Identification and Gene Expression Analysis of a Large Family of Transmembrane Kinases Related to the Gal/GalNac Lectin in Entamoeba histolytica†

David L. Beck, 1, ‡ Douglas R. Boettner, 1 Bojan Dragulev, 1 Kim Ready, 2, 3 Tomoyoshi Nozaki, 4, 5 and William A. Petri, Jr. 1, 2, 3*

Departments of Microbiology, 1 Medicine, 2 and Pathology, 3 University of Virginia, Charlottesville, Virginia 22908-1340, and Department of Parasitology, National Institute of Infectious Diseases, Tokyo 162-8640, 4 and Precursory Research for Embryonic Science and Technology, Japan Science and Technology Agency, 2-20-5 Akebonocho, Tachikawa, Tokyo 190-0012, 5 Japan

Received 22 November 2004/Accepted 9 February 2005

We identified in the Entamoeba histolytica genome a family of over 80 putative transmembrane kinases (TMKs). The TMK extracellular domains had significant similarity to the intermediate subunit (Igl) of the parasite Gal/GalNac lectin. The closest homolog to the E. histolytica TMK kinase domain was a cytoplasmic dual-specificity kinase, SplA, from Dictyostelium discoideum. Sequence analysis of the TMK family demonstrated similarities to both serine/threonine and tyrosine kinases. TMK genes from each of six phylogenetic groups were expressed as mRNA in trophozoites, as assessed by spotted oligoarray and real-time PCR assays, suggesting nonredundant functions of the TMK groups for sensing and responding to extracellular stimuli. Additionally, we observed changes in the expression profile of the TMKs in continuous culture. Antisera produced against the conserved kinase domain identified proteins of the expected molecular masses of the expressed TMKs. Confocal microscopy with anti-TMK kinase antibodies revealed a focal distribution of the TMKs on the cytoplastic face of the trophozoite plasma membrane. We conclude that E. histolytica expresses members of each subgroup of TMKs. The presence of multiple receptor kinases in the plasma membrane offers for the first time a potential explanation of the ability of the parasite to respond to the changing environment of the host.

The Gal/GalNac lectin of Entamoeba histolytica mediates parasite adherence to the host and signals the initiation of cytolysis (41, 44, 45, 49). It is a heterotrimer consisting of covalently linked heavy (Hgl) and light (Lgl) subunits with a noncovalently linked intermediate (Igl) subunit (9, 36, 37, 43, 46). The Igl subunit of the Gal/GalNac lectin has two known family members, Igll and IgI2. The Igl subunit has sequence similarity to the variant surface protein (VSP) of Giardia. We have previously identified a large number of proteins in the genome of E. histolytica containing CXXC motifs similar to those of Igl (8). Here we show that these CXXC-rich proteins form a large family of E. histolytica transmembrane kinases (TMKs) with highly variable extracellular domains homologous to IgI and VSPs of Giardia and with cytoplasmic kinase domains.

Amebic trophozoites have been demonstrated to persist in humans for longer than 6 months (21, 22). This prolonged period of infection suggests that the amebae evade the immune system. Other protozoan parasites, such as Plasmodium, Giardia, and Trypanosoma brucei, are also able to infect the host for long periods in spite of inducing robust immune responses. The mechanism(s) of persistence of these organisms is thought in part to be due to the variation of surface proteins. Plasmodium falciparum has three families of var genes that are independently expressed (29). The highest variation rate of these families is 2% per generation (52). Giardia encodes a family of 100 to 150 VSPs whose surface expression changes at a rate of one variation every 5 to 13 generations (38). T. brucei has a family of over 1,000 variant surface glycoproteins that change at a rate of $10^{-2}$ to $10^{-7}$ variations per generation (13, 51).

The discovery of the large family of CXXC-containing TMKs is of interest not only for their potential role in antigenic variation but also for their role in cell signaling. E. histolytica must respond to a wide variety of environmental stimuli as it excysts into a trophozoite in the intestinal lumen and enters the host by invasion of the intestinal mucosal epithelium. Invasion involves attaching to the epithelium and responding to that attachment event through signaling events via the E. histolytica Gal/GalNac adherence lectin that lead to host cell killing. The changing host environment should necessitate having a variety of ways of sensing and responding to the host.

Here we report sequence and expression analysis of the TMKs in laboratory-cultured trophozoites. An oligoarray and real-time PCR were used to measure the expression in cultured trophozoites of the TMK genes. We demonstrate that there are six families of TMKs, with each having one or more family members expressed. In addition, anti-TMK antibodies
were used to localize the TMKs to the plasma membrane of trophozoites, consistent with their proposed function in sensing the environment.

MATERIALS AND METHODS

Identification of genes homologous to the Igl subunit of the Gal/GalNac lectin. The genes were identified in the 7X assembly available from The Institute for Genomic Research (TIGR) and Sanger sequencing centers (http://www.tigr.org/tdb/ekl1/eh1 and http://www.sanger.ac.uk/Projects/E_historicata) by searching the database for homologs of Igl1. Genes with high sequence similarity to Igl1 were used to search the database and identify additional family members. Additionally, the Sanger assembly was translated in all six reading frames, and genes were identified by sequence similarity to known genes in the National Center for Biotechnology Information (NCBI) database. These sequences were then screened for genes containing sequences for three or more CXXC motifs, or kinase domains. Genes containing sequences for CXXC motifs but not kinase domains, transmembrane domains, or signal peptides were eliminated from the data set.

