A strategy for the identification of T-cell epitopes on Leishmania cysteine proteinases

C. R. Alves 1*, L. C. Pontes de Carvalho 1, A. L. A. Sousa 1 and S. Giovanni De Simone 1,3

1Laboratório de Bioquímica de Proteínas e Peptídeos, Departamento de Bioquímica e Biologia Molecular, Instituto Oswaldo Cruz, Rio de Janeiro, RJ, Brasil, 2Laboratório de Imunologia Molecular, Centro de Pesquisas Gonçalo Moniz (Salvador, Bahia, Brasil, FIOCRUZ, and 3Departamento de Biologia Celular e Molecular, Instituto de Biologia, Universidade Federal Fluminense, Niterói, RJ, Brasil (*Reprint address)

Key words: multiple sequence alignment, human leucocyte-associated antigen, cysteine proteinases, Leishmania, T-cell epitope

Abstract
In this study computational analysis was used to compile sequence alignments, construct a dendrogram and calculate physical data in order to predict potential T-cell epitopes of the Leishmania cysteine proteinase. Using multiple alignment of human and Leishmania proteinase sequences deposited on data bank sequences, it was possible to predict that the extreme C-terminus of cysteine proteinase (CySpp, 365-444) contained three peptides (pl 313-370, pl 415-422 and pl 431-444) with charge score, hydrophobicity and isolectric points compatible for human leucocyte-associated antigen (HLA) class II binding. The prediction was confirmed in vitro through the ability of synthetic peptides corresponding to the predicted regions to stimulate peripheral blood mononuclear cells of patients with leishmaniasis.

Introduction
Leishmania are obligate intracellular protozoan parasites responsible for a large spectrum of clinical manifestations in man, including self-healing cutaneous, diffuse cutaneous and mutilating mucosal lesions, as well as lethal visceral disease (Ridley, 1987). The parasites are dimorphic, existing as flagellate extracellular forms (promastigotes) in the sandfly vector, and as intracellular amastigotes (mastigotes), in the mammalian host.

The protective immunity against Leishmania is mediated by T-cells (Convit et al., 1972; Howard et al., 1980; Mitchell et al., 1981; Coutinho et al., 1996). Thus, the identification of linear motifs of the parasite proteins which can be processed by antigen-presenting cells associated with human leucocyte-associated antigen (HLA) class-II molecules to CD4+ T-lymphocytes, would be of great relevance for designing studies aiming at inducing immunity against Leishmania (Liew and O’Donnell, 1993).

Structural motifs in proteins of the cysteine proteinase family are particularly potent epitopes for T-cell or antibody responses (Urban et al., 1992; McKerrow, 1993). Cysteine proteinases have been identified and characterized in several species of Leishmania (Pupkus et al., 1986; Lockwood et al., 1987; Robertson and Coombs,
1990; Balanco et al., 1991; Alves et al., 1993; Alves et al., 1998; Traub-Ciecko et al., 1993). However, the potential of *Leishmania* cysteine proteinases for inducing T-cell response has not been ascertained.

Algorithms for the identification of protein sequences which may bind to major histocompatibility complex (MHC) molecules have been proposed (Rothbard and Taylor, 1988; Gornette et al., 1989). In the present study, potential T-cell epitopes on the COOH-terminus region of *Leishmania* cysteine proteinases were identified using (1) multiple sequence-alignment to compare human and *Leishmania* cysteine proteinases, (2) structural motif analysis and (3) physical datum determination.

**Materials and methods**

**Protein sequences**

Database searches for identical or homologous amino acid sequences were performed, using the software package of the Genetics Computer Group (GCG; Wisconsin University, Madison, Wisconsin, U.S.A.), version 7-UNIX. Most HLA class II and cysteine proteinase sequences were obtained from the PIR (Protein Identification Resource, National Biochemical Research Foundation, U.S.A.) and GenBank National Center for Biotechnology Information, U.S.A.). In the latter case only the coding sequence was used in the analyses.

**Sequence alignment and motif analysis**

Homology relationships between proteins was assessed using the GAP algorithm (Smith and Waterman, 1981). The multiple-sequence analysis (MSA) was undertaken using the PILEUP algorithm (Feng and Doolittle, 1987) and the PEPPLOT to predicted curves for secondary structures [α-helix and β-sheet, hydrophathy, hydrophilicity according to Chou and Fasman (1974) and Kyte and Doolittle (1982)]. The physical properties of the amino acid sequences, molar extinction coefficient, the isoelectric points and molecular weights were calculated using PEPTIDESORT algorithms.

