ELSEVIER

Biochimica et Biophysica Acta 1514 (2001) 14–20

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

www.bba-direct.com

Novel putative saposin-like proteins of *Entamoeba histolytica* different from amoebapores

Heike Bruhn, Matthias Leippe *

Bernhard Nocht Institute for Tropical Medicine, Bernhard-Nocht-Strasse 74, D-20359 Hamburg, Germany

Received 6 February 2001; received in revised form 18 April 2001; accepted 25 April 2001

Abstract

Amoebapores, the pore-forming proteins of *Entamoeba histolytica*, have been shown to play a pivotal role in the pathogenicity of the protozoan parasite. They belong to the functionally diverse family of saposin-like proteins (SAPLIPs) characterized by a conserved pattern of cysteine residues and the ability to interact with lipids. Here, we report the identification of genomic sequences encoding presumably novel SAPLIPs in *E. histolytica* and classify them in the structural and functional context provided by known family members. The genes of altogether 15 SAPLIPs are transcribed in the axenically cultured trophozoites as evidenced by reverse transcriptase–polymerase chain reaction. Interestingly, a remarkable sequence variety with a strong resemblance to that of known, functionally diverse SAPLIPs is present in this archaic, unicellular organism.

Keywords: Amoebapore; Saposin-like protein; Membrane interaction, pore-forming; Antimicrobial peptide; *Entamoeba histolytica*

1. Introduction

A family of pore-forming proteins, the amoebapores, is a major pathogenicity factor of *Entamoeba histolytica*, the causative agent of human amoebiasis and a parasitic protozoon with remarkable phagocytic and cytolytic capacities [1,2]. The amoebapores are membrane-interacting proteins which display pore-forming activity toward liposomes [3], antibacterial activity [4] and which are cytotoxic to human cell lines [5] in vitro. The primary function of the pore-forming proteins in vivo is assumed to be the killing of phagocytosed bacteria, but their cytolytic potency makes them a prominent instrument in the invasive behaviour of the amoeba [2,6]. In contrast to many other membrane-permeabilizing proteins, the amoebapores act by forming distinct ion channels which are recordable in artificial membranes [7].

Three isoforms, amoebapores A, B and C, are known. All of them are localized in cytoplasmic granules and show only some quantitative differences in their specific activities [4]. Recently, we characterized three amoebapore homologues, the disparpores, in the closely related organism *E. dispar* [8]. Structurally, they all belong to the family of saposin-like proteins (SAPLIPs) characterized by a conserved sequence motif of six cysteine residues involved in three disulfide bridges which comprises approximately 80 amino acid residues [9]. The various members of the SAPLIP family fulfil different functions: they possess activities as enzymes, i.e., acidic sphingomyelinases [10], acyloxyacyl hydrolases [11] and

* Corresponding author. Fax: +49-40-42818-512.
E-mail address: leippe@bni.uni-hamburg.de (M. Leippe).

0005-2736/01/$ – see front matter
PII: S0005-2736(01)00345-5

plant aspartic proteases [12], as cofactors of enzymes involved in lipid metabolism, i.e., saposins A to D [13], as components of pulmonary surfactant reducing the surface tension, i.e., pulmonary surfactant-associated protein B (SP-B) [14], as part of a high molecular protein complex involved in stage regulation of *Dictyostelium discoideum*, i.e., the recently identified and classified countin [15,16], or as antimicrobial effector molecules, i.e., the mammalian lymphocyte proteins NK-lysin [17] and granulysin [18], but their common feature is that they all are able to interact with lipids. The mature proteins consist of the SAPLIP domain only, as is found with amoebapores, saposins, SP-B, NK-lysin and granulysin, or contain the SAPLIP domain implemented in a larger structure, as observed in the aforementioned enzymes and countin. The saposins and SP-B are synthesized as large precursor proteins comprising the four known saposins A–D or three SAPLIP domains, respectively, and are proteolytically processed. In 1997, the first structure of a SAPLIP was solved with NK-lysin [19] and the presumably common fold, a five-helical bundle, was confirmed by the structure of the SAPLIP-domain of prophytepsin, an aspartate protease of barley, in 1999 [20].