Identification of other virulence genes and control genes. Genes were identified by sequence similarity to genes for amoebapores, cysteine proteinases, and the Gal/GalNac lectin IglL, IglG, and Hgl subunits. Additionally, genes were identified by examination of the translated Sanger assembly, which had been annotated to known genes in the NCBI database. Phagocytosis genes and control genes were similarly identified. BspA genes were identified in the translated Sanger assembly and then identified by sequence similarity in the TIGR assembly.

Phylogenetic analysis of the TMK proteins. A 260-amino-acid alignment of the kinase domains of the TMK proteins was made to Hank’s kinase alignment (Protein Kinase Resource [http://pkr.sdsc.edu/html/pk_classification/pk_catalytic/pk_hanks_class.html]) using CLUSTALX (20, 61). One representative per family, called the query panel of genes, was employed (http://pkr.sdsc.edu/html/pk_classification/pk_catalytic/query_panel.html). The alignment was manually optimized using GeneDoc (39), and then sequences were analyzed using the PHYLIP v3.6 package (15) and bootstrapped using Seqboot, PHYLIP v3.6 package (15) and bootstrapped using Seqboot, Protdist, Neighbor, and Consense. A subset of the sequences were then bootstrapped using Seqboot, PHYLIP v3.6 package (15) and bootstrapped using Seqboot, Protdist, Neighbor, and Consense. The alignment was manually optimized using GeneDoc (39), and then sequences were analyzed using the PHYLIP v3.6 package (15) and bootstrapped using Seqboot, Protdist, Neighbor, and Consense. A subset of the sequences were then bootstrapped using Seqboot, Protdist, Neighbor, and Consense. The alignment was manually optimized using GeneDoc (39), and then sequences were analyzed using the PHYLIP v3.6 package (15) and bootstrapped using Seqboot, Protdist, Neighbor, and Consense. A subset of the sequences were then bootstrapped using Seqboot, Protdist, Neighbor, and Consense.
used to ensure specificity, and any PCR primer pair that produced more than one
melt peak was discarded. PCR products that produced single melt peaks were
analyzed by gel electrophoresis in 1.5% agarose–Tris-borate-EDTA, and if mul-
tiple bands were observed, the primer pair was discarded. Finally, all PCR
products were sequenced using the forward amplification primer to verify spec-
ificity.

Real-time PCR validation of oligoarray results. RNA was reverse transcribed
using iQscript (Bio-Rad, Hercules, CA) according to the manufacturer’s direc-
tions. cDNA levels were measured using iQSYBRGreen super mix (Bio-
Rad, Hercules, CA), 1,100 pmol of forward primer (50 pmol/μl), and 88 μl of reverse primer (50 pmol/μl). To each well containing 2 μl
cDNA was added 25 μl of master mix. Duplicate assays were performed on
each sample. Each assay included standards, no-DNA-control wells, and no-RT-
control wells. The cycling conditions were 95°C for 5 min; 30 cycles of 95°C for
30 s, annealing for 30 s (see Table 1 for annealing temperatures), and 72°C for
30 s; and 1 cycle of 72°C for 2 min 30 s followed by a 90-step melt curve increasing
0.2°C with a 5-s hold.

| Gene product | Primer sequence | Annealing temp (°C) |
|-------------|----------------|-------------------|
| RNA Pol II  | AAAGAAGGTGTGTCCTGTGAAGGCAGGTTGGGAAGATGCTTCTTC | 66 |
| RNA Pol II L | GTAAGTTATGGAAGATGGATGCTTCTTC | 66 |
| Jacob | GCCAGTTACCAACTGGATGGGAGATTCTACA | 66 |
| Tmk 19 | TTTGAGGATGTTGGGAGAATGCTTCTTC | 66 |
| Tmk 21 | CACCCAGGGTTGGAAGATGCTTCTTC | 66 |
| Tmk 31 | CTCTTATGGAAGATGCTTCTTC | 66 |
| Tmk 71 | TTTGAGGATGTTGGGAGAATGCTTCTTC | 66 |
| Tmk 75 | AAACTACTGACGAGGTTGGAAGATGCTTCTTC | 66 |
| Tmk 79 | CACCCAGGGTTGGAAGATGCTTCTTC | 66 |
| Tmk 80 | CACCCAGGGTTGGAAGATGCTTCTTC | 66 |
| Tmk 83 | ACCTGTCAGTGGATGTTGGGAAGATGCTTCTTC | 66 |
| Tmk 96 | AAGATTGTTGAAGATGCTTCTTC | 66 |
| Tmk 97 | CACCCAGGGTTGGAAGATGCTTCTTC | 66 |
| Tmk 98 | CACCCAGGGTTGGAAGATGCTTCTTC | 66 |
| SA  | CACCCAGGGTTGGAAGATGCTTCTTC | 66 |
| Actin  | GCCACTGTCAGTGGATGTTGGGAAGATGCTTCTTC | 66 |
| Hgl  | GCCACTGTCAGTGGATGTTGGGAAGATGCTTCTTC | 66 |

