ELSEVIER

Acta Tropica 000 (2001) 000–000

ACTA TROPICA

www.parasitology-online.com

A cDNA encoding *Tc*-MUC-5, a mucin from *Toxocara canis* larvae identified by expression screening☆

Andrew Doedens [1], Alex Loukas [2], Rick M. Maizels *

Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK

Received 10 October 2000; received in revised form 29 November 2000; accepted 15 December 2000

Abstract

Toxocara canis is an ascarid nematode parasite of canids. Larvae infect a wide range of accidental hosts including humans, in whom they are the aetiologic agent of visceral and ocular *Larva migrans*. The labile surface coat of *T. canis* larvae consists of a family of mucin glycoproteins termed TES-120, for which the cDNAs have recently been cloned. In this paper, we describe the identification of a novel cDNA (*Tc-muc-5*) encoding an apomucin by expression screening of a cDNA library with antiserum raised to *T. canis* excretory/secretory products, and compare the predicted *Tc*-MUC-5 protein with those of other *T. canis* mucins (*Tc*-MUC-1–*Tc*-MUC-4) that include the TES-120 surface coat glycoproteins. *Tc*-MUC-5 has both a larger open reading frame and a more divergent sequence than the other *T. canis* mucins. It contains a putative signal peptide followed by two six-cysteine (SXC) domains, an extended threonine-rich central mucin core domain and two C-terminal SXC domains. Amino acid composition analysis of secreted TES-120 glycoproteins revealed a distinct lack of lysine residues; while this finding is in agreement with the primary sequences of *Tc*-MUC-1–*Tc*-MUC-4, *Tc*-MUC-5 is conspicuous by its relative abundance of lysines (6.7%), suggesting that this protein is not part of the TES-120 family of surface coat proteins.

Keywords: *Toxocara*; Nematode; Mucin; Glycoprotein; Excretory/secretory

1. Introduction

Mucins are heavily glycosylated proteins characterised by a core polypeptide backbone rich in serine and threonine residues that act as a scaffold for abundant *O*-linked carbohydrate side chains, generally accounting for more than one-half of the molecular weight of the glycoconjugate (Gendler and Spicer, 1995; Perez-Vilar and Hill, 1999). Assemblies of mucin macromolecules provide a viscous material that protects and lubri-

☆ The nucleotide sequence reported here has been deposited in the GenBank under accession number AF167710.

* Corresponding author. Tel: +44-131-6505511; Fax: +44-131-6505450.

E-mail address: r.maizels@ed.ac.uk (R.M. Maizels).

[1] Present address: Department of Medicine, Indiana University Medical Center, Indianapolis, IN 46202, USA.

[2] Molecular Parasitology Laboratory, Division of Infectious diseases and Immunology, Queensland Institute of Medical Research, 300 Herston Road, Queensland 4006, Australia

0001-706X/01/$ - see front matter
PII: S0001-706X(01)00137-1

cates epithelial surfaces, while also presenting an array of specific glycans involved in intercellular adhesion and cell trafficking. Most secreted mucins also contain flanking cysteine-rich domains, N- and/or C-terminal to the core Ser/Thr tract, thought to be responsible for concatemerisation into higher-order structures.

The surface of infective larvae of the parasitic roundworm *Toxocara canis* is a labile antigenic coat that consists primarily of a family of secreted mucins termed TES-120 (Page et al., 1992b). The surface coat is rapidly shed when the parasite is under immune attack along with adherent effector cells and antibodies (Badley et al., 1987b), providing *T. canis* with a dramatic means of evading immune destruction. The mucins are synthesised in large oesophageal and secretory glands within the larva (Page et al., 1992a), and are secreted in quantity by parasites maintained in vitro. Consequently, it has been postulated that they may provide an immunological diversion to protect parasites in vivo (Maizels and Loukas, 2001). With the definition of the dominant oligosaccharide side chains (Khoo et al., 1991), it has also been suggested that the secreted mucins of *Toxocara* compete with host sugars to block leukocyte infiltration into infected tissues. Thus, mucins may feature prominently in the immunobiology of nematode infections such as toxocariasis.

