PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Analysis of the protein kinome of *Entamoeba histolytica*

K. Anamika,[1] A. Bhattacharya,[2] and N. Srinivasan[1]*

[1] Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

[2] School of Life Sciences and Information Technology, Jawaharlal Nehru University, New Delhi 110067, India

ABSTRACT

Protein kinases play important roles in almost all major signaling and regulatory pathways of eukaryotic organisms. Members in the family of protein kinases make up a substantial fraction of eukaryotic proteome. Analysis of the protein kinase repertoire (kinome) would help in the better understanding of the regulatory processes. In this article, we report the identification and analysis of the repertoire of protein kinases in the intracellular parasite Entamoeba histolytica. Using a combination of various sensitive sequence search methods and manual analysis, we have identified a set of 307 protein kinases in E. histolytica genome. We have classified these protein kinases into different subfamilies originally defined by Hanks and Hunter and studied these kinases further in the context of noncatalytic domains that are tethered to catalytic kinase domain. Compared to other eukaryotic organisms, protein kinases from E. histolytica vary in terms of their domain organization and displays features that may have a bearing in the unusual biology of this organism. Some of the parasitic kinases show high sequence similarity in the catalytic domain region with calmodulin/calcium dependent protein kinase subfamily. However, they are unlikely to act like typical calcium/calmodulin dependent kinases as they lack noncatalytic domains characteristic of such kinases in other organisms. Such kinases form the largest subfamily of kinases in E. histolytica. Interestingly, a PKA/PKG-like subfamily member is tethered to pleckstrin homology domain. Although potential cyclins and cyclin-dependent kinases could be identified in the genome the likely absence of other cell cycle proteins suggests unusual nature of cell cycle in E. histolytica. Some of the unusual features recognized in our analysis include the absence of MEK as a part of the Mitogen Activated Kinase signaling pathway and identification of transmembrane region containing Src kinase-like kinases. Sequences which could not be classified into known subfamilies of protein kinases have unusual domain architectures. Many such unclassified protein kinases are tethered to domains which are Cysteine-rich and to domains known to be involved in protein–protein interactions. Our kinome analysis of E. histolytica suggests that the organism possesses a complex protein phosphorylation network that involves many unusual kinases.

Proteins 2008; 71:995–1006.

Key words: Entamoeba histolytica; domain structures; protein evolution; protein kinases; protein phosphorylation; signal transduction.

INTRODUCTION

A draft sequence of the genome of the protozoan parasite *Entamoeba histolytica*, has been reported.[1] *E. histolytica* is an etiological agent of amoebiasis causing large scale morbidity and mortality throughout the world, particularly in developing countries.[2] A large fraction of infected individuals do not display any symptoms of invasive disease and remain asymptomatic. It is not yet clear as to what factor(s) or signal(s) are particularly involved in turning a commensal *E. histolytica* into a virulent cell capable of tissue invasion. Studies of the signaling proteins and pathways may help us to understand the processes that may have an influence on the invasive disease process.

Though there are a number of molecules that are thought to be involved in pathogenic behavior of *E. histolytica*, such as galactose-*N*-acetyl-galactosamine-binding lectin,[3] pore-forming proteins amoebapores[4] and a number of cysteine proteinases[5] detailed mechanisms involved in the different stages of target tissue invasion and killing have not been explored. Further, it is not entirely clear the steps used in the array of receptors present mainly on the cell surface that transduce signals intracellularly using a variety of signal transduction pathways. The diversity of these pathways helps organisms to interact with differ-

The Supplementary Material referred to in this article can be found online at http://www.interscience.wiley.com/jpages/0887-3585/suppmat
Grant sponsors: Council of Scientific and Industrial Research (Government of India), Department of Biotechnology (Government of India).
*Correspondence to: N. Srinivasan, Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, Karnataka, India.
E-mail: ns@mbu.iisc.ernet.in
Received 26 April 2007; Revised 17 August 2007; Accepted 21 August 2007
Published online 14 November 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21790

ent signals and to respond appropriately. It is likely that some of the signaling pathways may play an important role in host-parasite relationship.

A number of signaling pathways have been reported in *E. histolytica*. Binding of extracellular matrix components (ECM) to amoeba is thought to be involved in generating signals needed for invasion.[6] Signal transduction pathways involving calcium have also been identified.[7] The major component of the signaling system of *E. histolytica* consists of a large number of protein kinases. Presence of protein tyrosine kinases and multiple Ser/Thr kinases have been previously identified in the protist genomes (*Leishmania major*, *Trypanosoma brucei*, *Trypanosoma cruzi*).[8–10] A large number of kinases with transmembraneous region related to the Gal/GalNac lectin and demonstrated to have similarity to both serine/threonine and tyrosine kinases have been identified in *E. histolytica*.[11] These transmembrane kinases have been shown to express in the trophozoites and at least one family of such kinases may be involved in cellular proliferation and serum response.[11,12] Gal/GalNac lectin has been known to mediate activation of mitogen activated protein kinase (MAPK) pathway.[13] Roles of reactive oxygen species (ROS), and of MAPK in the Entamoeba-induced apoptosis of human neutrophils have been investigated.[14] MAPK (Erk2) with focal adhesion kinase controls cytoskeleton integrity which plays an important role in adhesion and thus invasion of the host.[15] It has been seen that increment in c-adenosyl mono phosphate (cAMP) and adenylate cyclase activity produce a striking reorganization of actin into structures that consequently facilitate adhesive, locomotive, and secretory activity.[16] Presence of tyrosine kinases associated with the β1-integrin-like molecule has been confirmed previously by immunoprecipitation assays and these are likely to be involved in cellular proliferation.[17,18] As pointed out before by one of us and coworkers, *E. histolytica* has a large number of novel calcium binding proteins or CaBPs.[19] Both CaBP1 and 2 are known to activate unknown kinases. Though these CaBPs display high level of sequence similarity (about 79% identity) the activated kinases are thought to be different.[20]

All these indicate that the repertoire of protein kinases in *E. histolytica* may be large and that these are important to understand the biology of these organisms. Here, we present detailed analysis of protein kinases identified in *E. histolytica* genome using the bioinformatics approaches[21–26] employed by us earlier in analyzing kinomes of several other organisms[27–32] and also using 3D structural characteristics of kinases[32] wherever appropriate. Hanks and Hunter classification of protein kinases and the analysis of their domain organization have also been carried out to get an insight into their biological roles.

MATERIALS AND METHODS

The complete set of predicted protein sequences from the ORFs of the *E. histolytica* genome which is originated from TIGR [http://www.tigr.org] has been obtained from NCBI [http://www.ncbi.nlm.nih.gov]. Though the initial analysis was carried out using the first assembly and primary annotation, the results presented here has been checked against the genomic data recently assembled. We have surveyed the genome, for Ser/Thr and Tyr kinases using sensitive sequence profile matching algorithms. Briefly, we have employed multiple sensitive sequence search and analysis methods PSI-BLAST,[21] MulPSSM[22–24] involving extensive use of RPS-BLAST,[25] and HMMer[26] which match Hidden Markov models (HMMs) to identify protein kinase catalytic domain and their cooccurring domains. The search protocols are identical to those used in our earlier surveys for protein kinases in other organisms.[27–32] The criteria used to associate a given protein kinase in a given family on the basis of its primary structure include the degree of sequence identity and the presence of signature amino acids that are characteristics of protein kinase subfamilies.[33] In addition, search procedures such as PSI-BLAST have been used to detect sequences homologous to the kinase catalytic domain using an E-value cutoff of 0.0001 which is decided on the basis of previous benchmarking study.[34]

Initially, PSI-BLAST could pick up a total of 401 protein kinase (PK)-like sequences. CD-HITS[35] program was used to eliminate redundant sequences which are indicated by 100% sequence identity. After using CD-HIT and ignoring truncated sequences (which are less than 172 amino acids long), we arrived at a set of 338 PK-like sequences. Hits lacking significant sequence similarity with the query have been further examined manually and assessed by various fold prediction methods[36,37] for the compatibility of the query sequence with the 3D fold of Ser/Thr and Tyr kinases. Fold prediction methods did not associate protein kinase fold for seven of the sequences which were suspected to be protein kinases and so we did not consider these seven sequences as kinases.

