LMW-PTPs that have been identified in bacteria appear to be involved in the regulation of capsule polysaccharide synthesis [222]. The study of many non-mammalian LMW-PTPs is still at the phosphatase-activity testing stage with *p*NPP or other substrate analogs. For example, recombinant LMW-PTPs have been shown to have phosphatase activity by using wild-type and substrate-trapping mutants in *Staphylococcus aureus* [295], *Mycobacterium tuberculosis* [232], and *Campylobacter jejuni* [277]. In plants, expression of a LMW-PTP has been found to be induced during germination of lentil (*Lens esculenta*) seeds [296]. Unusually, LMW-PTP proteins in the kinetoplastid species *Trypanosoma brucei*, *T. cruzi*, *Leishmania major*, *L. donovani*, and *L. chagasi*, lack a catalytic

cysteine residue in the active site so these enzymes may be functionally inactive [297], so the dephosphorylative role of the LMW-PTP in these species is therefore likely played by a different protein tyrosine phosphatase.

Regulation of LMW-PTP Activity in Mammalian Cells

The regulation of LMW-PTP activity has been most studied in mammalian cells. LMW-PTPs have two adjacent tyrosine residues that are found in the conserved DPYY loop [212, 219]. This loop also contains the conserved aspartic acid residue which is catalytically critical in phosphate removal from the substrate [232, 271]. These tyrosines can be phosphorylated and dephosphorylated to regulate LMW-PTP activity [211, 228, 271]. In the human and mouse LMW-PTP these are residues Tyr131 and Tyr132, and can be phosphorylated by the protein tyrosine kinases Src, Lck, and Fyn [211, 298, 299]. Src phosphorylates both residues, but with different effects: Tyr131 phosphorylation induces a 25-fold increase in LMW-PTP activity (as measured with *p*NPP), while phosphorylation at Tyr132 has no effect on enzyme activity, but instead recruits Grb2, (growth factor receptor bound protein 2) an adaptor protein which binds to phosphorylated tyrosines via its SH2 (Src homology 2) domain, to the enzyme [211, 228, 298]. In mammalian LMW-PTPs, the Tyr132 residue is followed by the consensus sequence G-N-D, which conforms to the consensus Grb2 SH2 recognition motif X-N-X immediately following the phosphotyrosine residue [228, 300]. Since Tyr132 is close to the active site, the interaction of Grb2 or other SH2-domain-containing proteins with the phosphorylated Tyr132 could limit access to the active site, leading to selection of substrates by size, or even functional enzyme inactivation, if the adaptor protein occludes substrate access to the active site [211, 298]. The other Src-family tyrosine kinases Lck and Fyn can also phosphorylate the LMW-PTP at Tyr131, increasing enzymatic activity twofold, and to a lesser extent, phosphorylate Tyr132 [211, 228,

299]. The LMW-PTP may also form dimers as a regulatory mechanism; it forms dimers both in solution and in crystal structures, where Tyr131 and Tyr132 on one monomer insert into the active site of the second, with Tyr132 forming a hydrogen bond with Asp129 (which functions as the general acid/base in catalysis) thus occluding both the active site and the catalytic residues, preventing both binding of substrate and phosphorylation at Tyr131 and Tyr132 [219, 274].

There appear to be two pools of LMW-PTP in mammalian cells: a non-phosphorylated cytosolic/plasma membrane form associated with the PDGFr, and a Src-phosphorylated form, having higher catalytic activity, associated with p190RhoGap and the cytoskeleton (the Triton X-100 insoluble fraction) [211, 219, 301]. The LMW-PTPs in the cytosolic/plasma membrane-associated pool are phosphorylated and activated by Src, and bind and dephosphorylate activated PDGFr (which has autophosphorylated upon treatment with platelet-derived growth factor), while the cytoskeletal-associated pool acts on substrates that are tyrosine-phosphorylated upon PDGFr treatment [211, 219, 301]. The LMW-PTP molecules translocate from the cytosolic pool to the cytoskeletal-associated pool following stimulation of the cellular PDGF receptors with platelet-derived growth factor [302].

LMW-PTP Overexpression and Cancer

Overexpression of LMW-PTP can lead to unwanted dephosphorylation of its usual substrates, which can result in dysregulation of downstream signaling pathways. LMW-PTP has been found to be overexpressed in different human and other mammalian tumor types, and thus can be considered an oncogene [303]. Examples of overexpressed LMW-PTP involved in cancer include: colonic tumors which were induced in rats by DMH (1,2-dimethylhydrazine) had a threefold increase in the expression of LMW-PTP in colonic carcinomas that were formed in the proximal (right) part of the colon rather than in the distal (left colon) as compared with the

normal mucosa, in agreement with observations reported for human colon cancer [304]. The colonic adenomas (benign tumors), as opposed to the carcinomas, did not have a statistically significant level of LMW-PTP upregulation, suggesting that the overexpression of LMW-PTP is a phenomenon associated with the onset of malignancy, and that right and left colon cancers are different entities with different molecular and cellular characteristics [303, 304, 305]. In surgical samples of breast, colon and lung cancers, as well as neuroblastomas, the expression levels of LMW-PTP mRNA was evaluated using real-time PCR analysis, along with adjacent unaffected tissue samples as controls (except for the neuroblastoma samples which lacked unaffected tissue sample controls) [303]. LMW-PTP mRNA levels were significantly increased in breast and colon cancer samples, and also in neuroblastoma samples, but not in lung carcinoma [303]. EphA2 receptor tyrosine kinase is a substrate for LMW-PTP, and the oncogenic activities of LMW-PTP result from altered EphA2 expression and function [288, 306]. The EphA2 receptor tyrosine kinase is overexpressed in a large number of different human cancers including breast, prostate, colon, and lung carcinomas as well as metastatic melanomas, and this EphA2 is not tyrosine phosphorylated [288, 306]. Normally, tyrosine phosphorylation of EphA2 causes it to interact with downstream signaling components that function to negatively regulate cell growth and invasiveness; in contrast, unphosphorylated EphA2 appears to adopt a different subcellular localization, interacts with different substrates [288, 306], and functions as a powerful oncoprotein, whereas restoration of EphA2 phosphotyrosine content is sufficient to reverse the oncogenic potential of EphA2 [288, 306].

When wild-type and substrate-trapping mutant LMW-PTP-overexpressing NIH 3T3 cells (a mouse embryonic fibroblast cell line) were subcutaneously inoculated into nude mice, the wild-type overexpressing cells induced formation of sarcomas, while the mutant overexpressers did not [307]. The PDGF receptor and β -catenin were not the preferred wild-type LMW-PTP

substrates in these tumors, but the EphA2 receptor phosphorylation level due to wild-type LMW-PTP overexpression was dramatically reduced, and E-cadherin expression levels were increased [307]. In normal epithelial cells, E-cadherin mediates calcium-dependent cell-cell adhesions through interactions with E-cadherin on neighboring cells. EphA2 is tyrosine-phosphorylated upon E-cadherin-mediated adhesion and localizes to sites of cell-cell contact, while in breast cancer cells that lack E-cadherin, the phosphotyrosine content of EphA2 is decreased, and EphA2 is localized in membrane ruffles [308].

Why study this *Entamoeba histolytica* LMW-PTP?

The parasite *E. histolytica* has 20 genes annotated as protein tyrosine phosphatases or putative tyrosine phosphatases, according to the database at AmoebaDB, the amoeba genomics database [309, 310], far fewer than the 107 protein tyrosine phosphatases that the human genome contains [222, 256]. Two of these twenty are the low molecular weight protein tyrosine phosphatase genes in *E. histolytica*, XP_656359 and XP_653357, which are identical except for one amino acid change at position 85 from an alanine in XP_656359 to a valine in XP_653357. Only one was cloned (XM_651267) and had its protein (XP_656359) used for this study.

This *Entamoeba histolytica* enzyme seemed to be a suitable choice for study, since we are interested in virulence factors in *E. histolytica*. In mammalian cells, LMW-PTPs play roles in the control of cell proliferation, motility, and adhesion through dephosphorylation of such substrates as growth factor receptors [212, 228, 242, 281, 282, 284, 285, 307] and cytoskeleton-associated proteins [212, 219, 222, 228, 289, 307] and when overexpressed can cause tumors and thus is considered an oncogene [303, 307]. We were interested in the function of the LMW-PTP in *E. histolytica*, as the LMW-PTP might play a role in virulence, since proliferation, motility, and adhesion are necessary attributes for *E. histolytica* to cause disease; trophozoites need to bind

to the intestinal epithelium (or other tissue or intestinal mucus), need to translocate once they have destroyed the tissue (or ingested all the bacteria in that region) and need to proliferate. The questions that needed to be answered include what the substrates are for the LMW-PTP, if those substrates are similar to those of other LMW-PTPs, and how mutating the LMW-PTP could affect *E. histolytica* in its ability to cause disease: are such factors as motility, phagocytosis, or cell proliferation affected? Other questions we had were: what is the cellular location of the LMW-PTP? What other interaction partners does the LMW-PTP have, such as adaptor proteins and kinases that phosphorylate the LMW-PTP, and how are these involved in the overall cellular signal transduction pathways?

Materials and Methods

Alignment of LMW-PTP Protein Sequences

The sequence of the wild-type *E. histolytica* LMW-PTP protein XP_656359 was input into BLAST [311] in order to identify related LMW-PTP proteins from other species. BLAST [311] and LALIGN [312] were used to align the protein sequences, except for the *Entamoeba invadens* LMW-PTP protein identification and alignment; the ameba database [309] hosted by EuPathDB, Eukaryotic Pathogen Database Resources [313] was used. A phylogenetic tree for the relative distances between the LMW-PTPs were created by using the tools at the Phylogeny.fr web site [314] or mirror site [315].

Cloning of Untagged Wild-Type and Mutant LMW-PTPs

The wild-type and Cys7 to Ser7 active site LMW-PTP mutant LMW-PTPs were initially cloned via PCR amplification without epitope tags added. *E. histolytica* genomic DNA

was used as a template to clone the wild-type and mutant LMW-PTPs via PCR. Oligonucleotide primers were designed so the entire gene would be amplified, and to introduce the mutation in the active site from a cysteine to a serine at residue number 7 in the LMW-PTP protein sequence. To clone the untagged wild-type LMW-PTP via PCR, the forward primer *Bam* H1TyrPhosF and the reverse primer *Sal*ITyrPhosR were used (Table 3-5). To clone the untagged active site mutant LMW-PTP, the forward primer *Bam* H1TyrPhosFmut (Table 3-5) and the same reverse primer as used for the wild-type was used. The PCR conditions were per 100 μ l reaction volume: 10 μ l 10X SAHARA DNA Polymerase Buffer, 6 μ l 50 mM MgCl₂, 5 μ l 40 mM dNTP mix, 250 ng *E. histolytica* genomic DNA, 2 μ l each of 100 μ M forward and reverse oligo, 2 μ l SAHARA™ DNA polymerase (Bioline USA, Tauton, MA, USA), and sufficient water added to complete a 100 μ l volume. The cycling conditions were: 95°C for 10 minutes, followed by 35 cycles of: 95°C for 1 minute, 60°C for 1 minute, and 68°C for 3 minutes 30 seconds, with a final extension at 68°C for 10 minutes. The resulting PCR products, and also plasmid pBluescript II KS(+) (Agilent Technologies, Santa Clara, CA, USA), were digested with *Bam*H1 and *Sal*I, and subjected to agarose gel electrophoresis. DNA bands were excised from the gel, gel slices were purified with the QIAquick Gel Extraction Kit (Qiagen Inc.-USA, Valencia, CA, USA), and PCR products were ligated into pBluescript II KS(+) (Agilent Technologies, Santa Clara, CA, USA) with High Concentration T4 DNA Ligase (New England Biolabs, Ipswich, MA, USA). Ligations were transformed into chemically competent *E. coli*, colonies screened for inserts of the correct size, and were cultured overnight in liquid broth with 100 μ g/ml ampicillin added for selection. The plasmid DNA was extracted with the QIAprep Spin Miniprep Kit (Qiagen Inc.-USA, Valencia, CA, USA) and submitted for sequence analysis. Only one of the two LMW-PTP genes, GenBank: XM_651267, was cloned. DNA from sequence-verified active site Cys7 to Ser7 mutant LMW-PTP in pBluescript II KS(+) DNA was minipreped and sent to SSGCID for

recombinant protein expression and purification; they expressed and prepared the recombinant wild type and mutant LMW-PTP proteins for us as described in the Recombinant Protein Expression at SSGID section below.

Recombinant Protein Expression at SSGCID

The Seattle Structural Genomics Center for Infectious Disease (SSGCID) is a collaboration between four institutions based in the state of Washington: Seattle BioMed, Emerald BioStructures, the University of Washington, and the Pacific Northwest National Laboratory [316, 317] and is funded by NIAID (the National Institute of Allergy and Infectious Diseases) [316, 318]. The aim of SSGCID is to solve the three-dimensional structures of selected proteins from pathogens [319] that are in the NIAID Category A, B, and C priority lists, that are the cause of emerging and re-emerging infectious diseases, with proteins chosen that are potential vaccine or drug targets, or have an important biological function in the pathogen [318]. The expected rate of solving these structures is between 75 and 100 per year [319]. The University of Washington Protein Production Group, as part of the SSGCID, uses a non-automated approach to high-throughput screening of proteins, using auto-induction methods for the expression of recombinant proteins in *Escherichia coli*, immobilized metal-affinity chromatography (IMAC) for the purification of hexahistidine-tagged proteins and SDS-PAGE analyses to visually evaluate expression and recovery levels after IMAC purification [320]. Screening is performed in a 96-well format [320].

Protein genes were initially cloned by PCR amplification using either genomic DNA or cDNA as a template, and PCR primers contained a ligation-independent cloning (LIC) sequence at their 5' ends complementary to the LIC sequence in the plasmid vector [319, 320]. PCR products were cloned into the AVA0421 expression vector [321], a derivative of pET14b

containing a T7 promoter followed by an N-terminal six-histidine nickel-affinity tag (6xHis tag) followed by the modified human rhinovirus 3C protease (HRV-3C) recognition sequence Leu-Glu-Ala-Gln-Thr-Gln-Gly-Pro, so that the 6xHis tag can be later cleaved from the protein; the plasmid also confers resistance to ampicillin and carbenicillin [319, 320, 321]. Sequence-verified recombinant plasmids were transformed into the *E. coli* Rosetta Oxford strain [BL21*(DE3)-R3-pRARE2] cells for protein expression testing before scaling up to large-scale expression [319, 320]. Individual colonies were grown overnight with shaking in non-inducing medium at ambient temperature (about 20°C-22°C); an aliquot was removed from each culture and used to inoculate auto-induction medium [320, 322]. Inoculated cultures were grown at 20°C in a 96 well deep-well block on a plate shaker until the OD₆₀₀ readings were at least 0.6 [320]. Cells were harvested by centrifugation, the supernatant was removed and pellets stored at -80°C. Cell pellets were thawed and suspended in lysis buffer containing 20 mM HEPES pH 7.0, 500 mM NaCl, 5% glycerol, 0.5% CHAPS, 30 mM imidazole, 10 mM MgCl₂, 400 µg/ml lysozyme and 3 units Benzonase nuclease per ml, and were incubated at room temperature with moderate shaking for one hour [320]. A crude lysate sample was removed for later SDS-PAGE analysis; the remaining lysate was clarified by centrifugation and transferred into a 96-well block pre-loaded with pre-equilibrated Ni²⁺ Sepharose beads [320]. The protein/resin mixture was incubated with shaking at 4°C for 15 minutes, beads were washed three times with 1 ml wash buffer (20 mM HEPES pH 7.0, 500 mM NaCl, 5% glycerol, 30 mM imidazole) using a vacuum apparatus, and bound proteins were eluted with 100 µl elution buffer, same as the wash buffer but with 500 mM imidazole added [320]. An eluate aliquot was mixed with SDS-PAGE sample buffer, these and the crude lysate samples were subjected to SDS-PAGE, and gels were stained and visually scored for protein expression and solubility [320]. Target proteins that scored low to

high for total expression in the crude lysate, and medium to high for IMAC (immobilized metal-affinity chromatography) recovery continued to large-scale expression [319].

For large-scale protein expression, 3 ml of overnight culture for each sample was used to inoculate two liters of sterile auto-induction medium [322] in Pyrex bottles with the appropriate antibiotic and 400 μ l antifoam added to the medium [319]. Cultures were grown for ~24 hours at 25°C; the temperature was then reduced to 15°C for ~72 hours. Bacterial cells were harvested by centrifugation and the cell pellet (~25 grams) was flash-frozen in liquid nitrogen and stored at -80°C until lysed [319]. Frozen cell pellets were suspended in lysis buffer (25 mM HEPES, 500 mM NaCl, 5% glycerol, 30 mM imidazole, 0.025% sodium azide, 0.5% CHAPS, 10 mM MgCl₂, 1 mM TCEP, 250 μ g/ml AEBSF, 0.05 μ g/ml lysozyme pH 7.0) and sonicated on ice for 30 minutes in 15 second intervals alternating with 15 seconds of rest [319]. Sonicated lysate was incubated with 20 μ l Benzonase nuclease (25 units/ml) at room temperature for 45 minutes, a crude lysate sample was taken for SDS-PAGE analysis, lysate was clarified by centrifugation at 29 774 x g for 75 min at 4°C, and a soluble fraction sample was collected for SDS-PAGE analysis. Immobilized metal-affinity chromatography (IMAC) was used to remove the majority of the native *E. coli* proteins from the clarified lysate with HisTrap FF 5 ml columns equilibrated with wash buffer (25 mM HEPES, 500 mM NaCl, 5% glycerol, 30 mM imidazole, 0.025% sodium azide, 1 mM TCEP pH 7.0). A sample of the flowthrough was saved for analysis. Unbound protein was removed from the column by washing the column with 20 column volumes of wash buffer [319]. The His-tagged protein and any other nickel-binding proteins were then eluted with seven column volumes of elution buffer (25 mM HEPES, 500 mM NaCl, 5% glycerol, 1 mM TCEP, 250 mM imidazole and 0.025% azide pH 7.0) and collected in 3 ml fractions, and the OD₂₈₀ absorbance chromatogram was used to determine which fractions to pool [319]. The 6xHis tag was cleaved from the protein using 3C protease added to the protein

solution at a weight ratio of 1:50, and the protein solution dialyzed for 18 hours at 4°C in dialysis buffer (25 mM HEPES, 500 mM NaCl, 5% glycerol, 1 mM TCEP and 0.025% azide pH 7.5) [319]. To remove uncleaved protein, the 6xHis-tagged 3C protease, the 6xHis tag peptide, and Ni-binding *E. coli* contaminant proteins, the protein solution was loaded onto a gravity-flow column packed with 2.5 or 5 ml pre-equilibrated nickel-Sepharose, and the purified-protein-containing flowthrough was collected; two column volumes of wash buffer were used to wash the column, and this was also collected [319]. The remaining contaminating nickel-bound proteins (ideally, the 3C protease, uncleaved protein, and non-His-tagged *E. coli* protein contaminants) were eluted from the column with four column volumes of elution buffer. All flowthroughs, washes, and elutions of contaminants were analyzed by SDS-PAGE and protein concentrations were measured in these samples [319]. The cleaved target protein was concentrated using Amicon Ultra-15 Centrifugal Filter Units (Millipore, Billerica, MA, USA) to a volume of 10-15 ml and then subjected to size-exclusion chromatography (SEC) as a final purification step [319]. The protein was loaded onto a HiLoad 26/60 Superdex 75 preparative-grade column (GE Healthcare, Piscataway, New Jersey, USA) that had been equilibrated in SEC buffer (25 mM HEPES, 500 mM NaCl, 5% glycerol, 2 mM DTT, 0.025% azide pH 7.0) and the eluate was collected in 5 ml fractions [319]. The molecular weight of the eluted protein was estimated based on the elution volume and a standard calibration curve for the column [319]. The size-exclusion chromatography (SEC) fractions and in-process samples were analyzed by SDS-PAGE to confirm the success of purification and determine which SEC fractions to pool for final concentration [319]. After the appropriate SEC fractions were pooled, the protein was concentrated with Amicon Ultra-15 Centrifugal Filter Units (Millipore, Billerica, MA, USA) to 20 to 30 mg/ml [319]. Concentrated protein was flash-frozen in flexible eight-well strip tubes in 100-200 µl aliquots using liquid nitrogen, and stored at -80°C until needed [319].

***E. histolytica* Constructs for LMW-PTP Overexpression**

Genes for overexpression with N-terminally or C-terminally added epitope tags were cloned into the amebic expression vector pGIR209 [177, 179] conferring resistance to geneticin (G418) and to ampicillin in *E. coli* (Figure 3-18). This vector can function to constitutively express protein when transfected by itself into trophozoites. pGIR209 is also part of the two-vector tetracycline-inducible system with pGIR308 [177, 179]. pGIR308 confers resistance to hygromycin B in *E. histolytica* and to ampicillin in *E. coli*, and expresses the tetracycline repressor protein [177, 179]. The tetracycline repressor protein binds the tetracycline operator sequence at the 3' end of the *hgl5* promoter (Figure 3-18) when tetracycline is not present, and thus no transcription occurs; the addition of tetracycline will cause the repressor to release from the tetracycline operator, allowing gene transcription. Correct constructs in cloning vector pBluescript II KS(+) generated as described in the Cloning of Untagged Wild-Type and Mutant LMW-PTPs section, were digested with *Bam* H1 and *Sal* I. The amebic expression vector pGIR209 [177, 179] was digested with *Bgl* II and *Sal* I, and the digested construct and pGIR209 DNA was purified, ligated together, screened, and sequenced as described above. The oligos pGIR209seqF, 5'-CTTGGTTCTATTTTTATACTCTCCTCTTATTG-3' and pGIR209seqR, 5'-GAATATCTACAAATAACTTATTAATTCTTATTTGATTAAGGTG-3', were used to sequence the epitope-tagged constructs to confirm the tags and the rest of the cloned genes were correct and in frame. A diagram of the process of cloning the epitope-tagged LMW-PTP genes into pGIR209 is found in Figure 3-18. Correct plasmids were isolated using the Qiagen HiSpeed Maxiprep Kit (Qiagen Inc.-USA, Valencia, CA, USA), and transformed into HM1:IMSS strain trophozoites as described in the Transfection of Trophozoites section.

Testing Recombinant Wild-Type and Mutant LMW-PTPs for Phosphatase Activity

The purified recombinant wild-type protein prepared at SSGCID was tested for phosphatase activity using the SensoLyte *p*NPP Colorimetric Protein Phosphatase Assay Kit (AnaSpec, Fremont, CA, USA) as per the manufacturer's instructions. *p*NPP, *p*-nitrophenyl phosphate, is a colorimetric generic phosphatase substrate which turns yellow when dephosphorylated [265]. Purified LMW-PTP was diluted in the supplied Assay Buffer, with 50 μ l diluted LMW-PTP (10-200 ng purified protein per sample) aliquotted per well in a 96-well plate. 50 μ l *p*NPP reaction solution, containing 10 mM *p*NPP and 3 mM DTT added to the supplied Assay Buffer, was added per well with diluted LMW-PTP, to give a final concentration of 5 mM *p*NPP and 1.5 mM DTT per 100 μ l per well. All samples were done in quadruplicate, and wells containing no LMW-PTP were included as controls. Absorbance was measured at 405 nm over time. To test phosphatase inhibition of the wild-type LMW-PTP, HALT™ Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific, Waltham, MA, USA) was added to one set of samples at 1X. Values were plotted as nanograms of recombinant wild-type LMW-PTP vs the absorbance at 405 nm for the 60 minute time point using Excel. Test for phosphatase activity of the recombinant mutant LMW-PTP protein as compared with the recombinant wild-type LMW-PTP: the recombinant mutant LMW-PTP gene was cloned by us and the protein was prepared at SSGCID (see the Cloning of Untagged Wild-Type and Mutant LMW-PTPs section). The reactions were set up as described above, except a set amount of 100 ng of each recombinant protein (mutant and wild-type) was used. The reaction was allowed to proceed for two hours, with absorbance measured at 405 nm every five minutes, and plotted as A_{405} vs time (minutes) using Excel. The slope of the trendline and the R^2 value was also calculated using Excel, and these values are displayed on the graph.

Culture of Trophozoites

E. histolytica strain HM1:IMSS trophozoites were grown axenically in TYI-S-33 (trypticase-yeast extract-iron-serum) (TYI) medium supplemented with 1X Diamond's vitamins (SAFC Biosciences, Lenexa, KS, USA), 15% heat-inactivated bovine serum (Gemini Bio-Products, West Sacramento, CA), 100 U of penicillin/ml and 100 mg streptomycin sulfate/ml (Gibco/Invitrogen, Carlsbad, CA, USA), at 37°C in T-25 tissue culture flasks [175] in a volume of 50 ml.

Transfection of Trophozoites

Plasmid DNA was prepared for each construct using the HiSpeed Qiagen Maxi Kit (Qiagen, Valencia, CA, USA). Plasmid DNA was concentrated by ethanol precipitation overnight at -20°C, and was resuspended to a concentration of at least one $\mu\text{g}/\mu\text{l}$ in sterile 10 mM Tris pH 7.5. Alternately, 30 μg plasmid DNA was transferred into 2 ml sterile microcentrifuge tubes, precipitated with ethanol overnight at -20°C, and the plasmid DNA pelleted and dried. Medium 199 (M199) (Gibco BRL/Invitrogen, Carlsbad, CA, USA) was supplemented with 5.7 mM cysteine, 25 mM HEPES, and 0.6 mM ascorbic acid [72, 176], adjusted to between pH 6.8 and 7.0 and filter-sterilized. 20 to 30 μg plasmid DNA was diluted with supplemented M199 medium (M199S) in sterile 2-ml microcentrifuge tubes to a total volume of 200 μl ; if the plasmid DNA were a dry pellet, it was resuspended in 200 μl M199S. 30 μl of Attractene transfection reagent (Qiagen, Valencia, CA, USA) was added into the 200 μl of plasmid DNA in M199S, tubes were mixed by gently flicking, and were incubated at room temperature to allow transfection-complex formation as per the manufacturer's instructions. Heat-inactivated bovine serum was added to the remaining M199S to a final concentration of 15%. Amebae were harvested by tapping the tissue culture flasks on a benchtop, the contents were transferred to 50 ml conical tubes, tubes were

centrifuged at $200 \times g$ for 5 min at 4°C , and suspended in M199S with serum to 2.5×10^5 amebae/ml. Each tube containing transfection complexes was filled with the suspended trophozoites (1.8 to 1.9 ml), the tube contents were gently mixed by inversion, and the tubes were incubated horizontally for 3 hours at 37°C . Tube contents were added to warm TYI in 25 cm^2 tissue culture flasks, and incubated overnight at 37°C . $15 \mu\text{g/ml}$ hygromycin (Invitrogen, Carlsbad, CA, USA) or $6 \mu\text{g/ml}$ G418 (geneticin) (Gibco/Invitrogen, Carlsbad, CA, USA) was added for selection after the overnight incubation [177], with both hygromycin and G418 added for double transfectants. Three to five days after transfection, 25 ml of the TYI was removed to a new 25 cm^2 tissue culture flask to recover any transfected cells that had become unattached (they often will settle back down and reattach to the flask surface), and 25 ml fresh TYI with hygromycin and/or G418 was added to each of the flasks. Only half of the medium was changed every 2-3 days until transfectants were apparent and reproducing, between one and two weeks after transfection.

Western Immunoblotting

All trophozoites were cultivated in 25 cm^2 tissue culture flasks as described in the Culture of Trophozoites section. Transfected trophozoites were selected and maintained with $15 \mu\text{g/ml}$ of hygromycin and/or $6 \mu\text{g/ml}$ G418. For transfectants used in experiments, the levels of selection were increased to $30 \mu\text{g/ml}$ of hygromycin and/or $12 \mu\text{g/ml}$ G418 and maintained at that level for 48 hours before harvesting. Transfectants used in experiments were induced with either $1 \mu\text{g/ml}$ or $5 \mu\text{g/ml}$ tetracycline added to the culture medium 24 hours before harvesting [177, 179], with the exception of samples used for confocal imaging, which were induced with $10 \mu\text{g/ml}$ tetracycline. Trophozoites were harvested by either by tapping the culture flasks on the benchtop

or by icing the culture flasks for 15 minutes. The contents of the flasks were transferred to 50 ml conical tubes, counted using a hemacytometer, centrifuged at 200 x g for 5 minutes, and cell pellets were resuspended in ice cold lysis buffer. Several different lysis buffers were used in experiments, described under Lysis Buffers below. For experiments not requiring phosphatase inhibitors, regular lysis buffer without added phosphatase inhibitors was used. Lysis buffer was added to cell pellets such that there generally would be 2×10^6 – 1×10^7 amebae/ml in samples for immunoblotting. Samples lysed by sonication were lysed by pulsing twice for 10 seconds each with a 10 second rest on ice between pulses, using a Branson 150D sonifier set to 5 (Branson Ultrasonics Corporation, Danbury, CT, USA). If the amount of protein in the lysate needed quantification, it was measured using the Bio-Rad protein assay kit (Bradford) (Bio-Rad, Hercules, CA, USA). Samples were subjected to SDS-PAGE using 13%, 14% or 15% Tris-glycine gels in running buffer (25 mM Tris, 192 mM glycine, 1% SDS, pH 8.3). Gels were run under reducing conditions for all samples. After electrophoresis, sample proteins were transferred to Immobilon-P PVDF transfer membrane with pore size 0.45 μ m (Millipore, Billerica, MA, USA) using the Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad, Hercules, CA, USA) for 100 volts for 1 hour 45 minutes in transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol). After transfer, blots were blocked in either 5% nonfat dry milk or 3.5% BSA (bovine serum albumin) in TBST (43 mM Tris-HCl, 7 mM Tris base, 200 mM NaCl, 0.1% Tween 20, pH 7.35), and incubated with rabbit sera that we had made that recognized the LMW-PTP (please see the Immunoprecipitation of LMW-PTP with Rabbit Sera section on the next page), or commercially available antibodies: 4G10 Platinum antibody against phosphotyrosine [Millipore, Billerica, MA, USA], or the epitope tags V5 (Invitrogen, Carlsbad, CA, USA), FLAG (Sigma Life Sciences, Sigma-Aldrich, St. Louis, MO, USA), or cMyc (Santa Cruz Biotechnology, Santa Cruz, CA, USA). Blots were incubated with primary antibody either overnight at 4°C or at room

temperature for three hours. Secondary antibodies were from Sigma (Sigma-Aldrich, St. Louis, MO, USA); either goat anti-mouse Fc-specific peroxidase-conjugated secondary or goat anti-rabbit whole molecule peroxidase-conjugated secondary antibody was used. The ECL kit from Roche (Roche Applied Science, Indianapolis, IN, USA) was used to treat membranes after secondary antibody incubation, bands were visualized on film, and film images were electronically scanned.

Immunoprecipitation of LMW-PTP with Rabbit Sera

To generate an antibody that would recognize the untagged LMW-PTP in non-transfected HM1:IMSS strain amebae, two rabbits, code-named UVA417 and UVA418, were inoculated with the wild-type purified recombinant LMW-PTP to generate specific antibodies (Cocalico Biologicals, Inc., Reamstown, PA, USA). The two immune test bleeds from each rabbit, collected 1/11/2010 and 2/1/2010, were assayed to check that they recognized the native LMW-PTP from *E. histolytica* HM1:IMSS strain amebae, as compared with the pre-immune sera collected before immunization, and if the sera could be used to immunoprecipitate the native LMW-PTP from amebic lysate. 1 μ l (approximately 20 μ g total serum protein) of pre-immune serum or immune test bleed serum from each rabbit was diluted in a 200 μ l volume of PBS (150 mM NaCl, 5.6 mM Na₂HPO₄, 1 mM KH₂PO₄) with 0.05% Tween 20 added (PBS-Tween). 50 μ l of the magnetic Dynabeads Protein A (Invitrogen, Carlsbad, CA, USA) for each sample was aliquotted into microfuge tubes, and the supernatant was drawn off. The diluted pre-immune or immune serum in PBS-Tween was added to each microfuge tube with Dynabeads. To bind the antibodies in the sera to the beads, the samples were placed on a nutator to keep the beads suspended in buffer during the 15 minute incubation at room temperature. Beads were then washed with PBS-Tween to remove unbound antibody. Antibody was not chemically crosslinked

to the beads. Amebae were harvested as described above in western immunoblotting, counted using a hemacytometer, and were resuspended in regular lysis buffer (5X SigmaFast protease inhibitor cocktail (Sigma-Aldrich, St. Louis, MO, USA) 150 mM NaCl, 50 mM Tris, 1% NP-40 or 1% Triton X-100). Lysate from 6.75×10^6 trophozoites in 400 μ l volumes was added to each tube of beads with antibody attached, and tubes were incubated on the nutator for 25 minutes at room temperature to allow binding of bead-attached antibodies to target proteins. Samples were then washed three times each with 200 μ l PBS, the beads were resuspended in 100 μ l PBS, transferred to a new tube, the supernatant was removed, and the beads resuspended in SDS-PAGE sample buffer. For SDS-PAGE, beads in sample buffer were heated to 95°C for 5 minutes, Immunoblots were probed with either pre-immune or immune serum from the rabbit (at a 1:40 dilution) that was not used for that sample immunoprecipitation. 100 ng of the purified recombinant wild-type LMW-PTP protein was also included as a positive control.

Lysis Buffers

Regular lysis buffer did not contain phosphatase inhibitors, and consisted of 5 \times SigmaFast protease inhibitor cocktail (Sigma-Aldrich, St. Louis, MO, USA), 150 mM NaCl, 50 mM Tris, and either 1% Nonidet P-40 or 1% Triton X-100. Inhibitor lysis buffer with EDTA consisted of 5X SigmaFast protease inhibitor cocktail (Sigma-Aldrich, St. Louis, MO, USA), 150 mM NaCl, 50 mM Tris, either 1% Nonidet P-40 or 1% Triton X-100, 10% glycerol, 1 mM sodium orthovanadate (preparation of its stock solution is described in the next section), 5 mM iodoacetamide, and 100 μ M NaF. Inhibitor lysis buffer with no EDTA was of the same composition as inhibitor lysis buffer with EDTA, but contained a different protease inhibitor cocktail, 5X cOmplete EDTA-free Protease Inhibitor Cocktail (Roche Applied Science, Roche Diagnostics, Mannheim, Germany or Indianapolis, IN, USA) instead of the 5X SigmaFast

protease inhibitor cocktail. SigmaFast protease inhibitor cocktail tablets (Sigma-Aldrich, St. Louis, MO, USA) were used at a final concentration of 5X, which contains: 5 mM EDTA (ethylenediaminetetraacetic acid), which inhibits metalloproteases; 10 mM AEBSF [4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride], which inhibits serine proteases such as trypsin, chymotrypsin, plasmin, kallikrein, and thrombin; 650 μ M bestatin hydrochloride, which inhibits aminopeptidases such as leucine aminopeptidase and alanyl aminopeptidase; 70 μ M E-64 [N-(trans-Epoxy succinyl)-L-leucine 4-guanidinobutylamide], which inhibits cysteine proteases such as calpain, papain, cathepsin B, and cathepsin L; 5 μ M leupeptin, which inhibits serine and cysteine proteases such as trypsin, plasmin, trypsinogen, urokinase, and kallikrein; and 1.5 μ M aprotinin, which inhibits serine proteases such as chymotrypsin, trypsin, and elastase [323]. The EDTA inhibits metalloproteases [323] and serine-threonine phosphatases by chelating the metal ion cofactors these enzymes require as a cofactor [324, 325, 326, 327]. Protein tyrosine phosphatases do not require a metal ion for activity [326] and should not be affected by the presence of EDTA. cComplete EDTA-free Protease Inhibitor Cocktail Tablets (Roche Applied Science, Roche Diagnostics, Mannheim, Germany or Indianapolis, IN, USA) were dissolved in water and also used at a final concentration of 5X; according to a company representative, the composition of the cComplete tablets is confidential and proprietary.

Preparation of Sodium Orthovanadate Stock Solution

Sodium orthovanadate inhibits ATPase, alkaline phosphatase, and tyrosine phosphatases [328]. Sodium orthovanadate in solution at neutral pH polymerizes into yellow-orange decavanadate; to convert the polymeric solution into the colorless monomeric solution, it must be adjusted to pH 10 and either diluted or boiled [329] to “activate” it. Sodium orthovanadate powder (Na_3VO_4) (MP Biomedicals, LLC, Solon, OH, or Sigma-Aldrich, St. Louis, MO, USA)

was dissolved in water at a concentration of either 100 mM or 200 mM, and its pH was adjusted to 10 with HCl. The solution was then boiled until its color turned from orange to colorless [329, 330, 331, 332] and was allowed to cool to room temperature. The pH of the solution was measured again, re-adjusted to 10, and boiled again until colorless; this process of boiling, cooling, and re-adjusting the pH was repeated until the solution remained colorless and its pH remained stable after boiling and cooling, at which point the solution was aliquotted into 15 ml conical tubes and stored at -20°C [329, 331, 332].

Test for Phosphatase Inhibition in Transfected Amebae

We tested if the action of amebic tyrosine phosphatases could be inhibited by the addition of the phosphatase inhibitors sodium orthovanadate and iodoacetamide. Inhibition of the cellular tyrosine phosphatases would allow tyrosine kinases to phosphorylate their usual substrates, increasing the amount of tyrosine-phosphorylated proteins present in the amebae and thus the amount of possible substrate proteins available for the LMW-PTP substrate-trapping mutant. To assay if phosphatase inhibition would be possible even under the worst-case scenario when the wild-type LMW-PTP is constitutively overexpressed, we used transfectants expressing the N-terminally epitope-tagged wild-type LMW-PTP constitutively from the expression vector pGIR209 (single transfectant). Transfectants were grown in 25 cm² tissue culture flasks and selected with 6 µg/ml G418. One set of samples were pre-treated with 1 mM sodium orthovanadate by direct addition of 100 mM sodium orthovanadate pH 10 to their medium and incubating them at 37°C for 30 minutes before harvesting (please see Preparation of Sodium Orthovanadate Stock Solution above). Trophozoites were harvested, counted using a hemacytometer, and suspended either in regular lysis buffer (without phosphatase inhibitor added), or in inhibitor lysis buffer with or without EDTA added. The composition of these

buffers is detailed above under Lysis Buffers. Lysate from 5×10^5 amebae was loaded per lane. HALT™ Phosphatase Inhibitor (Thermo Fisher Scientific, Waltham, MA, USA) was also added to some of the lysed samples at a concentration of 1X. The cocktail contains a mixture of four broad-specificity phosphatase inhibitors: sodium fluoride, sodium orthovanadate, sodium pyrophosphate and β -glycerophosphate [333] but the specific concentrations of each inhibitor are not available. Cells were lysed via sonication as described in the Western Immunoblotting section, and 10 mM DTT was then added to the lysate. To check for enrichment of tyrosine-phosphorylated proteins, western immunoblots were probed with a 1:1000 dilution of 4G10 Platinum anti-phosphotyrosine antibody (Millipore, Billerica, MA, USA) as described in the Western Immunoblotting section.

Confocal Imaging

Confocal imaging of the cellular location of the C-terminal epitope-tagged wild-type and mutant LMW-PTP with induced or uninduced protein expression, and all other parts of the protocol except for inducing and harvesting the transfected amebae, was performed by technician Carrie Cowardin. Poly-L-lysine-coated coverslips were prepared beforehand. Poly-L-lysine hydrobromide (Sigma-Aldrich, St. Louis, MO, USA) was dissolved in PBS at a concentration of 0.1 $\mu\text{g}/\text{ml}$, and 100 μl drops were aliquotted as a liquid bead on small (12 mm size) round glass coverslips (Thermo Fisher Scientific, Waltham, MA, USA). Coverslips were allowed to dry for one hour, then were washed four times in MilliQ water by aliquotting the water onto the coverslip and aspirating off the water. Coverslips were then washed the same way with 70% ethanol, and allowed to air-dry in a cool dark area. To prepare samples for confocal imaging, dry coverslips were each transferred to a well in a 24-well plate. Transfected amebae overexpressing the C-terminally epitope-tagged wild-type and mutant LMW-PTP selecting with 12 $\mu\text{g}/\text{ml}$ G418 and 30

$\mu\text{g/ml}$ hygromycin and protein expression induced with $10 \mu\text{g/ml}$ tetracycline for 24 hours were harvested, counted, pelleted at $200 \times g$, and resuspended in TYI medium at a concentration of 1.3×10^7 cells/ml. $75 \mu\text{l}$ of the amoebae per sample were aliquotted as a bead onto the surface of the poly-L-lysine-coated coverslip. Amoebae were allowed to attach to coverslips for one hour at 37°C in an anaerobic bag, then coverslips were rinsed three times with M199S (M199, Medium 199 from Gibco BRL/Invitrogen, Carlsbad, CA, USA, supplemented with 5.7 mM cysteine, 25 mM HEPES, and 0.5% BSA (bovine serum albumin), pH adjusted to 6.8) warmed to 37°C , to remove amoebae that did not bind to the coverslips. Amoebae attached to the coverslips were fixed with 4% paraformaldehyde warmed to 37°C for 30 minutes. The reaction was quenched by rinsing 3 times, 5 minutes each, with 100 mM glycine in PBS. Quenched cells were permeabilized with 0.2% Triton X-100 in PBS for 1 minute and then washed three times with PBS + 0.1% Triton X-100. Samples were blocked overnight at 4°C in the dark in Blocking Solution: PBS with 20% goat serum, 5% BSA, and 0.1% Triton X-100, then rinsed three times in PBS with 0.1% Triton X-100. Coverslips were then removed to a small sheet of Parafilm set on top of a wetted sheet of Whatman paper in a hybridization tray (to keep the primary antibody solution from drying up during incubation). Mouse anti-V5-tag primary antibody (Invitrogen, Carlsbad, CA, USA) was diluted 1:1000 in Blocking Solution, and $50 \mu\text{l}$ of the primary antibody solution was aliquotted onto each coverslip. Coverslips were incubated at room temperature in the dark for 3.5 hours. After primary antibody incubation, coverslips were quickly washed four times, and then washed four more times for 5 minutes each in PBS with 0.1% Triton X-100. Secondary antibody (DyLight 649-conjugated goat anti-mouse secondary antibody (Jackson ImmunoResearch Laboratories, West Grove, PA, USA) was diluted 1:200 in Blocking Solution and aliquotted onto the surface of the coverslip, and incubated and washed the same way as the primary antibody. After the series of washes, the coverslips were washed once with a 1:5000 dilution of DAPI in

PBS for 5 minutes, and coverslips were mounted face down onto slides with a drop of fluoromount and then imaged.

Coupling Purified Mutant LMW-PTP and BSA to Affi-Gel 15 beads

Either 500 or 1000 μg purified recombinant *E. histolytica* substrate-trapping Cys7 to Ser7 mutant LMW-PTP protein, or an equal amount of BSA (bovine serum albumin), was coupled to 100 μl Affi-Gel 15 beads (Bio-Rad, Hercules, CA, USA). The LMW-PTP or BSA proteins were diluted in a total volume of 500 μl cold 100 mM HEPES (pH 7.5), 100 mM MOPS (pH 7.5), or 100 mM NaHCO_3 (pH 7.5); all coupling buffers appeared to work equally well. Beads-alone (mock-coupled) beads were also prepared, but with no protein added. To prepare Affi-Gel 15 beads for coupling, 100 μl of beads were transferred into microcentrifuge tubes and centrifuged at 2000 rpm for 10 seconds. The supernatant was removed, and beads were washed with 300 μl cold sterile water, centrifuged again, and the supernatant was removed. The 500 μl aliquots of diluted protein were added to the washed beads, and sample tubes were incubated overnight at 4°C on a nutator to maintain the beads in suspension. After incubation, beads were washed with cold 100 mM HEPES, MOPS, or NaHCO_3 solution until protein was undetectable in the wash buffer as measured by A_{280} . The amount of protein that successfully coupled to the beads was then calculated (about 70% of the total amount successfully coupled to the Affi-Gel 15 beads). If beads with coupled proteins were not to be immediately used, they were stored in microcentrifuge tubes at 4°C in either 100 mM HEPES or MOPS with 0.2% sodium azide added.

Pulldowns Using Mutant LMW-PTP and BSA Coupled to Affi-Gel 15 Beads

Inhibitor lysis buffer (containing 5 mM EDTA) was prepared as described as above in the Lysis Buffers section, and chilled on ice. HM1:IMSS strain non-transfected *E. histolytica* trophozoites were grown as described above in the Culture of Trophozoites section, and were pre-treated with 1 mM sodium orthovanadate by the direct addition of 200 mM sodium orthovanadate solution pH 10 to the TYI medium for 30 minutes before harvesting. Trophozoites were harvested by icing the flasks for 15 minutes, and the contents were transferred to 50 ml conical tubes. The conical tubes were centrifuged at $200 \times g$ for 5 min at 4°C , and pelleted trophozoites were suspended in a smaller volume of TYI medium (about 10 ml) and volumes combined into a single 50 ml conical tube. Trophozoites were then counted using a hemacytometer, centrifuged again as above to re-pellet them, the supernatant was removed and the trophozoite pellet was suspended in the inhibitor lysis buffer with EDTA as described in the Western Immunoblotting section. Lysates were incubated in microcentrifuge tubes on a nutator for 30 minutes at 4°C , then DTT was added to a final concentration of 10 mM and microcentrifuged at $10,000 \times g$ for 5 minutes [283, 334, 335, 336] to remove insoluble material. Clarified lysate was removed to a new microfuge tube. Total lysate protein was measured using the Bio-Rad Protein Assay Kit (Bio-Rad, Hercules, CA, USA), and was diluted to a concentration of $1 \mu\text{g}/\mu\text{l}$ with ice-cold inhibitor lysis buffer. $1000 \mu\text{l}$ lysate was added to each tube with prepared Affi-Gel 15 beads (described in the above section) and incubated on the nutator at 4°C overnight. The lysate was then removed and beads were washed five times with ice-cold PBS; beads were centrifuged at 2000 rpm for 10 seconds between washes to pull them to the bottom of the tubes. On the last wash, beads were transferred into new microfuge tubes, the supernatant was removed, and beads were suspended in SDS-PAGE sample buffer, frozen in liquid N_2 , and stored at -80°C , until subjected to SDS-PAGE and either immunoblotting or silver staining. Please see the Western Immunoblotting section for

sample preparation, SDS-PAGE and immunoblotting. Silver staining of samples was performed with the SilverQuest Staining Kit (Invitrogen, Carlsbad, CA, USA) as per the manufacturer's instructions; this kit is compatible with mass spectrometry. Silver-stained whole gels were submitted to the W.M. Keck Biomedical Mass Spectrometry Laboratory for trypsin digestion and mass spectrometry analysis of tryptic peptides of the two selected bands.

Mass Spectrometry Analysis

Mass spectrometry analysis was performed at the W.M. Keck Biomedical Mass Spectrometry Laboratory (Nicholas E. Sherman, Ph.D., Director). The analysis was returned as report number 1728. Samples were submitted on April 12, 2011, were digested on May 2, 2011, and analyzed by MS and MS/MS on May 6, 2011. The silver-stained bands were excised from the gel, and the gel pieces were transferred to a siliconized tube and washed and destained in 200 μ l 50% methanol overnight. The gel pieces were then dehydrated in acetonitrile, rehydrated in 30 μ L of 10 mM dithiothreitol in 0.1 M ammonium bicarbonate and reduced at room temperature for 0.5 h. The DTT solution was removed and the sample alkylated in 30 μ l 50 mM iodoacetamide in 0.1 M ammonium bicarbonate at room temperature for 0.5 hours. The reagent was removed and the gel pieces dehydrated in 100 μ l acetonitrile. The acetonitrile was removed and the gel pieces rehydrated in 100 μ l 0.1 M ammonium bicarbonate, then dehydrated in 100 μ l acetonitrile, the acetonitrile was removed and the pieces completely dried by vacuum centrifugation. The gel pieces were rehydrated in 20 ng/ μ l trypsin in 50 mM ammonium bicarbonate on ice for 10 min. Any excess enzyme solution was removed and 20 μ l 50 mM ammonium bicarbonate added. The sample was digested overnight at 37°C and the peptides formed extracted from the polyacrylamide in two 30 μ l aliquots of 50% acetonitrile/5% formic acid. These extracts were combined and evaporated to 15 μ l for MS analysis. The LC-MS system consisted of a Thermo

Electron Orbitrap Velos ETD mass spectrometer system with a Protana nanospray ion source interfaced to a self-packed 8 cm x 75 μm id Phenomenex Jupiter 10 μm C18 reversed-phase capillary column. 7.5 μl of the extract was injected and the peptides eluted from the column by an acetonitrile/0.1 M acetic acid gradient at a flow rate of 0.5 $\mu\text{l}/\text{min}$ over 0.5 hours. The nanospray ion source was operated at 2.5 kV. The digest was analyzed using the double play capability of the instrument acquiring full scan mass spectra to determine peptide molecular weights and product ion spectra to determine amino acid sequence in sequential scans. This mode of analysis produces approximately 10000 CAD spectra of ions ranging in abundance over several orders of magnitude. Not all CAD spectra were derived from peptides. The data were analyzed by database searching using the Sequest search algorithm against *E. histolytica*.

Cloning and Epitope-Tagging of Putative LMW-PTP Substrates Identified by Mass Spectrometry Analysis

The genes for selected proteins identified as putative LMW-PTP substrates in the pulldowns using untagged mutant LMW-PTP attached to Affi-Gel 15 beads (Bio-Rad, Hercules, CA, USA) by mass spectrometry analysis (Table 3-8) were cloned in a similar way to the wild-type and mutant LMW-PTPs. First, the GI number identified in the mass spectrometry analysis for each selected identified protein was input into the Protein Database at the National Center for Biotechnology Information website [337] to obtain the NCBI reference sequence protein accession number, and then the NCBI reference gene sequence; or alternately, the protein sequence itself obtained from the mass spectrometry data was input into BLAST [311] to obtain the NCBI reference sequence protein accession number, and then the NCBI reference gene sequence. The gene sequence was then used to design oligos (Tables 3-9) to amplify the entire coding region of the gene, except the start codon was not included in the gene that was to be N-

terminally tagged (it was included at the beginning of the tag) and the stop codon was not included in the gene that was to be C-terminally tagged (it was included at the end of the tag). Different epitope tags needed to be added than those that had been added to the LMW-PTPs. A combination of three epitope tags back-to-back were added to the possible substrate genes. These were the E epitope tag, the S epitope tag, and the KT3 epitope tag (Table 3-10). The tags were generated by annealing complementary oligos together. To anneal oligos, equal amounts of each oligo for each tag were added together into a microcentrifuge tube, and diluted with either water or TE buffer (10 mM Tris, 1 mM EDTA pH 7.5) to a final concentration of 2 mM of each oligo. Diluted oligos were aliquotted into PCR tubes, were heated to 95°C for 10 minutes and slowly cooled to 4°C in a thermalcycler. Tags were then ethanol-precipitated overnight at -20°C.

