J. Euk. Microbiol., 43(6), 1996 pp. 468–474

Linked Genes for Calmodulin and E2 Ubiquitin-Conjugating Enzyme in *Trichomonas vaginalis*

PATRICK J. KEELING,[1] AMANDA L. DOHERTY-KIRBY,[2] EVELYN M. TEH and W. FORD DOOLITTLE

Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia, Canada, B3H 4H7

ABSTRACT. In searching the genomes of early-diverging protists to study whether the possession of calmodulin is ancestral to all eukaryotes, the gene for calmodulin was identified in *Trichomonas vaginalis*. This flagellate is a member of the Parabasalia, one of the earliest lineages of recognized eukaryotes to have diverged. This sequence was used to isolate a homologous 1.250-kb fragment from the *T. vaginalis* genome by inverse polymerase chain reaction. This fragment was also completely sequenced and shown to contain the 3′ end of the single-copy calmodulin gene and the 3′ end of a gene encoding a protein with high similarity to E2 ubiquitin-conjugating enzymes, a family which has previously only been identified in animals, plants, and fungi. Phylogenetic analysis of 50 members of the E2 family distinguishes at least nine separate subfamilies one of which includes the *T. vaginalis* E2-homologue and an uncharacterized gene from yeast chromosome XII.

Supplementary key words. Molecular phylogeny, evolution, trichomonad, UBC.

THERE are numerous proteins, biochemical pathways, and cytological features that are widely thought of as being present in all eukaryotes but absent from any prokaryote. These characters are used to distinguish and sometimes define the two cellular types, but this must be done with care as an increasing number of such features that were once thought to be diagnostically useful have turned out to be more complex in their distribution. In some cases "eukaryotic" qualities have been found in prokaryotes as well. This is particularly true of the archaebacteria, which are the closest prokaryotic relatives to the eukaryotes, and are now known to share numerous molecular traits with their nucleated sisters [reviewed in 17]. On the other hand, there are a number of traits that are assumed to be ancestral in all eukaryotes but have not been examined in protist lineages that diverged early in eukaryotic evolution. For example, the mitochondrion is often assumed to be present in all eukaryotes, but the lack of mitochondria in several early diverging protists makes it difficult to distinguish whether it originated later than is assumed, or was lost in these lineages. Without specific information regarding such characters, any reconstruction of the events which took place around the origin of eukaryotes may be somewhat haphazard, and it becomes difficult even to formulate a complete definition of the differences between prokaryotic and eukaryotic cellular physiology.

The same arguments made for the mitochondrion can also be made for simpler molecular processes, and even individual proteins. Little is known about transcription or translation in the earliest-diverging protists, and there is practically nothing known about their DNA replication systems, cell cycle, recombination, intracellular signaling or protein turnover mechanisms. One way to approach this problem is to survey these taxa for the presence of genes encoding proteins involved in these systems, an approach we describe here for calmodulin.

Calmodulin is a member of the EF-hand family, proteins defined by the presence of one or more short folds which bind calcium with a very high specificity [12, 27]. EF-hand proteins are found in both prokaryotes and eukaryotes [30], but calmodulin has only been directly identified in eukaryotes where it is responsible for regulating a number of physiological functions by activating other proteins in response to changes in the local concentration of calcium ions [for review see 19, 21]. To determine if calmodulin-based regulation is ancestral in eukaryotes, we have sought to identify genes encoding calmodulin from early-diverging eukaryotic lineages. In this report the calmodulin gene was isolated from *Trichomonas vaginalis*, a human pathogen belonging to the Parabasalia [1, 11]. This gene was found to be linked to another interesting gene, a member of E2 ubiquitin-conjugating enzyme family. These enzymes play a central role in the conjugation of ubiquitin to target proteins in numerous processes including proteasome-dependent proteolysis, DNA repair, cell cycle control, and membrane translocation [reviewed in 14]. Ubiquitin-conjugation is another process which is thought to be ancestral in all eukaryotes, but the actual conjugation machinery has only been studied in animals, plants, and fungi. The identification of this E2 enzyme in *T. vaginalis* shows that this pathway did originate considerably earlier in evolution.

MATERIAL AND METHODS

Strains and general molecular techniques. All molecular manipulations were carried out using *E. coli* DH5αF′ which was grown on LB agar and in LB broth under ampicillin selection. Genomic DNA from *T. vaginalis* NIH-C1 was generously provided by M. Müller. Polymerase chain reaction (PCR) products were separated by aragose gel electrophoresis and isolated by freezing crushed gel slices in an equal volume of TE (10 mM TrisHCl, 1 mM EDTA, pH 8.0) and two volumes of phenol at −70° C followed by centrifugation and ethanol precipitation of the aqueous phase. Purified products were cloned using the TA cloning vector, pCRII (Invitrogen, San Diego, CA). Southern transfers were carried out to the manufacturer's specifications (DuPont, Boston, MA). Blots were hybridized at 65° C in 10% dextran sulfate, 1 M NaCl, 1% SDS, and washed twice at room temperature in 2× SSC (0.15 M NaCl, 0.015 M sodium citrate), twice at 65° C in 1% SDS, 2× SSC, and twice in 0.1× SSC at room temperature. Sequencing of double stranded templates was performed by dideoxy termination using the T7 polymerase (Pharmacia, Uppsala, Sweden). All clones were sequenced in triplicate to detect errors introduced by Taq polymerase.