The final data set of 331 putative PK-like sequences has been obtained from the compilation of hits obtained during various procedures. Out of these 331 PK-like sequences, 24 sequences lack aspartate in the catalytic loop. Kinase-like sequence which lack catalytic aspartate are unlikely to function as kinases and these are referred as kinase homology domain (KHD).[38,39] The 307 putative PK of *E. histolytica* with conserved catalytic aspartate, and hence expected to be functional, are referred to as potential protein kinases (PPKs) in the subsequent sections. The numbers of PK-like sequences and the final PPK dataset obtained at different steps of the analysis are shown in Figure 1.

Classification of *E. histolytica* putative protein kinases into subfamilies

Reverse PSI-BLAST (RPS-BLAST) was used to search each of the 401 PK-like sequences identified in the *E. histolytica* genome as a query against the database containing 55 position specific matrices (PSSMs) created for the various subgroups of protein kinases in each of the sub-

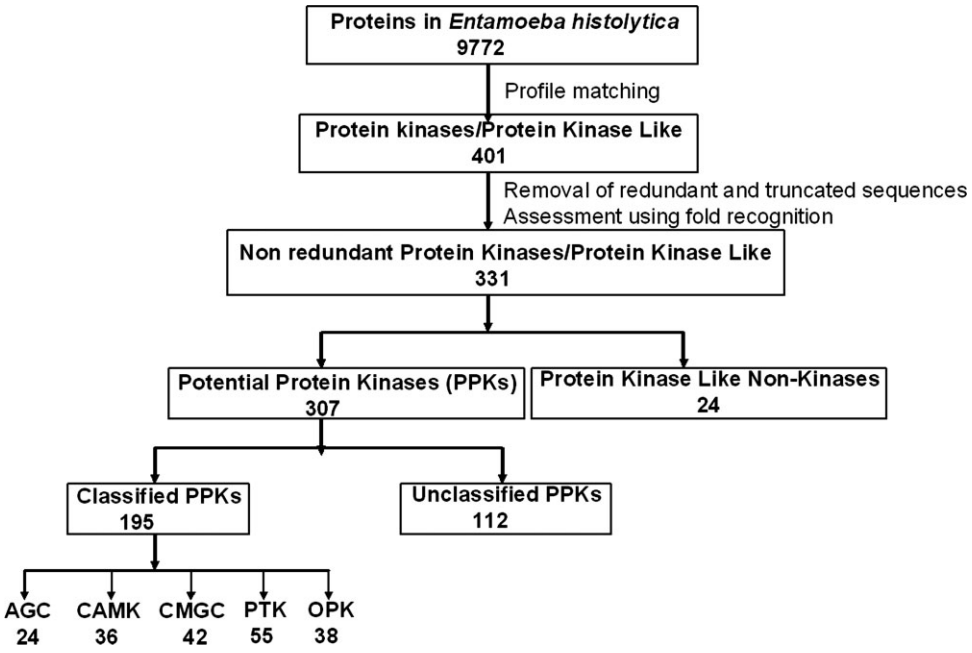

Figure 1

Distribution of E. histolytica *kinases and related sequences into different Hanks and Hunter groups. Various abbreviations followed in the diagram: AGC, protein kinase A, protein kinase G, protein kinase C; CAMK, Ca^{2+}/calmodulin dependent protein kinase; CMGC, cyclin dependent protein kinase, Mitogen activated protein kinase, glycogen synthase kinase, casein kinase-2; PTK, protein tyrosine kinase; OPK, other protein kinases.*

families. HMMer run against subfamily profile was mainly used and considered for subfamily assignment. A query kinase sequence was associated to its nearest subfamily based on the extent of sequence similarity. Sequences with greater than 30% identity with at least one of the members of a kinase subfamily have been considered as members of the subfamily concerned.

CLUSTALW[40] has been used to align the catalytic domains of 195 protein kinases of *E. histolytica* that were associated to known subfamilies. MEGA version 4[41] was used to generate the dendrogram showing the various subfamilies of protein kinases.

Domain assignments have been made for protein kinase catalytic domain containing gene products using the HMMer and RPS_BLAST search methods by querying each of the kinase domain containing proteins against the 7973 protein family HMMs available in the Pfam database[42] and MulPSSM profiles[23,24] of families in Pfam database, respectively. Trans-membrane segments were detected using TMHMM.[43]

RESULTS AND DISCUSSION

Distribution of protein kinases in the *E. histolytica* genome

Total number of functional protein kinases encoded in the genomes of some of the protozoan parasites[9,32] is shown in Figure 2. The genome of *E. histolytica* encodes 307 putative protein kinases (PPKs) which is about three times the number of kinases encoded in the genome of

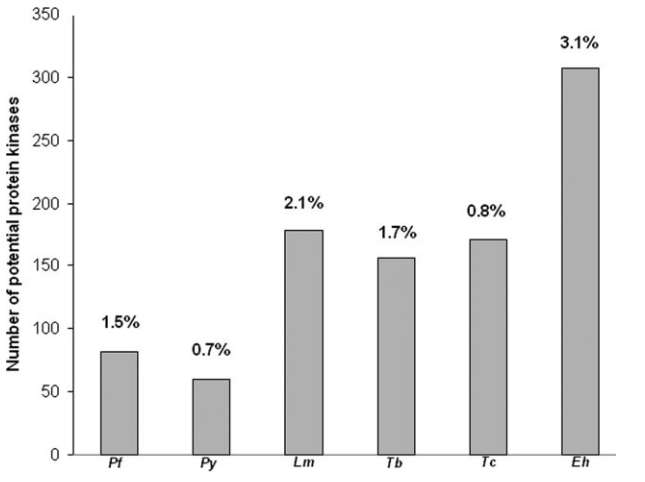

Figure 2

Bar diagram showing comparison of the potentially functional PK complement encoded in the genomes of protozoa. The percentage of protein kinase genes in the proteome complement is provided against every bar. Species abbreviations followed in the diagram: Pf, Plasmodium falciparum; Py, Plasmodium yoelii yoelii; Lm, Leishmania major; Tb, Trypanosoma brucei; Tc, Trypanosoma cruzi; Eh, Entamoeba histolytica.