The ongoing genome project of *E. histolytica* allowed us to identify 12 novel putative proteins containing the typical cysteine pattern of the SAPLIPs in this parasite. Here, we present their sequence characteristics in comparison to the known family members with special attention to the comprehensively characterized amoebapores, the only known SAPLIPs of *E. histolytica* to date.

2. Material and methods

2.1. Sequences and analysis

Homology searches for identification of the novel SAPLIP sequences were performed on the *E. histolytica* genome database of The Institute for Genome Research (http://www.tigr.org). Several known sequences of SAPLIP family members were used as a query. Successful identification of novel SAPLIP sequences were obtained by the following queries: accession nos. AAA29111, CYY54225, CAA54226, AAB06759, AF082528, AF154046, AF154047, AAB06759 and CAA59720. For completion, the newly identified sequences were then used itself as a query. Preliminary sequence data for *E. histolytica* is deposited regularly into the GSS division of GenBank. The Sequencing effort is part of the International *Entamoeba* Genome Sequencing Project and is supported by award from the National Institute of Allergy and Infectious Diseases, National Institutes of Health.

Homologues proteins to the newly identified SAPLIPs were searched using BLAST (http://www.ncbi.nlm.nih.gov/blast/) and Interpro (http://www.ebi.ac.uk/interpro/). Binary sequence comparisons and evaluation of alignment significance were performed with GAP [21] introducing 100 steps of random sequence shuffling for calculation of the average alignment score. Signal peptides and their cleavage sites were identified using signalP [22,23] and coiled coil structures were predicted by COILS [24]. The multiple alignment was performed by CLUSTAL W [25], coloured using MACBOXSHADE (K. Hofmann, M.D. Baron, http://www.isrec.isb-sib.ch/ftp-server/boxhade/) and the phylogenetic tree was visualized by TREEVIEW [26]. Homology modelling was performed using WHAT IF [27] on the structural template of NK-lysin (1NKL). The alignment was manually corrected to ensure the positioning of gaps in loop regions.

2.2. Transcriptional analysis

Trophozoites of *E. histolytica* HM1:IMSS were cultured axenically [28]. RNA was isolated directly or after incubation on chinase hamster ovary (CHO) cells for a period of 2 h from trophozoites in the late-logarithmic state which were suspended in Trizole Reagent (Life Technologies, Eggenstein, Germany) as recommended by the manufacturer. cDNA synthesis was performed according to Schramm et al. [29] and used as template in a PCR with gene specific oligonucleotides. The sense primers comprised 20 nucleotides coding for the first amino acids of the SAPLIP domain of the novel sequences according to Fig. 2A with the exception of gg73tr, which was amplified using the CapFinder Oligonucleotide GAGAGAACGCGTGACGAGAGACTGACAGGGGGGGG. The respective antisense primers consisted of the last 20 nucleotides preceding the stop