Table 1. RT-PCR primers

Table 2. Properties of the domains of the E. histolytica transmembrane kinase groups

| Group | Transmembrane kinase no. | Size | No. of amino acids, TM to kinase | Signature motif in kinase | No. of amino acids, kinase to end | No. of motifs |
|-------|--------------------------|------|---------------------------------|--------------------------|---------------------------------|--------------|
| A     | 4, 17, 23, 25, 52, 53, 55, 61, 65, 68, 69, 72, 85 | 517–532 | 220 | CC(U/V)KTDFFGTSR | 40 | 4 |
| B1    | 5, 12, 43, 76, 81, 95, 100, 101, 104, 107 | 897–916 | 135 | KLTDFGS(A/S)R | 0 | 1 |
| B2    | 2, 8, 10, 11, 14, 15, 31, 36, 41, 62, 74, 75, 77, 87, 89, 92, 94, 105 | 822–1763 | 133 | KLTDFGS(A/S)R | 0 | 0 |
| B3    | 21, 28, 29, 30, 32, 35, 37, 38, 42, 48, 51, 59 | 830–2117 | 220 | CC(U/V)KTDFFGTSR | 0 | 0 |
| C     | 9, 13, 39, 60, 63, 71 | 547–624 | 160 | C(A/G)KLTDFGS | 59–82 | 0 |
| D1    | 3, 18, 40, 56, 70, 79 | 520–619 | 234 | PTAKVTDFFGTS | 63 | 3 |
| D2    | 19, 27, 44, 50, 57, 64, 67, 82, 90 | 399–614 | 233 | V(T/V)(C/X)KTDFFGTS | 55 | 4 |
| E     | 22, 54, 66 | 401–412 | 150 | AKLSDFGSR | 60–97 | 1 |
| F     | 34, 45, 59, 80 | 231–368 | 24 | VKVSDFGLS and WXAPE | 0 | 0 |

a Group G transmembrane kinases are diverse and include kinases 1, 6, 16, 24, 47, 49, 73, 79, with no common signature motif or other features within the group or with other groups.

b Number of amino acid residues in the extracellular domain.

c Amino acids from the C-terminus of the transmembrane domain to the N-terminal glycine of the kinase domain.

d Amino acids from 10 residues after the conserved arginine to the stop codon.

e Number of times each motif is found in the protein sequence. The number of CXC, CXXC, CXXXC, and CXXCXXGYY motifs may vary between individual family members. Values with a slash indicate that some family members have 4 or 5, or 3 or 4, of the motif.

f Group A TMKs can additionally be identified by a moderately conserved (K/R)XXXDI(E/N)I(Y/F)KQQPXYYYYIXGSXXXPXKXX(K/R)Y motif C-terminal to the transmembrane domain.

g Group B1 TMKs can be identified by a KRKEKKEREKTTIFKTITQSNI(K/R)FI(S/P)LGDG sequence after their transmembrane domain.

h Group B2 TMKs can be identified by a RRR(K/R)XXKXXXXXIKPF(H/K)VSSD(L/V)ELXLL sequence following the transmembrane domain.

i Group B3 TMKs can be identified by a RRR(K/R)XXKXXXXXIKPF(H/K)VSSD(L/V)ELXLL sequence following the transmembrane domain.
Production of anti-ΔTMK96 rabbit serum production. The kinase region of Tmk96 (ΔTMK96) was PCR amplified with the primers 5′-CAATTAGAGAAGGAAATTCT-3′ (5′ primer) and 5′-TCACAAATGAGATGTGTTTTAAACAANCA-3′ (3′ primer). This 1,000-bp fragment was cloned into TOPO NT/T7 primer). This 1,000-bp fragment was cloned into TOPO NT/T7 vector. This 1,000-bp fragment was cloned into TOPO NT/T7 vector. This 1,000-bp fragment was cloned into TOPO NT/T7 vector.

Western blots using ΔTMK96. Soluble proteins were extracted from amebae by harvesting 5 × 10^7 trophozoites by incubation on ice for 10 min, followed by centrifugation (200 x g at 4°C for 5 min). The amebae were lysed in 10 mM sodium phosphate buffer with protease inhibitor cocktail I (Sigma, St. Louis, MO) per the manufacturer’s directions. Membranes were then clarified by centrifugation (10,000 x g at 4°C for 1 h). Whole-cell lysates were prepared by sonication of 10^6 amebae in three 5-min pulses on ice. Large intact particles were eliminated by centrifugation (20,000 x g at 4°C for 30 min). All samples were then separated on 10% polyacrylamide gels and then electrotransferred to polyvinylidene difluoride membranes (Millipore, Bedford, MA). Nonspecific binding was blocked by preincubation with Tris-buffered saline with 5% bovine serum albumin (Sigma) in PBS for 1 h at 37°C. After incubation with either the anti-ΔTMK96 rabbit polyclonal antibody (200 μg/ml) or anti-Gal/GalNAc lectin antibody (6 μg/ml) for 1 h at 37°C, the coverslips were washed twice before Cy3-conjugated goat anti-rabbit secondary antibodies (Jackson Laboratories, Bar Harbor, ME) were added at a 1:160 dilution for 1 h at 37°C. The coverslips were washed twice before Cy3-conjugated goat anti-rabbit secondary antibodies (Jackson Laboratories, Bar Harbor, ME) were added at a 1:160 dilution for 1 h at 37°C. The coverslips were washed twice before Cy3-conjugated goat anti-rabbit secondary antibodies (Jackson Laboratories, Bar Harbor, ME) were added at a 1:160 dilution for 1 h at 37°C. The coverslips were washed twice before Cy3-conjugated goat anti-rabbit secondary antibodies (Jackson Laboratories, Bar Harbor, ME) were added at a 1:160 dilution for 1 h at 37°C. The coverslips were washed twice before Cy3-conjugated goat anti-rabbit secondary antibodies (Jackson Laboratories, Bar Harbor, ME) were added at a 1:160 dilution for 1 h at 37°C. The coverslips were washed twice before Cy3-conjugated goat anti-rabbit secondary antibodies (Jackson Laboratories, Bar Harbor, ME) were added at a 1:160 dilution for 1 h at 37°C. The coverslips were washed twice before Cy3-conjugated goat anti-rabbit secondary antibodies (Jackson Laboratories, Bar Harbor, ME) were added at a 1:160 dilution for 1 h at 37°C. The coverslips were washed twice before Cy3-conjugated goat anti-rabbit secondary antibodies (Jackson Laboratories, Bar Harbor, ME) were added at a 1:160 dilution for 1 h at 37°C.

