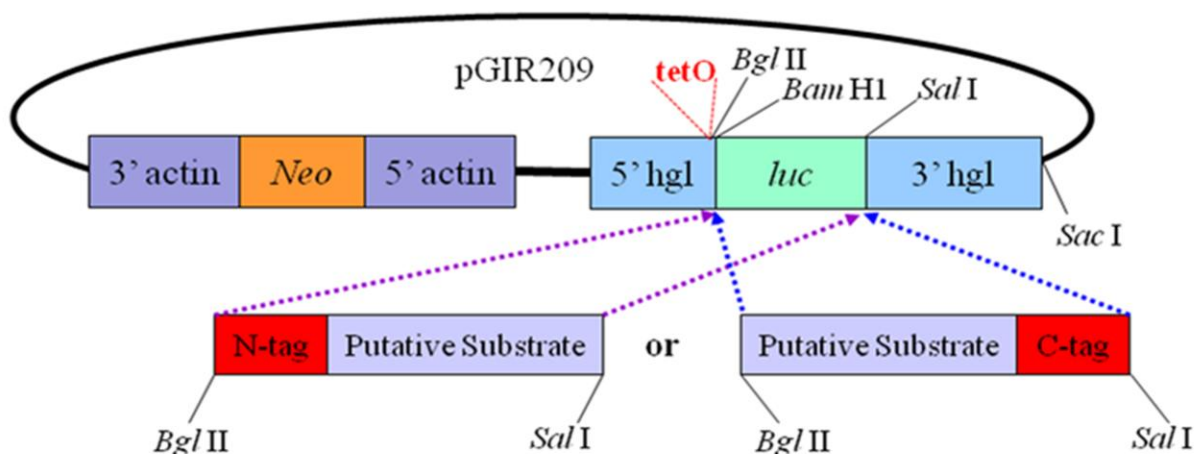


Supplemental Figures for Linford A. et al. Crystal structure and putative substrate identification for the *Entamoeba histolytica* low molecular weight tyrosine phosphatase.



Supplemental Fig. 1. Schematic representation of epitope-tagged putative substrate genes subcloned into pGIR209. Selected putative substrate genes identified via mass spectrometry (Table 1) were subcloned into pGIR209 with epitope tags E, S, and KT3 added. 5' actin and 3' actin, 5' and 3' flanking regions of the *Entamoeba histolytica* actin gene; *Neo*, neomycin resistance gene; *luc*, luciferase gene; 5' *hgl* and 3' *hgl*, 5' and 3' flanking regions of the *E. histolytica hgl* gene; tetO, tetracycline operator; Putative Substrate, genes for putative substrate proteins of the 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. As a control for transfection, drug selection, and expression of an epitope-tagged protein in amebae, URE3-BP was also cloned and expressed. The pGIR209 diagram is adapted from Ramakrishnan G, Vines RR, Mann BJ, Petri WA Jr. A tetracycline-inducible gene expression system in *Entamoeba histolytica*. Mol Biochem Parasitol 1997; 84(1):93-100, and Katz U, Bracha R, Nuchamowitz Y, Milstein O, Mirelman D. Comparison between constitutive and inducible plasmid vectors used for gene expression in *Entamoeba histolytica*. Mol Biochem Parasitol 2003; 128(2):229-33.

