Molecular cloning of actin genes in *Trichomonas vaginalis* and phylogeny inferred from actin sequences

Geneviève Bricheux*, Guy Brugerolle

Laboratoire de Biologie Comparée des Protistes, URA 1944, Université Blaise Pascal de Clermont-Ferrand, 63177 Aubière Cedex, France

Received 23 February 1997; revised 30 May 1997; accepted 30 May 1997

Abstract

The parasitic protozoan *Trichomonas vaginalis* is known to contain the ubiquitous and highly conserved protein actin. A genomic library and a cDNA library have been screened to identify and clone the actin gene(s) of *T. vaginalis*. The nucleotide sequence of one gene and its flanking regions have been determined. The open reading frame encodes a protein of 376 amino acids. The sequence is not interrupted by any introns and the promoter could be represented by a 10 bp motif close to a consensus motif also found upstream of most sequenced *T. vaginalis* genes. The five different clones isolated from the cDNA library have similar sequences and encode three actin proteins differing only by one or two amino acids. A phylogenetic analysis of 31 actin sequences by distance matrix and parsimony methods, using centriactin as outgroup, gives congruent trees with Parabasalida branching above Diplomonadida.

Keywords: *Trichomonas vaginalis*; Actin gene; Phylogenetic tree

1. Introduction

Actin is one of the most abundant proteins in all eukaryotic cells. In non-muscle cells, actin is involved in a variety of processes including cytoskeletal structure, cellular motility, cell surface motility, intracellular transport, cytoplasmic streaming, cytokinesis, endocytosis, exocytosis, microvillus movement and, possibly, chromosomal condensation and mitosis [1]. Actin is a conserved protein; actins from the most divergent sources still share amino acid identities in excess of 65–70%. These proteins are encoded by a multigene family in most eukaryotes, although there are some exceptions such as *Saccharyaomyces cerevisiae*, *Tetrahymena thermophila*, *Aspergillus nidulans* and *Phytophthora megasperma* where a single gene has been demonstrated.

The parabasalid protozoan *Trichomonas vaginalis* is a well known sexually transmitted infectious agent with a worldwide distribution. Upon adherence to host cells, it changes from the typical in vitro flagellate ellipsoid shape to a flattened amoeboid one, with modifications of the cytoskeleton. Extensive pseudopods form at the site of contact throughout the periphery of the organisms [2]. Membrane-membrane associations are prominent in the interactions between the host cell and the parasite, exhibiting many contact points and thin lamillipodia and filopodia on the transformed organisms. Adjacent trichomonads also present extensive membrane-membrane interactions. The morphological transfor-
mation of the parasite is proportional to the level of cytadherence and the amount of adhesins presented by the cells [3]. A previous biochemical study has shown that T. vaginalis possesses a remarkable amount of actin, with specific properties (molecular mass, acidic \( pI \), inability to bind DNase I, low affinity for phalloidin) [4]. This protein is dispersed in the cytoplasm in the swimming cells, and more concentrated in the cortical region of the cell and associated with pseudopod emission in the amoeboid forms. Actin filaments appear as a thin network and actin stress fibers have not been observed. These properties of actin may be implicit in the protein sequence itself or in its association with other proteins.

As a ubiquitous and well conserved structural protein, actin has been used to clarify the molecular phylogeny of various protist groups as well as phylogenetic relationships among plants, animals and fungi [5].

In the present paper we report on cDNA and genomic clones encoding the actin of T. vaginalis, and on comparison of actin sequences to discuss the level of emergence of T. vaginalis, a member of the Parasasala, which belongs to the earliest-diverging eu-karyotic lineages [6-9].

2. Materials and methods

2.1. Organism and culture

The T. vaginalis strain was isolated from a female patient and stored in liquid nitrogen. Cultures were grown at 37°C on tryptase/yeast extract/maltose (TYM) medium supplemented with 10% (v/v) heat-inactivated horse serum.