Restriction endonuclease recognition sites were included in the oligo sequences such that epitope tags and cloned genes would contain these sites on their ends for cloning purposes. Epitope tags that would be located on the 5' end of the protein included a *SpeI* site followed immediately by a *BglIII* site on their 5' ends and an *EcoRI* site on their 3' ends; those that would be located on the 3' end of the protein had an *EcoRI* site on their 5' ends and a *SalI* site on their 3' ends. The tags were cloned first into pBluescript II KS(+) (Agilent Technologies, Santa Clara, CA, USA) before the putative substrate protein genes, and were sequenced to verify they were correct and in frame. To clone the putative substrate genes, the oligo pairs listed in Table 3-9 were used. In general, cloned genes that would have a 5' tag added had an *EcoRI* site added to the 5' end and a *SalI* site to the 3' end, and those that would have a 3' tag added had a *SpeI* site followed immediately by a *BglIII* site added to the 5' end and an *EcoRI* site added to the 3' end. Alternate restriction endonuclease recognition sites were used if the gene sequence contained an internal *EcoRI*, *BglIII*, or *SpeI* site.

The oligo pairs listed in Table 3-9 were used to clone the selected genes via PCR in a 100 μ l reaction volume. *E. histolytica* genomic DNA was used as the PCR template. The PCR conditions were: 150 ng *E. histolytica* genomic DNA, 20 μ l 5X High Fidelity PCR Buffer, 2 μ l 50 mM MgCl₂, 1 μ l 100 μ M forward oligo, 1 μ l 100 μ M reverse oligo, 2 μ l 40 mM dNTP mix, 2 μ l (4 U) Phusion DNA polymerase (Thermo Fisher Scientific, Waltham, MA, USA), with sufficient water added to complete the volume to 100 μ l. The cycling conditions were: 98°C for two minutes, followed by 35 cycles of: 98°C for 10 seconds, 60.5°C for 30 seconds, and 72°C for one minute, followed by 72°C for 10 minutes, with the exception of the PCR products for the putative Arp 2/3 complex 34 kDa subunit and the conserved hypothetical protein gi56465028, which needed 200 ng genomic DNA template and a 66°C annealing temperature to yield a product. The PCR products were digested with the proper restriction endonucleases, ligated into pBluescript II KS(+) with the epitope tags previously added, transformed into chemically competent *E. coli*, colonies were screened for the correct size inserts, and plasmids extracted from overnight cultures and submitted for sequence analysis as described in the Cloning of Epitope-Tagged Wild-Type and Mutant LMW-PTPs section. Construct sequences were checked to verify that the epitope tags and protein genes were correct and in frame. Correct constructs were digested with the restriction endonucleases *Bgl*II and *Sal*I (or *Bam*H1 and *Sal*I in a few exceptions where the gene contained an internal *Bgl*II site). The amebic expression vector pGIR209 [177, 179] was also digested with the restriction endonucleases *Bgl* II and *Sal* I, and constructs subcloned into pGIR209 as described in the *E. histolytica* Constructs for LMW-PTP Overexpression section.

Results

Comparison of the *Entamoeba histolytica* LMW-PTP Protein Sequences with LMW-PTP Sequences from Other Organisms

Since there have been many LMW-PTP proteins sequenced from different organisms, we could compare other LMW-PTP protein sequences to the *Entamoeba histolytica* sequences. To examine how closely the two *E. histolytica* LMW-PTPs were related to LMW-PTPs from other organisms, and to review if typically conserved residues in other LMW-PTP proteins were also conserved in *E. histolytica*, the *E. histolytica* LMW-PTP XP_656359 protein sequence was input into BLAST [311] for comparison with other LMW-PTPs, with one exception: the amoeba database at AmoebaDB [309], which is part of EuPathDB, the Eukaryotic Pathogen Database [313] was used to compare the *E. histolytica* LMW-PTPs with the related species *Entamoeba invadens* LMW-PTP. The *Entamoeba invadens* IP1 strain was sequenced at the J. Craig Venter Institute (JCVI) [338]. *E. invadens* is a parasite of reptiles which is used as a model for the study of *E. histolytica* encystment, since *E. histolytica* trophozoites in culture do not differentiate into cysts [13, 339, 340, 341].

Selected LMW-PTP sequences from a number of different organisms were aligned with the *Entamoeba histolytica* LMW-PTPs using BLAST [311] and LALIGN [312]. The closely related *Entamoeba dispar* LMW-PTP and the less related *Entamoeba invadens* LMW-PTP were included, as were representative LMW-PTPs from eight species of plants, seven species of bacteria, five species of cyanobacteria, three species of yeast, the unicellular eukaryotes *Giardia lamblia*, *Giardia intestinalis*, and *Dictyostelium discoideum*, and LMW-PTPs from the multicellular species *Homo sapiens*, *Mus musculus*, *Bos taurus*, *Gallus gallus*, *Xenopus laevis*, *Drosophila melanogaster* and *Danio rerio*, so 40 representative LMW-PTP protein sequences in

total. Alignments for *E. histolytica* LMW-PTPs with those of selected plant LMW-PTPs are shown in Figure 3-3, alignments with selected LMW-PTPs from prokaryotic organisms are in Figure 3-4, alignments with selected unicellular eukaryote LMW-PTPs are shown in Figure 3-5, and alignments with selected LMW-PTPs from multicellular animals are shown in Figure 3-6. The full species names and strains, accession numbers, protein descriptions, length of the LMW-PTP, which residues of the *E. histolytica* LMW-PTPs are compared, the percent identity, similarity, and gaps present for the selected aligned LMW-PTPs are shown in Table 3-1 (comparison with plants), Table 3-2 (comparison with prokaryotes), Table 3-3 (comparison with single-celled eukaryotes) and Table 3-4 (comparison with multicellular organisms).

In terms of having the most identical or similar residues on a percentage basis, the most closely related LMW-PTP proteins to *Entamoeba histolytica*, outside of the *Entamoeba* genus, are found in plant species or Bacteroides species, with slightly more than half of the residues being identical and about two thirds being identical or similar (conservative substitutions). From the alignments, it is clear that certain regions of the protein sequence are highly conserved across kingdoms. The active site (P loop) consensus sequence (I/V)CLGNICRS, as expected, is almost completely conserved; out of the 40 selected LMW-PTP proteins, only two have one residue altered in the active site: *Vibrio cholerae* has a conservative substitution of a methionine residue for the leucine residue, and *Saccharomyces cerevisiae* has a nonconservative substitution of a phenylalanine for the isoleucine in the middle of the active site. The DPYY loop contains the conserved aspartic acid residue, which acts as a general acid and then as a general base during phosphate removal from the substrate, and the two back-to back tyrosines that have been shown to be phosphorylated in mammalian LMW-PTPs and regulate enzyme activity and recruit adaptor proteins [212, 219, 228]. The only LMW-PTP protein out of the 40 selected LMW-PTPs that

lacks a completely conserved DPYY loop is *Saccharomyces cerevisiae*, which has a substitution of a tryptophan for the first tyrosine (Figure 3-5).

There overall is quite a high degree of residue conservation between the LMW-PTP from *E. histolytica* and the other LMW-PTPs. The *E. histolytica* LMW-PTP protein, when compared to those from two other species within its own genus, the closely related LMW-PTP from *E. dispar* had four residues changed, one of them being a conservative substitution, so 97.5% identical and 98.1% similar; the less closely related *E. invadens* LMW-PTP was only 60.9% identical and 74.8% similar (Table 3-3, Figure 3-5). Plant and bacterial species had the next most similar LMW-PTPs; the plant LMW-PTPs were at least 50% identical and up to 66.9% similar to the *E. histolytica* LMW-PTP; most of the bacterial species were at least 50% identical and up to 67.6% similar to the *E. histolytica* LMW-PTP (Tables 3-1 and 3-2; Figures 3-3 and 3-4).

The *Entamoeba histolytica* LMW-PTPs were least related to those of multicellular animals, and interestingly, to the other unicellular eukaryotes examined; all of these were less than 40% identical to the *E. histolytica* LMW-PTPs (Tables 3-3 and 3-4; Figures 3-5 and 3-6). It is interesting also that the of the LMW-PTP proteins selected for comparison with the *E. histolytica* LMW-PTPs, those from multicellular animals have a smaller size range (155-159 residues) (Table 3-4) than those of bacteria (154-171 residues) (Table 3-2), and other unicellular eukaryotes (154-161 residues) (Table 3-3). Plant LMW-PTPs tend to be larger than any of the others (172-279 residues) (Table 3-1).

A Phylogenetic Tree for Selected LMW-PTP Protein Relationships

To analyze the relationship between these selected LMW-PTPs, the “one-click” phylogenetic tree maker using the bioinformatic tools at the web site www.phylogeny.fr/ [314] or its mirror site [315] were used to construct a phylogenetic tree comparing LMW-PTP protein

sequences from *Entamoeba* species with those selected from bacterial species, plant species, other unicellular eukaryotes, and multicellular eukaryotes (Figure 3-7). This site offers “robust phylogenetic analysis for the non-specialist” which is a “free, simple-to-use web service dedicated to reconstructing and analyzing phylogenetic relationships” [314, 315]. The tree shows *Entamoeba* and bacteria (except for cyanobacteria and *Pseudomonas aeruginosa*) branching from a common ancestor, with plants branching off slightly earlier; the cyanobacteria, except for *Prochlorus marinus*, branch off earlier than plants, and *Prochlorus marinus* and *Pseudomonas aeruginosa* branch off before the cluster of cyanobacteria. The LMW-PTPs from other eukaryotic organisms are on their own branch, clustering in groups of *Giardia*, *Saccharomyces cerevisiae*/*Kluyveromyces lactis*, *Dictyostelium discoideum*/*Schizosaccharomyces pombe*, and multicellular animals (Figure 3-7). The *Entamoeba histolytica* LMW-PTP proteins appear to be most closely related to those of plants or Bacteroides/Parabacteroides species, and are the least similar to those of eukaryotic multicellular animals or to the compared non-*Entamoeba* unicellular organisms.

Structures of the *Entamoeba histolytica* LMW-PTP

At the beginning of this project, this *E. histolytica* LMW-PTP had been already expressed as a recombinant protein and successfully crystallized by the Seattle Structural Genomics Center for Infectious Disease (SSGCID) [316] as described in the Recombinant Protein Expression at SSGCID section under Materials and Methods. Several different structures for the LMW-PTP were solved at SSGCID following crystallization and X-ray diffraction. These structures were: enzyme with no ligand bound (apo structure) (Figure 3-8), enzyme with the substrate mimic HEPES bound in the active site (HEPES is present in the crystallization buffer and bound in the active site as a substrate mimic) (Figure 3-9) and enzyme with the product mimic sulfate bound in

the active site (Figure 3-10). Structures were deposited in the RSCB Protein Data Bank (RSCB PDB) [342].

The non-ligand-bound (apo) structure of the LMW-PTP has a disordered P-loop (active site), presumably due to the absence of substrate (Figure 3-8). The disordered active site residues and those residues that interact with the active site residues did not crystallize, as the disordered flexible loops were not in a single configuration. The PDB identifier for the enzyme with no ligand bound is 3ILY [342]. Its resolution is 2.2 Å and its space group was $P4_3$, which means it is “primitive”: the crystals contain only one repeating unit [343] and the crystal shape of the crystallized protein was that of a tetragonal pyramid (Figure 3-12) [344, 345, 346, 347].

For the structure obtained of the LMW-PTP with substrate mimic bound, two structures (of 2.2 Å and 1.95 Å resolution) were obtained with a HEPES molecule bound tightly in the active site mimicking the substrate (Figure 3-9). Since the standard buffer for SSGCID proteins that go into crystallization trials contains HEPES at 25mM, HEPES is occasionally present in the structures of biological macromolecules. In this structure, the P-loop active site residues are well-ordered, as are the residues that interact with the P-loop residues (Figure 3-9). The PDB identifiers for these structures are 3IDO and 3JS5 respectively [342]. The space group for both is $P2_12_12_1$, which means the crystal structure is “primitive”, contains only one repeating unit, and forms an orthorhombic sphenoidal crystal [345, 346, 347, 348] (Figure 3-12). A *Saccharomyces cerevisiae* LMW-PTP crystal structure of 2.2 Å resolution has this same space group [270]. A structure was also obtained of the enzyme with product mimic bound, of 1.8 Å resolution, with sulfate, a product mimic, bound in the P-loop (active site). In this structure, the P-loop of the phosphatase is well ordered, but additional residues on the surface of the active site are disordered (Figure 3-10). The PDB identifier of this structure is 3JVI. The space group for this structure is

P4₃2₁2, which means the crystal structure shape is a “primitive” tetragonal trapezohedron (Figure 3-12) [346, 347, 349].

An overlay of all three structures for comparison is shown in Figure 3-11. Note how the yellow apo structure and the red sulfate (product mimic bound) structures appear to be discontinuous in sections, especially in the lower right of the molecule where the residues that interact with the P-loop residues are disordered, and therefore did not crystallize since the protein structure was not in a single configuration in all the individual protein molecules in the crystal. All structure images are courtesy of Thomas E. Edwards (Emerald BioStructures). The expression, crystallization, and X-ray diffraction work was performed by the SSGCID protein crystallization team: Angela Kelley (Cloning, University of Washington), Erwin Analau (Screening, University of Washington), Wenjin Guo (Purifications, Seattle BioMed), Janhavy Bhandari (Purifications, University of Washington), Becky Poplawski (Crystallization, Emerald BioStructures), and Shelly Dietrich (Crystallization, Emerald BioStructures). This series of structures provides a picture of the ligand-induced conformational changes that occur and are necessary for the binding of substrate and thus can be used as a starting point for drug design.

There have been many other LMW-PTP proteins from diverse organisms crystallized and their structures determined, which can be compared with the *Entamoeba histolytica* LMW-PTP structure. These include *Homo sapiens* [223, 262, 271], *Bos taurus* [272, 273, 274], *Saccharomyces cerevisiae* [270], *Mycobacterium tuberculosis* [275], *Bacillus subtilis* [276], *Campylobacter jejuni* [277], and the bovine parasite *Tritrichomonas foetus* [278]. The *Entamoeba histolytica* LMW-PTP structure has the same structure as other LMW-PTP structures from other organisms: the enzyme is characterized by a central twisted four-stranded β -sheet surrounded by five α -helices [270, 275, 278]. The two β - α - β motifs form a Rossmann fold, and the active site (P loop) is in the first β - α - β motif, at the C-terminus of β 1 and the N-terminus of α 1 [270, 275].

The P-loop conformation is stabilized by a network of hydrogen bonding and with substrate, substrate analogs or inhibitors bound, is the most rigid, and ordered, part of the enzyme [270, 272, 273, 278].

The configuration of active site residues (P-loop) for the *E. histolytica* LMW-PTP with the substrate mimic HEPES bound are shown in Figure 3-13 with the individual residues labeled and numbered according to which residue they are in the *E. histolytica* LMW-PTP protein sequence. Figure 3-14 shows the active site as a space-filling model. The residues that make up the active site are completely conserved, presenting the characteristic CLGNICR sequence [203, 228]. The DPYY loop [212, 219] contains the conserved aspartic acid residue which is catalytically critical in phosphate removal from the substrate [232, 271] and the two back-to back tyrosines which are known to be phosphorylated for regulating the LMW-PTP in the mammalian LMW-PTP [211, 228, 271]. The histidine residue at position 45 in the *E. histolytica* LMW-PTP is conserved in the selected plant, bacterial, and in most of the other unicellular eukaryotic LMW-PTPs, and is part of a variable loop which forms a wall of the active site crevice and may be involved in determining substrate specificity [275]. The catalytic mechanism for LMW-PTPs (and PTPs in general) is shown in Figure 3-2.

Figure 3-15 shows a space-filling model of top and side views of the *E. histolytica* LMW-PTP structure with HEPES bound in the active site; this is the same structure that is shown in Figure 3-9 in ribbon form. Note how the active site is an obvious feature in the space-filling model and forms a pocket or crevice where the substrate binds. This pocket is quite deep in all phosphotyrosine-specific phosphatase enzyme structures, about 9 Å, allowing specificity for phosphotyrosine substrates [224, 279; 280], but these enzymes then lack the ability to catalyze the removal of phosphate groups from phosphoserine and phosphothreonine, which have shorter side

chains [203]. In contrast, the dual-specificity phosphatases have shallower active site pockets of about 6 Å deep that can accommodate phosphotyrosine, -serine, and -threonine [224, 280].

Testing the Recombinant Wild-Type *Entamoeba histolytica* LMW-PTP Protein for Phosphatase Activity

The *E. histolytica* LMW-PTP is annotated as being a putative phosphatase. The purified recombinant wild-type LMW-PTP prepared at SSGCID was used in these assays to test if it actually had phosphatase activity. Recombinant protein preparation is described in detail in the Recombinant Protein Expression at SSGCID section under Materials and Methods. Briefly, the protein was expressed with a 6xHis tag followed by the 3C protease recognition sequence Leu-Glu-Ala-Gln-Thr-Gln-Gly-Pro. Recombinant protein was expressed in *E. coli* and purified by immobilized metal-affinity chromatography using nickel beads. The 6xHis tag was cleaved off with 6xHis-tagged 3C protease, the protein was further purified by subtractive nickel-immobilized metal-affinity chromatography to remove the 6xHis tag peptide, the 3C protease, and any nickel-binding *E. coli* proteins, further purified by size-exclusion chromatography, concentrated to 20-30 mg/ml and eluted in buffer containing 25 mM Tris pH 7.0, 500 mM NaCl, 5% (v/v) glycerol, 2 mM dithiothreitol and 0.025% (w/v) sodium azide [319].

The recombinant LMW-PTP was tested by us for phosphatase activity using the SensoLyte *p*NPP Colorimetric Protein Phosphatase Assay Kit (AnaSpec, Fremont, CA, USA). *p*NPP, *p*-nitrophenyl phosphate, is a colorimetric generic phosphatase substrate which turns yellow when dephosphorylated. *p*-nitrophenyl phosphate has been used in many studies to test PTP catalysis [203, 225, 265], especially if the physiological substrate of a PTP is unknown, and it can also be a challenge to obtain enough of the properly phosphorylated substrate for enzymological testing even if the substrate is known [225].

To assay the phosphatase activity of the recombinant wild-type LMW-PTP, different amounts of recombinant wild-type LMW-PTP, ranging from 10 ng to 200 ng, were diluted in supplied buffer with a final concentration of 5 mM *p*NPP. The reaction was allowed to proceed, and absorbance was measured at 405 nm at the 60 minute time point. To test inhibition of phosphatase activity, HALT™ Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific, Waltham, MA, USA) was added to the reaction at 1X as described in the Testing Recombinant Wild-Type and Mutant LMW-PTPs for Phosphatase Activity section under Materials and Methods. Values were plotted as nanograms of recombinant wild-type LMW-PTP versus absorbance at 405 nm using Excel (Figure 3-16). The slope of the trendlines and the R^2 values were also calculated using Excel, and these values are displayed on the graph (Figure 3-16). The wild-type recombinant LMW-PTP definitely had phosphatase activity, and was inhibited by the addition of the HALT™ Phosphatase Inhibitor Cocktail by an average of 92.5 ± 1.0 %.

Testing of Rabbit Immune Sera

To test that the immune sera from rabbits inoculated with the purified LMW-PTP recombinant protein, the two immune test bleeds from each of the two rabbits were assayed to check that they recognized the native LMW-PTP from *E. histolytica* HM1:IMSS strain amebae, as compared with the pre-immune serum collected before immunization, and if the sera could be used to immunoprecipitate the native LMW-PTP from amebic lysate. Please see Immunoprecipitation of LMW-PTP with Rabbit Serum section under Materials and Methods for details. A representative immunoblot is shown in Figure 3-17. The expected size of the *E. histolytica* LMW-PTP is 17 kDa, and we were able to see a band of the expected size after immunoprecipitating with serum from one rabbit and probing the immunoblot with serum from the other rabbit, so the sera appeared to be recognizing the native LMW-PTP protein (100 ng of

the purified protein was included as a positive control), and the pre-immune sera are not recognizing amebic proteins. There was a band of about 50 kDa apparent in the immunoprecipitated sample, which was likely IgG heavy chain, since there was other rabbit IgG in the rabbit sera that could have been binding to the Protein A Dynabeads and eluting with the rest of the IP, since the antibody was not crosslinked to the beads.

Cloning of the Substrate-Trapping Low Molecular Weight Protein Tyrosine

Phosphatase Mutant

A mutant LMW-PTP was cloned as a tool to help identify the cellular substrates of this LMW-PTP. This mutant phosphatase was expected to “trap” the substrate: it should bind substrate but not release it or catalyze the conversion to product [220, 232]. The active site of the LMW-PTP is close to the N-terminus of the protein [219] so this facilitated generating site-directed mutations via PCR. The wild type LMW-PTP was cloned as a control, while the mutant version had a Cys7 to Ser7 mutation. This mutation is catalytically inactive in other LMW-PTPs [228, 250, 252, 281, 282, 283]. The absence of the nucleophilic center in the active site should render the enzyme inactive [220, 232]. The wild-type and substrate-trapping active site mutant LMW-PTPs were first cloned without epitope tags as described in the Cloning of Untagged Wild-Type and Mutant LMW-PTPs section under Materials and Methods, and oligonucleotides used in cloning are shown in Table 3-5. The plasmid DNA for the untagged mutant LMW-PTP clone in pBluescript II KS(+) was minipreped and sent to SSGCID for recombinant protein expression and purification (please see the Recombinant Protein Expression at SSGCID section under Materials and Methods for details on protein expression and purification). The recombinant mutant LMW-PTP protein was used to both test for phosphatase activity and was covalently attached to beads for pulldown of putative LMW-PTP substrate proteins. The untagged wild-type

and mutant LMW-PTP clones had epitope tags added to either the 5' or 3' end of the cloned gene so the expressed proteins would have N-terminal or C-terminal epitope tags. The process of cloning these is described in the Cloning of Epitope-Tagged Wild-Type and Mutant LMW-PTPs section under Materials and Methods.

To overexpress the wild-type and mutant LMW-PTP proteins, we used a two-vector tetracycline-inducible system to express the LMW-PTP wild-type and mutant forms inducibly using the two expression vectors pGIR209 and pGIR308 originally developed by Girija Ramakrishnan (Figure 3-18) [177, 179]. The epitope-tagged wild-type and mutant LMW-PTPs were subcloned into pGIR209 [177, 179] using the *Bgl* II and *Sal* I restriction sites, which removed the luciferase reporter gene. pGIR308 [177, 179] expresses the tetracycline repressor protein (Figure 3-18). When these two plasmids are co-transfected into amoebae, the tetracycline repressor is expressed from pGIR308, and the repressor binds to the tetracycline operator at the transcription start site in pGIR209, so in this case the LMW-PTP gene (or luciferase reporter gene) is not transcribed. In the presence of tetracycline, the repressor protein is not bound to the operator and transcription can occur. pGIR209 when transfected by itself functions as a constitutive expression vector. Amoebae were transfected first with plasmid pGIR308, and then pGIR308 transfectants or non-transfected HM1:IMSS amoebae were transfected with sequence-verified wild-type and mutant epitope-tagged LMW-PTP clones in pGIR209, giving the option of either tetracycline-inducible or constitutive expression. Please see the section on Transfection of Amoebae under Materials and Methods.

Testing the Recombinant Substrate-Trapping Mutant *Entamoeba histolytica* LMW-PTP Protein for Phosphatase Activity

To test that the active-site substrate-trapping mutant was catalytically nonfunctional, the mutant LMW-PTP recombinant protein for which we had cloned the gene, and was expressed and prepared at SSGCID (please see the Recombinant Protein Expression at SSGCID section under Materials and Methods for details on protein expression and purification) was tested for phosphatase activity with the SensoLyte *p*NPP Colorimetric Protein Phosphatase Assay Kit (AnaSpec, Fremont, CA, USA). To assay the phosphatase activity of the mutant recombinant protein as compared with the recombinant wild-type LMW-PTP, 100 ng of each protein per well was diluted in supplied buffer with a final concentration of 5 mM *p*NPP. The reaction was allowed to proceed, and absorbance was measured at 405 nm every five minutes for 120 minutes. Please see the Testing Recombinant Wild-Type and Mutant LMW-PTPs for Phosphatase Activity section under Materials and Methods for more detail. Values were plotted as nanograms of recombinant wild-type LMW-PTP vs absorbance at 405 nm using Excel (Figure 3-19). The slope of the trendline and the R^2 value was also calculated using Excel and these values are displayed on the graph (Figure 3-19). The mutant recombinant LMW-PTP as compared with the wild-type recombinant LMW-PTP did not possess any phosphatase activity. This is what was expected, since the equivalent Cys12 to Ser12 mutation in the mammalian LMW-PTP abolishes enzymatic activity, but still allows substrate binding [228, 250, 252, 281, 282].

Test for Tetracycline Induction of Epitope-Tagged Overexpressed Wild-Type LMW-PTP Protein in Transfectants.

Western immunoblotting was performed to test that: (i) tetracycline induces expression of the epitope-tagged wild-type and mutant LMW-PTP transfectants, (ii) the added FLAG, cMyc,

and V5 epitope tags are recognized by commercial antibodies, and (iii) the antibodies do not recognize amebic proteins. Amebae doubly transfected with both pGIR308 and either pGIR209, with the original luciferase reporter gene present, or the wild-type N-terminally or C-terminally tagged LMW-PTP cloned into pGIR209 were selected with both hygromycin and G418 (neomycin). Nontransfected amebae did not have hygromycin or G418 added. To induce gene expression, tetracycline was added to the medium to induce gene expression at a concentration of either 1 $\mu\text{g/ml}$ or 5 $\mu\text{g/ml}$, and amebae were left to grow for 24 hours. Control samples did not have tetracycline added, in order to test gene induction. Non-transfected amebae were also included as a control to check that antibody directed against the epitope tags would not recognize amebic proteins. Amebae were then harvested, counted, and lysed, and lysates from 5×10^4 amebae were loaded per lane and subjected to SDS-PAGE. Western immunoblotting was performed with antibodies recognizing the V5, FLAG, and c-Myc epitope tags as described in the Western Immunoblotting section. Epitope tag sequences are listed in Table 3-10. Please see Figure 3-20 for a representative immunoblot for wild-type LMW-PTP transfectants. The commercial antibodies against the V5, FLAG, and c-Myc epitopes all recognized a band of the correct size (with the epitope tags added, the LMW-PTPs are 20 kDa rather than 17 kDa) and the expression of the LMW-PTP was induced well. There were no bands visible for either the non-transfected amebae or for the transfectants expressing untagged luciferase from pGIR209 (Figure 3-20), so there was good induction of expression that was not leaky, and other amebic proteins were not recognized by antibodies directed against the epitope tags.

Testing if Treatment of *E. histolytica* Trophozoites with the Protein Phosphatase Inhibitor Sodium Orthovanadate Increases the Cellular Level of Phosphorylated Proteins

Before substrate trapping studies could be performed, we needed to test if the amount of tyrosine-phosphorylated possible protein substrates available for binding to the recombinant mutant LMW-PTP could be increased by inhibiting cellular phosphatase function. Therefore, we tested if the action of amebic cellular tyrosine phosphatases could be inhibited by the addition of the tyrosine phosphatase inhibitors sodium orthovanadate and iodoacetamide. Inhibition of the cellular tyrosine phosphatases should allow tyrosine kinases to phosphorylate their usual substrates, increasing the amount of tyrosine-phosphorylated proteins present in the amebae and thus the amount of possible substrate proteins available for the LMW-PTP substrate-trapping mutant. To test if phosphatase inhibition would be possible even under the worst-case scenario when the wild-type LMW-PTP was overexpressed constitutively, we used transfectants constitutively expressing the N-terminal epitope-tagged wild-type LMW-PTP from the expression vector pGIR209 (single transfectant). Transfectants in 25 cm² tissue culture flasks were pre-treated with 1 mM sodium orthovanadate by direct addition of 100 mM sodium orthovanadate pH 10 to the transfectants in TYI medium for 30 minutes before harvesting. The vanadate is a phosphate analog acting as a competitive inhibitor for the normal substrate of phosphorylated tyrosine residues [283, 350]. For details, please see the Test for Phosphatase Inhibition in Transfected Amebae section under Materials and Methods. Trophozoites were harvested, counted using a hemacytometer, and suspended either in regular lysis buffer (without phosphatase inhibitor added), or in inhibitor lysis buffer with or without EDTA added. The EDTA chelates metal ions such as Mg²⁺, Mn²⁺, and Zn²⁺, thus inhibiting the activity of metalloproteases [323] and serine-threonine protein phosphatases, since these enzymes require a

metal ion cofactor for catalysis [324, 325, 326, 327]. Protein tyrosine phosphatases do not require a metal ion for activity [326] and should therefore be unaffected by the presence of EDTA, but we performed the assay in the presence and absence of EDTA to test this just in case there was an effect. The composition of these buffers is detailed in the Lysis Buffers section under Materials and Methods. Please see Figure 3-21 for a representative immunoblot.

Pre-treatment of amebae with 1 mM sodium orthovanadate solution increased the amount of tyrosine-phosphorylated proteins, especially of higher-molecular-weight species. The 30-minute pre-incubation with vanadate added to the medium successfully increased the tyrosine-phosphorylated protein population. Iodoacetamide was also included in the lysis buffer where indicated. Iodoacetamide and iodoacetic acid are alkylating agents used to inhibit protein tyrosine phosphatases: they chemically modify thiol groups irreversibly [335]. The catalytic cysteine in the active site is normally found in the thiolate anion form ($-S^-$) and is highly susceptible to oxidation [235]. Iodoacetamide and iodoacetic acid react specifically with the thiolate anion, which attacks and displaces the iodine atom in a nucleophilic reaction: $-S^- + I-CH_2CONH_2 \rightarrow S-CH_2CONH_2 + I^-$ [351].

Cellular Localization of Epitope-Tagged Wild-Type and Mutant Low Molecular Weight Protein Tyrosine Phosphatase

The C-terminally epitope-tagged wild-type and mutant LMW-PTP overexpressing transfectants were utilized to visualize the cellular location of the LMW-PTP via confocal microscopy. Confocal microscopy was performed by technician Carrie Cowardin. Overexpressers had to be used because confocal imaging of the cellular location of native LMW-PTP using nontransfected amebae probed with anti-LMW-PTP rabbit sera, not only did the control pre-immune rabbit serum recognize many amebic cellular proteins, but the native level of the LMW-

PTP was sufficiently low that both the pre-immune and immune anti-LMW-PTP rabbit sera gave unacceptably high backgrounds (data not shown). The C-terminally epitope-tagged wild-type and mutant tetracycline-inducible LMW-PTP transfectants were selected with 12 $\mu\text{g/ml}$ G418 and 30 $\mu\text{g/ml}$ hygromycin, and protein expression was induced with 10 $\mu\text{g/ml}$ tetracycline for 24 hours before harvesting. Control transfectants were not induced with tetracycline. Amebae were harvested, counted, resuspended in M199S, allowed to attach to polylysine-coated coverslips, fixed with 4% formaldehyde and blocked overnight. Mouse anti-V5 tag primary antibody (Invitrogen, Carlsbad, CA, USA) was used at a dilution of 1:1000, and DyLight 649-conjugated goat anti-mouse secondary antibody (Jackson ImmunoResearch Laboratories, West Grove, PA, USA) was used at a dilution of 1:200. Amebae were also stained with DAPI at 1:5000 to visualize the nuclei (blue). Please see the Confocal Imaging section under Materials and Methods for details.

Representative wild-type and mutant LMW-PTP induced and non-induced amebae are shown in Figure 3-22. Protein expression appeared to be induced in both the wild-type and mutant overexpressers due to the addition of tetracycline. The wild-type epitope-tagged LMW-PTP with tetracycline-induced expression showed increased expression of the protein as compared with no induction, and the distribution appeared to be vesicular with possibly some in the cytoplasm, with the protein appearing to have a somewhat punctate distribution. The mutant LMW-PTP overexpresser had a similar protein distribution. Interestingly, individual trophozoites appeared to be expressing the protein at different levels. One caveat is that since the wild-type and mutant LMW-PTP proteins are being overexpressed, the intracellular location of the overexpressed proteins may not be completely representative of the endogenous cellular location; for example, the “extra” protein may be being transferred into vesicles for disposal.

Substrate Trapping Using Recombinant Substrate-Trapping Mutant LMW-PTP Protein Coupled to Beads

To test if it were possible to trap LMW-PTP substrates by using the substrate-trapping Cys7 to Ser7 LMW-PTP mutant coupled to beads, pulldowns were performed using mutant LMW-PTP or BSA (bovine serum albumin) coupled to Affi-Gel 15 beads (Bio-Rad, Hercules, CA, USA). Either 500 or 1000 μg recombinant mutant LMW-PTP or BSA in 100 mM HEPES, 100 mM MOPS, or 100 mM NaHCO_3 in water (all pH 7.5) was attached to 100 μl Affi-Gel 15 beads overnight at 4°C on a nutating mixer. Mock protein attachments to beads were done using buffer and beads only (no protein was coupled). HM1:IMSS amoebae were pre-treated with 1 mM sodium orthovanadate for 30 minutes, harvested, and lysed. Beads with either LMW-PTP or BSA attached or beads alone were incubated overnight at 4°C on a nutating mixer with either 1000 μg lysate or with lysis buffer alone. According to Garton *et al* (1999), lysates are stable with regard to their phosphotyrosine content during “prolonged” (at least 20 hours) incubation at 4°C in the absence of exogenously added phosphatase [335]. After incubation and washing, half of each sample was loaded on each of two SDS-PAGE gels. After electrophoresis, one of the gels was silver-stained with the SilverQuest Staining Kit (Invitrogen, Carlsbad, CA, USA) and the other blotted to membrane and probed with 4G10 Platinum anti-phosphotyrosine antibody (Millipore, Billerica, MA, USA) to assay for phosphorylated proteins. Please see the Coupling Purified Mutant LMW-PTP and BSA to Affi-Gel 15 beads and Pulldowns Using Mutant LMW-PTP and BSA Coupled to Affi-Gel 15 Beads sections under Materials and Methods for more detail.

Figure 3-23 shows a representative immunoblot probed with anti-phosphotyrosine antibody. There are three protein bands that appear to be phosphorylated: one at ~37 kDa, one about ~45 kDa, and one about ~55 kDa. Some of the attached mutant LMW-PTP protein seems to have detached from the Affi-Gel 15 beads, and the protein is running higher than expected,

between 20 and 25 kDa rather than its usual 17 kDa. This is likely due to a portion of the recombinant mutant LMW-PTP protein that was attached to the Affi-Gel 15 beads detaching from the beads; the beads were boiled in the presence of SDS-PAGE sample buffer with DTT added before being subjected to SDS-PAGE.

In the silver-stained gels, there appeared to be two unique bands that were the same size as two of the phosphorylated bands in the immunoblots, of about ~37 kDa and ~45 kDa. The LMW-PTP itself also appears to be phosphorylated. A representative silver-stained gel is shown in Figure 3-24, and these bands are indicated by red arrows. The larger apparent size of the mutant recombinant LMW-PTP for these substrate-trapping studies is also seen in these silver stained gels, likely because some of the recombinant mutant LMW-PTP is being released from the Affi-Gel 15 beads when the sample is boiled. The BSA is also running at a larger apparent size, about ~75 kDa, rather than its expected ~66 kDa [352] likely for the same reason. The recombinant mutant LMW-PTP was expressed and purified for us at SSGCID, and the protein preparation appears to have some contaminating *E. coli* proteins that are apparent in the silver stained gel of the lanes containing the mutant LMW-PTP attached to Affi-Gel 15 beads and incubated only with buffer. Since there were two unique bands in the silver stained gels that corresponded to two of the three phosphorylated bands in the immunoblots, it appeared that there were indeed possible substrates that had been recovered. The silver-stained gels were submitted for mass spectrometry analysis of the two bands at ~37 kDa and ~45 kDa. (Figure 3-25).

Possible LMW-PTP Substrate Proteins Identified by Mass Spectrometry Analysis

A total of eighty-four proteins were identified in the ~37 and ~45 kDa bands, with some proteins being present in both bands. Please see the Mass Spectrometry Analysis section under Materials and Methods for details on sample preparation and treatment. Sample digestion and

analysis was performed at the W.M. Keck Biomedical Mass Spectrometry Laboratory. The peptides, and their corresponding proteins, were identified for the samples submitted. Several dozen proteins were identified in each sample. Most proteins matched the predicted molecular masses, but the most abundant, of significantly lower molecular mass, was not surprisingly identified as the mutant *E. histolytica* LMW-PTP. The data were searched for serine, threonine, or tyrosine phosphorylation but none were found. There are a number of possible reasons for this. The phosphotyrosine sites could have been at low stoichiometry, or on proteins at low relative abundance, or located in regions not covered by the digest.

The proteins identified from peptides in the ~37 kDa band are listed in Table 3-6, and the proteins identified from peptides found in the ~45 kDa band are listed in Table 3-7. Each table includes the rank of the protein (under #) in terms of the number of unique peptides generated for that protein (the lower the number, the more unique peptides were found), the description of the protein, the total number of assigned spectra for that protein, the number of unique peptides in those spectra (there can be more than one of the same peptide), the percent peptide coverage of the protein, whether or not it was found in both bands that were sequenced (the ~37 kDa and ~45 kDa bands) and if the protein is close to being the correct size to be present in that band (plus or minus 5 kDa). Many proteins were identified with only one unique peptide being found; and a number of proteins were not the correct size to be found in that band. Proteins that were not of the correct size could represent degraded proteins (proteins of larger molecular mass) or proteins with post-translational modifications (proteins of smaller molecular mass).

There were eight possible substrate proteins that we selected as likely putative substrate proteins to be examined further (Table 3-8) based on how many unique peptides had been identified for each, if they were in the window of being the correct size, had good peptide coverage of the protein, and had tyrosines in their sequence that would likely be phosphorylated

and thus substrates for the LMW-PTP. One of these proteins, URE3-BP, was not the correct size in itself, which is 22 kDa, but it was found in the ~45 kDa band and would be the correct size if it were bound as a dimer. It was included because we already have antibodies that recognize it, and the mammalian LMW-PTP is known to have a transcription factor substrate. The other proteins were to be cloned, have an epitope tag added, and overexpressed using the pGIR209/pGIR308 inducible system.

The proteins selected for cloning, epitope tagging, and overexpression were: hypothetical protein 328.t00002, a type A flavoprotein, a putative protein kinase, conserved hypothetical protein gi56465028, a putative ARP2/3 complex 34 kDa subunit, hypothetical protein 503.t00001, and a putative ribose-phosphate pyrophosphokinase (Table 3-8). Hypothetical protein 328.t00002 has no conserved domains according to BLAST [311] and only matches *E. histolytica* and *E. dispar* proteins. The type A flavoprotein contains a lactamase B domain [353] and a flavodoxin (NADPH-dependent FMN reductase) domain [354]. In bacteria, flavoproteins catalyze electron transfer reactions [355] and β -lactamases allow for resistance to β -lactam antibiotics such as penicillins and cephalosporins [356] so this flavoprotein could be involved in electron transfer and/or antibiotic resistance. The putative protein kinase appears to be a serine/threonine kinase related to the mammalian Ste20-like protein kinase MST1, which in mammalian cells activates the mitogen-activated protein kinases (MAPKs) p38 and JNK through MKK7 (a MAPK kinase) and MEKK1 (a MAPK kinase kinase) by acting as a MAPK kinase kinase kinase; activation of JNK leads to caspase activation and apoptosis [357]. MST1 action has also been implicated in cell proliferation and differentiation [357]. Conserved hypothetical protein gi56465028 belongs to the TLD superfamily (pfam07534); this domain is “predicted to be an enzyme” [358]. Hypothetical protein 503.t00001 contains both a BAR domain (Bin/Amphi-physin/Rvs) domain, which is involved in dimerization, lipid binding and membrane curvature sensing modules and are found

in many different proteins with diverse functions including organelle biogenesis, membrane trafficking or remodeling, and cell division and migration, according to NCBI [359], and an SH3 (Src homology 3) domain, which binds to proline-rich ligands, especially PXXP motifs; SH3 domains “play a role in the regulation of enzymes by intramolecular interactions, changing the subcellular localization of signal pathway components and mediate multiprotein complex assemblies” [360]. The putative Arp 2/3 complex 34 kDa subunit is interesting since the Arp 2/3 protein complex has been implicated in the control of actin polymerization in cells [361]. Mammalian LMW-PTP has been found to be associated with the cytoskeleton [211, 219, 301], and since actin was a protein identified in the mass spectrometry analysis (Table 3-6 and Table 3-7), the *E. histolytica* LMW-PTP may also be associated with the cytoskeleton.

The locations of the identified tryptic peptides were diagrammed onto the protein sequences of the selected putative substrate proteins to show where in the protein sequences the tryptic peptides were located. The individual unique peptides are also shown in a table next to the protein sequence for each protein. Figure 3-26 shows these sequences for putative substrate proteins 328t.00002 (12 unique peptides with 36% coverage of the protein sequence) and type A flavoprotein (15 unique peptides with 44% coverage of the protein sequence) (Table 3-8). Figure 3-27 shows these sequences for substrate proteins putative protein kinase gi56466736 (9 unique peptides with 31% coverage of the protein sequence) and conserved hypothetical protein gi56465028 (7 unique peptides with 21% coverage of the protein sequence) (Table 3-8). Figure 3-28 shows these sequences for putative substrate proteins putative ARP2/3 complex 34 kDa subunit (7 unique peptides with 27% coverage of the protein sequence) and the hypothetical protein 503.t00001 (7 unique peptides, with 19% coverage of the protein sequence) (Table 3-8). Figure 3-29 shows these sequences for putative substrate proteins putative ribose-phosphate pyrophosphokinase (4 unique peptides with 13% coverage of the protein sequence) and the

calcium-binding and URE3-sequence specific DNA binding protein (URE3-BP) (2 unique peptides with 15% coverage of the protein sequence) (Table 3-8).

To check that these seven putative substrate proteins (plus URE3-BP) contained tyrosine residues in their protein sequences that likely would be phosphorylated, and thus could be LMW-PTP substrates, the NetPhos 2.0 server [362] was used to predict tyrosine phosphorylation sites in these proteins. The NetPhos 2.0 server produces neural network predictions for serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins [362]. Specific recognition of a phosphotyrosine residue occurs in the context of a peptide sequence, with the flanking residues conferring additional binding affinity of the enzyme to its substrate [279]. The higher the score is to a value of one, the more likely the phosphorylation of that tyrosine residue. All these selected putative substrate proteins contained at least two predicted phosphorylation sites, with hypothetical protein 503.t00001 containing a high of six (Figure 3-30).

To clone the genes for these proteins, oligo pairs were designed to amplify their open reading frames. Oligo pairs used to amplify these genes via PCR are shown in Table 3-9. The PCR conditions, epitope-tagging, and cloning process is explained in detail in the Cloning and Epitope-Tagging of Putative LMW-PTP Substrates Identified by Mass Spectrometry Analysis section under Materials and Methods. The epitope tags added to these putative substrate genes are shown in Table 3-10. Epitope tags were cloned into pBluescript II KS(+), and genes were cloned via PCR using *E. histolytica* genomic DNA as a template, ligated into pBluescript II KS(+) with the previously cloned epitope tag, the gene sequence was verified, and 5' - and 3' -epitope-tagged genes were subcloned into pGIR209 [177, 179]. The luciferase reporter gene in pGIR209 also had 5' epitope tags added as a control to be used for transfection and antibiotic selection. Please see the schematic diagram of subcloning epitope-tagged putative substrate genes into pGIR209 in Figure 3-31.

Discussion

Two of the twenty genes annotated as protein tyrosine phosphatases or putative tyrosine phosphatases in the parasite *E. histolytica* are low molecular weight protein tyrosine phosphatases, XP_656359 and XP_653357, which are identical except for one amino acid change at position 85 from an alanine in XP_656359 to a valine in XP_653357. Only one was cloned (XM_651267) and had its protein (XP_656359) used for this study.

We have partially characterized the *E. histolytica* LMW-PTP protein. Several different structures for this amebic tyrosine phosphatase have been solved at SSGCID, showing the conformational changes that occur when a substrate or product analog is bound. The structures of the *E. histolytica* LMW-PTP show that it has the same configuration as LMW-PTP structures from other organisms; LMW-PTPs all have a central twisted four-stranded β -sheet surrounded by five α -helices. The protein sequences of the two LMW-PTP proteins in *E. histolytica* have all the features of known LMW-PTPs: they contain the conserved active site with the catalytic cysteine, the stabilizing arginine, the second cysteine in the active site which can form a disulfide bond with the catalytic cysteine to prevent irreversible oxidation, and the DPYY loop which contains the general acid/base asparagine followed by two tyrosines. These features, which are required for catalysis, are completely conserved. The *E. histolytica* LMW-PTP protein sequence is most similar to selected plant or Bacteroides LMW-PTPs, with about 50% sequence identity, and is the least similar to those from other unicellular eukaryotes or multicellular animals (33%-40% identity). Some of the plant LMW-PTPs were quite a bit larger than canonical LMW-PTPs, and it may be that the plant LMW-PTPs at the larger end of the spectrum may not have the same function or substrate types as their smaller relatives.

The *E. histolytica* LMW-PTP, annotated as a putative phosphatase, and according to its structural features and conserved sequences should be an actual phosphatase, was found to have

phosphatase activity. Recombinant wild-type *E. histolytica* LMW-PTP protein had phosphatase activity when tested with the generic substrate *p*NPP, and this activity was inhibited by a commercial blend of phosphatase inhibitors. The substrate-trapping mutant recombinant LMW-PTP protein with a Cys to Ser mutation of the catalytic cysteine in the active site had no catalytic activity when tested with *p*NPP, as expected.

We were able to successfully generate rabbit sera that recognized the native protein in trophozoites, and immunoprecipitated the LMW-PTP from amebae using the rabbit sera. We cloned the wild-type and mutant LMW-PTPs with added epitope tags, overexpressed them inducibly from an amebic expression vector, and both the wild-type and mutant LMW-PTP could be detected in lysate by immunoblotting without an immunoprecipitation step. These epitope-tagged LMW-PTPs were used to visualize the cellular location in the amebae via confocal microscopy using antibody against the epitope tags, and its location appears to be cytoplasmic and vesicular, which agrees with the general location of LMW-PTPs in other organisms. In preparation for substrate-trapping experiments, we tested that pre-treatment of amebae with sodium orthovanadate allowed the inhibition of cellular protein tyrosine phosphatases, thus increasing the amount of available tyrosine-phosphorylated substrates available for the substrate-trapping LMW-PTP mutant.

To trap possible LMW-PTP substrates, the recombinant substrate-trapping mutant LMW-PTP protein was coupled to Affi-Gel 15 beads, beads were incubated with lysate from sodium orthovanadate-treated HM1:IMSS amebae, and samples were subjected to SDS-PAGE. Gels were either silver-stained or blotted to membrane and membranes probed with anti-phosphotyrosine antibody. Two of the three phosphorylated bands in the anti-phosphotyrosine blots corresponded with unique bands of the same sizes in the silver-stained gels, so these two bands of about 37 and 45 kDa were submitted for analysis via mass spectrometry. There were no peptides with tyrosine

phosphorylation found, but this could be because those sites are at low stoichiometry, in regions not covered by the trypsin digest, or on proteins at low relative abundance.

Of the 84 total proteins present in both bands, seven putative substrate proteins were chosen for further analysis based on their being the correct size to be present in the band, having a high number of unique peptides and percent peptide coverage of the protein sequence, and if there were tyrosine residues in the sequence predicted to be phosphorylated. An additional identified protein, URE3-BP, was also included since it fit all the criteria except for size, and we had existing antibodies against it. The open reading frames for these proteins were cloned into the tetracycline-inducible expression vector pGIR209 with epitope tags added. These seven proteins were: hypothetical protein 328.t00002, a type A flavoprotein, a putative protein kinase, conserved hypothetical protein gi56465028, a putative ARP2/3 complex 34 kDa subunit, hypothetical protein 503.t00001, and a putative ribose-phosphate pyrophosphokinase. None of these proteins appeared to contain transmembrane domains. Hypothetical protein 328.t00002 contains no conserved domains and only matches *E. histolytica* and *E. dispar* proteins. The type A flavoprotein could be involved in electron transfer and/or antibiotic resistance. The putative protein kinase is related to a mammalian serine/threonine MAPK kinase kinase which is involved in caspase activation and apoptosis as well as cell proliferation and differentiation [357]. Conserved hypothetical protein gi56465028 belongs to the TLD superfamily; TLD domains are predicted to have enzymatic activity [358]. Hypothetical protein 503.t00001 contains both a BAR domain (Bin/Amphiphysin/ Rvs) domain, which is involved in dimerization, lipid binding and membrane curvature sensing [359], and an SH3 (Src homology 3) domain, which binds preferentially to PXXP motifs; proteins with these motifs “play a role in the regulation of enzymes by intramolecular interactions, changing the subcellular localization of signal pathway components and mediate multiprotein complex assemblies” [360], so likely it is involved in

protein-protein interactions, and perhaps in membrane remodeling, cell division, or migration [359]. The putative Arp 2/3 complex 34 kDa subunit is an interesting find, since the Arp 2/3 protein complex has been implicated in the control of actin polymerization in cells [361]. Actin was also identified in the mass spectrometry analysis, and as mammalian LMW-PTP is associated with the cytoskeleton [211, 219, 301], the *E. histolytica* LMW-PTP might also be associated with cytoskeletal components.

The function of human (and mammalian) LMW-PTPs has been studied for much longer than those from other organisms, so many substrates have been identified for those LMW-PTPs. There are many similarities between the *E. histolytica* LMW-PTPs and other LMW-PTPs. There is substantial sequence conservation, with complete conservation of the active site sequence and the DPYY loop which contain the catalytic cysteine, the stabilizing arginine, the second cysteine for disulfide bonding to prevent irreversible oxidation, and the two tyrosines that can be phosphorylated to regulate phosphatase activity.