Three genes for members of the TES-120 family have been cloned from *T. canis* cDNA, and have been shown to encode surface coat/secreted mucins (Gems and Maizels, 1996; Loukas et al., 2000a); these have been designated *Tc-muc-1* to *Tc-muc-3*. In addition, a fourth mucin gene (*Tc-muc-4*) has been identified, whose sequence does not correspond to any of the identified surface/secreted molecules (Tetteh et al., 1999; Loukas et al., 2000a). Intriguingly, all four genes include two or four copies of a conserved 36-amino acid, six-cysteine domain, which may fulfil the cross-linking function of cysteine-rich flanking domains in vertebrate mucins. This motif, first designated NC6 (Gems et al., 1995; Gems and Maizels, 1996) and latterly as SXC (Blaxter, 1998; Loukas et al., 2000a), is now considered an evolutionarily mobile cassette fused to a variety of proteins of different origin and function, linked only by a possible association with secretion or surface expression by nematodes (Blaxter, 1998; Maizels et al., 2000).

Previous mucins from *T. canis* have been identified by molecular biological approaches that favour highly-expressed transcripts (Gems and Maizels, 1996; Tetteh et al., 1999). We now describe a fifth mucin-encoding cDNA, *Tc-muc-5*, which has been isolated by an immunological screen using antibodies to the excretory/secretory antigens of *T. canis* (TES). This new mucin gene encodes a product that is both larger and more divergent than the other *T. canis* mucins, but that is unlikely to be a member of the TES-120 family of surface coat glycoproteins. Thus, *Tc-muc-5* represents either a somatic mucin, or a new type of secreted mucin not previously described.

2. Materials and methods

2.1. Production of mouse anti-TES serum

BALB/c mice were immunised with 10 μg TES products emulsified in Freund's complete adjuvant followed by two boosts at days 28 and 42 with 10 μg TES in Freund's incomplete adjuvant, administered via the subcutaneous route. Sera were collected at day 49. Antiserum was precleared of antibodies to *Escherichia coli* according to the manufacturer's instructions (Stratagene PicoBlue Immunoscreening protocol).

2.2. cDNA library: general methods and immunoscreening

A cDNA library made from mRNA of *T. canis* infective larvae cloned into the λ-phage vector Uni-ZAP XR (Stratagene) was a generous gift from Dr C. Tripp and Dr R. Grieve (Heska Inc., Fort Collins, CO). The library was used to infect *E. coli* strain XL-1 blue, and plated at semi-confluency onto 9 cm agar plates containing tetracycline (12.5 μg ml^{-1}) according to the manufacturer's instructions (Stratagene PicoBlue Immunoscreening protocol). Plates were incubated at 42°C until plaques became visible, then nitrocellulose filters presoaked in 10 mM IPTG were placed on the agar plates and incubated at

37°C for 3 h to induce production of recombinant proteins. Filters were then removed, rinsed in phosphate-buffered saline (PBS) and blocked overnight in 5% skimmed milk powder in PBS at 4°C. Filters were rinsed twice with PBS/0.05% Tween-20 (PBS-T) and probed with pre-cleared mouse anti-TES serum diluted 1:1000 in PBS-T/3% skimmed milk. Filters were incubated in primary sera at room temperature for 2 h with gentle rocking, and washed three times in PBS-T for 15 min before being incubated with horseradish peroxidase-conjugated rabbit anti-mouse immunoglobulin G (BioRad; diluted 1:2000 in PBS-T/3% skimmed milk) for 1 h at room temperature. Filters were washed as earlier and developed using TMB substrate (Kirkegaard and Perry Laboratories). Putative positive plaques were cored from the agar plate, phage eluted from the agar plug and replated for secondary screening according to the manufacturer's instructions (Stratagene).

2.3. Preparation of plasmid, sequencing and sequence analysis

Individual plaques that were positive by secondary screen were cored out and phage DNA rescued in vivo using ExAssist™ helper phage and non-suppressing *E. coli* strain SOLR™ (Stratagene). Plasmid containing the insert was prepared from bacterial cultures using the Qiagen spin miniprep kit (Qiagen). Double-stranded *Tc-muc-5* cDNA was sequenced using universal forward and reverse primers, and internal gene-specific primers. Sequencing was performed using the AmpliTaq DyeDeoxy Terminator Cycle Sequencing System (Applied Biosystems) and an automated ABI Prism 377 DNA sequencer (Applied Biosystems). Analyses of nucleotide and deduced amino acid sequences were performed using the MacVector 6.5.1 software program (Oxford Molecular). Database searches were performed using the BLAST server to search a non-redundant set of databases (Altschul et al., 1990). Sequence alignments were performed using MacVector 6.5.1. The translated protein sequence was analysed using the SignalP web server (http://www.cbs.dtu.dk/services/SignalP/) in order to identify a putative signal peptide and corresponding cleavage point.