2.2. DNA extraction and Southern hybridization

The cells were harvested in late exponential phase by centrifugation and washed with phosphate buffered saline. High molecular mass DNA was isolated by the DEPC-Triton X-100 method as described in [10]. DNA (5 \( \mu g \)) was digested with restriction endonucleases, size-fractionated in agarose gels and transferred onto Hybond-N+ membrane (Amersham). The membranes were hybridized at 42°C to randomly primed \(^{32}\)P-labeled DNA probes in 50% formamide, 5×SSPE (1×SSPE = 180 mM NaCl/10 mM sodium phosphate/1 mM EDTA, pH 7.4), 0.5% SDS, 5×Denhard’s solution (1×Denhard’s solution = 0.02% Ficoll/0.02% polyvinylpyrrolidone/0.02% bovine serum albumin) with 100 \( \mu g \) of denatured salmon sperm DNA per ml and washed under stringent conditions.

2.3. Library screening and clone isolation

A genomic and a cDNA library in λZAPII (Stratagene) were kindly provided by P.J. Johnson (University of California, School of Medicine, Department of Microbiology and Immunology, Los Angeles, CA). These were screened with a 1200 bp chicken actin probe in the same conditions as above, except that 30% formamide was used to decrease the stringency. Positive clones were isolated and the pBluescript phagemids carrying the insert were excised with R408 helper phage according to the instructions of the manufacturer (Stratagene). The excised phagemid was grown in Escherichia coli XL1-Blue and purified from lysed cells by extraction.

2.4. DNA sequencing and sequence analysis

The nucleotide sequences of genomic and cDNA clones coding for actin were determined by the dideoxy chain termination method, using a T7 polymerase kit (Pharmacia). Sequencing was performed on both strands first with universal primers then with synthetic primers (Eurogentec) on the basis of results obtained in previous cycles of sequencing. The DNA nucleotide sequences obtained have been submitted to the Genbank database with accession numbers: U63122, U63123, U63124, U63125, U63126.

The derived amino acid sequences of T. vaginalis actin genes were aligned with related sequences retrieved from data bases [11]. The sequences from various species and their Genbank accession numbers are: Saccharomyces cerevisiae centrinactin X79811, MDCK cell centrinactin S45367, Pneumocystis carinii centrinactin L21184, Neurospora crassa centrinactin L31505, Giardia lamblia L29032, Oxystricha nova M22480, Euplotes crassus J04533, Leishmania major L16961, Trypanosoma brucei M20310, Tetra-
hymena thermophila M13939, Cryptosporidium parvum M86241, Plasmodium falciparum M19146, Achlya bisexualis X59936, Phytophthora megasperma X15900, Naegleria fowleri M90311, Entamoeba histolytica M16339, Volvox carteri M33963, Arabidopsis thaliana M20016, Zea mays J01238, Saccharomyces cerevisiae V01288, Schizosaccharomyces pombe Y00447, Pneumocystis carinii L21183, Acanthamoeba castellanii J01016, Dictyostelium discoideum M14146, Physarum polycephalum X07792, Strongylocentrotus franciscanus X03075, Drosophila melanogaster J01065, Caenorhabditis elegans X16797, Homo sapiens J00068. Alignment was obtained by visual editing using the ED program. The program NET was then used to choose boundaries and exclude regions where the alignment is difficult. Distance matrices were then formatted for the program NJBOOT which calculates the bootstrap proportion (BP) from 1000 replicates in each case. Finally the trees were drawn with the program TREEPLT. These different programs are part of the MUST package [12].

The same alignment was used to infer a phylogenetic tree with the maximum parsimony method using the PAUP software package (ver. 3.1.1) with a heuristic search procedure. Monophyly of groups was assessed with the bootstrap method (100 replicates).

Alignment and boundaries will be sent upon request.