Table I
Distribution of E. histolytica Kinases and Related Proteins into Various Hanks and Hunter Groups and Subfamilies

Protein kinase like sequences	331
Protein kinase like nonkinase (or KHD)	24
Potential protein kinases (PPKs)	307
Protein kinase subfamilies	
AGC group	24
CAMK group	36
CDK	9
CLK	22
MAPK	2
GSK	4
CK2	3
CMGC_other	2
Casein kinase 1	6
MEKK	4
MLK	2
NimA	1
PAK	17
Polo	2
PTK	55
Raf	6
Unclassified protein kinases	112

the malaria parasite *Plasmodium falciparum*.[29] The distribution of PPKs into various Hanks and Hunter subfamilies is summarized in Table I. The *E. histolytica* kinome, which is less than half the size of the human kinome,[27,44] differs in numerous ways from kinases in the mammalian host. These kinases are listed in the Supplementary Table I and are also included in our KinG database (http://hodgkin.mbu.iisc.ernet.in/~king/)[30] along with subfamily assignments and other domains that are tethered to kinase catalytic domains. Among the 307 PPKs that have been analyzed, 296 kinase domains contain at least one glycine in the ATP anchoring glycine-rich loop motif GXGXXG present in subdomain I of kinase catalytic domain. Although in 11 sequences not even one glycyl residue is present in the "glycine-rich" loop it is not conclusive that ATP cannot be accommodated in these PPKs[45] as many kinases from other organisms with no glycyl residue in the "Gly-rich" region are shown to bind ATP. The Lys-Glu salt bridge localizing the gamma phosphate in ATP is not well conserved in 40 PPKs. The variations thus observed in the key nucleotide binding residues suggest a high divergence of the *E. histolytica* kinases compared to the other eukaryotic kinases known so far. This is an encouraging observation from the point of view of designing inhibitors of *E. histolytica* kinases as most successful inhibitors of kinases target ATP binding region. The phosphorylation of the activation segment is required for activation in most protein kinases that contain an arginine preceding the catalytic aspartate. This activation segment, which is located between the two lobes of the kinase structure, adopts characteristic active conformation as a result of phosphorylation in this segment in many kinases. These kinases are also referred to as "RD" kinases.[46,33] The *E. histolytica* genome encodes a total of 228 "RD" kinases suggesting a potential requirement for phosphorylation in their activation loop for regulation. Interestingly, 50 of these 228 "RD" kinases do not have S/T or Y residue in the activation segment suggesting an unusual mechanism of their activation. The alignment of catalytic domain of each kinase group in *E. histolytica* is provided in the Supplementary Table II.

Among the putative protein kinases in the dataset, 140 and 55 PPKs are likely to be Ser/Thr and tyrosine kinases, respectively. Out of 140 putative Ser/Thr protein kinases, six are putative receptor Ser/Thr kinases as membrane spanning regions are predicted in these sequences. Similarly out of 55 putative Tyr kinases, 43 PPKs are putative receptor tyrosine kinases.

Among the analyzed PK-like sequences, 24 sequences are considered nonfunctional as protein kinase (KHD) as they lack crucial residue (aspartate) in the catalytic loop. There are two KHDs which are tethered with the potentially functional kinase domain (EAL52204 and EAL45554). The roles of such PK-like sequences in *E. histolytica* are currently unclear. The presence of KHDs in other unicellular parasites suggests an important role, possibly in the regulation of phosphorylation network as proposed for inactive mammalian protein kinases.[47] In the case of human, most protein kinases (about 80% of

Table II
List of E. histolytica Protein Kinases with Unusual Domain Combinations

Protein kinase subfamily assignment to *E. histolytica* kinase	Unusual domain(s) tethered N-terminal to kinase domain	Unusual domain(s) tethered C-terminal to kinase domain
PKA/PKG like protein kinase	PH (Pleckstrin homology) domain	
CAMK like protein kinase		Endonuclease_5 domain
Raf	VSP (variant surface protein) domain, transmembrane domain	
PAK	Transmembrane domain, PH (Pleckstrin homology) domain	Calpain III domain
Src kinase	Transmembrane domain Kelch domain repeats	
Abl tyrosine kinase	Kelch domain repeats	

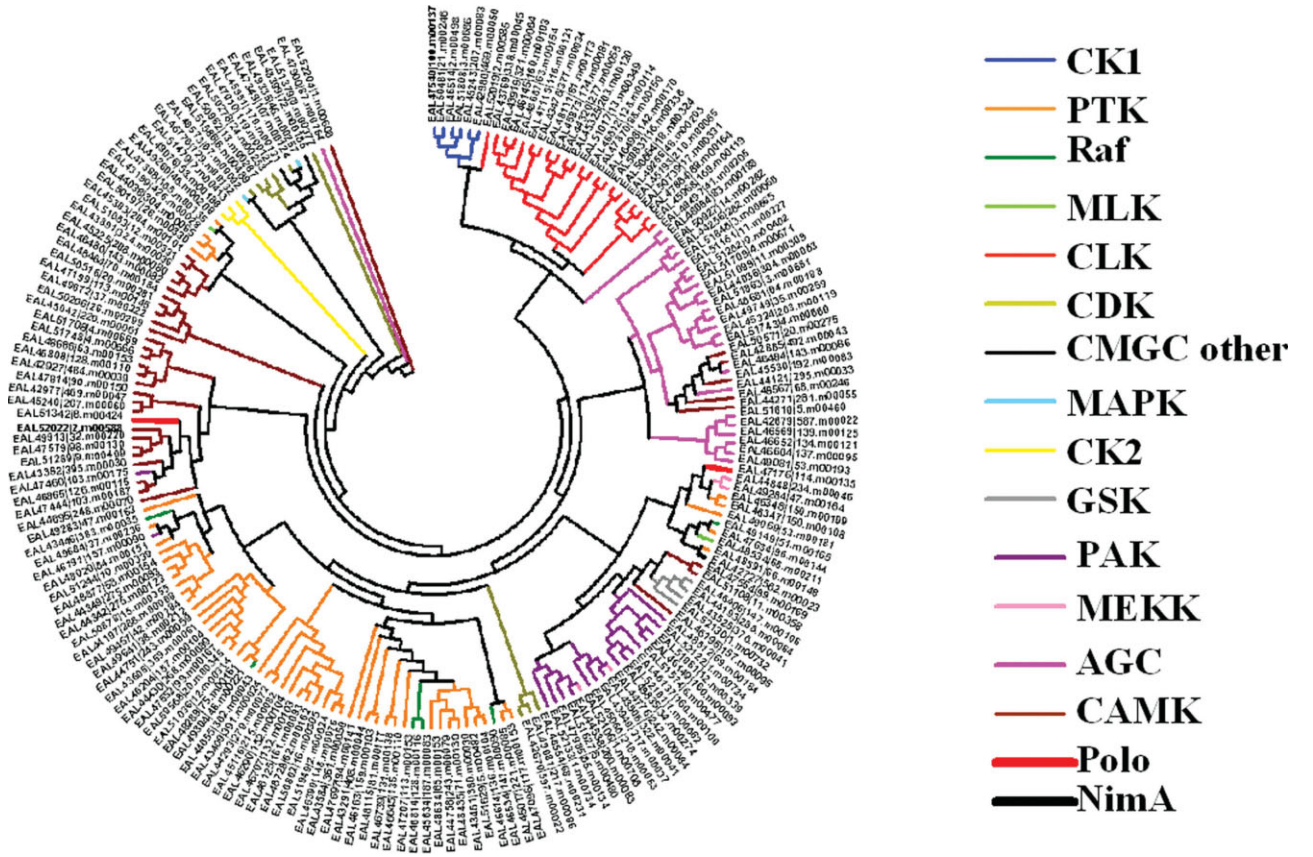

Figure 3

Dendrogram showing classified members of protein kinases. Various protein kinase subfamilies are represented in different colors. Various abbreviations followed in the dendrogram: CK1, casein kinase-1; PTK, protein tyrosine kinase; MLK, mixed lineage kinase; CDK, cyclin dependent kinase; CLK, CDK like kinase; MAPK, mitogen activated kinase; GSK, glycogen synthase kinase; CK2, casein kinase-2; CMGC other, CDK, MAPK, GSK, CK2 other; PAK, p-21 activated protein kinase; MEKK, MAP kinase kinase kinase; AGC, protein kinase A, protein kinase G, protein kinase C; CAMK, calcium/calmodulin dependent protein kinase.

the total kinome) contain at least one recognizable domain other than catalytic kinase domain[27]; but many *E. histolytica* kinases generally lack additional domains tethered to kinase domain.