2.4. Cloning of the 5′ end of Tc-muc-5

The original immunopositive *Tc-muc-5* clone was truncated at a point later determined to be 274 nucleotides (nt) from the 5′ end. The cDNAs encoding other *T. canis* apomucins all contain the spliced leader-1 (SL-1) sequence at their 5′ termini (Gems and Maizels, 1996; Loukas et al., 2000a). The 5′ end of *muc-5* was amplified from the cDNA library under standard polymerase chain reaction conditions using an internal gene-specific primer (5′-CAGAAGCCGCAGGTTTTCCTGC-3′, complementary to nt 305–327 in Fig. 1) in conjunction with a primer designed to bind to SL-1. A single amplicon was generated and cloned into the T-ended vector pGEM-T (Promega) and sequenced.

3. Results

Full-length *Tc-muc-5* cDNA was sequenced and found to contain 1458 base pairs including 30 and 479 nucleotides of 5′ and 3′ untranslated regions, respectively (Fig. 1). The 5′ end is *trans*-spliced with SL-1 and there is a potential polyadenylation signal (AATAAA) that begins 23 nucleotides upstream of the poly-A tail. The predicted protein contains a signal peptide with a potential cleavage site between Ala-23 and Ser-24. After removal of the signal peptide, but before any post-translational modification, the *Tc*-MUC-5 protein has a predicted molecular weight of 29.9 kDa. Significantly, in comparison with known features of the TES-120 mucins, MUC-5 contains 21 lysine residues comprising some 6.7% of all amino acids in the protein.

The mature *Tc*-MUC-5 polypeptide can be divided into three segments, with cysteine-rich SXC domains at either terminus, and a central mucin-like core. This central domain (amino acids 100–245) is rich in threonine (45%) and alanine (27%) residues, typical of apomucins, and is likely to form the scaffold for post-translational addition of *O*-linked saccharides. There are four repeats of the heptapeptide PTTTTAA, but while the remainder of the central domain is rich in Thr/Ala, the sequence is non-repetitive.

There are four six-cysteine (SXC) domains in the mature protein, two at the N-terminus beginning shortly after the signal peptide cleavage site, and two at the C-terminus after the Thr-rich mucin core domain (Fig. 2A). As with other *T. canis* proteins known to contain SXC domains,