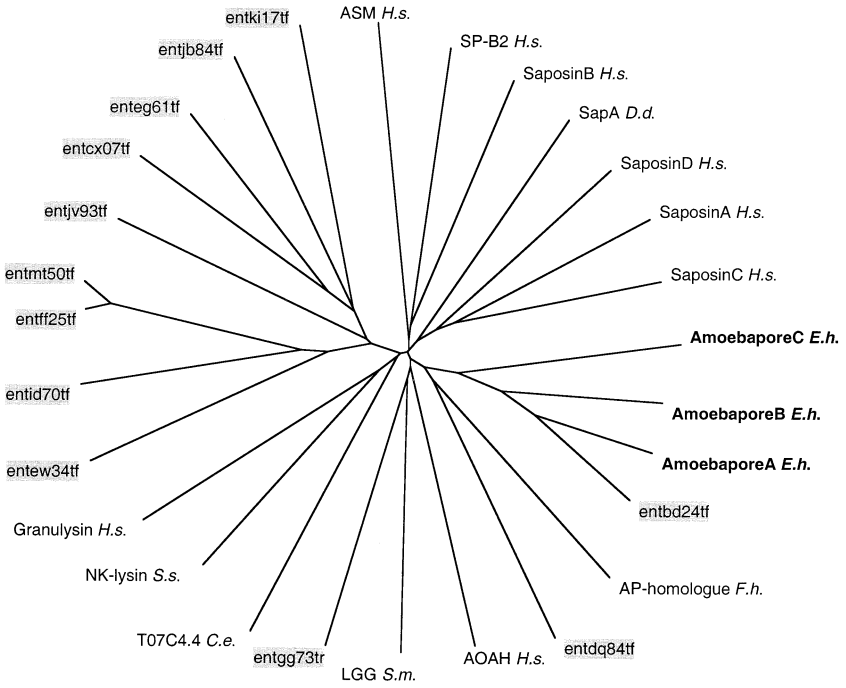

Fig. 1. Phylogenetic tree of the SAPLIP domains of the novel putative *E. histolytica* proteins and representative family members. The new sequence names are shaded in grey and the amoebapores are written in bold. The scale bar represents 0.1 mutations per site. The accession numbers of the various proteins are: amoebapore A, B and C (*E. histolytica*) (AAA29111, CAA54225 and CAA54226), saposin A–D precursor (*Homo sapiens*) AAA36595, NK-lysin (*Sus scrofa*) CAA59720, granulysin (*H. sapiens*) AAA59935, SP-B (*H. sapiens*) AAA60212, acyloxyacyl hydrolase (AOAH) (*H. sapiens*) AAA35506, acidic sphingomyelinase (ASM) (*H. sapiens*) AAA75009, LGG (*Schistosoma mansoni*) AAB81008, SapA (*Dictyostelium discoideum*) AAB06759, hypothetical protein T07C4.4 (*C. elegans*) S41017, and amoebapore (AP)-homologue (*Fasciola hepatica*) AAB02579.

codon. The polymerase chain reaction (PCR) signals were analysed by agarose gel electrophoresis.

3. Results and discussion

By searching the *E. histolytica* genome database, we newly identified 12 putative gene products with similarity to SAPLIPs. All of the novel SAPLIP sequences of *E. histolytica* contain only one SAPLIP-domain which is always located at the C-terminus of the putative proteins. Sequences containing the typical cysteine pattern were: entbd24tf (nt 194–424; frame +2; acc. no. AZ529784), entcx07tf (nt 430–185; frame −1; acc. no. AZ530711), enteg61tf (nt 779–426; frame −1; acc. no. AZ550036), entff25tf (nt 30–389; frame +3; acc. no. AZ546503), entmt50tf (nt 776–522; frame −2; acc. no. AZ689362), entid70tf (nt 142–387; frame +1; acc. no. AZ685704), entew34tf (nt 289–62; frame −1; acc. no. AZ546519), entgg73tr (nt 3–212; frame +3; acc. no. AZ544462), entdq84tf (nt 664–449; frame −1; acc. no. AZ546992), entjb84tf (nt 573–824; frame +3; acc. no. AZ690015), entjv93tf (nt 701–462; frame −2; acc. no. AZ687176) and entki17tf (nt 507–268; frame −3; acc. no. AZ692153).

In the SAPLIP family, an enormous sequence variety is observed. Pairwise sequence identities of the SAPLIP domains are mostly below a homology threshold of about 30% with only a few exceptions, e.g., human saposin A compared to saposin D (38.1%) and the amoebapore isoforms among each other (35.1–57.1%) (Fig. 1). Because of the high similarity to their orthologues (95%, 91% and 88%), the disparpores are omitted in this analysis. Interestingly, among the novel *E. histolytica* sequences only entbd24tf can be clearly classified as an amoebapore-like protein with a sequence identity of 64.9%, 50.6% and 39.0% to amoebapore A, B and C, respectively. Other remarkable identities of the novel sequences to known SAPLIPs do not exist. Accordingly, the evaluation of binary sequence comparisons