RESULTS

Phylogenetic analysis of the E. histolytica TMK family. Previously we reported a family of CXXC-containing proteins with sequence similarity to Igl1 and Igl2 (8). With the completion of the genome project, it became apparent that the majority of these CXXC-containing proteins were members of an approximately 80-gene family with intracellular kinase domains. The extracellular domains shared sequence similarity to Igl and to VSPs of Giardia. The kinase domain of these proteins in most cases contained all five conserved kinase motifs: the key conserved glycine-rich motif, the K residue, and the HRDL, DFG, and APE motifs (Table 2). However, a few lacked the glycine-rich motif (Tmk01, Tmk24, Tmk45, and Tmk73), which helps coordinate the second phosphate of ATP (14). The glycine, K, and DFG motifs are implicated in the binding and orientation of ATP. The HRDL motif is involved in catalysis. The APE motif is responsible for anchoring the substrate and thereby influencing the specificity of interaction with the protein that is being phosphorylated (14, 20, 60). Despite the conservation of the essential kinase motifs, overall the kinase domains were

FIG. 2. Diagram of Igl and the TMKs. This diagram shows the approximate sizes of the different proteins and the distribution of the CXXCXXGYY motifs in the extracellular domain (indicated by a black circle). The CXXCXXGYY motifs are part of a larger motif, CXXCXXG(Y)(Y/F)(L/V/F/Y/M)-Polar-Polar, which also can begin with CXC instead of CXXC. GPI represents a putative GPI anchor. A black rectangle indicates a transmembrane domain, and a black oval indicates a putative kinase domain. The serine (SSS)- and serine/threonine (SST)-rich regions found in groups D1 and D2 are shown. Numbers in brackets indicate numbers of known family members.
divergent from other known kinases. The TMK family branched closely with the other protein kinase group IX (OPK IX), which are the TGFβ/H9252 receptor and activin family of serine/threonine kinases, OPK VIII, which are a Raf family of serine/threonine kinases, and OPK XII, which are the casein kinase I family of kinases (Fig. 1). It was not possible to determine by sequence analysis alone if the TMKs were serine/threonine or tyrosine kinases. Sp1A from Dictyostelium discoideum was the most closely related kinase identified by sequence similarity in the NCBI database that has been functionally characterized. The Sp1A kinase has been shown to be a dual-specificity kinase that phosphorylates both tyrosine and serine/threonine residues (40). The Sp1A kinase was found to be phylogenetically within the TMK family (Fig. 1). Unlike the TMK family, Sp1A is a cytoplasmic kinase rather than a transmembrane kinase.

Two TMKs (Tmk58 and Tmk89) clearly did not group with the TMK family but grouped closely with other protein kinase groups VI and VII (data not shown), suggesting that these may represent serine/threonine transmembrane kinases that have evolved separately from the TMK family.

The TMK family members were grouped based on sequence and phylogenetic analysis of their cytoplasmic kinase domains. A specific signature motif between the conserved HRDL and APE motifs was identified for each family (Table 2). The motif was 15 to 24 amino acids C-terminal to the histidine in the HRDL motif. The motif sequences were CC(I/V)KITDFGTSR (group A), KLTDFGS(A/S)R (group B), C(A/G)KLTDFGTC (group C), PITAKVTDFGTS (group D1), V(T/V)(C/X)KV(T/S)DFGTS (group D2), AKLSDFGTSR (group E), and VKVSDFGLS with a conserved tryptophan two residues.