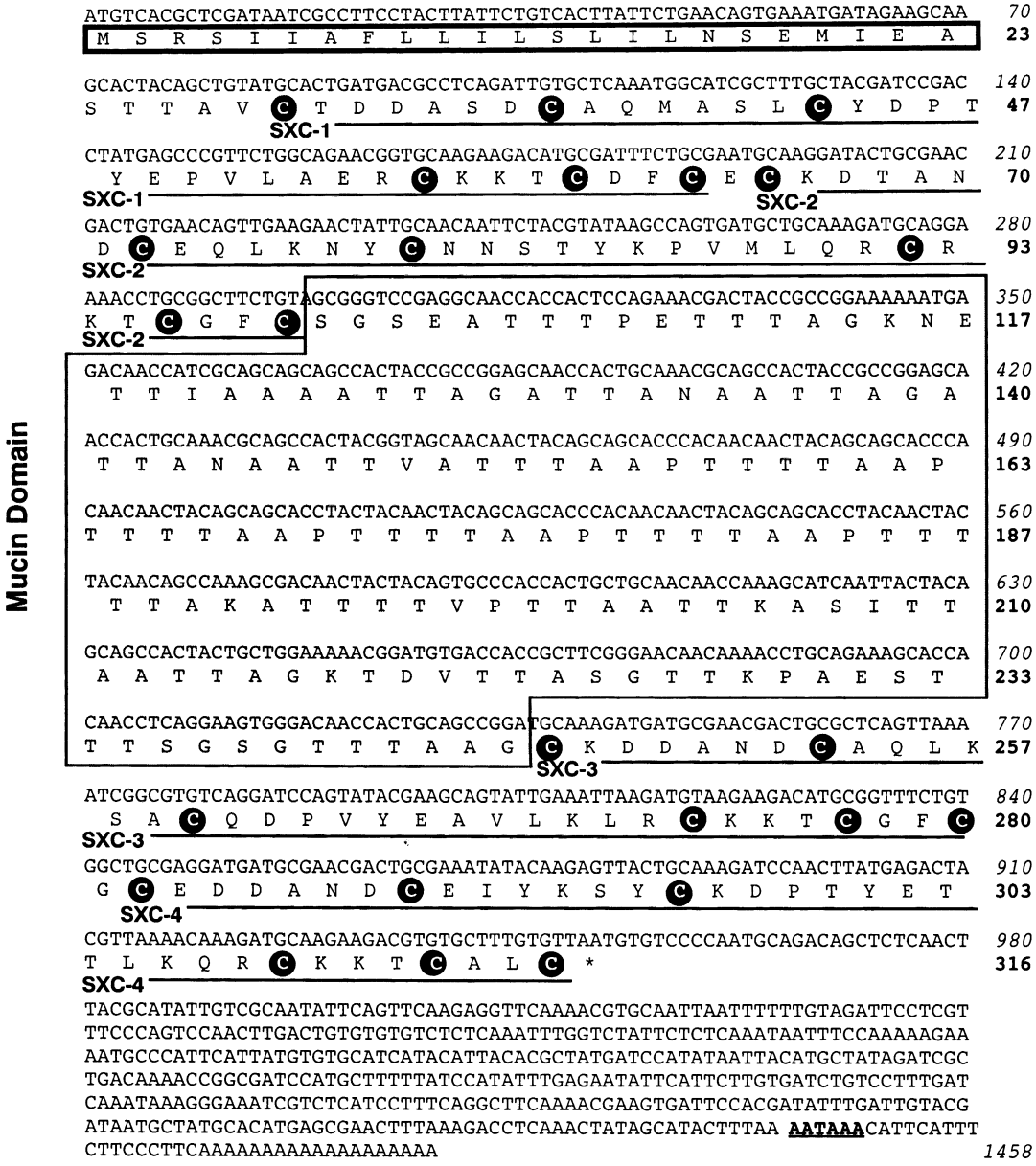

Fig. 1. THe nucleotide and deduced amino acid sequence of *Tc-muc-5* cDNA. Nucleotide numbers are indicated on the right in italics, while amino acid numbers of the predicted MUC-5 protein are in bold. The putative signal peptide of the predicted protein (amino acids 1–23) is shown in an open, thick-lined box. The four SXC domains are indicated by heavy underlining, and the six cysteines in each domain identified by solid circles. The mucin domain is enclosed by a shaded box. The putative polyadenylation signal is underlined in bold type close to the 3′ end of the sequence.