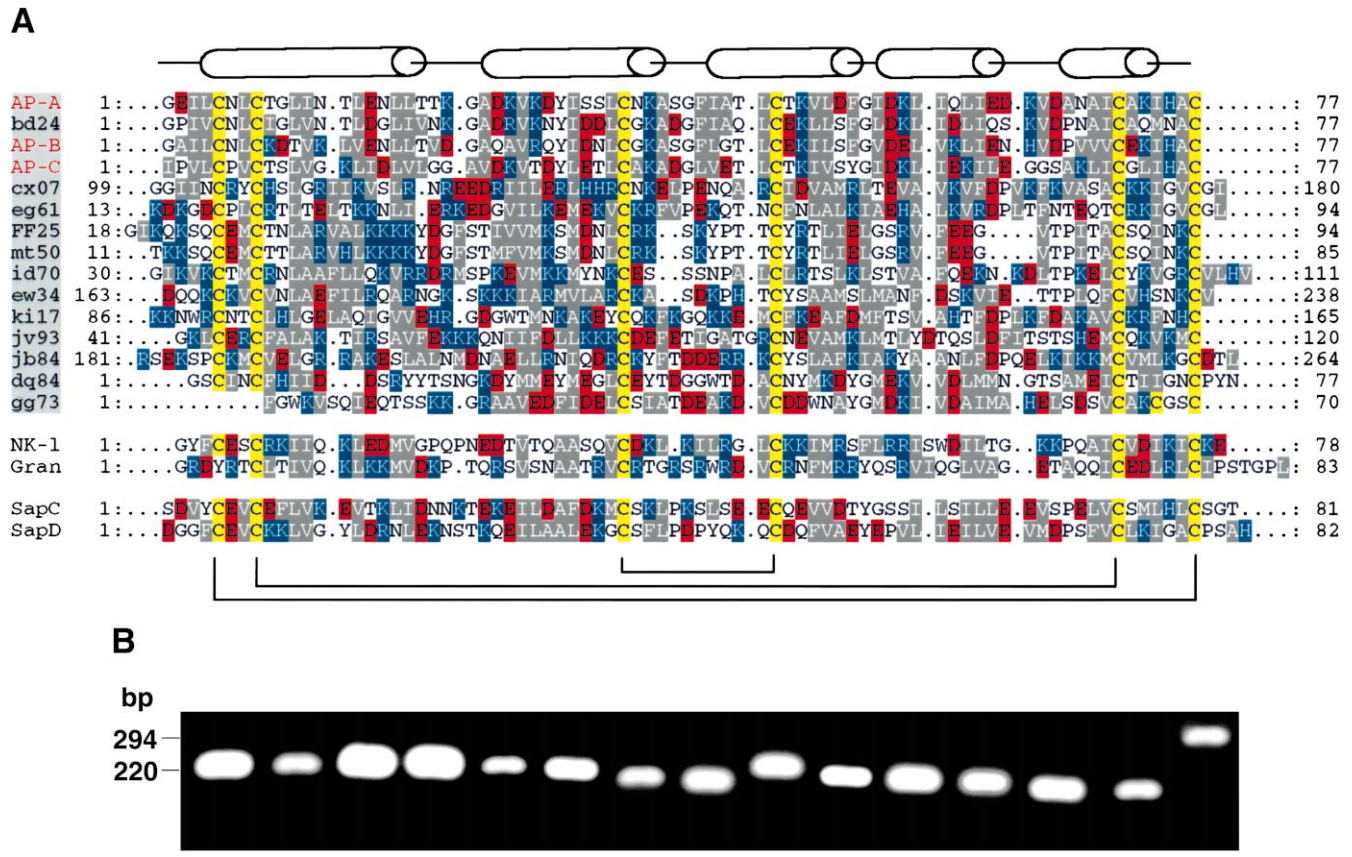

Fig. 2. Sequence alignment of all SAPLIP domain sequences of *E. histolytica* (A) and their transcriptional analysis (B). (A) For comparison, the amoebapores, NK-lysin (NK-l), granulysin (Gran) and saposin (Sap) C and D were included. The numbering refers to the mature proteins if the processing site is known or predicted or to the translational start methionine. In the case of entmt50tf, entgg73tf, entki17tf and entjv93tf the definite translational start signal is not identified yet. The helices according to the NK-lysin structure are presented by cylinders and the disulfide bonds are represented by horizontal lines below the sequences. Positively charged amino acid residues (K,R,H) are coloured in blue, negatively charged amino acid residues (D,E) are shown in red, hydrophobic amino acid residues (I,L,V,M,F,W,A) are shaded in grey and the cysteines are presented in yellow. Histidine residues are positively charged provided that the pH is <6.5 in its microenvironment. (B) Agarose gel analysis of RT–PCR reactions performed on *E. histolytica* RNA with specific primers for the SAPLIP domains of the novel putative proteins. The different reactions are shown in the same order as in the alignment above. The position of marker DNAs of defined lengths is given at the left-hand side.

between all novel SAPLIP domains and known members of this family reflects the equivalence of the natural sequences and identically composed but random sequences concerning the alignment significance, indicating the absence of substantial sequence similarities beyond the cysteine motif. Concerning the identity of the novel sequences among each other, the sequences entff25tf and entmt50tf are closely related (89.6% identity) and both show a significant similarity to entid70tf (34% identity). The only other similarity worth mentioning is between enteg61tf and entcx07tf (36.6.% identity).

Only two of the putative proteins, entdq84tf and the amoebapore-like entbd24tf, comprise a typical signal sequence of 15 amino acid residues for targeting into the endoplasmic reticulum [23]. The predicted cleavage sites are located 2 and 4 amino acid residues N-terminal of the first cysteine residue of the SAPLIP motif. The absence of such a signal sequence in the other putative SAPLIPs of *E. histolytica* may rather point to the fact that protein targeting in *E. histolytica* is not completely understood yet than infer a cytosolic localization. A typical feature of the SAPLIP fold is its stabilization by the three distinct disulfide bonds, which cannot exist in cytosolic proteins. The length of the sequence stretch