Ent his 1	1	-----MKLLFVCLGNICRSPAAEAVMKKVIQNHHLTEKYI-----CDSAGTCSYHGQQQDSRM-----	54	Ent his 1	55	-----RK-VGKSR-GYQVDSI-----SRPVSDDFKNFYIIFAMNDNYDEL-----L-----D-RC--P	99
Ent his 2	1	-----MKLLFVCLGNICRSPAAEAVMKKVIQNHHLTEKYI-----CDSAGTCSYHGQQQDSRM-----	54	Ent his 2	55	-----RK-VGKSR-GYQVDSI-----SRPVSDDFKNFYIIFAMNDNYDEL-----L-----D-RC--P	99
Ent dispar	1	-----MKLLFVCLGNICRSPAAEAVMKKVIQNHHLTEKYI-----CDSAGTCSYHGQQQDSRM-----	54	Ent dispar	55	-----RK-VGKSR-GYQVDSI-----SRPVSDDFKNFYIIFAMNDNYDEL-----L-----D-RC--P	99
Ent invad	1	-----MKLLFVCLGNICRSPAAEAVMKKVIQNHHLTEKYI-----CDSAGTCSYHGQQQDSRM-----	54	Ent invad	55	-----RK-VGKSR-GYQVDSI-----SRPVSDDFKNFYIIFAMNDNYDEL-----L-----D-RC--P	99
Gia lam P15	1	-----MKLLFVCLGNICRSPAAEAVMKKVIQNHHLTEKYI-----CDSAGTCSYHGQQQDSRM-----	54	Gia lam P15	55	-----RK-VGKSR-GYQVDSI-----SRPVSDDFKNFYIIFAMNDNYDEL-----L-----D-RC--P	99
Giard intes	1	-----MKLLFVCLGNICRSPAAEAVMKKVIQNHHLTEKYI-----CDSAGTCSYHGQQQDSRM-----	54	Giard intes	55	-----RK-VGKSR-GYQVDSI-----SRPVSDDFKNFYIIFAMNDNYDEL-----L-----D-RC--P	99
Dictyo disc	1	-----MSADSKNQKLVFLVCLGNICRSTMAEIVLRLGVLHRSRLLDQF-----IDSAGTCSYHIGDTPPDRV-----	63	Dictyo disc	64	QSCNQNMGR-A-ISEES-LKHFKSIPLHRAQQTDEDTSKFDYIIFAMDESINLSMKKVL-----KHST	125
Klu lactis	1	-----MSEQVSVAFVCLGNICRSPAAEAVFKHVEKGLQKFKRIEESFGTAGYHIVGESSPDSRS-----	59	Klu lactis	60	-----AQYKSKHFKFDYIIFAMDESINLSMKKVL-----KHST	125
Sacch cerev	1	-----MTIEKPKSVAFIICLGNICRSPAAEAVFKHVEKGLQKFKRIEESFGTAGYHIVGESSPDSRS-----	59	Sacch cerev	63	-----V-LCKQH-GVKTHK-----GKQKTKHFDYIIFAMDESINLSMKKVL-----KHST	125
Schiz pombe	1	-----MTKNIQVLFVCLGNICRSPAAEAVFNRVEKAGLEAFRTDSDSCGTGAHWGNRPDR-----	62	Schiz pombe	60	-----LE-VLKN-GVTKHL-----ARKLSDFKNFYIIFAMDESINLSMKKVL-----KHST	125
Sorg bicol	97	MAAEASTADAAAAAEPFAVLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	168	Sorg bicol	169	-----RA-ASKKR-GIEVTSI-----SRPVPDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	222
Zea mays 1	89	MAAEASTAGAAAAPFAVLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	160	Zea mays 1	161	-----RA-ASKKR-GIEVTSI-----SRPVPDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	214
Zea mays 2	1	MAAEASTAGAAAAPFAVLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	72	Zea mays 2	73	-----RA-ASKKR-GIEVTSI-----SRPVPDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	126
Oryza sat	85	MAAEASAAAEGA-QAKPFSVLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	155	Oryza sat	156	-----RS-ASKKR-GIEVTSI-----SRPVPDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	209
Hord vulg	8	MAAEASTAPAEA-SAKPFAVLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	148	Hord vulg	149	-----IS-TSKKR-GIEVTSI-----SRPVPDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	202
Arab thal	1	MATPPPTQ-SS-----ETKPYSLVFLVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	68	Arab thal	69	-----RS-AAKKR-GIEVTSI-----SRPVPDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	122
Vit vinif	1	MASGPSIDT-----QSKPYSVLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	67	Vit vinif	68	-----RA-ASKKR-GIEVTSI-----SRPVPDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	121
Pop tricho	1	-----MASS-TETKPPAVLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	63	Pop tricho	64	-----RA-ASKKR-GIEVTSI-----SRPVPDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	117
Ricin comm	1	-----MASSTETKPPAVLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	66	Ricin comm	66	-----RA-ASKKR-GIEVTSI-----SRPVPDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	119
Nost punct	1	-----MTIDRGKQTSMPYKLLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	55	Nost punct	67	-----SA-AAAKLGKIKRQ-----ARQPKKSDFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	113
Lynghya sp	1	-----MPYKLLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	56	Lynghya sp	57	-----TK-AASR-GILKQG-----ARQPKKSDFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	102
Cyano sp	1	-----MPYKLLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	56	Cyano sp	57	-----NK-AASR-GILKQG-----ARQPKKSDFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	101
Arthro max	1	-----MSYKLLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	56	Arthro max	57	-----TM-AAKR-GIVLKG-----ARQPKKSDFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	102
Prochlo mar	1	-----MIRKVLFLVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	56	Prochlo mar	57	-----RS-AAKR-GIEVTSI-----SRPVPDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	101
Bdello bact	1	-----MTRGNNSPKIIMKRSKLLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	69	Bdello bact	70	-----IL-HGERR-GYDTSI-----SRPVPDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	114
Bact caccae	1	-----MKLLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	55	Bact caccae	56	-----RA-HAARR-GYQVHR-----SRPVPDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	102
Bact theta	1	-----MKLLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	55	Bact theta	56	-----RA-HAARR-GYQVHR-----SRPVPDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	102
Bact fragil	1	-----MKLLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	56	Bact fragil	57	-----RA-HAARR-GYQVHR-----SRPVPDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	103
Parab merd	1	-----MMEKKKGEKLLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	62	Parab merd	64	-----RS-HAARR-GYQVHR-----SRPVPDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	108
Vibrio chol	1	-----MKRVLVVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	55	Vibrio chol	56	-----KA-AGER-GYFDSI-----KARKIRDFDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	101
Pseudo aeru	1	-----MKRVLVVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	55	Pseudo aeru	55	-----RL-AAQR-GYDTSI-----RGRQVDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	104
Homo isof b	1	-----MAEQATKSVLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	60	Homo isof b	61	-----VS-CLRNH-GIHTAH-----ARQTKDFDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	104
Homo isof c	1	-----MAEQATKSVLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	60	Homo isof c	61	-----QS-CMKRH-GIHTAH-----ARQTKDFDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	104
Mus musc 1	1	-----MAEVGSKSVLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	60	Mus musc 1	61	-----VS-CLRNH-GIHTAH-----ARQTKDFDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	104
Mus musc 2	1	-----MAEVGSKSVLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	60	Mus musc 2	61	-----QS-CMKRH-GIHTAH-----ARQTKDFDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	104
Bos taurus	1	-----MAEQATKSVLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	60	Bos taurus	61	-----QS-CMKRH-GIHTAH-----ARQTKDFDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	104
Xenop laev	1	-----MAEQATKSVLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	61	Xenop laev	62	-----LK-CLSKH-SIHTAH-----ARQTKDFDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	106
Gal gallus	1	-----MAEQATKSVLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	60	Gal gallus	61	-----QT-CMKRH-GIHTAH-----ARQTKDFDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	104
Dros melan	1	-----MVRKVLFLVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	55	Dros melan	56	-----IS-TLQKH-GIHTAH-----ARQTKDFDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	100
Danio rerio	1	-----MAASSGKSVLFVCLGNICRSPAAEAVFRTLVRKRGDLSKFL-----IDSAGTIGYHEGNKADSRM-----	60	Danio rerio	61	-----LA-CLRKH-GIHTAH-----ARQTKDFDFRFDLIIAMDRQNYDEL-----LNSFERWRKKE-PL--P	105