PCR and inverse PCR amplification conditions. All amplification reactions consisted of 35 cycles of 92° C denaturing for 1 min, 55° C annealing for 2 min, and 72° C extension for 2 min, all followed by a 5-min extension at 72° C in a 100-μl volume containing 10 mM TrisHCl, 50 mM KCl, 1.5 mM $MgCl_2$, 0.1% Triton X100, and 0.2 mg/ml BSA at pH 9. Calmodulin genes were amplified using the degenerate primers TGGGGTACCCAAGATATGATHAAYGARGT and GGACTAGTATCATTTCRTCNACYTCYTC. For inverse PCR, 0.8 μg of genomic *T. vaginalis* DNA was digested to completion with *Hin*dIII, purified by phenol extraction and ethanol precipitated. Digested DNA was ligated in a 100-μl volume to promote intramolecular reactions. The ligation was again purified and precipitated as before and one tenth was used in PCR reactions as described with CCTTCATCTGACGGTTCATC and

[1] To whom correspondence should be addressed. Telephone: 902-494-3569; Fax: 902-494-1355.

[2] Present address: *Department of Chemistry, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.*

CAAGATTACAGCTGCAGAGC, exact-match primers designed to prime DNA synthesis in opposite directions.

Sequence analysis. Amino acid sequences of other calmodulin and E2 ubiquitin-conjugating enzymes were retrieved from existing databases and aligned using the PileUp program from the GCG package [4]. Alignment of the E2 ubiquitin-conjugating enzyme was edited by eye and the conserved core used to infer phylogenetic trees to determine gene families and their relationships to one another. Unweighted parsimony trees were searched using PAUP version 3.1.1 [31] by conducting 10 random addition heuristic searches with tree bisection and reconnection. Distance trees were inferred by performing neighbor-joining analysis on corrected distances based on the PAM 250 substitution matrix using the NEIGHBOR and PROTDIST programs from the PHYLIP 3.57c package [6]. The statistical confidence of nodes in the distance trees was estimated by conducting 100 bootstrap resampling replicates.

RESULTS

Isolation of the *Trichomonas vaginalis* calmodulin gene. Genomic DNA from *T. vaginalis* was used as a template in PCR amplification reactions with primers specific for all known calmodulin sequences. A single product of the expected size was isolated and three individual clones sequenced. All were shown to encode identical open reading frames with an extremely high sequence similarity to calmodulin (on the amino acid level 70% identical to human calmodulin).

This PCR fragment was used as a probe against a Southern blot of genomic DNA from *T. vaginalis* (Fig. 1). In DNA cut with *Eco*RI, *Hin*dIII and *Sau*3A a single fragment hybridized to the probe, suggesting that it recognizes a single-copy locus (this is also supported by the agreement between this pattern and the map of the genomic clone described below). The largest of these fragments, a 1.25-kb *Hin*dIII fragment, was sought by inverse PCR amplification using primers based on the known sequence of the small fragment, and circularized *Hin*dIII-digestion products of *T. vaginalis* genomic DNA as a template. A product of the expected size was isolated, cloned, and three individual copies completely sequenced.

The calmodulin gene was found to lay at one extreme end of the fragment, truncated at the amino terminus by a *Hin*dIII site at what corresponds to codon 15 of the majority of known calmodulin genes (Fig. 2). The 3′ end of the gene is marked by a termination codon at exactly the same position as most known calmodulin homologues, and the sequence is extremely conserved throughout the length of the gene (over the 134 amino acid positions, the *T. vaginalis* calmodulin is identical to the human gene at 89 positions).

Downstream of the termination codon there are 461 bp of apparently noncoding DNA which is extremely AT-rich (78% AT as opposed to an AT content of 61% in coding regions), followed by another open reading frame on the opposite strand. This open reading frame encodes a sequence with high similarity to E2 ubiquitin-conjugating (UBC) enzymes.

***Trichomonas vaginalis* UBC1, a member of the E2 ubiquitin-conjugating enzyme family**. This second open reading frame is unfortunately also truncated at the amino terminus by the other *Hin*dIII site, but over the 128 codons which it does cover, it shares many similarities with other E2 ubiquitin-conjugating enzymes. These enzymes have until now only been found in animals, plants, and fungi where they make up a large multigene family. The best sampling of diverse *UBC* genes is from *S. cerevisiae,* where 12 individual members have been identified. In Fig. 3 these 12 genes are aligned to the *T. vaginalis* sequence (named *Tv*UBC1 to conform to one existing nomenclature [28]). From the alignment the similarity of the *T.*

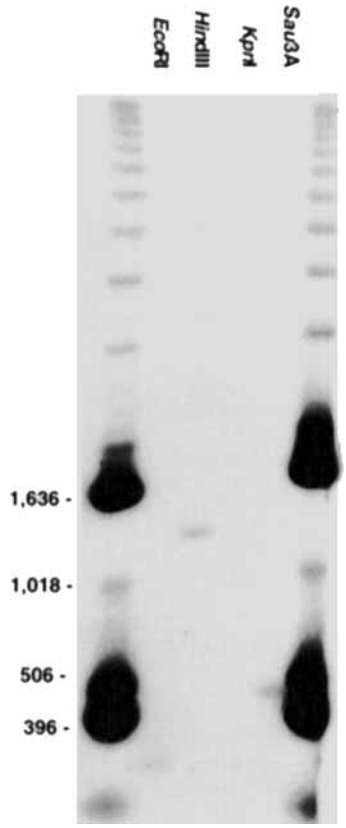

Fig. 1. Southern blot of *T. vaginalis* genomic DNA probed with calmodulin PCR fragment. The single band in the *Hin*dIII lane corresponds to a 1.250-kb fragment which was subsequently isolated by inverse PCR and sequenced. The small *Eco*RI and *Sau*3A fragments which hybridized to the calmodulin probe were calculated to be congruent with the restriction map of the 1,250-bp *Hin*dIII clone. The *Kpn*I lane yielded only a very large, faintly hybridizing fragment.

vaginalis sequence to other homologues can be seen to concentrate around the otherwise highly conserved domains, especially surrounding the catalytic cysteine residue at position 127 (boxed in Fig. 3), which forms the actual thioester bond to ubiquitin [14, 28, 29].