### Table 3. Consensus sequence comparison of subdomains of malian tyrosine kinases, serine/threonine kinases, and the E. histolytica transmembrane kinase groups

| Kinase | Consensus sequence in subdomain<sup>a</sup> |
|--------|----------------------------------------|
| Tyrosine | LGRxGxFGxVxxG | VA(1/V)K | Exxxx(R/K) | GxKooO | GxL(F/Y)lx |
| Ser/Thr | LGGRxGxFGxVxxG | VAIK | Exxxx(K/R/H) | Gxxxxx | GxX(Y/F)(L/I)H |
| A | IAEGAMG-VY-G | VA(1/V)K | ExxxxL | GSVTYI | GSLe~gYR |
| B1 | IGGExFGIVYVG | VAIK | Exe(S/N)DK | GAVFiP | GSIQDINMK |
| B2 | oGExFGooY(K/o)G | VAIK | EVoMLDK | GavfiP | GSLxoxoXK |
| B3 | oGExFGoV(Y/F)KG | VAIK | ExomLdk | Gavoop | GSooxo(x)k |
| C | IGEAFGMVFRG | VAIK | EQKEEF | GAVYTE | GS(M/L)SKOWXK |
| D1 | IGGEFGIVYRA | VAVK | EAEMMER | GSVoTx | GSLRKFMK |
| D2 | IGGEFGoVY(K/R)(G/A) | VA(V/I)K | ExxxxEx | GSVxxx | GSLRK(Y/f)ok |
| E | IGEg(S/T)FgLV(Y/F)xG | VAIK | ExFxxE | Gxxxox | GSeKxOY |
| F | oGxg(S/T)xgxxvxxa | oavK | evxlmK | Gsgxdx | gxoxoL |

<sup>a</sup> Subdomain is according to Hanks's alignment of kinases (20). Underlined in tyrosine and serine/threonine kinases indicate highly conserved residues. Letters in uppercase are conserved. Letters in lowercase indicate moderate conservation. x, any residue; o, hydrophobic residue; —, sequence of varying length.

### Table 4. Significantly changed genes during growth

| Growth period<sup>a</sup> | Probe(s)<sup>b</sup> | Protein | Group | Change | P value |
|-------------------------|------------------|---------|-------|--------|---------|
| Early (12 h p.i.) to mid-log (48 h p.i.) | 1, 91, 99 | Actin | Gal/GalNAc lectin | Increased 2.17-fold | <0.05 |
| | 190 | Lgl13 | | Increased 1.37-fold | <0.05 |
| Late log (96 h p.i.) to nonadherent (144 h p.i.) | 1, 91, 99 | Actin | Cysteine proteinase | Decreased 1.88-fold | <0.01 |
| | 321 | EHCPl&2 | | Decreased 1.79-fold | <0.05 |
| | 320 | EHCPl20 | Cysteine proteinase | Increased 1.37-fold | <0.05 |
| | 88 | Hgl family | Gal/GalNAc lectin | Decreased 1.58-fold | <0.01 |
| | 189 | Lgl1 | Gal/GalNAc lectin | Decreased 2.87-fold | <0.05 |
| | 280 | Lgl4 | Gal/GalNAc lectin | Decreased 2.28-fold | <0.01 |
| | 232 | Lgl5 | Gal/GalNAc lectin | Decreased 4.26-fold | <0.05 |
| | 224 | EN3539 | Unknown—surface | Increased 1.61-fold | <0.05 |
| | 145 | TSA | Thiolredoxin | Decreased 1.65-fold | <0.05 |
| | 239, 277 | SREHP | Surface protein | Decreased 3.08-fold | <0.05 |

<sup>a</sup> Three hybridizations (biological replicates) were analyzed in two dye experiments comparing 12 h to 48 h, 48 h to 96 h (no significant differences were found), and 96 h, to 148 h postinoculation (p.i.).

<sup>b</sup> Probe numbers are from Table S1 in the supplemental material.
Table 3—Continued

| Vla | Vlb | VII | VIII | IX | X | XI |
|-----|-----|-----|------|----|----|----|
| GMxY1x | o(I/V)HRDLaar | C-KxDPoGl | EIVxK(R) | SDxVx(S/A)(F/Y) | — | CIxxxxxxxxxxRPxF |
| (G/A)ox | (Y/F)OH | (G/A)ox | — | GxxxOE—PyP | — | — |
| GMxFLH | IoHDLKPNlOluV | KITDFGt | GTPxYxxAPEXY | — | — | — |
| GIxY1x | IHLRLIKPDNoLo | KLTDFGS | GTPxYxxAPEoL | — | — | — |
| GIxY1x | IHLRLIKPDNoLo | KLTDFGS | GTPxYxxAPEoL | — | — | — |
| ALx(Y/F) | IIRHDxVxGEnoLo | KLTDFGt | GTPxYxxAPEoL | — | — | — |
| GMEY1x | IIRHDxKTDxNVlY | KVPxDFG1 | GTPxYxxAPEoL | — | — | — |
| GMxY1x | IoHDLxKtENVlO | KVx(T/S) | GTPxYxxAPEoL | — | — | — |
| GMxFLH | IIRHDxKPDNoLo | KLSDFxT | GTPxYxxAPEoL | — | — | — |
| GMxY1x | IoHDLxKsxNoLVx | KvSDFgl | gtxWxAPEoL | — | — | — |

Consensus sequence in subdomain:

Genes N-terminal to the APE motif (group F) (Table 2 and Fig. 2), Tmk58 and Tmk89, which did not group with the rest of the TMKs, contained an ITDFG/LAKK motif. Group G TMKs lacked a conserved motif in the kinase domain and also lacked one or more of the conserved kinase domain motifs. Sequence similarity between family members was not limited to the kinase domain. Additionally, that can be used to identify group members between the cytoplasmic and kinase domain (Table 2 footnotes). Additionally, that for the cyst-specific transcript hybridization values in trophozoites significantly greater than