Classification of the *E. histolytica* protein kinases

Based on the amino acid sequence of the catalytic domain, *E. histolytica* kinases could be classified into various subfamilies (see Fig. 3) as described by Hanks and coworkers.[48] One hundred ninety-five PPKs could be placed in one of the known subfamilies. Remaining 112 PPKs are unclassified protein kinases and are discussed separately. The distribution and the diversity of the various classified subfamilies are discussed below. A list of unusual domain organizations of various PPKs identified in *E. histolytica* genome is presented in Table II.

Ser/Thr kinase subfamilies

Close homologues of calcium/calmodulin dependent protein kinases

Calcium-dependent protein kinases (CDPK) are activated by the direct binding of calcium to regulatory domains, which contain four EF-hand calcium binding domains and are similar to calmodulin (CaM). CDPKs have both catalytic and regulatory domains joined by an autoinhibitory domain. Calmodulin-dependent protein kinases (CAMKs) are activated by the binding of calmodulin to calmodulin-binding segment preceding the kinase catalytic region in the primary structure. PPKs with catalytic regions showing high sequence similarity to CAMKs are the most represented subfamily of PKs in *E. histolytica*. While bioinformatics analysis on *E. histolytica* genome has enabled recognition of a number of putative calcium binding proteins, at least couple of them has

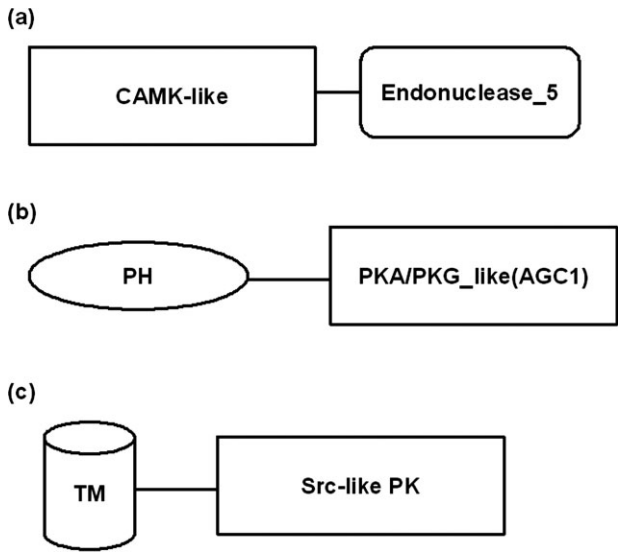

Figure 4

*Unusual domain combinations in some of the E. histolytica kinases. (**a**) CAMK-like subfamily member with an endonuclease domain. (**b**) PKC like PKA homologue. (**c**) Homologue of Src-like kinase with transmembrane spanning domain.*

been experimentally characterized.[19] *E. histolytica* genome contains 36 PPK with more than 30% sequence identity in their kinase domain with CAMK. But these sequences do not contain calmodulin binding region and hence are unlikely to function as classical CAMKs. This suggests an evolutionary paradigm in the highly conserved signaling pathway of eukaryotes. This subfamily of PPKs occurs with various domain combinations that are observed in other kinomes. Four of these PPKs (EAL52022, EAL49672, EAL47814, and EAL47579) have fork head associated (FHA) domain preceding the protein kinase domain in the sequence. FHA is a phosphopeptide recognition domain ~80–100 amino acids long, found in many regulatory proteins. One of these PPKs (EAL50571) has Endonuclease_5 domain succeeding the PK domain [Fig. 4(a)]. This protein is likely to be specific for single stranded DNA or duplex DNA that contains uracil or that is damaged by a variety of agents.[49] This domain combination is unusual for protein kinases in general. The kinase domain of six of these PPKs (EAL50516, EAL45303, EAL45225, EAL44271, EAL51053, and EAL51708) show more than 30% sequence identity in their kinase domains with CDPK-I subgroup of kinases from *Plasmodium falciparum*. However, no EF-hand motifs have been found tethered to these kinase catalytic domains suggesting that no specific mechanism can be assigned at present on their mode of regulation. One of the PPKs (EAL51108) has two tandem kinase domains and both show high sequence similarity in the catalytic domain to CAMK group. Earlier bioinformatics analysis[19] on *E. histolytica* genome and in the present analysis, we recognize a number of putative CaM-like multi-EF hand proteins (EAL46322, EAL46978, EAL51814, EAL50371, EAL50341, EAL45614, EAL46660, EAL49262, and EAL43751) which show sequence identity greater than 17% among themselves and sequence identity of about 30% or more with calmodulins of higher eukaryotes. It is tempting to speculate that some of the CDPKs may be regulated by these CaM-like multi-EF hand proteins. Indeed, a putative CaM is known to evoke electron-dense granules (EDGs) secretion which has a role in host cell invasion,[50] although involvement of a protein kinase in this process is not yet reported.

Close homologues of cyclin dependent kinases

Cyclin dependent kinases (CDKs) which control the progression of the eukaryotic cell cycle in conjunction with various regulatory proteins can be categorized based on the extent of similarity in the cyclin binding regions and the inhibitory phosphorylation sites into putative CDKs and CDK-like sequences (CLKs). There are nine close homologues of CDKs (EAL48554, EAL50962, EAL45081, EAL51586, EAL47010, EAL51379, EAL47349, EAL50278, and EAL42670) identified in *E. histolytica*. Threonine 160 of human CDK located in the activation segment serves as a permissive phosphorylation site and it is conserved in eight putative CDKs of *E. histolytica* and is substituted by cysteine in one of the potential CDKs (EAL47349). The inhibitory phosphorylation sites, threonine-14 and tyrosine-15 are conserved in six and seven potential CDKs, respectively. In one of the CDKs (EAL51379), phenylalanine is present at the Tyr-15 position and in another CDK (EAL50278) no suitable residue was found that could be aligned with Tyr-15. Cyclin-binding motif PSTAIRE is conserved in eight CDKs with minimal residue variation. Absence of cyclin binding motif in one of the CDKs (EAL50278) leaves an open question about its regulation by cyclin.

The CDK-like kinase (CLK) family is relatively large in this organism with 22 members closely related to CLK. CLKs are implicated in cell growth and differentiation. A large number of CLKs might add complexity to the cell cycle regulating system of the parasite. It has been previously seen that the N-terminal region of CLKs is playing a regulatory role.[51] There is a putative nuclear localization sequence found prior to the kinase catalytic domain in many close homologues of CLKs.

CDKs require an activating cyclin partner for activity, and analysis of the *E. histolytica* genome reveals eight putative cyclins (EAL43280, EAL46881, EAL46402, EAL43526, EAL47872, EAL50559, EAL49793, and EAL44871) which in conjunction with CDKs probably carry out cell-cycle regulation in eukaryotes. Sequence identity between these cyclins ranges from 17 to 94%. The cyclin A1 of human is not very closely related to

amoeba cyclins (about 31% sequence identity). Other regulatory subunits of CDKs have not been detected. This analysis supports the experimental findings suggesting unusual cell cycle organization and cellular duplication in *E. histolytica*.[52–54]

The homologues of the CMGC subfamily of eukaryotic kinases, putative glycogen synthase kinase (GSK), and putative casein kinase-2 (CK2) have also been identified. *E. histolytica* encodes four potential GSKs (EAL44193, EAL52130, EAL46406, and EAL43525). Sequence similarity between *E. histolytica* GSKs and human GSK3 is ~46%.