Different types of E2 ubiquitin-conjugating enzymes have been classified by different physical properties and activities, including the presence of a long, often acidic carboxy terminal extension, the ability to ubiquitinate histone, and the requirement of E3 ubiquitin-protein ligase [reviewed in 14, 16, 23]. However, the strong correlation between having a carboxy terminal extension and being able to ubiquitinate histones without E3 suggests that they may really be testing the same characteristic in different ways. We sought, therefore, to make a phylogenetic classification of E2 enzymes and in this way determine the relationship of the *T. vaginalis* sequence to other E2 sequences.

The alignment in Fig. 3 is a sample of a larger one composed of 50 *UBC* genes from animals, plants, and fungi which was used to infer phylogenetic trees by both distance and parsimony methods. A distance tree based on 154 positions of the conserved UBC core is shown in Fig. 4 where suggested subfamilies are indicated. Each subfamily comprises a group of sequences separated from other subfamilies by a significant branch. Also, the branching order of taxa within subfamilies

```
Trichomonas   ---------- ---------- -AFNIFDKDG DGRITAKELG TVMRSLGQNP SEAELQDMIN EIDLDGNGTI EFDEFLYMMN
N.gruberi     MSREAISNNE LTEEQIAEFK EAFSLFDKDG DGTITTSELG TVMRSLGQNP TEAELHDMIN EVDADGNGTI DFTEFLTMMA
Solanum       ......MAEQ LTEEQIAEFK EAFSLFDKDG DGCITTKELG TVMRSLGQNP TEAELQDMIS EADADQNGTI DFPEFLNLMA
Oryza         ......MADQ LTDDQIAEFK EAFSLFDKDG DGCITTKELG TVMRSLGQNP TEAELQDMIN EVDADGNGTI DFPEFLNLMA
T.cruzi       ......MADQ LSNEQISEFK EAFSLFDKDG DGTITTKELG TVMRSLGQNP TEAELQDMIN EVDQDGSGTI DFPEFLTLMA
Euglena       ......MAEA LTHEQIAEFK EAFSLFDKDG DGTITTKELG TVMRSLGQNP TEAELQDMIN EVDQDGSGTI DFPEFLTLMS
Plasmodium    ......MADK LTEEQISEFK EAFSLFDKDG DGTITTKELG TVMRSLGQNP TEAELQDMIN EIDTDGNGTI DFPEFLTLMA
Stylonychia   ......MADN LTEEQIAEFK EAFSLFDKDG DGTITTKELG TVMRSLGQNP TEAELQDMIN EVDADGNGTI DFPEFLSLMA
Tetrahymena   ......MADQ LTEEQIAEFK EAFSLFDKDG DGTITTKELG TVMRSLGQNP TEAELQDMIN EVDADGNGTI DFPEFLSLMA
Drosophila    ......MADQ LTEEQIAEFK EAFSLFDKDG DGTITTKELG TVMRSLGQNP TEAELQDMIN EVDADGNGTI DFPEFLTMMA
Homo          ......MADQ LTEEQVTEFK EAFSLFDKDG DGCITTRELG TVMRSLGQNP TEAELRDMMS EIDRDGNGTV DFPEFLGMMA
Aspergillus   ......MADS LTEEQVSEYK EAFSLFDKDG DGQITTKELG TVMRSLGQNP SESELQDMIN EVDADNNGTI DFPEFLTMMA
Neurospora    ......MADS LTEEQVSEFK EAFSLFDKDG DGQITTKELG TVMLSLGQNP SESELQDMIN EVDADNNGTI DFPEFLTMMA
Saccharomyces ......MSSN LTEEQIAEFK EAFALFDKDN NGSISSSELA TVMRSLGLSP SEAEVNDLMN EIDVDGNHQI EFSEFLALMS

Trichomonas   RQMKEGDTEE EIKDAFRVFD KDGDGKITAA ELAHIMKNLG EPLTQEEVDE MIAQADTNKD GIIDYGEFVH LMLTS*
N.gruberi     KKMKDTDNEE EIKEAFKVFD KDGNGFISAQ ELRHVMCNLG EKLTDEEVDE MIREADIDGD NQINYTEFVK MMMQK*
Solanum       RKMKDTDSEE ELKEAFKVFD KDQNGFISAA ELRHVMTNLG EKLTDEEVDE MIREADIDGD GQVNYEEFVR MMLAK*
Oryza         RKMKDTDSEE ELKEAFRVFD KDQNGFISAA ELRHVMTNLG EKLTDEEVDE MIREADVDGD GQINYEEFVK VMMAK*
T.cruzi       RKMQDSDSEE EIKEAFRVFD KDGNGFISAA ELRHVMTNLG EKLTDEEVDE MIREADVDGD GQINYEEFVK MMMSK*
Euglena       RKMHDTDTEE EIKEAFRVFD KDGNGFISAA ELRHVMTNLG EKLTDEEVDE MIREADVDGD GQINYEEFVK MMMSK*
Plasmodium    RKLKDTDTEE ELIEAFRVFD RDGDGYISAD ELRHVMTNLG EKLTNEEVDE MIREADIDGD GQINYEEFVK MMIAK*
Stylonychia   RKMKDTDTEE ELVEAFKVFD RDGNGLISAA ELRHVMTNLG EKLTDEEVDE MIREADVDGD GHINYEEFVR MMMAK*
Tetrahymena   RKMKDTDTEE ELIEAFKVFD RDDNDLISAA ELRHVMTNLG EKLTDEEVDE MIREADIDGD GHINYEEFVR MMMAK*
Drosophila    RKMKDTDSEE EIREAFRVFD KDGNGFISAA ELRHVMTNLG EKLTDEEVDE MIREADIDGD GQVNYEEFVT MMTSK*
Homo          RKMKDTDNEE EIREAFRVFD KDGNGFVSAA ELRHVMTRLG EKLSDEEVDE MIRAADTDGD GQVNYEEFVR VLVSK*
Aspergillus   RKMKDTDSEE EIREAFKVFD RDNNGFISAA ELRHVMTSIG EKLTDDEVDE MIREADQDGD GRIDYNEFVQ LMMQK*
Neurospora    RKMKDTDSEE EIREAFKVFD RDNNGFISAA ELRHVMTSIG EKLTDDEVDE MIREADQDGD GRIDYNEFVQ LMMQK*
Saccharomyces RQLKSNDSEQ ELLEAFKVFD KNGDGLISAA ELKHVLTSIG EKLTDAEVDD MLREVS.DGS GEINIQQFAA LLS.K*
```