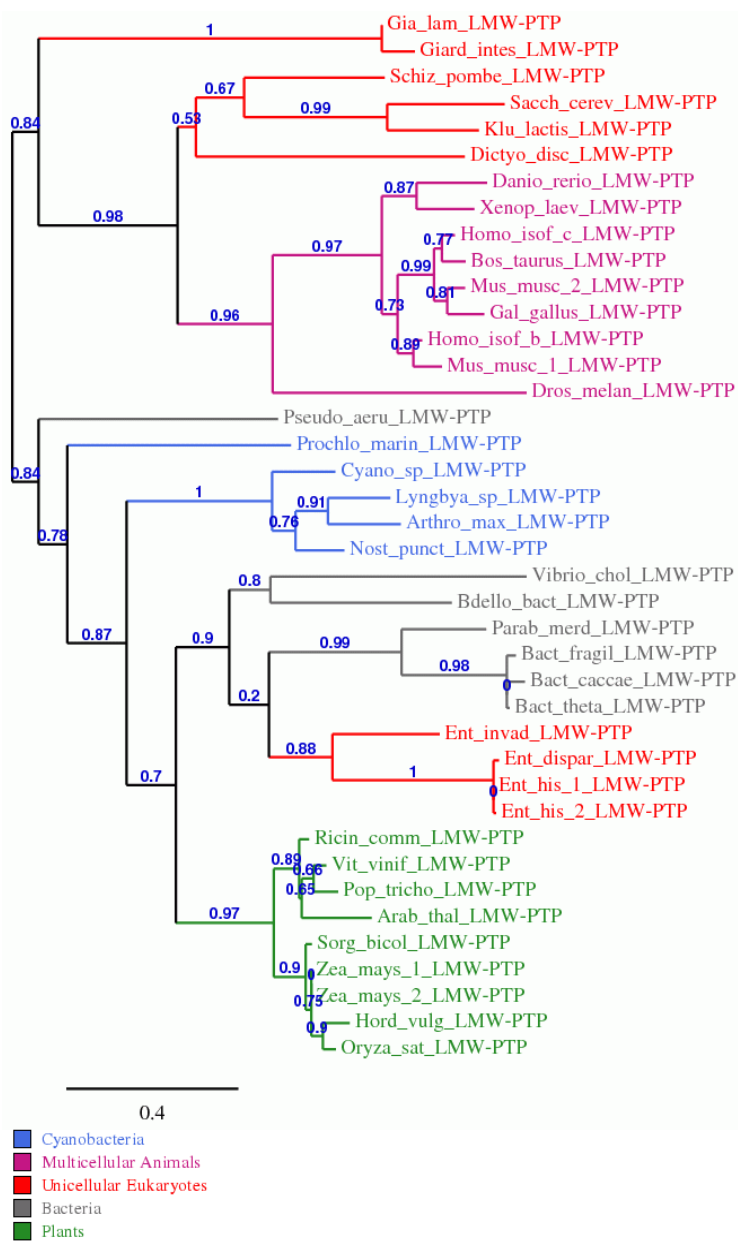
Supplemental Fig. 2. Alignment of the protein sequence of the *E. histolytica* LMW-PTPs with other selected LMW-PTP sequences.

The *E. histolytica* LMW-PTP protein XP_656359 sequence (Ent his 1) was aligned with other LMW-PTPs. As compared with

XP_656359: **Identical amino acids**, **conservative substitutions**, **active site**, **DPYY loop**. Supplemental Figs 3 and 4 show phylogenetic trees for all selected LMW-PTPs without and with the PTP1B outgroup.

Supplemental Data Set 2 lists tabulates species, LMW-PTP accession numbers, and percent identity and similarity to the *E. histolytica* LMW-PTP protein XP_656359 for each LMW-PTP and PTP1B homologous sequence for each organism.

Ent his 1	100	E-Q-YKQK-IFRMVDFCTTIKTE-VDPDYY-GGEGFHRVIDILEDACENLIILK-EEGKLN	157
Ent his 2	100	E-Q-YKQK-IFRMVDFCTTIKTE-VDPDYY-GGEGFHRVIDILEDACENLIILK-EEGKLN	157
Ent dispar	100	E-Q-YKQK-IFRMVDFCTTIKTE-VDPDYY-GGEGFHRVIDILEDACENLIILK-EQGLTN	157
Ent invad	100	K-E-YQDK-IHRMVEYCKKSDSE-VDPDYY-SGLKGEQVIDLEDGKMLLDQL-ENAGDE	156
Gia lam P15	100	K-R-DQSK-VRMFRDYDPEGR-GD-VDPDYY-CGR--FELVYSIVERTCRSLIDRT-KSGKAL	154
Giard intes	100	K-R-DRSK-IRMFDRDYDPEGR-GD-VDPDYY-CGR--FELVYSIVERTCRSLIDRT-KNGKAL	154
Dictyo disc	126	D-N-HIAT-IKRLGEVHTH-KKIN-VEPDYY-GMNSNINCNHVVHDCVNFVNLTE-ESLN	179
Klu lactis	105	K-E-PPAQ-LSLFGKWNNDKFDQ-IVDPDYY-GDSGFEYNFKQIYTFYSKQFKKEL	157
Sacch cerev	108	E-G-SKAK-VCLFGEVNTNDCTVQTIIEFDW-GDIQDFEYNFKQIYTFYSKQFKKEL	161
Schiz pombe	105	C-G-SRAK-VMLFGEYASPGVSKI-VDPDYY-GSDGFGDQYIQLVDFSNFKLSIA	156
Sorg bicol	223	E-N-APDK-VKLMCSYCKRHTESE-VDPDYY-GGPGQFEKVLDDLEDACESLIDSTV--ASISG	279
Zea mays 1	215	D-S-APNK-VKLMCSYCKQHTESE-VDPDYY-GGPGQFEKVLDDLEDACESLIDSTVANNASISG	274
Zea mays 2	127	D-S-APNK-VKLMCSYCKQHTESE-VDPDYY-GGPGQFEKVLDDLEDACESLIDSTVANNASISG	186
Oryza sat	210	D-S-SGPK-VKLMCSYCKRHTESE-VDPDYY-GGPGQFEKVLDDLEDACESLIDSTVAENASVEA	268
Hord vulg	203	D-S-SGPK-VKLMCSYCKRHTESE-VDPDYY-GGPGQFEKVLDDLEDACESLIDSTVAENASVEA	262
Arab thal	123	P-D-ADKK-VKLMCSYCKKHDFE-VDPDYY-GGPGQFEKVLDDLEDACESLIDSTVAQS	177
Vit vinif	122	A-D-AHKK-VKLMCSYCKKHDFE-VDPDYY-GGPGQFEKVLDDLEDACESLIDSTVAENSHLDS	182
Pop tricho	118	D-D-AHKK-VKLMCSYCKKHDFE-VDPDYY-GGPGQFEKVLDDLEDACESLIDSTVAENSHLDS	173
Ricin comm	120	D-D-AHKK-VKLMCSYCKKHDFE-VDPDYY-GGPGQFEKVLDDLEDACESLIDSTVAENSHLDS	172
Nost punct	114	E-Q-YQKH-VRLMCFECSRHTLKE-VDPDYY-GGQDFGNOVIDLILDACAGLLETKVKEEF	169
Lynghya sp	103	R-Q-YQDK-VKLMCFECSRHHNVQE-VDPDYY-GGPGGNOVIDLILDACAGLLETKVQKPPFKAF	162
Cyano sp	102	SG-Q-YQDK-VRLICEFATHYSIKE-VDPDYY-GGPGGNOVIDLILDACAGLLETKVQKPPFKAF	161
Arthro max	103	K-K-YQDK-VKLMCFECSRHTLKE-VDPDYY-GGPGGNOVIDLILDACAGLLETKVQKPPFKAF	160
Prochlo mar	103	ELNISYKAK-VKPLLEYAINTDLVE-VDPDYY-GGEGFEDVNLLENALIEGLRIDIK	156
Bdello bact	115	D-KTLDD-LSLVTDYCESEFVKV-VDPDYY-GGVDGDFHVDLILEDALIEGLRIDIK-ODRSKI	171
Bact caccae	103	E-E-WK-K-IHRMTEYCNRI PADH-VDPDYY-GGAGGEFVLDLEDACAGLLETKVQKPPFKAF	156
Bact theta	103	E-E-WK-K-IHRMTEYCNRI PADH-VDPDYY-GGAGGEFVLDLEDACAGLLETKVQKPPFKAF	156
Bact fragil	104	A-E-WK-K-IHRMTEYCNRI PADH-VDPDYY-GGAGGEFVLDLEDACAGLLETKVQKPPFKAF	157
Parab merd	109	DL-E-SVEK-IHRMTEYCNRI PADH-VDPDYY-GGAGGEFVLDLEDACAGLLETKVQKPPFKAF	169
Vibrio chol	102	C-S-HQHK-LSLMLSHSDS-EYQE-IPDPDYY-GGERGFEVLDLVEDAAEQFLIKL-CCQGQR	155
Pseudo aeru	102	R-G-AABL-DLFLRRYGS-ERDE-VDPDYY-GGEGFEDVLDLIEDALIEGLRIDIK-BAKGR	154
Homo isof b	105	N-Q-VKTRAKIELLGSYDPOKQLI-IEDPYY-GNDSDFETVYQCCVRCRAFLEKAH	158
Homo isof c	105	N-Q-VKTRAKIELLGSYDPOKQLI-IEDPYY-GNDSDFETVYQCCVRCRAFLEKAH	158
Mus musc 1	105	N-Q-VKTRAKIELLGSYDPOKQLI-IEDPYY-GNDSDFETVYQCCVRCRAFLEKAH	158
Mus musc 2	105	N-Q-VKTRAKIELLGSYDPOKQLI-IEDPYY-GNDSDFETVYQCCVRCRAFLEKAH	158
Bos taurus	105	N-Q-VKTRAKIELLGSYDPOKQLI-IEDPYY-GNDSDFETVYQCCVRCRAFLEKAH	158
Xenop laev	107	QV-Q-NC-KAKIELLGSYDPOKQLI-IEDPYY-GRDDEFTVYQCCVRCRAFLEKSS	159
Gal gallus	105	N-Q-VKTRAKIELLGSYDPOKQLI-IEDPYY-GNDSDFETVYQCCVRCRAFLEKAH	158
Dros melan	101	K-G-SK-AEILMLGDFGLEKNRI-IEDPYY-ERGEAF-GEFTAYQCCVVAAMFKERLQK	155
Danio rerio	106	SVE-N-SK-AKI-ELLGSYDPOKQLI-IEDPYY-GSDKDFETVYQCCVRCRAFLEKAH	158