While there are many similarities between the *E. histolytica* and mammalian LMW-PTPs, there are also a number of important differences which indicate that the *E. histolytica* and mammalian LMW-PTPs have different features, and thus different substrate specificities. For example, the histidine residue at position 45 in the *E. histolytica* LMW-PTP is conserved in all the compared plant, bacterial, and unicellular eukaryotic LMW-PTPs except for *Giardia lamblia* and *Giardia intestinalis*. This histidine residue is not conserved in any of the multicellular animals except for *Drosophila melanogaster* (please see Figure 3-6 for sequence alignments), with that residue being either an asparagine or a glutamic acid. This histidine, which is also conserved in the *Mycobacterium tuberculosis* LMW-PTP, is part of a variable loop which forms a wall of the active site pocket or crevice [232]. The residues that are found in this variable loop affect the pattern of charge distribution around the wall of the crevice, and residues which are in

close proximity to the active site, where the phosphotyrosine binds, may be crucial in determining substrate specificity [275] since it is the local conformation of the active site and associated regions which affect substrate specificity [276]. The phosphorylated tyrosine residue is not just recognized on its own; specific recognition of a phosphotyrosine residue occurs in the context of a peptide sequence, with the flanking residues conferring additional binding affinity of the enzyme to its substrate [279] and interacting with the residues surrounding the active site. The charge distribution would thus be different around the *E. histolytica* active site as compared with the mammalian one, which likely indicates different substrate specificity.

A second difference between mammalian and all other selected plant, bacterial, and unicellular eukaryotic LMW-PTPs including that from *E. histolytica*, is that only in mammalian LMW-PTPs is the Tyr132 residue followed by the consensus sequence G-N-D which conforms to the consensus motif X-N-X for phosphotyrosine recognition by the Grb2 SH2 domain [228, 300]; in the *Entamoeba histolytica* LMW-PTP the corresponding sequence is GGE. This sequence, GGX, is conserved among all the selected compared plant and bacterial LMW-PTPs except for that of *Parabacteroides merdae*. This would indicate that the non-mammalian LMW-PTPs do not interact with Grb2 and are regulated differently in this regard than the mammalian LMW-PTPs. However, since the two tyrosines in the DPYY loop are completely conserved among all LMW-PTPs, with the sole exception of *Saccharomyces cerevisiae*, these two tyrosines likely have an important function, likely regulatory, in all LMW-PTP enzymes, but the regulatory kinases and adaptor proteins involved likely differ between the mammalian LMW-PTPs and those of *Entamoeba*, plants, and bacteria.

A third difference is that the mammalian LMW-PTP is phosphorylated by the cytoplasmic Src family protein kinases Src, Lck, and Fyn at residues Tyr131 and Tyr132 [211, 298, 299]; however, in *E. histolytica*, the most closely related protein kinases to Src family

kinases contain transmembrane regions, and therefore are unlikely to be functionally analogous to Src kinases [363]. It is also possible that the Src-type kinase in *E. histolytica* is highly diverged from its homolog in multicellular eukaryotes [363] but does exist and performs the same function, or the *E. histolytica* LMW-PTP could be phosphorylated on those Tyr residues by a different kinase other than a Src family kinase.

A fourth difference is that the *E. histolytica* LMW-PTP lacks the conserved consensus caveolin-1 binding motif (I/L)XXXXFXXF, corresponding to residues 77-85 in the human and mouse LMW-PTP [228, 258, 364]. However, Caselli *et al* (2007) found that this region may not be required for caveolin-1 interaction, since mammalian LMW-PTP with mutations in this region still could interact with caveolin-1 [258].

Overall, it appears there are some significant differences between mammalian and non-mammalian LMW-PTPs in terms of the presence of certain regulatory residues or residues located close to the active site, which would influence substrate specificity, indicating that the *E. histolytica* LMW-PTP may have different substrates and its function may be regulated differently than the better-studied mammalian LMW-PTPs. It may be that the role that the *E. histolytica* LMW-PTP plays in signal transduction in the cell is very different from that of the mammalian LMW-PTP.

The seven putative substrate proteins that were chosen for cloning, epitope-tagging, and expression were chosen based on the criteria that they were the correct size, had tyrosines in their sequences that were predicted to be phosphorylated, and had a high number of unique peptides generated. There are likely other proteins in the list of the possible substrate proteins identified by mass spectroscopy analysis that were less abundant than the seven that were chosen for initial screening that are perfectly viable LMW-PTP substrates that need to be further investigated.

Possible substrates identified by mass spectrometry analysis for the *E. histolytica* LMW-PTP did not include any obvious receptor kinase proteins or other integral membrane proteins, unlike what has been found in mammalian cells. A difference between our studies and some of those performed in mammalian cells involves how possible substrates (phosphorylated proteins) are generated in the cells before harvesting. The normal level of protein phosphotyrosines in a resting cell is low [212, 213]. Two approaches are used to increase tyrosine phosphorylation of cellular proteins: treatment of cells either with growth factors, or with phosphatase inhibitors. Stimulation of cells with growth factors trigger cascades of tyrosine phosphorylation, but results in tyrosine phosphorylation of the limited number of proteins involved in that signaling pathway, while treatment of cells with phosphatase inhibitors such as vanadate inhibits the intracellular PTPs, resulting in the phosphorylation of a large number of proteins [258]. Many of the studies using mammalian cells stimulated cultured cells with growth factors to increase phosphotyrosine protein levels. For studies of *E. histolytica*, we used pre-incubation of amebae with sodium orthovanadate to inhibit endogenous tyrosine phosphatases to generate a large number of phosphorylated proteins, since it was unknown in which signaling pathway the *E. histolytica* LMW-PTP is involved. The differences in how possible substrate proteins are generated could strongly affect the outcome of an experiment.

Another possible reason why no transmembrane kinases or other receptor or signaling-involved integral membrane proteins were identified as possible substrates could be the detergent choice for solubilizing the amebic lysate. It is standard to use non-ionic detergents such as Triton X-100 or Nonidet P-40 at a 1% concentration in the lysis buffer, which was also used in our studies. However, it appears that transmembrane proteins may not solubilize well using these non-ionic detergents, including receptor kinases like the PDGF receptor, a known substrate of the mammalian LMW-PTP. For example, Coats *et al* (1996) [365] compared two commonly used

lysis buffers in their ability to extract PDGF receptors from a variety of cultured mouse and human cell lines using either Triton-glycerol buffer (1% Triton X-100, 10% glycerol, 1 mM PMSF, 100 μ M sodium vanadate, 1 mM NaF, 20 mM HEPES, pH 7.2) or RIPA buffer (containing 1% Triton X-100 and 0.1% SDS) and found that neither of these buffers recovered more than 50% of the PDGF receptor present, and the addition of 1% SDS was required to solubilize the remaining Triton X-100-insoluble fraction [365]. The insolubility may be in part due to protein-protein interactions among cytoskeleton-associated proteins, and since activation of the PDGF receptor stimulates cell-surface “ruffling”, this suggested the direct interaction of the receptors with cytoskeletal elements [365]. Tyrosine-phosphorylated PDGF receptors were recovered in both the soluble and insoluble fractions after PDGF stimulation, but the tyrosine-phosphorylated levels were consistently higher in the Triton X-100 insoluble fraction than in the soluble fraction [365]. Cirri *et al* (1998) used RIPA lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% Nonidet P-40, 2 mM EGTA, 1 mM sodium orthovanadate, 1 mM phenylmethanesulfonyl fluoride, 10 μ g/ml aprotinin, 10 μ g/ml leupeptin) to extract cytosol-associated proteins, but found that “complete” RIPA lysis buffer had to include 0.5% sodium deoxycholate and 0.1% SDS in order to extract cytoskeleton-associated proteins [284]. 60 mM n-octyl glucoside, a non-denaturing detergent for the solubilization and reconstitution of membrane bound proteins [366] was added to lysis buffer with 1% Triton X-100 to solubilize caveolin-1 for immunoprecipitation [258]. The choice of detergent used in the lysis buffer is extremely important, because if possible substrate proteins are not soluble in the standard buffer, they are removed with the pellet when the cellular lysate is clarified by centrifugation to remove insoluble material. It is standard to clarify the cell lysate by centrifugation before immunoprecipitation or pulldown experiments [227, 243, 282, 283, 301, 334, 335, 336, 367] and this was done in most of our studies as well.

Certain structures in the cell membrane, such as lipid rafts, are generally more difficult to solubilize in buffer. Most plasma membrane-associated proteins such as transmembrane and GPI-anchored proteins cluster in lipid rafts, which are enriched in sterols (especially cholesterol) and sphingolipids such as sphingomyelin and glycosphingolipids [368, 369]. Lipid rafts can serve as platforms in which adhesion and signaling molecules accumulate in a signal-dependent fashion [370, 371]; these would include transmembrane kinases [67], phosphatases, and other proteins regulated by phosphorylation and dephosphorylation. Actin staining and inhibitor studies indicate that lipid rafts are connected to the cytoskeleton, and at least partially depend on it for their formation and/or maintenance [369]. These rafts are resistant to detergents such as Triton X-100 [369, 372]. Caveolae are a subset of lipid rafts [258, 371] and the mammalian LMW-PTP is known to have caveolin-1 as a substrate [213, 258]. Many of the proteins found in caveolae are involved in signal transduction, including receptor protein tyrosine kinases, receptor protein tyrosine phosphatases, and Src family members [258]. The major structural proteins in caveolae are the three isoforms of caveolin which are integral membrane proteins with a transmembrane domain [258, 373] and may function to concentrate signaling molecules within caveolae [258, 374]. The mammalian LMW-PTP associated with caveolae is able to bind and dephosphorylate the caveolin-1 that had been phosphorylated by Src kinases [211, 213, 374].

It is likely that transmembrane kinases and other receptor kinases similar to those found in mammalian cells are located in lipid rafts in *Entamoeba histolytica* and are used for spatial control of signaling molecules and other virulence factors. *Entamoeba histolytica* was found to have raft-like plasma membrane domains by using fluorescent lipid analogs that partition either into raft or non-raft membrane regions [372]. Membrane proteins such as the Gal/GalNAc lectin, an important adhesin and virulence factor of *E. histolytica*, are enriched in these raft regions [372]. The transmembrane kinases have a non-uniform distribution in the membrane and have a

focal staining pattern, suggesting they are part of a multi-molecular signaling complex located in lipid rafts in the plasma membrane [67]. Lipid rafts are important in the ability of *E. histolytica* to bind to extracellular matrix proteins, such as fibronectin and collagen, which trophozoites encounter in the submucosa of the human colon if they are able to breach the mucus-secreting colonic epithelium [370]. Since virulence-related integral membrane proteins and signaling molecules are enriched in lipid rafts, and these rafts tend not to solubilize well in non-ionic detergents used in standard cell lysis buffers, a whole subset of possible substrate proteins could have been missed in many of the LMW-PTP studies previously undertaken.

In conclusion, we have structures for the *E. histolytica* LMW-PTP that show changes that occur due to the binding of substrate and product, and its overall structure is the same as that found for other LMW-PTPs in other species. The active site and DPYY loop that participates in catalysis are completely conserved. The *E. histolytica* wild-type recombinantly expressed LMW-PTP has phosphatase activity, as measured with the generic phosphatase substrate *p*NPP, and the recombinant substrate-trapping Cys7 to Ser7 mutant protein is catalytically dead. The *E. histolytica* LMW-PTP is most closely related to LMW-PTPs in selected plant species and *Bacteroides* species, and least related to mammalian LMW-PTPs, which have been the most studied. There are a number of differences between the *E. histolytica* and mammalian LMW-PTPs in their sequences, which indicate they may be regulated differently and have different substrate specificities. We generated rabbit sera that recognize the native LMW-PTP and which can be used to immunoprecipitate it. Epitope-tagged wild-type and substrate-trapping mutant LMW-PTPs were inducibly overexpressed and used to visualize the LMW-PTP cellular location via confocal microscopy using antibody against the epitope tag. Substrate-trapping of putative substrate proteins was performed using recombinant mutant LMW-PTP covalently bound to beads incubated with sodium orthovanadate-treated *E. histolytica* lysate, and putative substrate

proteins were identified via mass spectrometry analysis. Seven of the proteins identified by mass spectrometry analysis that were the correct size to be in the chosen bands that had the most unique associated peptides and had predicted phosphotyrosine sites were cloned with epitope tags in order to overexpress them in trophozoites to test for protein-protein interactions. None of these seven is predicted to have a transmembrane domain, and it may be that signaling molecules that are integral membrane proteins, such as receptor protein kinases such as PDGF, which has been found to be a LMW-PTP substrate in mammalian cells, were not solubilized sufficiently in standard lysis buffers using the non-ionic detergents Triton X-100 or Nonidet P-40. Alternately, the amebic LMW-PTP could have exclusively soluble (cytosolic) substrates. It may also be that there are other LMW-PTP substrates not identified by mass spectrometry analysis because they are expressed at too low a level to be detected.

Identification of the substrates for this LMW-PTP is important because in mammalian cells, LMW-PTPs play roles in the control of cell proliferation, motility, and adhesion through dephosphorylation of growth factor receptors [212, 228, 242, 281, 282, 284, 285, 307], cytoskeleton-associated proteins [212, 219, 222, 228, 289, 307], and the transcription factor STAT5 [212, 222, 291, 292]. Once we have been able to confirm that identified putative substrates of the LMW-PTP are truly substrates, if we know what the function of those proteins is, we can predict in which signaling pathways the *E. histolytica* LMW-PTP is involved, and if those ultimately play a role in proliferation, motility, adhesion or phagocytic ability of this parasite. Since the *Entamoeba histolytica* LMW-PTP is more closely related to bacterial and plant LMW-PTPs, it may be more functionally similar to those of bacteria than to those of mammals, and likely will have some novel substrates.

Future Directions

Since LMW-PTPs are important in regulating growth, motility, and adhesion in mammalian cells, the *E. histolytica* LMW-PTP may play a similar role. Overexpression of LMW-PTP in mammalian cells can result in dysregulation of its substrates and in uncontrolled growth of cells, so this protein may play a role in *E. histolytica* virulence in enhancing its growth and motility. There are still a number of possible studies that can be performed for this project. Confirmation that the identified putative LMW-PTP substrates are actual substrates still needs to be accomplished; once this is known, we may be able to predict in which signaling pathways the *E. histolytica* LMW-PTP is involved. Seven of the putative substrate proteins have been cloned with epitope tags. The next step with these is to transfect amoebae and overexpress these proteins to assay for their interaction with the substrate-trapping mutant LMW-PTP. This could be performed either *in vitro* or *in vivo*. For example, lysates from epitope-tagged putative substrate protein overexpressers pre-treated with the phosphatase inhibitor sodium orthovanadate, and lysates from epitope-tagged substrate-trapping LMW-PTP overexpressers could be combined, incubated together, and then the mutant LMW-PTP immunoprecipitated with its trapped substrate(s). There were total of eighty-four proteins that were identified by mass spectrometry analysis, but even when the ones of the incorrect sizes are eliminated, there are still many of them that could be investigated further to see if they are good candidates for being substrates. For example, since kinases and phosphatases are often regulated themselves by phosphorylation and dephosphorylation, these would be good candidates for analysis.

To identify that a putative substrate is an actual substrate of the LMW-PTP, we need to show an interaction between the two, so we may have to overexpress additional substrate candidates with epitope tags, or express recombinant putative substrate protein for *in vivo* or *in vitro* experiments, or have antibodies made against the recombinant proteins. Once we have

shown putative substrates do interact with the substrate-trapping LMW-PTP and are true substrates, this should allow us to begin mapping the LMW-PTP signal transduction pathways in *Entamoeba histolytica*, what other phosphatases and kinases may be involved, and what role they play in virulence by phenotypic testing (upregulation or downregulation of motility, phagocytosis, cell killing etc).

Some of the possible LMW-PTP substrates, if they are located in lipid rafts, may be insoluble in the lysis buffer containing 1% Triton X-100 or 1% Nonidet-P40, but might be solubilized in buffer with 1% SDS [365]. To identify these, the pulldown experiments could be repeated using the substrate-trapping recombinant mutant protein attached to beads, but beads could be incubated with lysate that contains 1% SDS in the buffer to solubilize integral membrane proteins such as receptor protein tyrosine kinases, or incubated with buffer containing purified recombinant putative substrate protein. Some of the *E. histolytica* transmembrane kinases [67] appear to be receptor protein kinases and could possibly be LMW-PTP substrates, since many of the mammalian LMW-PTP substrates, such as PDGFR [228, 282, 284], insulin receptor [228, 281], fibroblast growth factor receptor [212, 228, 285] and EphA2 receptor tyrosine kinase [228, 288] are receptor protein tyrosine kinases. Another way to perform this experiment in vivo could be to overexpress both the mutant epitope-tagged substrate-trapping LMW-PTP (and the wild-type as a control) and the epitope-tagged transmembrane kinase in the same amebic cell (or alternately, overexpress in two different transfectant lines and then incubate the lysates together) and then do either a pulldown or an immunoprecipitation to check for an interaction. The level of possible LMW-PTP substrates (phosphorylated proteins) could possibly be additionally increased. It seems to be standard to pretreat cells for 30 minutes before harvesting with 1 mM vanadate, but amebae have been treated with as much as 100 mM orthovanadate before harvesting them for lysate [336].

Another approach is that since the epitope-tags on the wild-type and mutant LMW-PTP clones include 6xHis tags, the wild-type and mutant LMW-PTP over-expressing transfectants could be induced, harvested, and the lysate incubated with nickel beads to pull down the substrate-trapping LMW-PTP protein and any trapped substrate. Another possible way of performing these experiments could be to express the mutant LMW-PTP as a GST fusion protein and bind it to glutathione-agarose beads either before or after the addition of lysate to trap substrates [367]. An additional *in vitro* approach could be to express recombinant putative LMW-PTP substrate proteins in *E. coli*, purify the proteins, treat the recombinant proteins with the proper kinase to phosphorylate them, and assay if the wild-type recombinant LMW-PTP can dephosphorylate the kinased putative substrate proteins.

Studies could be performed to measure if overexpression of the wild-type LMW-PTP and substrate-trapping mutants alters the phenotype of the transfected trophozoites. For example, when wild-type LMW-PTP is overexpressed in NIH 3T3 cells (a mouse embryonic fibroblast cell line), it leads to a strong increase in cell motility and chemoattraction, as assayed by cell migration through 8 μm pore transwell inserts toward medium containing serum, and overexpression of the mutant substrate-trapping LMW-PTP had the opposite effect [307]. An increase in trophozoite migration using the same transwell inserts is seen when the dominant-positive calcium-insensitive mutant URE3-BP transcription factor is inducibly overexpressed in trophozoites [69] and these had an enhanced ability to invade intestinal epithelium and liver tissue [77]. It may be that the *E. histolytica* LMW-PTP does not directly interact with URE3-BP, but could be part of the same signaling pathway. Both motility and the ability to bind to a surface are required for amebic virulence. In order to invade host intestinal epithelium, *E. histolytica* trophozoites must first adhere to its surface and then penetrate the intestinal epithelium. The

regulation of trophozoite motility may be a factor in the ability of this parasite to cause invasive disease or colonize the host asymptotically.

It may be that the activity of the *E. histolytica* LMW-PTP is regulated by phosphorylation of the Tyr124 and Tyr125 residues (equivalent to the mammalian Tyr131 and Tyr132) found in the DPYY loop. In the mammalian LMW-PTP, phosphorylation of Tyr131 induces a 25-fold increase in LMW-PTP activity [211, 298, 299] while phosphorylation at Tyr132 recruits Grb2 to bind to the enzyme [211, 228, 298]. These tyrosines could be mutated to residues that are incapable of being phosphorylated, and the mutant proteins assayed either in vitro or in vivo to study if these tyrosines play the same or a similar role in the regulation of LMW-PTP enzymatic activity or in recruiting adaptor proteins to bind to the LMW-PTP.

Another study that could be done is to repeat the enzyme activity experiment with the recombinant wild-type LMW-PTP, but using different amounts of *p*NPP substrate, so V_{max} , K_m , and k_{cat} could be calculated for comparison with other LMW-PTPs.

Since the confocal imaging of the cellular location of the wild-type and mutant LMW-PTPs was performed with overexpressed epitope-tagged protein and which may not be representative of the endogenous cellular location, the confocal experiment could be repeated, but with less tetracycline added (perhaps 1 μ g/ml or 0.5 μ g/ml rather than the 10 μ g/ml used originally) to induce protein expression of the LMW-PTP, so that sufficient LMW-PTP would be overexpressed to visualize its location but not so much so that it could be improperly localized.

There are two LMW-PTP genes in *Entamoeba histolytica* (GenBank: XM_651267 and XM_648265) which are the same length and are identical except for a single residue change at position 85 from an alanine to a valine. Only one of them was cloned (XM_651267) and had its protein (XP_656359) used for this study. It is interesting that there are two *E. histolytica* LMW-PTP genes, while the nonpathogenic related species *Entamoeba dispar* and the reptile parasite

Entamoeba invadens both have a single LMW-PTP gene. Even though there is only one residue change between the two proteins, it could be that the two genes are differently regulated, expressed, localized, or have different enzyme kinetics. This may be something to be further investigated.

Finally, one other approach to take is that we could also knock down expression of the LMW-PTP using the shRNA system as described in Chapter 2; select sequences in the LMW-PTP genes for targeting, design oligos and perform PCR to clone constructs for expression of hairpins, transfect amebae, and assay if there is protein knockdown of the LMW-PTP and if there is any resulting phenotypic difference in the knockdown transfectants.

Figure 3-1. The protein tyrosine phosphatase superfamily. All protein tyrosine phosphatases (PTPs) contain the conserved motif CX₅R, comprising the enzyme active site, and are divided into three class types based on the substrate specificities and structures of the members [212, 219, 220, 221, 222, 223]. They all share the same catalytic mechanism, which requires a conserved cysteine residue in the active site [212, 220, 221, 222, 227]. Otherwise, these three class types have little sequence homology [203, 220; 222, 223, 227]. Type I PTPs include the dual-specificity phosphatases, which dephosphorylate phospho-serine, -threonine and -tyrosine residues; and the classical PTPs, which are subdivided into either a non-receptor cytosolic group or a receptor-like group [203, 212, 220, 221, 222]. Subgroups in the dual-specificity group include MAPK phosphatases, Slingshots, Cdc14s, PTENs, myotubularins, and atypical dual-specificity phosphatases [212, 222]. Subgroups in the classical non-receptor group include PTP1B, SHPs, TC-PTP, PTP-PEST, and PTPL1 [212, 219, 222]. Subgroups in the classical receptor-like group include CD45, LAR, PTP α , RPTP σ , and RPTP γ [212, 219, 222]. The Type III PTP class contains the Cdc25 protein tyrosine phosphatases: Cdc25A, Cdc25B, and Cdc25C [212, 221, 222]. The Type II PTP class also contains only one subgroup: the family of low molecular weight protein tyrosine phosphatases (LMW-PTPs) [212, 221, 222, 223]. LMW-PTPs have their active site close to the N-terminus of the protein; in the other PTPs the active site is towards the C-terminus [219, 222]. Figure adapted from Samet and Tal (2010) [212].

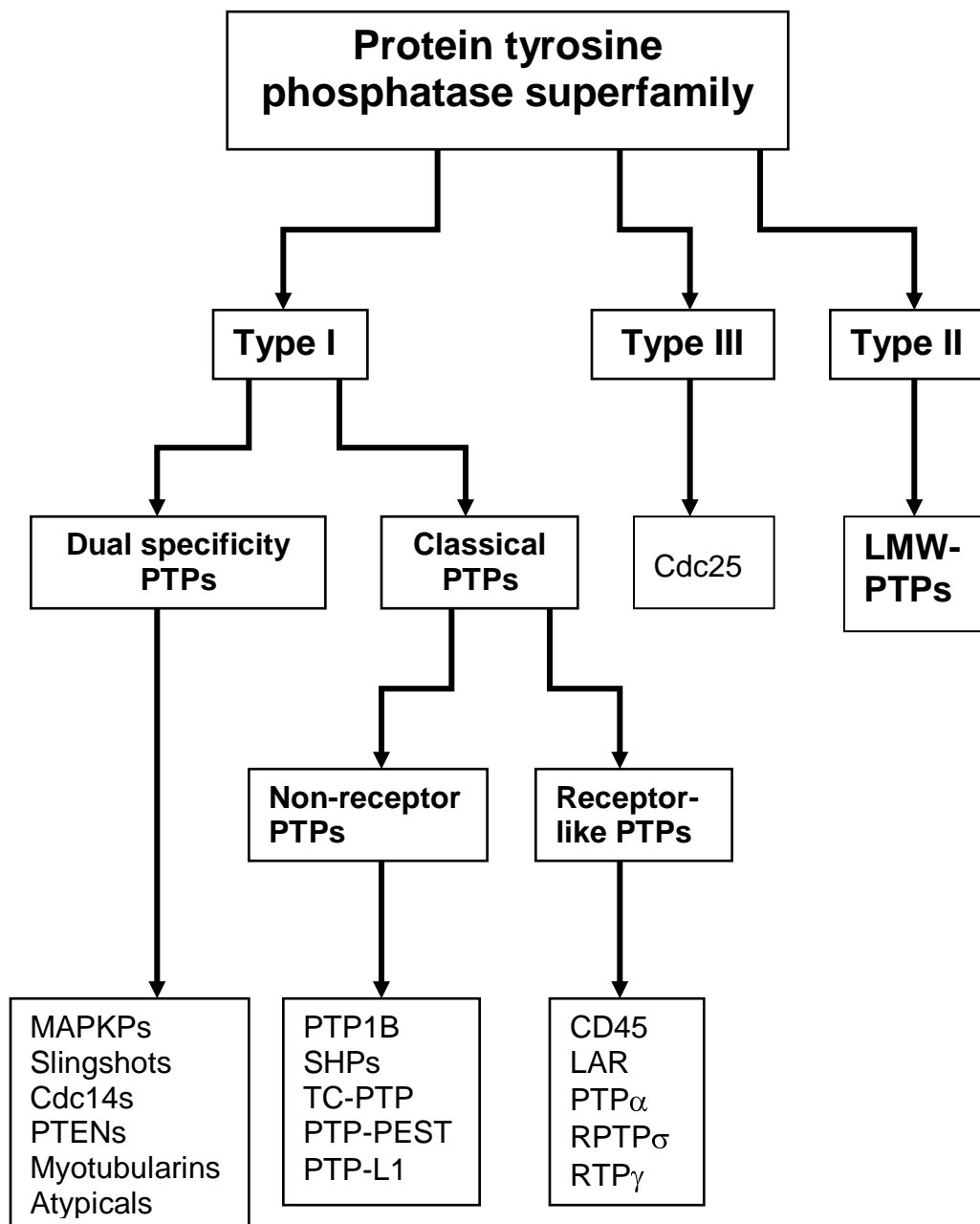


Figure 3-2. The catalytic mechanism of protein tyrosine phosphatases. Three key structural features are conserved among all protein tyrosine phosphatases: the catalytic cysteine and arginine in the active site, along with the general acid/base containing surface loop, which contains the conserved aspartic acid residue [224, 225, 228]. (A) The cysteine, arginine, and aspartic acid residues for the active site of the *E. histolytica* LMW-PTP are shown. (B) A phosphotyrosine residue of a substrate protein enters the active site of the phosphatase. (C) The arginine residue stabilizes the transition state during hydrolysis, and the aspartic acid acts as a general acid by providing a proton to the oxygen of the tyrosine residue leaving group. (D) The thiol side chain of the active site catalytic cysteine serves as a nucleophile and attacks the phosphoryl group of the substrate, forming a thiol-phosphate intermediate, and the dephosphorylated tyrosine residue is released. (E) The phosphoryl group is now bound to the thiol side chain of the active site catalytic cysteine. (F) A water molecule enters the active site. (G) The aspartic acid residue acts as a general base and accepts a proton from the water molecule. (H) The carboxyl group on the aspartic acid has been regenerated, allowing the OH⁻ ion from the water to nucleophilically attack the phosphorus of the thiol-phosphate intermediate, and the electrons return to the thiol side chain. The inorganic phosphate is released and the enzyme active site has been returned to its resting state (A). Figure adapted from Ramponi and Stefani (1997) [203]. Also see Figure 3-13 for the structure of the active site and diagram of the active site residues for the *E. histolytica* LMW-PTP.

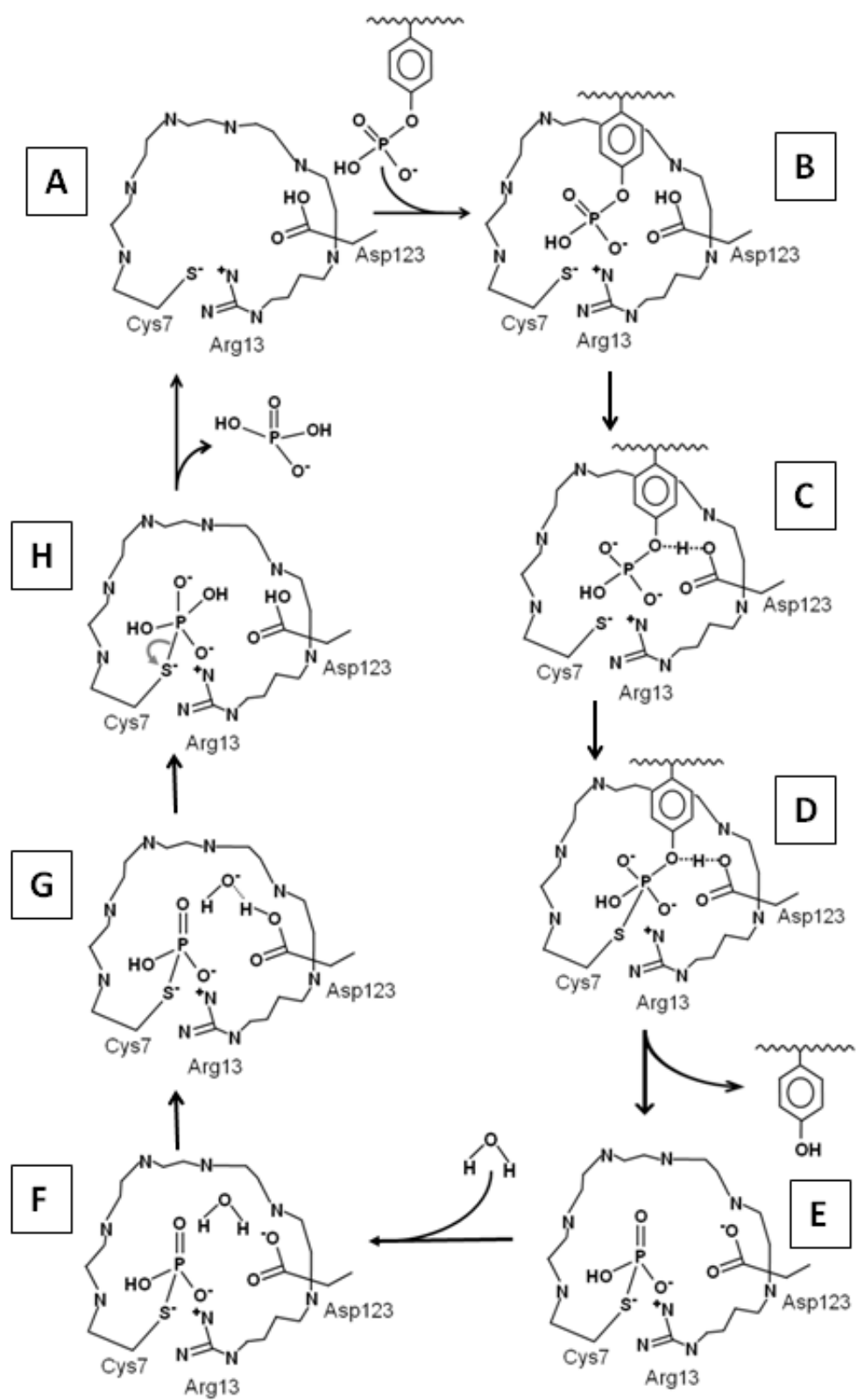


Figure 3-3. Protein sequence alignments comparing the *Entamoeba histolytica* LMW-PTPs with selected plant LMW-PTPs.

Identical amino acids are highlighted in yellow, and conservative substitutions are highlighted in light blue. The conserved active site of each LMW-PTP is highlighted in pink, and the conserved DPYY loop is highlighted in red.

The amino acid sequence of the *E. histolytica* LMW-PTP protein XP_656359 was aligned with other related proteins using BLAST [311] and LALIGN [312]. Table 3-1 lists species abbreviations, LMW-PTP accession numbers, and percent identity and similarity to the *E. histolytica* LMW-PTP protein XP_656359 for each plant LMW-PTP sequence. Figure 3-7 shows a phylogenetic tree for all selected LMW-PTPs.

Ent his 1		-----	
Ent his 2		-----	
Sorg bicol	1	MPGAVAPAPTLAPTRSSAYTSAAAASTLSPNPTSRGDPLT	40
Zea mays 1	1	MPGAVA----LAPTFAAASSSAAAASTLAPNPTSRGDPLI	36
Zea mays 2		-----	
Oryza sat	1	MAGAAAASSAACSSPAPNPTTRYPVRRRVPRPPLVAAASR	40
Hord vulg	1	MVGATV-----ASTSTAAAYSSLTHPTAR-DPFK	28
Arab thal		-----	
Vit vinif		-----	
Pop tricho		-----	
Ricin comm		-----	
Ent his 1		-----	
Ent his 2		-----	
Sorg bicol	41	LSSCRNAPASATPLVPLAATRRHVGRGFRGAPL-VASPSPH	79
Zea mays 1	37	IRLCRNAPASATPLVPLVATRRYVRGCRGAAL-VASP--H	73
Zea mays 2		-----	
Oryza sat	41	HCTASPFPTTISISSSGLGQARPRDPFLNPRLRFA-----	75
Hord vulg	29	RRTTH-----PLV--AATNRHVHGRTVTVSLRQIHPRDH	59
Arab thal		-----	
Vit vinif		-----	
Pop tricho		-----	
Ricin comm		-----	
Ent his 1	1	-----MKLL	4
Ent his 2	1	-----MKLL	4
Sorg bicol	80	HARNPRLRFASAAAAEG-MAAEASTADAAAAAEAKPFAVL	118
Zea mays 1	74	HAPNPRLRFASAA--EG-MAAEASTAGAASAAEAKPFAVL	110
Zea mays 2	1	-----MAAEASTAGAASAAEAKPFAVL	22
Oryza sat	76	-----AAAAAA-EGEMAAEASAAAAEGA-QAKPFSVL	120
Hord vulg	60	FANPPRPV FVASAAEAEEMAAEASTPAPAEA-SAKPFAVL	105
Arab thal	1	-----MATPPPTQ-SS---ETKPFSVL	18
Vit vinif	1	-----MASGPSIDT----QSKPFSVL	17
Pop tricho	1	-----MASS--TETKPFAVL	13
Ricin comm	1	-----MASSTETETKPFSVL	15
Ent his 1	5	FVCLGNICRS PAAEAVMKKVIQNHHLTEKYICDSAGTCSY	44
Ent his 2	5	FVCLGNICRS PAAEAVMKKVIQNHHLTEKYICDSAGTCSY	44
Sorg bicol	119	FVCLGNICRS PAAEAVFRTLVSKRGLDSKFLIDSAGTIGY	158
Zea mays 1	111	FVCLGNICRS PAAEAVFRTLVSKRGLDSKFLIDSAGTIGY	150
Zea mays 2	23	FVCLGNICRS PAAEAVFRTLVSKRGLDSKFLIDSAGTIGY	62
Oryza sat	121	FVCLGNICRS PAAEAVFRNLVSKRGLSKFLIDSAGTIGY	160
Hord vulg	106	FVCLGNICRS PAAEAVFRNLVSKRGLKSKFQIDSAGTIGY	145
Arab thal	19	FVCLGNICRS PAAEGVFRDIVKKRGLDSKFNIDSAGTIDY	58
Vit vinif	18	FVCLGNICRS PAAEGVFTDAVKRGLDSKFKIDSAGTINY	57
Pop tricho	14	FVCLGNICRS PAAEGVFTDIVNKRGLDSKFKIDSAGTINY	53
Ricin comm	16	FVCLGNICRS PAAEGVFRDIVKKRSLDSKFNIDSAGTINY	55

Figure 3-3 is continued on page 185.

Figure 3-3, continued from page 184.

Ent his 1	45	HEGQQADSRMRKVGKSRGYQVDSISRPPVSSDFKNFDYIF	84
Ent his 2	45	HEGQQADSRMRKVGKSRGYQVDSISRPPVSSDFKNFDYIF	84
Sorg bicol	159	HEGNKADSRMRAASKKRGIEVTSISRPIRPSDFRDFDLIL	198
Zea mays 1	151	HEGNKADSRMRAASKKRGIEVTSISRPIKPSDFRDFDLIL	190
Zea mays 2	63	HEGNKADSRMRAASKKRGIEVTSISRPIKPSDFRDFDLIL	102
Oryza sat	161	HEGNKADSRMRSASKKRGIEVTSISRPKPSDFRDFDLIL	200
Hord vulg	146	HEGNKADSRMISTSKKRGIEVTSISRPIKPSDFRDFDLIL	185
Arab thal	59	HEGNMADPRMRSAAKRRGIEITSLSRPIKASDFRDFDLIL	98
Vit vinif	58	HEGNLADPRMRAASKRRGIEITSISRPIQPSDFRDFDLIL	97
Pop tricho	54	HEGNPADSRMRAASKRRGIEITSISRPIRPSDFRDFDIIL	93
Ricin comm	56	HEGNPADPRMRAASKRRGIEITSISRPIRPSDFRDFDIIL	95
Ent his 1	85	AMDNDNYEYEL-----DRCPEQYKQKIFKMVDFCTT	115
Ent his 2	85	VMDNDNYEYEL-----DRCPEQYKQKIFKMVDFCTT	115
Sorg bicol	199	AMDRQNYEDIILNSFERWRRKEPLPENAPDKVKLMCSYCKR	238
Zea mays 1	191	AMDRQNYEDIILNSFERWRRKEPLPSAPNKVKLMCSYCKQ	230
Zea mays 2	103	AMDRQNYEDIILNSFERWRRKEPLPSAPNKVKLMCSYCKQ	142
Oryza sat	201	AMDRQNYEDLLSSFERWRHREPLPSGPKVKVKLMCSYCKR	240
Hord vulg	186	AMDRQNYEDIILSSFDRWQHKETLPSGPKVKVKLMCSYCKR	225
Arab thal	99	AMDDQNKEDIILKAYNVWKARGNFPDADKVKVKLMCSYCKK	138
Vit vinif	98	AMDKQNRDDIMEAFDRWRFRETLPADAHKKVKVKLMCSYCKK	137
Pop tricho	94	AMDNQNRHEHIMEAFNKWKFKETLPDADAHKKVKVKLMCSFCKK	133
Ricin comm	96	AMDKQNREDIILEAFNRWKVRDSLPPDAHEKVKVKLMCSYCKK	135
Ent his 1	116	IKTTEVPDPYYGGEKGFHRVIDILEDACENLIIKLEEGKLIN	157
Ent his 2	116	IKTTEVPDPYYGGEKGFHRVIDILEDACENLIIKLEEGKLIN	157
Sorg bicol	239	HTESEVPDPYYGGPQGFVKVLDILEDACESSLDSIV---ASISG	279
Zea mays 1	231	HTESEVPDPYYGGPQGFVKVLDILEDACESSLDSIVANNASISG	274
Zea mays 2	143	HTESEVPDPYYGGPQGFVKVLDILEDACESSLDSIVANNASISG	186
Oryza sat	241	HTESEVPDPYYGGPQGFVKVLDILEDACESSLDSIVAENASVFA	268
Hord vulg	226	HTESEVPDPYYGGAKGFVKVLDILEDACESSLDSIMAENENISA	262
Arab thal	139	HNDKFVDPDPYYGGAQGFVKVLDILEDACESSLDSITAQS	177
Vit vinif	138	HDETEVPDPYYGGPQGFVKVLDILEDACESSLDSILAENSHLLDS	182
Pop tricho	134	HDETEVPDPYYGGPQGFVKVLDILEDACESSLDSILAENK	173
Ricin comm	136	HDETEVPDPYYGGPQGFVKVLDILEDACESSLDSIVA	172

Figure 3-4. Protein sequence alignments comparing the *Entamoeba histolytica* LMW-PTPs with selected bacterial LMW-PTPs.

Identical amino acids are highlighted in yellow, and conservative substitutions are highlighted in light blue. The conserved active site of each LMW-PTP is highlighted in pink, and the conserved DPYY loop is highlighted in red.

The amino acid sequence of the *E. histolytica* LMW-PTP protein XP_656359 was aligned with other related proteins using BLAST [311] and LALIGN [312]. Table 3-2 lists species abbreviations, LMW-PTP accession numbers, and percent identity and similarity to the *E. histolytica* LMW-PTP protein XP_656359 for each bacterial LMW-PTP sequence. Figure 3-7 shows a phylogenetic tree for all selected LMW-PTPs.

Ent his 1	1	-----M K L L F V C L G N I C R S	PAAEAVM K K V I	25
Ent his 2	1	-----M K L L F V C L G N I C R S	PAAEAVM K K V I	25
Nost punct	1	---MTIDKGQKTSMPY K L L F V C L G N I C R S	PSAENIM N H L I	37
Lynghya sp	1	-----MPY Q L L F V C L G N I C R S	PSAENIM N H L I	27
Cyano sp	1	-----MPY K L L F V C L G N I C R S	PSAENIM N H L I	27
Arthro max	1	-----MSY R L L F V C L G N I C R S	PSAENIM N H L I	27
Prochlo mar	1	-----MIRK V L F I C L G N I C R S	PAAEAI F L H H L	27
Bdello bact	1	MTRQNNSPKIIQMKSQ K L L F V C L G N I C R S	PTAEAV A A H L I	40
Bact caccae	1	-----M K K I L F V C L G N I C R S	STAE G V M L H L I	26
Bact theta	1	-----M K K I L F V C L G N I C R S	STAE G V M L H L I	26
Bact fragil	1	-----M K K K I L F V C L G N I C R S	STAE G V M L H L I	27
Parab merd	1	-----M M E E K K G E Y K I L F V C L G N I C R S	PSAEAVM K K L V	33
Vibrio chol	1	-----M Q K V L V V C M G N I C R S	PTAEAV L R--A	24
Pseudo aeru	1	-----M R V L F V C L G N I C R S	PTAE G V F R R K V	25
Ent his 1	26	QNHHLTEKYICDSAGTCSYHEGQQADSRMRKVGKSR-	GYQ	64
Ent his 2	26	QNHHLTEKYICDSAGTCSYHEGQQADSRMRKVGKSR-	GYQ	64
Nost punct	38	EQAGLSDAIICDSAGTSSYHVGSPDRRMSAAAATKLGK		77
Lynghya sp	28	ELAHLGRSILCDSAGTSSYHIGSPDARMTKAASSR-	GIL	66
Cyano sp	28	EQAGLSEKILCDSAGTSNYHIGEPDRRMNAAASGR-	GIP	66
Arthro max	28	EEANLSDRIVCDSAGTSGYHIGSAPDRMTMAAKKR-	GIV	66
Prochlo mar	28	REIDLINEFIVDSAGTGGWHVGKADSRMRSAALS-	GIV	66
Bdello bact	41	KQRDLP--WVVDSDAGTSGAHDGEMADPRSILHGERR-	GYD	77
Bact caccae	27	EEAGLEKEFVIDSAGILSYHQGELPDSRMRAHAARR-	GYQ	65
Bact theta	27	KEAGLEKEFVIDSAGILSYHQGELPDSRMRAHAARR-	GYQ	65
Bact fragil	28	KEAGLEKEFVIDSAGILAYHQGELPDSRMRAHAARR-	GYE	66
Parab merd	34	QDAGLDGRIKIDSAGIIGYHAGEKADPRMRSHAARR-	GYK	72
Vibrio chol	25	KAAQLKVDVEIDSAGTIGYHQGNPPDARSKAAGEKR-	GYD	63
Pseudo aeru	26	EEAGLASRIHVDSAGTAGWHAGKAPDQRTRLAAQRR-	GYD	64
Ent his 1	65	VDSI-SRPVVSDFKNFDYIFAMDNDNYEEL--LD-RC--		98
Ent his 2	65	VDSI-SRPVVSDFKNFDYIFVMDNDNYEEL--LD-RC--		98
Nost punct	78	LRGQ-ARQFQKSDFDLILAMDQENYENILTLD-RT--		113
Lynghya sp	67	LKGQ-ARQFGREDFENFDLILAMDYDNYESIVRLD-SA--		102
Cyano sp	67	LKGG-ARQFQPSDFENFDLILAMDRENYRDI--RR-LD--		100
Arthro max	67	LKGS-ARQFTASDFDNFDLILAMDQDNYQQILYLD-R---		101
Prochlo mar	67	IESR-ARQISLNDFNFDLILTMDQSNLDDV--NS-LA--		100
Bdello bact	78	LTSI-SRAIRESDYDFDWILAMDASNLEHL--RQ-RC--		111
Bact caccae	66	LVHR-SRPVTEDFYNFDLIIGMDDRNIEDL--KD-KAPS		99
Bact theta	66	LVHR-SRPVTEDFYNFDLIIGMDDRNIDDL--KE-KAPS		99
Bact fragil	67	LVHR-SRPVTEDFYNFDLIIGMDDRNMDL--KE-KAPS		100
Parab merd	73	LDSV-SRPVCTEDFFDFDLIIGMDNRNIDDL--KR-KA--		107
Vibrio chol	64	FSGIKARKIRDEDFVKFDWILAADQENLAEL--KA-RC--		98
Pseudo aeru	65	LSALRGRQVGIEDFSRHDLILAMDLSNLADL--ETLRA--		100

Figure 3-4 is continued on page 188.

Figure 3-4, continued from page 187.

Ent his 1	99	PE--Q-YKQKIFKMVDFCTTIKTTEVPDPYYGGGKGFHRV	135
Ent his 2	99	PE--Q-YKQKIFKMVDFCTTIKTTEVPDPYYGGGKGFHRV	135
Nost punct	114	-E--Q-YQHKVRLMCEFCRHTLKEVDPDYYGGQDGFNQV	149
Lyngbya sp	103	-R--Q-YQDKVKLMCEFCRHHNVQEVDPDYYGGPEGFNQV	138
Cyano sp	101	PSG-Q-YQDKVRLICEFATHYSIKEVDPDYYGGPEGFNQV	138
Arthro max	102	GK--K-YGDKVKLMCDFCSHHTLKEVDPDYYGGPEGFNQV	138
Prochlo mar	101	KELNISYKAKVKPLLEYATNTDLVEVDPDYYGGGKGFEDV	140
Bdello bact	112	PD--KTLLDKISLVTDYCSFKVKGVPDPYYGGVDGFDHV	149
Bact caccae	100	PE--E-WK-KIHRMTEYCNRI PADHVPDPYYGGAE GFEYV	137
Bact theta	100	TE--E-WK-KIHRMTEYCTRI PADHVPDPYYGGAE GFEYV	137
Bact fragil	100	PA--E-WK-KIHRMTEYCTRI PADHVPDPYYGGAE GFEYV	138
Parab merd	108	PDL-E-SVEKIHQMTEYSQNKLYDHVPDPYYSGAE GFEYV	145
Vibrio chol	99	PQ--S-HQHKLSLMLSHSDS-EYQETDPDYYGGERGFELV	134
Pseudo aeru	101	GR--G-AAELDLFLRRYGS--ERDEVDPDYYGGEE GFEQV	135
Ent his 1	136	IDILEDACENLI IKL--EEGKLIN	157
Ent his 2	129	IDILEDACENLI IKL--EEGKLIN	157
Nost punct	150	IDLLIDACEGLLTKVKSEEF	169
Lyngbya sp	139	IDILLDACEGLLEYVQKQPQKAV	162
Cyano sp	139	IDLLLDACSGLL-----EEVKLKGLQEH	161
Arthro max	139	IDLLLDACEGLLEFVNVQPSVN	160
Prochlo mar	141	LNLLENAIEGLIRDIK	156
Bdello bact	150	LDILEDACGLIDKV--QDRSDKI	171
Bact caccae	138	LDILEDACSGLLISLT-QDN	156
Bact theta	138	LDVLEDACAGLLTSLT-QDN	156
Bact fragil	139	LDILEDACAGLLTSLT-QDS	157
Parab merd	146	LDLLEDACAGLLDELV-QFISSNPDN	169
Vibrio chol	135	LDLVEDAAEQFLKLL--QQQGQR	155
Pseudo aeru	136	LDLIEHACDRLLIEAKGRL	154

Figure 3-5. Protein sequence alignments comparing the *Entamoeba histolytica* LMW-PTPs with selected unicellular eukaryotic LMW-PTPs.

Identical amino acids are highlighted in yellow, and conservative substitutions are highlighted in light blue. The conserved active site of each LMW-PTP is highlighted in pink, and the conserved DPYY loop is highlighted in red.

The amino acid sequence of the *E. histolytica* LMW-PTP protein XP_656359 was aligned with other related proteins using BLAST [311] and LALIGN [312], except for the alignment with the *Entamoeba invadens* LMW-PTP, which was aligned using the amoeba database at AmoebaDB [309]. Table 3-3 lists species abbreviations, LMW-PTP accession numbers, and percent identity and similarity to the *E. histolytica* LMW-PTP protein XP_656359 for each unicellular eukaryotic LMW-PTP protein sequence. A phylogenetic tree comparing all selected LMW-PTPs from all organisms is shown in Figure 3-7.