---

Fig. 1. The nucleotide sequence of *T. vaginalis* actin gene (type 1) and its 5' and 3' flanking regions. The derived amino acid sequence is shown above the nucleotide sequence. Nucleotide and amino acid numbers are indicated on each line. An asterisk above TAA indicates a stop codon. Open boxes indicate the four regions of higher divergence. The putative conserved promoter element is underlined. These sequence data will appear in Genbank with the accession number U63122.
3. Results and discussion

3.1. Cloning of *T. vaginalis* actin gene and its deduced amino acid sequence

To isolate the actin gene from *T. vaginalis*, a genomic λZAPII library was screened with the chicken actin probe. One clone was identified and the pBluescript phagemid carrying the insert was excised. The sequence of the 2500 bp insert was determined. As shown in Fig. 1, this sequence contains an open reading frame (ORF) of 1128 bp corresponding to 376 amino acids. Alignment of the obtained protein sequence with the actin of diverse eukaryotes demonstrated that the sequenced portion of DNA was typical of actin. The estimated molecular mass of the translation product was 41.8 kDa, which is close to the 43 kDa previously determined by SDS-PAGE analysis [4], with a pI value of 5.37, slightly more acidic than the most common actins (pI=5.4) [1]. We have designated the *T. vaginalis* actin gene corresponding to this sequence type 1. It is the first available sequence of actin in Trichomonadida.

The non-coding 5' and 3' flanking regions of the actin gene are A+T rich, which is in agreement with previous observations made on other *T. vaginalis* sequences. Examination of the DNA sequence upstream of the initiator element failed to reveal appropriately spaced TATA boxes frequently associated with eukaryotic RNA polymerase II. Closely preceding the coding region is the motif TCATTTTCT which resembles the consensus sequence TCAYT-WYTCATTA found in the genes of *T. vaginalis*, and is assumed to act as initiator element (Inr) in transcription initiation of higher eukaryotic genes [13].

Codon usage is highly biased, but in the same manner as noted for the β-tubulin and ferredoxin genes of *T. vaginalis*. Twenty-five codons are never used in this gene, 13 amino acids are encoded by one significantly preferred codon and five amino acids are encoded by two preferred codons (Table 1). As in *Giardia lamblia* [5] and *Entamoeba histolytica*, the protein coding region of *T. vaginalis* actin is not interrupted by intervening sequences.

A comparison with actins from other organisms reveals general sequence conservation with identity scores around 65–70%, but if conservative amino acid changes are considered, the similarity with various actins increases to 80–85%. The lowest value is with *Giardia* actin which presents only 58% identity and 76% similarity. Four regions of higher divergence (positions 1–8, 41–51, 187–191 and 237–240) have been found (Fig. 1). The first 6–8 amino acids of actin are generally highly variable among species [1] and represent sequences for which the constraints are reduced and which are important for isoform-specific functions in the cell. The residues 38–52 represent a region of association with DNase I [14], the modification of this region in *T. vaginalis* may explain the low inhibition of DNase I activity detected by Cappuccinelli et al. [4].

3.2. Isolation of cDNA encoding *T. vaginalis* actin

A *T. vaginalis* λZAPII cDNA library was screened under low stringency conditions with the 1200 bp chicken actin clone. Ten clones were isolated. The insert from each clone was excised and sequenced. These cDNAs correspond to five different actin sequences of 1102–1124 bp which lack the 5' end and

| Table 1 |
| Codon usage in the Trichomonas actin gene type 1 |

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>First and second base</th>
<th>Third base</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TC</td>
<td>T</td>
</tr>
<tr>
<td>S</td>
<td>TC</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>AG</td>
<td>0</td>
</tr>
<tr>
<td>L</td>
<td>CT</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>TT</td>
<td>-</td>
</tr>
<tr>
<td>R</td>
<td>CG</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>AG</td>
<td>-</td>
</tr>
<tr>
<td>F</td>
<td>TT</td>
<td>0</td>
</tr>
<tr>
<td>Y</td>
<td>TA</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>TG</td>
<td>0</td>
</tr>
<tr>
<td>P</td>
<td>CC</td>
<td>0</td>
</tr>
<tr>
<td>H</td>
<td>CA</td>
<td>0</td>
</tr>
<tr>
<td>Q</td>
<td>CA</td>
<td>-</td>
</tr>
<tr>
<td>I</td>
<td>AT</td>
<td>2</td>
</tr>
<tr>
<td>T</td>
<td>AC</td>
<td>0</td>
</tr>
<tr>
<td>N</td>
<td>AA</td>
<td>3</td>
</tr>
<tr>
<td>K</td>
<td>AA</td>
<td>-</td>
</tr>
<tr>
<td>V</td>
<td>GT</td>
<td>18</td>
</tr>
<tr>
<td>A</td>
<td>GC</td>
<td>16</td>
</tr>
<tr>
<td>D</td>
<td>GA</td>
<td>17</td>
</tr>
<tr>
<td>E</td>
<td>GA</td>
<td>-</td>
</tr>
<tr>
<td>G</td>
<td>GG</td>
<td>13</td>
</tr>
</tbody>
</table>
Table 2  
Comparison of the five types (T) of cDNA actin nucleotide sequences