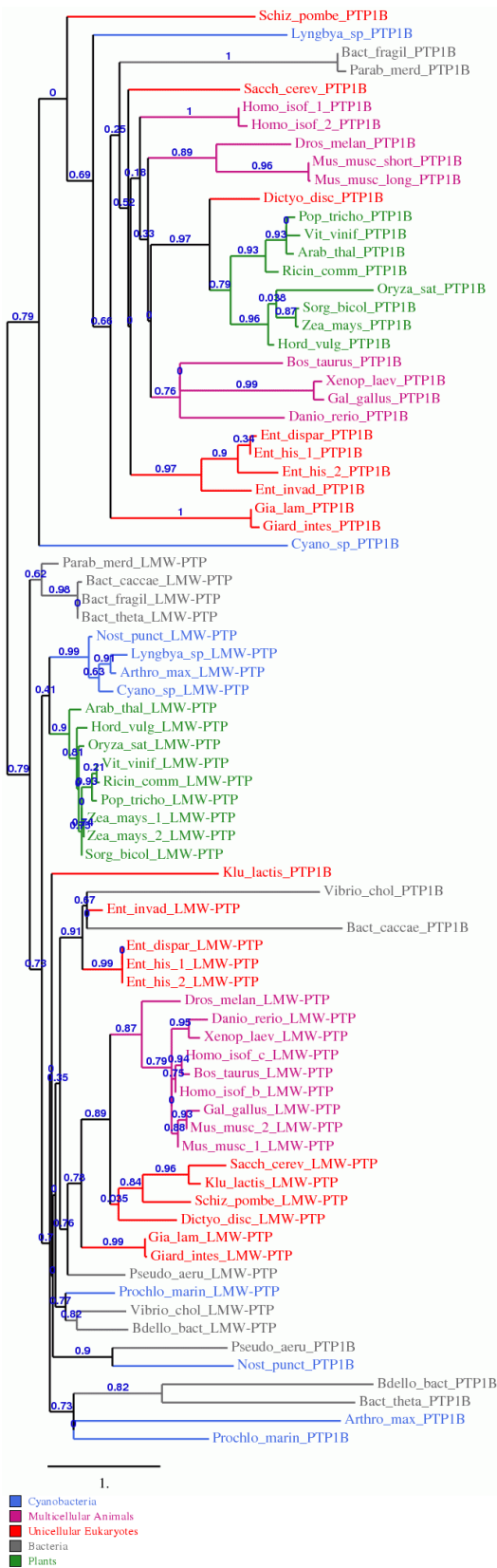


Supplemental Fig. 3. A phylogenetic tree comparing the *E. histolytica* LMW-PTP protein XP_656359 and selected LMW-PTP proteins from different species.

