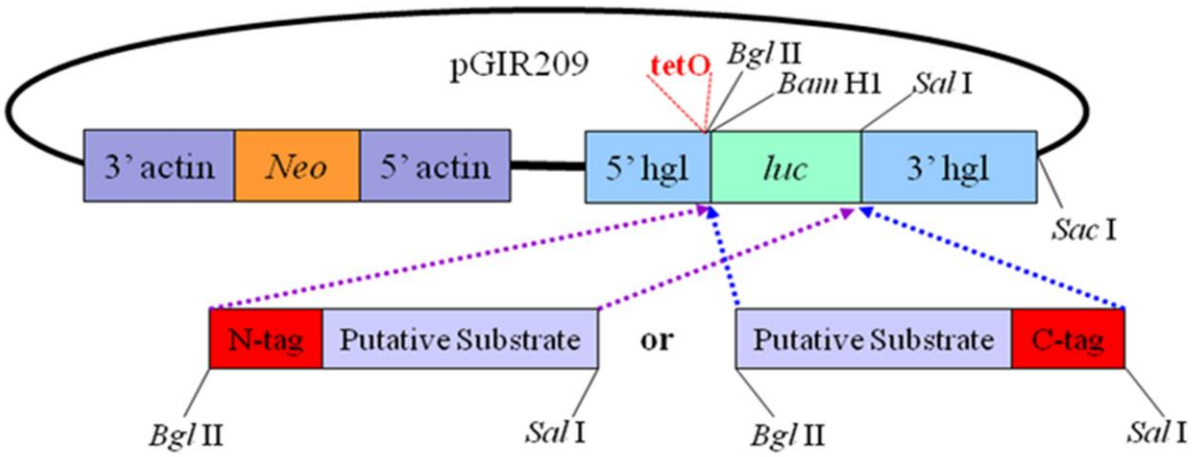
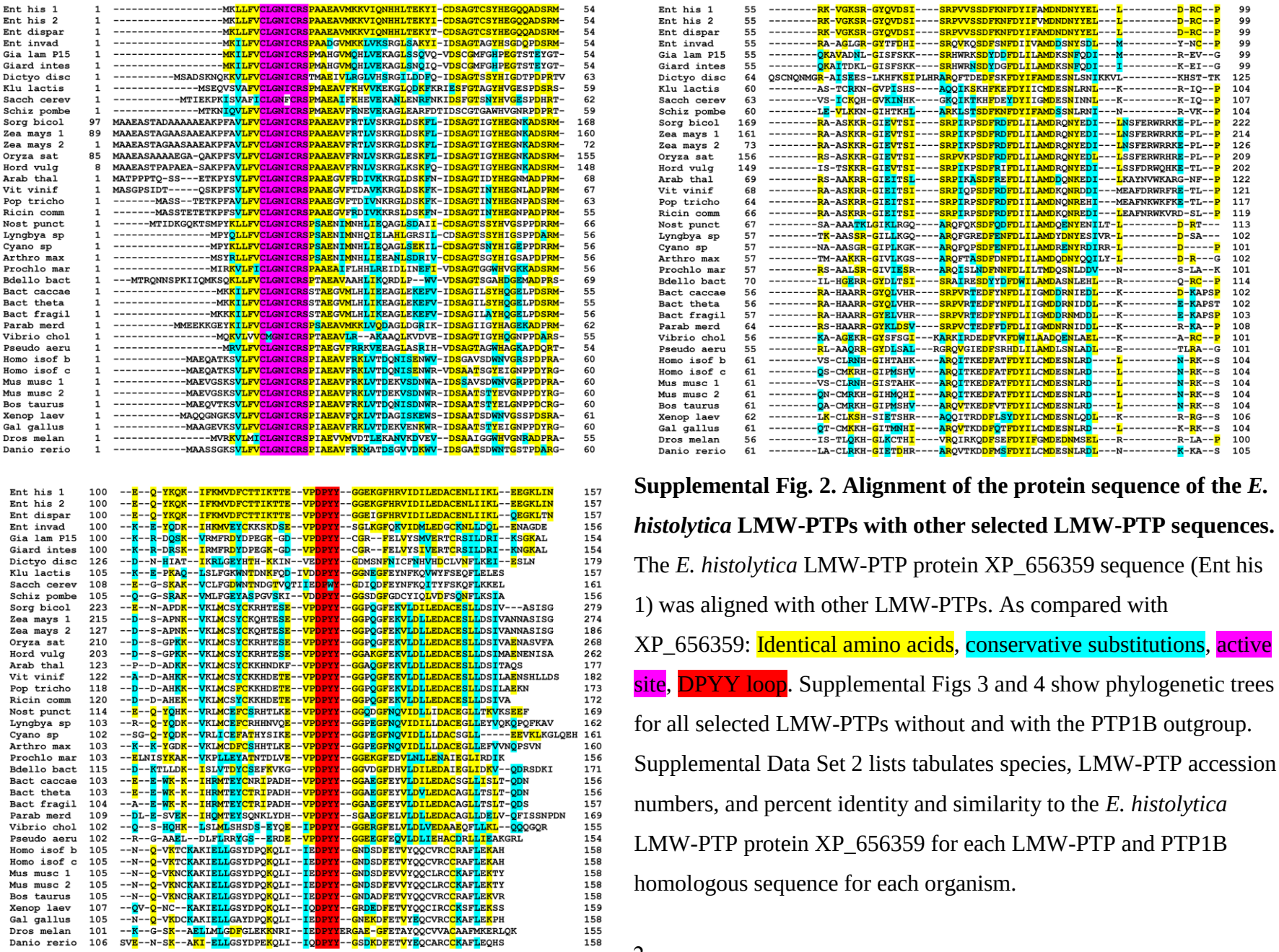
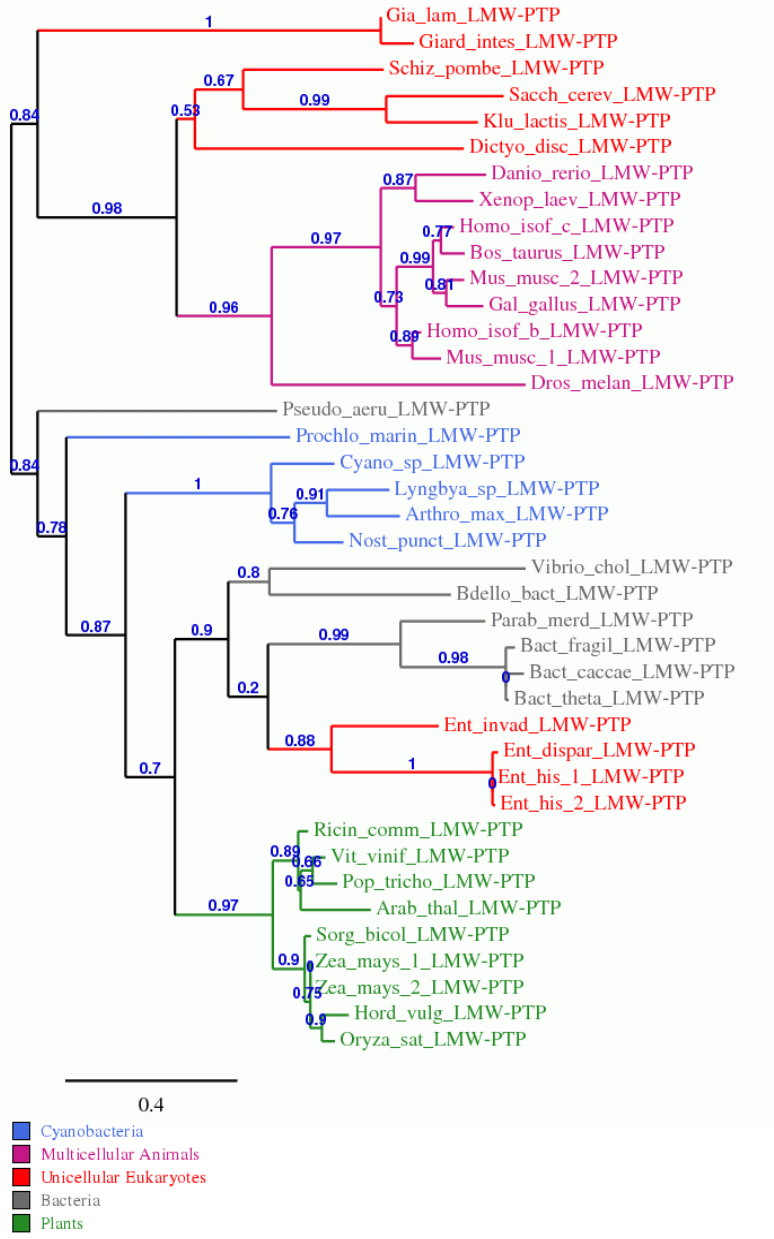
Supplemental Figures for Linford A. et al. Crystal structure and putative substrate identification for the *Entamoeba histolytica* low molecular weight tyrosine phosphatase.