Ent his 1	1	-----MKLLFV CLGNICRS PAAEAVMKKVIQNHHLTE	32
Ent his 2	1	-----MKLLFV CLGNICRS PAAEAVMKKVIQNHHLTE	32
Ent dispar	1	-----MKLLFV CLGNICRS PAAEAVMKKVIQNHHLTE	32
Ent invad	1	-----MKILFV CLGNICRS PAADGVMKKLVKSRGLSA	32
Gia lam P15	1	-----MKILFV CLGNICRS PMAHGVMQHLVEKAGLSS	32
Giard intes	1	-----MKILFV CLGNICRS PMAHGVMQHLVEKAGLSN	32
Dictyo disc	1	MSADSKNQKKVLFV CLGNICRS TMAEIVLRGLVHSRGILD	40
Klu lactis	1	----MSEQVSVAFV CLGNICRS PMAEAVFKHVVEKGLQD	36
Sacch cerev	1	-MTIEKPKISVAFT CLGNFCRS PMAEATFKHEVEKANLEN	39
Schiz pombe	1	----MTKNIQVLFV CLGNICRS PMAEAVFRNEVEKAGLEA	36
Ent his 1	33	KYI-CDSAGTCSYHEGQQADSRM-----RK-VGKSR	61
Ent his 2	33	KYI-CDSAGTCSYHEGQQADSRM-----RK-VGKSR	61
Ent dispar	33	KYT-CDSAGTCSYHEGQQADSRM-----RK-VGKSR	61
Ent invad	33	KYI-IDSAGTAGYHSGDQPD SRM-----RA-AGLGR	61
Gia lam P15	33	QVQ-VDSCGMFGHP EGTST EYGT-----QKAVADNL	62
Giard intes	33	QIQ-VDSCGMFGHP EGTST EYGT-----QKAITDKL	62
Dictyo disc	41	DFQ-IDSAGTSSYHIGDTPDPRTVQSCNQNMGR-AISEES	78
Klu lactis	37	KFKRIESFGTAGYHVGESPD SRS-----AS-TCRKN	66
Sacch cerev	40	RFNKIDSFGTSNYHVGE SPDHRT-----VS-ICKQH	69
Schiz pombe	37	RFDTIDSCGTGAWHVGNRPDPRT-----LE-VLKKN	66
Ent his 1	62	GYQVDSI----SRPVVSSDFKNFDYIFAMDNDNYEEL---	94
Ent his 2	62	GYQVDSI----SRPVVSSDFKNFDYIFVMDNDNYEEL---	94
Ent dispar	62	GYQVDSI----SRPVVSSDFKNFDYIFAMDNDNYEEL---	94
Ent invad	62	GYTFDHI----SRQVKQSDFSNFDIIVAMDSDSNYS DL---	94
Gia lam P15	63	GISFSKK----SRHWRKSDYDDFDLILAMDKSNFQDI---	95
Giard intes	63	GISFSKK----SRHWRNSDYDGFDLILAMDKSNFQDI---	95
Dict disc	79	LKHFKSIPLHRARQFTDEDFSKFDYIFAMDES NLSNIKKV	118
Klu lactis	67	GVPI SHS----AQQIKSKHFKEFDYIICMDES NLRNL---	99
Sacch cerev	70	GVKINHK----GKQIKTKHFDEYDYIIGMDES NINNL---	102
Schiz pombe	67	GIHTKHL----ARKLSTSDFKNFDYIFAMDSS NLRNI---	99

Figure 3-5 is continued on page 191.

Figure 3-5, continued from page 190.

Ent his 1	95	--LDRCP EQYKQKIFKMVDFCTTIKTTE --VP DPYYGGEK	130
Ent his 2	95	--LDRCP EQYKQKIFKMVDFCTTIKTTE --VP DPYYGGEK	130
Ent dispar	95	--LDRCP EQYKQKIFKMVDFCTTIKTTE --VP DPYYGGEI	130
Ent invad	95	--MYNCP KEYQDKIHKMVEYCKKSKDSE --VP DPYYSGLK	130
Gia lam P15	96	--MREVGKRD QSKVRMFRDYDPEGK -GD--VP DPYYCGR -	130
Giard intes	96	--I KEIGKRDRSKIR MFRDYDPEGK-GD--VP DPYYCGR -	130
Dictyo disc	119	LKHSTTKDNHIATIKRL GEYH TH-KKIN--VE DPYYGDMS	155
Klu lactis	100	--KRIQ PKEPKAQLSLFGKWNTDNKFQD -IVD DPYYGGNE	136
Sacch cerev	103	--KKIQ PEGSKAKVCLFGDWNTNDGTVQTIIE DP WYGDIQ	140
Schiz pombe	100	--NRVK PQGSRAKVMLFG EYASPGVSKI--V DPYYGGSD	135
Ent his 1	131	GFHRVIDILEDACENLI IKLEEGKLIN	157
Ent his 2	131	GFHRVIDILEDACENLI IKLEEGKLIN	157
Ent dispar	131	GFHRVIDILEDACENLI IKL QEGKLTN	157
Ent invad	131	GFQKVIDMLEDGCKNLLDQ LENAGDE	156
Gia lam P15	131	-FELVYS MVERTCRSILDRIKSGKAL	154
Giard intes	131	-FELVYS IVERTCRSILDRIKNGKAL	154
Dictyo disc	156	NFNICFN HVHDCLVNFLKEI ESLN	179
Klu lactis	137	GFEYNFKQVWYFSEQFLELES	157
Sacch cerev	141	DFEYNFKQ ITYFSKQFLK KEL	161
Schiz pombe	136	FGDCYIQLVDF SONFLKSIA	156

Figure 3-6. Protein sequence alignments comparing the *Entamoeba histolytica* LMW-PTPs with selected LMW-PTPs from multicellular animals.

Identical amino acids are highlighted in yellow, and conservative substitutions are highlighted in light blue. The conserved active site of each LMW-PTP is highlighted in pink, and the conserved DPYY loop is highlighted in red.

The amino acid sequence of the *E. histolytica* LMW-PTP protein XP_656359 was aligned with other related proteins using BLAST [311] and LALIGN [312]. Homo isof b corresponds to IF2 and Homo isof c corresponds to IF1 [203]. Table 3-4 lists species abbreviations, LMW-PTP accession numbers, and percent identity and similarity to the *E. histolytica* LMW-PTP protein XP_656359 for each multicellular animal LMW-PTP sequence. A phylogenetic tree comparing all selected LMW-PTPs is shown in Figure 3-7.

Ent his 1	1	-----M KL LFV CLGNICRS PAAEAVM KKVIQ NHHL	30
Ent his 2	1	-----M KL LFV CLGNICRS PAAEAVM KKVIQ NHHL	30
Homo isof b	1	-----MAEQATKSV LFVCLGNICRS PIAEAVFR KLVT DQNI	36
Homo isof c	1	-----MAEQATKSV LFVCLGNICRS PIAEAVFR KLVT DQNI	36
Mus musc 1	1	-----MAEVGSKSV LFVCLGNICRS PIAEAVFR KLVT DEKV	36
Mus musc 2	1	-----MAEVGSKSV LFVCLGNICRS PIAEAVFR KLVT DEKV	36
Bos taurus	1	-----MAEQVTKSV LFVCLGNICRS PIAEAVFR KLVT DQNI	36
Xenop laev	1	-----MAQQNGKSV LFVCLGNICRS PIAEAVF QKLVT DAGI	37
Gal gallus	1	-----MAAGEVKSV LFVCLGNICRS PIAEAVFR KLVT DEKV	36
Dros melan	1	-----MVRK VLMICLGNICRS PIAEVVM VDT LEKANV	32
Danio rerio	1	-----MAASSGKSV LFVCLGNICRS PIAEAVFR KMAT DSGV	36
Ent his 1	31	TEKYICDSAGTCSYHEGQQADSRMRKVGKSRGYQVDSISR PVVSS	75
Ent his 2	31	TEKYICDSAGTCSYHEGQQADSRMRKVGKSRGYQVDSISR PVVSS	75
Homo isof b	37	SENWVIDSGAVSD WNVGRSPDPRAV SCLRNHGIHTAHKARQITKE	81
Homo isof c	37	SENWRVDSAATSGYEIGNPPDYRGQSCMKRHGIPMSHVARQITKE	81
Mus musc 1	37	SDNWAIDS SAVSDWNVGRPPDPRAV SCLRNHGIHTAHKARQITKE	81
Mus musc 2	37	SDNWRIDSAATSTYEVGNPPDYRGQNCMRKHGIHMQH IARQITKE	81
Bos taurus	37	SDNWRIDSAATSTYELGNPPDCRGQACMRKHGIPMSHVARQV TKE	81
Xenop laev	38	SKEWSIDSAATSD WNVGSSPDSRAL KCLKSHSIETSHRAQQITRD	82
Gal gallus	37	ENKWRIDSAATSTYEIGNPPDYRGQTCMKKHGITMNH IARQVTKD	81
Dros melan	33	KDVEV- DSAAIGGWHVGN RADPRAISTLQKHGLKCTHIVRQIRKQ	76
Danio rerio	37	VDKWVIDSGATSD WNTGSTP DARGLACL RKHGIETD HRARQVTKD	81
Ent his 1	76	DFKNFDYIFAMDNDNYELLDRCP--E-QYK--QKIFKMVDFCTT	115
Ent his 2	76	DFKNFDYIFAMDNDNYELLDRCP--E-QYK--QKIFKMVDFCTT	115
Homo isof b	82	DFATFDYILCMDESNLRD-LNRKS--N-QVKTCKAKIELLGSYDP	122
Homo isof c	82	DFATFDYILCMDESNLRD-LNRKS--N-QVKTCKAKIELLGSYDP	122
Mus musc 1	82	DFATFDYILCMDESNLRD-LNRKS--N-QVKNCKAKIELLGSYDP	122
Mus musc 2	82	DFATFDYILCMDESNLRD-LNRKS--N-QVKNCKAKIELLGSYDP	122
Bos taurus	82	DFVTFDYILCMDESNLRD-LNRKS--N-QVKNCRAKIELLGSYDP	122
Xenop laev	83	DFLSYDYILCMDESNLQDLKRRGS--QVQNC--KAKIELLGSYDP	123
Gal gallus	82	DFQTFDYILCMDESNLRD-LKRKS--N-QVKDCKAKIELLGAYDP	122
Dros melan	77	DFSEFDYIFGMDEDNMSELRR LAP --K-GSK--AELLM L GDFGLE	116
Danio rerio	82	DFMSFDYILCMDESNLRDLNKKASSVE-NSK--AKI-ELLGSYDP	122
Ent his 1	116	IKTTEVPDPYY--GGEKGFHRVIDILEDACENLI IKLEEG KLIN	157
Ent his 2	116	IKTTEVPDPYY--GGEKGFHRVIDILEDACENLI IKLEEG KLIN	157
Homo isof b	123	QKQLI I EDPYY--GNDSDFETVYQQCVRC CRAF LEKAH	158
Homo isof c	123	QKQLI I EDPYY--GNDSDFETVYQQCVRC CRAF LEKAH	158
Mus musc 1	123	QKQLI I EDPYY--GNDSDFEVVYQQCLRC CRAF LEKTY	158
Mus musc 2	123	QKQLI I EDPYY--GNDSDFEVVYQQCLRC CRAF LEKTY	158
Bos taurus	123	QKQLI I EDPYY--GNDADFETVYQQCVRC CRAF LEKVR	158
Xenop laev	124	QKQLI I QDPYY--GRDEFETVYQQCIRC CSF LEKSS	159
Gal gallus	123	QKQLI I EDPYY--GNEKDFETVYE Q CVRC CRAF LEKPH	158
Dros melan	117	KKNRI I EDPYYERGAEGFETAYQQCVVACAAFMKERLQK	155
Danio rerio	123	EKQLI I QDPYY--GSDKDFETVYE Q CARC CRAF LEQHS	158

Figure 3-7. A phylogenetic tree for the comparison between the *Entamoeba histolytica* LMW-PTP protein XP_656359 and all selected LMW-PTP proteins from different species (multicellular eukaryotes, unicellular eukaryotes, and bacteria).

Species abbreviations are listed in Tables 3-1 through 3-4, and LMW-PTP protein alignments are shown in Figures 3-3 through 3-6. The protein sequences for the selected LMW-PTPs were input into www.phylogeny.fr [314] or its mirror site [315] and the phylogenetic tree was created with “one-click” phylogenetic analysis using MUSCLE for protein sequence alignment, Gblocks to eliminate divergent regions and poorly aligned positions PhyML for maximum likelihood phylogeny determination, and TreeDyn for tree rendering [314, 375, 376, 377, 378, 379]. The branch length is proportional to the number of substitutions per site.

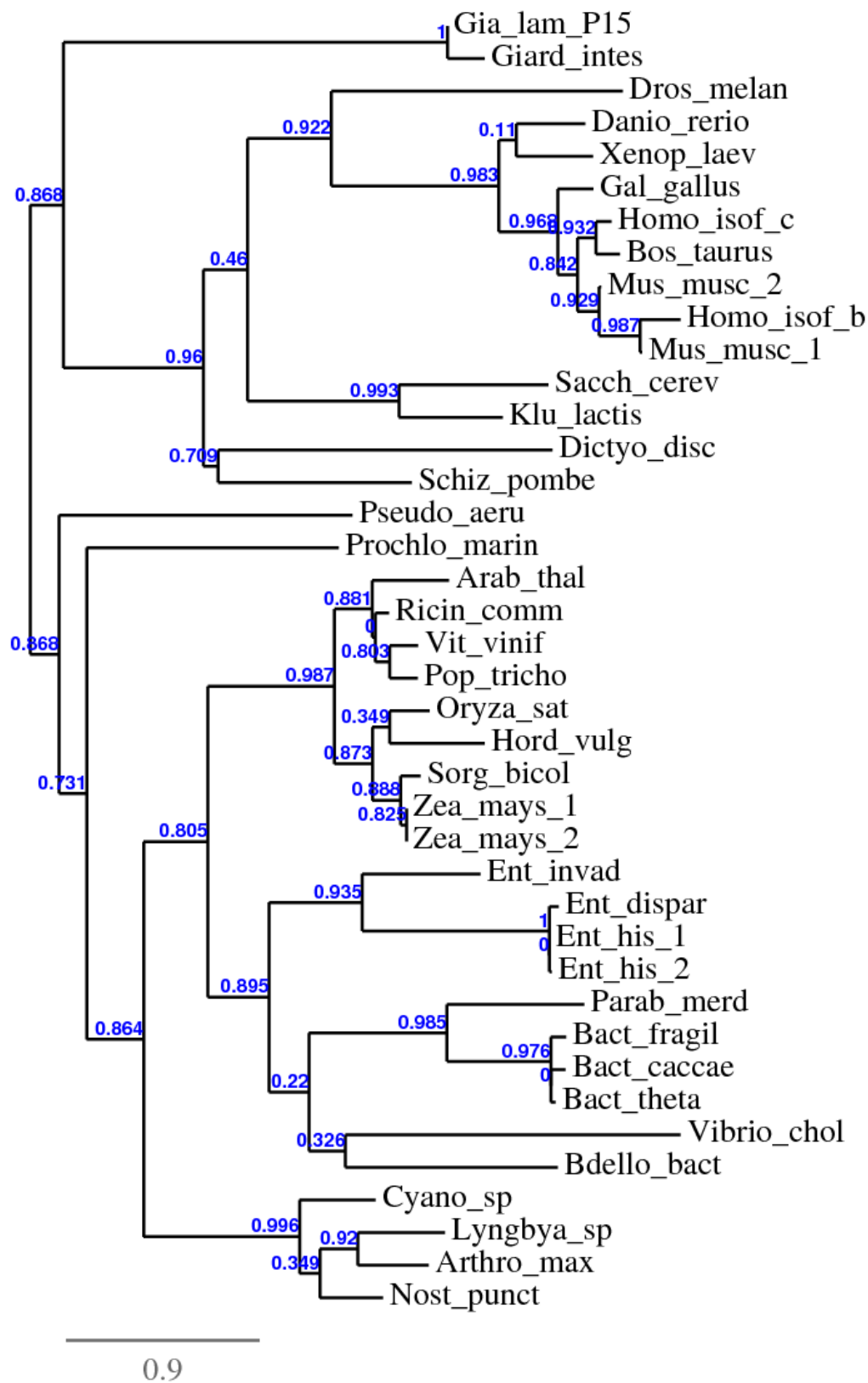


Figure 3-8. Structure of the *Entamoeba histolytica* LMW-PTP apoenzyme with no bound ligand. The crystallized enzymes were subjected to X-ray diffraction and the structure solved for each given resolution, and were deposited in the RSCB Protein Data Bank (PDB) [342]. The structure of the LMW-PTP apoenzyme (resolution: 2.2 Å) shows a disordered active site, presumably due to the absence of bound substrate; as a result the structure appears to be discontinuous since the P-loop (active site) and P-loop interacting parts of the molecule (located in the lower right of the figure) are not visible in this structure. Its PDB identifier [342] is 3ILY.

The crystal structure images are courtesy of Thomas E. Edwards (Emerald BioStructures).

Crystallization, X-ray diffraction, and structure determination were performed at the SSGCID consortium.

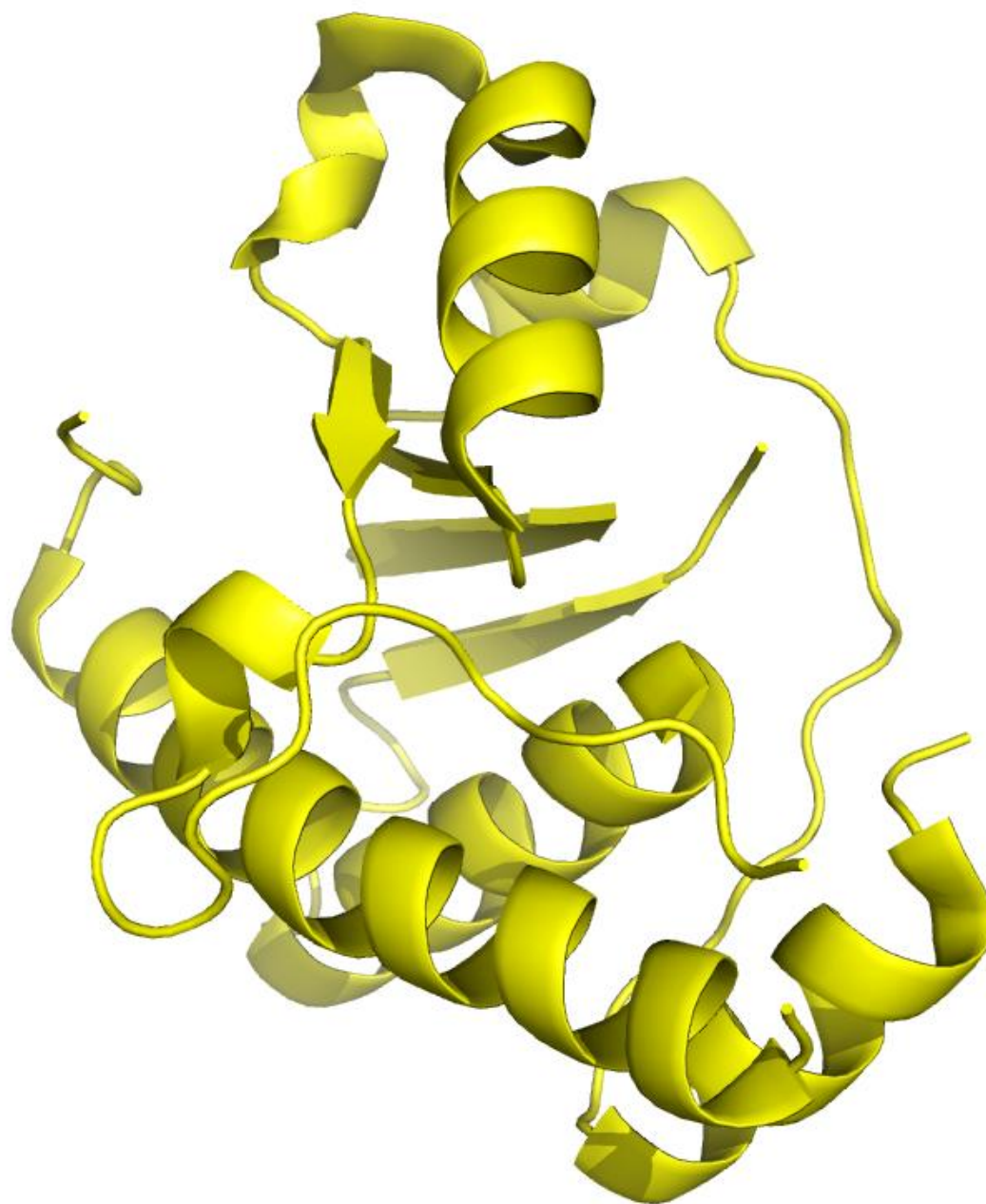


Figure 3-9. Structure of the *Entamoeba histolytica* LMW-PTP with the substrate mimic

HEPES bound. The crystallized enzyme was subjected to X-ray diffraction, the structures solved, and deposited in the RSCB Protein Data Bank (RSCB PDB) [342]. Two structures, of 2.2 Å and 1.95 Å resolution, were obtained with HEPES (a non-hydrolyzable phosphotyrosine substrate mimic) bound tightly in the active site. In this structure, the P-loop (active site) residues are well-ordered, as are the residues that interact with the residues of the P-loop. The PDB identifiers [342] for these are 3IDO and 3JS5 respectively. This structure is also shown in space-filling model form in Figure 3-15.

The crystal structure images are courtesy of Thomas E. Edwards (Emerald BioStructures).

Crystallization, X-ray diffraction, and structure determination were performed at the SSGCID consortium.

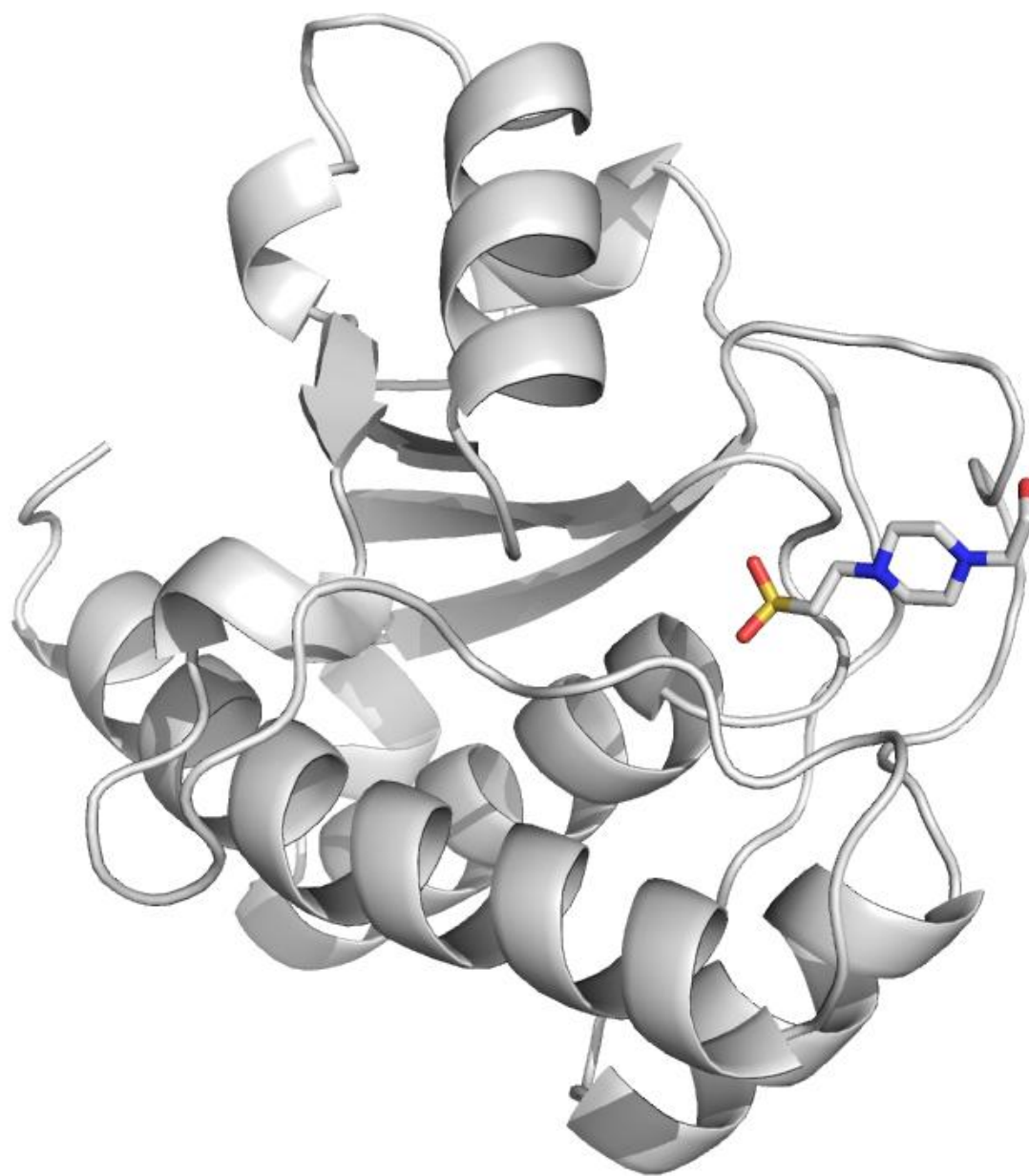


Figure 3-10. Structures of the *Entamoeba histolytica* LMW-PTP with sulfate product mimic bound. The crystallized enzyme was subjected to X-ray diffraction and the structure was solved and deposited in the RSCB Protein Data Bank (RSCB PDB) [342]. A structure of resolution 1.8Å was obtained with sulfate as a product mimic bound in the active site. This structure has a well-ordered P-loop, but additional residues on the surface of the active site are disordered (the discontinuity of the structure is apparent in the lower right corner of the image). The PDB identifier [342] for this molecule is 3JVI.

The crystal structure images are courtesy of Thomas E. Edwards (Emerald BioStructures). Crystallization, X-ray diffraction, and structure determination were performed at the SSGCID consortium.

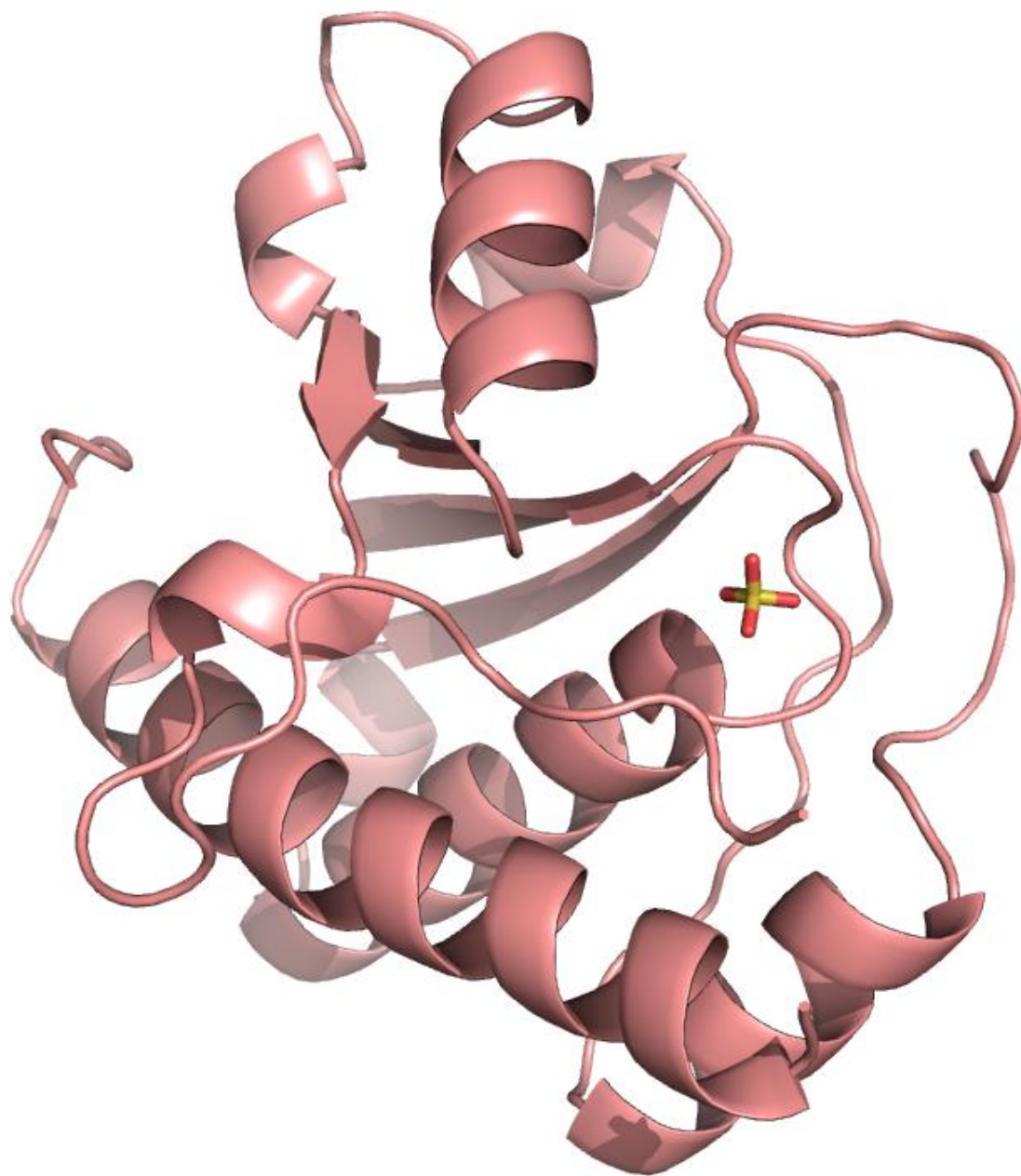


Figure 3-11. Triple overlay of the three structures of the *Entamoeba histolytica* LMW-PTP.

Triple overlay of the three structures shown in Figures 3-8, 3-9, and 3-10: the apoenzyme structure with no bound ligand (yellow), the enzyme structure with HEPES bound as a substrate mimic (gray), and the enzyme with the product mimic sulfate bound (red). To obtain the structure of the enzyme, the crystallized protein was subjected to X-ray diffraction. The apo structure (yellow) has a disordered active site, presumably due to the lack of bound substrate. Note how the structure of the apoenzyme appears to be discontinuous in sections, since the molecule would have more than one structural conformation in those sections. The discontinuity is especially apparent in the lower right of the overlaid images, the location of the active site and the DPYY loop. In the structure with bound HEPES mimicking a substrate molecule (gray), the active site residues are well-ordered. In the structure (red) with sulfate bound in the active site, mimicking a product, the P-loop (active site) of the phosphatase is well ordered, but additional residues on the surface of the active site are disordered.

All crystal structure images are courtesy of Thomas E. Edwards (Emerald BioStructures).

Crystallization, X-ray diffraction, and structure determination were all performed at the SSGCID consortium.

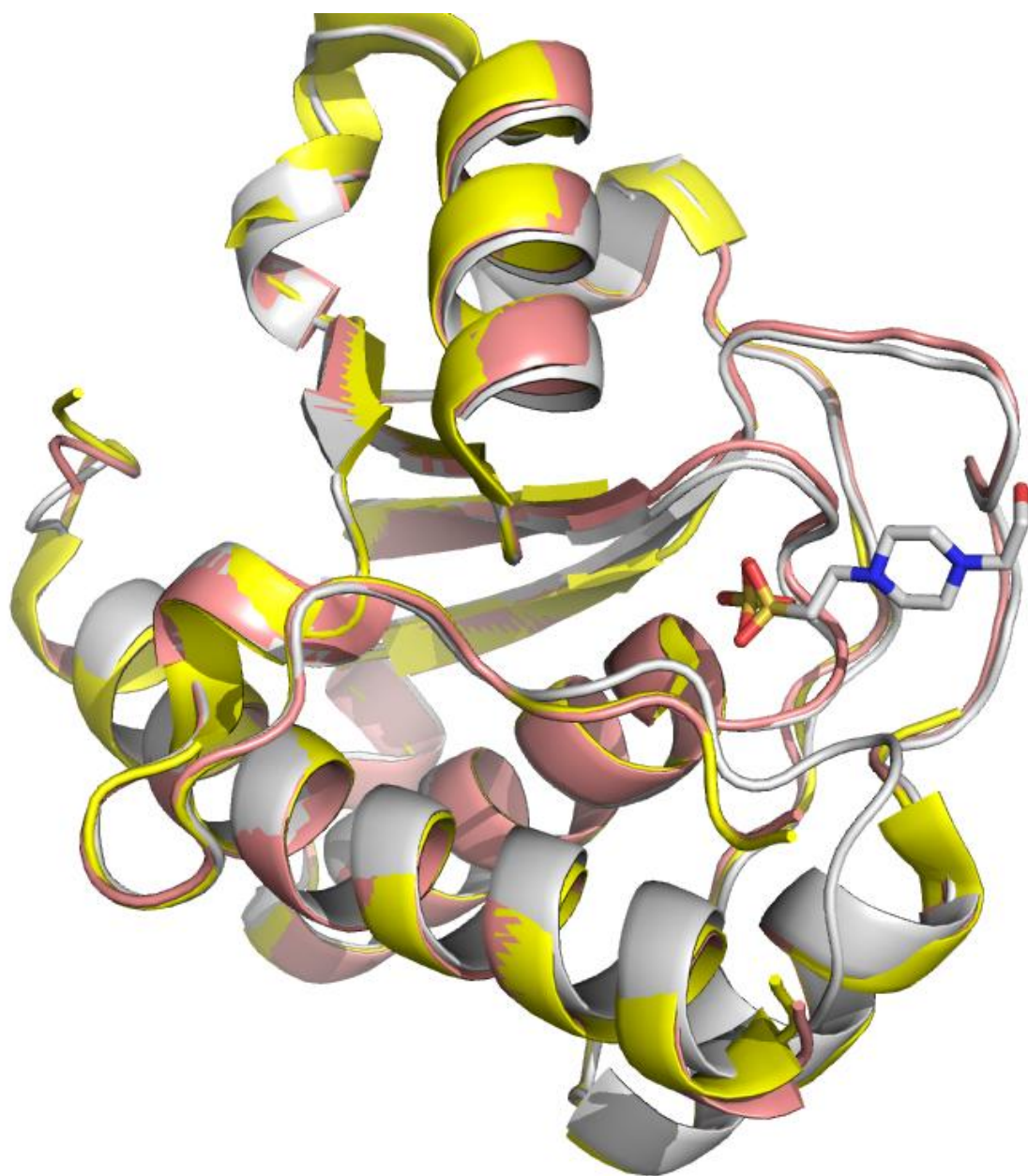


Figure 3-12. Crystal shapes of the crystallized LMW-PTP proteins. (A). The non-ligand-bound (apo) structure of the LMW-PTP had the space group $P4_3$, which means it is “primitive”: the crystals contain only one repeating unit [343]. The crystal shape of the crystallized protein was that of a tetragonal pyramid [344, 345, 346, 347]. Figure adapted from <http://www.metafysica.nl/turing/vierzijdigepiramide.gif> [380]. (B). The LMW-PTP with substrate mimic bound had the space group $P2_12_12_1$, containing only one repeating unit, and forming an orthorhombic sphenoidal crystal [345, 346, 347, 348]. Figure adapted from Salisbury and Ford (1922) [381]. (C). The crystal structure of the enzyme with the product mimic sulfate bound had the space group $P4_32_12$, which corresponds to a “primitive” tetragonal trapezohedral crystal [346, 347, 349]. Figure adapted from http://www.metafysica.nl/turing/tetragonal_trapezohedron.jpg [382].

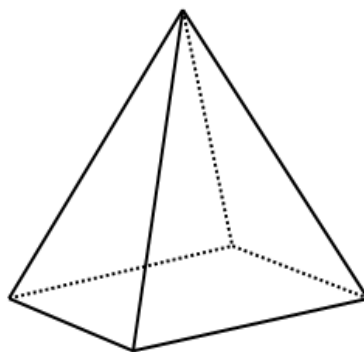
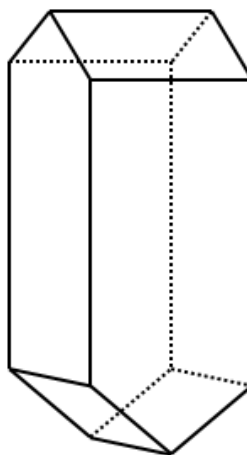
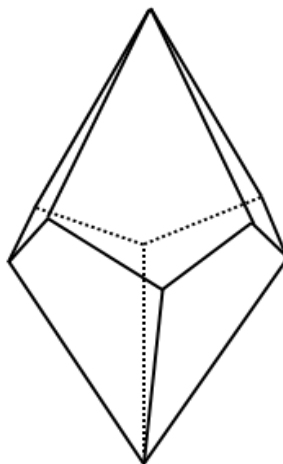
A**B****C**

Figure 3-13. Active site residues for the *E. histolytica* LMW-PTP with the substrate mimic HEPES bound. The residues that form the active site are shown in green. For R groups of the residues, thiol (sulfhydryl) groups (-SH) are shown in yellow, amino (-NH₂) groups are shown in blue, and hydroxyl groups (-OH) in red. Residues are numbered according to which residue they are in the *E. histolytica* LMW-PTP protein sequence. Residue abbreviations: Arg, arginine; Asn, asparagine; Asp, aspartic acid; Cys, cysteine; Gly, glycine; His, histidine; Ile, isoleucine; Leu, leucine; Tyr, tyrosine. Key residues in the active site include the catalytic residue Cys7; Cys12 can form a disulfide bond with Cys7 to prevent irreversible oxidation of Cys7 in the presence of reactive oxygen species, and Arg13 stabilizes the transition state via hydrogen bonding. The surface loop contains the conserved Asp123, which functions as a general acid/base in catalysis, as well as Tyr124 and Tyr125 residues, which have been shown to be phosphorylated by Src family kinases in mammalian LMW-PTPs. The His45 residue appears to function in substrate specificity recognition. The space-filling model is shown in Figure 3-14. Figure 3-2 shows the catalytic mechanism of the enzyme.

The crystal structure images are courtesy of Thomas E. Edwards (Emerald BioStructures).

Crystallization, X-ray diffraction, and structure determination were performed at the SSGCID consortium.

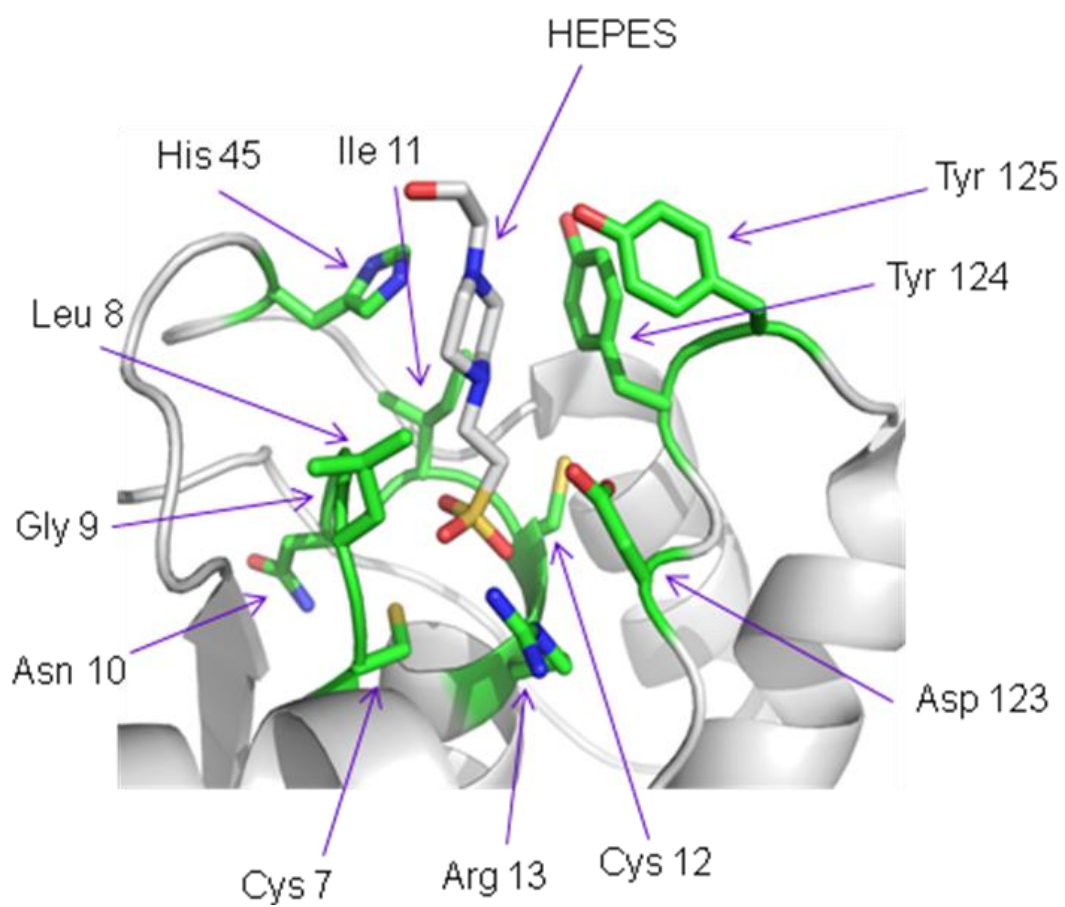


Figure 3-14. Active site residues for the *E. histolytica* LMW-PTP with the substrate mimic HEPES bound, using a space-filling model. This is the same structure that is shown in Figure 3-13 in ribbon form.

The crystal structure images are courtesy of Thomas E. Edwards (Emerald BioStructures).

Crystallization, X-ray diffraction, and structure determination were performed at SSGCID.

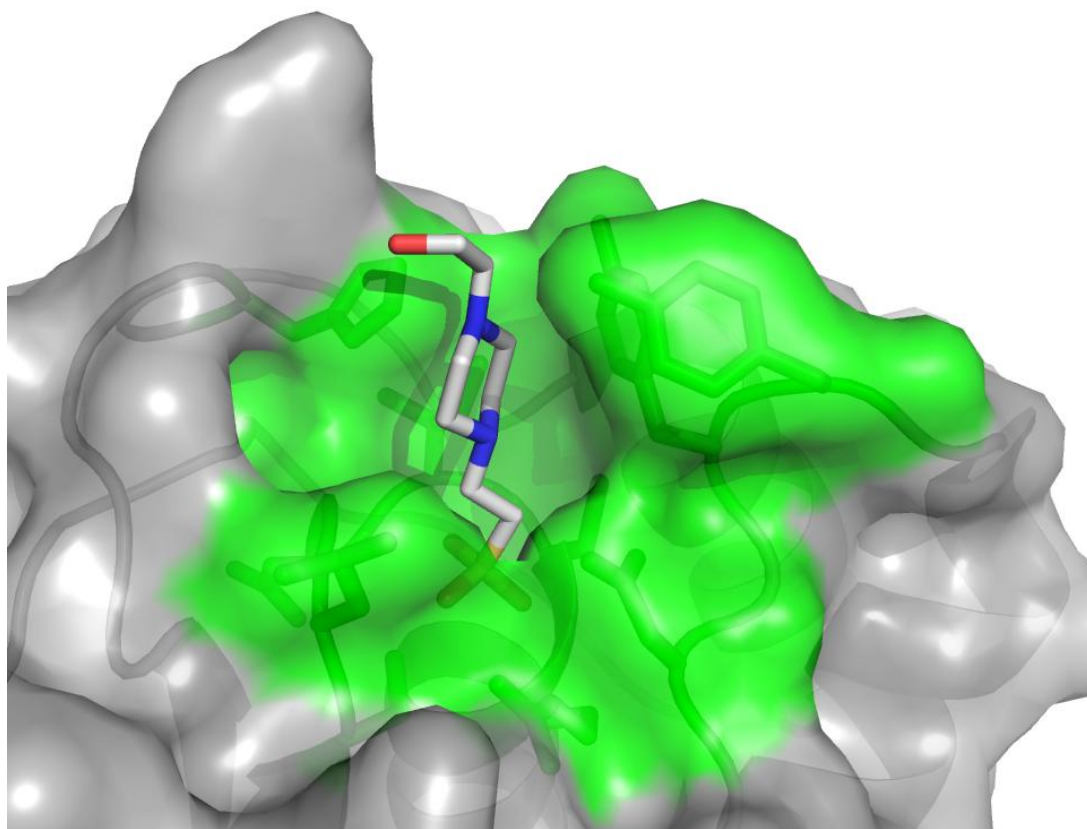


Figure 3-15. Space-filling model of top and side views of the *E. histolytica* LMW-PTP structure with HEPES bound in the active site. This is the same structure that is shown in Figure 3-9 in ribbon form. Active site residues are shown in green, and the rest of the residues in gray, with a HEPES molecule (a non-hydrolyzable phosphotyrosine mimic) bound in the active site, shown in sticks representation. Note the feature of the pocket formed by the active site, DPYY loop, and other surrounding residues, where the phosphotyrosine substrate binds.

The crystal structure images are courtesy of Thomas E. Edwards (Emerald BioStructures).

Crystallization, X-ray diffraction, and structure determination were all performed at the SSGCID consortium.

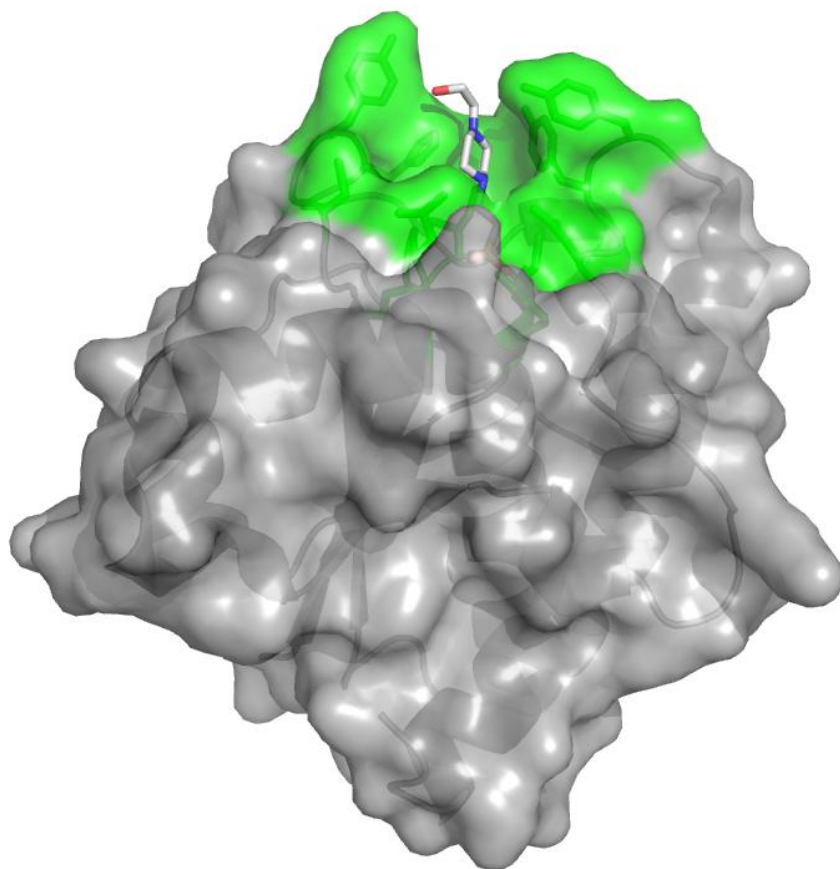
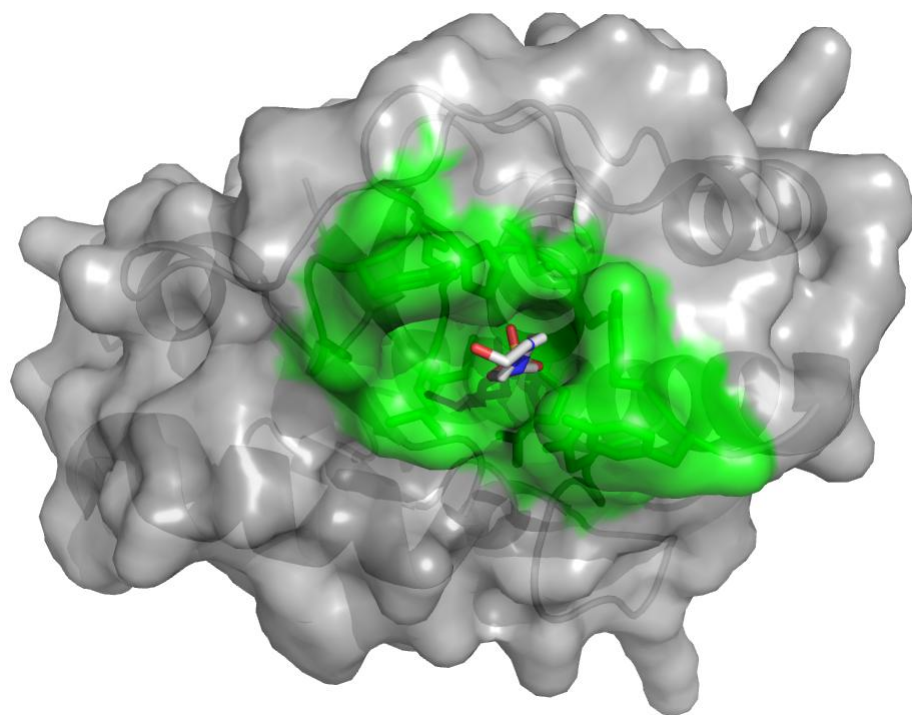


Figure 3-16. Assay for phosphatase activity with purified recombinant LMW-PTP, generic phosphatase substrate, and phosphatase inhibitor. Wild-type recombinant *E. histolytica* LMW-PTP was tested for phosphatase activity using the Sensolyte *p*NPP Colorimetric Protein Phosphatase Assay Kit (AnaSpec, Fremont, CA, USA). *p*NPP, *p*-nitrophenyl phosphate, is a colorimetric generic phosphatase substrate that turns yellow when dephosphorylated. Recombinant wild-type LMW-PTP, in amounts from 10 ng to 200 ng, were diluted in supplied buffer with 5 mM *p*NPP. The reaction was allowed to proceed, and absorbance was measured at 405 nm at the 60 minute time point. To test inhibition, HALT™ Phosphatase Inhibitor (Thermo Fisher Scientific, Waltham, MA, USA) was added to the reaction at a concentration of 1X. Values were plotted as nanograms of recombinant wild-type LMW-PTP vs absorbance at 405 nm using Excel. The slope of the trendline and the R² value was also calculated using Excel and these values are displayed on the graph. The wild-type recombinant LMW-PTP definitely had phosphatase activity, and was inhibited by the addition of the HALT™ Phosphatase Inhibitor Cocktail by an average of 92.5 ± 1.0 %.