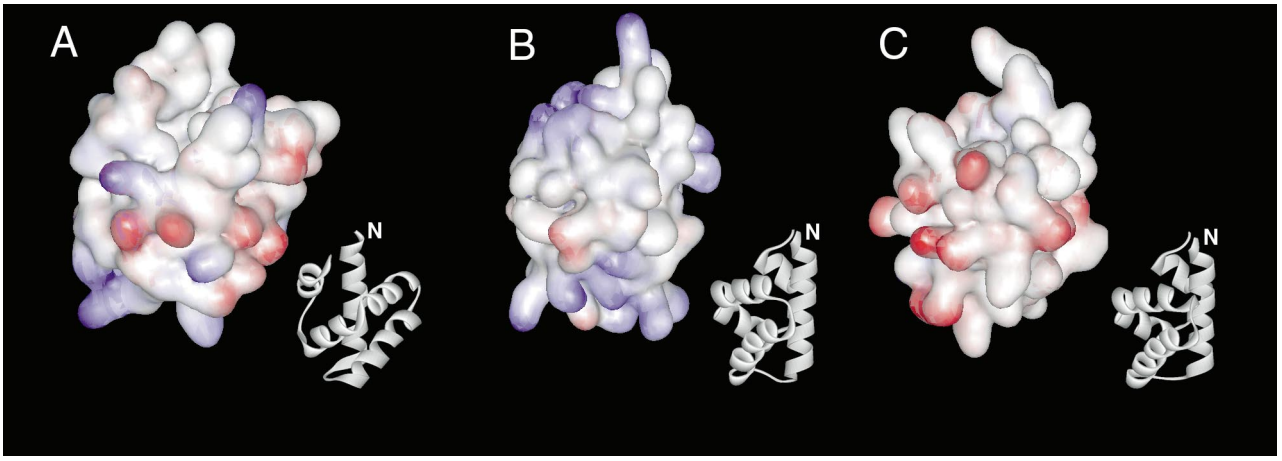

Fig. 3. Protein surfaces coloured according to their electrostatic potential. The protein surfaces of structural models of the SAPLIP domains of entbd24tf (A), entid70tf (B) and entdq84tf (C) are coloured according to their electrostatic potential. Red colour represents negatively charged amino acid residues and blue colour represents positively charged amino acid residues. The orientation of the helices are given right of the corresponding surface presentations in grey colour.

N-terminal to the cysteine motif of these putative proteins differs from 12 to 180 amino acid residues (Fig. 2). For none of these sequences preceding the SAPLIP domain, homology searches by sequence similarities or by family profiles and patterns could identify a significant homology. Notably, similarities to coiled coil structures could be detected by sequence comparison and can also be predicted by use of the respective algorithms for entcx07tf (amino acid residues (aa) 41–80), entjb84tf (aa 69–132 and 143–185), entew34tf (aa 1–31) and entki17tf (aa 37–82). In general, coiled coil sequences can serve as a dimerization interface to build rigid rod-shaped structures, but they are also found intramolecular. Therefore, the function of the here identified coiled coils cannot be predicted.

The genomic sequences of 36.1 Mb of *E. histolytica* available to date (www.tigr.org/tdb/edb2/enta/htmls/) comprise approximately twofold the genome [30]. Due to errors in the sequencing procedure not corrected at this early stage, some of the here presented new sequences may bear minor uncertainties especially in the far N-terminal (e.g.entff25tf, entmt50tf and entgg73tr) or C-terminal region (e.g., entjb84tf). The same reason might be responsible for the absence of the typical translational start signal ATG in entki17tf, entgg73tr, entmt50tf and entjv93tf.

In the SAPLIP-family, even proteins which fulfil similar functions, such as amoebapores and the mammalian NK-lysin and granulysin, do not reveal a significant sequence homology beyond the cysteine motif and, moreover, apparently achieve their activities by different mechanisms of membrane interaction. As conserved functional motifs in the SAPLIPs are absent, the distinct distribution of hydrophobic and charged amino acid residues in the structure may be the key factor in determining the function. All of the novel SAPLIP domains are polar, charged molecules with substantial differences in the charge density from 16 (entbd24tf and entdq84tf) up to 30 charged amino acid residues (entjb84tf) (Fig. 2A). Only entbd24tf resembles the amoebapores in its low net charge and dispersed charge distribution, which is most similar to amoebapore C. An essential step in channel formation by amoebapores, the oligomerization, is dependent on the sole histidine residue at the very C-terminus of the sequence as could be shown for amoebapore A [31]. Notably, this histidine residue is changed to asparagine in entbd24tf which may imply that this putative protein does not possess the pronounced – if any – pore-forming activity of amoebapores. Correspondingly, during the amoebapore purification procedure, we could not identify any other pore-forming protein in *E. histolytica* [4].

Transcriptional analysis of the genes, however, of axenically grown *E. histolytica* and amoebae incubated with CHO-cells prior to harvest gave evidence that all of the here identified genes are actively tran-

scribed under both of the conditions tested. A reverse transcriptase–PCR employing gene specific oligonucleotides for all newly identified SAPLIPs and the amoebapores yielded amplification products of the expected size (Fig. 2B).