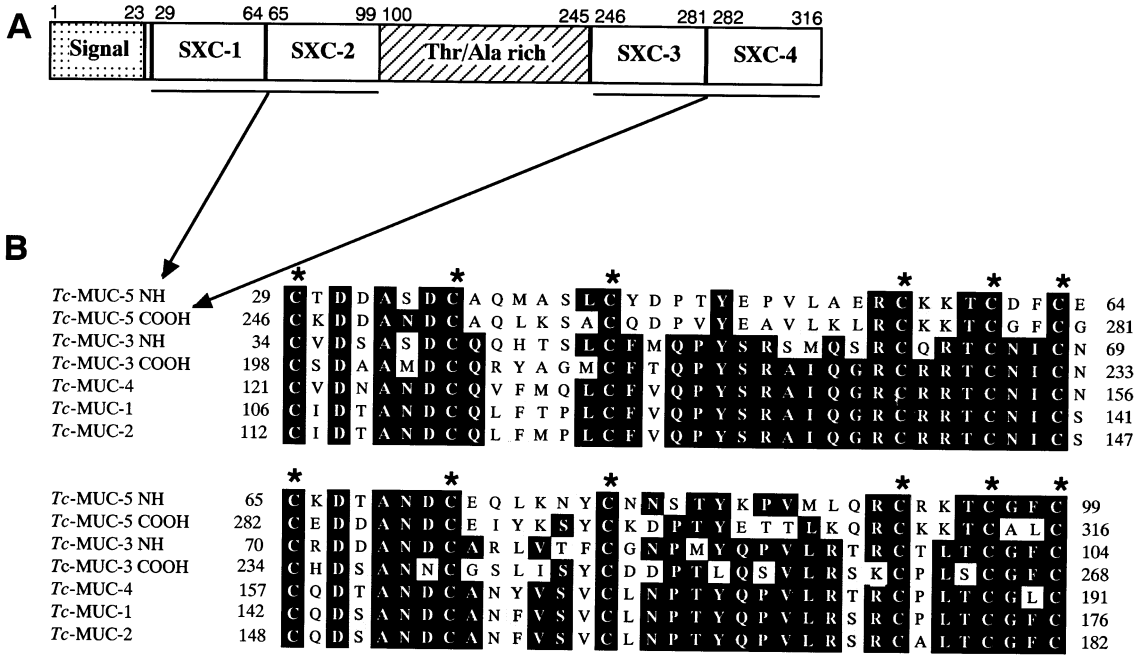

Fig. 2. Schematic representation of the domain structure of the *Tc*-MUC-5 protein (A) and alignment of the SXC domains of MUC-5 with those from other known *T. canis* mucins (B). Numbering indicates the amino acid positions of the domains within their respective proteins. Asterisks mark the cysteine residues of each paired SXC domain.

these 36 amino acid motifs occur in tandem pairs, each with six identically spaced cysteine residues (Fig. 2B). The N-terminal SXC domains of *Tc*-MUC-5 share 49% identity with their closest homologue, the N-terminal SXC domains of *Tc*-MUC-3, while the C-terminal SXC domains of *Tc*-MUC-5 share 45% identity with their closest homologue, the *Tc*-SXC domains of *Tc*-MUC-4 (Fig. 2B). A pattern has previously been noted that the first SXC domain of each pair is more similar to other first domains than to any second domain. Interestingly, this is not observed in *Tc*-MUC-5, in which all SXC domains are more related to the typical second SXC. For example, the first domains of *Tc*-MUC-1 to *Tc*-MUC-4 contain residues QPY at positions 18–20 of the motif, while the second domains contain PTY. In the case of *Tc*-MUC-5, PTY and PVY are found in this position in the first domains of both N- and C-terminal motifs.

4. Discussion

Mucins are conspicuous gene products of certain parasitic pathogens (Di Noia et al., 1998; Freitas-Junior et al., 1998; Loukas et al., 2000a). Previous studies have highlighted the TES-120 family of *T. canis* mucins that compose the surface coat of infective larvae. Biochemical analysis of the TES-120 mucins has identified four closely migrating *O*-linked glycoproteins. Using a combination of amino acid composition analysis, mass determination, tryptic peptide sequence and Western blot analysis, three genes for TES-120 products (*Tc-muc-1*–*Tc-muc-3*) have been unequivocally identified, while a fourth remains to be found (Loukas et al., 2000a). Amino acid composition data showed an absence of Lys residues from all four of the purified TES-120 proteins, consistent with earlier metabolic labelling studies showing that [^{14}C]-lysine is not incorporated into

TES-120 (Gems and Maizels, 1996). Because the new gene described here, *Tc-muc-5*, contains a total of 21 Lys (6.7%) residues, we conclude that it cannot encode one of the four known TES-120 components.

Tc-MUC-5 is in many respects the most divergent of the five mucins so far characterised in *T. canis*. First, it alone has non-repetitive sequence in the mucin domain, in contrast to the other four mucins that comprise 10–12 tandem repeats each of six to seven amino acids. Secondly, the 146 amino acid mucin domain is significantly longer than the others, which range from 84 amino acids (*Tc*-MUC-1) to 99 amino acids (*Tc*-MUC-4). Third, although *Tc*-MUC-5 shares with *Tc*-MUC-3 the presence of paired SXC domains at both the N and C termini, the amino acid sequence within this domain is only distantly related to that present in any of the other mucins. Thus, the *Tc*-MUC-5 SXC domains show 34–60% identity to their closest relatives in MUC-1/MUC-4, while within the latter group identities range from 51 to 97%. Despite this divergence, it remains the case that all *T. canis* mucins contain recognisable SXC domains, which have hitherto been linked with the export of proteins to the surface or to secretory pathways (Gems et al., 1995; Gems and Maizels, 1996; Blaxter, 1998; Loukas et al., 2000a).