Species abbreviations, LMW-PTP accession numbers, and percent identity and similarity to the *E. histolytica* LMW-PTP protein XP_656359 for each LMW-PTP sequence are found in Supplemental Data Set 2, and LMW-PTP protein alignments are shown in Supplemental Fig. 2. The phylogenetic tree was created with “one-click” phylogenetic analysis using online software [www.phylogeny.fr] by inputting the protein sequences for the selected LMW-PTPs. Selected LMW-PTPs are color-coded in terms of the general phylogenetic groups into which they fall: cyanobacteria, multicellular animals, unicellular eukaryotes, bacteria, and plants. The scale bar at the bottom of the figure indicates the branch length representing the amount of genetic change

(the number of nucleotide substitutions divided by the length of the protein sequence)

[http://epidemic.bio.ed.ac.uk/how_to_read_a_phylogeny]. The numbers in blue (between 0 and 1) at each node indicate the measure of support for the node (the node represents a putative ancestor), with 1 being maximal support. The phylogeny.fr web site uses the approximate likelihood-ratio test (aLRT) for branches to calculate this figure [Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst Biol 2006; 55(4):539-52]. LMW-PTP sequences from cyanobacteria are shown in blue, those from multicellular animals are shown in violet, unicellular eukaryotic LMW-PTP sequences are red, bacterial LMW-PTPs are gray, and plant LMW-PTPs are green.



Supplemental Figure 4. A phylogenetic tree comparing the *E. histolytica* LMW-PTP protein and selected LMW-PTP proteins from different species, and including as an outgroup proteins homologous to PTP1B from those same species. Species abbreviations are in Supplemental Data Set 2, and LMW-PTP protein alignments are shown in Supplemental Fig. 2. The phylogenetic tree was created using online software (www.phylogeny.fr) by inputting the protein sequences for the selected LMW-PTP and PTP1B homologues. The human PTP1B (isoform 1) sequence was input into BLAST [<http://blast.ncbi.nlm.nih.gov/Blast.cgi>] to identify the *E. histolytica* homologue; BLAST was then used to identify PTP1B homologues in the same organisms chosen for LMW-PTP comparison using the *E. histolytica* homologue. Sequences are labeled as LMW-PTP or PTP1B following the species abbreviation, and are color-coded in terms of the general phylogenetic groups into which they fall: blue for cyanobacteria, violet for multicellular animals, red for unicellular eukaryotes, gray for bacteria, and green for plants. The scale bar at the bottom of the figure indicates the branch length representing the amount of genetic change (the number of nucleotide substitutions divided by the length of the protein sequence). The numbers in blue (between 0 and 1) at each node indicate the measure of support for the node (the node represents a putative ancestor), with 1 being maximal support. Accession numbers, and percent identity and similarity to the *E. histolytica* LMW-PTP protein XP_656359 for each LMW-PTP and PTP1B sequence are found in Supplemental Data Set 2.

Supplemental Fig. 3 shows the tree without the PTP1B outgroup, and more information is found in its legend.

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

MNAFILLFIALATAEDIYSPFKRSCVIQ
 NRINTI**TAQTKA**IHNKIRADEQQERNLN
 DLSYLYRDLKYATVAEDRARLELE**TEA**
IQKLHSVRIHKVNMLRDIRKETDKITAF
 PFRDSIARSNKIE**RHVGLDEGNFVDER**
LKTQATI**KKVTSQLANKYATLAAKIAAQ**
SNVAKAAKKNNDAAKQAFVAKKAEAA
 YTKVHSKIVKAI**SAAVENGVARNNVE****TV**
AKTAIEQIVDKLMMKVAQPSVEAADKV
 AQNIVKSQLKQIKKAVKVT**PKVVKV**PKV
 KKAVAKKVVKKAAPKPAAKKI**AKK**

Peptide Sequence
(K) YATVAEDR (A)
(K) VTSQLANK (Y)
(K) VAAQPSVEAADK (V)
(K) TATEQIVDK (L)
(R) NNVE TVAK (T)
(R) LELE IEAIQK (K)
(K) KVTSQLANK (Y)
(K) KFNNDAAK (V)
(R) INTI TAQTK (A)
(K) IAAQSNVAK (A)
(R) HVGLEDEGNFVDER (L)
(K) AISAAVENGVAR (N)

Protein sequence of type A flavoprotein (gi56474182)

MKALEVVDKLYWVGVDFDKELRVFDIIMTTP
 YGTSYNSFLKSE**KGNVLFETCKENFAGEC**
LERIEDVIGKEGKLDYIVLNHTEPDHSGSL
 VHILE**KYPEATVIGTMAALNNIKY**IGHIKE
 NTKTNSGKI**KTLDLGNVHLKFLIQPFLHW**
 PDTMMTVIEEMKVLVSCD**VFGGHYADERVF**
NDQMMERIKDMDAYKHDFDCIFGPFKNVY
 IKGLDMIE**TQMGFPSEDELKAI**CCSHGPVLR
 THIKENIERQWQ**PIALKNKVV**IAYGSA
 YGY**TQMAEQISEGIKSTGVEV**MFNIVES
 SVGDV**LKE**FEDAKGLL**GTP**TLVNDT**IPPI**
 MQIACSLNPT**IHCNRFVQC**FGSFGWS**GEGV**
 KNL**SARIVQLKHQ**PVEPL**SIKQ**FPNSNEL
QTCFEWGGKFAEALKA

Peptide Sequence
(K) ENFAGEcLER (I)
(K) FQPSNELQTcFEWGGK (K)
(R) FVQcFGSFGWSGEGVK (N)
(K) GLDMIE TQMGFPSEDELK (A)
(K) GNVLFETcK (E)
(K) GNVLFETcKENFAGEcLER (I)
(K) HYFDcLFGPFK (N)
(R) IKDMDAYK (H)
(K) MAEQISEGIK (S)
(K) MFNIVESVGDVLEK (E)
(K) TLDLGNVHLK (F)
(R) VFNDQmMER (I)
(K) VLVScdVFGGHYADER (V)
(K) VVIAYGSAYGYTQK (M)
(K) YPEATVIGTmAALNNIK (Y)

Protein sequence of the putative serine/threonine protein kinase (gi56466736)

MTTPTKLQVSMQKINVAECS**KDFEVVEK**
LEGSGVGFHAIHRPTGTHAAVKT**VRIDG**
DTTDVKEIVLMKQCDTDYIVRF**FGHFFVD**
 KDLWVMEICGAGSVCIDMNNLS**LDERE**
 IQIIL**KD**TLHGLAYLH**KMH**HRDI**KAANI**
LVSDDGVCKLADFGV**STK**GGTKLQ**KT**LI
 GTPYMAPEVILESGHNEKAD**IWS**VGI**TAI**
 EMFDGRPPHSELH**AMRAFLI**PNRPP**PLS**
TFEEVSKAFNDFIALCCVKNFRERPSALEL
LRKCDFMKPNGR**KRI**TDL**IEKENE**VIA**FCG**
SRKKALEQ**GEDDE**GDDEZDDDD**TEEDT**
 EEEEE**DDTE**EDSDDG**DDTMVIRK**