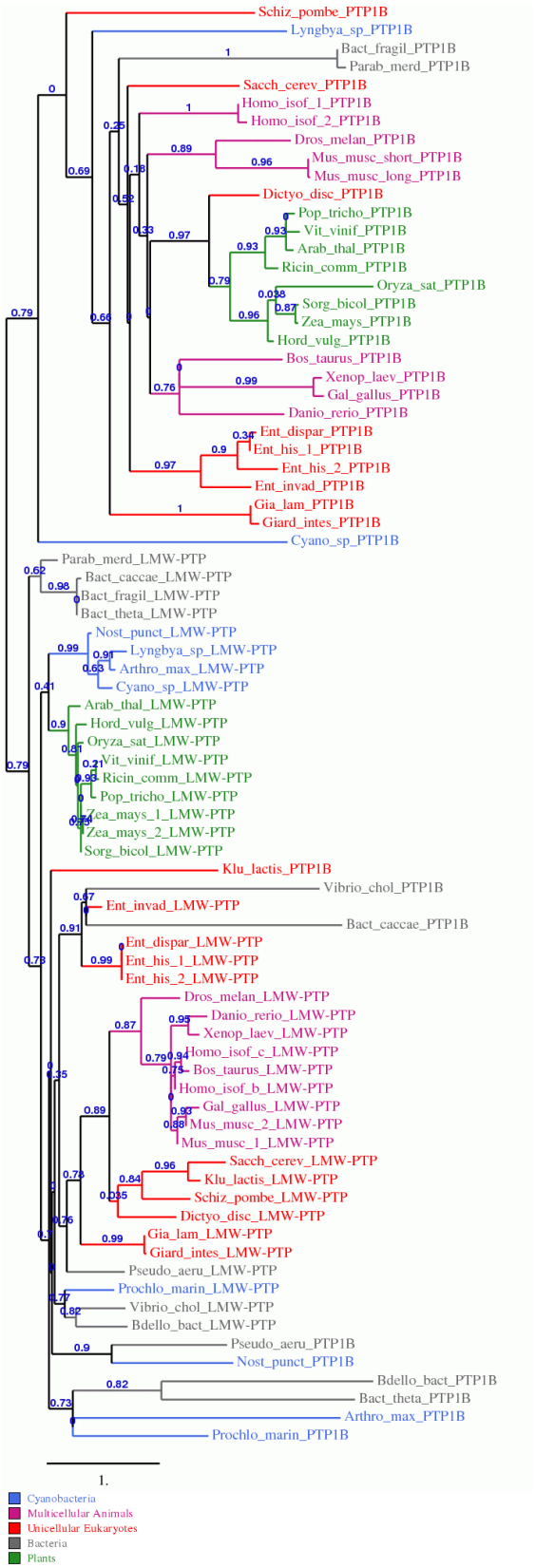
Supplemental Fig. 1. Schematic representation of epitope-tagged putative substrate genes subcloned into pGIR209. Selected putative substrate genes identified via mass spectrometry (Table 1) were subcloned into pGIR209 with epitope tags E, S, and KT3 added. 5′ actin and 3′ actin, 5′ and 3′ flanking regions 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.