Three putative casein kinase-2 (CK2) (EAL46776, EAL49076, and EAL51479) proteins are encoded in the amoeba genome. The CK2 is known to be active as both monomer and as tertrameric holoenzymes. The functional holoenzyme comprises of two CK2-β and two CK2-α (catalytic) subunits.[55] Five homologues (EAL43388, EAL51513, EAL43183, EAL50781, and EAL44722) of the regulatory subunit (CK2-β) have also been identified that share greater than 35% sequence identity with other eukaryotic CK2-β subunits. Six close homologues of CK1 (EAL47540, EAL46514, EAL51808, EAL45243, EAL50481, and EAL42980) have also been identified in our analysis.

Close homologues of MAP kinases of E. histolytica

The core components of MAPK signaling pathway are MAP kinase kinase kinase (MEKK), MAP kinase kinase (MEK), and MAP kinase (MAPK). The dual phosphorylation in the activation segment of MAPK on Threonine and Tyrosine (TXY motif) by MEK leads to activation of MAPK. Of the two MAPKs identified in the *E. histolytica* genome sharing a sequence identity of 31% in their catalytic domain, the TXY motif is completely conserved in one of them as TDY (EAL48573), and is altered to TIQ in the other (EAL49335). This close homologue of MAPK is 340 amino acid residues long. One of the MAPKs (EAL48573) was previously identified. Phylogenetic study has shown that this MAPK belongs to the extracellular signal-regulated kinase (ERK) subfamily.[56] The close homologues of MEK were however, not identified previously in the amoeba genome. This provides a clear example that MAPK pathway may not require MEK since no member of the latter has yet been identified in this organism. However, four close homologues of MEKK (EAL44848, EAL49284, EAL44776, and EAL52106) and six sequences associated with Raf-subfamily (EAL47207, EAL46814, EAL43446, EAL49149, EAL46534, and EAL49304) of kinases have been identified. Four homologues of the Raf subfamily contain transmembrane regions. One Raf subfamily PPK (EAL49149) has six ankyrin repeats situated preceding the kinase domain in the sequence. Two putative protein kinases (EAL48534 and EAL47388) with close similarity to the mixed lineage kinases (MLK), have been identified. These eight kinases are likely to serve as upstream proteins of the MEK. One of the MLKs (EAL47388) has Leucine Rich Repeats (LRR) which is likely to mediate multimerization.

The CMGC family includes CDK, MAPK, GSK3, and dual specificity CLK which are relatively abundant in this protozoan. This fact can possibly be explained by the requirement of members of CMGC protein kinase to control the complex life cycle and cell cycle of the parasite by correct replication and segregation of putative organelles.

AGC subfamily homologues in E. histolytica

The binding of second messengers like cAMP, cGMP, and diacyl glycerol (DAG) regulates the protein kinase members of the AGC family. The amoeba genome encodes 24 PPKs that are closely related to the catalytic domain of AGC kinases and thus AGC subfamily is second most represented subfamily of kinases in *E. histolytica*. The parasite encodes eight close homologues of PKA/PKG-like (AGC1) of which four PPKs (EAL46604, EAL49081, EAL51099, and EAL44036) have PH domain preceding the kinase domain in the primary structure. As this domain combination is common for PKCs but unusual for PKA/PKG-like it appears that these kinases share the properties of PKC and PKA/PKG-like molecules [Fig. 4(b)]. Another three close homologues of PKA/PKG (EAL45324, EAL51743, and EAL48681) have protein kinase C terminal domain succeeding the kinase domain in the sequence. No PKA regulatory subunit has been identified in the genome suggesting that PKC-PKA/PKG-like hybrid sequences are more likely to function similar to PKC rather than PKA. PKC activity has been identified in various strains of *E. histolytica* and relocation of the amoebic PKC activity from the cytosol to the membrane was observed in trophozoites actively phagocytizing bacteria.[57] Fast release of mucins by a PKC-dependent mechanism was also observed.[58]

Other *E. histolytica* protein kinases

p21 activated kinases (PAK) are known to mediate cytoskeleton actin assembly and regulates cell motility. PAK proteins are categorized further into PAK1, PAK2, PAK3, and PAK4. These proteins serve as targets for the small GTP binding proteins Cdc-42 and Rac and have been implicated in a wide range of biological activities. PAK 1 regulates cell motility and morphology. It was previously shown that *E. histolytica* PAK is involved in its migration and phagocytosis.[59] PAK domain is usually tethered with PAK binding domain (PBD, also known as CRIB for Cdc42/Rac 1 interactive binding) in its N-terminal end. This binds Cdc42 and/or Rho-like small GTPase and stimulates the kinase activity by a mecha-

nism involving autophosphorylation.[60] We have identified 17 close homologues of PAK in *E. histolytica* genome. While four such PPKs (EAL43940, EAL46137, EAL52161, and EAL52122) have PH domain N -terminal to kinase domain, six homologues have PBD (EAL51574, EAL51061, EAL46137, EAL52161, EAL52122, and EAL46149) domain preceding kinase domain. Interestingly, three homologues (EAL46137, EAL52161, and EAL52122) have both PH domain and P21-Rho-binding domain (PBD) preceding kinase domain. A PAK homologue (EAL43906) has Calpain III domain succeeding kinase domain. Calpain III domain is believed to participate in intracellular signaling pathways mediated by calcium ions.[61] There are two PAK-like kinases (EAL43940 and EAL46191) with transmembrane region. While one of these homologues (EAL46191) has VSP domain in the extracellular region, the other (EAL47460) has FHA domain preceding the kinase domain. The NIMA family of protein kinases plays an important role in the eukaryotic cell cycle division process involving chromatin condensation and formation of mitotic spindles. A close homologue (EAL42727) of NimA protein kinase has been identified in *E. histolytica* genome. The "FXXT" motif of NIMA family kinase is partially conserved as "IXXT" where threonine is phosphorylated by various kinases.

Polo protein kinases play important role in mitosis and meiosis. In our analysis, two putative polo kinases (EAL51342 and EAL47176) have been identified. Only one of them (EAL47176) has Polo box domain C-terminal to the kinase domain.

Close homologues of protein tyrosine kinases

There are two types of tyrosine kinases, one with transmembrane region (receptor kinases) and other one is cytoplasmic in nature. Receptor tyrosine kinases (RTK) comprise of transmembrane region sandwiched by extracellular and intracellular regions. A common mechanism of activation involves dimerization of the receptor subsequent to ligand binding. Representative of the tyrosine kinase in humans include receptor protein kinases such as insulin receptor and cytosolic kinases such as Src. There are 55 protein kinases from the parasite that are placed in the protein tyrosine kinase (PTK) subfamily. Among 55 such PPKs, 43 sequences are putative receptor tyrosine kinases having single pass transmembrane region. Occurrence of a good number of putative receptor tyrosine kinases is unlike other protozoa parasites, such as trypanosomatids which have very few receptor tyrosine kinases.[62] While no cytoplasmic tyrosine kinases have been reported in trypanosomids,[62] there are 11 ORFs which potentially encode for Src-like protein kinases. The Src kinase has a structure composed of two peptide binding domains (SH2 and SH3), in addition to a catalytic kinase domain. SH3 domain recognizes regions with polyproline helical conformation. The N-terminal region is myristylated, so the protein is associated with the cell membrane. Surprisingly, neither HMMer nor RPS-BLAST search against Pfam profiles identified SH2 or SH3 domain in spite of large size (greater than 650 amino acids) of the gene products although the catalytic kinase region shows a good similarity to the catalytic region of Src kinases.