Peptide sequence
(K) AANILVSDDGcK (L)
(K) AFNDFIALcCVK (N)
(R) AIFLIPNRPPPLSTPEEVSK (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)

MEELINKNNQLIKEVETLPNIQQSIDDVN
 NNYQSYLESIKNEVFFELKENE**TYSEAI**KRG
 EGI**FDR**ME**TERRR**KL**VI**ESQ**ETIN**QIEI
 MKNKI**SKIIEEC**NNDE**ETFGK**TRNELIDKI
 ISIEEMIMKDKL**FRTEE**TD**ITSEFN**KKK
 EE**WK**HYSDYLE**RK**RRKE**EEEE**KER**KKQE**
 ERLQIMKGMNSMEEMRQIEEW**TKK**VGNVL
FESDVDDWNN**KNTSVFDQK**LNKEHIMIVIE
 DEEGNKFGGYVNSKID**EV**ESCITDQ**KS**FLF
 SLE**SKGR**M**KRM**KFD**IK**EPQ**CAF**Y**LN**SQTD
 DCL**FE**FG**GH**DI**CV**F**KN**NTQ**SYCY**Q**RSF**
EYNGIS**NALCGK**QLPER**FTQ**KRIMV**IE**MK

Peptide Sequence
(K) VGNVLFESDVEDDWNK (N)
(K) GmNSMEEMR (Q)
(K) IIEECNNDEETFGK (T)
(K) KVGNVLFESDVEDDWNK (N)
(K) NTSVFDQK (I)
(R) RTEE TDTITSEFNK (K)
(R) SFYEYNGISNALcGK (Q)

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

MAPNLLLESYNRIVYARVLEKLTTEGGIKK
 GLKE**KLADFNQSIWYIERT**KEKMTISVSTP
 CWKQ**LKENRSLEFLQ**KVYGS**DL**VEADNS**E**
 YNLIT**PI**EKHDKPEEFAMNA**AKLLTNMLI**
GPAVV**L**AD**EV**KNKADEKL**VQIDYR**PGESY
 WLK**PN**GDRL**TVIF**SI**K**FDD**KD**DA**V**FG**RVFI**
NEFSK**SAAG**CP**SCD**V**TR**KN**AP**PPSE**L**KT**V**
 E**GL**AD**NCY**IS**FL**LE**KR**HL**NP**Q**KT**LE**VLM**
SCRNYIN**F**HI**K**CSKA

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)

MSGIGDAFGRKFYQIKTHVGA**QK**TMDSDV
QYAKNK**LE**SY**K**FK**N**LDV**IK**LAP**TV**HA
 TNL**M**Q**VE**VL**T**SL**GD**CV**V**NT**SP**ET**K**SD**IDS****I**
ISTF**Q**IDE**GV**NT**YE**TR**IES**DI**IV**PL**K**TY**M**
 EQ**F**VM**E**K**RF**ICH**N**RR**V**DM**D**RY**H**DS**V**LS**I**
SKK**PP**G**K**Q**S**GL**GEA**Q**N**K**Y**N**V**ARD**LY**N**LR**N
 E**I**AD**VE**KL**T**SS**A**EV**VS**PI**CG**TL**IV**SY**TD**
 YLNHLNNY**W**GE**A**SE**V**SN**F**RI**C**AL**D**P**API**
 TP**G**ESS**M**VID**AN**V**S**KK**S**SD**V**MG**T**Y**E**P**SE**
 F**EN**SG**S**TT**T**ST**T**TT**T**TT**V**ST**S**SG**R**PP**P**
 P**PP**SR**KE**Q**VR**CE**Y**EQ**AE**Q**GE**LS**F**KE**GD**I
 I**TV**L**K**KE**GD**W**W**L**GE**L**K**D**N**K**D**T**S**H**T**IM**FL**H**F**
 K**FT**FL**K**D**Y**F**IS**SS**L**F**IQ**N**K**EL

Peptide Sequence
(K) EGDWLLGELK (D)
(K) IDEGVNTYETR (I)
(R) IESDIIVPLK (T)
(K) QSGLGEAQNK (Y)
(K) SDIDSISTFQK (I)
(K) TmSDVQYAK (N)
(R) YHDSVLSISK (K)

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

MSVTTLSAEQ**W**HDHFIVHTSKAGYLAQS
 ISAVLGYPI**IP**I**HK**FS**D**GE**CY**RI**AI**PE
 RTSLVGRD**VVV**SS**IV**DD**Q**EL**LE**V**V**R**IG**GE
 LAELG**TK**R**R**I**F**VI**P**Y**LM**Y**ST**ME**RAVHPGEV**
VTCK**S**TV**R**ML**CG**IP**SS**GL**GN**LF**ML**MD**L**HT**S**
 GI**I**HY**F**EG**T**VQ**AM**E**LY**A**ES**T**LE**Q**AI**EN**ID**
 F**SE**P**VI**FG**S**AD**L**GR**PL**W**VE**TF**AN**H**F**GV**G**IA
 F**IR**K**S**RS**F**ED**TH**VI**GE**PI**GD**V**K**G**H**V**VI**Y**D**
 DL**TR**S**AG**SL**IK**AC**N**AY**DL**NG**AI**K**V**T**AV**V**TH**
 L**SL**IDD**V**VQ**V**IDS**R**ID**RV**IA**T**NS**N**VR**S**Q
 MD**AV**K**NS****KRF**I**IC**DI**SP**V**FA**Q**I**RA**CL**NN
 ND

Peptide Sequence
(R) AVHPGEVVTcK (S)
(K) FSDGEcYYR (I)
(K) KFIICDISPVFAK (Q)
(R) VIATNSNVR (S)

Supplemental Fig. 5. Sequences of unique peptides and their locations in the protein sequences of putative substrate proteins identified by mass spectrometry and selected for further evaluation. Locations of the unique peptides are shown in yellow in the protein sequence, and the list of unique peptides identified for each protein is located in the table to the right of each protein sequence.