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.



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" phylogenic 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
NRINTITAQTKAIHNKIRADEQQERNLN
DSLYYLYRDLKYATVAEDRARLELEIEA
IQKKLHSVRIHKVNMLRDIRKETDKITA
PFRDSIARSNKIERHVGLEDEGNFVDER
LKTQATIKKVTSQLANKYATLAAKIAAQ
SNVAKAAKKFNNDAAKVQAFVAKKAEAA
YTKVHSKIVKAISAAVENGVARNNVETV
AKTAIEQIVDKLMMKVAAQPSVEAADKV
AQNIVKSQLKQIKKAVKVTPKKVVKKPV
KKAVAKKVVKKAAPKPAAKKIAKK

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

Protein sequence of the conserved hypothetical protein (gi56465028)

MEELINKNNQLIKEVVETLPNIQQSIDDVN
NNYQSYLESIKNEVFELKENETYSEAIKRG
EGIFDRMDETERRRKLVIEKSQETINQIEI
MKNKISKIIEECNNDEETFGKTRNELIDKI
ISIEEMIMKDKLFRRTEETDTITSEFNKKK
EEWKKHYSDYLERKKRKEEEEEKERKKQEE
ERLQIMKGMNSMEEMRQIEEWTNKKVGNVL
FESDVDDWNKNTSVFDQKIKNKEHIMIVIE
DEEGNKFGGYVNSKIDEVESCITDQKSFLF
SLESKGRMKRMMKFDIKEPQCAFYLNSQTD
DCLFEFGFGHDICVFKENNKTQSYCYQRSF
EYNGISNALCGKQLPERFTQKRIMVIEMK

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

Protein sequence of type A flavoprotein (gi56474182)

MKALEVVKDLYWVGVFDKELRVFDIIMTTP
YGTSYNSFLLKSEKGNVLFETCKENFAGEC
LERIEDVIGKEGKLDYIVLNHTEPDHSGSL
VHILEKYPEATVIGTMAALNNIKYIGHIKE
NTKTLNSGKIKTLDLGNYHLKFLIQPFLHW
PDTMMTVIEEMKVLVSCDVFGGHYADERVF
NDQMMERIKDMDDAYKHYFDCIFGPFKNYV
IKGLDMIETQMGFPSDELKAICCSHGPVLR
THIKENIERYRQWAQPIALKNKVVIAYGSA
YGYTQKMAEQISEGIKSTGVEVKMFNIVES
SVGDVLKEFEDAKGLLLGTPTLVNDTIPPI
MQIACSLNPTIHCNRFVQCFGSFGWSGEGV
KNLSARIVQLKVHQPVEPLSIKFQPNSNEL
QTCFEWGKKFAEALKA

Peptide Sequence
(K) ENFAGEcLER (I)
(K) FQPNSNELQTcFEWGK (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) VLVScDVFGGHYADER (V)
(K) VVIAYGSAYGYTQK (M)
(K) YPEATVIGTmAALNNIK (Y)

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

MAPNLLLESYNRIVYARVLEKLTTEGGIKK
GLKEKLADFNQSIWYIERTKEKMTISVSTP
CWKQLKENRSLEFLQKVYGSDLVEADSNSE
YNLILTIPEKHDKPEEFAMNAAKLLTNMLI
GPAVVLADEVKNNKADEKLVQIDYRPGESY
WLKPNGDRLTVIFSIKFDDKDDAVFGRVFI
NEFSKSAAGCPSCDVVTRKNAPPPSELKTV
EGLAEDNCYISFLLEKRHLNNPQKTLEVLM
SCRNYINFHIKCSKA