Src is known to be a cytoplasmic protein tyrosine kinase but our analysis has shown that seven close homologues of Src proteins have transmembrane region [Fig. 4(c)]. This indicates that Src kinase-like kinase of *E. histolytica* is highly diverged from its closest homologues in higher eukaryotes. One of the cytoplasmic Src (EAL46348) has four Kelch_1 domains preceding the kinase domain. This domain family has diverse functionality, but, broadly involved in regulatory and cytoskeletal function.[63,64] Hence, these are unlikely to be functionally analogous to src kinases in spite of close sequence similarity of the kinase domain with src kinases.

Two putative Jak kinases (EAL51036 and EAL49260) have also been identified in our analysis. One Jak (EAL49260) has seven LRR (leucine rich repeats) domains preceding the kinase catalytic domain. LRR appears to provide a structural framework for the formation of protein–protein interaction. The domain which is usually tethered with the protein tyrosine kinase catalytic domain are Giardia variant surface protein (VSP), Ankyrin repeats, Kelch 1 domain, Furine like domain, and TIG (domain found in some cell surface receptors). VSP is a major surface antigen first identified in the intestinal protozoan parasite *Giardia lamblia*. The ankyrin repeat is one of the most common protein–protein interaction motifs mainly found in eukaryotes. Overall, this suggests that many pathways characterized in other eukaryotic systems also operate in this parasite.

Unclassified protein kinases

Several open reading frames in the *E. histolytica* genome encode polypeptides with a clearly recognizable kinase catalytic domain but could not be ascribed to specific known protein kinase subfamilies on the basis of the criteria used in the present work. There are 112 such PPKs which could not be placed into any one of the Hanks and Hunter subfamilies. Sequence identity of the kinase domain of unclassified kinases varies from 13 to 92%. Most of the unclassified kinases are "RD" kinases. Thirty-seven out of 112 unclassified kinases have single span transmembrane regions. Dendrogram generated on the basis of the amino acid sequences of the kinase domain regions of these unclassified kinases shows many clusters (see Fig. 5). This means that some of the unclassified kinases form close paralogues and hence may be performing important functions.

Unclassified protein kinases have different domain architecture from classified protein kinase members.

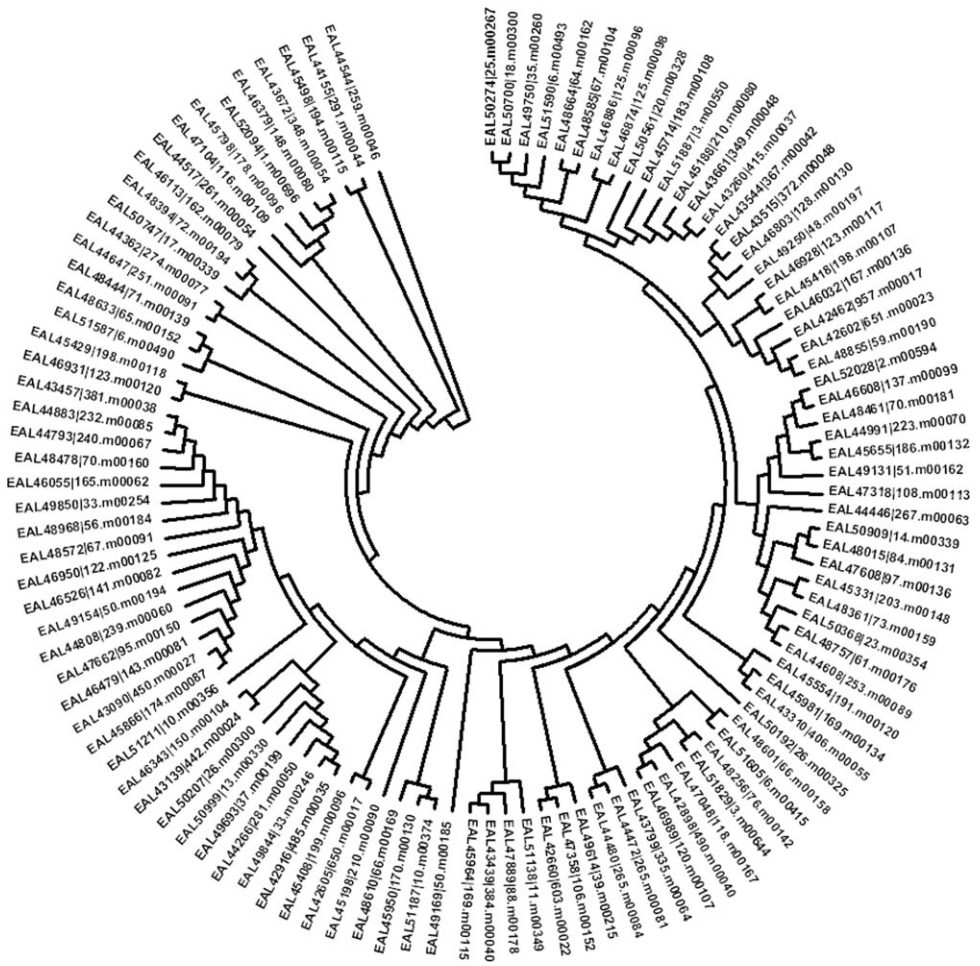

Figure 5

Clustering of unclassified protein kinases of E. histolytica genome.

Some of the common domains tethered to the kinase catalytic domains are SH2 domain, Laminin EGF domain, GCC2_GCC3, Tetratricopeptide repeats (TPR), C triplex, Rho-GEF, Paralytic/Growth-Blocking Peptide/PSP peptide (GBP PSP) domains, Leucine Rich Repeats, and Ankyrin Repeats. Most of these protein domains are involved in protein–protein interactions and are cysteine rich. Laminin EGFs mediate cell adhesion, growth migration, and differentiation.[65] GCC2_GCC3 domain contains five conserved cysteine residues that are likely to be involved in disulphide bond formation. They are found in a wide variety of extracellular proteins. The TPR is a structural motif present in a wide range of proteins and mediates protein–protein interactions and the assembly of multiplex proteins.[66] C triplex is also rich in cysteine residues and its function is currently unknown. Rho-GEF is involved in protein–protein interaction and hence regulates a diverse array of cellular processes. GBP-PSP peptides have been first detected in insects and are known to halt metamorphosis.[67,68] Two putative protein kinases, that are likely to be functionally equivalent of Src kinases (EAL50909 and EAL49250), have tethered SH2 domain C-terminal to the kinase domain. However the catalytic region is not highly similar, in terms of amino acid sequence, to the catalytic region of src kinases. There are other (EAL48757 and EAL44608) PPKs that have Ribonuclease 2–5A domains C-terminal to kinase domain. This arrangement is found in a group of endoribonucleases, involved in cleaving introns in higher eukaryotes.[69] One of the unclassified PPKs (EAL43139) has a transmembrane region and nine antifreeze protein (AFP) domains following the kinase domain in the sequence. This family of extracellular proteins is involved in stopping the formation of ice crystals in low temperature and hence permit organism to survive in adverse conditions. All the 112 PPKs were further queried against Non-Redundant Database [NRDB available at the National Center for Biotechnology Information (NCBI)] to find

out their closest homologues. Highest similarity of almost all unclassified kinases is either from plant or from other protozoa with 30% or higher sequence identity.