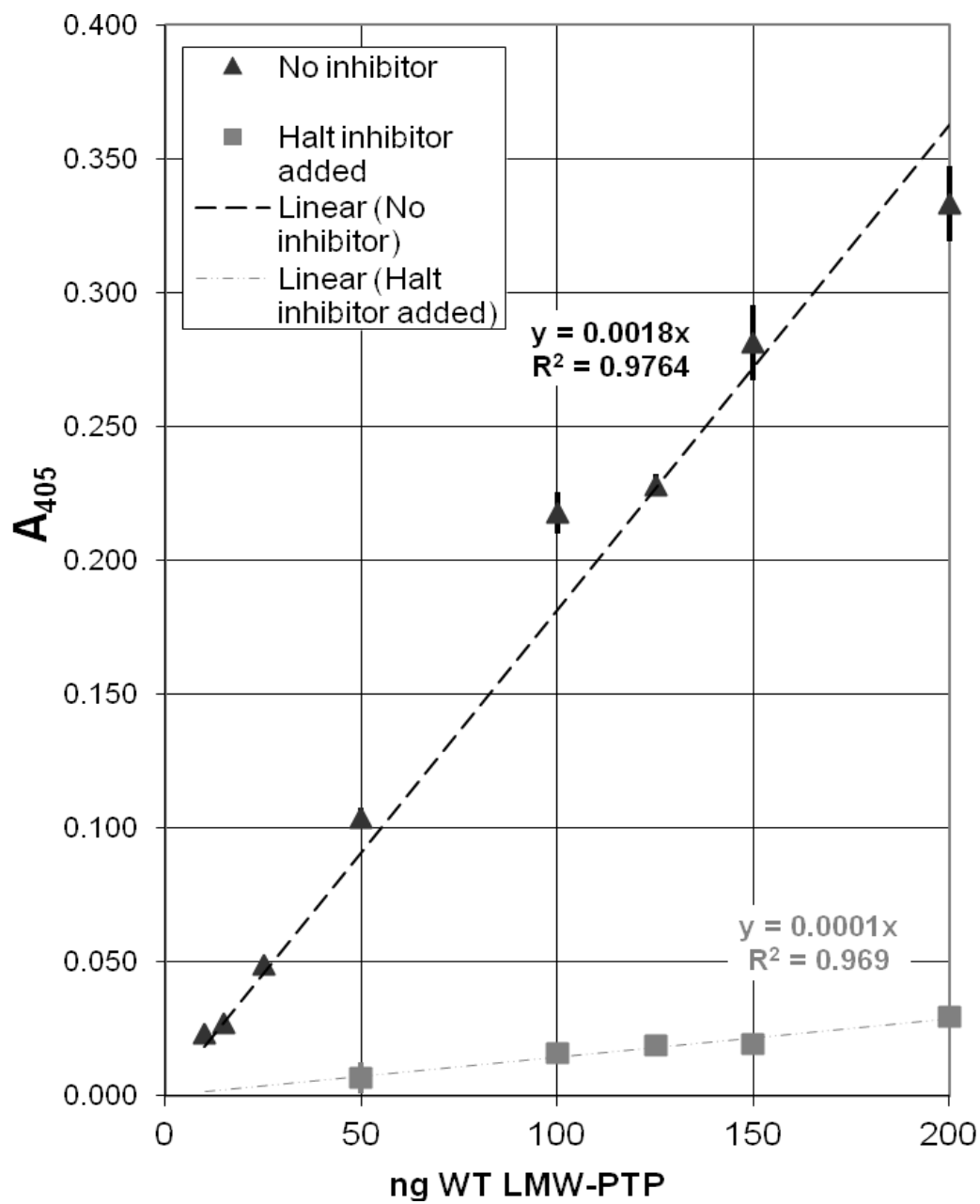


Figure 3-17. Test of anti-LMW-PTP sera from inoculated rabbits. We assayed if the native LMW-PTP is recognized by immune sera from rabbits inoculated with wild type recombinant LMW-PTP, and also if the sera can be used to immunoprecipitate the native LMW-PTP from amebic lysate. Two rabbits were immunized with wild-type recombinant LMW-PTP to generate antibodies that would recognize the native LMW-PTP. Lysate from 6.75×10^6 trophozoites was immunoprecipitated with either pre-immune or immune rabbit serum from each of two test bleeds, using Invitrogen Protein A Dynabeads (Invitrogen, Carlsbad, CA, USA). Antibody was not crosslinked to the beads. Immunoblots were probed with either pre-immune or immune serum from the rabbit that was not used for that IP. 100 ng of the purified recombinant wild-type LMW-PTP protein was also included as a positive control. Arrow: expected size of the LMW-PTP (17 kDa).

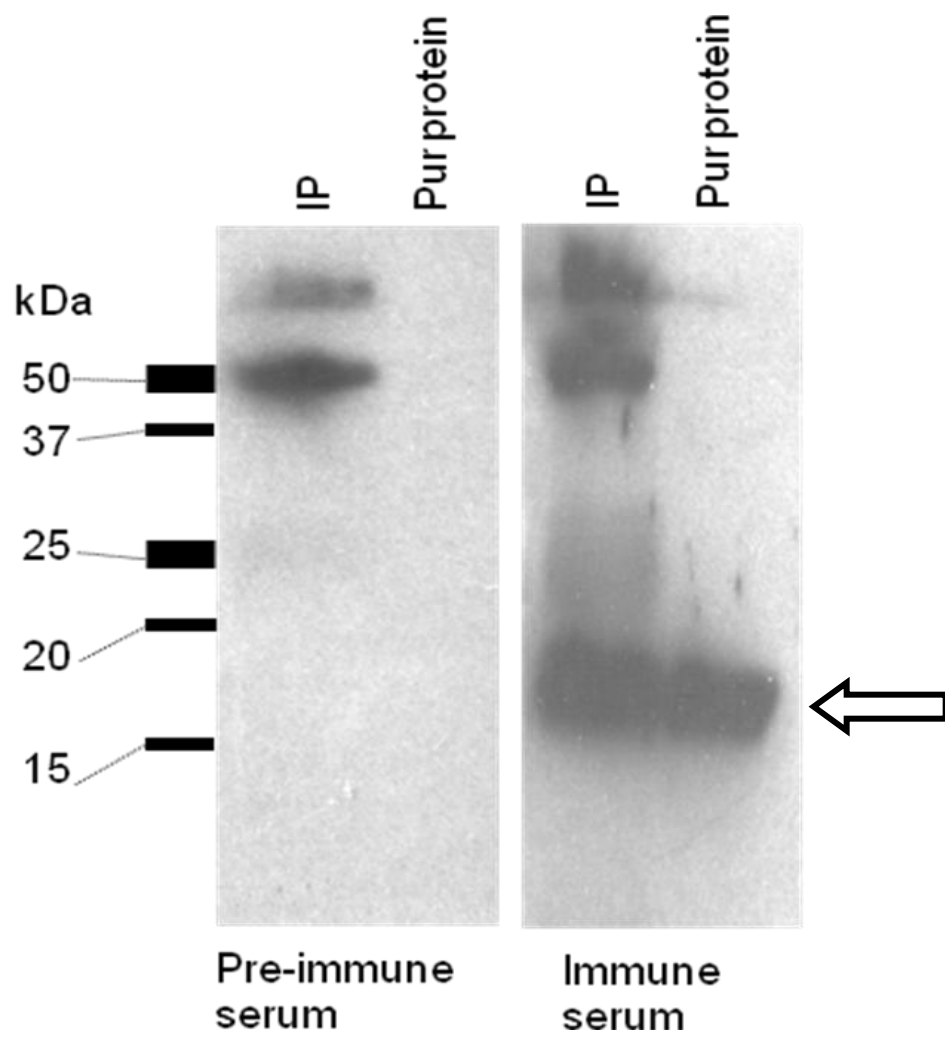


Figure 3-18. Diagram of the two-vector tetracycline-inducible system for expressing wild-type and mutant epitope-tagged LMW-PTP. 5' actin and 3' actin, 5' and 3' flanking regions of the *Entamoeba histolytica* actin gene; *Neo*, ORF (open reading frame) of the neomycin resistance gene; *luc*, ORF of the luciferase gene; *hyg*, ORF of the hygromycin resistance gene; 5' *hgl* and 3' *hgl*, 5' and 3' flanking regions of *Entamoeba histolytica hgl* gene; *tetR*, ORF of tetracycline repressor gene; tetO, tetracycline operator, sequence bound by the tetracycline repressor, with sequence TCCCTATCAGTGATAGAGA; 5' *fdx* and 3' *fdx*, 5' and 3' flanking regions of *Entamoeba histolytica* ferredoxin gene; N-tag, N-terminally added cMyc-FLAG-V5-6xHis epitope tag; C-tag, C-terminally added cMyc-FLAG-V5-6xHis epitope tag. The epitope-tagged wild-type and mutant LMW-PTPs were subcloned into pGIR209 [177, 179] and pGIR308 [177, 179] expresses the tetracycline repressor protein. The pGIR209 and pGIR308 diagrams are adapted from Ramakrishnan *et al* (1997) [177] and Katz *et al* (2003) [179].

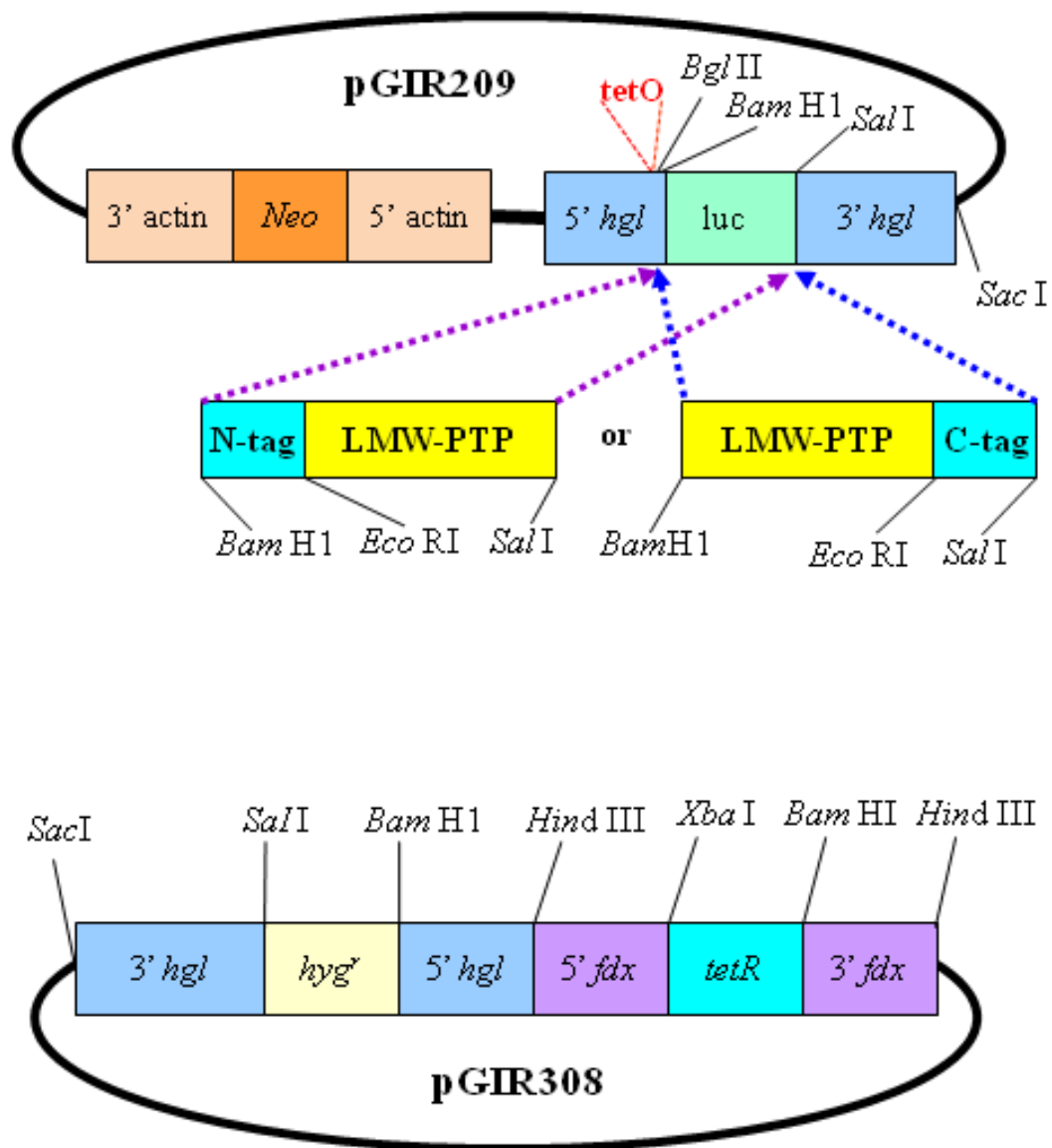


Figure 3-19. Assay for phosphatase activity with purified recombinant wild-type and substrate-trapping active-site mutant LMW-PTP proteins using the SensoLyte *p*NPP Colorimetric Protein Phosphatase Assay Kit (AnaSpec, Fremont, CA, USA). *p*NPP, *p*-nitrophenyl phosphate, is a colorimetric generic phosphatase substrate that turns yellow when dephosphorylated. 100 ng of either wild-type (WT) and mutant (Mut) recombinant purified protein was diluted in supplied buffer with 5 mM *p*NPP. The reaction was allowed to proceed for 120 minutes, with absorbance measured at 405 nm every five minutes, and plotted as A_{405} vs time (minutes) using Excel. The slope of the trendline and the R^2 value was also calculated using Excel, and these values are displayed on the graph.

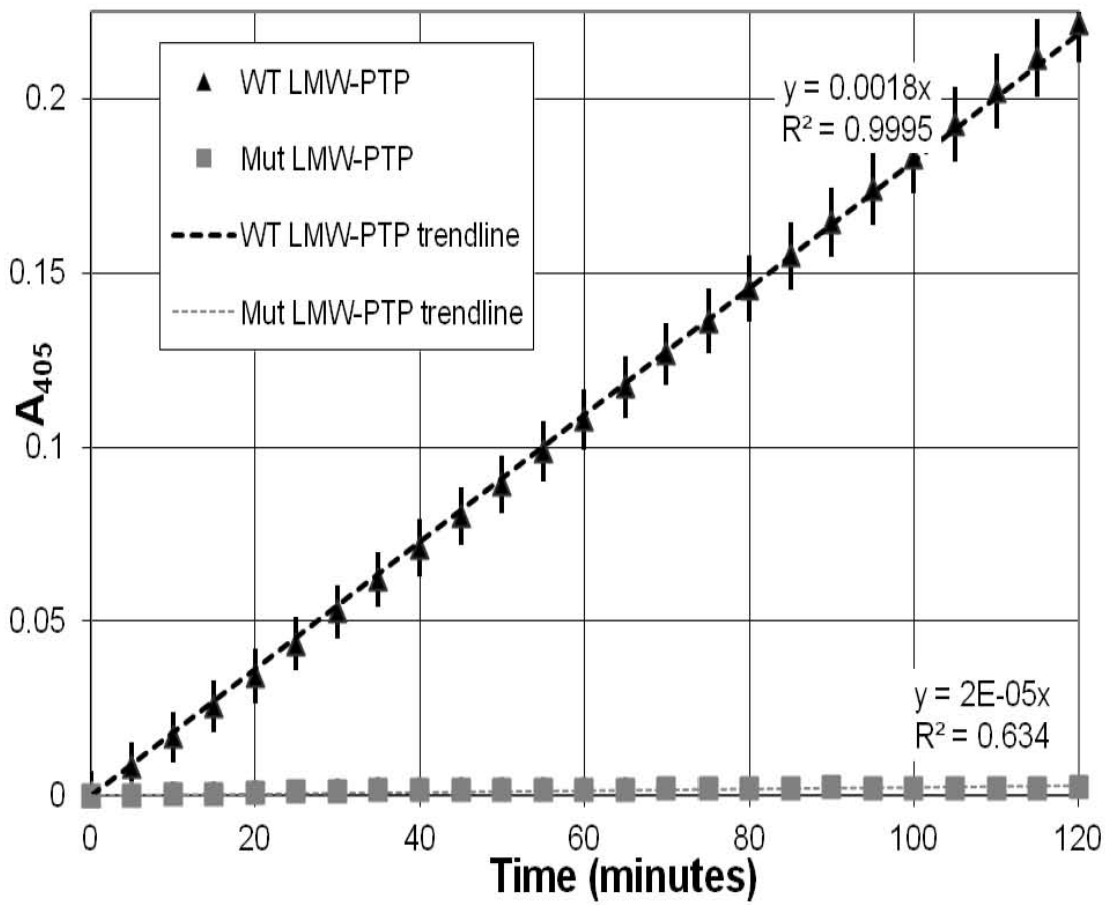


Figure 3-20. Assay for tetracycline induction of epitope-tagged wild-type LMW-PTP

overexpression. Amebae doubly transfected with both pGIR308 and either pGIR209, with the original (untagged) luciferase reporter gene present [177], or the wild-type N-terminally or C-terminally tagged wild-type LMW-PTP cloned into pGIR209, were selected with both hygromycin and G418 (neomycin). Tetracycline was added to the medium to induce gene expression at a concentration of either 1 $\mu\text{g/ml}$ or 5 $\mu\text{g/ml}$, and amebae were left to grow for 24 hours. Control samples did not have tetracycline added, in order to test gene induction. Non-transfected HM1:IMSS strain amebae were also included as a control to check that antibody directed against the epitope tags would not recognize amebic proteins. Amebae were then harvested, counted, and lysed. Lysates from 5×10^4 amebae were loaded per lane and subjected to SDS-PAGE. Western immunoblotting was performed with antibodies recognizing the V5, FLAG, and c-Myc epitope tags.

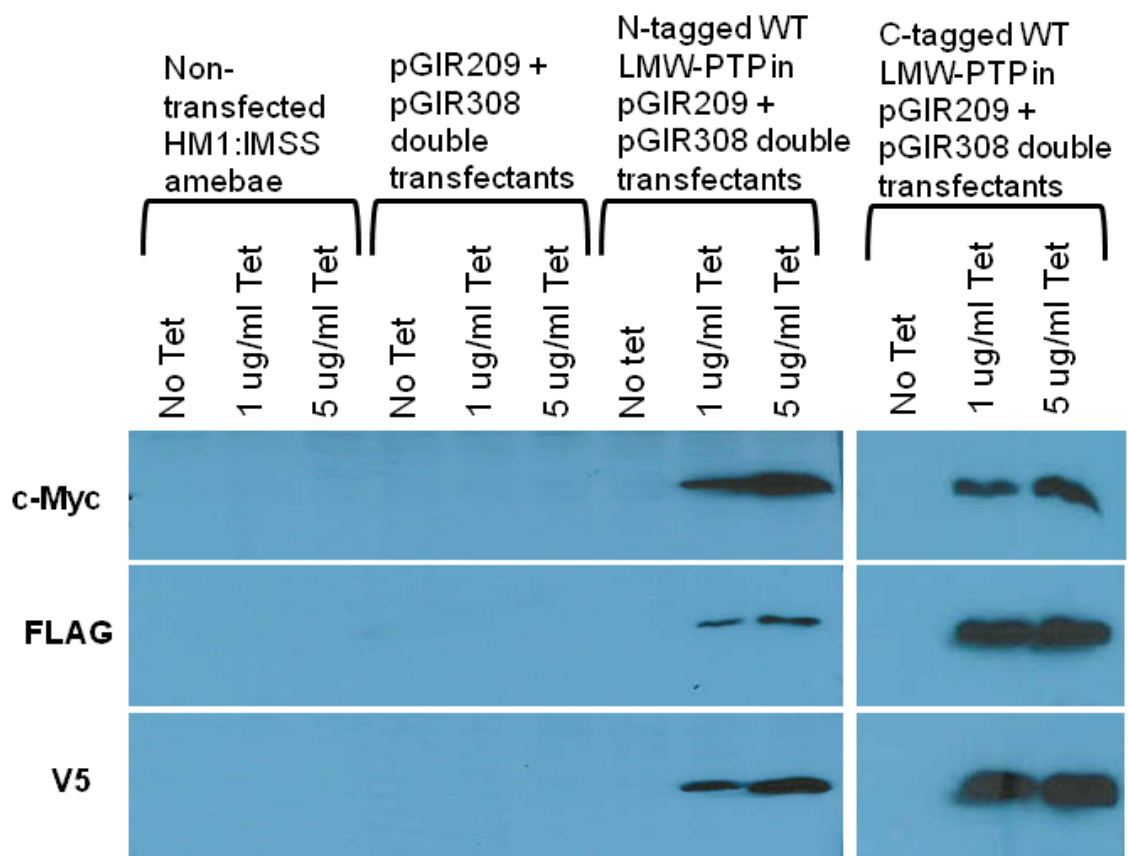


Figure 3-21. Test for inhibition of cellular protein tyrosine phosphatases by the addition of

sodium orthovanadate and iodoacetamide. Lysate from transfected amoebae constitutively expressing N-terminally epitope-tagged wild-type LMW-PTP from the expression vector pGIR209 was used to assay if general inhibition of phosphatase action would be possible even under the scenario when the wild-type LMW-PTP is overexpressed in transfected amoebae. With phosphatases inhibited, the action of tyrosine kinases should increase the level of tyrosine-phosphorylated proteins present in the amoebae. One set of transfectants was pre-treated for 30 minutes with 1 mM sodium orthovanadate before harvesting; the other set was not pre-treated. Transfectants were suspended in either:

- (1) regular lysis buffer without any phosphatase inhibitor added (contained 5 mM EDTA from the protease inhibitor)
 - (2) inhibitor lysis buffer with 1 mM sodium orthovanadate and 5 mM iodoacetamide (contained 5 mM EDTA from the protease inhibitor)
 - (3) inhibitor lysis buffer with 1 mM sodium orthovanadate, 5 mM iodoacetamide, but no EDTA.
- Halt Protease Inhibitor was also added to some of the lysed samples where indicated. Cells were lysed via sonication and 10 mM DTT was then added to the lysate. Lysate from 5×10^5 amoebae was loaded per lane. To check for enrichment of tyrosine-phosphorylated proteins, western blots were probed with anti-phosphotyrosine antibody.

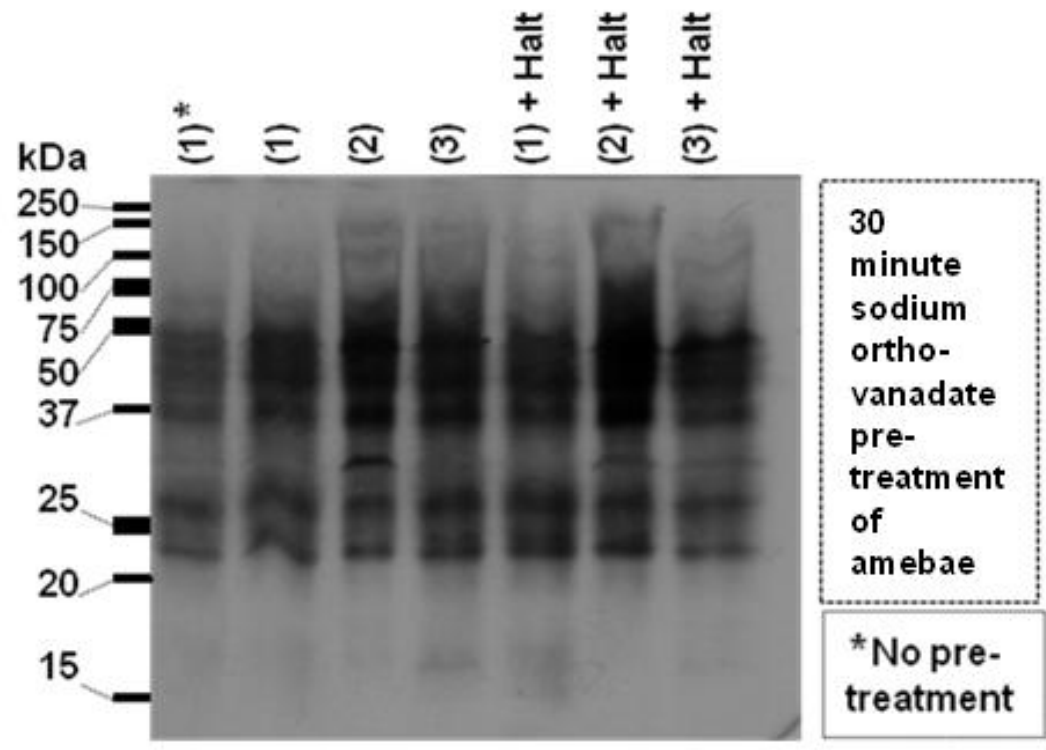
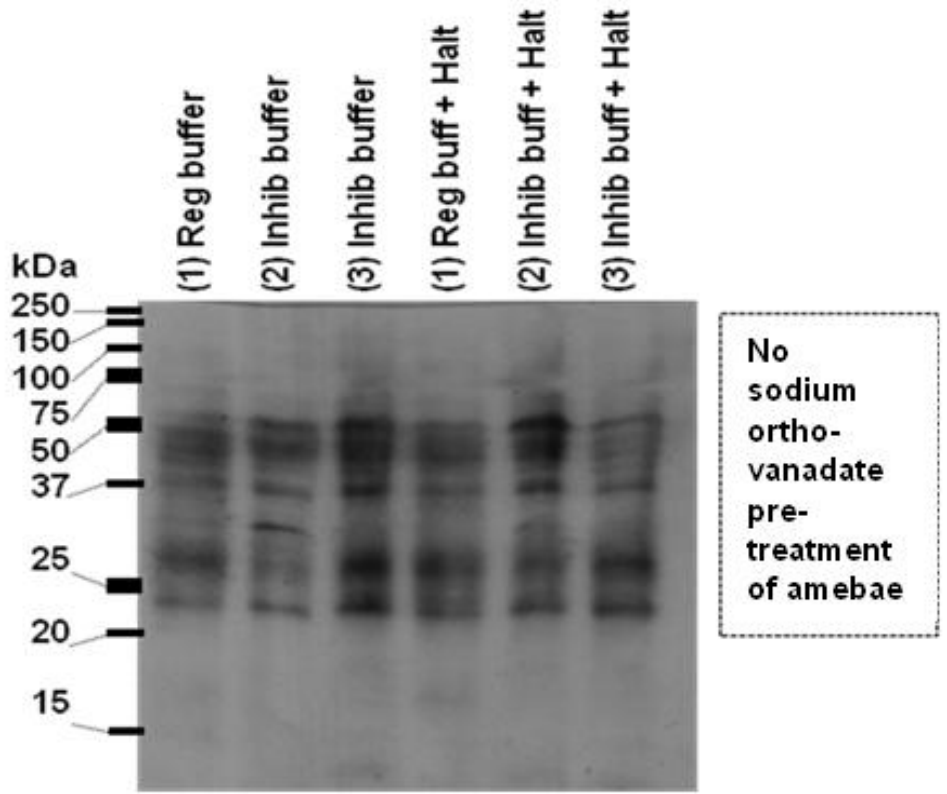
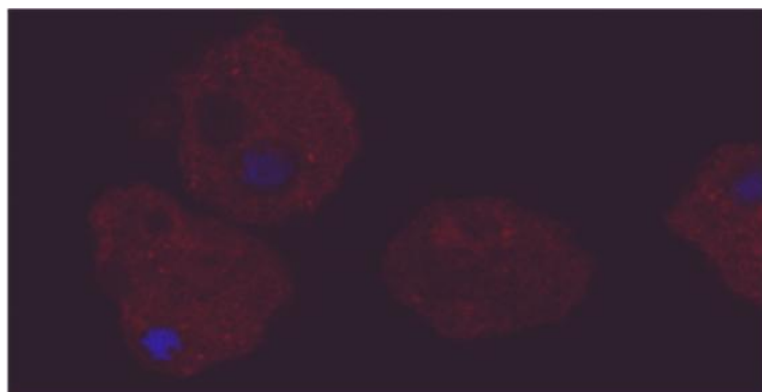
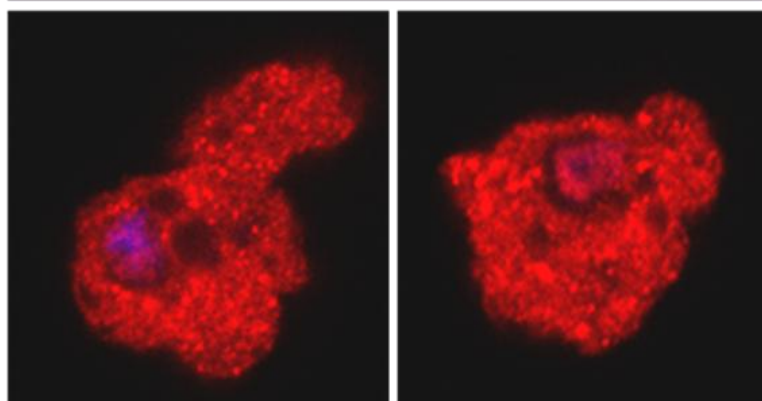


Figure 3-22. Confocal images of the cellular location of the C-terminal epitope-tagged wild-type and mutant LMW-PTP with induced or uninduced LMW-PTP expression. Since the immune anti-LMW-PTP rabbit serum recognizes amebic proteins other than only the LMW-PTP, we utilized the C-terminally epitope-tagged LMW-PTP wild-type and mutant overexpressers to visualize the protein cellular location via confocal microscopy. Confocal microscopy of these samples was performed by technician Carrie Cowardin. The C-terminally epitope-tagged wild-type and mutant tetracycline-inducible LMW-PTP transfectants were selected with 12 $\mu\text{g/ml}$ G418 and 30 $\mu\text{g/ml}$ hygromycin (double the usual baseline selection concentration) and were induced with 10 $\mu\text{g/ml}$ tetracycline for 24 hours before harvesting. V5 antibody was used at a dilution of 1:1000 and the secondary antibody was DyLight 649-conjugated anti-mouse IgG from goat. Amebae were also stained with DAPI at 1:5000 to visualize the nuclei (blue). The wild-type epitope-tagged LMW-PTP with tetracycline-induced expression showed increased expression of the protein as compared with no induction. Individual trophozoites appeared to be expressing the protein at different levels.

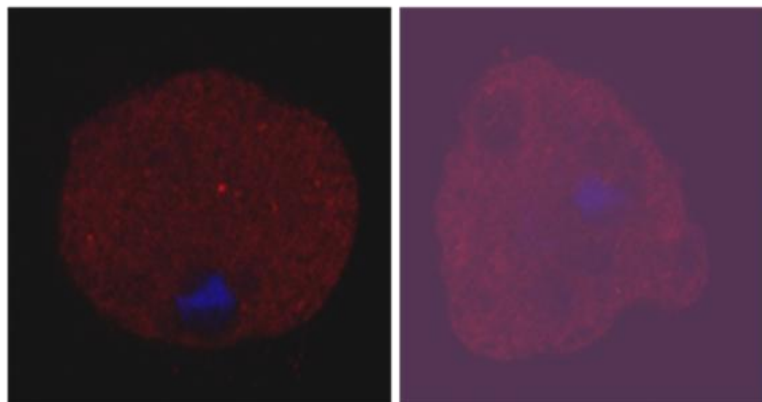
Uninduced
C-tagged
WT LMW-
PTP



Induced
C-tagged
WT LMW-
PTP



Uninduced
C-tagged
Mut LMW-
PTP



Induced
C-tagged
Mut LMW-
PTP

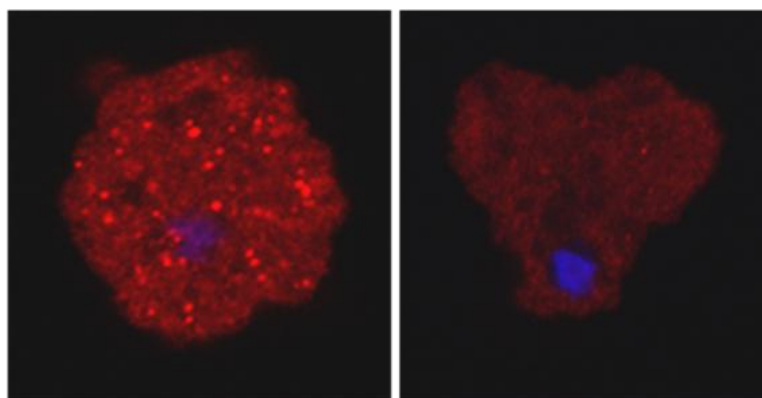


Figure 3-23. A representative immunoblot of pulldowns using substrate-trapping mutant LMW-PTP or BSA coupled to Affi-Gel 15 beads. Recombinant mutant LMW-PTP or BSA protein was attached to Affi-Gel 15 beads overnight at 4°C on a nutating mixer. Mock protein attachments were done with beads alone. HM1:IMSS amoebae were pre-treated with sodium orthovanadate, harvested, and lysed. Beads with LMW-PTP or BSA attached or mock attachment (beads alone) were incubated overnight at 4°C on a nutating mixer with amoebic lysate from non-transfected amoebae pre-treated with sodium orthovanadate, or with lysis buffer alone (the LMW-PTP and BSA attached to Affi-Gel 15 beads alone are not shown on this blot). Beads alone were incubated with lysate or with lysis buffer alone. Each sample was split into two sets. One set of samples was subjected to SDS-PAGE, blotted to membrane, and probed with anti-phosphotyrosine antibody to assay for phosphorylated proteins. There are three bands that appear to be phosphorylated: one at ~37 kDa, one about ~45 kDa, and one about ~55 kDa. Some of the attached mutant LMW-PTP seems to be detaching from the Affi-Gel 15 beads and the protein is running higher than expected, between 20 and 25 kDa rather than its usual 17 kDa. This is likely due to the large amount of the recombinant mutant LMW-PTP protein that was attached to the Affi-Gel 15 beads (500 µg recombinant protein per sample was used); the beads, after incubation with lysate, were boiled before being subjected to SDS-PAGE. The BSA is also running higher than its expected size of ~66 kDa [352], likely for the same reason. The larger apparent size of the mutant recombinant LMW-PTP for these substrate-trapping studies is also seen in the silver stained gels in Figure 3-24 and Figure 3-25.

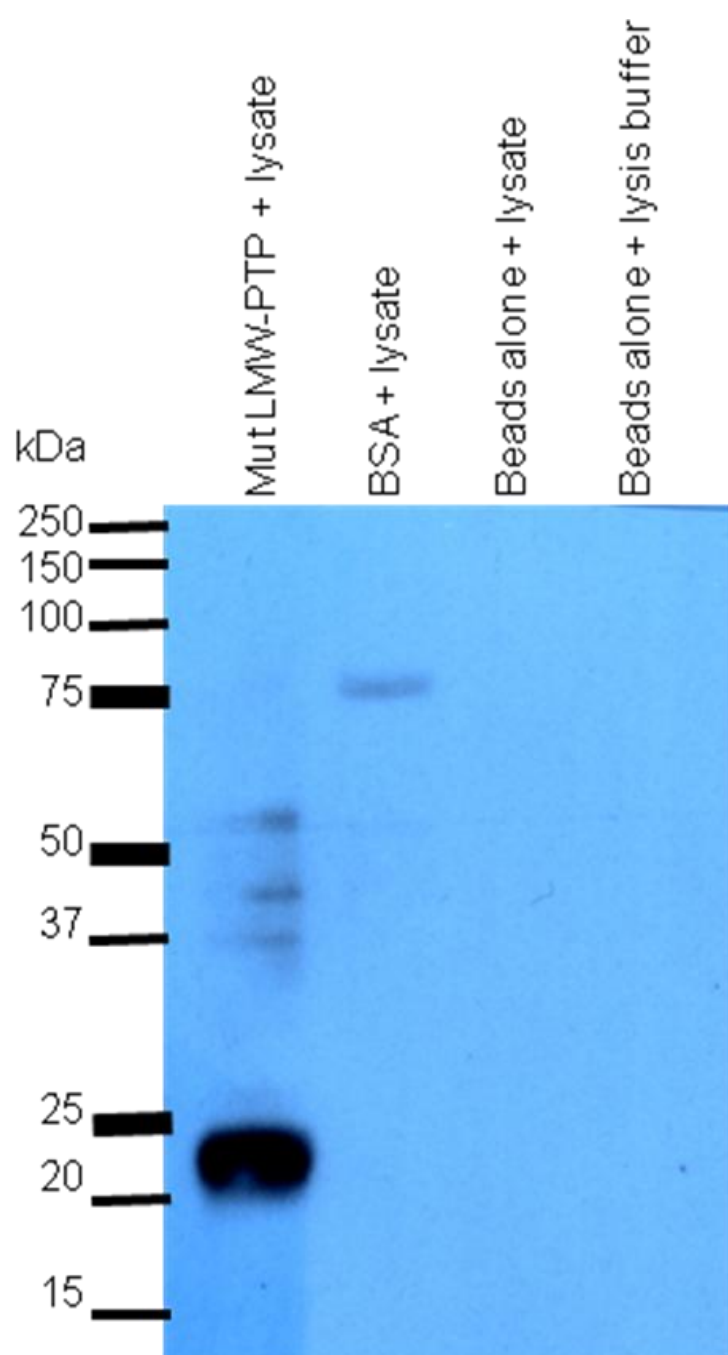


Figure 3-24. A representative silver stain of pulldowns using mutant LMW-PTP or BSA coupled to Affi-Gel 15 beads and incubated with amebic lysate or lysis buffer only.

Recombinant mutant LMW-PTP or BSA was attached to Affi-Gel 15 beads overnight at 4°C on a nutating mixer. Mock protein attachments were done with beads alone. HM1:IMSS amebae were pre-treated with sodium orthovanadate, harvested, and lysed. Beads with LMW-PTP or BSA attached or mock attachment were incubated overnight at 4°C on a nutating mixer with either amebic lysate or with lysis buffer only. Each sample was split into two sets. One set of samples was immunoblotted and probed with anti-phosphotyrosine antibody to assay for phosphorylated proteins (Figure 3-23). The second set was then subjected to SDS-PAGE, and silver-stained.

There appear to be two unique bands (indicated by red arrows) in the silver-stained samples of the Affi-Gel 15 beads + mutant LMW-PTP + amebic lysate, that are not present in the Affi-Gel 15 beads + mutant LMW-PTP + lysis buffer only or the BSA or beads-only samples, that are the same size, ~37 and ~45 kDa, as two of the three bands in the phosphotyrosine immunoblot (Figure 3-23). Some of the attached recombinant mutant LMW-PTP was released from the Affi-Gel 15 beads when samples were boiled before loading, and is running between 20 and 25 kDa rather than its usual 17 kDa. This is likely due to the large amount of protein (500 µg per sample) that was attached to the Affi-Gel 15 beads. The BSA is also running at a larger apparent size, about ~75 kDa, rather than its expected ~66 kDa [352] for likely the same reason. The larger apparent size of the mutant recombinant LMW-PTP for these substrate-trapping studies is also seen in the immunoblot in Figure 3-23 and the silver stained gels in Figure 3-25.

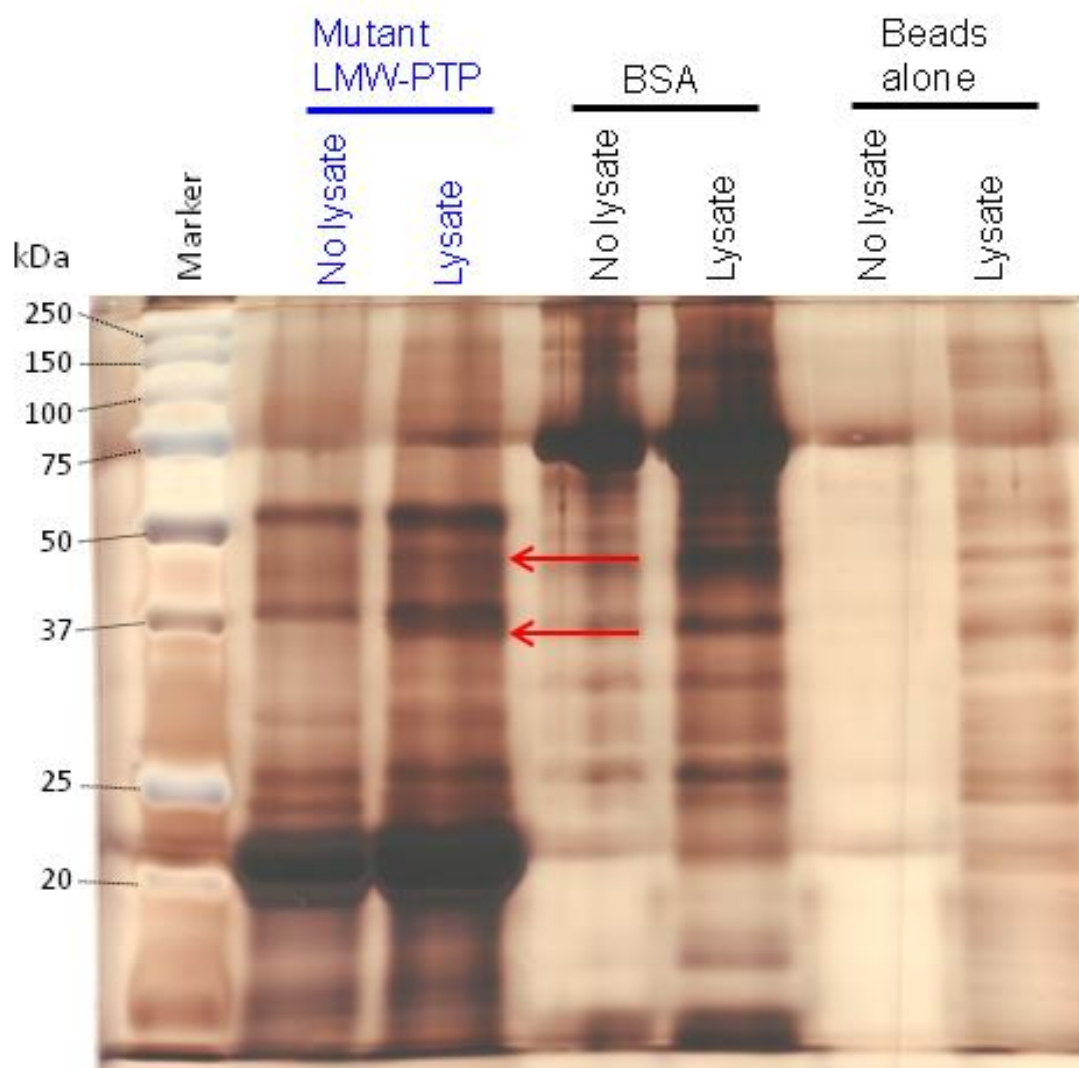


Figure 3-25. Silver stains submitted for mass spectrometry analysis of pulldowns using mutant LMW-PTP or BSA coupled to Affi-Gel 15 beads and incubated with amebic lysate or lysis buffer only. Red arrows = bands unique to the Affi-Gel 15 + mutant LMW-PTP + lysate (Figure 3-24) that are also phosphorylated (Figure 3-23) with sizes of about ~37 and ~45 kDa. Recombinant mutant LMW-PTP or BSA was attached to Affi-Gel 15 beads overnight at 4°C on a nutating mixer. Mock protein attachments were done with beads alone. HM1:IMSS amebae were pre-treated with sodium orthovanadate, harvested, and lysed. Beads with LMW-PTP or BSA attached or mock attachment (beads alone) were incubated overnight at 4°C on a nutating mixer with either amebic lysate or with lysis buffer alone. Samples were subjected to SDS-PAGE, the gels silver-stained, and entire gels submitted for mass spectrometry analysis. Some of the attached recombinant mutant LMW-PTP was released from the Affi-Gel 15 beads when samples were boiled before loading, and is running between 20 and 25 kDa rather than its usual 17 kDa. This is likely due to the large amount of protein (500 µg per sample) that was attached to the Affi-Gel 15 beads. The larger apparent size of the mutant recombinant LMW-PTP for these substrate-trapping studies is also seen in the immunoblot in Figure 3-23 and the silver stained gels in Figure 3-24. In Figure 3-24, the control BSA attached to beads also runs at a larger apparent size, about ~75 kDa, rather than its expected ~66 kDa [352] for likely the same reason.

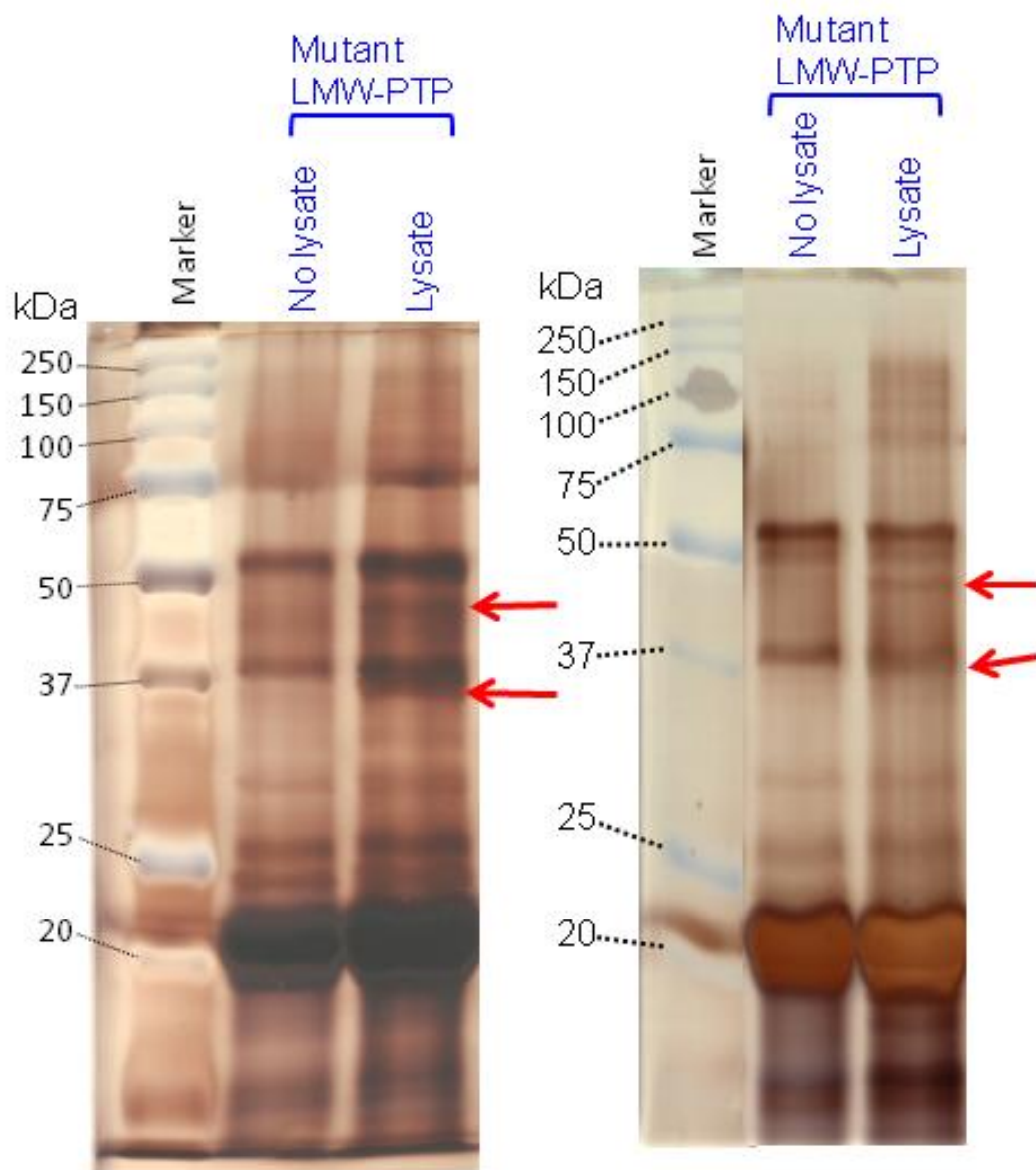


Figure 3-26. Sequences of unique peptides and their locations in the protein sequences of putative substrate proteins 328t.00002 and type A flavoprotein identified by mass spectrometry (Tables 3-6, 3-7, and 3-8). Locations of the unique peptides are highlighted in the protein sequence, and the list of unique peptides identified for each protein is in a table to the right of the protein sequence. The hypothetical protein 328t.00002 had 12 unique peptides representing 36% coverage of the protein sequence; the type A flavoprotein had 15 unique peptides representing 44% coverage of the protein sequence (Table 3-8).

Protein sequence of hypothetical protein 328t.00002 (gi56465624)

MNAFILLFIALATAEDIYSPFKRSCVIQ
 NRINTITAQTKAIHNKIRADEQQERNLN
 DSLYYLYRDLKYATVAEDRARLELEIEA
 IQKKLHSVRIHKVNMLRDIRKETDKITA
 PFRDSIARSNKIERHVGLEDEGNFVDER
 LKTQATIKKVTSQLANKYATLAAKIAAQ
 SNVAKAAKFFNNDAAKVQAFVAKKAEAA
 YTKVHISKIVKAISAAVENGVARNNVETV
 AKTAIEQIVDKLMMKVAAQPSVEAADKV
 AON

Peptide Sequence

(K) YATVAEDR (A)
(K) VTSQLANK (Y)
(K) VAAQPSVEAADK (V)
(K) TAIEQIVDK (L)
(R) NNVETVAK (T)
(R) LELEIEA IQK (K)
(K) KVTSQLANK (Y)
(K) KFNNDAAK (V)
(R) INTITAQTK (A)
(K) IAAQSNVAK (A)
(R) HVGLEDEGNFVDER (L)
(K) AISAAVENGVAR (N)

Protein sequence of type A flavoprotein (gi56474182)

MKALEVVKDLYWVGVDKELRVFDIIMTT
 PYGTSYNSFLLKSEKGNVLFETCKENFAG
 ECLERIEDVIGKEGKLDYIVLNHTEPDHS
 GSLVHILEKYPEATVIGTMAALNNIKYIG
 HIKENTKTLNSGKIKTLDLGNYHLKFLIQ
 PFLHWPDTMMTVIEEMKVLVSCDVFGGHY
 ADERVFNDQMMERIKDMDDAYKHYFDCIF
 GPFKNYVIKGLDMIETQMGFPSDELKAIC
 CSHGPVLRTHIKENIERYRQWAQPIALKN
 KVVIAYGSA YGYTQKMAEQISEGIKSTGV
 EVKMFNIVESSVGDVLKEFEDAKGLLLGT
 PTLVNDTIPPIMQIACSLNPTIHCNRFVQ
 CFGSFGWSGEGVKNLSARIVQLKVHPVE
 PLSIKFQPN SNELQTCFEWGKKFAEALKA

Peptide sequence

(K) ENFAGEcLER (I)
(K) FQPN SNELQTCFEWGK (K)
(R) FVQcFGSFGWSGEGVK (N)
(K) GLDMIETQMGFPSDELK (A)
(K) GNVLFETcK (E)
(K) GNVLFETcKENFAGEcLER (I)
(K) HYFDcIFGPFK (N)
(R) IKDMDDAYK (H)
(K) MAEQISEGIK (S)
(K) MFNIVESSVGDVLK (E)
(K) TLDLGNYHLK (F)
(R) VFNDQmmER (I)
(K) VLVS cDVFGGHYADER (V)
(K) VVIAYGSA YGYTQK (M)
(K) YPEATVIGTmAALNNIK (Y)

Figure 3-27. Sequences of unique peptides and their locations in the protein sequences of the putative substrate proteins putative protein kinase and conserved hypothetical protein gi56465028 identified by mass spectrometry (Tables 3-7 and 3-8). Locations of the unique peptides are **highlighted** in the protein sequence, and the list of unique peptides identified for each protein is in a table to the right of the protein sequence. The putative protein kinase gi56466736 had 9 unique peptides representing 31% coverage of the protein sequence; the conserved hypothetical protein gi56465028 had 7 unique peptides representing 21% coverage of the protein sequence (Table 3-8).