Hypothetical protein 328.t00002 (gi56465624; XM_644144)

1 MNAFILLFIALATAEDIYSPFKRSCVIQNRINTITAQTKAHNRKIRADEQ
 51 QERNLNDLSLYLYRDLKATVAEDRARLELEIEAIQKHLHVRHKNVNL
 101 RDIRKETDKITAPFRDSIARSNKIERHVGLEDEGNFVDERLKTQATIKKV
 151 TSQLANKYATLAAKIAAQNVAKAAKFNNDAAKVQAFVAKKAEAAATYKV
 201 HSKIIVKAIASAAVENGAVRNNEVETAKTAIEQIVDKLMMKVAQPSVEAAD
 251 KVQNIIVKSQLKQIKKAVKVTFKVKVVKPVKAVAKVVKVKAAPKPAAKK
 301 IAKK

Phosphorylation sites predicted: Tyr: 3

Name	Pos	Context	Score	Pred
328.t00002	18	AEDTYSPPFK	0.948	*Y*
328.t00002	68	RLDKYATVA	0.970	*Y*
328.t00002	197	AEAAATYKVVH	0.829	*Y*

Type A flavoprotein (gi56474182; XM_651854)

1 MKALEVVKDLYVWGVDFKELRVFDIIMTTPYGTSYNSFLKSEKGNVLFEE
 51 TCKENFAGECLERIEDVIGKEGKLDIVLNHTEPDHSGSLVHILEKYPEA
 101 TVIGTMAALNNIKYIGHIKENTKTLNSGKIKTLDLGNVHLKFLIQPFLHW
 151 PDTMTVIEEMKVLVSCDVFVGGHADERVFNQMMERIKDMDDAKHYFD
 200 CIFGPFKNYVIKGLDMIETQMGGFSDDELKAIACCSHGVLTRTHIKENIERY
 251 RQWAQPIALKNRVVIAYGSAYGYTQKMAEQISEGIKSTGVEVKMFNIVES
 301 SVGDVLEKFEFADAKGLLLGTPPTLVNDTIPPIQIACSLNPTIHCNRFVQCF
 351 GSPGWSGEGVKNLSARIVQLKVHQPVEPLSIFQPNNSNLQTCFEWGGKFF
 401 AEALKA

Phosphorylation sites predicted: Tyr: 4

Name	Pos	Context	Score	Pred
Flavopr	76	GKLDYIVLN	0.821	*Y*
Flavopr	174	FGGHYADER	0.795	*Y*
Flavopr	195	MDDAYKHYP	0.513	*Y*
Flavopr	209	PFKNYVIKG	0.574	*Y*

Serine/threonine protein kinase, putative (gi56466736; XM_645070)

1 MTTPTKLQVSMQKINWVAECSKDFVEVKEKLEGGSSVFAIHRPTGTH
 51 AAVKTVRIDGDTTVEKEIVLMKQCDTYIVRFFGHFFVDKDLWVMEIC
 101 GAGSVCIDIMNINLSTLDEREIQIILKDTLHGLAYLHKHMHHRDIKAANI
 151 LVSDDGVCKLADFGVSTKMGGLNKQKTLIGTPYWMAPVILESGHNEKA
 201 DIWSVGITAIEMFDRPPELHARRAIPLIPNRPPTLSTPEEVSKAFN
 251 DFIALCCVKNFRERFSALELLKCCDFMKNPGRKRITDLEKENEVIAEFG
 301 SRKALAEQGEDDDEGDDEQDDDDTEEDTEEEEDTEEDSDDGDDTM
 351 VIRK

Phosphorylation sites predicted: Tyr: 2

Name	Pos	Context	Score	Pred
ProtKina	36	GEGSYGGSVF	0.816	*Y*
ProtKina	79	CDTYIVRF	0.809	*Y*

Conserved hypothetical protein (gi56465028, XP_001914335)

1 MEELINKNNQLIKEVETPLNQQSIDDDVNNYQSYLESIKNEVFELKEN
 51 ETYSEAIKRGEGIFDRMDETERRRKLVEKSQETINQIEIMNKISKIIE
 101 ECNNDDETFGKTRNELIDKIISEEMIMKDKLFRRTTEETDTITSEFNKKK
 151 EEWKKHYSYDLERKKRKEEBEERKQEEERLQIMKGMNSMEEMRQIEE
 201 WTNKKGNVLFESVDVDDWKNSTVFDQKIKNKEHIMIVIEDEEGNKFGGY
 251 VNSKIDEVESCITDQKSLFLESKGRMKRMMKFDKEPQCAFYLNSTQTD
 301 DCLFEFFGFDICVFKENKNTQSYCYQRSFENYINGISNALCGKQLPERFTQ
 351 KRIMVIEEMK

Phosphorylation sites predicted: Tyr: 3

Name	Pos	Context	Score	Pred
Cons hyp	53	ENETYSEAI	0.927	*Y*
Cons hyp	157	WKKHYSYDL	0.562	*Y*
Cons hyp	250	KFGYVNSK	0.953	*Y*

Ribose-phosphate pyrophosphokinase, putative (gi56473321, XP_656141)

1 MSVTTLSLAEQWQHDHFVHTSKAGYLAQSIASVAVLGYPIIPIHKKFFSDG
 51 ECYRIAPERTSLVGRDVVVSSIVDDQELLEVVRIIGGELAEALGTRRI
 101 FVIPIYLMYSTMERAVHPGEVVTCKSTVRMLCGIPSSGLNLFMLMDLHTS
 151 GIHYFEGTVQAMELAEASTLEQAIAENIDFSEPVIFGSADLGRPLWVET
 201 FANHFVGVIAPIKRSRSEDTHVIGEPIDGVKGGHVVIYDDLTRISAGSLI
 251 KACNAYLDNGAIVKVTAVVTHLSLIDDDVQKVIDSRIDRVATNSNVRVRSQ
 301 MDAVKNSKKFIICDISPVFAKQIRACLNNND

Phosphorylation sites predicted: Tyr: 3

Name	Pos	Context	Score	Pred
Rib-phos	53	DGECYRIA	0.784	*Y*
Rib-phos	166	AMELYAEST	0.956	*Y*
Rib-phos	256	ACNAYLDNG	0.947	*Y*