The *Tc-muc-5* cDNA was isolated by screening with antiserum raised to TES products. Although *Tc*-MUC-5 might represent a minor constituent of TES-120, its proportion would have to be very low for it not to have been previously detected (Loukas et al., 2000a). Alternatively, *T. canis*, like many parasitic protozoa (Hoeijmakers et al., 1980; Smith et al., 1995), might selectively vary expression of surface proteins to change the antigenicity of its surface coat. This explanation would require *Tc-muc-5* to be expressed in the larval *T. canis* used to generate both the cDNA library and the TES used for immunization, but not in the parasites taken in separate studies for biochemical analysis of TES. This appears to us to be unlikely.

A second possibility is that *Tc*-MUC-5 represents a different TES protein. While a number of TES products, including TES-26 (Gems et al., 1995), TES-32 (Loukas et al., 1999) and TES-70, have been cloned (Loukas et al., 2000b), it has been estimated that up to 50 distinct proteins are secreted by *T. canis* larvae (Badley et al., 1987a). Because mucins undergo heavy post-translational processing, mostly in the form of *O*-glycosylation, the molecular weight of mature *Tc*-MUC-5 cannot be estimated. However, *T. canis* soluble extracts and secretions contain large, heavily glycosylated molecules other than TES-120, such as TES-400 (Page and Maizels, 1992), where the corresponding cDNAs have not been identified.

Alternatively, the new mucin may be employed internally; for example, in the lining of the gut, as suggested for parasitic trematodes (Carlier et al., 1980; Marin et al., 1992). In an attempt to immunolocalise *Tc*-MUC-5, we cloned the N-terminal SXC domains one and two into the expression vector pET-15b to induce expression of recombinant protein in *E. coli*. Recombinant *Tc*-MUC-5 protein was produced in extremely low amounts (data not shown) despite attempts to optimise production, suggesting that the SXC domains of *Tc*-MUC-5 are toxic to the bacteria. Future work will focus on production of larger amounts of protein and generation of an antiserum so that the native molecular weight and site of expression of mature *Tc*-MUC-5 can be determined.

Parasite mucins might be involved in creating an immunological smokescreen by forming antigen–antibody complexes. Larval *T. canis* shed their mucinous surface coat when under immunological attack (Badley et al., 1987b), and experimental murine infections with *T. canis* larvae result in the formation of granulomas whether parasites are present or absent (Parsons et al., 1986), suggesting the deposition of antigens to divert the immune response from migrating larvae. *Tc*-MUC-5 is the most divergent of the *T. canis* mucins discovered to date and its localisation to structures within the larval parasite will provide insight into its role in nematode infections. Given that mucins provide a lubricating, protective barrier over underlying structures, chemotherapy or vaccines aimed at inhibiting or interfering with the deposition of mucins on these surfaces offer an attractive target for novel approaches in controlling helminth infections.

Acknowledgements

The authors thank Kevin Tetteh for comparisons between SXC domains in *T. canis* proteins. This work was supported by a grant from the Medical Research Council.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Badley, J.E., Grieve, R.B., Bowman, D.D., Glickman, L.T., Rockey, J.H., 1987a. Analysis of *Toxocara canis* larval excretory-secretory antigens: physicochemical characterization and antibody recognition. J. Parasitol. 73, 593–600.

Badley, J.E., Grieve, R.B., Rockey, J.H., Glickman, L.T., 1987b. Immune-mediated adherence of eosinophils to *Toxocara canis* infective larvae: the role of excretory–secretory antigens. Parasite Immunol. 9, 133–143.

Blaxter, M., 1998. *Caenorhabditis elegans* is a nematode. Science 282, 2041–2046.