**Protein sequence of putative
protein kinase (gi56466736)**

MTTPTKLQVSQMOKINWVAECSKDF
 EVVEKLGEGSYGSVFHAIHRPTGTH
 AAVKTVRIDGDTTDVEKEIVLMKQC
 DTDYIVRFFGHFFVDKDLWIVMEIC
 GAGSVCDIMNINSTLDEREQIIL
 KDTLHGLAYLHKMHMIHRDIKAANI
 LVSDDGVCCKLADFGVSTKMGGTNKL
 QKTLIGTPYWMapevileSGHNEKA
 DIWSVGITAIEMFDGRPPHSELHAM
 RAIFLIPNRPPPTLSTPEEVSKAFN
 DFIALCCVKNFRERPSALELLKCDF
 MKPNGNRKRITDLIEKENEVIAEFG
 SRKKALEQGEDDDEGDDEZDDDDDT
 EEDDTEEEEDDTEEDDSDDGDDTM
 VIRK

Peptide sequence

(K) AANILVSDDGVcK (L)
(K) AFNDFIALccVK (N)
(R) AIFLIPNRPPPTLSTPEEVSK (A)
(K) DFEVVEK (L)
(K) DTLHGLAYLHK (M)
(R) ERPSALELLK (C)
(R) IDGDTTDVEK (E)
(R) ITDLIEKENEVIAEFGSR (K)
(K) LADFGVSTK (M)

**Protein sequence of the conserved
hypothetical protein (gi56465028)**

MEELINKNNQLIKEVVETLPNIQQSIDD
 VNNNYQSYLESIKNEVFELKENETYSEA
 IKRGEIGIFDRMDETERRRKLVIKESQET
 INQIEIMKNKISKIIEECNNDEETFGKT
 RNELIDKIIISIEEMIMKDKLFRRTEETD
 TITSEFNKKKEEWKHHYSDYLERKKRKE
 EEEKERKKQEEERLQIMKGMNSMEEMR
 QIEEWTNKKVGNVLFESDVDDWNKNTSV
 FDQKIKNKEHIMIVIEDEEGNKFGGYVN
 SKIDEVESCITDQKSFLSLESKGRMKR
 MMKFDIKEPQCAFYLNSTDDCLFEFGF
 GHDICVFKENNKTSYCYQRSEFYNGIS
 NALCGKQLPERFTQKRIMVIEMK

Peptide Sequence

(K) VGNVLFESDVDDWNK (N)
(K) GmNSMEEMR (Q)
(K) IIEEcNNDEETFGK (T)
(K) KVGNVLFESDVDDWNK (N)
(K) NTSVFDQK (I)
(R) RTEETDTITSEFNK (K)
(R) SFEYNGISNALcGK (Q)

Figure 3-28. Sequences of unique peptides and their locations in the protein sequences of putative substrate proteins putative ARP2/3 complex 34 kDa subunit and the hypothetical protein 503.t00001 identified by mass spectrometry (Tables 3-6, 3-7 and 3-8). Locations of the unique peptides are highlighted in the protein sequence, and the list of unique peptides identified for each protein is in a table to the right of the protein sequence. The putative ARP2/3 complex 34 kDa subunit had 7 unique peptides representing 27% coverage of the protein sequence; the hypothetical protein 503.t00001 also had 7 unique peptides, representing 19% coverage of the protein sequence (Table 3-8).

Protein sequence of the putative ARP2/3 complex 34 kda subunit (gi56471818)

MAPNLLLESYNRIVYARVLEKLTTEGGIKK
 GLKEKLADFNQSIWYIERTKEKMTISVSTP
 CWKQLKENRSLEFLQKVYGS DLVEADSNSE
 YNLILTIPKHKDPPEEFAMNAAKLLTNMLI
 GPAVVLADEVKNNKADEKLVQIDYRPGESY
 WLKPNGDRLTVIFSIKFDDKDDAVFGRVFI
 NEFSKSAAGCPSCDVVTRKNAPPPSELKTV
 EGLAEDNCYISFLLEKRHLNNPQKTLEVL
 MSCRNYINFHIKCSKA

Peptide sequence

(K) FDDKDDAVFGR (V)
(K) LADFNQSIWYIER (T)
(K) LLTNMLIGPAVVLADEVK (N)
(K) SAAGcPScDVVTR (K)
(R) SLEFLQK (V)
(K) TLEVLmScR (N)
(R) VFINEFSK (S)

Protein sequence of the hypothetical protein 503.t00001 (gi56464309)

MSGIGDAFGRKFYQIKTHVGAGQKTMDSDVQ
 YAKNKLSESYKKFKNILDVIKKLAPT VHATN
 LMQVEVLTS LGDCVVNTSPETKSDIDS IIST
 FQKIDEGVNTYETRIESDIIVPLKTYMEQFK
 VMEKRFEICHNRRVDMDRYHDSVLSISKPP
 GKQSGLGEAQNKYNVARDLYNYLRNEI IADV
 EKLTSSAEVVS PICGTLIVSYTDYLNHLNN
 YWGEASEVSSNFRICALDPAPIITPGESSMV
 IDANVFSKKSSDVMGGTYEPSEFENSGSSTT
 TSTTTTTTTTVSTSSGKRPPPPPPSRKEQVR
 CEYEYQAQEQGELSFKEGDIITVLKKEGDWW
 LGELKDNKDTSHTIMFLHFKFTFELKDYFIS
 FSSLFIFQNKEL

Peptide sequence

(K) EGDWWLGELK (D)
(K) IDEGVNTYETR (I)
(R) IESDIIVPLK (T)
(K) QSGLGEAQNK (Y)
(K) SDIDSIIISTFQK (I)
(K) TmDSDVQYAK (N)
(R) YHDSVLSISK (K)

Figure 3-29. Sequences of unique peptides and their locations in the protein sequences of putative substrate proteins putative ribose-phosphate pyrophosphokinase and the calcium-binding and URE3-sequence specific DNA binding protein (URE3-BP) identified by mass spectrometry (Tables 3-6, 3-7 and 3-8). Locations of the unique peptides are **highlighted** in the protein sequence, and the list of unique peptides identified for each protein is in a table to the right of the protein sequence. The putative ribose-phosphate pyrophosphokinase had 4 unique peptides representing 13% coverage of the protein sequence; the URE3-BP had 2 unique peptides, representing 15% coverage of the protein sequence (Table 3-8).

Protein sequence of the putative ribose-phosphate pyrophosphokinase (gi56473321)

MSVTTLSAEQWQHDFIVHTSKAGYLAQS
 ISAVLGYPIIPIIHKKFS~~SDGEC~~YYRIAIPE
 RTSLVGRDVVVVSSIVDDQELLEVVRI
 LAELGTRRIFVPIPYLMYSTMERAVHPGEV
 VTCKSTVRMLCGIPSSGLGNLFMLMDLHTS
 GIIHYFEGTVQAMELYAESTLEQAIENID
 FSEPVIFGSADLGRPLWVETFANHFGV
 FIRKSRSEFEDTHVIGEPIGDVKGKHV
 VIYD DLTRSAGSLIKACNAYDLNGAIKVTAVVTH
 LSLIDDDVVQKVIDSRIDRVIATNSN
 VRSQMDAVKNSKKFII~~CDISPVFAKQ~~IRACLNN
 ND

Peptide sequence

(R) AVHPGEVVTcK (S)

(K) FSDGEcYYR (I)

(K) KFIIcDISPVFAK (Q)

(R) VIATNSNVR (S)

Protein sequence of the calcium-binding and URE3-sequence specific DNA binding protein (URE3-BP) (gi56463918)

MQPPVANFCLWNLQPIQGSWMGAA
 CIYQMPPSVRNTWWFLLNTIPLD
 QYTRIYQWFMGVDRDRSGTLEINE
 LMMGQFPGGIRLSPQTALRMMRIF
 DTDFNHISFYEFMAMYKFMELAY
 NLFVMNDRNRSGTLEPHEILPALQ
 QLGFYINQRTSLLLHRLFARGMAF
 CDLNCWIAICAFAAQTRSAYQMIF
 MNPYYGPMKPFNPMEFGKFLDVVT
 SLLE

Peptide sequence

(K) FLDVVTSLLE (-)

(R) SGTLEPHEILPALQQLGFYINQR (T)

Figure 3-30. Protein sequences and predicted tyrosine phosphorylation sites in the selected proteins from mass spectrometry analysis selected for overexpression and further analysis.

Please see Table 3-8 for a list of these proteins. The NetPhos 2.0 server [362] was used to predict tyrosine phosphorylation sites in these proteins. The NetPhos 2.0 server produces neural network predictions for serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins.

Specific recognition of a phosphotyrosine residue occurs in the context of a peptide sequence, with the flanking residues conferring additional binding affinity of the enzyme to its substrate [279]. The protein sequences are shown, and predicted phosphorylated tyrosines are shown in **bold** and are **highlighted**, with their surrounding residues in a table below. The higher the score is (closer to a value of one), the more likely the phosphorylation of that tyrosine residue.

hypothetical protein 328.t00002 (gi56465624; XM_644144)

1 MNAFILLFIALATAEDIYSPFKRSCVIQNRINTITAQTKAIHNKIRADEQ
 51 QERNLNDSLYLYRDLKYATVAEDRARLELEIEAIQKKLHSVRIHKVNML
 101 RDIRKETDKITAPFRDSIARSNKIERHVGLEDEGNFVDERLKTQATIKKV
 151 TSQLANKYATLAAKIAAQSNVAKAAKFFNNDAAKVQAFVAKKAEAA^vTKV
 201 HSKIVKAISAAVENGVARNNVETVAKTAIEQIVDKLMMKVAAQPSVEAAD
 251 KVAQNIVKSQKQIKKAVKVTPKKVVKKPVKKAVAKKVVKKAAPKPAACK
 301 IAKK

Phosphorylation sites predicted: Tyr: 3

Name	Pos	Context	Score	Pred
328.t00002	18	AEDI ^v YSPFK	0.948	*Y*
328.t00002	68	RDLKYATVA	0.970	*Y*
328.t00002	197	AEAA [^] YTKVH	0.829	*Y*

type A flavoprotein (gi56474182; XM_651854)

1 MKALEVVKDLYWVGVDKELRVFDIIMTTPYGTSYNSFLLKSEKGNVLFEE
 51 TCKENFAGECLERIEDVIGKEGKLDYIVLNHTEPDHSGSLVHILEKYPEA
 101 TVIGTMAALNNIKYIGHIKENTKTLNSGKIKTLDLGNHYHLKFLIQPFLHW
 151 PDTMMTVIEEMKVLVSCDVFGGHYADERVFNQMMERIKDMDDAY^vKHYFD
 200 CIFGPFKN^vVIKGLDMIETQMGFSPDELKAICCSHG^vPVLRTHIKENIERY
 251 RQWAQPIALKNKVVIAYGSAYGYTQKMAEQISEGIKSTGVEVKMFNIVES
 301 SVGDVLKEFEDAKGLLLGTPTLVNDTIPPIMQIACSLNPTIHCNRFVQCF
 351 GSGFWSGEGVKNLSARIVQLKVHQ^vPVEPLSIKFQPN^vSNELQTCFEWGK^vKF
 401 AEALKA

Phosphorylation sites predicted: Tyr: 4

Name	Pos	Context	Score	Pred
Flavopr	76	GKLD ^v YIVLN	0.821	*Y*
Flavopr	174	FGGHYADER	0.795	*Y*
Flavopr	195	MDDAY ^v KHYF	0.513	*Y*
Flavopr	209	PFKN ^v VIKG	0.574	*Y*

Figure 3-30 is continued on page 242.

Figure 3-30, continued from page 241.

protein kinase, putative (gi56466736; XM_645070)

```

1   MTTPTKLQVSQMOKINWVAECSKDFEVVEKLGEGSYGSVFHAIHRPTGTH
51  AAVKTVRIDGDTTDVEKEIVLMKQCCTDYIVRFFGHFFVDKDLWIVMEIC
101 GAGSVCDIMNILNSTLDEREIQIILKDTLHGLAYLHKMHMIHRDIKAANI
151 LVSDDGVCKLADFGVSTKMGGTNKLQKTTLIGTPYWMAPEVILESGHNEKA
201 DIWSVGITAIEMFDGRPPHSELHAMRAIFLIPNRPPPTLSTPEEVSKAFN
251 DFIALCCVKNFRERPSALELLKCFMKNRKRITDLIEKENEVIAEFG
301 SRKKALEQGEDDDEGDDEQDDDDTEEDDTEEEEDDTEEDDSDDGDDTM
351 VIRK

```

Phosphorylation sites predicted: Tyr: 2

Name	Pos	Context	Score	Pred
		v		
ProtKina	36	GEGSYGSVF	0.816	*Y*
ProtKina	79	CTDYIVRF	0.809	*Y*
		^		

conserved hypothetical protein (gi56465028, XP_001914335)

```

1   MEELINKNNQLIKEVVETLPNIQQSIDDVNNNYQSYLESIKNEVFELKEN
51  ETYSEAIKRGEIGFDRMDETERRRKLVIEKSQETINQIEIMKNKISKIIE
101 ECNNDEETFQKTRNELIDKIIISIEEMIMKDKLFRRTEETDTITSEFNKKK
151 EEWKHYSDYLERKKRKEEEEEKERKKQEEERLQIMKGMNSMEEMRQIEE
201 WTNKKVGNVLFESDVDDWNKNTSVFDQKIKNKEHIMIVIEDEEGNKFGGY
251 VNSKIDEVESCITDQKSFLFSLESKGRMKRMMKFDIKEPQCAFYLNSTQD
301 DCLFEFGFGHDICVFKENKNTQSYCYQRSFEYNGISNALCGKQLPERFTQ
351 KRIMVIEMK

```

Phosphorylation sites predicted: Tyr: 3

Name	Pos	Context	Score	Pred
		v		
Cons hyp	53	ENETYSEAI	0.927	*Y*
Cons hyp	157	WKKHYSDYL	0.562	*Y*
Cons hyp	250	KFGGYVNSK	0.953	*Y*
		^		

Figure 3-30 is continued on page 243.

Figure 3-30, continued from page 242.

ARP2/3 complex 34 kDa subunit, putative (gi56471818, DS571335)

```

1   MAPNLLLESYNRIVYARVLEKLTTEGGIKKGLKEKLADFNQSIWYIERTK
51  EKMTISVSTPCWKQLKENRSLEFLQKVYGSDDLVEADSNSEYvNLILTIPEK
101 HDKPEEFAMNAAKLLTNMLIGPAVVLADDEVKNNKADEKLVQIDYRPGESYv
151 WLKPNGDRLTVIFSIKFDKDDAVFGRVFINEFSKSAAGCPSCDVVTRKN
201 APPPSELKTVEGLAEDNCYvISFLLEKRHLNPNQKTLEVLMSCRNYvINFHI
251 KCSKAFLHIRMRNKVSHLQLVLNRAKPEREVEKKTASGRTFKK

```

Phosphorylation sites predicted: Tyr: 4

Name	Pos	Context	Score	Pred
		v		
Arp 2/3	91	SNSEY ^v NLIL	0.767	*Y*
Arp 2/3	150	PGESY ^v WLKP	0.721	*Y*
Arp 2/3	219	EDNCY ^v ISFL	0.884	*Y*
Arp 2/3	245	SCRNY ^v INFH	0.737	*Y*
		^		

hypothetical protein 503.t00001 (gi56464309, XP_656019)

```

1   MSGIGDAFGRKFYQIKTHVGAGQKTMSDVQYvAKNKLSSEYvKKKFKNILDV
51  IKKLAPTvVHATNLMQVEVLTSLGDCVVNTSPETKSDIDSIIISTFQKIDEG
101 VNTYETRIESDIIIVPLKTYMEQFKVMEKRFEICHNRRVDMDRYvHDSVLSI
151 SKKPPGKQSGLGvEAQNKNVARDLYNLRNEIIADVEKLTSSAAEEVVSPI
201 CGTLIVSYTDYvLNHLNvNYWGEASEVSSNFRICALDPAPIITPGESSMVID
251 ANVFSKKSvSDVMDGTvYEPSEFENSGSSTTTSTTTTTTTVSTSSGKRPPPP
301 PPSRKEQVRCEYEYvQAQEQGELSFKEGDIITVLKKEGDWWLGELKDNKDT
351 SHTIMFLHFKFTFELKDYFISSSLFIFQNKEL

```

Phosphorylation sites predicted: Tyr: 6

Name	Pos	Context	Score	Pred
		v		
501.t	32	SDVQY ^v AKNK	0.957	*Y*
501.t	41	LSESY ^v KKFK	0.519	*Y*
501.t	143	DMDRY ^v HDSV	0.971	*Y*
501.t	211	SYTDY ^v LNHL	0.927	*Y*
501.t	266	MDGT ^v YEPSE	0.607	*Y*
501.t	314	CEYEY ^v QAQE	0.575	*Y*
		^		

Figure 3-30 is continued on page 244.

Figure 3-30, continued from page 243.

ribose-phosphate pyrophosphokinase, putative (gi56473321, XP_656141)

```

1   MSVTTLSLAEQWQHDHFIVHTSKAGYLAQSSISAVLGYPIIPIIHKKFS
51  ECYYYRIAIPERTSLVGRDVVVVSSIVDDQELLEVVVRIGGELAE
101 FVIPYLMYSTMERAVHPGEVVTCKSTVRMLCGIPSSGLGNLFMLMDL
151 GIIHYFEGTVQAMELYAESTLEQAI AENIDFSEPVIFGSADLGRPLW
201 FANHFGVGI AFIRKSR SFEDTHVIGEPIGDVKGKHVVIIYDDLTR
251 KACNAYLDNGAIKVTAVVTHLSLIDDDVVQKVIDSRIDRVIATNSN
301 MDAVKNSKKFIICDISPVFAKQIRACLNNND

```

Phosphorylation sites predicted: Tyr: 3

Name	Pos	Context	Score	Pred
		v		
Rib-phos	53	DGEC ^Y YRIA	0.784	*Y*
Rib-phos	166	AMEL ^Y AEST	0.956	*Y*
Rib-phos	256	ACNA ^Y LDNG	0.947	*Y*
		^		

calcium-binding protein, putative, URE3-sequence specific DNA binding protein (URE3-BP) (gi56463918, AAG18423)

```

1   MQPPVANFCLWNLQPIQGSWMGAACIYQMPPSVRNTWWFPLLNTIPLDQY
51  TRIYQWFMGVDRDRSGTLEINELMMGQFPGGIRLSPQTALRMMRIFD
101 NGHISFYEFMAMYKFMELAYNLFVMNDRNRSGTLEPHEILPALQQLGF
151 NQRTSLLLHRLFARGMAFCDLNCWIAICAFAAQTRSAYQMI FMPNY
201 KPFNPMEFGKFLDVVTSLE

```

Phosphorylation sites predicted: Tyr: 2

Name	Pos	Context	Score	Pred
		v		
URE3-BP	50	PLDQ ^Y TRII	0.804	*Y*
URE3-BP	197	MNP ^Y GPMK	0.698	*Y*
		^		

Figure 3-31. Schematic representation of epitope-tagged putative substrate genes subcloned into pGIR209. Putative substrate genes identified via mass spectrometry were cloned into pBluescript II KS(+) with epitope tags E, S, and KT3 added, sequence-verified, and subcloned into pGIR209. 5' actin and 3' actin, 5' and 3' flanking regions of the *Entamoeba histolytica* actin gene; *Neo*, ORF (open reading frame) of neomycin resistance gene; *luc*, ORF of luciferase gene; 5' *hgl* and 3' *hgl*, 5' and 3' flanking regions of *Entamoeba histolytica hgl* gene; tetO, tetracycline operator; Putative Substrate, ORFs (open reading frames) of genes for putative substrate proteins of LMW-PTP identified by mass spectrometry. N-tag, N-terminally added E-S-KT3 epitope tag; C-tag, C-terminally added E-S-KT3 epitope tag. The epitope tags and the putative substrate genes were cloned first into pBluescript II KS (+). After sequence verification, constructs were then subcloned into pGIR209 [177, 179]. As a control for transfection, drug selection, and expression of an epitope-tagged protein in amebae, the luciferase reporter gene had the epitope tag E-S-KT3 N-terminally added. The pGIR209 and pGIR308 diagrams are adapted from Ramakrishnan *et al* (1997) [177] and Katz *et al* (2003) [179].

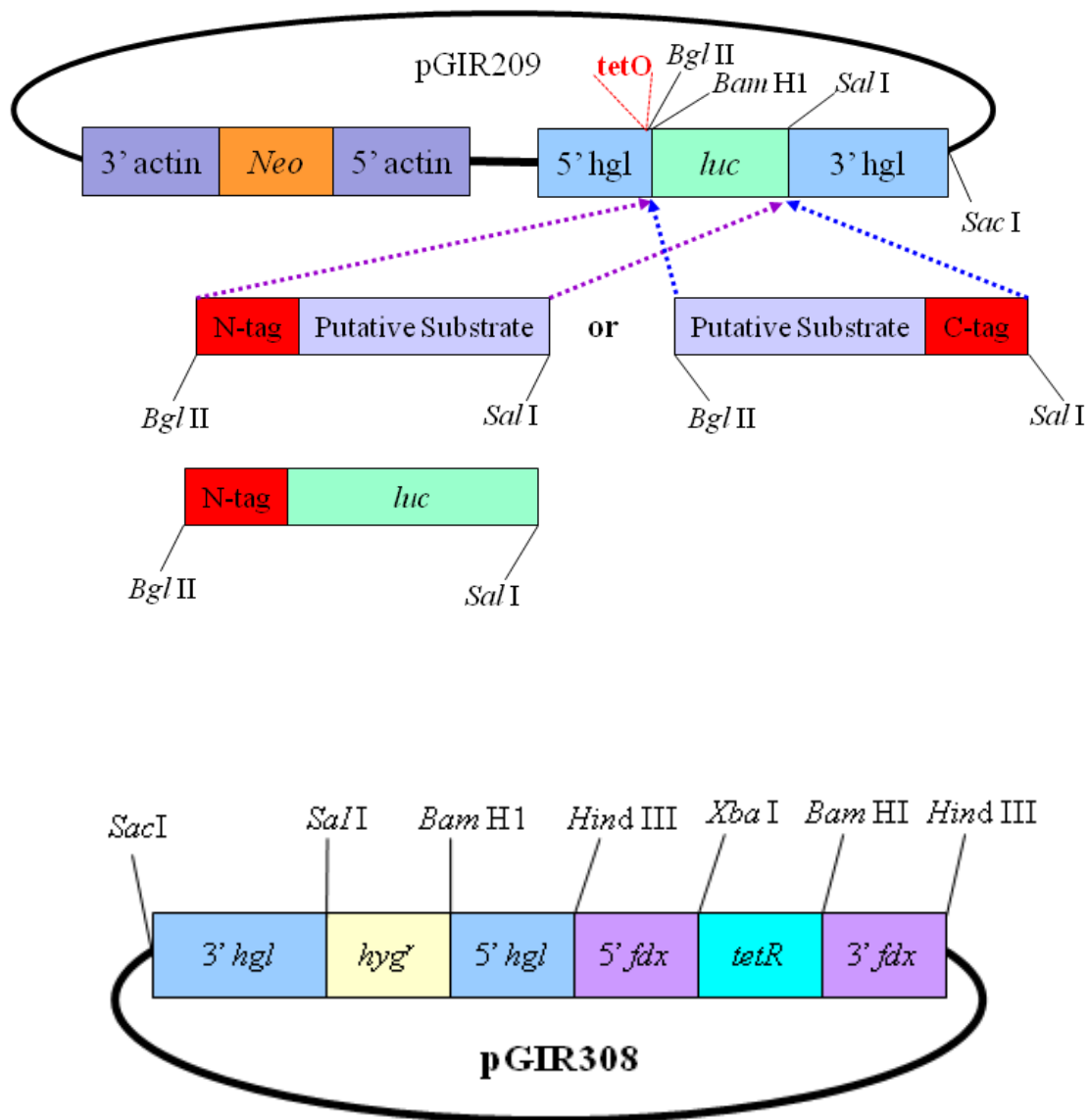


Table 3-1. Comparison of *Entamoeba histolytica* LMW-PTPs to selected representative

LMW-PTPs found in plants. The *Entamoeba histolytica* LMW-PTP protein sequence for XP_656359 was input into BLAST [311] in order to identify related LMW-PTP protein sequences and align protein sequences of related LMW-PTPs from other species. LALIGN [312] was also used to check alignments. The first part of the table gives the abbreviation of the species name, the GenBank accession number, GI number, the species and strain to which the compared LMW-PTP belongs, and the protein description and annotation. The second part of the table shows which residues of the *E. histolytica* LMW-PTP protein XP_656359 are aligned with each compared protein sequence, the level of sequence identity to the *E. histolytica* LMW-PTP sequence (identical amino acid residues for that position in the sequence), the level of sequence similarity to the *E. histolytica* LMW-PTP sequence (either an identical or a conservatively substituted residue for that position in the sequence), and the presence of gaps in the alignment between the *E. histolytica* LMW-PTP sequence and the compared LMW-PTP sequence. Please see Figure 3-3 for sequence alignments for these LMW-PTPs. A phylogenetic tree comparing all selected LMW-PTPs from all organisms is shown in Figure 3-7.

Abbreviation	Protein Accession Number	GI Number	Species/strain	Protein Description/Annotation
Ent his 1	XP_656359	gi67482019	<i>Entamoeba histolytica</i> HM1:IMSS	protein tyrosine phosphatase, putative
Ent his 2	XP_653357	gi67475326	<i>Entamoeba histolytica</i> HM1:IMSS	protein tyrosine phosphatase, putative
Sorg bicol	XP_002445697	gi242081857	<i>Sorghum bicolor</i> (sorghum)	hypothetical protein SORBIDRAFT_07g024360
Zea mays 1	NP_001149867	gi226501280	<i>Zea mays</i> (maize)	low molecular weight protein-tyrosine-phosphatase slr0328
Zea mays 2	ACN25957	gi223943747	<i>Zea mays</i> (maize)	Unknown low molecular weight phosphatase family protein
Oryza sat	AAO72543	gi29367341	<i>Oryza sativa</i> Japonica Group (rice)	Japonica cultivar-group isolate 20132 protein tyrosine phosphatase-like protein
Hord vulg	BAJ87374	gi326510315262	<i>Hordeum vulgare</i> subsp. <i>vulgare</i> (domesticated barley)	predicted protein
Arab thal	CAB72464	gi6967110	<i>Arabidopsis thaliana</i> (mouse-ear cress)	protein-tyrosine-phosphatase-like protein
Vit vinif	CBI17062	gi297734828	<i>Vitis vinifera</i> (wine grapevine)	unnamed protein product
Pop tricho	XP_002328065	gi224131612	<i>Populus trichocarpa</i> (black cottonwood tree)	<i>Populus trichocarpa</i> predicted protein
Ricin comm	XP_002529074	gi255576363	<i>Ricinus communis</i> (castor bean plant)	low molecular weight protein-tyrosine-phosphatase, putative

Table 3-1 is continued on page 249.

Table 3-1, continued from page 248.

Abbreviation	Protein Length	Residues of XP_656359 aligned	% Identity to XP_656359	% Similarity to XP_656359	Gaps in the sequence alignment
Ent his 1 (XP_656359)	157 aa	1-157	157/157 (100.0%)	157/157 (100.0%)	0/157 (0.0%)
Ent his 2	157 aa	1-157	156/157 (99.4%)	156/157 (99.4%)	0/157 (0.0%)
Sorg bicol	279 aa	3-147	81/154 (52.6%)	103/154 (66.9%)	9/154 (5.8%)
Zea mays 1	274 aa	3-147	79/154 (51.3%)	103/154 (66.9%)	9/154 (5.8%)
Zea mays 2	186 aa	3-147	79/154 (51.3%)	103/154 (66.9%)	9/154 (5.8%)
Oryza sat	268 aa	3-147	81/154 (52.6%)	104/154 (67.5%)	9/154 (5.8%)
Hord vulg	262 aa	3-147	81/154 (52.6%)	102/154 (66.2%)	9/154 (5.8%)
Arab thal	177 aa	2-152	73/154 (47.4%)	97/154 (63.0%)	9/154 (5.8%)
Vit vinif	182 aa	3-147	78/154 (50.6%)	100/154 (64.9%)	9/154 (5.8%)
Pop tricho	173 aa	3-147	78/154 (50.6%)	100/154 (64.9%)	9/154 (5.8%)
Ricin comm	172 aa	3-147	77/154 (50.0%)	101/154 (65.6%)	9/154 (5.8%)

Table 3-2. Comparison of *Entamoeba histolytica* HM1:IMSS strain LMW-PTPs to selected representative LMW-PTPs found in bacteria. The *E. histolytica* LMW-PTP protein sequence for XP_656359 was input into BLAST [311] in order to identify and align protein sequences of related LMW-PTPs from other species. LALIGN [312] was also used to check alignments. The first part of the table gives the abbreviation of the species name, the GenBank accession number, GI number, the species and strain to which the compared LMW-PTP belongs, and the protein description and annotation. The second part of the table shows the residues of the *E. histolytica* LMW-PTP protein XP_656359 which are aligned with each compared protein sequence, the level of sequence identity (identical amino acid residues for that position in the sequence), the level of sequence similarity (either an identical or a conservatively substituted residue for that position in the sequence), and if there are gaps, and how many, between the compared protein sequences. Please see Figure 3-4 for sequence alignments for these LMW-PTPs. A phylogenetic tree comparing all selected LMW-PTPs from all organisms is shown in Figure 3-7.

Abbreviation	GenBank Accession Number	GI Number	Species/strain	Protein Description/Annotation
Ent his 1	XP_656359	gi67482019	<i>Entamoeba histolytica</i> HM1:IMSS	protein tyrosine phosphatase, putative
Ent his 2	XP_653357	gi67475326	<i>Entamoeba histolytica</i> HM1:IMSS	protein tyrosine phosphatase, putative
Pseudo aeru	ZP_04935007	gi254241685	<i>Pseudomonas aeruginosa</i> 2192	phosphotyrosine protein phosphatase
Nost punct	YP_001864171	gi186680975	<i>Nostoc punctiforme</i> PCC 73102 (cyanobacterium)	protein tyrosine phosphatase
Lyngbya sp	ZP_01622631	gi119490007	<i>Lyngbya</i> sp. PCC 8106 (cyanobacterium)	protein tyrosine phosphatase
Cyano sp	YP_003887408	gi307152024	<i>Cyanothece</i> sp. PCC 7822 (cyanobacterium)	protein tyrosine phosphatase
Arthro max	ZP_03274593	gi209526061	<i>Arthrospira maxima</i> CS-328 (cyanobacterium)	protein tyrosine phosphatase
Prochloro marin	NP_876138	gi33241196	<i>Prochlorococcus marinus</i> subsp. <i>marinus</i> str. CCMP1375 (marine cyanobacterium)	protein tyrosine phosphatase
Bdello bact	NP_967768	gi42522388	<i>Bdellovibrio bacteriovorus</i> HD100	phosphotyrosine protein phosphatase
Bact caccae	ZP_01958600	gi153805932	<i>Bacteroides caccae</i> ATCC 43185	hypothetical protein BACCAC_00172
Bact theta	NP_811662	gi29348159	<i>Bacteroides thetaiotaomicron</i> VPI-5482	putative protein tyrosine phosphatase
Bact fragil	YP_213610	gi60683466	<i>Bacteroides fragilis</i> NCTC 9343	putative low molecular weight protein tyrosine phosphatase
Parab merd	ZP_02031159	gi154491533	<i>Parabacteroides merdae</i> ATCC 43184	hypothetical protein PARMER_01144
Vibrio chol	ZP_04415055	gi229525650	<i>Vibrio cholerae</i> bv. <i>albensis</i> VL426	low molecular weight protein tyrosine phosphatase

Table 3-2 is continued on page 252.

Table 3-2, continued from page 251.

Abbreviation	Protein Length	Residues of XP_656359 aligned	% Identity to XP_656359	% Similarity to XP_656359	Gaps in the sequence alignment
Ent his 1 (XP_656359)	157 aa	1-157	157/157 (100%)	157/157 (100%)	0/157 (0.0%)
Ent his 2	157 aa	1-157	156/157 (99.4%)	156/157 (99.4%)	0/157 (0.0%)
Pseudo aeru	154 aa	1-149	61/151 (40.4%)	88/151 (58.3%)	4/151 (2.6%)
Nost punct	169 aa	2-151	77/153 (50.3%)	103/153 (67.3%)	4/153 (2.6%)
Lyngbya sp	162 aa	2-147	78/148 (52.7%)	95/148 (64%)	3/148 (2.0%)
Cyano sp	161 aa	2-155	78/155 (50.3%)	100/155 (64.5%)	4/155 (2.6%)
Arthro max	160 aa	2-147	76/148 (51.4%)	100/148 (67.6%)	3/148 (2.0%)
Prochlo marin	156 aa	2-147	70/149 (47.0%)	96/149 (64.4%)	3/149 (2.0%)
Bdello bact	171 aa	2-152	77/152 (50.7%)	98/152 (64.5%)	3/152 (2.0%)
Bact caccae	156 aa	2-152	79/153 (51.6%)	102/153 (66.7%)	3/153 (2.0%)
Bact theta	156 aa	2-152	77/153 (50.3%)	102/153 (66.7%)	3/153 (2.0%)
Bact fragil	157 aa	2-152	74/152 (48.7%)	100/152 (65.8%)	1/152 (0.7%)
Parab merd	169 aa	2-150	75/150 (50.0%)	101/150 (67.3%)	1/150 (0.7%)
Vibrio chol	155 aa	2-152	66/152 (43.4%)	95/152 (62.5%)	4/152 (2.6%)

Table 3-3. Comparison of *E. histolytica* strain HM1:IMSS LMW-PTPs to representative phosphatases found in selected unicellular eukaryotes. These include two other *Entamoeba* species, *E. dispar* and *E. invadens*; *Giardia lamblia* and *Giardia intestinalis*; *Dictyostelium*

discoideum; and the yeasts *Kluyveromyces lactis*, *Saccharomyces cerevisiae* and

Schizosaccharomyces pombe). The *Entamoeba histolytica* LMW-PTP protein sequence for

XP_656359 was input into BLAST [311] in order to identify and align protein sequences of related LMW-PTPs from other species, with one exception: the amoeba database at AmoebaDB

[309] was used to compare *E. histolytica* and the related species *Entamoeba invadens* LMW-PTPs, as the *E. invadens* LMW-PTP sequence was not available in the BLAST database [311].

LALIGN [312] was also used to check alignments. The first part of the table gives the species name abbreviation, the GenBank accession number, GI number, the species and strain to which the LMW-PTP belongs, and the protein description and annotation. The second part of the table

shows the residues of the *E. histolytica* LMW-PTP protein XP_656359 that are aligned with each compared protein sequence, the level of sequence identity (identical amino acid residues for that position in the sequence), the level of sequence similarity (either an identical or a conservatively substituted residue for that position in the sequence), and the presence of gaps between the

compared protein sequences. Please see Figure 3-5 for sequence alignments for these LMW-PTPs. A phylogenetic tree comparing all selected LMW-PTPs from all organisms is shown in

Figure 3-7.

Abbreviation	GenBank Accession Number	GI Number	Species/strain	Protein Description/Annotation
Ent his 1	XP_656359	gi67482019	<i>Entamoeba histolytica</i> HM1:IMSS	protein tyrosine phosphatase, putative
Ent his 2	XP_653357	gi67475326	<i>Entamoeba histolytica</i> HM1:IMSS	protein tyrosine phosphatase, putative
Ent dispar	XP_001736331	gi167382924	<i>Entamoeba dispar</i> SAW760	low molecular weight protein tyrosine phosphatase, putative
Ent invad	EIN_156590	not available	<i>Entamoeba invadens</i> IP1	low molecular weight protein-tyrosine-phosphatase, putative
Gia lam P15	EFO65743	gi308163403	<i>Giardia lamblia</i> P15	low molecular weight protein tyrosine phosphatase
Giard intes	EET01406	gi253745561	<i>Giardia intestinalis</i> ATCC 50581	low molecular weight protein-tyrosine-phosphatase
Dictyo disc	XP_647072	gi66827435	<i>Dictyostelium discoideum</i> AX4	acid phosphatase 1; low molecular weight phosphotyrosine protein phosphatase
Klu lactis	XP_451937	gi50303975	<i>Kluyveromyces lactis</i> NRRL Y-1140	hypothetical protein
Sacch cerev	AAA80146	gi1044794	<i>Saccharomyces cerevisiae</i> S288c	Ltp1p; low molecular weight protein tyrosine phosphatase
Schiz pombe	P41893	gi1172568	<i>Schizosaccharomyces pombe</i>	low molecular weight phosphotyrosine protein phosphatase; small tyrosine phosphatase

Table 3-3 is continued on page 255.

Table 3-3, continued from page 254.

Abbreviation	Protein Length	Residues of XP_656359 aligned	% Identity to XP_656359	% Similarity to XP_656359	Gaps in the sequence alignment
Ent his 1 (XP_656359)	157 aa	1-157	157/157 (100%)	157/157 (100%)	0/157 (0.0%)
Ent his 2	157 aa	1-157	156/157 (99.4%)	156/157 (99.4%)	0/157 (0.0%)
Ent dispar	157 aa	1-157	153/157 (97.5%)	154/157 (98.1%)	0/157 (0.0%)
Ent invad	156 aa	1-151	92/151 (60.9%)	113/151 (74.8%)	0/151 (0.0%)
Giard lamb	154 aa	1-156	53/157 (33.8%)	85/157 (54.1%)	4/157 (2.5%)
Giard intes	154 aa	1-156	54/157 (34.4%)	84/157 (53.5%)	4/157 (2.5%)
Dictyo disc	179 aa	10-176	54/168 (32.1%)	79/168 (47%)	19/168 (11.3%)
Klu lactis	157 aa	1-132	53/134 (39.6%)	73/134 (54.5%)	2/134 (1.5%)
Sacch cerev	161 aa	1-132	45/135 (33.3%)	70/135 (51.9%)	3/135 (2.2%)
Schiz pombe	156 aa	1-147	57/148 (38.5%)	84/148 (56.8%),	1/148 (0.7%)

Table 3-4. Comparison of *E. histolytica* strain HM1:IMSS LMW-PTPs to representative phosphatases found in multicellular organisms. These include human, mouse, cattle, frog, chicken, fruit fly, and zebrafish LMW-PTPs. The *Entamoeba histolytica* LMW-PTP protein sequence for XP_656359 was input into BLAST [311] in order to identify and align protein sequences of related LMW-PTPs. LALIGN [312] was also used to check alignments. The first part of the table gives the abbreviation of the species name, the GenBank accession number, GI number, the species and strain to which the compared LMW-PTP belongs, and the protein description and annotation. The second part of the table shows the residues of the *E. histolytica* LMW-PTP protein XP_656359 which are aligned with each compared protein sequence, the level of sequence identity (identical amino acid residues for that position in the sequence), sequence similarity (the presence of either an identical residue or a conservatively substituted residue for that position in the sequence), and the presence of gaps between the compared protein sequences and the *E. histolytica* LMW-PTP. Please see Figure 3-6 for sequence alignments for these LMW-PTPs. A phylogenetic tree comparing all selected LMW-PTPs from all organisms is shown in Figure 3-7.

Abbreviation	GenBank Accession Number	GI Number	Species/strain	Protein Description/ Annotation
Ent his 1	XP_656359	gi67482019	<i>Entamoeba histolytica</i> HM1:IMSS	protein tyrosine phosphatase, putative
Ent his 2	XP_653357	gi67475326	<i>Entamoeba histolytica</i> HM1:IMSS	protein tyrosine phosphatase, putative
Homo isof b	NP_009030	gi6005988	<i>Homo sapiens</i> (human being)	low molecular weight phosphotyrosine protein phosphatase isoform b (transcript variant 2); acid phosphatase 1, ACP1
Homo isof c	NP_004291	gi4757714	<i>Homo sapiens</i> (human being)	low molecular weight phosphotyrosine protein phosphatase isoform c (transcript variant 3); acid phosphatase 1, ACP1
Mus musc 1	NP_001103709	gi159032062	<i>Mus musculus</i> (house mouse)	low molecular weight phosphotyrosine protein phosphatase isoform 1
Mus musc 2	NP_067305	gi31542070	<i>Mus musculus</i> (house mouse)	low molecular weight phosphotyrosine protein phosphatase isoform 2
Bos taurus	AAI03028	gi74267860	<i>Bos taurus</i> (cattle)	ACP1 protein; acid phosphatase 1
Xenop laev	NP_001108241	gi168693451	<i>Xenopus laevis</i> (frog)	acid phosphatase 1, soluble isoform b
Gal gallus	NP_001034380	gi86129490	<i>Gallus gallus</i> (chicken)	low molecular weight phosphotyrosine protein phosphatase
Dros melan	NP_001027186	gi78706766	<i>Drosophila melanogaster</i> (fruit fly)	primo-1; low molecular weight phosphotyrosine protein phosphatase 1; low molecular weight cytosolic acid phosphatase 1
Danio rerio	NP_001014324	gi62122785	<i>Danio rerio</i> (zebrafish)	low molecular weight phosphotyrosine protein phosphatase; acid phosphatase 1

Table 3-4 is continued on page 258.

Table 3-4, continued from page 257.

Abbreviation	Protein Length	Residues of XP_656359 aligned	% Identity to XP_656359	% Similarity to XP_656359	Gaps in the sequence alignment
Ent his 1 (XP_656359)	157 aa	1-157	157/157 (100%)	157/157 (100%)	0/157 (0.0%)
Ent his 2	157 aa	1-157	156/157 (99.4%)	156/157 (99.4%)	0/157 (0.0%)
Homo isof b	158 aa	3-135	46/135 (34.1%)	71/135 (52.6%)	3/135 (2.2%)
Homo isof c	158 aa	3-135	50/135 (37.0%)	72/135 (53.3%)	3/135 (2.2%)
Mus musc 1	158 aa	3-135	45/135 (33.3%)	70/135 (51.9%)	3/135 (2.2%)
Mus musc 2	158 aa	3-135	49/135 (36.3%)	72/135 (53.3%)	3/135 (2.2%)
Bos taurus	158 aa	3-135	49/135 (36.3%)	73/135 (54.1%)	3/135 (2.2%)
Xenop laev	159 aa	3-149	49/148 (33.1%)	79/148 (53.4%)	1/148 (0.7%)
Gal gallus	158 aa	3-149	56/149 (37.6%)	78/149 (52.3%)	3/149 (2.0%)
Dros melan	155 aa	2-133	51/133 (38.3%)	70/133 (52.6%)	2/133 (1.5%)
Danio rerio	158 aa	3-154	52/147 (35.4%)	77/147 (52%)	3/147 (2.0%)

Table 3-5. Oligos used to clone wild-type and mutant untagged and epitope-tagged LMW-PTPs. Oligo pairs were used to amplify via PCR the gene for the *E. histolytica* LMW-PTP protein XP_656359. HM1:IMSS strain amebic genomic DNA was used initially as the PCR template to clone the untagged wild-type and the Cys7 to Ser7 active site mutant sequences. The codon mutated to generate the Cys7 to Ser7 mutation is shown in **boldface** type in the oligo sequence. For cloning the untagged wild-type and mutant LMW-PTP genes, each had its own forward oligonucleotide for PCR amplification, *Bam*H1TyrPhosF or *Bam*H1TyrPhosFmut, but the same reverse oligo, *Sal*I TyrPhosR, was used for both reactions. Restriction endonuclease recognition sites were added to the oligos such that PCR products could be digested with *Bam*H1 and *Sal*I and ligated into pBluescript II KS(+). To clone wild-type and active site mutant LMW-PTPs that would have epitope tags added on either the 5' or the 3' end of the protein, DNA from the untagged wild-type or mutant LMW-PTP clones were used as PCR templates. To generate clones that would have an epitope tag added to the 3' end of the gene, *Bam*H1 and *Eco*RI restriction endonuclease recognition sites were added to the oligo sequences; those clones that would have an epitope tag added to the 5' end of the gene, *Eco*RI and *Sal*I sites were included, for cloning into pBluescript II KS(+), and ultimately the clones with tags were subcloned into pGIR209.

	Oligo Name	Oligo Sequence
Untagged	<i>Bam</i> H1TyrPhosF	ATG GGA TCC ATG AAG TTG TTG TTT GTA TGT TTA GGC AAC ATT TGT CGA TCT CCT
	<i>Bam</i> H1TyrPhosFmut	ATG GGA TCC ATG AAG TTG TTG TTT GTA AGT TTA GGC AAC ATT TGT CGA TCT CCT
	<i>Sal</i> I TyrPhosR	GGC GTC GAC TTA ATT AAT AAG TTT TCC TTC TTC TAG TTT AAT GAT TTA ATT CTC ACA AGC ATC

Epitope tags to be added to the 3' end of the gene	Tyr Phos WT <i>Bam</i> H1 F	CAT CAG GGA TCC ATG AAG TTG TTG TTT GTA TGT TTA GGC AAC ATT T
	Tyr Phos Mut <i>Bam</i> H1 F	CAT CAG GGA TCC ATG AAG TTG TTG TTT GTA AGT TTA GGC AAC ATT TT
	Tyr Phos <i>Eco</i> RI R	CGG CTG GAA TTC ATT AAT AAG TTT TCC TTC TTC TAG TTT AAT GAT TAA ATT CTC ACA AG

Epitope tags to be added to the 5' end of the gene	Tyr Phos WT <i>Eco</i> RI F	CAT CAG GAA TTC AAG TTG TTG TTT GTA TGT TTA GGC AAC ATT TGT
	Tyr Phos Mut <i>Eco</i> RI F	CAT CAG GAA TTC AAG TTG TTG TTT GTA AGT TTA GGC AAC ATT TGT
	Tyr Phos <i>Sal</i> I R	CCG CAG GTC GAC TTA ATT AAT AAG TTT TCC TTC TTC TAG TTT AAT GAT TAA ATT CTC

Table 3-6. Proteins identified by mass spectrometry in the ~37 kDa band from silver-stained gels of the pulldowns using the mutant LMW-PTP attached to Affi-Gel 15 beads incubated with amebic lysate. These are all the proteins that were identified in the ~37 kDa band. The table includes the rank of the protein (under #) in terms of the number unique peptides generated (the lower the number, the more unique peptides were found), the total number of assigned spectra, the number of unique peptides in those spectra (there can be more than one of the same peptide), the percent peptide coverage of the protein, whether or not it was found in both bands that were sequenced (the ~37 kDa and ~45 kDa bands) and if the protein is close to being the correct size to be present in the ~37 kDa band (plus or minus 5 kDa). Silver-stained gels from which the bands were cut are shown in Figures 3-24 and 3-25.

#	Identified protein(s)	GI number(s)	Molecular mass	Total number of assigned spectra	Number of unique peptides	Percent coverage of protein	Found in both 37 and 45 kDa bands?	Correct size (+/- 5 kDa)?
1	protein tyrosine phosphatase, putative	gi56470300	17 kDa	74	12	64%	yes	no
2	glyceraldehyde-3-phosphate dehydrogenase, putative	gi56465309	35 kDa	39	12	43%	yes	yes
3	actin	gi56473822	33 kDa	7	6	25%	yes	yes
4	actin or putative actin	gi56473961, gi56472634, gi56470711, gi56469321, gi56465111, gi56463962, gi113294, gi158916, gi158912, gi158910, gi158908	42 kDa	2	2	5.6%	yes	yes
5	hypothetical protein 328.t00002	gi56465624	34 kDa	16	12	36%	no	yes
7	actin-like protein, putative	gi56474849	45 kDa	1	1	3.0%	yes	no
8	40S ribosomal protein S4, putative	gi56471403, gi56468916, gi56465310	35 kDa	8	7	22%	yes	yes
11	ARP2/3 complex 34 kDa subunit, putative	gi56471818, gi56467208	34 kDa	8	7	27%	no	yes
14	hypothetical protein 37.t00026	gi56472161	143 kDa	1	1	0.5%	no	no

Table 3-6 is continued on page 263.

Table 3-6, continued from page 262.

#	Identified protein(s)	GI number(s)	Molecular mass	Total number of assigned spectra	Number of unique peptides	Percent coverage of protein	Found in both 37 and 45 kDa bands?	Correct size (+/- 5 kDa)?
16	protein disulfide isomerase, putative	gi56467297	42 kDa	2	2	7.1%	no	yes
18	ribose-phosphate pyrophosphokinase, putative	gi56473321	37 kDa	6	4	13%	no	yes
19	glyeraldehyde-3-phosphate dehydrogenase, putative	gi56466989	35 kDa	1	1	9.4%	no	yes
20	protein tyrosine phosphatase, putative	gi56473557	18 kDa	3	1	12%	yes	no
21	NAD ⁺ dependent alcohol dehydrogenase-related	gi56464391	22 kDa	2	2	11%	no	no
24	conserved hypothetical protein	gi56472450	330 kDa	1	1	0.65%	no	no
30	UDP-glucose 4-epimerase, putative	gi56466956	38 kDa	2	2	7.6%	no	yes
32	cysteine synthase A; cysteine synthase type II	gi56464392, gi2366771	37 kDa	6	5	20%	no	yes
34	alcohol dehydrogenase 3, putative	gi56467041	42 kDa	3	2	7.6%	no	yes
36	hypothetical protein 60.t00038	gi56471239	23 kDa	1	1	10%	yes	no

Table 3-6 is continued on page 264.