Most of the putative proteins are positively charged, first and foremost entid70tf with a net charge of +13. These basic amino acid residues are especially concentrated on the first helix and the following loop (Fig. 3), a feature which is also observed in entff25tf, entmt50tf, entew34tf and entcx07tf. In entew34tf this positive region is continued all over the second helix, whereas in entff25tf and entmt50tf, a distinct second region comprising the loop between helix two and three is also positively charged. The charge distribution of none of these putative proteins match with that of NK-lysin or granulysin. In these molecules, the positive charges concentrated on helix three are reported to be essential components for the antimicrobial and cytolytic function of the molecules [32–35]. Nevertheless, one may suggest an antimicrobial activity for the highly positively charged novel SAPLIPs found here according to a mechanism of molecular electroporation [32] or to a carpet-like membrane perturbation observed in smaller, mostly cationic antibacterial peptides [36].

A noticeably unique feature of entew34tf is a unpolar sequence stretch comprising the whole third helix and the following loop. Two of the sequences, entgg73tr and entdq84tf, are highly negatively charged with a concentration of acidic amino acid residues at helix two and three (Fig. 3), thereby resembling the saposins. As all genes are transcribed, an interesting question arising is whether the lipid metabolism of *E. histolytica* comprises reactions comparable to those in the mammalian pathway which require the cofactor activity and the specific lipid interactions of SAPLIPs. Apparently, this unicellular eukaryote – which may be considered relatively primitive – possesses an impressive variety of SAPLIP sequences. In mammals, SAPLIPs display various functions. Thus, it is tempting to suggest that these amoebic proteins are ancient ancestors leading to the SAPLIP diversity observed in higher eukaryotes. Alternatively, the novel putative proteins may belong to a very fine-tuned and comprehensive antimicrobial system and may complement the activity of amoebapores toward phagocytosed bacteria, particularly Gram-negatives. The 'arms race' with potential pathogens and their high mutation rate may have led to a strong selection pressure towards sequence diversity, a hypothesis that has been formulated to explain the higher rate of mutations in defensive molecules compared to that of other proteins [37].

Concerning the role of amoebic SAPLIPs in the pathogenicity, the amoebapores apparently are the only pore-forming proteins of *E. histolytica* and hence they most likely are the unrivalled members of this protein family that contribute to the destruction of host tissues.

The multiplicity of natural variants found here in sequence and structural features provide the chance to identify general structure–function correlations for the SAPLIP family. Ultimately, ongoing studies should develop a system for recombinant expression of all amoebic SAPLIP members for their functional characterization.

Acknowledgements

We thank Verena Diehm for providing the cDNA. The work was supported by the Deutsche Forschungsgemeinschaft (DFG) by Grant (Le 1075/2-3) and by a Heisenberg fellowship to M.L. Preliminary sequence data for *E. histolytica* is deposited regularly into the GSS division of GenBank. The sequencing effort is part of the International *Entamoeba* Genome Sequencing Project and is supported by an award from the National Institute of Allergy and Infectious Diseases, National Institutes of Health.

References

[1] M. Leippe, Parasitol. Today 13 (1997) 178–183.
[2] R. Bracha, Y. Nuchamowitz, M. Leippe, D. Mirelman, Mol. Microbiol. 34 (1999) 463–472.
[3] M. Leippe, S. Ebel, O.L. Schoenberger, R.D. Horstmann, H.J. Müller-Eberhard, Proc. Natl. Acad. Sci. USA 88 (1991) 7659–7663.
[4] M. Leippe, J. Andrä, R. Nickel, E. Tannich, H.J. Müller-Eberhard, Mol. Microbiol. 14 (1994) 895–904.
[5] M. Leippe, J. Andrä, H.J. Müller-Eberhard, Proc. Natl. Acad. Sci. USA 91 (1994) 2602–2606.