Gene expression analysis of E. histolytica during growth. To characterize the expression profile of these genes, we constructed an oligoarray of the TMK genes, Gal/GalNAc lectin genes, and other putative surface virulence genes. Few changes in gene expression were seen when early phase (12 h) was compared with mid-log phase (48 h) and late log phase (96 h) was compared with nonadherent phase (144 h) (Table 4). When mid-log-phase (48 h) and late-log-phase (96 h) amebic cultures were compared, very little change was seen (data not shown). actin was clearly growth regulated (Table 4), as were the Gal/GalNAc lectin hgl genes, many but not all of which decreased significantly during late log phase (Table 4 and data not shown). This decrease is consistent with previous observations of hgl1, hgl2, and hgl3 gene expression (48). Expression of known genes was, in general, consistent with RT-PCR, Western, and/or Northern analysis (data not shown). We concluded that E. histolytica did not appear to growth-phase regulate expression of most putative virulence genes when grown under lab culture conditions.

Expression analysis of the TMKs. The expression profile of 67 of the TMK genes during mid-log phase (72 h) is shown in Fig. 3. At this time point, 19 genes (group A, tmk61, tmk65, and tmk72; group B2, tmk02, tmk08, and tmk74; group B3, tmk21 and tmk28; group C, tmk39 and tmk63; group D1, tmk40 and tmk56; group D2, tmk19, tmk44 and tmk46; group E, tmk22 and tmk54; group F, tmk39; group G, tmk06) (Fig. 3) showed hybridization values in trophozoites significantly greater than that for the cyst-specific transcript jacob (P < 0.05).

Analysis of TMK expression by RT-PCR. Real-time PCR was conducted on jacob, tmk31, tmk80, tmk96, and tmk98, all of
which were predicted by oligoarray not to be expressed, and tmk19, tmk21, tmk63, tmk65, tmk71, sa, hgl, and actin, all of which were predicted to be expressed. Gene expression was monitored sequentially in trophozoites in laboratory culture over a 12-day period. To allow comparison between time points, results were normalized to the average of three RNA polymerase II genes (Fig. 4). All real-time PCR results of the TMK genes were consistent with the oligoarray results. Significant variations in expression during laboratory culture were observed for tmk19, tmk63, and tmk79 (P < 0.01) (Fig. 4D, 4F, and 4I).

Expression analysis of the TMKs during erythrophagocytosis. One million trophozoites were grown in 50 ml of medium with or without a vast excess of erythrocytes (24 million/ml of medium) for 24 h. We did not observe significant changes in TMK gene expression during erythrophagocytosis (data not shown).

Detection of expression of TMK family members with polyclonal antibodies. The kinase domain of Tmk96 was expressed in Escherichia coli and used to generate polyclonal antibodies. Multiple trophozoite proteins were detected with polyclonal anti-kinase domain antisera (Fig. 5A). No bands were observed with preimmune sera (data not shown). The recognition of multiple proteins by the antisera raised against the Tmk96 kinase domain was not surprising given that the kinase domain is conserved between different TMKs. The predicted TMKs vary in size from 482 to 2,577 amino acids, and Tmk96 is in the largest subfamily (group B) of TMKs with multiple members expressed. Of the TMK genes that we detected with expression significantly above jacob the size range was from 686 to 2,577 amino acids (78 kDa to 294 kDa not accounting for potential posttranslational modifications). The smallest proteins observed on Western blots were of a mass consistent with that of a typical kinase domain (270 amino acids) and may represent

FIG. 3. Expression of TMKs during log-phase culture as determined by oligoarrays. The TMKs were classified into groups: A, B2, B3, C, D1, D2, E, F, G, and “other” (tmk58). Expression of jacob (encoding a cyst protein not expected to be expressed in trophozoites), ebp1, and ebp2 genes is shown for reference. TMK genes expressed at a higher level than jacob (P < 0.05) are indicated by an asterisk and were as follows: group A, tmk61 (391.t00004-AAFB01000774, 279.t000010-AAFB01000993), tmk65 (62.t00013-AAFB01000240), and tmk72 (302.t00003-AAFB01000819); group B2, tmk2 (70.t00014-AAFB01000264), tmk68 (10.t00040-AAFB010000051), and tmk74 (6.t00088-AAFB010000031); group B3, tmk21 (42.t000019-AAFB010000175), and tmk28 (66.t00027-AAFB01000251); group C, tmk39 (359.t00009-AAFB01000933), and tmk63 (20.t00067-AAFB01000094); group D1, tmk40 (65.t000015-AAFB010000027), and tmk56 (5.t00091-AAFB010000028); group D2, tmk19 (135.t00017-AAFB010000458), tmk4 (159.t000012-AAFB010000511), and tmk46 (131.t00015-AAFB01000449); group E, tmk22 (12.t00043-AAFB01000464) and tmk54 (75.t00011-AAFB01000285); group F, tmk59 (304.t00008-AAFB01000821); and group G, tmk96 (274.t00000-AAFB01000764). Error bars represent the standard error of the mean of three hybridizations (biological replicates).
TMKs that do not have an extracellular domain or whose extracellular domain has been cleaved off.