**Synthetic peptides**

All sequences were assembled manually using N-fluorenylmethoxy-carbonyl (Fmoc)-protected amino acids (Novabiochem, Läufelfingen, Switzerland) (Ferreira da Cruz et al., 1996). The purity of the three peptides, as determined by analytical reverse phase high performance liquid chromatography (HPLC; C18-column; water/acetonitrile/trifluoroacetic acid, gradient from 99:1:0:0 to 30:70:0:1), was >70%.
Small aliquots of peptides were purified by semi-preparative HPLC and subjected to a sequencer protein (PSQ-1, Shimadzu, Kyoto, Japan). The purified peptides were deposited on TFA-polybrene treated glass filters and the sequence conducted as described previously (Giovanni De Simone et al., 1994). The peptides were dissolved (1 mg ml⁻¹) in (PBS; pH 7) or (DMSO; pH 7) for the proliferation assay.

Peripheral blood mononuclear cells (PBMC) cultures

The PBMC, obtained from ten infected and ten uninfected individuals, were incubated for 5 days at 37°C in 5% CO₂ with Roswell Park Memorial Institute (RPMI) 1640 medium. The latter supplemented with 10% human AB serum contained 0, 3, 10 or 30 μg ml⁻¹ of the different peptides, Leishmania brasilienensis lysate (30 μg ml⁻¹) or 1 μg ml⁻¹ of concanavalin A (Con A) in triplicate. The proliferative response was assessed by measuring the incorporation of radiolabelled thymidine (DeLuca et al., 1999) and expressed as the stimulation index calculated from mean counts per minute (cpm) of experimental cultures divided by the mean of the control cpm cultures, i.e. cultures without antigen or Con A.

Results and discussion

The dendrogram generated by PILEUP algorithm (Figure 1) shows one principal cluster containing Lpcys1, Lpcys2, Lmacpa, Lmecpb, Lmecpc, Ldccys1, Ldccys2, CatL and CatS, and another cluster disclosing a relationship among Lmecpa, Lmacpb and CatB. A similarity graph plotted with sequence data from all molecules studied revealed that the COOH-terminal regions in the alignment had scores close to zero in regard to similarity, whereas most of the remaining sections had varying degrees of similarity (Figure 2). Indeed, sequences showed that the COOH-terminal regions of Leishmania cysteine proteinases did not relate to the primary sequence of our human cysteine proteinase studies (not shown). Therefore, it was likely that the COOH-terminal sequence (355-444) of Leishmania cysteine proteinases possessed appropriate epitopes for human T-cell activation, more than the other regions. Indeed, since it did not cross-react with their human proteinase counterparts, there was no immunological tolerance to them. This sequence had three stretches with a hydrophobic score above 0.6 (Figure 3). The peptides corresponding to these stretch sequences were synthesized and tested as T-cell epitopes: εε ε ε (361-370), pH (416-423) and pH (431-444).
Figure 1: Structural dendrogram profile of proteins of the cystatine-proteinase family. The dendrogram was built with the PILEUP algorithm based on the amino acid sequence of human cathepsin L (CatL: A77913), cathepsin B (CatB: a55432) and cathepsin S (CatS: A89553) and L. pifanoi Lpcys1 (B84766) and Lpcys2 (P06894), L. amazonensis Lmecpa (U73793) and Lmecph (P06400) and Lmecpc (P68999), L. major Lmecpa (P06770) and Lmecph (U73793), L. donovani Ldcys1 (A80499) and L. chagasi Ldcys2 (A80499). The closeness of branches represents the degree of homology between different molecules.

It has been demonstrated that hydrophobic residues are required for the interactions of peptides with a conserved hydrophobic pocket in HLA class II molecules (Hill et al., 1991; Brown et al., 1993). The chosen sequences may have a β-sheet (pI and pII) or an α-helix (pII) configuration (Figure 3). The secondary structure of the peptides may well play a role in determining the extent to which individual amino acids are exposed to interactions with sites within the MHC pocket.
Figure 2 Average similarity amongst all sequences. The similarity score was calculated from PILEUP algorithms (Feng and Doolittle, 1987; Higgs and Sharp, 1989), based on the amino acid sequence of human and Leishmania cysteine proteinases, using the PLOTSIMILARITY algorithm (Smith and Waterman, 1981). Shaded areas indicate the amino acid positions with lower similarity scores.