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
</tr>
</thead>
<tbody>
<tr>
<td>33*</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>45*</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>69*</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>90*</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>A</td>
<td>T</td>
</tr>
<tr>
<td>147*</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>150*</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>156*</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>159*</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>165*</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>177*</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>231*</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>387*</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>417*</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>426*</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>435*</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>636*</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>660*</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>708*</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>816*</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>846*</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>910</td>
<td>A</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>922</td>
<td>A</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>996*</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>1113*</td>
<td>T</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
</tbody>
</table>

* indicates nucleotide position corresponding to the third base of a codon.

Changes and that an identical gene was found in both cDNA and gDNA libraries indicate that the variations in actin sequences are not artifacts introduced by the reverse transcriptase enzyme.

3.3. Organization and expression of the T. vaginalis actin genes

To estimate the copy number of the actin gene in T. vaginalis, genomic DNA was digested with various restriction enzymes and probed under high stringency conditions with the type 1 actin cDNA clone. The same pattern of labeling is observed when the other four clones are used as probes (not shown). A complex array of nine bands ranging from 9 to 2.5 kb is observed on EcoRI digest (Fig. 2A). One band (3.7 kb) has a higher intensity, suggesting that it could represent more than one copy. Note that the 2.5 kb fragment corresponds to the original type 1 clone.

Fig. 2. Estimation of actin gene copy number in T. vaginalis. A Southern blot of genomic DNA cleaved with EcoRI (lane A), NsiI (lane B) BamHI (lane C) and BssXI (lane D) was hybridized under stringent conditions to the type 1 cDNA clone. The positions of the molecular weight standards (in kb) are indicated.
actin found in the gDNA library. Digestion with 
NsiI (Fig. 2B) shows nine bands plus a faint signal 
at 7.5 kb which could represent a distantly related 
sequence rather than bona fide actin gene. Thus, the 
digestion is consistent with the existence of more 
than nine gene copies in *T. vaginalis*. The presence 
of fewer bands, with higher size, detected by diges-
tion with *Bam*III (Fig. 2C) and *Bst*XI (Fig. 2D) may 
indicate that these genes are closely localized. Deter-
mination of gene copy number as presented above 
assumes that there is no cleavage site for the enzymes 
within all the actin clones. These sites are absent in 
the five actin types so far characterized. Southern 
blots and cloning experiments suggest that actin is 
encoded by a multigenic family composed of more 
than five isoforms (sequenced) and at least nine cop-
ies. As we have only isolated five different expressed 
gegens in the cDNA library, we cannot exclude the 
presence of pseudogenes or cycle-specific expressed 
gegens. It is difficult to speculate on the organization 
of this actin gene family, given the present lack of 
data concerning the organization of the genome. 

Nevertheless, the genome of *T. vaginalis* is known 
to contain several families of repeated sequences: 
the TV-E650 family, the seven β-tubulin genes and 
the 17 ubiquitin sequences including an 11-repeat 
fragment of a polyubiquitin gene.

Actins are most commonly encoded by multi-copy 
gegens. Multiple expressed actin genes may provide a 
selective advantage by permitting the production of 
slightly different actin isoforms that function in spe-
cific ways during the cell cycle or life cycle and may 
facilitate the regulation of actin gene expression in 
response to various developmental signals or envi-
ronmental stimuli. In *T. vaginalis*, the products of 
the different genes are essentially identical in primary 
sequence and diversity of actin function may be con-
ferred through post-translational modification as ob-
served in the 17 genes of *Dictyostelium discoideum* 
[15]. In any case, the high copy number of actin 
gegens in *T. vaginalis* would provide a simple mecha-
nism for high level expression of actin necessitated 
by the extensive use of actin microfilaments in the 
amoeboid forms, i.e. the formation of pseudopods/ 
filopods and phagocytosis/pinocytosis. According to 
the large amount of actin found in swimming cells, 
the actin molecules could be sequestered in the cyto-
plasm, associated with actin binding proteins. Our 
study represents an initial step in the development of 
a molecular approach to understand the process of 
invasion by *T. vaginalis*, related to cytoskeleton 
modifications.