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 putative serine/threonine protein kinase (gi56466736)

MTTPTKLQVSQMQKINWVAECSKDFEVVEK
LGEGSYGSVFHAIHRPTGTHAAVKTVRIDG
DTTDVEKEIVLMKQCDTDYIVRFFGHFFVD
KDLWIVMEICGAGSVCDIMNILNSTLDERE
IQIILKDTLHGLAYLHKMHMIHRDIKAANI
LVSDDGVCKLADFGVSTKMGGTNKLQKTLI
GTPYWMAPEVILESGHNEKADIWSVGITAI
EMFDGRPPHSELHAMRAIFLIPNRPPPTLS
TPEEVSKAFNDFIALCCVKNFRERPSALEL
LKCDFMKPNGNRKRITDLIEKENEVIAEFG
SRKKALEQGEDDDEGDDEZDDDDDTEEDDT
EEEEEDDTEEDDSDDGDDTMVIRK

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 hypothetical protein 503.t00001 (gi56464309)

MSGIGDAFGRKFYQIKTHVGAGQKTMDSDV
QYAKNKLSESYKKFKNILDVIKKLAPTVHA
TNLMQVEVLTSLGDCVVNTSPETKSDIDSI
ISTFQKIDEGVNTYETRIESDIIVPLKTYM
EQFKVMEKRFEICHNRRVDMDRYHDSVLSI
SKKPPGKQSGLGEAQNKYNVARDLYNYLRN
EIIADVEKLTSSAEEVVSPICGTLIVSYTD
YLNHLNNYWGEASEVSSNFRICALDPAPII
TPGESSMVIDANVFSKKSSDVMGGTYEPSE
FENSGSSTTTSTTTTTTTTVSTSSGKRPPP
PPPSRKEQVRCEYEYQAQEQGELSFKEGDI
ITVLKKEGDWWLGELKDNKDTSHTIMFLHF
KFTFELKDYFISFSSLFIFQNKEL

Peptide Sequence
(K) EGDWWLGELK (D)
(K) IDEGVNTYETR (I)
(R) IESDIIVPLK (T)
(K) QSGLGEAQNK (Y)
(K) SDIDSIISTFQK (I)
(K) TmDSDVQYAK (N)
(R) YHDSVLSISK (K)

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

MSVTTSLSAEQWQHDHFIVHTSKAGYLAQS
ISAVLGYPIIPIIHKKFSDGECYYRIAIPE
RTSLVGRDVVVVSSIVDDQELLEVVRIGGE
LAELGTKRRIFVIPYLMYSTMERAVHPGEV
VTCKSTVRMLCGIPSSGLGNLFMLMDLHTS
GIIHYFEGTVQAMELYAESTLEQAIAENID
FSEPVIFGSADLGRPLWVETFANHFGVGIA
FIRKSRSFEDTHVIGEPIGDVKGKHVVIYD
DLTRSAGSLIKACNAYDLNGAIKVTAVVTH
LSLIDDDVVQKVIDSRIDRVIATNSNVRSQ
MDAVKNSKKFIICDISPVFAKQIRACLNNN
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   MNAFILLFIALATAEDIYSPFKRSCVIQNRINTITAQTKAIHNKIRADEQ
 51   QERNLNDSLYYLYRDLKYATVAEDRARLELEIEAIQKKLHSVRIHKVNML
101   RDIRKETDKITAPFRDSIARSNKIERHVGLEDEGNFVDERLKTQATIKKV
151   TSQLANKYATLAAKIAAQSNVAKAAKKFNNDAAKVQAFVAKKAEAAYTKV
201   HSKIVKAISAAVENGVARNNVETVAKTAIEQIVDKLMMKVAAQPSVEAAD
251   KVAQNIVKSQLKQIKKAVKVTPKKVVKKPVKKAVAKKVVKKAAPKPAAKK
301   IAKK
```

Phosphorylation sites predicted: Tyr: 3

Name	Pos	Context	Score	Pred
328.t00002	18	AEDIYSPFK	0.948	*Y*
328.t00002	68	RDLKYATVA	0.970	*Y*
328.t00002	197	AEAAYTKVH	0.829	*Y*

Type A flavoprotein (gi56474182; XM_651854)