Fig. 2. Calmodulin sequence from *Trichomonas* aligned with those of representative eukaryotes. Length heterogeneity at the amino terminus and gaps within the alignment are indicated by dots (·····), missing data by dashes (– – –), and termination codons as asterisks (*).

does not strongly contradict what is known of the organismal phylogeny, for instance the animals, plants, and fungi are not interspersed among one another. Different datasets varying in the inclusion of positions of ambiguous alignment were also examined (most excluding the short region missing from *Tv*UBC1), and in all cases the subfamilies shown were conserved and the order between families was slightly variable. These subfamilies are also supported by parsimony analysis which resulted in 528 equally parsimonious trees, the strict consensus of which includes all these groups. All analyses, however, resulted in low levels of significance for most inter-subfamily nodes, casting considerable doubt on any meaning that this branching order might have.

In Fig. 4 the *Trichomonas* enzyme is affiliated with a *S. cerevisiae* open reading frame known only from the sequence of chromosome XII. This enzyme has not been functionally characterized and its role in the cell is completely unknown.

Of the clearly cohesive groups, two (represented by *Sc*UBC8 and *Sc*UBC1 proteins) are composed entirely of enzymes with carboxy-terminal extensions, and several families are totally void of this type of enzyme. However, two subfamilies comprise a mixture of enzymes with and without carboxy-terminal extensions. This is particularly obvious in the group which includes *Sc*UBC3 and *Sc*UBC7; this family shares a conserved insertion in the UBC core, but only *Sc*UBC3 and *Hs*CDC34 have carboxy extensions. In general the enzymes with carboxy-terminal extensions are not all directly related to one another, which implies that classification based on these physical properties in some cases does not reflect relatedness.

DISCUSSION

Among the list of characteristics that seem to define the eukaryotes, one is the use of calmodulin as a receptor of intracellular calcium. In an attempt to clarify whether calmodulin was present in the ancestor of extant eukaryotes, we have sequenced a fragment of the calmodulin gene from the parabasalian, *Trichomonas vaginalis,* and found it to be extremely similar in sequence to homologues from other eukaryotes.

There is biochemical evidence for the presence of calmodulin in another protist taxon which perhaps diverged even earlier, the diplomonad *Giardia lamblia,* where a protein with many similar characteristics has been observed but no sequence reported [22]. This together with the present observation of a highly conserved calmodulin gene in the deeply branching *T. vaginalis* suggests that calmodulin was likely a feature of the common ancestor of extant eukaryotes. There has also been a report of a protein with traits resembling calmodulin in the archaebacterium *Halobacterium halobuim* [24], but once gain no sequence has been identified and these organisms lack an endomembrane system. It appears, moreover, that there is no eu-