Localization of TMK kinase domains to plasma membrane microdomains. Permeabilized amebae showed a focal plasma membrane staining pattern with anti-TMK kinase domain antibodies (Fig. 5B). No staining was seen with nonpermeabilized cells or with permeabilized cells stained with preimmune sera (data not shown). We concluded that the kinase domain was on the cytoplasmic side of the plasma membrane. The TMKs therefore appeared to be typical type I transmembrane proteins with an amino-terminal signal sequence and a predicted transmembrane domain preceding the kinase domain. The focal staining pattern contrasted with the uniform plasma membrane staining pattern seen with anti-Gal/GalNAc lectin antisera (Fig. 5C).

DISCUSSION

The most important finding of this work is the identification of a large family of over 80 transmembrane kinases in *E. histolytica*. Although eukaryotic-type transmembrane kinases are found in organisms from bacteria to humans (4, 64, 68), large families of TMKs have been previously described only in multicellular organisms (20). For example, we were not able to find any TMKs in the partially completed genome of the ameba *D. discoideum* (17), and only a few TMKs were found in the apicomplexan *P. falciparum* (4). The TMKs were also found in *Entamoeba invadens*, *Entamoeba dispar*, and *Entamoeba moshkovskii*, indicating that this family is conserved within *Entamoeba* (data not shown) (66). Multiple TMKs were observed to be expressed at the mRNA and protein levels,

FIG. 4. Expression of TMK family genes during culture as determined by real-time PCR. Quantitative real-time PCR was performed on (A) the RNA polymerase II gene (*rna pol ii*) (27.t00035-AAFB01000114), (B) the RNA polymerase II 13 gene (*rna pol ii 13*) (344.t00001-AAFB010000458), (C) the RNA polymerase II L gene (*rna pol ii l*) (147.t00002-AAFB010000482), (D) *tmk19* (345.t00002-AAFB010000482), (E) *tmk21* (42.t00001-AAFB01000175), (F) *tmk63* (20.t000067-AAFB01000004), (G) *tmk65* (62.t000013-AAFB010000420), (H) *tmk71* (268.t000077-AAFB010000754), (I) *tmk79* (71.t000002-AAFB01000266), and (J) *tmk98* (361.t00001-AAFB01000937). Two sequential growth curves are shown. For growth curve A (triangles and solid line), samples were collected at 12, 48, 96, and 144 h postinoculation. For growth curve B (squares and dashed line), samples were collected at 12, 24, 48, 72, 96, 120, and 144 h postinoculation. Triplicate samples were collected at each time point. Culture B was established by transferring 300,000 amebae from culture A at 144 h. The standard errors of three biological samples, with each sample analyzed in duplicate, are shown.
consistent with nonoverlapping biological functions for individual members of the TMK family.

The existence of multiple different extracellular domains of the TMKs suggests that each interacts with the host environment and signals into the parasite in distinct ways. The extracellular domains of the group B, C, and D TMKs had sequence similarity to Igl of *E. histolytica* (8), laminin LE domains (27, 57, 63), and VSPs of *Giardia lamblia* (1, 38). The sequence similarity is largely limited to a repeated CXXCXXGYY motif. In laminin the LE domains function as mini-globular folds arranged in tandem to form a rod-like structure. If the CXXCXXGYY motifs in the TMKs take on a similar conformation, then these motifs may function to help the extracellular domain of the TMKs project off the surface of the cell in a pilus-like manner. This may make them available for interaction with host cell factors. In each TMK subfamily there was one or more members expressed; however, most family members appeared not to be expressed under the conditions of laboratory culture. By RT-PCR we did observe that the expression of some TMKs varied between growth curves. This indicates that the expression of these genes may be dynamic. Whether the TMKs share with the *Giardia* VSPs the process of antigenic variation under different biological conditions remains to be determined.

The most significant feature of the TMKs is the kinase domain that, with the exception of two TMKs, is distinct from other known kinases. It is not possible, based on sequence analysis, to predict activity, as most have similarity to both the serine/threonine and tyrosine kinases. Interestingly, a closely related kinase, SplA from *D. discoideum*, is a dual-specificity kinase with both tyrosine and serine/threonine kinase activity (40). All of the essential kinase motifs were conserved, suggesting that these are functional kinases. We were not able to demonstrate kinase activity when the kinase domain was expressed in *E. coli*. Since all of the functional residues were conserved, the most likely explanation for this is that the kinase is not functional in *E. coli* or that the kinase domain is not able to phosphorylate the substrates we have used. It has been previously shown that some kinases were not functional when expressed in *E. coli* (18). Additionally, even if the kinase is functional, identification of a substrate is often the rate-limiting step in characterizing a kinase (5, 23). Further experimentation will be necessary to identify the substrate or interacting partners of the TMKs and determine if the TMKs are serine/threonine and/or tyrosine kinases. Phylogenetic and sequence analysis shows that there are six subfamilies of kinases with distinct motifs within the kinase domains. Some families had additional conserved motifs outside of the kinase domain. This
would imply that the subfamilies may represent functionally different families of kinases in sensing (differences in extracellular domains) and signaling (differences in kinase domains).