```
  1   MKALEVVKDLYWVGVFDKELRVFDIIMTTPYGTSYNSFLLKSEKGNVLFE
 51   TCKENFAGECLERIEDVIGKEGKLDYIVLNHTEPDHSGSLVHILEKYPEA
101   TVIGTMAALNNIKYIGHIKENTKTLNSGKIKTLDLGNYHLKFLIQPFLHW
151   PDTMMTVIEEMKVLVSCDVFGGHYADERVFNDQMMERIKDMDDAYKHYFD
200   CIFGPFKNYVIKGLDMIETQMGFPSDELKAICCSHGPVLRTHIKENIERY
251   RQWAQPIALKNKVVIAYGSAYGYTQKMAEQISEGIKSTGVEVKMFNIVES
301   SVGDVLKEFEDAKGLLLGTPTLVNDTIPPIMQIACSLNPTIHCNRFVQCF
351   GSFGWSGEGVKNLSARIVQLKVHQPVEPLSIKFQPNSNELQTCFEWGKKF
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	MDDAYKHYF	0.513	*Y*
Flavopr	209	PFKNYVIKG	0.574	*Y*

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

```
  1   MTTPTKLQVSQMQKINWVAECSKDFEVVEKLGEGSYGSVFHAIHRPTGTH
 51   AAVKTVRIDGDTTDVEKEIVLMKQCDTDYIVRFFGHFFVDKDLWIVMEIC
101   GAGSVCDIMNILNSTLDEREIQIILKDTLHGLAYLHKMHMIHRDIKAANI
151   LVSDDGVCKLADFGVSTKMGGTNKLQKTLIGTPYWMAPEVILESGHNEKA
201   DIWSVGITAIEMFDGRPPHSELHAMRAIFLIPNRPPPTLSTPEEVSKAFN
251   DFIALCCVKNFRERPSALELLKCDFMKPNGNRKRITDLIEKENEVIAEFG
301   SRKKALEQGEDDDEGDDEQDDDDDTEEDDTEEEEEDDTEEDDSDDGDDTM
351   VIRK
```

Phosphorylation sites predicted: Tyr: 2

Name	Pos	Context	Score	Pred
ProtKina	36	GEGSYGSVF	0.816	*Y*
ProtKina	79	CDTDYIVRF	0.809	*Y*

Conserved hypothetical protein (gi56465028, XP_001914335)

```
  1   MEELINKNNQLIKEVVETLPNIQQSIDDVNNNYQSYLESIKNEVFELKEN
 51   ETYSEAIKRGEGIFDRMDETERRRKLVIEKSQETINQIEIMKNKISKIIE
101   ECNNDEETFGKTRNELIDKIISIEEMIMKDKLFRRTEETDTITSEFNKKK
151   EEWKKHYSDYLERKKRKEEEEEKERKKQEEERLQIMKGMNSMEEMRQIEE
201   WTNKKVGNVLFESDVDDWNKNTSVFDQKIKNKEHIMIVIEDEEGNKFGGY
251   VNSKIDEVESCITDQKSFLFSLESKGRMKRMMKFDIKEPQCAFYLNSQTD
301   DCLFEFGFGHDICVFKENNKTQSYCYQRSFEYNGISNALCGKQLPERFTQ
351   KRIMVIEMK
```