→

Fig. 3. Alignment of *Tv*UBC1 amino acid sequence with that of twelve known *UBC* genes from *S. cerevisiae.* Genes are named where possible by one conventional nomenclature where the first two letters are the organisms initials (in this case *Sc* is *S. cerevisiae*) followed by UBC for ubiquitin-conjugating enzyme, and a number to distinguish paralogous enzymes from the same genome. Two homologues in *S. cerevisiae,* which are only known from genome sequencing, are also shown: L2142.3 and YD6652.4 (GenBank accession numbers U17247 and Z50111, respectively). Length heterogeneity is indicated by spaces, gaps in the alignment are shown as dots (·····), missing data in *T. vaginalis* as dashes (– – –), and termination codons as asterisks (*). Four of the *S. cerevisiae* proteins—UBC1, 3, 6, and 8—have considerable carboxy terminal extensions, which are not shown, but the length is given in square brackets. The cysteine residue that forms the thioester bond with ubiquitin is boxed.

```
TvUBC1   ---------- ---------- ---------- ---------- ---------K LSFPEAN...
ScUBC8                                MSSSKR RIETDVMKLL MSDHQVDLIN DS........
L2142.3  MLKLRQLQKK KQKENENSSS IQPNLSAARI RLKRDLDSLD LPPTVTLNVI TSPDSADR..
ScUBC7                              MSKTAQK RLLKELQQL. .IKDSPPGIV AGPKSEN...
ScUBC3                            MSSRKSTASS LLLRQYRELT DPKKAIPSFH IELEDDS...
ScUBC9                              MSSLCLQ RLQEERKKW. .RKDHPFGFY AKPVKKAD..
ScUBC2                              MSTPARR RLMRDFKRM. .KEDAPPGVS ASPLPD....
YD6652.4                            MASLPK RIIKETEKL. .VSDPVPGIT AEPHDD....
ScUBC4                                MSSSK RIAKELSDL. .ERDPPTSCS AGPVGD....
ScUBC5                                MSSSK RIAKELSDL. .GRDPPASCS AGPVGD....
ScUBC1                                MSRAK RIMKEIQAV. .KDDPAAHIT LEFVSE....
ScUBC10             MPNFWILENR RSYTSDTCMS RIVKEYKVIL KTLASDDPIA NPYRGIIESL
ScUBC6                              MATKQAHK RLTKEYKLM. .VENPPPYIL ARP.......

TvUBC1   ......NIQK FIVLIQ..PA TGHWRAGKFE FEFTIPDDWP ITKPDIKILT ..RVWHPNID
ScUBC8   .......MQE FHVKFLG.PK DTPYENGVWR LHVELPDNYP YKSPSIGFVN ..KIFHPNID
L2142.3  .......SQS PKLEVIVRPD EGYYNYGSIN FNLDFNEVYP IEPPKVVCLK ..KIFHPNID
ScUBC7   ......NIFI WDCLIQG.PP DTPYADGVFN AKLEFPKDYP LSPPKLTFTP ..SILHPNIY
ScUBC3   ......NIFT WNIGVMVLNE DSIYHGGFFK AQMRFPEDFP FSPPQFRFTP ..AIYHPNVY
ScUBC9   ...GSMDLQK WEAGIPG.KE GTNWAGGVYP ITVEYPNEYP SKPPKVKFPA ..GFYHPNVY
ScUBC2   ......NVMV WNAMIIG.PA DTPYEDGTFR LLLEFDEEYP NKPPHVKFLS ..EMFHPNVY
YD6652.4 ......NLRY FQVTIEG.PE QSPYEDGIFE LELYLPDDYP MEAPKVRFLT ..KIYHPNID
ScUBC4   ......DLYH WQASIMG.PA DSPYAGGVFF LSIHFPTDYP FKPPKISFTT ..KIYHPNIN
ScUBC5   ......DLYH WQASIMG.PS DSPYAGGVFF LSIHFPTDYP FKPPKVNFTT ..KIYHPNIN
ScUBC1   .....SDIHH LKGTFLG.PP GTPYEGGKFV VDIEVPMEYP FKPPKMQFDT ..KVYHPNIS
ScUBC10  NPIDETDLSK WEAIISG.PS DTPYENHQFR ILIEVPSSYP MNPPKISFMQ .NNILHCNVK
ScUBC6   ...NEDNILE WHYIITG.PA DTPYKGGQYH GTLTFPSDYP YKPPAIRMIT PNGRFKPNT.

TvUBC1   E.NGAV C LSILRDN..... ........YL ATLSISQFVA G.LQYLFIEP NPNSP......
ScUBC8   IASGSI C LDVINST..... ........WS PLYDLINIVE WMIPGLLKEP NGSDP......
L2142.3  L.KGNV C LNILRED..... ........WS PALDLQSIIT G.LLFLFLEP NPNDP......
ScUBC7   P.NGEV C ISILHSPGDDPN MYELAEERWS PVQSVEKILL S.VMSMLSEP NIESG......
ScUBC3   R.DGRL C ISILHQSG.DPM TDEPDAETWS PVQTVESVLI S.IVSLLEDP NINSP......
ScUBC9   P.SGTI C LSILNEDQD... ........WR PAITLKQIVL G.VQDLLDSP NPNSP......
ScUBC2   A.NGEI C LDILQNR..... ........WT PTYDVASILT S.IQSLFNDP NPASP......
YD6652.4 R.LGRI C LDVLKTN..... ........WS PALQIRTVLL S.IQALLASP NPNDP......
ScUBC4   A.NGNI C LDILKDQ..... ........WS PALTLSKVLL S.ICSLLTDA NPDDP......
ScUBC5   S.SGNI C LDILKDQ..... ........WS PALTLSKVLL S.ICSLLTDA NPDDP......
ScUBC1   SVTGAI C LDILKNA..... ........WS PVITLKSALI S.LQALLQSP EPNDP......
ScUBC10  SATGEI C LNILKPEE.... ........WT PVWDLLHCVH A.VWRLLREP VCDSP......
ScUBC6   ....RL C LSMSDYHPDT.. ........WN PGWSVSTILN G....LLSFM TSDEATTGSIT

TvUBC1   .LNTEAA.TM FKNEPAKFQE KVDDYIEKYC PK*
ScUBC8   .LNNEAA.TL QLRDKKLYEE KIKEYIDKYA TKEKYQQMFG GDNDSD   [55 AA Tail]
L2142.3  .LNKDAA.KL LCEGEKEFAE AVRLTMSGGS IEHVKYDNIV SP*
ScUBC7   .ANIDAC.IL WRDNRPEFER QVKLSILKSL GF*
ScUBC3   .ANVDAA.VD YRKNPEQYKQ RVKMEVERSK QDIPKGFIMP TSESAY  [111 AA Tail]
ScUBC9   .AQEPAW.RS FSRNKAEYDK KVLLQAKQYS K*
ScUBC2   .ANVEAA.TL FKDHKSQYVK RVKETVEKSW EDDMDDMDDD DDDDDDDDDDEAD*
YD6652.4 .LANDVA.ED WIKNEQGAKA KAREWTKLYA KKKPE*
ScUBC4   .LVPEIA.HI YKTDRPKYEA TAREWTKKYA V*
ScUBC5   .LVPEIA.QI YKTDKAKYEA TAKEWTKKYA V*
ScUBC1   .QDAEVA.QH YLRDRESFNK TAALWTRLYA SETSNGQKGN VEESDL   [50 AA Tail]
ScUBC10  .LDVDIGNII RCGDMSAYQG IVKYFLAERE RINNH*
ScUBC6   TDEGDAANTG DETEDPFTKA AKEKVISLEE ILDPEDRIRA EQALRS   [78 AA Tail]
```