Table 3-6, continued from page 263.

#	Identified protein	GI number	Molecular mass	Number of assigned spectra	Number of unique peptides	Percent coverage of protein	Found in both 37 and 45 kDa bands?	Correct size (+/- 5 kDa)?
37	galactose-inhibitable lectin 35 kDa subunit precursor	gi56473325	34 kDa	2	2	8.3%	no	yes
38	60S acidic ribosomal protein P0, putative	gi56466556	34 kDa	3	3	14%	no	yes
45	hypothetical protein 263.t00003	gi56466411	43 kDa	2	2	8.6%	no	no
46	WD repeat protein	gi56471140	38 kDa	3	3	10%	no	yes
48	conserved hypothetical protein	gi56468642	123 kDa	1	1	2.6%	no	no
49	LIM domain protein	gi56474276	33 kDa	2	1	4%	no	yes
58	LIM domain protein	gi56467468	36 kDa	2	2	6.8%	no	yes
63	hypothetical protein 33.t00049	gi56472388	33 kDa	1	1	3.2%	no	yes
64	arsenite-translocating ATPase, putative	gi56465686	36 kDa	1	1	2.8%	no	yes
66	hypothetical protein 312.t00002	gi56465791	290 kDa	1	1	0.76%	no	no
67	hypothetical protein 10.t00068	gi56473774	89 kDa	1	1	2%	no	no

Table 3-6 is continued on page 265.

Table 3-6, continued from page 264.

#	Identified protein	GI Number	Molecular mass	Total number of assigned spectra	Number of unique peptides	Percent coverage of protein	Found in both 37 and 45 kDa bands?	Correct size (+/- 5 kDa)?
68	BspA-like leucine rich repeat protein, putative	gi56469895	59 kDa	1	1	5.5%	no	no
69	hypothetical protein 19.t00028	gi56473145	126 kDa	1	1	1.6%	no	no
70	protein phosphatase, putative	gi56463741	14 kDa	1	1	10%	no	no
71	RNA methyltransferase, putative	gi56467236	55 kDa	1	1	3.8%	no	no
76	eukaryotic translation initiation factor 3 subunit 2, putative	gi56471147	36 kDa	1	1	3.7%	no	yes
77	heat shock transcription factor, putative	gi56468077	32 kDa	1	1	3.4%	no	yes
78	BspA-related protein, putative	gi56465923	125 kDa	1	1	1.7%	no	no
79	protein kinase, putative	gi56469007	105 kDa	1	1	2.1%	no	no

Table 3-6 is continued on page 266.

Table 3-6, continued from page 265.

#	Identified protein	GI Number	Molecular mass	Total number of assigned spectra	Number of unique peptides	Percent coverage of protein	Found in both 37 and 45 kDa bands?	Correct size (+/- 5 kDa)?
80	hypothetical protein 103.t00024	gi56469733	59 kDa	1	1	2.2%	no	no
81	hypothetical protein 99.t00016	gi56469849	51 kDa	1	1	4.8%	no	no
82	conserved hypothetical protein	gi56470043	42 kDa	1	1	3.1%	no	yes
83	hypothetical protein 9.t00033	gi56473872	59 kDa	1	1	2.2%	no	no
84	hypothetical protein 2.t00094	gi56474656	154 kDa	1	1	1.5%	no	no

Table 3-7. Proteins identified by mass spectrometry in the ~45 kDa band from silver-stained gels of the pulldowns using the mutant LMW-PTP attached to Affi-Gel 15 beads incubated with amebic lysate. These are all the proteins that were identified in the ~45 kDa band. The table includes the rank of the protein (under #) in terms of the number of unique peptides generated (the lower the number, the more unique peptides were found) total number of assigned spectra, the number of unique peptides in those spectra (there can be more than one of the same peptide), the percent peptide coverage of the protein, whether or not it was found in both bands that were sequenced (the ~37 kDa and ~45 kDa bands) and if the protein is actually close to being the correct size to be in the ~45 kDa band (plus or minus 5 kDa). Silver-stained gels from which the bands were cut are shown in Figures 3-24 and 3-25.

#	Identified protein(s)	GI number(s)	Molecular mass	Total number of assigned spectra	Number of unique peptides	Percent coverage of protein	Found in both 37 and 45 kDa bands?	Correct size (+/- 5 kDa)?
1	protein tyrosine phosphatase, putative	gi56470300	17 kDa	59	12	64%	yes	no
2	glyceraldehyde-3-phosphate dehydrogenase, putative	gi56465309	35 kDa	2	2	9.3%	yes	no
3	actin	gi56473822	33 kDa	45	16	68%	yes	no
4	actin; actin, putative	gi56473961, gi56472634, gi56470711, gi56469321, gi56465111, gi56463962, gi113294, gi158916, gi158912, gi158910, gi158908	42 kDa	27	6	15%	yes	yes
6	type A flavoprotein	gi56474182	46 kDa	24	15	44%	no	yes
7	actin-like protein, putative	gi56474849	45 kDa	20	13	39%	yes	yes
8	40S ribosomal protein S4, putative	gi56471403, gi56468916, gi56465310	35 kDa	5	3	12%	yes	no
9	protein kinase, putative	gi56466736	40 kDa	12	9	31%	no	yes

Table 3-7 is continued on page 269.

Table 3-7, continued from page 268.

#	Identified protein(s)	Accession number(s)	Molecular mass	Total number of assigned spectra	Number of unique peptides	Percent coverage of protein	Found in both 37 and 45 kDa bands?	Correct size (+/- 5 kDa)?
10	conserved hypothetical protein	gi56465028, gi56465031	43 kDa	11	7	21%	no	yes
12	hypothetical protein 503.t00001	gi56464309	43 kDa	8	7	19%	no	yes
13	translation elongation factor EF-1 alpha, putative	gi56464491, gi56464963	48 kDa	6	5	13%	no	yes
15	Sec61 alpha subunit, putative	gi56464518	49 kDa	3	3	6.7%	no	yes
17	clathrin-adaptor medium chain, putative	gi56471326	48 kDa	3	3	9.2%	no	yes
20	protein tyrosine phosphatase, putative	gi56473557	18 kDa	4	1	12%	yes	no
22	clathrin coat assembly protein, putative	gi56465959	46 kDa	2	2	5.2%	no	yes
23	conserved hypothetical protein	gi56463451	43 kDa	1	1	2.8%	no	yes
25	calcium-binding protein, putative, URE3-sequence specific DNA binding protein	gi56463918, gi10505244	22 kDa	2	2	15%	no	no
26	alcohol dehydrogenase 3, putative	gi56466329	28 kDa	2	2	9.6%	no	no
27	galactokinase, putative	gi56465403	43 kDa	4	3	9%	no	yes

Table 3-7 is continued on page 270.

Table 3-7, continued from page 269.

#	Identified protein(s)	GI number(s)	Molecular Mass	Total number of assigned spectra	Number of unique peptides	Percent coverage of protein	Found in both 37 and 45 kDa bands?	Correct size (+/- 5 kDa)?
28	cysteine desulfurase, putative; iron-sulfur cluster assembly protein; NifS/IscS-related protein	gi56472379, gi33415910, gi46367711	43 kDa	4	3	7.2%	no	yes
29	peptidase, putative	gi56466416	42 kDa	4	4	12%	no	yes
31	26s proteasome subunit P45 family protein, putative	gi56471787	44 kDa	1	1	3.1%	no	yes
33	SH3 domain protein	gi56474795, gi56463909	44 kDa	3	3	8.4%	no	yes
35	aldehyde-alcohol dehydrogenase 2, putative	gi56465626	97 kDa	1	1	1.3%	no	no
36	hypothetical protein 60.t00038	gi56471239	23 kDa	2	1	10%	yes	no
39	G protein alpha subunit, putative	gi56468509	36 kDa	1	1	3.7%	no	no
40	protein kinase, putative	gi56473305	43 kDa	2	2	5.9%	no	yes
41	protein phosphatase, putative	gi56473624	35 kDa	2	2	8.4%	no	no
42	SH3 domain protein	gi56471301	42 kDa	4	3	9.8%	no	yes
43	hypothetical protein 388.t00006	gi56465087	191 kDa	1	1	1.1%	no	no

Table 3-7 is continued on page 271.

Table 3-7, continued from page 270.

#	Identified protein(s)	Accession Number(s)	Molecular mass	Total number of assigned spectra	Number of unique peptides	Percent coverage of protein	Found in both 37 and 45 kDa bands?	Correct size (+/- 5 kDa)?
44	alcohol dehydrogenase 3, putative	gi56470756	39 kDa	3	3	15%	no	no
47	hypothetical protein 131.t00006	gi56468929	42 kDa	1	1	3.3%	no	yes
50	pyruvate: ferredoxin oxidoreductase	gi56469894	127 kDa	1	1	0.96%	no	no
51	oxysterol binding protein, putative	gi56472782	48 kDa	1	1	2.9%	no	yes
52	type A flavoprotein	gi56471042	47 kDa	2	1	2.7%	no	yes
53	translation elongation factor EF-1 alpha, putative	gi56468654	48 kDa	3	2	7.9%	no	yes
54	clathrin-adaptor medium chain, putative	gi56464949	50 kDa	2	2	5.6%	no	yes
55	BspA-like leucine rich repeat protein, putative	gi56468030	76 kDa	1	1	3.1%	no	no
56	peptidyl-prolyl cis-trans isomerase, putative	gi56470350, gi56470326	45 kDa	1	1	2.8%	no	yes
57	hypothetical protein 180.t00020	gi56467858	42 kDa	1	1	2.9%	no	yes
59	conserved hypothetical protein	gi56468182	164 kDa	1	1	1.4%	no	no
60	peroxiredoxin	gi56463575	26 kDa	2	2	9.9%	no	no

Table 3-7 is continued on page 272.

Table 3-7, continued from page 271.

#	Identified protein(s)	GI Number	Molecular mass	Total number of assigned spectra	Number of unique peptides	Percent coverage of protein	Found in both 37 and 45 kDa bands?	Correct size (+/- 5 kDa)?
61	pleckstrin homology domain protein	gi56473241	47 kDa	1	1	3.1%	no	yes
62	proteasome regulatory subunit, putative 26S protease regulatory subunit 8, putative	gi56466175, gi56465447	45 kDa	1	1	3.5%	no	yes
65	hypothetical protein 199.t00003	gi56467435	40 kDa	1	1	3.8%	no	yes
69	hypothetical protein 19.t00028	gi56473145	126 kDa	1	1	1.6%	no	no
70	protein phosphatase, putative	gi56463741	14 kDa	0	0	0	no	no
72	SPRY domain protein	gi56473673	37 kDa	1	1	3.4%	no	no
73	RNA-binding protein, putative	gi56474367	37 kDa	1	1	3.9%	no	no
74	eukaryotic initiation factor 4A, putative	gi56474536	44 kDa	1	1	3.1%	no	yes
75	hypothetical protein 628.t00001	gi56463881	44 kDa	1	1	2.5%	no	yes

Table 3-8. Proteins identified by mass spectrometry selected for cloning and epitope-tagging, overexpression, and further characterization. Candidates from the ~37 kDa and ~45 kDa bands were selected based on whether they were close to the correct size for being in that position on the gel, by the number of unique peptides identified, and the percent coverage of the protein (Tables 3-6 and 3-7). Three were chosen from the ~37 kDa band and four from the ~45 kDa band. URE3-BP, which was identified in the ~45 kDa band, was also included as a candidate for further investigation, since we have antibodies against it, and it had two unique peptides and 15% protein coverage. This protein is 22 kDa, so by itself it is not the correct size, but would be if it were functioning as a dimer.

#	Identified protein	GenBank accession numbers	Molecular mass	Number of unique peptides		Percent coverage of protein		Gene size (bp)
				37 kDa band	45 kDa band	37 kDa band	45 kDa band	
5	hypothetical protein 328.t00002	XM_644144, XP_649236	34 kDa	12	0	36%	0%	915
6	type A flavoprotein	XM_651854, XP_656946	46 kDa	0	15	0	44%	1221
9	protein kinase, putative	XM_645070, XP_650162	40 kDa	0	9	0	31%	1065
10	conserved hypothetical protein gi56465028	XM_001914300, XP_001914335	43 kDa	0	7	0	21%	1041
11	ARP2/3 complex 34 kDa subunit, putative	DS571335, EAL49360	34 kDa	7	0	27%	0	882
12	hypothetical protein 503.t00001	XM_650927, XP_656019	43 kDa	0	7	0	19%	1080
18	ribose-phosphate pyrophosphokinase, putative	XM_651049, XP_656141	37 kDa	4	0	13%	0	999
25	calcium-binding protein, putative, URE3-sequence specific DNA binding protein	AF291721, AAG18423	22 kDa	0	2	0%	15%	663

Table 3-9. Oligo pairs used to amplify via PCR the genes for the proteins selected from the mass spectrometry data in Figure 3-8. HM1:IMSS strain amebic genomic DNA was used as

the PCR template for PCR, and restriction endonuclease recognition sites were added to the oligos such that PCR products could be digested and ligated into pBluescript II KS(+) and ultimately into pGIR209, the amebic expression vector. The oligo name indicates if either a 5' or 3' epitope tag is to be added, which restriction sites are included, and the protein name.

Sequences of the oligos are shown, and also if there are internal restriction sites within the gene.

The epitope tags, E, S, and KT3 (protein sequences are shown in Table 3-10) were made by annealing oligos together, and these tags were added at the 5' end or the 3' end of the cloned gene. For clones with an epitope tag added to the 3' end of the gene, the restriction sites used for cloning were a *SpeI* site immediately followed by a *BglII* site at the 5' end, and an *EcoRI* site at the 3' end (unless the cloned gene contained any of these sites internally; then an alternate site was used). 3' epitope tags themselves had *EcoRI* and *SalI* restriction sites. For clones with epitope tags added to the 5' end of the gene, the cloned genes had *EcoRI* and *SalI* restriction sites, unless the gene had an internal *EcoRI* or *SalI* site present, then an alternate site was used. The added 5' tags themselves had a *SpeI* site immediately followed by a *BglII* site at the 5' end and an *EcoRI* site at the 3' end. After being cloned into pBluescript II KS(+) with epitope tags added, clones were sequence-verified, and were digested with *BglII* (or *BamHI* if there was an internal *BglII* site) and *SalI*, and ligated into the amebic expression vector pGIR209 (Figure 3-26).

Oligo name	Oligo sequence	Internal sites?
5' tag <i>EcoRV</i> 328.t00002 F	CAGCAT GATATC AACGCAT TTATTC TCCTCTT	<i>Bgl</i> III, <i>EcoRI</i>
5' tag <i>Sal</i> I 328.t00002 R	CATCAT GTCGAC TTA TT TCTTA GCGAT CTTT TTG	
5' tag <i>EcoRI</i> type A flavoprot F	CAGCAG GAATTC AAAGC ATTGG AAGTA GTAAA AG	no
5' tag <i>Sal</i> I type A flavoprot R	CATAAT GTCGAC TTAA GCTT TAAGG GCCTC	
5' tag <i>EcoRI</i> Put prot kinase F	CATCAG GAATTC ACAACAC CAACT AACT TCA	no
5' tag <i>Sal</i> I Put prot kinase R	CATCAG GTCGAC TTATT TTCTA ATAAC CATTG TATCAT	
5' tag <i>EcoRI</i> ConsHypothProt gi56465028 F	CAGCAG GAATTC GAAGAGT TAATA AACAA GAATAAC	no
5' tag <i>Sal</i> I ConsHypothProt gi56465028 R	CATCAT GTCGAC TCATTT CATTT CAATT ACTAT GATTC	
5' tag <i>EcoRV</i> Put ARP2/3 34kDa subu F	CATCAT GATATC GCACCA AACTTG CTTCTT	<i>Spe</i> I, <i>EcoRI</i>
5' tag <i>Sal</i> I Put ARP2/3 34kDa subu R	CATAAT GTCGAC TTA TTTTG AAAGT TCTTC CTG	
5' tag <i>EcoRI</i> HypothProt 503.t00001	CATAAT GAATTC AGCG GAATT GGAGA CGCAT	<i>Spe</i> I
5' tag <i>Sal</i> I HypothProt 503.t00001	CATAAT GTCGAC TTAAAG TGGAGA AACATA ATTGT ATGGG	
5' tag <i>EcoRI</i> ribphos ppkinase F	CATCAG GAATTC AGTGTA ACAAC TTCTT TATCC	<i>Bgl</i> III
5' tag <i>Sal</i> I ribphos ppkinase R	CATCAT GTCGAC TTAGT CATT ATTGT TATTT AAC ATGC	
5' tag <i>EcoRI</i> URE3-BP F	CATCAT GAATTC CAACC ACCTG TAGCT AAT	<i>Bgl</i> III
5' tag <i>Sal</i> I URE3-BP R	CATTAT GTCGAC TTATT CCAAG AGGGA AGTAA	

Table 3-9 is continued on page 277.

Table 3-9, continued from page 276.

Oligo name	Oligo sequence	Internal sites?
3' tag <i>SpeI</i> - <i>Bam</i> H1 328.t00002 F	CATCAT ACTAGT GGATCC ATGAA CGCAT TTATT CTCCT CTTC ATTG	<i>Bgl</i> II, <i>Eco</i> RI
3' tag <i>Pst</i> I 328.t00002 R	CATCAG CTGCAG TTTCTT AGCGAT CTTTT TGGCA GCA	
3' tag <i>SpeI</i> - <i>Bgl</i> II type A flavoprotein F	CGGCGG ACTAGT AGATCT ATG AAAGC ATTGG AAGTA GTAAA AGA CTTA	no
3' tag <i>Eco</i> RI type A flavoprotein R	CCGCAG GAATTC AGCTT TAAGG GCCTC AGCA AATTT	
3' tag <i>SpeI</i> - <i>Bgl</i> II Put prot kinase F	CAGCAG ACTAGT AGATCT ATGAC AACAC CAACT AAAC TCAA GTTTC	no
3' tag <i>Eco</i> RI Put prot kinase R	GCCGCGG GAATTC TT TTCTA ATAAC CATTG TATCA TCTCC ATCATCA	
3' tag <i>SpeI</i> - <i>Bgl</i> II cons hypoth protein gi56465028 F	CCGCCG ACTAGT AGATCT ATGGA AGAGT TAATA AACAA GAATA ACCAA TTG	no
3' tag <i>Eco</i> RI cons hypoth protein gi56465028 R	CGGCGGC GAATTC TTTCATTT CAATT ACTAT GATTC GTTTT GGTGT	
3' tag <i>Xba</i> I- <i>Bgl</i> II Put ARP2/3 complex 34 kDa subunit F	CAGCAG TCTAGA AGATCT ATGGC ACCAA ACTTG CTTCT TGAA	<i>Spe</i> I, <i>Eco</i> RI
3' tag <i>Pst</i> I Put ARP2/3 complex 34 kDa subunit R	CAGCAC CTGCAG CTTTT TGAA AGTTC TTCCT GAAGCAG	
3' tag <i>Xba</i> I- <i>Bgl</i> II hypoth prot 503.t00001 F	CAGCAT TCTAGA AGATCT ATGAGCG GAATT GGAGA CGCAT	<i>Spe</i> I
3' tag <i>Eco</i> RI hypoth prot 503.t00001 R	CCGCGG GAATTC AAGTG GAGAA ACATA ATTGT ATGGG AAGTAT	
3' tag <i>SpeI</i> - <i>Bam</i> H1 Put ribose-phosphate pyrophosphokinase F	CATCAT ACTAGT GGATCC ATGAG TGTA CAACT TCTTT ATCCGC	<i>Bgl</i> II
3' tag <i>Eco</i> RI Put ribose-phosphate pyrophosphokinase R	GCCGCCG GAATTC GTCATT ATTGT TATTT AAACA TGCTC TAATC TGTTT	
3' tag <i>SpeI</i> - <i>Bam</i> H1 URE3-BP clone F	CATCAT ACTAGT GGATCC ATGC AACC ACCTG TAGCT AATTT	<i>Bgl</i> II
3' tag <i>Eco</i> RI URE3-BP clone R	CCGCTG GAATTC TT CCAAG AGGGA AGTAA CAACGT	

Table 3-10. Epitope tags added to LMW-PTP and putative substrate protein genes. The epitope tags c-Myc, FLAG, V5, and 6xHis were added to the 5' end or 3' end of the cloned wild-type or mutant LMW-PTP. The epitope tags E, S, or KT3 were added to the 5' end or 3' end of the cloned putative LMW-PTP substrate protein genes. The epitope-tagged genes were initially cloned into pBluescript II KS(+) and subcloned into the amebic expression vector pGIR209 for overexpression (Figure 3-31).

Epitope Tag	Peptide Sequence
c-Myc	EQKLISEEDL
FLAG	DYKDDDDK
V5	GKPIPNPLLGLDST
6xHis	HHHHHH
E tag	GAPVPYDPLEPR
S tag	KETAAAKFERQHMS
KT3	KPPTPPPEPET

Literature Cited

1. Stanley SL Jr. (2003). Amoebiasis. *The Lancet* 361:1025-34.
2. Gilchrist CA, Petri WA Jr. (1999). Virulence factors of *Entamoeba histolytica*. *Curr Opin Microbiol* 2:433-437.
3. Katz U, Ankri S, Stolarsky T, Nuchamowitz Y, Mirelman D. (2002). *Entamoeba histolytica* expressing a dominant negative N-truncated light subunit of its gal-lectin are less virulent. *Mol Biol Cell* 13:4256-4265.
4. Pacheco J, Shibayama M, Campos R, Beck DL, Houpt E, Petri WA Jr, Tsutsumi V. (2004). In vitro and in vivo interaction of *Entamoeba histolytica* Gal/GalNAc lectin with various target cells: an immunocytochemical analysis. *Parasitol Internat* 53:35-47.
5. Ravdin JI, Guerrant RL. (1981). Role of adherence in cytopathic mechanisms of *Entamoeba histolytica*. *J Clin Invest* 68:1305-1313.
6. Haque R, Huston CD, Hughes M, Houpt E, Petri WA Jr. (2003). Amebiasis. *N Eng J Med* 348(16):1565-1573.
7. Ayeh-Kumi PF, Ali IM, Lockhart LA, Gilchrist CA, Petri WA Jr, Haque R. (2001). *Entamoeba histolytica*: genetic diversity of clinical isolates from Bangladesh as demonstrated by polymorphisms in the serine-rich gene. *Exp Parasitol* 99:80-88.
8. Blessmann J, Van LP, Nu PAT Thi HD, Muller-Myhsok B, Buss H, Tannich E. (2002). Epidemiology of amebiasis in a region of high incidence of amebic liver abscess in central Vietnam. *Amer J Trop Med Hyg* 66(5):578-583.
9. Diamond LS, Clark CG. (1993). A redescription of *Entamoeba histolytica* Schaudinn, 1903 (Emended Walker, 1911) separating it from *Entamoeba dispar* Brumpt, 1925. *J Euk Microbiol* 40(3):340-344.

10. World Health Organization (No authors listed). (1997). Amoebiasis. *Weekly Epidemiol Record* 72(14):97-100.
11. Kucik CJ, Martin GL, Sortor BV. (2004). Common intestinal parasites. *Amer Fam Physician* 69(5):1161-1168.
12. Petri WA Jr, Haque R, Mann BJ. (2002). The bittersweet interface of parasite and host: lectin-carbohydrate interactions during human invasion by the parasite *Entamoeba histolytica*. *Annu Rev Microbiol* 56:39-64.
13. Eichinger D. (2001a). A role for a galactose lectin and its ligands during encystment of *Entamoeba*. *J Eukaryot Microbiol* 48(1):17-21.
14. Lohia A. (2003). The cell cycle of *Entamoeba histolytica*. *Mol Cell Biochem* 253:217-222.
15. Eichinger D. (2001b). Encystation in parasitic protozoa. *Curr Opin Microbiol* 4:421-426.
16. Dodson JM, Lenkowski PW Jr., Eubanks AC, Jackson TFGH, Napodano J, Lyerly DM, Lockhart LA, Mann BJ, Petri WA Jr. (1999). Infection and immunity mediated by the carbohydrate recognition domain of the *Entamoeba histolytica* Gal/GalNAc lectin. *J Infect Dis* 179:460-6.
17. Seigneur M, Mounier J, Prevost MC, Guillén N. (2005). A lysine- and glutamic acid-rich protein, KERP1, from *Entamoeba histolytica* binds to human enterocytes. *Cell Microbiol* 7(4):569-579.
18. Houpt E, Barroso L, Lockhart L, Wright R, Cramer C, Lyerly D, Petri WA Jr. (2004). Prevention of intestinal amebiasis by vaccination with the *Entamoeba* Gal/GalNAc lectin. *Vaccine* 22:611-617.

19. Houpt ER, Glembocki DJ, Obrig TG, Moskaluk CA, Lockhart LA, Wright RL, Seaner RM, Keepers TR, Wilkins TD, Petri WA Jr. (2002). The mouse model of amebic colitis reveals mouse strain susceptibility to infection and exacerbation of disease by CD4⁺ T cells. *J Immunol* 169(8):4496-4503.
20. Niehaus MD, Moore SR, Patrick PD, Derr LL, Lorntz B, Lima AA, Guerrant RL. (2002). Early childhood diarrhea is associated with diminished cognitive function 4 to 7 years later in children in a northeast Brazilian shantytown. *Amer J Trop Med Hyg* 66(5):590-593.
21. Petri WA Jr, Mondal D, Peterson KM, Duggal P, Haque R. (2009). Association of malnutrition with amebiasis. *Nutr Rev* 67 Suppl 2:S207-S215.
22. Loftus B, Anderson I, Davies R, Alsmark UCM, Samuelson J, Amedeo P, Roncaglia P, Berriman M, Hirt RP, Mann BJ, Nozaki T, Suh B, Pop M, Duchene M, Ackers J, Tannich E, Leippe M, Hofer M, Bruchhaus I, Willhoeft U, Bhattacharya A, Chillingworth T, Churcher C, Hance Z, Harris B, Harris D, Jagels K, Moule S, Mungall K, Ormond D, Squares D, Whitehead S, Quail MA, Rabbinowitsch E, Norbertczak H, Price C, Wang Z, Guillén N, Gilchrist C, Stroup SE, Bhattacharya S, Lohia A, Foster PG, Sicheritz-Ponten T, Weber C, Singh U, Mukherjee C, El-Sayed NM, Petri WA Jr, Clark CG, Embley TM, Barrell B, Fraser CM, Hall N. (2005). The genome of the protist parasite *Entamoeba histolytica*. *Nature* 433:865-868.
23. Das S, Lohia A. (2002). Delinking of S phase and cytokinesis in the protozoan parasite *Entamoeba histolytica*. *Cell Microbiol* 4(1):55-60.
24. Gangopadhyay SS, Ray SS, Kennady K, Pande G, Lohia A. (1997). Heterogeneity of DNA content in axenically growing *Entamoeba histolytica* HM1:IMSS clone A. *Mol Biochem Parasitol* 90:9-20.

25. Willhoeft U, Tannich E. (1999). The electrophoretic karyotype of *Entamoeba histolytica*. *Mol Biochem Parasitol* 99:41-53.
26. Ghosh S, Frisardi M, Ramirez-Avila L, Descoteaux S, Sturm-Ramirez K, Newton-Sanchez OA, Santos-Preciado JI, Ganguly C, Lohia A, Reed S, Samuelson J. (2000). Molecular epidemiology of *Entamoeba* spp.: evidence of a bottleneck (demographic sweep) and transcontinental spread of diploid parasites. *J Clin Microbiol* 38(10):3815-3821.
27. Kaur G, Lohia A. (2004). Inhibition of gene expression with double strand RNA interference in *Entamoeba histolytica*. *Biochem Biophys Res Commun* 320:1118-1122.
28. Vayssié L, Vargas M, Weber C, Guillén N. (2004). Double-stranded RNA mediates homology-dependent gene silencing of γ -tubulin in the human parasite *Entamoeba histolytica*. *Mol Biochem Parasitol* 138:21-28.
29. Mai Z, Ghosh S, Frisardi M, Rosenthal B, Rogers R, Samuelson J. (1999). Hsp60 is targeted to a cryptic mitochondrion-derived organelle ("crypton") in the microaerophilic protozoan parasite *Entamoeba histolytica*. *Mol Cell Biol* 19:2198-2205.
30. Tovar J, Fischer A, Clark CG. (1999). The mitosome, a novel organelle related to mitochondria in the amitochondriate parasite *Entamoeba histolytica*. *Mol Microbiol* 32:1013-1021.
31. Doolittle WF. (1998). You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet* 14(8):307-311.
32. Petri WA Jr, Schnaar RL. (1995). Purification and characterization of galactose and N-acetylgalactosamine-specific adhesin lectin of *Entamoeba histolytica*. *Methods Enzymol* 253:98-104.

33. Mitchell PD, Stern E, Tepper Y. (2008). Dysentery in the crusader kingdom of Jerusalem: an ELISA analysis of two medieval latrines in the city of Acre (Israel). *J Archaeol Science* 35:1849-1853.
34. Goncalves ML, da Silva VL, de Andrade CM, Reinhard K, da Rocha GC, Le Bailly M, Bouchet F, Ferreira LF, Araujo A. (2004). Amoebiasis distribution in the past: first steps using an immunoassay technique. *Trans R Soc Trop Med Hyg* 98(2):88-91.
35. Barwick RS, Uzicanin A, Lareau S, Malakmadze N, Imnadze P, Iosava M, Ninashvili N, Wilson M, Hightower AW, Johnston S, Bishop H, Petri WA Jr, Juraneck DD. (2002). Outbreak of amebiasis in Tbilisi, Republic of Georgia, 1998. *Amer J Trop Med Hyg* 67(6):623-631.
36. Markell EK. (1986). The 1933 Chicago outbreak of amebiasis. *West J Med* 144(6):750.
37. Select Committee. (1934). Amebiasis outbreak in Chicago: Report of a special committee. *JAMA* 102: 369.
38. (No authors listed). (1934). Amebic dysentery in Chicago. *Am J Public Health Nations Health* 24(7):756-758.
39. Long JD. (1909). Amebic dysentery. *Cal State J Med* 7(6):199-201.
40. Haque R, Duggal P, Ali IM, Hossain MB, Mondal D, Sack RB, Farr BM, Beaty T, Petri WA Jr. (2002). Innate and acquired resistance to amebiasis in Bangladeshi children. *J Infect Dis* 186:547-552.
41. Vinall LE, Hill AS, Pigny P, Pratt WS, Toribara N, Gum JR, Kim YS, Porchet N, Aubert J-P, Swallow DM. (1998). Variable number tandem repeat polymorphism of the mucin genes located in the complex on 11p15.5. *Hum Genet* 102:357-366.

42. Duggal P, Haque R, Roy S, Mondal D, Sack RB, Farr BM, Beaty TH, Petri WA Jr. (2004). Influence of human leukocyte antigen class II alleles on susceptibility to *Entamoeba histolytica* infection in Bangladeshi children. *J Infect Dis* 189(3):520-526.
43. Haque R, Ali IM, Sack RB, Farr BM, Ramakrishnan G, Petri WA Jr. (2001). Amebiasis and mucosal IgA antibody against the *Entamoeba histolytica* adherence lectin in Bangladeshi children. *J Infect Dis* 183(12):1787-1793.
44. Haque R, Mondal D, Duggal P, Kabir M, Roy S, Farr BM, Sack RB, Petri WA Jr. (2006). *Entamoeba histolytica* Infection in children and protection from subsequent amebiasis. *Infect Immun* 74(2):904-909.
45. Duggal P, Guo X, Haque R, Peterson KP, Ricklefs S, Mondal D, Alam F, Noor Z, Verkerke HP, Marie C, Leduc CA, Chua SC Jr, Myers MG Jr, Leibel RL, Houtp E, Gilchrist CA, Sher A, Porcella SF, Petri WA Jr. (2011). A mutation in the leptin receptor is associated with *Entamoeba histolytica* infection in children. *J Clin Invest* 121(3):1191-1198.
46. Braga LL, Ninomiya H, McCoy JJ, Eacker S, Wiedmer T, Pham C, Wood S, Sims PJ, Petri WA Jr. (1992). Inhibition of the complement membrane attack complex by the galactose-specific adhesin of *Entamoeba histolytica*. *J Clin Invest* 90:1131-1137.
47. Huston CD. (2004). Parasite and host contributions to the pathogenesis of amebic colitis. *Trends Parasitol* 20(1):23-26.
48. Ankri S, Padilla-Vaca F, Stolarsky T, Koole L, Katz U, Mirelman D. (1999a). Antisense inhibition of expression of the light subunit (35 kDa) of the Gal/GalNAc lectin complex inhibits *Entamoeba histolytica* virulence. *Mol Microbiol* 33(2):327-337.

49. Huston CD, Houpt ER, Mann BJ, Hahn CS, Petri WA Jr. (2000). Caspase 3-dependent killing of host cells by the parasite *Entamoeba histolytica*. *Cell Microbiol* 2(6):617-625.
50. MacFarlane RC, Shah PH, Singh U. (2005). Transcriptional profiling of *Entamoeba histolytica* trophozoites. *Int J Parasitol* 35:533-542.
51. Que X, Kim SH, Sajid M, Eckmann L, Dinarello CA, McKerrow JH, Reed SL. (2003). A surface amebic cysteine protease inactivates interleukin-18. *Infect Immun* 71:1274-1280.
52. Chadee K, Petri WA Jr, Innes DJ, Ravdin JI. (1987). Rat and human colonic mucins bind to and inhibit adherence lectin of *Entamoeba histolytica*. *J Clin Invest* 80:1245-1254.
53. Moncada D, Yu Y, Keller K, Chadee K. (2000). *Entamoeba histolytica* cysteine proteinases degrade human colonic mucin and alter its function. *Arch Med Res* 31:S224-S225.
54. McCoy JJ, Mann BJ, Vedvick TS, Petri WA Jr. (1993). Sequence analysis of genes encoding the light subunit of the *Entamoeba histolytica* galactose-specific adhesin. *Mol Biochem Parasitol* 61:325-328.
55. Springer TA, Lasky LA. (1991). Sticky sugars for selectins. *Nature* 349:196-197.
56. Cheng XJ, Hughes MA, Huston CD, Loftus B, Gilchrist CA, Lockhart LA, Ghosh S, Miller-Sims V, Mann BJ, Petri WA Jr, Tachibana H. (2001). Intermediate subunit of the Gal/GalNAc lectin of *Entamoeba histolytica* is a member of a gene family containing multiple CXXC sequence motifs. *Infect Immun* 69(9):5892-5898.
57. McCoy JJ, Mann BJ, Petri WA Jr. (1994). Adherence and cytotoxicity of *Entamoeba histolytica* or how lectins let parasites stick around. *Infect Immun* 62(8):3045-3050.
58. Ragland BD, Ashley LS, Vaux DL, Petri WA Jr. (1994). *Entamoeba histolytica*: target cells killed by trophozoites undergo DNA fragmentation which is not blocked by Bcl-2. *Exper Parasitol* 79:460-467.

59. Guerrant RL, Brush J, Ravdin JI, Sullivan JA, Mandell GL. (1981). Interaction between *Entamoeba histolytica* and human polymorphonuclear neutrophils. *J Infect Dis* 143:83-93.
60. Leroy A, de Bruyne G, Mareel M, Nokkaew C, Bailey G, Nelis H. (1995). Contact-dependent transfer of the galactose-specific lectin of *Entamoeba histolytica* to the lateral surface of enterocytes in culture. *Infect Immun* 63(11):4253-4260.
61. Tannich E, Ebert F, Horstmann RD. (1991). Primary structure of the 170-kDa surface lectin of pathogenic *Entamoeba histolytica*. *Proc Natl Acad Sci USA* 88:1849-1853.
62. Vines RR, Purdy JE, Ragland BD, Samuelson J, Mann BJ, Petri WA Jr. (1995). Stable episomal transfection of *Entamoeba histolytica*. *Mol Biochem Parasitol* 71:265-267.
63. Pillai DR, Wan PS, Yau YC, Ravdin JI, Kain KC. (1999). The cysteine-rich region of the *Entamoeba histolytica* adherence lectin (170-kilodalton subunit) is sufficient for high-affinity Gal/GalNAc-specific binding in vitro. *Infect Immun* 67:3836-3841.
64. Ramakrishnan G, Ragland BD, Purdy JE, Mann BJ. (1996). Physical mapping and expression of gene families encoding the N-acetyl-D-galactosamine adherence lectin of *Entamoeba histolytica*. *Mol Microbiol* 19:91-100.
65. Ramakrishnan G, Lee S, Mann BJ, Petri WA Jr. (2000). *Entamoeba histolytica*: deletion of the GPI anchor signal sequence on the Gal/GalNAc lectin light subunit prevents its assembly into the lectin heterodimer. *Exp Parasitol* 96:57-60.
66. Cheng XJ, Tsukamoto H, Kaneda Y, Tachibana H. (1998). Identification of the 150-kDa surface antigen of *Entamoeba histolytica* as a galactose- and N-acetyl-D-galactosamine inhibitable lectin. *Parasitol Res* 84:632-639.
67. Beck DL, Boettner DR, Dragulev B, Ready K, Nozaki T, Petri WA Jr. (2005). Identification and gene expression analysis of a large family of transmembrane kinases related to the Gal/GalNAc lectin in *Entamoeba histolytica*. *Eukaryot Cell* 4(4):722-732.

68. Gilchrist CA, Holm CF, Hughes MA, Schaenman JM, Mann BJ, Petri WA Jr. (2001). Identification and characterization of an *Entamoeba histolytica* upstream regulatory element 3 sequence-specific DNA-binding protein containing EF-hand motifs. *J Biol Chem* 276(15):11838-11843.
69. Gilchrist CA, Baba DJ, Zhang Y, Crasta O, Evans C, Caler E, Sobral BW, Bousquet CB, Leo M, Hochreiter A, Connell SK, Mann BJ, Petri WA. (2008). Targets of the *Entamoeba histolytica* transcription factor URE3-BP. *PLoS Negl Trop Dis* 2(8):e282.
70. Purdy JE, Pho LT, Mann BJ, Petri WA Jr. (1996). Upstream regulatory elements controlling expression of the *Entamoeba histolytica* lectin. *Mol Biochem Parasitol* 78(1-2):91-103.
71. Petri WA Jr, Singh U. (1999). Diagnosis and management of amebiasis. *Clin Infect Dis* 29(5):1117-1125.
72. Linford AS, Moreno H, Good KR, Zhang H, Singh U, Petri WA Jr. (2009). Short hairpin RNA-mediated knockdown of protein expression in *Entamoeba histolytica*. *BMC Microbiol* 9:38.
73. Boettner DR, Huston CD, Linford AS, Buss SN, Houpt E, Sherman NE, Petri WA Jr. (2008). *Entamoeba histolytica* phagocytosis of human erythrocytes involves PATMK, a member of the transmembrane kinase family. *PLoS Pathogens* 4(1):e8.
74. Moreno H, Linford AS, Gilchrist CA, Petri WA Jr. (2010). Phospholipid-binding protein EhC2A mediates calcium-dependent translocation of transcription factor URE3-BP to the plasma membrane of *Entamoeba histolytica*. *Eukaryot Cell* 9(5):695-704.
75. Gilchrist CA, Mann BJ, Petri WA Jr. (1998). Control of ferredoxin and Gal/GalNAc lectin gene expression in *Entamoeba histolytica* by a cis-acting DNA sequence. *Infect Immun* 66:2383-2386.

76. Gilchrist CA, Leo M, Line CG, Mann BJ, Petri WA Jr. (2003). Calcium modulates promoter occupancy by the *Entamoeba histolytica* Ca²⁺-binding transcription factor URE3-BP. *J Biol Chem* 278:4646-4653.
77. Gilchrist CA, Moore ES, Zhang Y, Bousquet CB, Lannigan JA, Mann BJ, Petri WA Jr. (2010). Regulation of virulence of *Entamoeba histolytica* by the URE3-BP transcription factor. *mBio* 1(1):e00057-10.
78. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391(6669):806-811.
79. Ketting RF, Haverkamp THA, van Luenen HGAM, Plasterk RHA. (1999). *mut-7* of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* 99(2):133-141.
80. Keene KM, Foy BD, Sanchez-Vargas I, Beaty BJ, Blair CD, Olson KE. (2004). RNA interference acts as a natural antiviral response to O'nyong-nyong virus (*Alphavirus*; *Togaviridae*) infection of *Anopheles gambiae*. *Proc Natl Acad Sci U S A.* 101(49): 17240–17245.
81. Downward J. (2004). RNA interference. *British Medical Journal* 328:1245-1248.
82. Hannon GJ. (2002). RNA interference. *Nature* 418(6894):244-251.
83. Novina CD, Sharp PA. (2004). The RNAi revolution. *Nature* 430(6996):161-164.
84. Plasterk RHA. (2002). RNA silencing: the genome's immune system. *Science* 296:1263-1265.
85. Waterhouse PM, Wang MB, Lough T. (2001). Gene silencing as an adaptive defence against viruses. *Nature* 411(6839):834-842.
86. Baulcombe D. (2004). RNA silencing in plants. *Nature* 431(7006):356-363.

87. Lau NC, Lim LP, Weinstein EG, Bartel DP. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294(5543):858-862.
88. Napoli C, Lemieux C, Jorgenson R. (1990). Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. *Plant Cell* 2:279-289.
89. Zamore PD. (2002). Ancient pathways programmed by small RNAs. *Science* 296(5571):1265-1269.
90. Hamilton AJ, Baulcombe DC. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 286:950-952.
91. Zamore PD, Tuschl T, Sharp PA, Bartel DP. (2000) RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101:25-33.
92. Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T. (2001a). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411:494-498.
93. Elbashir SM, Lendeckel W, Tuschl T. (2001b). RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev* 15:188-200.
94. Parrish S, Fleenor J, Xu SQ, Mello C, Fire A. (2000). Functional anatomy of a dsRNA trigger: differential requirement for the two trigger strands in RNA interference. *Mol Cell* 6:1077-1087.
95. Yang D, Lu H, Erickson JW. (2000). Evidence that processed small dsRNAs may mediate sequence-specific mRNA degradation during RNAi in *Drosophila* embryos. *Curr Biol* 10(19):1191-1200.

96. Bernstein E, Caudy AA, Hammond SM, Hannon GJ. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409, 363-366.
97. Dykxhoorn DM, Lieberman J. (2005). The silent revolution: RNA interference as basic biology, research tool, and therapeutic. *Annu Rev Med* 56:401-423.
98. Dykxhoorn DM, Novina CD, Sharp PA. (2003). Killing the messenger: short RNAs that silence gene expression. *Nature Reviews Molecular Cell Biology* 4:457-467.
99. Liu Q, Rand TA, Kalidas S, Du F, Kim HE, Smith DP, Wang X. (2003). R2D2, a bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. *Science* 301(5641):1921-1925.
100. Tabara H, Yigit E, Siomi H, Mello CC. (2002). The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in *C. elegans*. *Cell* 109(7):861-871.
101. Pham JW, Pellino JL, Lee YS, Carthew RW, Sontheimer EJ. (2004). A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in *Drosophila*. *Cell* 117(1):83-94.
102. Tomari Y, Du T, Haley B, Schwarz DS, Bennett R, Cook HA, Koppetsch BS, Theurkauf WE, Zamore PD. (2004). RISC assembly defects in the *Drosophila* RNAi mutant *armitage*. *Cell* 116(6):831-841.
103. Hammond SM, Boettcher S, Caudy AA, Kobayashi R, Hannon GJ. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* 293(5532):1146 – 1150.
104. Song JJ, Smith SK, Hannon GJ, Joshua-Tor L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* 305(5689):1434-1437.
105. Sioud M. (2007). RNA interference and innate immunity. *Adv Drug Deliv Rev* 59(2-3):153-163.

106. Miyoshi M, Tsukumo H, Nagami T, Siomi H, Siomi MC. (2005). Slicer function of *Drosophila* Argonautes and its involvement in RISC formation. *Genes Dev* 19:2837-2848.
107. Agrawal N, Dasaradhi PVN, Mohammed A, Malhotra P, Bhatnagar RK, Mukherjee SK. (2003). RNA interference: biology, mechanism, and applications. *Microbiol Mol Biol Rev* 67(4): 657–685.
108. Nykänen A, Haley B, and Zamore PD. (2001). ATP Requirements and small interfering RNA structure in the RNA interference pathway. *Cell* 107(3):309-321.
109. Okamura K, Ishizuka A, Siomi H, Siomi MC. (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev* 18(14):1655-66.
110. Aravin AA, Klenov MS, Vagin VV, Bantignies F, Cavalli G, Gvozdev VA. (2004). Dissection of a natural RNA silencing process in the *Drosophila melanogaster* germ line. *Mol Cell Biol* 24(15):6742-6750.
111. Sen GC. (2001). Viruses and interferons. *Annu Rev Microbiol.* 55(1):255-281.
112. Ui-Tei K, Zenno S, Miyata Y, Saigo K. (2000). Sensitive assay of RNA interference in *Drosophila* and Chinese hamster CHO-K1 cultured cells using firefly luciferase gene as target. *FEBS Letters* 479(3):79-82.
113. Williams BRG. (2005). Dicing with siRNA. *Nat Biotechnol* 23:181-182.
114. Walton SP, Wu M, Gredell JA, Chan C. (2010). Designing highly active siRNAs for therapeutic applications. *FEBS J* 277(23):4806-4813.
115. Dykxhoorn DM, Chowdhury D, Lieberman J. (2008). RNA interference and cancer: endogenous pathways and therapeutic approaches. *Adv Exp Med Biol* 615:299-329.
116. Fabian MR, Sonenberg N, Filipowicz W. (2010). Regulation of mRNA translation and stability by microRNAs. *Annu Rev Biochem* 79:351-379.

117. Du T, Zamore PD. (2005). microPrimer: the biogenesis and function of microRNA. *Development* 132(21):4645–4652.
118. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN. (2004). MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* 23(20):4051-4060.
119. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. (2001). Identification of novel genes coding for small expressed RNAs. *Science* 294:853–858.
120. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425(6956):415-419.
121. Lee Y, Jeon K, Lee JT, Kim S, Kim VN. (2002). MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J* 21(17):4663-4670.
122. Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ. (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature* 432(7014):231–235.
123. Han J, Lee Y, Yeom KH, Kim YK, Jin H, KimVN. (2004). The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* 18(24):3016-3027.
124. Bohnsack MT, Czaplinski K, Gorlich D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* 10:185-191.
125. Yi R, Qin Y, Macara IG, Cullen BR. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* 17:3011–3016.
126. Forstemann K, Tomari Y, Du T, Vagin VV, Denli AM, Bratu DP, Klattenhoff C, Theurkauf WE, Zamore PD. (2005). Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol* 3(7):e236.

127. Chendrimada TP, Gregory RI, Kumaraswamy E, Norman J, Cooch N, Nishikura K, Shiekhattar R. (2005). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* 436(7051):740-744.
128. Doench JG, Petersen CP, Sharp PA. (2003). siRNAs can function as miRNAs. *Genes Dev* 17:438-442.
129. Saxena S, Jonsson ZO, Dutta A. (2003). Small RNAs with imperfect match to endogenous mRNA repress translation. Implications for off-target activity of small inhibitory RNA in mammalian cells. *J Biol Chem* 278:44312-44319.
130. Easow G, Teleman AA, Cohen SM. (2007). Isolation of microRNA targets by miRNP immunopurification. *RNA* 13:1198-1204.
131. Gu S, Jin L, Zhang F, Sarnow P, Kay MA. (2009). Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nat Struct Mol Biol* 16:144-150.
132. Pak J, Fire A. (2007). Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science* 315(5809):241-244.
133. Sijen T, Fleenor J, Simmer F, Thijssen KL, Parrish S, Timmons L, Plasterk RHA, Fire A. (2001). On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* 107:465-476.
134. Wassenegger M, Krczal G. (2006). Nomenclature and functions of RNA-directed RNA polymerases. *Trends Plant Sci* 11:142-151.
135. Cerutti H, Casas-Mollano JA. (2006). On the origin and functions of RNA-mediated silencing: from protists to man. *Curr Genet* (2006) 50:81-99.

136. Lipardi C, Wei Q, Paterson BM. (2001). RNAi as Random Degradative PCR: siRNA Primers convert mRNA into dsRNAs that are degraded to generate new siRNAs. *Cell* 107(3):297-307.
137. Aoki K, Moriguchi H, Yoshioka T, Okawa K, Tabara H. (2007). In vitro analyses of the production and activity of secondary small interfering RNAs in *C. elegans*. *EMBO J* 26(24):5007-5019.
138. Schiebel W, Haas B, Marinković S, Klanner A, Sängner HL. (1993). RNA-directed RNA polymerase from tomato leaves. II. Catalytic in vitro properties. *J Biol Chem* 268(16):11858-11867.
139. Schiebel W, Péliissier T, Riedel L, Thalmeir S, Schiebel R, Kempe D, Lottspeich F, Sängner HL, Wassenegger M. (1998). Isolation of an RNA-directed RNA polymerase-specific cDNA clone from tomato. *Plant Cell* 10:2087-2101.
140. Dalmay T, Hamilton A, Rudd S, Angell S, Baulcombe DC. (2000). An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* 101(5): 543-553.
141. Vaistij FE, Jones L, Baulcombe DC. (2002). Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. *Plant Cell* 14:857-867.
142. Cogoni C, Maciano G. (1999). Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature* 399:166-169.
143. Martens H, Novotny J, Oberstrass J, Steck TL, Postlethwait P, Nellen W. (2002). RNAi in *Dictyostelium*: the role of RNA-directed RNA polymerases and double-stranded RNase. *Mol Biol Cell*.13(2): 445-453.