Figure 3 Hydrophobic, hydrophilic and secondary structure profile of Cyspap. The peptide sequence was predicted by exclusion analysis of human cysteine proteinase sequence hydrophobicity and hydrophilicity values were calculated according to Kyte and Doolittle (1982). For plotting α-helix and β-sheet structures, the method of Chou and Fasman (1978) was used. The shaded regions identify the sequences of the selected peptides with hydrophobic scores of >0.6: pl (VMVEVDVCFDi, pl (VGGGLCFE) and plII (PYLGGVSTCIVYT).
Determination of conformational and structural characteristics of the peptides binding to HLA molecules is crucial for defining potential T-cell epitopes. The algorithms used in this study helped us to determine which peptides possessed the structural features which may be important for the binding of small peptides from *Leishmania* cysteine proteinases to HLA molecules. The selected peptides pl, pII and pIII, from the cysteine sequence, have highly hydrophobic scores which is a significant feature of binding, comparable with peptides which bind to HLA-DR1Dw1 molecules (Jordečky et al., 1990; Hill et al., 1991).

Charged residues in both HLA class II and peptide sequences are also important for binding. The charge, calculated by PEPTIDESORT algorithm, shows a motif containing a negatively charged residue in pI and pH (data not shown), which may bind to the positively charged HLA class II molecules: DPw4b, DQb, DQw9 and DRw4 alleles (Boehmcke et al., 1993).

The isolectric points (IP) of the predicted epitopes, calculated by PEPTIDESORT algorithm, were above the processing and the binding pH. The IP were calculated from the amino acid sequence alone and do not take into account the changes in pKa of the side groups in response to tertiary environment. Moreover, it did not account for glycosylation. It was observed that pl, pH and pII had IP of 3.58, 3.90 and 7.29, respectively, while the IP of fifteen different HLA sequences, calculated in this work, varied from 4.4 to 11.26 (data not shown). The IP may be a crucial factor for the binding of peptides to MHC, since it may depend on peptide solubility. In accordance with this view, all HLA molecules examined in this paper, although having large IP, were soluble at the endocytic pathway pH.

The importance of pH for peptide-MHC interactions was further supported by observations on quantitative loading of peptides to class II in low pH conditions (Sadegh-Nasseri and Germain, 1991), and by the fact that the binding of some peptides was accelerated by the low pH of endosomal compartments (Jensen, 1990; Harding et al., 1991; Reay et al., 1992).

The T-cell antigenicity of the peptides was determined by testing their stimulatory effect on lymphocytes from patients with active cutaneous leishmaniasis, in vitro. Diagnosis was confirmed by visualization of *Leishmania* amastigotes in histopathological examination. The mean cpm in control cultures from normal donors ± standard deviation was 595 ± 30, and in the control cultures from infected donors was 280 ± 50. The proliferative responses to Con A, were intense and comparable in patients and normal donors (mean stimulation index (SI) ± standard deviation = 64 ± 25). The lympho-
proliferative response to *Leishmania braziliensis* antigen was positive (SI > 2.5 for 30 μg ml⁻¹ of antigen) in three out of the ten patients tested (Figure 4). Comparable proliferative responses (SI > 2.5) were observed, using the maximal peptide concentration (30 μg ml⁻¹) with peripheral blood mononuclear cells (PBMC) from two patients (Figure 4).

Our results demonstrate the feasibility of the proposed strategy for the identification of antigenic peptides and, as previously discussed, indicated that no single peptide will be immunogenic for all individuals. In consequence a peptide vaccine aimed at protecting a population against a micro-organism would need to be composed of several peptides.

**Acknowledgements**

This investigation received financial support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo a Pesquisa do Estado do Rio de Janeiro (FAPERJ) and FIOCRUZ. We thank the Multi-Media Sector of FIOCRUZ for expert photographic work. This paper is dedicated to the Oswaldo Cruz Institute on the occasion of the centenary of its foundation on 25 May 1900.
References


FERREIRA DE CRUZ M. F., Giovine De Simone S., Baric D., Cervelheiro M. C., Camus D. and Dentel Ribeiro C. 1996. Can software be used to predict antigenic regions in Plasmodium falciparum? Parasite Immunol. 18 159-61.