The focal staining pattern of the TMKs distinctly contrasts with the uniform plasma membrane staining pattern seen with the Gal/GaINac lectin of *E. histolytica* or VSPs in *G. lamblia*, both of which lack cytoplasmic kinase domains (38, 42, 45). This localization suggests that the TMKs form a focal multi-molecular signaling complex in the plasma membrane (31, 33).

In conclusion, the work presented here may begin to explain how *E. histolytica* is able to persist in the host for long periods of time despite immune surveillance, as well as sensing and responding to host stimuli. The large families of TMKs described here could serve in both biological sensing and antigenic variation. The distinct extracellular and kinase domains of the TMKs suggest that each TMK may sense or interact with different host factors and cause a distinct signaling event in response to that environmental cue.

**ACKNOWLEDGMENTS**

We thank Aaron J. Mackey and William R. Pearson for guidance with the bioinformatics analyses and Brendan Loftus and Neil Hall for access to the *E. histolytica* genome sequencing project data at the TIGR and Sanger sequencing centers. Barbara Mann provided the Sp1 sequence.

This study was supported by NIH grant AI26649 to W.A.P. B.D. was supported by the Biomolecular Research Facility of the University of Virginia. T.N. was supported by a grant for Precursory Research for Embryonic Science and Technology, Japan Science and Technology Agency, a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science and Technology of Japan (16017307, 16044259, 15590378), and a grant from the Japan Health Sciences Foundation.

**REFERENCES**

1. Adam, R. D., A. Aggarwal, A. A. Lai, V. F. Cruz, T. McCutchan, and T. E. Nash, 1988. Antigenic variation of a cysteine-rich protein in *Giardia lamblia*. J. Exp. Med. 167:109–118.

2. Ali, V., T. Hashimoto, Y. Shigeta, and T. Nozaki, 2004. Molecular and biochemical characterization of α-phosphoglycerate dehydrogenase from *Entamoeba histolytica*. J. Biochem. 135:649–656.

3. Brinkworth, R. I., R. A. Breinl, and B. Kobe, 2003. Structural basis and prediction of substrate specificity in protein serine/threonine kinases. Prot. Sci. 12:2420–2431.

4. Chavez-Rios, R., and M. Vargas-Mejia, 2000. Isolation and identification of a 150-kDa surface antigen of *Entamoeba histolytica*. Biochem. J. 348:533–538.

5. Cheng, X. J., H. Tsukamoto, Y. Kaneda, and H. Tachibana, 1991. Chitin synthase in encysting *Entamoeba histolytica*. J. Exp. Med. 167:109–118.

6. Haque, R., P. Duggal, I. K. M. Ali, M. B. Hossain, D. R. B. Sack, B. M. Farr, T. H. Beaty, and W. A. Petri, Jr., 2002. Invasion and acquired resistance to amebicidal compounds in *Entamoeba histolytica*. J. Infect. Dis. 183:1787–1795.

7. Hardie, D. G., 1999. Protein phosphorylation: a practical approach, 2nd ed. Oxford University Press, Oxford, United Kingdom.

8. Haug, K., T. A. Jatkoe, C. R. Stumpf, J. Lu, J. D. Thomas, and S. J. Leibler, 2001. Development of a gene inactivation system for *Bacteroides forsythus*. Construction and characterization of a BspA mutant. Infect. Immun. 69:4066–4069.

9. Huber, M., L. Garfinkel, C. Gütler, D. Mirelman, R. Revel, and E. Rozenblat, 1988. Nucleotide sequence analysis of *Entamoeba histolytica* fimbriae. Mol. Biochem. Parasitol. 33:27–33.

10. Jave, M., W. S. Modi, G. A. Ricca, R. Mudd, I. M. Chiu, S. J. O’Brien, and W. N. Drohan, 1987. Isolation of a DNA clone for the human lamin-B1 chain and its gene localization. Am. J. Hum. Genet. 41:605–615.

11. Kraemer, S. M., and J. D. Smith, 2003. The intestinal *Entamoeba histolytica* genome. *Parasitol. Trans.* 200:741–748.

12. Laughlin, R., C. G. Mcgegan, R. P. Powell, B. H. Welter, and L. A. Temesvari, 2004. Involveent of raft-like plasma membrane domains of *Entamoeba histolytica* in pinoysis and adhesion. Infect. Immun. 72:5349–5357.

13. Leippe, M., and H. J. Müller-Eberhard, 1994. The pore-forming peptide of *Entamoeba histolytica*, the protozoan parasite causing human amoebiasis. Toxicon 32:75–81.

14. Lucero, H. A., and P. W. Robbins, 2004. Lipid raft-protein association and the regulation of protein activity. Arch. Biochem. Biophys. 426:208–224.

15. Mai, Z., and J. Samuelson, 1998. A new gene family (*arel*) encodes asparagine-rich *Entamoeba histolytica* antigens, which resemble the amebic vaccine candidate serine-rich *E. histolytica* protein. Infect. Immun. 66:353–355.

16. Nakashima, H., T. Sato, T. Iwatsuki, T. Nishimura, and J. Samuelson, 1991. Identification of a cysteine-rich galactose-specific lectin of *Entamoeba histolytica*. Proc. Natl. Acad. Sci. USA 88:5348–5352.

17. Nash, T. E., 2002. Surface antigenic variation in *Giardia lamblia*. Mol. Microbiol. 45:585–590.