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

1 MAPNLLLESYNRIVYARVLEKLTTEGGIKKGLKEKLADEFNOSIWIYERTK
 51 EKMTISVSTPCWKQLKENRSLFELQKVVYSGDLVEADSENENLILTIPEK
 101 HDKPEEFAMNAAKLLTNLIGPAVVLADVKNKKADEKLVQIDYRPGESY
 151 WLKPNGRDLTVIFSLKFDKDDAVFGRVFNFEFSKSAAGCPSCDVVTRKN
 201 APPPELKTVEGLAEDNCISFLLEKRHLNNPQKTLVLMSCRNYINFHI
 251 KCSKAFHIRMNRKVSHLQVLNRAKPERVEKKTASGRFTFK

Phosphorylation sites predicted: Tyr: 4

Name	Pos	Context	Score	Pred
Arp 2/3	91	SNSYLNLL	0.767	*Y*
Arp 2/3	150	PGESYWLKP	0.721	*Y*
Arp 2/3	219	EDNYSISFL	0.884	*Y*
Arp 2/3	245	SCRNYINFH	0.737	*Y*

Hypothetical protein 503.t00001 (gi56464309, XP_656019)

1 MSGIGDAFGRKFIQIKTHVAGQKMTSDSVYAKNKLSEYKFKFNILDV
 51 IKKLAPTVAHATNLQVEVLTSLGDCVNTSPETKSDIDSIIISFQKIDEG
 101 VNTYETRIESDIIIVPLKTYMEQPKVMEKRFIECHNRVDMDRYHDSVLSI
 151 SKKPPGKQSLGEAQNRYNVARLDYLNREIADVKELTSSAEVVSPI
 201 CGTLIVSYTDYLNHNNYWGAESEVSSNFRICALDPAPITPGESSMVID
 251 ANVFSKSSDVMMDTYEPSEFENSGSSTTTSTTTTTTSTSSGKRPPPP
 301 PPSRKEQVRCEYEQAQEQGELSFKREGDIIITVLKREGDWLGLKDNKDT
 351 SHTIMFLHFKFTFELKDYFISFSLFTFNKEL

Phosphorylation sites predicted: Tyr: 6

Name	Pos	Context	Score	Pred
501.t	32	SDVQYAKNK	0.957	*Y*
501.t	41	LSEYKFKF	0.519	*Y*
501.t	143	DMDRYHDSV	0.971	*Y*
501.t	211	SYTDYLNHL	0.927	*Y*
501.t	266	MDGTYEPSE	0.607	*Y*
501.t	314	CEYEQAQE	0.575	*Y*

Calcium-binding protein, putative, URE3-sequence specific DNA binding protein (URE3-BP) (AAG18423).

1 MQPPVANFCLWNLQPIQGSWMGAACIYQMPPSVNTWVFLPNTLPLDQY
 51 TRIYQWFMGVDRDRSGTLEINELMMGQFPGGIRLSPQTLRMMRIFDTFD
 101 NGHISFYEFMAMYKFMELAYNLVFMNDRNRSGTLEPHEILPALQLGFYI
 151 NQRTSLLLHRLFARGMAFCDLNCWIAICAFAAQTRSAQMIQMPNYGPM
 201 KPFNMEFGKFLDVTLSLLE

Phosphorylation sites predicted: Tyr: 2

Name	Pos	Context	Score	Pred
URE3-BP	50	PLDQYTRIY	0.804	*Y*
URE3-BP	197	MNPYQGMK	0.698	*Y*

Supplemental Fig. 6. Protein sequences and predicted tyrosine phosphorylation sites in selected putative substrate proteins. Tyrosine phosphorylation sites were predicted using the NetPhos 2.0 server at

<http://www.cbs.dtu.dk/services/NetPhos/>. URE3-BP was included as a control. Protein sequences are shown with predicted phosphorylated tyrosines in bold and highlighted, with their surrounding residues in a table below. The closer the score is to a value of one the more likely the phosphorylation of that tyrosine residue.

Supplemental Table 1. Crystallographic X-ray data and refinement statistics. Crystallized recombinant *E. histolytica* LMW-PTP representing different states along the catalytic pathway were subjected to X-ray diffraction in order to solve their structures: the enzyme with no ligand bound (apo structure), with the substrate mimic HEPES bound, and with the phosphate analog sulfate bound. Citation for MolProbity: Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 2007; 35(Web Server issue):W375-W383.

	Apo structure	HEPES (form 1)	HEPES (form 2)	Phosphate analog
PDB ID	3ILY	3IDO	3JS5	3JVI
Crystal parameters				
Space group	$P4_3$	$P2_12_12_1$	$P2_12_12_1$	$P4_32_12$
Cell dimensions a=b=c (Å), $\alpha=\beta=\gamma$ (°)	45.7, 45.7, 132.3 90, 90, 90	60.6, 70.4, 79.2 90, 90, 90	38.6, 60.3, 72.56 90, 90, 90	45.0, 45.0, 133.3 90, 90, 90
Data set				
X-ray Source	FRE+ Superbright	ALS 5.0.3	MicroMax 007 HF	FRE+ Superbright
Wavelength (Å)	1.54	0.9765	1.5418	1.5418
Resolution (Å)	50-2.2 (2.28-2.20)	50-2.2 (2.28-2.20)	50-1.95 (2.02-1.95)	50-1.80 (1.86-1.80)
Rmerge	0.063 (0.316)	0.140 (0.492)	0.066 (0.198)	0.059 (0.231)
I/sigma (I)	23.6 (4.3)	14.0 (2.4)	19.6 (4.7)	17.7 (4.1)
Completeness	98.3% (85.4%)	98.6% (89.5%)	97.9% (84.9%)	94.6% (98.2%)
# reflections, unique	13,504 (1154)	17,494 (1539)	12,697 (1069)	23,136
Multiplicity	4.5 (3.4)	8.1 (5.7)	4.4 (2.2)	4.3 (2.5)
Refinement statistics				
R _{work}	0.211 (0.245)	0.206 (0.281)	0.169 (0.215)	0.206 (0.260)
R _{free}	0.264 (0.325)	0.254 (0.315)	0.212 (0.244)	0.249 (0.361)
RMSD bond lengths (Å)	0.016	0.010	0.009	0.012
RMSD bond angles (Å)	1.529	1.211	1.087	1.273
Ramachandran:				
Preferred	96.2%	98.7%	98.1%	97.4%
allowed	99.3%	100%	100%	99.4%
disallowed	0.7%	0%	0%	0.6%
MolProbity clash score	7.82 (95 th)	4.94 (98 th)	4.00 (98 th)	10.22 (72 nd)
Molprobity score	21.8 (75 th)	1.56 (98 th)	1.19 (99 th)	2.01 (62 nd)