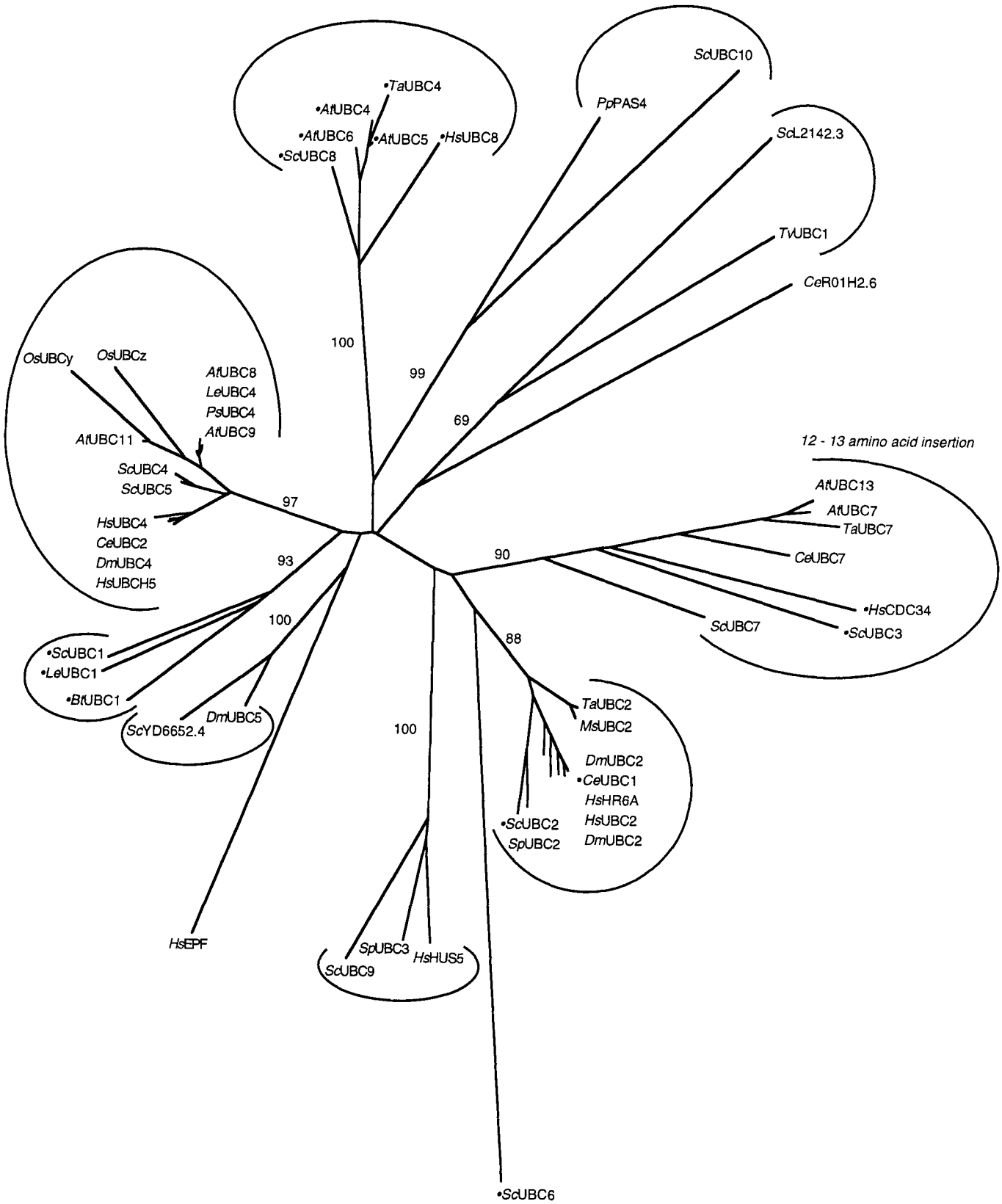

Fig. 4. Neighbor-joining tree of 50 UBC proteins with bootstrap percents shown for major nodes over 50%. Nine significant clusters are defined, each of which is bracketed. The group represented by *Sc*UBC3/7 is also defined by a shared 12–13 amino acid insertion unique to this group. Sequences which have carboxy terminal extensions are identified by a dot (•) preceding their name. The nomenclature is the same as in Fig. 3, consisting of the initials of the organism (*Tv, T. vaginalis*; *Sc, S. cerevisiae*; *Sp, Schizosaccharomyces pombe*; *Pp, Pichia pastoris*; *Ce, Caenorhabditis elegans*; *Dm, Drosophila melanogaster*; *Hs, Homo sapiens*; *Bt, Bos taurus*; *At, Arabidopsis thaliana*; *Le, Lycopersicon esculentum*;

bacterial orthologue as no calmodulin-like sequence appears in the genomes of either *Haemophilus influenzae,* or *Mycoplasma genitalium,* the only two eubacterial genomes which have been fully sequenced and released to public databases [7, 8].