40 Cytobios

C. R. Alves et al.


Accepted 4 June 2000

41 T-cell epitopes on Leishmania cysteine proteases
Mapping of the N terminus of the *Schistosoma mansoni* tegumental antigen Sm15 to its predicted precursor protein

Roberto P. Werkhäuser, Salvatore G. De Simone, Edeneide Xavier, Silvia M.L. Montenegro, Frederico G.C. Abath

Departamento de Bioquímica e Biologia Molecular, Instituto Oswaldo Cruz, Av. Brazil 4305, 21045-900 Rio de Janeiro, RJ, Brazil

Laboratório de Bioquímica de Proteínas e Peptídeos, Departamento de Bioquímica e Biologia Molecular, Instituto Oswaldo Cruz, Av. Brazil 4305, 21045-900 Rio de Janeiro, RJ, Brazil

Received 30 March 2000; revised in revised form 15 May 2000; accepted 15 May 2000

Abstract

Sm15 is a major *Schistosoma mansoni* 15 kDa tegumental antigen, resulting from the proteolytic processing of a larger precursor. The amino terminus of Sm15 was identified by direct amino acid sequencing, and the antigen was tentatively mapped to the segment spanning amino acids 362–497 of the precursor. This will allow subsequent studies to elucidate the possible immunological role of proteolytic processing in schistosomiasis. © 2000 Australian Society for Parasitology Inc. Published by Elsevier Science Ltd. All rights reserved.

Keywords: *Schistosoma mansoni*; Tegumental antigen; Sm15; Precursor protein; Purification; Mapping

As previously demonstrated by antibodies from mice protectively vaccinated with adult worm tegumental membranes [1,2], Sm15 is a partially characterised major 15 kDa tegumental antigen present in 2 to 6-week-old *Schistosoma mansoni* worms [2,3]. The gene for Sm15 encodes a much longer protein than that identified in tegumental membranes, indicating that it encodes a precursor that is subsequently highly processed [2,3]. In addition, there is evidence that the precursor is differentially processed during maturation, providing the parasite with the capability of expressing different products of the gene at different life cycle stages [2]. In order to further characterize the biosynthesis of Sm15 and allow subsequent studies to elucidate the possible immunological role of proteolytic processing in schistosomiasis, we report the mapping of Sm15 to its precursor.

Although the precise coding region of Sm15 was unknown, previous work had shown that some of the epitopes were present between amino acids 692–1032 (λ, gt11 A70 cDNA) [3]. Thus, constructs overlapping this cDNA, were subcloned in pGEM™, expressed in *Escherichia coli* as GST-(glutathione-S-transferase)-fusion proteins, and analyzed by immunoblotting with monospecific antibodies to Sm15 (A70 RasM). These antibodies were obtained by affinity purifying anti-adult tegumental membrane antigen by adsorption to the λ gt11 A70 cDNA expressing a β-galactosidase-fusion protein [2]. In addition to recognising native Sm15, these antibodies recognized specifically the following GST-fusion proteins: 2TR7A (corresponding to the amino acids 444–788 of the Sm15 precursor), A70HG2 (amino acids 692–858) and A70-20 (amino acids 692–1032) (Figs. 1 and 2). Fig. 2 shows a region of repeats previously described in detail [3], and the relationship of the fusion proteins with this region. These results suggested that at least some major epitopes of Sm15 are located between amino acids 692–788, which represents the region of overlap of the three fusion proteins. Further work involved the purification and partial amino acid sequencing of native Sm15. The protein was isolated from adult tegumental membrane proteins obtained essentially according to the freezing-thawing method [4]. The preparation was analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)/immunoblotting, and the band, identified as Sm15 by rabbit anti-adult tegumental serum, was sequenced [5]. Partial sequencing resulted in the 43 amino acid sequence GTVGEm (with some uncertainty respecting the 6th residue), which maps to amino acids 362–367 of the sequence of the Sm15 precursor (Fig. 2),
Fig. 1. Western blot probed with ATU RoM, specific to Smp15. 1. GST-fusion protein 2TR7A; 2. GST-fusion protein A38H1-02; 3. GST-fusion protein A38H1-20; 4. GST; 5. S. luteovirens mosaics infected with S. luteovirens. M. molecular mass markers (kDa).

Fig. 2. Derived amino acid sequence of the Smp15 precursor (GenBank accession number L07513). See Ref. 3 showing the putative mapping of Smp15 (in bold) and the region of repeats (boxed). The amino acids that define the position of the fusion protein 2TR7A (amino acid 467-788), A38H1-02 (amino acids 692-852) and A38H1-20 (amino acids 662-1032) are indicated by arrows.

Acknowledgements

We are grateful to Paul Hagan for critical reading of the manuscript. We thank Mc Luiz C. O. Figueredo, Mrs Rosineide J. de Lima and Dr M. Canedo for their technical assistance. This communication was supported by WHO Special Program for Research and Training in Tropical Disease (TDR), Brazilian Research Council (CNPq), FIOCRUZ, FACEPE and FACEPE.