[6] O. Berninghausen, M. Leippe, Infect. Immun. 65 (1997) 3615–3621.
[7] M. Leippe, H.J. Müller-Eberhard, Toxicology 87 (1994) 5–18.
[8] R. Nickel, C. Ott, T. Dandekar, M. Leippe, Eur. J. Biochem. 265 (1999) 1002–1007.
[9] R.S. Munford, P.O. Sheppard, P.J. O'Hara, J. Lipid Res. 36 (1995) 1653–1663.
[10] E.H. Schuchman, M. Suchi, T. Takahashi, K. Sandhoff, R.J. Desnick, J. Biol. Chem. 266 (1991) 8531–8539.
[11] F.S. Hagen, F.J. Grant, J.L. Kuijper, C.A. Slaughter, C.R. Moomaw, K. Orth, P.J. O'Hara, R.S. Munford, Biochemistry 30 (1991) 8415–8423.
[12] P. Runeberg-Roos, K. Tormakangas, A. Ostman, Eur. J. Biochem. 202 (1991) 1021–1027.
[13] Y. Kishimoto, M. Hiraiwa, J.S. O'Brien, J. Lipid Res. 33 (1992) 1255–1267.
[14] J. Johansson, T. Curstedt, Eur. J. Biochem. 244 (1997) 675–693.
[15] D.A. Brock, R.H. Gomer, Genes Dev. 13 (1999) 1960–1969.
[16] Y. Zhai, M.H. Saier Jr., Biochim. Biophys. Acta 1469 (2000) 87–99.
[17] M. Andersson, H. Gunne, B. Agerberth, A. Boman, T. Bergman, R. Sillard, H. Jornvall, V. Mutt, B. Olsson, H. Wigzell, A. Dagerlind, H.G. Boman, G.H. Gudmundsson, EMBO J. 14 (1995) 1615–1625.
[18] S.V. Peña, D.A. Hanson, B.A. Carr, T.J. Goralski, A.M. Krensky, J. Immunol. 158 (1997) 2680–2688.
[19] E. Liepinsh, M. Andersson, J.M. Ruysschaert, G. Otting, Nat. Struct. Biol. 4 (1997) 793–795.
[20] J. Kervinen, G.J. Tobin, J. Costa, D.S. Waugh, A. Wlodawer, A. Zdanov, EMBO J. 18 (1999) 3947–3955.
[21] J. Devereux, P. Haeberli, O. Smithies, Nucleic Acids Res. 12 (1984) 387–395.
[22] H. Nielsen, J. Engelbrecht, S. Brunak, G. von Heijne, Int. J. Neural Syst. 8 (1997) 581–599.
[23] H. Nielsen, J. Engelbrecht, S. Brunak, G. von Heijne, Protein Eng. 10 (1997) 1–6.
[24] A. Lupas, M. Van Dyke, J. Stock, Science 252 (1991) 1162–1164.
[25] J.D. Thompson, D.G. Higgins, T.J. Gibson, Nucleic Acids Res. 22 (1994) 4673–4680.
[26] R.D.M. Page, Comput. Appl. Biosci. 12 (1996) 357–358.
[27] G. Vriend, J. Mol. Graph. 8 (1990) 52–56.
[28] L.S. Diamond, C.C. Cunnick, Trans. R. Soc. Trop. Med. Hyg. 72 (1978) 431–432.
[29] G. Schramm, I. Bruchhaus, T. Roeder, Nucleic Acids Res. 28 (2000) 1–4.
[30] U. Willhoeft, E. Tannich, Mol. Biochem. Parasitol. 99 (1999) 41–53.
[31] J. Andrä, M. Leippe, FEBS Lett. 354 (1994) 97–102.
[32] M. Miteva, M. Andersson, A. Karshikoff, G. Otting, FEBS Lett. 462 (1999) 155–158.
[33] D. Andreu, C. Carreno, C. Linde, H.G. Boman, M. Andersson, Biochem. J. 344 (1999) 845–849.
[34] J. Andrä, M. Leippe, Med. Microbiol. Immunol. 188 (1999) 117–124.
[35] W.A. Ernst, S. Thoma-Uszynski, R. Teitelbaum, C. Ko, D.A. Hanson, C. Clayberger, A.M. Krensky, M. Leippe, B.R. Bloom, T. Ganz, R.L. Modlin, J. Immunol. 165 (2000) 7102–7108.
[36] Y. Shai, Biochim. Biophys. Acta 1462 (1999) 55–70.
[37] P.M. Murphy, Cell 72 (1993) 823–826.