The finding of E2 ubiquitin-conjugating enzyme in *T. vaginalis* also helps to date the advent of an important system in eukaryotes, but is quite a bit more complex in its implications. Ubiquitin-conjugation is predominantly involved in degradation of misfolded or short-lived proteins, but is also observed in chromatin structure, DNA repair, cell-cycle control, membrane translocation, and a host of other cellular activities [3, 9, 10, 13, 15, 26]. In the conjugation pathway [reviewed in 14], ubiquitin is first converted to an adenylated intermediate by E1 ubiquitin-activating enzyme, which proceeds to covalently bind the ubiquitin molecule through a thioester linkage. The ubiquitin moiety is subsequently transferred to E2 ubiquitin-conjugating enzyme by transesterification. This enzyme may then catalyse the formation of an isopeptide bond between ubiquitin and the target protein, in some cases through an E3-ubiquitin thioester intermediate [25]. There are numerous families of conjugating enzymes, and the substrate choice for ubiquitination is to some extent specified by the different physical characteristics and activities of the E2 involved. Of the known *UBC* genes, the *T. vaginalis* E2 enzyme is most similar in sequence to L2142.3, an uncharacterized gene on chromosome XII of *S. cerevisiae,* so little can be inferred about the function of *Tv*UBC1. Although this tree cannot be rooted, the relatively firm relationship of *Tv*UBC1 with a particular *S. cerevisiae* sequence gives some indication that *T. vaginalis* likely also has multiple E2 ubiquitin-conjugating enzymes. Otherwise the root of the tree would have to lie in the branch leading to *T. vaginalis,* and *Tv*UBC1 would be unlikely to have a specific affinity to any particular subfamily. This implies that some distribution of function among UBC proteins may be found in early diverging eukaryotes, but this cannot be known for sure until more *UBC* genes from these taxa are identified and assigned to other subfamilies.

Ubiquitin itself has also been identified in a wide variety of eukaryotes, among which are several deep-branching protists including *T. vaginalis* and *G. lamblia* [18, 20], and in *Trypanosoma cruzi* where it is actually linked to calmodulin [2]. The presence of E2 in *T. vaginalis* is the first evidence that the conjugation machinery was also in place early in eukaryotic evolution, but exactly how early these proteins first arose remains uncertain. There is evidence that ubiquitin might predate the prokaryote-eukaryote divergence as ubiquitin peptide sequences have been identified in archaebacteria and eubacteria [5, 33], and the genes for the proteasome have been identified and well characterized in the archaebacterium, *Thermoplasma acidophilum* [32]. However, there still have been no prokaryotic ubiquitin *genes* identified, and like calmodulin, sequences for neither ubiquitin nor the ubiquitin-conjugating enzymes appear in the genomes of *H. influenzae,* or *M. genetalium* [7, 8]. Until such sequences are produced, or until additional prokaryotic genome sequences rule out their presence, the date for the advent of both the calmodulin-based regulation system and the ubiquitin-conjugating pathway may only be said to be no later than very early in the evolution of eukaryotes.

ACKNOWLEDGMENTS

This work was supported by operating grant MT4467 from the Medical Research Council of Canada (MRC). PJK is the recipient of an MRC studentship. WFD is a Fellow of the Canadian Institute for Advanced Research. We would like to thank M. Müller for providing *T. vaginalis* genomic DNA, and R. D. Virestra and S. van Nocker for their interest and helpful advice. The new sequence has been deposited in GenBank under the accession number U38786.

LITERATURE CITED

1. Cavalier-Smith, T. 1993. The kingdom Protozoa and its 18 phyla. *Microbiol. Rev.,* **57**:953–994.

2. Chung, S. H. & Swindle, J. 1990. Linkage of the calmodulin and ubiquitin loci in *Trypanosoma cruzi. Nucl. Acids Res.,* **18**:4561–4569.

3. Davie, J. R. & Murphy, L. C. 1990. Level of ubiquitinated histone HB2 in chromatin is coupled to ongoing transcription. *Biochemistry,* **29**:4752–4757.

4. Devereux, J., Haeberli, P. & Smithies, O. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.,* **12**: 387–395.

5. Durner, J. & Börger, P. 1995. Ubiquitin in the prokaryote *Anabaena variabilis. J. Biol. Chem.,* **270**:3720–3725.

6. Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package), J. Felsenstein, University of Washington, Seattle.

7. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J.-F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C. A., Gocayne, J. D., Scott, J. D., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O. & Venter, J. C. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science,* **269**:496–512.

8. Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., Fritchman, J. L., Weidman, J. F., Small, K. V., Sandusky, M., Fuhrmann, J. L., Nguyen, D. T., Utterback, T. R., Saudek, D. M., Phillips, C. A., Merrick, J. M., Tomb, J.-F., Dougherty, B. A., Bott, K. F., Hu, P.-C., Lucier, T. S., Peterson, S. N., Smith, H. O., Hutchison, C. A., III & Venter, J. C. 1995. The minimal gene complement of *Mycoplasma genitalium. Science,* **270**:397–403.

9. Goebl, M. G., Yochem, J., Jentsch, S., McGrath, J. P., Varshavsky, A. & Byers, B. 1988. The yeast cell cycle gene CDC34 encodes a ubiquitin-conjugating enzyme. *Science,* **241**:1331–1335.

10. Goldknopf, I. L. & Busch, H. 1977. Isopeptide linkage between nonhistone and histone 2A polypeptides of chromosomal conjugate-protein A24. *Proc. Nat. Acad. Sci.* USA, **74**:864–868.