144. De S. Pal D, Ghosh SK. (2006). *Entamoeba histolytica*: computational identification of putative microRNA candidates. *Exp Parasitol* 113(4):239-243.
145. Abed M, Ankri S. (2005). Molecular characterization of *Entamoeba histolytica* RNase III and AGO2, two RNA interference hallmark proteins. *Exp Parasitol* 110(3):265–269.
146. Kolev NG, Tschudi C, Ullu E. (2011). RNA interference in protozoan parasites: achievements and challenges. *Eukaryot Cell* 10(9):1156-1163.
147. Zhang H, Kolb FA, Jaskiewicz, Westhof LE, Filipowicz W. (2004). Single processing center models for human Dicer and bacterial RNase III. *Cell* 118:57-68.
148. Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR, Bartel DP. (2009). RNAi in budding yeast. *Science* 326(5952):544-550.
149. Zhang H, Ehrenkauf GM, Pompey JM, Hackney JA, Singh U. (2008). Small RNAs with 5'-polyphosphate termini associate with a Piwi-related protein and regulate gene expression in the single-celled eukaryote *Entamoeba histolytica*. *PLoS Pathog* 4(11):e1000219.
150. Petri WA Jr, Ramakrishnan G. (1999). Applying antisense technology to the study of *Entamoeba histolytica* pathogenesis. *Trends Microbiol* 7(12):471-473.
151. Ankri S, Bracha R, Padilla-Vaca F, Mirelman D. (1999b). Applying antisense technology to the study of *Entamoeba histolytica* pathogenesis: response. *Trends Microbiol* 7(12):473-474.
152. Bracha R, Nuchamowitz Y, Mirelman D. (2000). Inhibition of gene expression in *Entamoeba* by the transcription of antisense RNA: effect of 5' and 3' regulatory elements. *Mol Biochem Parasitol* 107(1):81-90.
153. Dastidar PG, Majumder S, Lohia A. (2007). Eh Klp5 is a divergent member of the kinesin 5 family that regulates genome content and microtubular assembly in *Entamoeba histolytica*. *Cell Microbiol* 9(2):316-328.

154. MacFarlane RC, Singh U. (2007). Identification of an *Entamoeba histolytica* serine-, threonine-, and isoleucine-rich protein with roles in adhesion and cytotoxicity. *Eukaryot Cell* 6(11):2139–2146.
155. Yu JH, DeRuiter SL, Turner DL. (2002). RNA interference by expression of short-interfering RNAs and hairpin RNAs in mammalian cells. *Proc Natl Acad Sci USA* 99(9):6047–6052.
156. Gondai T, Yamaguchi K, Miyano-Kurosaki, Habu Y, Takaku H. (2008). Short-hairpin RNAs synthesized by T7 phage polymerase do not induce interferon. *Nucleic Acids Research* 36(3):e1.
157. Kunkel GR, Pederson T. (1989). Transcription of a human U6 small nuclear RNA gene in vivo withstands deletion of intragenic sequences but not of an upstream TATATA box. *Nucleic Acids Res.* 17(18):7371–7379.
158. Das G, Henning D, Wright D, Reddy R. (1988). Upstream regulatory elements are necessary and sufficient for transcription of a U6 RNA gene by RNA polymerase III. *EMBO J.* 7(2): 503–512.
159. Noonberg SB, Scott GK, Garovoy MR, Benz CC, Hunt CA. (1994). In vivo generation of highly abundant sequence-specific oligonucleotides for antisense and triplex gene regulation. *Nucleic Acids Res* 22(14):2830-2836.
160. Tazi J, Forne T, Jeanteur P, Cathala G, Brunel C. (1993). Mammalian U6 small nuclear RNA undergoes 3' end modifications within the spliceosome. *Mol Cel Biol* 13(3):1641-50.
161. Booth BL Jr, Pugh BF. (1997). Identification and characterization of a nuclease specific for the 3' end of the U6 small nuclear RNA. *J Biol Chem* 272(2):984-991.
162. Brummelkamp TR, Bernards R, Agami R. (2002). A system for stable expression of short interfering RNAs in mammalian cells. *Science* 296:550-553.

163. Siolas D, Lerner C, Burchard J, Ge W, Linsley PS, Paddison PJ, Hannon GJ, Cleary MA. (2004). Synthetic shRNAs as potent RNAi triggers. *Nat Biotechnol* 23:227-231.
164. Li L, Lin X, Khvorova A, Fesik SW, Shen Y. (2007). Defining the optimal parameters for hairpin-based knockdown constructs. *RNA* 13:1765-1774.
165. Silva JM, Li MZ, Chang K, Ge W, Golding MC, Rickles RJ, Siolas D, Hu G, Paddison PJ, Schlabach MR, Sheth N, Bradshaw J, Burchard J, Kulkarni A, Cavet G, Sachidanandam R, McCombie WR, Cleary MA, Elledge SJ, Hannon GJ. (2005). Second-generation shRNA libraries covering the mouse and human genomes. *Nat Genet* 37(11): 1281-1288.
166. Zeng Y, Wagner EJ, Cullen BR. (2002). Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol Cell* 9:1327-1333.
167. Kim DH, Behlke MA, Rose SD, Chang MS, Choi S, Rossi JJ. (2005). Synthetic dsRNA Dicer substrates enhance RNAi potency and efficiency. *Nat Biotechnol* 23(2):222-226.
168. Petri WA Jr., Smith RD, Schlesinger PH, Murphy CF, Ravdin JI. (1987). Isolation of the galactose-binding lectin that mediates the in vitro adherence of *Entamoeba histolytica*. *J Clin Invest* 80:1238-1244.
169. Clark CG, Alsmark UCM, Tazreiter M, Saito-Nakano Y, Ali V, Marion S, Weber C, Mukherjee C, Bruchhaus I, Tannich E, Leippe M, Sicheritz-Ponten T, Foster PG, Samuelson J, Noël CJ, Hirt RP, Embley TM, Gilchrist CA, Mann BJ, Singh U, Ackers JP, Bhattacharya S, Bhattacharya A, Lohia A, Guillén N, Duchêne M, Nozaki T, Hall N. (2007). Structure and content of the *Entamoeba histolytica* genome. *Adv Parasitol* 65:51-190.
170. Tachibana H, Kobayashi S, Cheng XJ, Hiwatashi E. (1997). Differentiation of *Entamoeba histolytica* from *E. dispar* facilitated by monoclonal antibodies against a 150-kDa surface antigen. *Parasitol Res* 83(5):435-439.

171. Nash TE. (2002). Surface antigenic variation in *Giardia lamblia*. *Mol Microbiol* 45:585–590.
172. Okada M, Huston CD, Mann BJ, Petri WA Jr, Kita K, Nozaki T. (2005). Proteomic analysis of phagocytosis in the enteric protozoan parasite *Entamoeba histolytica*. *Eukaryot Cell* 4:827-831.
173. Carrión AM, Link WA, Ledo F, Mellström B, Naranjo JR. (1999). DREAM is a Ca²⁺-regulated transcriptional repressor. *Nature* 398:80-84.
174. Nalefski EA, Falke JJ. (1996). The C2 domain calcium-binding motif: structural and functional diversity. *Protein Sci* 5:2375-2390.
175. Diamond LS, Harlow DR, Cunnick CC. (1978). A new medium for the axenic cultivation of *Entamoeba histolytica* and other *Entamoeba*. *Trans R Soc of Trop Med Hyg* 72(4):431-432.
176. Huston CD, Boettner DR, Miller-Sims V, Petri WA Jr. (2003). Apoptotic killing and phagocytosis by the parasite *Entamoeba histolytica*. *Infect Immun* 71(2):964-972.
177. Ramakrishnan G, Vines RR, Mann BJ, Petri WA Jr. (1997). A tetracycline-inducible gene expression system in *Entamoeba histolytica*. *Mol Biochem Parasitol* 84(1):93-100.
178. Miranda R, Salgado LM, Sánchez-López R Alagón A, Lizardi PM. (1996). Identification and analysis of the u6 small nuclear RNA gene from *Entamoeba histolytica*. *Gene* 180(1-2):37-42.
179. Katz U, Bracha R, Nuchamowitz Y, Milstein O, Mirelman D. (2003). Comparison between constitutive and inducible plasmid vectors used for gene expression in *Entamoeba histolytica*. *Mol Biochem Parasitol* 128(2):229-233.
180. Gou D, Jin N, Liu L. (2003). Gene silencing in mammalian cells by PCR-based short hairpin RNA. *FEBS Letters* 548:113-118.

181. The Ambion/Applied Biosystems siRNA Target Finder
http://www.ambion.com/techlib/misc/siRNA_finder.html
182. *E. histolytica* Genome Project database at the J.Craig Venter Institute (JCVI, formerly TIGR) at <http://www.tigr.org/tdb/e2k1/eha1/> (**Note: SITE IS NO LONGER ACTIVE**).
183. GraphPad QuickCalcs P-value calculator
<http://www.graphpad.com/quickcalcs>
184. Guide to Performing Relative Quantitation of Gene Expression Using Real-Time Quantitative PCR, Applied Biosystems
http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_042380.pdf
185. Čikoš Š, Bukovská A, Koppel J. (2007). Relative quantification of mRNA: comparison of methods currently used for real-time PCR data analysis. *BMC Mol Biol* 8:113-126.
186. Livak KJ, Schmittgen TD. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* 25: 402-408.
187. Real-Time PCR: M. Teyfik Dorak, MD, PhD
<http://www.dorak.info/genetics/realtime.html>
188. Hamann L, Nickel R, Tannich E. (1995). Transfection and continuous expression of heterologous genes in the protozoan parasite *Entamoeba histolytica*. *Proc Natl Acad Sci USA* 92(19):8975-8979.
189. Hamann L, Buss H, Tannich E. (1997). Tetracycline-controlled gene expression in *Entamoeba histolytica*. *Mol Biochem Parasitol* 84(1):83-91.
190. Shao Y, Chan CY, Maliyekkel A, Lawrence CE, Roninson IB, Ding Y. (2007). Effect of target secondary structure on RNAi efficiency. *RNA* 13:1631-1640.

191. Gredell JA, Berger AK, Walton SP. (2008). Impact of target mRNA structure on siRNA silencing efficiency: a large-scale study. *Biotechnol Bioeng* 100:744-755.
192. Gilchrist CA, Houpt E, Trapaidze N, Fei Z, Crasta O, Asgharpour A, Evans C, Martino-Catt S, Baba DJ, Stroup S, Hamano S, Ehrenkaufner G, Okada M, Singh U, Nozaki T, Mann BJ, Petri WA Jr. (2006). Impact of intestinal colonization and invasion on the *Entamoeba histolytica* transcriptome. *Mol Biochem Parasitol* 147(2):163-176.
193. Miyagishi M, Sumimoto H, Miyoshi H, Kawakami Y, Taira K. (2004). Optimization of an siRNA-expression system with an improved hairpin and its significant suppressive effects in mammalian cells. *J Gene Med* 6(7):715-723.
194. Ohkawa J, Taira K. (2000). Control of the functional activity of an antisense RNA by a tetracycline-responsive derivative of the human U6 snRNA promoter. *Hum Gene Ther* 11:577-585.
195. Matsukura S, Jones PA, Takai D. (2003). Establishment of conditional vectors for hairpin siRNA knockdowns. *Nucl Acids Res* 31(15):e77.
196. Sahoo N, Bhattacharya S, Bhattacharya A. (2003). Blocking the expression of a calcium binding protein of the protozoan parasite *Entamoeba histolytica* by tetracycline regulatable antisense-RNA. *Mol Biochem Parasitol* 126(2):281-284.
197. Santi-Rocca J, Weber C, Guigon G, Sismeiro O, Coppée J-Y, Guillén N. (2008). The lysine- and glutamic acid-rich protein KERP1 plays a role in *Entamoeba histolytica* liver abscess pathogenesis. *Cell Microbiol* 10(1):202-217.
198. Cheng TL, Teng CF, Tsai WH, Yeh CW, Wu MP, Hsu HC, Hung CF, Chang WT. (2009). Multitarget therapy of malignant cancers by the head-to-tail tandem array multiple shRNAs expression system. *Cancer Gene Ther* 16(6):516-531.

199. Boudreau RL, Monteys AM, Davidson BL. (2008). Minimizing variables among hairpin-based RNAi vectors reveals the potency of shRNAs. *RNA* 14(9):1834-1844.
200. Shan ZX, Lin QX, Yang M, Deng CY, Kuang SJ, Zhou ZL, Xiao DZ, Liu XY, Lin SG, Yu XY. (2009a). A quick and efficient approach for gene silencing by using triple putative microRNA-based short hairpin RNAs. *Mol Cell Biochem* 323(1-2):81-89.
201. Shan Z, Lin Q, Deng C, Li X, Huang W, Tan H, Fu Y, Yang M, Yu XY. (2009b). An efficient method to enhance gene silencing by using precursor microRNA designed small hairpin RNAs. *Mol Biol Rep* 36(6):1483-1489.
202. Zhou H, Huang C, Xia XG. (2008). A tightly regulated Pol III promoter for synthesis of miRNA genes in tandem. *Biochim Biophys Acta* 1779(11):773-779.
203. Ramponi G, Stefani M. (1997). Structure and function of the low M_r phosphotyrosine phosphatases. *Biochim Biophys Acta* 1341:137-156.
204. Hunter T. (1995). Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. *Cell* 80(2):225-238.
205. Eckhart W, Hutchinson MA, Hunter T. (1979). An activity phosphorylating tyrosine in polyoma T antigen immunoprecipitates. *Cell* 18: 925-933.
206. Hunter T, Sefton BM. (1980). Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. *Proc Natl Acad Sci USA* 77(3): 1311–1315.
207. Witte ON, Dasgupta A, Baltimore D. (1980). Abelson murine leukaemia virus protein is phosphorylated in vitro to form phosphotyrosine. *Nature* 283(5750):826-831.
208. Hunter T. (1998). The Croonian Lecture 1997. The phosphorylation of proteins on tyrosine: its role in cell growth and disease. *Philos Trans R Soc Lond B Biol Sci.* 353(1368):583-605.

209. Charbonneau H, Tonks NK, Walsh KA, Fischer EH. (1988). The leukocyte common antigen (CD45): a putative receptor-linked protein tyrosine phosphatase. *Proc Natl Acad Sci USA* 85(19):7182-7186.
210. Charbonneau H, Tonks NK, Kumar S, Diltz CD, Harrylock M, Cool DE, Krebs EG, Fischer EH, Walsh KA. (1989). Human placenta protein-tyrosine-phosphatase: amino acid sequence and relationship to a family of receptor-like proteins. *Proc Natl Acad Sci USA* 86(14):5252-5256.
211. Souza ACS, Azoubel S, Queiroz KC, Peppelenbosch MP, Ferreira CV. (2009). From immune response to cancer: a spot on the low molecular weight protein tyrosine phosphatase. *Cell Mol Lif Sci* 66(7):1140-1153.
212. Samet JM, Tal TL. (2010). Toxicological disruption of signaling homeostasis: tyrosine phosphatases as targets. *Annu Rev Pharmacol Toxicol* 50:215-235.
213. Caselli A, Taddei ML, Manao G, Camici G, Ramponi G. (2001). Tyrosine-phosphorylated caveolin is a physiological substrate of the low M_r protein-tyrosine phosphatase. *J Biol Chem* 276(22):18849-18854.
214. Bishop JM. (1991). Molecular themes in oncogenesis. *Cell* 64(2):235-48.
215. Aleshin A, Finn RS. (2010). SRC: a century of science brought to the clinic. *Neoplasia* 12(8):599-607.
216. Martin GS. (2001). The hunting of the Src. *Nat Rev Mol Cell Biol* 2(6):467-475.
217. Mustelin T (2006). A brief introduction to the protein phosphatase families. In Protein phosphatase protocols, Humana Press Inc., Totowa, NJ, USA. p. 9-22.
218. Saha S, Bardelli A, Buckhaults P, Velculescu VE, Rago C, St. Croix B, Romans KE, Choti MA, Lengauer C, Kinzler KW, Vogelstein B. (2001). A phosphatase associated with metastasis of colorectal cancer. *Science* 294(5545):1343-1346.

219. Tabernero L, Aricescu AR, Jones EY, Szedlacsek SE. (2008). Protein tyrosine phosphatases: structure-function relationships. *FEBS Journal* 275:867-882.
220. Neel BG, Tonks NK. (1997). Protein tyrosine phosphatases in signal transduction. *Curr Opin Cell Biol* 9:193-204.
221. Fauman EB, Saper MA. (1996). Structure and function of the protein tyrosine phosphatases. *Trends Biochem Sci* 21:413-417.
222. Alonso A, Sasin J, Bottini N, Friedberg I, Friedberg I, Osterman A, Godzik A, Hunter T, Dixon J, Mustelin T. (2004). Protein tyrosine phosphatases in the human genome. *Cell* 117(6):699-711.
223. Zabell APR, Schroff AD Jr, Bain BE, Van Etten RL, Wiest O, Stauffacher CV. (2006). Crystal structure of the human B-form low molecular weight phosphotyrosyl phosphatase at 1.6-Å resolution. *J Biol Chem* 281(10):6520-6527.
224. Zhang ZY, Zhou B, Xie L. (2002). Modulation of protein kinase signaling by protein phosphatases and inhibitors. *Pharmacol Ther* 93:307-317.
225. Zhang ZY. (2003). Mechanistic studies on protein tyrosine phosphatases. *Prog Nucleic Acid Res Mol Biol* 73:171-220.
226. Zhang ZY, Wang Y, Dixon JE. (1994). Dissecting the catalytic mechanism of protein tyrosine phosphatases. *Proc Natl Acad Sci USA* 91:1624-1627.
227. Xing K, Raza A, Löfgren S, Fernando MR, Ho YS, Lou MF. (2007). Low molecular weight protein tyrosine phosphatase (LMW-PTP) and its possible physiological functions of redox signaling in the eye lens. *Biochim Biophys Acta* 1774:545-555.
228. Raugei G, Ramponi G, Chiarugi P. (2002). Low molecular weight protein tyrosine phosphatases: small but smart. *Cell Mol Life Sci* 941:949.

229. Wang F, Li W, Emmet MR, Hendrickson CL, Marshall AG, Zhang YL, Wu L, Zhang ZY. (1998). Conformational and dynamic changes of *Yersinia* protein tyrosine phosphatase induced by ligand binding and active site mutation and revealed by H/D exchange and electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Biochemistry* 37:15289-15299.
230. Wo YYP, Zhou MM, Stevis P, Davis JP, Zhang ZY, Van Etten, RL. (1992a). Cloning, expression, and catalytic mechanism of the low molecular weight phosphotyrosyl protein phosphatase from bovine heart. *Biochemistry* 31:1712-1721.
231. Guan KL, Dixon JE. (1991). Evidence for protein-tyrosine phosphatase catalysis proceeding via a cysteine-phosphate intermediate. *J Biol Chem* 266:17026-17030.
232. Madhurantakam C, Chavali VR, Das AK. (2008). Analyzing the catalytic mechanism of MPtpA: a low molecular weight protein tyrosine phosphatase from *Mycobacterium tuberculosis* through site-directed mutagenesis. *Proteins* 71(2):706-714.
233. Barford D. (1999). Structural studies of reversible protein phosphorylation and protein phosphatases. *Biochem Soc Trans* 27:751-766.
234. Groen A, Lemeer S, van der Wilk T, Overvoorde J, Heck AJ, Ostman A, Barford D, Slijper M, den Hertog G. (2005). Differential oxidation of protein-tyrosine phosphatases. *J Biol Chem* 280(11):10298-10304.
235. den Hertog J, Groen A, van der Wijk T. (2005). Redox regulation of protein-tyrosine phosphatases. *Arch Biochem Biophys* 434(1):11-15.
236. Denu JM, Tanner KG. (1998). Specific and reversible inactivation of protein tyrosine phosphatases by hydrogen peroxide: evidence for a sulfenic acid intermediate and implications for redox regulation. *Biochemistry* 37(16):5633-5642.

237. Salmeen A, Andersen JN, Myers MP, Meng TC, Hinks JA, Tonks NK, Barford D. (2003). Redox regulation of protein tyrosine phosphatase 1B involves a sulphenyl-amide intermediate. *Nature* 423(6941):769-773.
238. van Montfort RL, Congreve M, Tisi D, Carr R, Jhoti H. (2003). Oxidation state of the active-site cysteine in protein tyrosine phosphatase 1B. *Nature* 423(6941):773-777.
239. Sundaresan M, Yu ZX, Ferrans VJ, Irani K, Finkel T. (1995). Requirement for generation of H₂O₂ for platelet-derived growth factor signal transduction. *Science* 270(5234):296-299.
240. Lee SR, Yang KS, Kwon J, Lee C, Jeong W, Rhee SG. (2002). Reversible inactivation of the tumor suppressor PTEN by H₂O₂. *J Biol Chem* 277(23):20336-20342.
241. Savitsky PA, Finkel T. (2002). Redox regulation of Cdc25C. *J Biol Chem* 277:20535-20540.
242. Caselli A, Marzocchini R, Camici G, Manao G, Moneti G, Pieraccini G, Ramponi G. (1998). The inactivation mechanism of low molecular weight phosphotyrosine-protein phosphatase by H₂O₂. *J Biol Chem* 273:32554-32560.
243. Chiarugi P, Fiashi T, Taddei ML, Talini D, Giannoni E, Raugei G, Ramponi G. (2001). Two vicinal cysteines confer a peculiar redox regulation to low molecular weight protein tyrosine phosphatase in response to platelet-derived growth factor receptor stimulation. *J Biol Chem* 276(36):33478-33487.
244. Chiarugi P. (2001). The redox regulation of LMW-PTP during cell proliferation or growth inhibition. *IUBMB Life* 52(1-2):55-59.
245. Bae YS, Kang SW, Seo MS, Baines IC, Tekle E, Chock PB, Rhee SG. (1997). Epidermal growth factor (EGF)-induced generation of hydrogen peroxide. Role in EGF receptor-mediated tyrosine phosphorylation. *J Biol Chem* 272(1):217-221.

246. Caselli A, Chiarugi P, Camici G, Manao G, Ramponi G. (1995). In vivo inactivation of phosphotyrosine protein phosphatases by nitric oxide. *FEBS Letters* 374: 249-252.
247. Zhou G, Denu JM, Wu L, Dixon JE. (1994). The catalytic role of Cys¹²⁴ in the dual specificity phosphatase VHR. *J Biol Chem* 269(45):28084-28090.
248. Zhang ZY, Dixon JE. (1993). Active site labeling of the *Yersinia* protein tyrosine phosphatase: The determination of the pK_a of the active site cysteine and the function of the conserved histidine 402. *Biochemistry* 32:9340-9345.
249. Pot DA, Woodford TA, Remboutsika E, Haun RS, Dixon JE. (1991). Cloning, bacterial expression, purification, and characterization of the cytoplasmic domain of rat LAR, a receptor-like protein tyrosine phosphatase. *J Biol Chem* 266(29):19688-19696.
250. Cirri P, Chiarugi P, Camici G, Manao G, Raugei G, Cappugi G, Ramponi G. (1993). The role of Cys12, Cys17, and Arg18 in the catalytic mechanism of low *Mr* cytosolic phosphotyrosine protein phosphatase. *Eur J Biochem* 214:647-657.
251. Davis JP, Zhou MM, Van Etten RL. (1994). Kinetic and site-directed mutagenesis studies of the cysteine residues of bovine low molecular weight phosphotyrosyl protein phosphatase. *J Biol Chem* 269(12):8734-8740.
252. Chiarugi P, Marzocchini R, Raugei G, Pazzagli C, Berti A, Camici G, Manao G, Cappugi G, Ramponi G. (1992). Differential role of four cysteines on the activity of a low *Mr* phosphotyrosine protein phosphatase. *Febs Lett* 310(1):9-12.
253. Zhang YL, Yao ZJ, Sarmiento M, Wu L, Burke TR Jr, Zhang ZY. (2000). Thermodynamic study of ligand binding to protein-tyrosine phosphatase 1B and its substrate-trapping mutants. *J Biol Chem* 275:34205-34212.
254. Moorhead GBG, De Wever V, Templeton G, Kerk D. (2009). Evolution of protein phosphatases in plants and animals. *Biochem J* 417:401-409.

255. Kerk D, Bulgrien J, Smith DW, Barsam B, Veretnik S, Gribskov M. (2002). The complement of protein phosphatase catalytic subunits encoded in the genome of *Arabidopsis*. *Plant Physiol* 129(2):908-925.
256. Kerk D, Templeton G, Moorhead GB. (2008). Evolutionary radiation pattern of novel protein phosphatases revealed by analysis of protein data from the completely sequenced genomes of humans, green algae, and higher plants. *Plant Physiol* 146(2):351-367.
257. Miller DT, Read R, Rusconi J, Cagan RL. (2000). The *Drosophila primo* locus encodes two low-molecular-weight tyrosine phosphatases. *Gene* 243(1-2):1-9.
258. Caselli A, Taddei ML, Bini C, Paoli P, Camici G, Manao G, Cirri P, Ramponi G. (2007). Low molecular weight protein tyrosine phosphatase and caveolin-1: interaction and isoenzyme-dependent regulation. *Biochemistry* 46(21):6383-6392.
259. Bryson GL, Massa H, Trask BJ, Van Etten RL. (1995). Gene structure, sequence, and chromosomal localization of the human red cell-type low-molecular weight acid phosphotyrosyl phosphatase gene, ACP1. *Genomics* 30:133-140.
260. Tailor P, Gilman J, Williams S, Mustelin T. (1999). A novel isoform of the low molecular weight phosphotyrosine phosphatase, LMPTP-C, arising from alternative mRNA splicing. *Eur J Biochem* 262:277-282.
261. Modesti A, Marzocchini R, Raugei G, Chiti F, Sereni A, Magherini G, Ramponi G. (1998). Cloning, expression and characterisation of a new human low M_r phosphotyrosine protein phosphatase originating by alternative splicing. *FEBS Lett* 431(1):111-115.
262. Wo YYP, McCormack AL, Shabanowitz J, Hunt DF, Davis JP, Mitchell GL, Van Etten RL. (1992b). Sequencing, cloning, and expression of human red cell-type acid phosphatase, a cytoplasmic phosphotyrosyl protein phosphatase. *J Biol Chem* 267(15):10856-10865.

263. Hopkinson DA, Spencer N, Harris H. (1963). Red cell acid phosphatase variants: a new human polymorphism. *Nature* 199:969-971.
264. Hopkinson DA, Spencer N, Harris H. (1964). Genetical studies on human red cell acid phosphatase. *Am J Hum Gen* 16(1):141-154.
265. Bessey OA, Lowry OH, Brock MJ. (1946). A method for the rapid determination of alkaline phosphatase with five cubic millimeters of serum. *J Biol Chem* 164(1):321-329.
266. Heinrickson RL. (1969). Purification and characterization of a low molecular weight acid phosphatase from bovine liver. *J Biol Chem* 244(2):299-307.
267. Waheed A, Laidler PM, Wo YY, Van Etten RL. (1988). Purification and physicochemical characterization of a human placental acid phosphatase possessing phosphotyrosyl protein phosphatase activity. *Biochemistry* 27(12):4265-4273.
268. Okada M, Owada K, Nakagawa H. (1986). [Phosphotyrosine]protein phosphatase in rat brain: a major [phosphotyrosine]protein phosphatase is a 23 kDa protein distinct from acid phosphatase. *Biochem J* 239:155-162.
269. Ostanin K, Pokalsky C, Wang S, Van Etten RL. (1995). Cloning and characterization of a *Saccharomyces cerevisiae* gene encoding the low molecular weight protein-tyrosine phosphatase. *J Biol Chem* 270(31):18491-18499.
270. Wang S, Taberero L, Zhang M, Harms E, Van Etten RL, Stauffacher CV. (2000). Crystal structures of a low-molecular weight protein tyrosine phosphatase from *Saccharomyces cerevisiae* and its complex with the substrate *p*-nitrophenyl phosphate. *Biochemistry* 39:1903-1914.
271. Zhang M, Stauffacher CV, Lin D, Van Etten RL. (1998). Crystal structure of a human low molecular weight phosphotyrosyl phosphatase: implications for substrate specificity. *J Biol Chem* 273(34):21714-21720.

272. Zhang M, Van Etten RL, Stauffacher CV. (1994). Crystal structure of bovine heart phosphotyrosyl phosphatase at 2.2 Å resolution. *Biochemistry* 33(37):11097-11105.
273. Su XD, Taddei N, Stefani M, Ramponi G, Nordlund P. (1994). The crystal structure of a low-molecular-weight phosphotyrosine protein phosphatase. *Nature* 370(6490):575-578.
274. Taberner L, Evans BN, Tishmack PA, Van Etten RL, Stauffacher CV. (1999). The structure of the bovine protein tyrosine phosphatase dimer reveals a potential self-regulation mechanism. *Biochemistry* 38:11651-11658.
275. Madhurantakam C, Rajakumara E, Mazumdar PA, Saha B, Mitra D, Wiker HG, Sankaranarayanan R, Das AK. (2005). Crystal structure of low-molecular-weight protein tyrosine phosphatase from *Mycobacterium tuberculosis* at 1.9-Å resolution. *J Bacteriol* 187(6):2175-2181.
276. Xu H, Xia B, Jin C. (2006). Solution structure of a low-molecular-weight protein tyrosine phosphatase from *Bacillus subtilis*. *J Bacteriol* 188(4):1509-1517.
277. Tolkathev D, Shavkhutdinov R, Xu P, Plamondon J, Watson DC, Young NM, Ni F. (2006). Three-dimensional structure and ligand interactions of the low molecular weight protein tyrosine phosphatase from *Campylobacter jejuni*. *Protein Sci* 15(10):2381-2394.
278. Gustafson CLT, Stauffacher CV, Hallenga K, Van Etten RL. (2005). Solution structure of the low-molecular-weight protein tyrosine phosphatase from *Trichomonas foetus* reveals a flexible phosphate binding loop. *Protein Sci* 14(10):2515-2525.
279. Jia Z, Barford D, Flint AJ, Tonks NK. (1995). Structural basis for phosphotyrosine peptide recognition by protein tyrosine phosphatase 1B. *Science* 268(5218):1754-1758.
280. Yuvaniyama J, Denu JM, Dixon JE, Saper MA. (1996). Crystal structure of the dual specificity protein phosphatase VHR. *Science* 272(5266):1328-1331.

281. Chiarugi P, Cirri P, Marra F, Raugei G, Camici G, Manao G, Ramponi G. (1997). LMW-PTP is a negative regulator of insulin-mediated mitotic and metabolic signaling. *Biochem Biophys Res Commun* 238(2):676-682.
282. Chiarugi P, Cirri P, Raugei G, Camici G, Dolfi F, Berti A, Ramponi G. (1995). PDGF receptor as a specific in vivo target for low *Mr* phosphotyrosine protein phosphatase. *FEBS Lett* 372:49-53.
283. Blanchetot C, Chagnon M, Dubé N, Hallé M, Tremblay ML. (2005). Substrate-trapping techniques in the identification of cellular PTP targets. *Methods* 35(1):44-53.
284. Chiarugi P, Cirri P, Marra F, Raugei G, Fiaschi T, Camici G, Manao G, Romanelli RG, Ramponi G. (1998). The Src and signal transducers and activators of transcription pathways as specific targets for low molecular weight phosphotyrosine-protein phosphatase in platelet-derived growth factor signaling. *J Biol Chem* 273:6776-6785.
285. Rigacci S, Rovida E, Bagnoli S, Dello Sbarba P, Berti A. (1999). Low M(r) phosphotyrosine protein phosphatase activity on fibroblast growth factor receptor is not associated with enzyme translocation. *FEBS Lett* 459(2):191-194.
286. Chiarugi P, Cirri P, Taddei L, Giannoni E, Camici G, Manao G, Raugei G, Ramponi G. (2000). The low M(r) protein-tyrosine phosphatase is involved in Rho-mediated cytoskeleton rearrangement after integrin and platelet-derived growth factor stimulation. *J Biol Chem* 275(7):4640-4646.
287. Taddei ML, Chiarugi P, Cirri P, Buricchi F, Fiaschi T, Giannoni E, Talini D, Cozzi G, Formigli L, Raugei G, Ramponi G. (2002). Beta-catenin interacts with low-molecular-weight protein tyrosine phosphatase leading to cadherin-mediated cell-cell adhesion increase. *Cancer Res* 62(22):6489-6499.

288. Kikawa KD, Vidale DR, Van Etten RL, Kinch MS. (2002). Regulation of the EphA2 kinase by the low molecular weight tyrosine phosphatase induces transformation. *J Biol Chem* 277:39274–39279.
289. Rigacci S, Rovida E, Dello Sbarba P, Berti A. (2002). Low Mr phosphotyrosine protein phosphatase associates and dephosphorylates p125 focal adhesion kinase, interfering with cell motility and spreading. *J Biol Chem* 277(44):41631-41636.
290. Stein E, Lane AA, Cerretti DP, Schoecklmann HO, Schroff AD, Van Etten RL, Daniel TO. (1998). Eph receptors discriminate specific ligand oligomers to determine alternative signaling complexes, attachment, and assembly responses. *Genes Dev* 12(5):667-678.
291. Rigacci S, Talini D, Berti A. (2003). LMW-PTP associates and dephosphorylates STAT5 interacting with its C-terminal domain. *Biochem Biophys Res Commun* 312(2):360-366.
292. Rigacci S, Guidotti V, Parri M, Berti A. (2008). Modulation of STAT5 interaction with LMW-PTP during early megakaryocyte differentiation. *Biochemistry* 47(6):1482-1489.
293. Mukhopadhyay A, Kennelly PJ. (2011). A low molecular weight protein tyrosine phosphatase from *Synechocystis* sp. strain PCC 6803: enzymatic characterization and identification of its potential substrates. *J Biochem* 149(5):551-562.
294. Bugert P, Geider K. (1997). Characterization of the *amsI* gene product as a low molecular weight acid phosphatase controlling exopolysaccharide synthesis of *Erwinia amylovora*. *FEBS Lett* 400(2):252-256.
295. Mukherjee S, Dhar R, Das AK. (2009). Analyzing the catalytic mechanism of protein tyrosine phosphatase PtpB from *Staphylococcus aureus* through site-directed mutagenesis. *Int J Biol Macromol* 45(5):463-469.

296. Bose SK, Taneja V. (1998). Induction of a germination specific, low molecular weight, acid phosphatase isozyme with specific phosphotyrosine phosphatase activity in lentil (*Lens esculenta*) seeds. *Biochem Biophys Res Commun* 250:629-634.
297. Szöör B. (2010). Trypanosomatid protein phosphatases. *Mol Biochem Parasitol* 173(2):53-63.
298. Bucciantini M, Chiarugi P, Cirri P, Taddei L, Stefani M, Raugei G, Nordlund P, Ramponi G. (1999). The low M_r phosphotyrosine protein phosphatase behaves differently when phosphorylated at Tyr¹³¹ or Tyr¹³² by Src kinase. *FEBS Lett* 456(1):73-78.
299. Tailor P, Gilman J, Williams S, Couture C, Mustelin T. (1997). Regulation of the low molecular weight phosphotyrosine phosphatase by phosphorylation at tyrosines 131 and 132. *J Biol Chem* 272(9):5371-5374.
300. Songyang Z, Shoelson SE, McGlade J, Olivier P, Pawson T, Bustelo XR, Barbacid M, Sabe H, Hanafusa H, Yi T, Ren R, Baltimore D, Ratnofsky S, Feldman RA, Cantley LC. (1994). Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk, and Vav. *Mol Cell Biol* 14(4):2777-2785.
301. Cirri P, Chiarugi P, Taddei L, Raugei G, Camici G, Manao G, Ramponi G. (1998). Low molecular weight protein-tyrosine phosphatase tyrosine phosphorylation by c-Src during platelet-derived growth factor-induced mitogenesis correlates with its subcellular targeting. *J Biol Chem* 273:32522-32527.
302. Rigacci S, Bucciantini M, Marzocchini R, Berti A. (1998). Low molecular weight phosphotyrosine protein phosphatase translocation during cell stimulation with platelet-derived growth factor. *FEBS Lett* 432:145-149.

303. Malentacchi F, Marzocchini R, Gelmini S, Orlando C, Serio M, Ramponi G, Raugei G. (2005). Up-regulated expression of low molecular weight protein tyrosine phosphatases in different human cancers. *Biochem Biophys Res Commun* 334(3):875-883.
304. Marzocchini R, Malentacchi F, Biagini M, Cirelli D, Luceri C, Caderni G, Raugei G. (2008). The expression of low molecular weight protein tyrosine phosphatase is up-regulated in 1,2-dimethylhydrazine-induced colon tumours in rats. *Int J Cancer* 122(7):1675-1678.
305. Glebov OK, Rodriguez LM, Nakahara K, Jenkins J, Cliatt J, Humbyrd CJ, DeNobile J, Soballe P, Simon R, Wright G, Lynch P, Patterson S, Lynch H, Gallinger S, Buchbinder A, Gordon G, Hawk E, Kirsch IR. (2003). Distinguishing right from left colon by the pattern of gene expression. *Cancer Epidemiol Biomarkers Prev* 12(8):755-762.
306. Zelinski DP, Zantek ND, Stewart JC, Irizarry AR, Kinch MS. (2001). EphA2 overexpression causes tumorigenesis of mammary epithelial cells. *Cancer Res* 61:2301-2306.
307. Chiarugi P, Taddei ML, Schiavone N, Papucci L, Giannoni E, Fiaschi T, Capaccioli S, Raugei G, Ramponi G. (2004). LMW-PTP is a positive regulator of tumor onset and growth. *Oncogene* 23:3905-3914.
308. Zantek ND, Azimi M, Fedor-Chaiken M, Wang B, Brackenbury R, Kinch MS. (1999). E-cadherin regulates the function of the EphA2 receptor tyrosine kinase. *Cell Growth Differ* 10(9):629-638.
309. AmoebaDB: the Amoeba Genomics Resource
<http://amoebadb.org/amoeba/>
310. Aurrecochea C, Barreto A, Brestelli J, Brunk BP, Caler EV, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Iodice J, Kissinger JC, Kraemer ET, Li W, Navak

- V, Pennington C, Pinney DF, Pitts B, Roos DS, Srinivasamoorthy G, Stoeckert CJ Jr, Treatman C, Wang H. (2011). AmoebaDB and MicrosporidiaDB: functional genomic resources for Amoebozoa and Microsporida species. Nucl Acids Res 39 (Database issue):D612-619.
311. BLAST (Basic Local Alignment Search Tool) (National Center for Biotechnology Information)
<http://blast.ncbi.nlm.nih.gov/Blast.cgi>
312. LALIGN finds internal duplications by calculating non-intersecting local alignments of protein or DNA sequences.
http://fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi?rm=lalign
313. EuPathDB: Eukaryotic Pathogen Database Resources
<http://eupathdb.org/eupathdb/>
314. Phylogeny.fr: robust phylogenetic analysis for the non-specialist (main site)
<http://www.phylogeny.fr/>
315. Phylogeny.fr: robust phylogenetic analysis for the non-specialist (mirror site)
<http://phylogeny.lirmm.fr/phylo.cgi/index.cgi>
316. The Seattle Structural Genomics Center for Infectious Disease (SSGCID)
<http://www.ssgcid.org/home/index.asp>
317. The Seattle Structural Genomics Center for Infectious Disease (SSGCID): Participating Institutions
http://www.ssgcid.org/home/Participating_Institutions.asp
318. National Institute of Allergy and Infectious Diseases: Structural Genomics Centers for Infectious Diseases
<http://www.niaid.nih.gov/labsandresources/resources/dmid/sg/Pages/default.aspx>

319. Bryan CM, Bhandari J, Napuli AJ, Leibly DJ, Choi R, Kelley A, Van Voorhis WC, Edwards TE, Stewart LJ. (2011). High-throughput protein production and purification at the Seattle Structural Genomics Center for Infectious Disease. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 67(Pt 9):1010-1014.
320. Choi R, Kelley A, Leibly D, Nakazawa Hewitt S, Napuli A, Van Voorhis W. (2011). Immobilized metal-affinity chromatography protein-recovery screening is predictive of crystallographic structure success. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 67(Pt 9):998-1005.
321. Quartley E, Alexandrov A, Mikucki M, Buckner F, Hol WG, DeTitta G, Phizicky EM, Grayhack E. (2009). Heterologous expression of *L. major* proteins in *S. cerevisiae*: a test of solubility, purity, and gene recoding. *J Struct Funct Genomics* 10(3):233-247.
322. Studier FW. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41(1):207-234.
323. Sigma-Aldrich: SigmaFast protease inhibitor cocktail tablet technical bulletin <http://www.sigmaaldrich.com/etc/medialib/docs/Sigma/Bulletin/s8820bul.Par.0001.File.tmp/s8820bul.pdf>
324. Hayashi Y, Nakamura S, Takemiya A, Takahashi Y, Shimazaki K, Kinoshita T. (2010). Biochemical characterization of in vitro phosphorylation and dephosphorylation of the plasma membrane H⁺-ATPase. *Plant Cell Physiol* 51(7):1186-1196.
325. Whisnant AR, Gilman SD. (2002). Studies of reversible inhibition, irreversible inhibition, and activation of alkaline phosphatase by capillary electrophoresis. *Anal Biochem* 307(2):226-234.

326. Silverman RB. (2002). p.82, in The organic chemistry of enzyme-catalyzed reactions, revised edition. Academic Press, London, UK, or San Diego, CA, USA. 717 pp.
327. Invitrogen: An engineered thermosensitive alkaline phosphatase for dephosphorylating DNA
<http://www.invitrogen.jp/focus/173093.pdf>
328. Sigma-Aldrich: Sodium Orthovanadate
http://www.sigmaaldrich.com/catalog/ProductDetail.do?lang=en&N4=S6508|SIGMA&N5=SEARCH_CONCAT_PNO|BRAND_KEY&F=SPEC
329. Gordon JA. (1991). Use of vanadate as protein-phosphotyrosine phosphatase inhibitor. *Methods Enzymol* 201:477-482.
330. Senna R, Simonin V, Silva-Neto MAC, Fialho E. (2006). Induction of acid phosphatase activity during germination of maize (*Zea mays*) seeds. *Plant Physiol Biochem* 44:467-473.
331. The Science Advisory Board: sodium orthovanadate activation
http://www.scienceboard.net/resources/protocols.asp?action=article&protocol_id=584
332. AppliChem: Sodium orthovanadate *BioChemica* (scroll to bottom of page)
<http://www.applichem.com/en/shop/product-detail/as/natriumorthovanadat-ibiochemical/>
333. Thermo Scientific: Instructions, HALT™ Phosphatase Inhibitor Cocktail
<http://www.piercenet.com/instructions/2161754.pdf>
334. Garton AJ, Flint AJ, Tonks NK. (1996). Identification of p130^{cas} as a substrate for the cytosolic protein tyrosine phosphatase PTP-PEST. *Mol Cell Biol* 16(11):6408-6418.
335. Garton AJ, Flint AJ, Tonks NK. (1999). Identification of substrates for protein-tyrosine phosphatases. In Protein phosphorylation: a practical approach. D.G. Hardie, editor; second edition. Oxford University Press Inc., New York, NY, USA.

336. Herrera-Rodríguez SE, Baylón-Pacheco L, Talamás-Rohana, Rosales-Encina, JL. (2006). Cloning and partial characterization of *Entamoeba histolytica* PTPases. *Biochem Biophys Res Commun* 342:1014-1021.
337. National Center for Biotechnology Information Protein Database
<http://www.ncbi.nlm.nih.gov/protein/>
338. J. Craig Venter Institute: Genomic sequencing center for infectious diseases: *Entamoeba invadens* IP1
http://gsc.jcvi.org/projects/msc/entamoeba/entamoeba_invadens_ip1/index.shtml
339. Ganguly A, Lohia A. (2001). The cell cycle of *Entamoeba invadens* during vegetative growth and differentiation. *Mol Biochem Parasitol* 112(2):277-285.
340. Rengpien S, Bailey GB. (1975). Differentiation of *Entamoeba*: a new medium and optimal conditions for axenic encystations of *E. invadens*. *J Parasitol* 61(1):24-30.
341. Coppi A, Eichinger D. (1999). Regulation of *Entamoeba invadens* encystation and gene expression with galactose and *N*-acetylglucosamine. *Mol Biochem Parasitol* 102;67-77.
342. RSCB Protein Data Bank (RSCB PDB)
<http://www.rcsb.org/pdb/home/home.do>
343. UCLA Molecular Instrumentation Center: crystallography in a nutshell: the lattice and the unit cell
<http://www.mic.ucla.edu/Xray/tutorials.htm#The%20Lattice%20and%20the%20Unit%20Cell>
344. Wolfram|Alpha computational knowledge engine for $P4_3$
<http://www.wolframalpha.com/input/?i=crystallographic+space+group+P43&lk=1>
345. UCLA Molecular Instrumentation Center: crystallography in a nutshell: the space group
<http://www.mic.ucla.edu/X-ray/tutorials.htm#The%20Space%20Group>

346. High-Resolution Space Group Diagrams and Tables
<http://img.chem.ucl.ac.uk/sgp/large/sgp.htm>
347. Space Group Notation
<http://img.chem.ucl.ac.uk/sgp/misc/notation.htm>
348. Wolfram|Alpha computational knowledge engine for $P2_12_12_1$
<http://www.wolframalpha.com/input/?i=crystallographic+space+group+P212121>
349. Wolfram|Alpha computational knowledge engine for $P4_32_12$
<http://www.wolframalpha.com/input/?i=crystallographic+space+group+P43212>
350. Huyer G, Liu S, Kelly J, Moffat J, Payette P, Kennedy B, Tsaprailis G, Gresser MJ, Ramachandran C. (1997). Mechanism of inhibition of protein-tyrosine phosphatases by vanadate and pervanadate. *J Biol Chem* 272:843-851.
351. What-when-how in-depth information: Iodoacetamide, Iodoacetate (Molecular Biology)
<http://what-when-how.com/molecular-biology/iodoacetamide-iodoacetate-molecular-biology/>
352. New England Biolabs BSA FAQ. See Q8 for the molecular weight of BSA
<http://www.neb.com/nebecomm/products/faqproductB9001.asp>
353. NCBI Conserved Domains: smart00849: Lactamase_B
<http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?ascbin=8&maxaln=10&seltype=2&uid=197917>
354. NCBI Conserved Domains: pfam03358:FMN_red
<http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=pfam03358&islf=1>
355. Wasserfallen A, Ragetti S, Jouanneau Y, Leisinger T. (1998). A family of flavoproteins in the domains Archaea and Bacteria. *Eur J Biochem* 254(2):325-332.

356. Bush K, Jacoby GA, Medeiros AA. (1995). A functional classification scheme for beta-lactamases and its correlation with molecular structure. *Antimicrob Agents Chemother* 39(6):1211-1233.
357. NCBI Conserved Domains: cd06612: STKc_MST1_2
<http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?ascbin=8&maxaln=10&seltype=2&uid=132943>
358. NCBI Conserved Domains: cl02144: TLD superfamily
<http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=cl02144>
359. NCBI Conserved Domains: cd07307: The Bin/Amphiphysin/Rvs (BAR) domain
<http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?ascbin=8&maxaln=10&seltype=2&uid=153271>
360. NCBI Conserved Domains: cd00174: SH3 domains
<http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?ascbin=8&maxaln=10&seltype=2&uid=29136>
361. Welch MD, DePace AH, Verma S, Iwamatsu A, Mitchison TJ. (1997). The human Arp2/3 complex is composed of evolutionarily conserved subunits and is localized to cellular regions of dynamic actin filament assembly. *J Cell Biol* 138(2):375-384.
362. The NetPhos 2.0 server for prediction of phosphorylation sites
<http://www.cbs.dtu.dk/services/NetPhos/>
363. Anamika K, Bhattacharya A, Srinivasan N. (2008). Analysis of the protein kinome of *Entamoeba histolytica*. *Proteins* 71:995-1006.
364. Carman CV, Lisanti MP, Benovic JL. (1999). Regulation of G protein-coupled receptor kinases by caveolin. *J Biol Chem* 274(13):8858-8864.

365. Coats SR, Pledger WJ, Awazu M, Daniel TO. (1996). Detergent solubility defines an alternative itinerary for a subpopulation of PDGF beta receptors. *J Cell Physiol* 168(2):412-423.
366. Santa Cruz Biotechnology, Inc: n-Octylglucoside: sc-29103
<http://www.scbt.com/datasheet-29103-n-octylglucoside.html>
367. Bugga L, Ratnaparkhi A, Zinn K. (2009). The cell surface receptor Tartan is a potential in vivo substrate for the receptor tyrosine phosphatase Ptp52F. *Mol Cell Biol* 29(12):3390-3400.
368. Simons K, Gerl MJ. (2010). Revitalizing membrane rafts: new tools and insights. *Nat Rev Mol Cell Biol* 11(10):688-699.
369. Lillemeier BF, Pfeiffer JR, Surviladze Z, Wilson BS, Davis MM. (2006). Plasma membrane-associated proteins are clustered into islands attached to the cytoskeleton. *Proc Natl Acad Sci USA* 103(50):18992-18997.
370. Mittal K, Welter BH, Temesvari LA. (2008). *Entamoeba histolytica*: lipid rafts are involved in adhesion of trophozoites to host extracellular matrix components. *Exp Parasitol* 120(2):127-34.
371. Simons K, Toomre D. (2000). Lipid rafts and signal transduction. *Nat Rev Mol Cell Biol* 1:31-39.
372. Laughlin RC, McGugan GC, Powell RR, Welter BH, Temesvari LA. (2004). Involvement of raft-like plasma membrane domains of *Entamoeba histolytica* in pinocytosis and adhesion. *Infect Immun* 72:5349-5357.
373. Razani B, Woodman SE, Lisanti MP. (2002). Caveolae: from cell biology to animal physiology. *Pharmacol Rev* 54(3):431-467.

374. Sargiacomo M, Scherer PE, Tang Z, Kübler E, Song KS, Sanders MC, Lisanti MP. (1995). Oligomeric structure of caveolin: implications for caveolae membrane organization. *Proc Natl Acad Sci USA* 92(20):9407-9411.
375. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucl Acids Res* 36 (suppl 2):W465-W469.
376. Anisimova M, Gascuel O. (2006). Approximate likelihood ratio test for branches: a fast, accurate and powerful alternative. *Syst Biol* 55(4):539-552.
377. Castresana J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17(4):540-552.
378. Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl Acids Res* 32(5):1792-1797.
379. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307-321.
380. Regular tetragonal pyramid
<http://www.metafysica.nl/turing/vierzijdigepyramide.gif>
381. Salisbury DE, Ford WE. (1922). p.128 in A Text-Book of Mineralogy, 3rd Edition. John Wiley & Sons, New York, New York, USA.
382. Tetragonal trapezohedron
http://www.metafysica.nl/turing/tetragonal_trapezohedron.jpg