CONCLUSIONS

The analysis presented here shows that *E. histolytica* possesses a substantial complement of protein kinases accounting for about 3% of the total proteome size which is slightly more than the size of kinome of most other eukaryotes. Presence of a large number of protein kinases indicates that protein phosphorylation is a key mechanism of regulation of this parasite. The CAMK-like subfamily which traditionally includes Ca^{2+}/calmodulin regulated kinases and AMP dependent kinase, is the most represented subfamily of protein kinases in the parasite. But the noncatalytic domain combinations of CAMK-like kinases are not consistent with the functions of CAMKs and hence these kinases are radically different from classical CAMKs. The second most represented protein kinase group is the AGC which requires secondary messenger for their activation and action. These appear to be more similar to PKC than PKA in terms of domain architecture. PKC subfamily members have not been picked but close homologues of PKA with PH domain may function analogous to PKC that are activated by phospholipids. The MAP kinase pathway appears to deviate from the canonical cascades, with absence of MEK. Presence of close homologues of CDKs without their regulatory proteins, except cyclins, leaves ample room for divergence of cell cycle regulation in the parasite. PAK subfamily members are also abundant in the parasite and they are found with various domain combinations tethered with catalytic domain. Close homologues of Src Tyrosine kinases exist in *E. histolytica* without SH3 and SH2 tethered to kinase domain and this peculiar feature leaves an open question about their regulation. The unclassified PKs identified in the analysis share high sequence identity in their catalytic region with plant kinases, suggesting a close evolutionary link between the plant and the parasite kinases. The PKs of *E. histolytica* represent a highly divergent class of PKs that are distinct in various features from other eukaryotic kinases known so far. For example, absence of typical glycine-rich motif at the ATP binding site in some of the identified kinases could be due to the potential use of pyrophosphate instead of ATP. Some of the recent data from one of our laboratories suggests that some of *E. histolytica* kinases are functional and not relics of past evolution.[12] We anticipate that the biological roles and regulatory process of many of these novel protein kinases identified with variant modular organization will be investigated in the future.

REFERENCES

1. Loftus B, Anderson I, Davies R, Alsmark UC, 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 R, 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. The genome of the protist parasite *Entamoeba histolytica*. Nature 2005;433:865–868.
2. Walsh JA. Amebiasis in the world. Arch Invest Med (Mex) 1986;17: 385–389.
3. Frederick JR, Petri WA, Jr. Roles for the galactose-/*N*-acetylgalactosamine-binding lectin of Entamoeba in parasite virulence and differentiation. Glycobiology 2005;15:53R–59R.
4. Leippe M, Ebel S, Schoenberger OL, Horstmann RD, Muller-Eberhard HJ. Pore-forming peptide of pathogenic *Entamoeba histolytica*. Proc Natl Acad Sci USA 1991;88:7659–7663.
5. Stanley SL, Jr, Zhang T, Rubin D, Li E. Role of the *Entamoeba histolytica* cysteine proteinase in amebic liver abscess formation in severe combined immunodeficient mice. Infect Immun 1995;63:1587–1590.
6. Meza I. Extracellular matrix-induced signaling in *Entamoeba histolytica*: its role in invasiveness. Parasitol Today 2000;16:23–28.
7. Yadava N, Chandok MR, Prasad J, Bhattacharya S, Sopory SK, Bhattacharya A. Characterization of EhCaBP, a calcium-binding protein of *Entamoeba histolytica* and its binding proteins. Mol Biochem Parasitol 1997;84:69–82.
8. Parsons M, Valentine M, Carter V. Protein kinases in divergent eukaryotes: identification of protein kinase activities regulated during trypanosome development. Proc Natl Acad Sci USA 1993; 90:2656–2660.
9. Parsons M, Worthey EA, Ward PN, Mottran JC. Comparative analysis of the kinomes of three pathogenic trypanosomatids: *Leishmania major*, *Trypanosoma brucei*, *Trypanosoma cruzi*. BMC Genomics 2005;6:127.
10. Que X, Samuelson J, Reed S. Molecular cloning of a rac family protein kinase and identification of a serine/threonine protein kinase gene family of *Entamoeba histolytica*. 1993;60:161–170.
11. Beck DL, Boettner DR, Dragulev B, Ready K, Nozaki T, Petri WA, Jr. Identification and gene expression analysis of a large family of transmembrane kinases related to the Gal/GalNAc lectin in *Entamoeba histolytica*. Eukaryot Cell 2005;4:722–732.
12. Mehra A, Fredrick J, Petri WA, Jr, Bhattacharya S, Bhattacharya A. Expression and function of a family of transmembrane kinases from the protozoan parasite *Entamoeba histolytica*. Infect Immun 2006;74:5341–5351.
13. Rawal S, Majumdar S, Vohra H. Activation of MAPK kinase pathway by Gal/GalNAc adherence lectin of *E. histolytica*: gateway to host response. Mol Cell Biochem 2005;268:93–101.
14. Sim S, Yong TS, Park SJ, Im KI, Kong Y, Ryu JS, Min DY, Shin MH. NADPH oxidase-derived reactive oxygen species-mediated activation of ERK1/2 is required for apoptosis of human neutrophils induced by *Entamoeba histolytica*. J Immunol 2005;174:4279–4288.
15. Robledo T, Cruz-Vera J, Reyes M, de la Garza M, Ortega A, Perez Salazar E. *Entamoeba histolytica*: focal adhesion kinase and Erk phosphorylation are altered in the cytoskeleton mutant BG-3. Parasitol Res 2005;95:231–235.
16. Franco E, Manning-Cela R, Meza I. Signal transduction in *Entamoeba histolytica* induced by interaction with fibronectin: presence and activation of phosphokinase A and its possible relation to invasiveness. Arch Med Res 2002;33:389–397.
17. Hernandez-Ramirez VI, Anaya-Ruiz M, Rios A, Talamas-Rohana P. *Entamoeba histolytica*: tyrosine kinase activity induced by fibronectin through the β1-integrin-like molecule. Exp Parasitol 2000;95:85–95.
18. Chaudhuri S, Choudhury N, Raha S. Growth stimulation by serum in *Entamoeba histolytica* is associated with protein tyrosine dephosphorylation. FEMS Microbiol Lett 1999;178:241–249.