11. Gunderson, J., Hinkle, G., Leipe, D., Morrison, H. G., Stickel, S. K., Odelson, D. A., Breznak, J. A., Nerad, T. A., Müller, M. & Sogin, M. L. 1995. Phylogeny of trichomonads inferred from small-subunit rRNA sequences. *J. Euk. Microbiol.,* **42**:411–415.

12. Heizmann, C. W. & Hunziker, W. 1991. Intracellular calcium-binding proteins: more sites than insights. *Trends Biochem. Sci.,* **16**:98–103.

13. Holloway, S. L., Glotzer, M., King, R. W. & Murray, A. W. 1993. Anaphase is initiated by proteolysis rather than by the inactivation of a maturation-promoting factor. *Cell,* **73**:1393–1402.

14. Jentsch, S. 1992. The ubiquitin-conjugation system. *Annu. Rev. Genet.,* **26**:179–207.

15. Jentsch, S., McGrath, J. P. & Varshavsky, A. 1987. The yeast

←

Ts, Triticum aestivum; *Os, Oryza sativa*; *Ms, Medicago sativa*) followed by their number where specified by the original authors in database entries. However, not all genes have been assigned such names: exceptions and their GenBank accession numbers are: *Hs*HUS5 (U29092), *Hs*EPF (M91670), *Hs*HR6A (M74522), *Hs*CDC34 (L22005), *Ce*R01H2.6 (U00035), *Os*UBCy (U15971), *Os*UBCz (D17786), and *Pp*PAS4 (U12511).

DNA repair gene RAD6 encodes a ubiquitin-conjugating enzyme. *Nature,* **329**:131–134.

16. Jentsch, S., Seufert, W., Sommer, T. & Reins, H.-A. 1990. Ubiquitin-conjugating enzymes: novel regulators of eukaryotic cells. *Trends Biochem. Sci.,* **15**:195–198.

17. Keeling, P. J. & Doolittle, W. F. 1995. Archaea: narrowing the gap between prokaryotes and eukaryotes. *Proc. Nat. Acad. Sci.* USA, **92**:5761–5764.

18. Keeling, P. J. & Doolittle, W. F. 1995. Concerted evolution in protists: a polyubiquitin gene from *Trichomonas vaginalis. J. Mol. Evol.,* **41**:556–562.

19. Klee, C. B., Crouch, T. H. & Richman, P. G. 1980. Calmodulin. *Annu. Rev. Biochem.,* **49**:489–515.

20. Krebber, H., Wöstmann, C. & Bakker-Grunwald, T. 1994. Evidence for the existance of a single ubiquitin gene in *Giardia lamblia. FEBS Lett.,* **343**:234–236.

21. Means, A. R. & Dedman, J. R. 1980. Calmodulin—an intracellular calcium receptor. *Nature,* **285**:73–77.

22. Munoz, M. L., Weinbach, E. C., Wieder, S. C., Claggett, C. E. & Weinbach, E. C. 1987. *Giardia lamblia*: detection and characterization of calmodulin. *Exp. Parasitol.,* **63**:42–48.

23. Qin, S., Nakajima, B., Nomura, M. & Arfin, S. M. 1991. Cloning and characterization of a *Saccharomyces cerevisiae* gene encoding a new member of the ubiquitin-conjugating protein family. *J. Biol. Chem.,* **266**:15549–15554.

24. Rothärmel, T. & Wagner, G. 1995. Isolation and characterization of a calmodulin-like protein from *Halobacterium salinarium. J. Bacteriol.,* **177**:864–866.

25. Scheffner, M., Nuber, U. & Huibregtse, J. M. 1995. Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. *Nature,* **373**:81–83.

26. Sommer, T. & Jentsch, S. 1993. A protein translocation defect linked to ubiquitin conjugation at the endoplasmic reticulum. *Nature,* **365**:176–179.

27. Strynadka, N. C. & James, M. N. 1989. Crystal structures of the helix-loop-helix calcium-binding proteins. *Annu. Rev. Biochem.,* **58**: 951–998.

28. Sullivan, M. L. & Vierstra, R. D. 1991. Cloning of a 16-kDa ubiquitin carrier protein from wheat and *Arabidopsis thaliana.* Identification of functional domains by in vitro mutagenesis. *J. Biol. Chem.,* **266**:23878–23885.

29. Sung, P., Prakash, S. & Prakash, L. 1990. Mutation of cysteine-88 in the *Saccharomyces cerevisiae* RAD6 protein abolishes its ubiquitin-conjugating activity and its various biological functions. *Proc. Nat. Acad. Sci.* USA, **87**:2695–2699.

30. Swan, D. G., Hale, R. S., Dhillon, N. & Leadlay, P. F. 1987. A bacterial calcium-binding protein homologous to calmodulin. *Nature,* **329**:84–85.

31. Swofford, D. L. 1993. PAUP: Phylogenetic Analysis Using Parsimony, Vers. 3.1, Illinois Natural History Survey, Champaign, Illinios.

32. Wenzel, T. & Baumeister, W. 1993. *Thermoplasma acidophilum* proteasomes degrades partially unfolded and ubiquitin-associated proteins. *FEBS Lett.*, **326**:215–218.

33. Wolf, S., Lottspeich, F. & Baumeister, W. 1993. Ubiquitin found in the archaebacterium *Thermoplasma acidophilum. FEBS Lett.*, **326**: 42–44.

Received 12-20-95, 4-4-96, 6-7-96, 7-8-96; accepted 7-9-96.