Phosphorylation sites predicted: Tyr: 3

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

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

```
  1   MAPNLLLESYNRIVYARVLEKLTTEGGIKKGLKEKLADFNQSIWYIERTK
 51   EKMTISVSTPCWKQLKENRSLEFLQKVYGSDLVEADSNSEYNLILTIPEK
101   HDKPEEFAMNAAKLLTNMLIGPAVVLADEVKNNKADEKLVQIDYRPGESY
151   WLKPNGDRLTVIFSIKFDDKDDAVFGRVFINEFSKSAAGCPSCDVVTRKN
201   APPPSELKTVEGLAEDNCYISFLLEKRHLNNPQKTLEVLMSCRNYINFHI
251   KCSKAFLHIRMRNKVSHLQLVLNRAKPEREVEKKTASGRTFKK
```

Phosphorylation sites predicted: Tyr: 4

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

Hypothetical protein 503.t00001 (gi56464309, XP_656019)

```
  1   MSGIGDAFGRKFYQIKTHVGAGQKTMDSDVQYAKNKLSESYKKFKNILDV
 51   IKKLAPTVHATNLMQVEVLTSLGDCVVNTSPETKSDIDSIISTFQKIDEG
101   VNTYETRIESDIIVPLKTYMEQFKVMEKRFEICHNRRVDMDRYHDSVLSI
151   SKKPPGKQSGLGEAQNKYNVARDLYNYLRNEIIADVEKLTSSAEEVVSPI
201   CGTLIVSYTDYLNHLNNYWGEASEVSSNFRICALDPAPIITPGESSMVID
251   ANVFSKKSSDVMDGTYEPSEFENSGSSTTTSTTTTTTVSTSSGKRPPPP
301   PPSRKEQVRCEYEYQAQEQGELSFKEGDIITVLKKEGDWWLGELKDNKDT
351   SHTIMFLHFKFTFELKDYFISFSSLFIFQNKEL
```

Phosphorylation sites predicted: Tyr: 6

Name	Pos	Context	Score	Pred
501.t	32	SDVQYAKNK	0.957	*Y*
501.t	41	LSESYKKFK	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	CEYEYQAQE	0.575	*Y*

Ribose-phosphate pyrophosphokinase, putative (gi56473321, XP_656141)

```
  1   MSVTTSLSAEQWQHDHFIVHTSKAGYLAQSISAVLGYPIIPIIHKKFSDG
 51   ECYYRIAIPERTSLVGRDVVVVSSIVDDQELLEVVRIGGELAELGTKRRI
101   FVIPYLMYSTMERAVHPGEVVTCKSTVRMLCGIPSSGLGNLFMLMDLHTS
151   GIIHYFEGTVQAMELYAESTLEQAIAENIDFSEPVIFGSADLGRPLWVET
201   FANHFGVGIAFIRKSRSFEDTHVIGEPIGDVKGKHVVIYDDLTRSAGSLI
251   KACNAYLDNGAIKVTAVVTHLSLIDDDVVQKVIDSRIDRVIATNSNVRSQ
301   MDAVKNSKKFIICDISPVFAKQIRACLNNNND
```

Phosphorylation sites predicted: Tyr: 3

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

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

```
  1   MQPPVANFCLWNLQPIQGSWMGAACIYQMPPSVRNTWWFPLLNTIPLDQY
 51   TRIYQWFMGVDRDRSGTLEINELMMGQFPGGIRLSPQTALRMMRIFDTDF
101   NGHISFYEFMAMYKFMELAYNLFVMNDRNRSGTLEPHEILPALQQLGFYI
151   NQRTSLLLHRLFARGMAFCDLNCWIAICAFAAQTRSAYQMIFMNPYYGPM
201   KPFNPMEFGKFLDVVTSLLE
```

Phosphorylation sites predicted: Tyr: 2

Name	Pos	Context	Score	Pred
URE3-BP	50	PLDQYTRIY	0.804	*Y*
URE3-BP	197	MNPYYGPMK	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), α=β=γ (°)	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})