19. Bhattacharya A, Padhan N, Jain R, Bhattacharya S. Calcium-binding proteins of *Entamoeba histolytica*. Arch Med Res 2006;37:221–225.
20. Chakrabarty P, Sethi DK, Padhan N, Kaur KJ, Salunke DM, Bhattacharya S, Bhattacharya A. Identification and characterization of EhCaBP2. A second member of the calcium-binding protein family of the protozoan parasite *Entamoeba histolytica*. J Biol Chem 2004;279:12898–12908.
21. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
22. Anand B, Gowri VS, Srinivasan N. Use of multiple profiles corresponding to a sequence alignment enables effective detection of remote homologues. Bioinformatics 2005;21:2821–2826.
23. Gowri VS, Krishnadev O, Swamy CS, Srinivasan N. MulPSSM: a database of multiple position-specific scoring matrices of protein domain families. Nucleic Acids Res 2006;34:D243–D246.
24. Gowri VS, Tina KG, Krishnadev O, Srinivasan N. Strategies for the effective identification of remotely related sequences in multiple PSSM search approach. Proteins 2007;67:789–794.
25. Marchler-Bauer A, Anderson JB, DeWeese-Scott C, Fedorova ND, Geer LY, He S, Hurwitz DI, Jackson JD, Jacobs AR, Lanczycki CJ, Liebert CA, Liu C, Madej T, Marchler GH, Mazumder R, Nikolskaya AN, Panchenko AR, Rao BS, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Vasudevan S, Wang Y, Yamashita RA, Yin JJ, Bryant SH. CDD: a curated Entrez database of conserved domain alignments. Nucleic Acids Res 2003;31:383–387.
26. Eddy SR. Profile hidden Markov models. Bioinformatics 1998;14:755–763.
27. Krupa A, Srinivasan N. The repertoire of protein kinases encoded in the draft version of the human genome: atypical variations and uncommon domain combinations. Genome Biol 2002;3:RESEARCH0066.
28. Krupa A, Srinivasan N. Diversity in domain architectures of Ser/Thr kinases and their homologues in prokaryotes. BMC Genomics 2005;6:129–148.
29. Anamika K, Srinivasan N, Krupa A. A genomic perspective of protein kinases in *Plasmodium falciparum*. Proteins 2005;58:180–189.
30. Krupa A, Abhinandan KR, Srinivasan N. KinG: a database of protein kinases in genomes. Nucleic Acids Res 2004;32:D153–D155.
31. Krupa A, Anamika K, Srinivasan N. Genome-wide comparative analyses of domain organisation of repertoires of protein kinases of *Arabidopsis thaliana* and *Oryza sativa*. Gene 2006;380:1–13.
32. Anamika K, Srinivasan N. Comparative kinomics of plasmodium organisms: unity in diversity. Protein Pept Lett 2007;14:509–517.
33. Krupa A, Preethi G, Srinivasan N. Structural modes of stabilization of permissive phosphorylation sites in protein kinases: distinct strategies in Ser/Thr and Tyr kinases. J Mol Biol 2004;339:1025–1039.
34. Muller A, MacCallum RM, Sternberg MJ. Benchmarking PSI-BLAST in genome annotation. J Mol Biol 1999;293:1257–1271.
35. Li W, Jaroszewski L, Godzik A. Clustering of highly homologous sequences to reduce the size of large protein databases. Bioinformatics 2001;17:282–283.
36. Kelley LA, MacCallum RM, Sternberg MJ. Enhanced genome annotation using structural profiles in the program 3D-PSSM. J Mol Biol 2000;299:499–520.
37. Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, Jones DT. Protein structure prediction servers at University College London. Nucleic Acids Res 2005;3:W36–W38.
38. Koller KJ, de Sauvage FJ, Lowe DG, Goeddel DV. Conservation of the kinaselike regulatory domain is essential for activation of the natriuretic peptide receptor guanylyl cyclases. Mol Cell Biol 1992;12:2581–2590.
39. Saharinen P, Takaluoma K, Silvennoinen O. Regulation of the Jak2 tyrosine kinase by its pseudokinase domain. Mol Cell Biol 2000;20:3387–3395.
40. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. Multiple sequence alignment with the clustal series of programs. Nucleic Acids Res 2003;31:3497–3500.
41. Kumar S, Tamura K, Nei M. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. Comput Appl Biosci 1994;10:189–191.
42. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL. The Pfam protein families database. Nucleic Acids Res 2002;30:276–280.
43. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001;305:567–580.
44. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science 2002;298:1912–1934.
45. Robinson MJ, Harkins PC, Zhang J, Baer R, Haycock JW, Cobb MH, Goldsmith EJ. Mutation of position 52 in ERK2 creates a nonproductive binding mode for adenosine 5′-triphosphate. Biochemistry 1996;35:5641–5646.
46. Johnson LN, Noble ME, Owen DJ. Active and inactive protein kinases: structural basis for regulation. Cell 1996;85:149–158.
47. Pils B, Schultz J. Inactive enzyme-homologues find new function in regulatory processes. J Mol Biol 2004;340:399–404.
48. Hanks SK, Quinn AM, Hunter T. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science 1988;241:42–52.
49. Demple B, Linn S. On the recognition and cleavage mechanism of Escherichia coli endodeoxyribonuclease V, a possible DNA repair enzyme. J Biol Chem 1982;257:2848–2855.
50. Munoz Mde L, Moreno MA, Perez-Garcia JN, Tovar GR, Hernandez VI. Possible role of calmodulin in the secretion of *Entamoeba histolytica* electron-dense granules containing collagenase. Mol Microbiol 1991;5:1707–1714.
51. Hanks SK, Quinn AM. Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. Methods Enzymol 1991; 200:38–62.
52. Lohia A, Samuelson J. Cloning of the Eh cdc2 gene from *Entamoeba histolytica* encoding a protein kinase p34cdc2 homologue. Gene 1993;127:203–207.
53. Das S, Lohia A. Delinking of S phase and cytokinesis in the protozoan parasite *Entamoeba histolytica*. Cell Microbiol 2002;4:55–60.
54. Lohia A. The cell cycle of *Entamoeba histolytica*. Mol Cell Biochem 2003;253:217–222.
55. Niefind K, Guerra B, Ermakowa I, Issinger OG. Crystal structure of human protein kinase CK2: insights into basic properties of the CK2 holoenzyme. EMBO J 2001;20:5320–5331.
56. Ray D, Dutta S, Banerjee S, Banerjee R, Raha S. Identification, structure, and phylogenetic relationships of a mitogen-activated protein kinase homologue from the parasitic protist *Entamoeba histolytica*. Gene 2005;346:41–50.
57. De Meester F, Mirelman D, Stolarsky T, Lester DS. Identification of protein kinase C and its potential substrate in *Entamoeba histolytica*. Comp Biochem Physiol B 1990;97:707–711.
58. Keller K, Olivier M, Chadee K. The fast release of mucin secretion from human colonic cells induced by *Entamoeba histolytica* is dependent on contact and protein kinase C activation. Arch Med Res 1992;23:217–221.
59. Labruyere E, Zimmer C, Galy V, Olivo-Marin JC, Guillen N. EhPAK, a member of the p21-activated kinase family, is involved in the control of *Entamoeba histolytica* migration and phagocytosis. J Cell Sci 2003;116:61–71.
60. Manser E, Leung T, Salihuddin H, Zhao Z-S, Lim L. A brain serine/threonine protein kinase activated by Cdc42 and Rac1. Nature 1994;367:40–46.
61. Sorimachi H, Ishiura S, Suzuki K. Structure and physiological function of calpains. Biochem J 1997;328:721–732.

62. Naula C, Parsons M, Mottram JC. Protein kinases as drug targets in trypanosomes and Leishmania. Biochim Biophys Acta 2005;1754: 151–159.
63. Bork P, Doolittle RF. Drosophila kelch motif is derived from a common enzyme fold. J Mol Biol 1994;236:1277–1282.
64. Li X, Zhang D, Hannink M, Beamer LJ. Crystal structure of the Kelch domain of human Keap1. J Biol Chem 2004;279:54750–54758.
65. Engel J. Laminins and other strange proteins. Biochemistry 1992;31: 10643–10651.
66. Lamb JR, Tugendreich S, Hieter P. Tetratrico peptide repeat interactions: to TPR or not to TPR? Trends Biochem Sci 1995;20:257–259.
67. Volkman BF, Anderson ME, Clark KD, Hayakawa Y, Strand MR, Markley JL. Structure of the insect cytokine peptide plasmatocyte-spreading peptide 1 from *Pseudoplusia includens*. Biol Chem 1999; 274:4493–4496.
68. Skinner WS, Dennis PA, Li JP, Summerfelt RM, Carney RL, Quistad GB. Isolation and identification of paralytic peptides from hemolymph of the lepidopteran insects *Manduca sexta*, *Spodoptera exigua*, and *Heliothis virescens*. J Biol Chem 1991;266:12873–12877.
69. Tirasophon W, Welihinda AA, Kaufman RJ. A stress response pathway from the endoplasmic reticulum to the nucleus requires a novel bifunctional protein kinase/endoribonuclease (Ire1p) in mammalian cells. Genes Dev 1998;12:1812–1824.