Northern blots of *T. vaginalis* RNA were hy-
bridized with the five cDNA types of actin and the 
autoradiograms indicate that transcripts of the 
different genes have the same size (1.2 kb) (not 
shown).

3.4. Phylogenetic reconstruction

Based on known actin gene sequences, phyloge-
etic trees were constructed with the neighbor-join-
ing method (Fig. 3A) and the parsimony method 
(Fig. 3B). As actin is not present in bacteria, the 
trees were rooted on centraclin (arp1) which is the 
result of an ancient gene duplication.

In our trees, the branching pattern was fairly 
reasonable with major divisions corresponding to 
conventional taxonomic classification: the Kinet-
oplastida, Apicomplexa, Oomycetidae, Plantae, Fungi 
and Animalia are monophyletic groups as in the 
recent actin-based phylogeny of Drouin et al. [5], 
and the animals and fungi are sister groups while 
plants constitute an independent evolutionary line-
age, in agreement with the results of Baldauf and 
Palmer [16].

Some discrepancies exist such as the paraphy-
cy of ciliates, observed only on actin and EF-1α phy-
logenies [9]. The monophyly of Ciliates is currently 
accepted. On the actin trees, the two ciliates *Oxy-
tricha* and *Euplotes* have very long branches com-
pared to other taxa. This may reflect a long period 
of independent evolution for these ciliates or 
more probably a differing rate of sequence diver-
gence of actin coding regions; the faster a gene has evolved in a given species, the earlier it may emerge in a phylogeny due to the long branch attraction artifact. In the neighbor-joining tree (Fig. 3A), T. vaginalis branches off after the Kinetoplastida, but this is not supported by a high bootstrap value (72%) and the addition of an Euglena actin sequence resolves the Parabasala before the Euglenozoa (unpublished results). More species are necessary to resolve this branching node. In the PAUP tree (Fig. 3B), the ciliates are paraphyletic as well but T. vaginalis emerges directly above Diplomonadida with a bootstrap value of 77%.

Three groups are currently accepted at the base of the eukaryotic tree: Microsporidia, Diplomonadida and Parabasala. Their genes are more prokaryotic in organization than those of higher eukaryotes: short promoter, lack of introns, structure of the rDNAs. Microsporidia have recently been determined as degenerate fungi [17]. The emergence of Trichomonas and Giardia inferred from the different molecular phylogenies is inconsistent and often poorly supported by bootstrap values. On trees based on SSU rRNA [7] or tubulin [17], Parabasala branch off before the Diplomonadida whereas on EF-1α, EF-2 [9] and our actin phylogenies, Diplomonadida branch off before Parabasala. Correct branching orders in rRNA molecular phylogenies may have been obscured by unusual features in the molecules from the earliest branching lineages and protein trees always present a risk of paralogous comparison and serious inequalities in rates of evolution of protein genes in some taxa. It seems impossible to determine, with the currently available data, the order of emergence of Diplomonadida and Parabasala.

The lack of mitochondria in these two basal groups was initially interpreted as a strong argument in favor of their primitive character. This view must now be moderated for trichomonads in which genes related to chaperonin 60 [18–20] and mitochondrial-type HSP70 [19,21] have been obtained, placing the mitochondrial endosymbiosis much earlier in eukaryote evolution than previously thought. Typical mitochondrial genes have not yet been clearly demonstrated in diplomonads. In conclusion, T. vaginalis no more belongs to the so-called base of the tree [21] but nevertheless has an earlier emergence than the large majority of Protists.

Acknowledgments

The authors thank A. Gernot for critical reading of the manuscript, and P. Johnson for giving us the two libraries. This work was supported by the Centre National de la Recherche Scientifique.

References