Carlier, Y., Bout, D., Strecker, G., Debray, H., Capron, A., 1980. Purification, immunochemical, and biologic characterization of the *Schistosoma* circulating M antigen. J. Immunol. 124, 2442–2450.

Di Noia, J.M., D'Orso, I., Aslund, L., Sanchez, D.O., Frasch, A.C., 1998. The *Trypanosoma cruzi* mucin family is transcribed from hundreds of genes having hypervariable regions. J. Biol. Chem. 273, 10843–10850.

Freitas-Junior, L.H., Briones, M.R., Schenkman, S., 1998. Two distinct groups of mucin-like genes are differentially expressed in the developmental stages of *Trypanosoma cruzi*. Mol. Biochem. Parasitol. 93, 101–114.

Gems, D., Maizels, R.M., 1996. An abundantly expressed mucin-like protein from *Toxocara canis* infective larvae: the precursor of the larval surface coat glycoproteins. Proc. Natl. Acad. Sci. USA 93, 1665–1670.

Gems, D., Ferguson, C.J., Robertson, B.D., Nieves, R., Page, A.P., Blaxter, M.L., Maizels, R.M., 1995. An abundant, trans-spliced mRNA from *Toxocara canis* infective larvae encodes a 26-kDa protein with homology to phosphatidylethanolamine-binding proteins. J. Biol. Chem. 270, 18517–18522.

Gendler, S.J., Spicer, A.P., 1995. Epithelial mucin genes. Annu. Rev. Physiol. 57, 607–634.

Hoeijmakers, J.H., Frasch, A.C., Bernards, A., Borst, P., Cross, G.A., 1980. Novel expression-linked copies of the genes for variant surface antigens in trypanosomes. Nature 284, 78–80.

Khoo, K.-H., Maizels, R.M., Page, A.P., Taylor, G.W., Rendell, N., Dell, A., 1991. Characterisation of nematode glycoproteins: the major *O*-glycans of *Toxocara* excretory secretory antigens are methylated trisaccharides. Glycobiology 1, 163–171.

Loukas, A., Mullin, N.P., Tetteh, K.K.A., Moens, L., Maizels, R.M., 1999. A novel C-type lectin secreted by a tissue-dwelling parasitic nematode. Current Biol. 9, 825–828.

Loukas, A., Hintz, M., Tetteh, K.K.A., Mullin, N.P., Maizels, R.M., 2000. A family of secreted mucins from the parasitic nematode *Toxocara canis* bear diverse mucin domains but share similar flanking six-cysteine (SXC) repeat motifs. J. Biol. Chem. (in press).

Loukas, A., Doedens, A., Hintz, M., Maizels, R.M., 2000b. Identification of a new C-type lectin, TES-70, secreted by infective larvae of *Toxocara canis*, which binds to host endothelial cells. Parasitology 121, 545–554.

Maizels, R.M., Loukas, A., 2001. In: Kennedy, M.W., Harnett, W. (Eds.), Parasitic Nematodes — Molecular Biology, Biochemistry and Immunology. CABI Publishing (in press).

Maizels, R.M., Tetteh, K.K.A., Loukas, A., 2000. *Toxocara canis*: genes expressed by the arrested infective larval stage of a parasitic nematode. Int. J. Parasitol. 30, 495–508.

Marin, M.S., Prieto, M., Martin, J.M., Casais, R., Boga, J.A., Parra, F., 1992. Identification and expression of a *Fasciola hepatica* gene encoding a gut antigen protein bearing repetitive sequences. Mol. Biochem. Parasitol. 55, 155–165.

Page, A.P., Maizels, R.M., 1992. Biosynthesis and glycosylation of serine/threonine-rich secreted proteins from *Toxocara canis* larvae. Parasitology 105, 297–308.

Page, A.P., Hamilton, A.J., Maizels, R.M., 1992a. *Toxocara canis*: monoclonal antibodies to carbohydrate epitopes of secreted (TES) antigens localize to different secretion-related structures in infective larvae. Exp. Parasitol. 75, 56–71.

Page, A.P., Rudin, W., Fluri, E., Blaxter, M.L., Maizels, R.M., 1992b. *Toxocara canis*: a labile antigenic surface coat overlying the epicuticle of infective larvae. Exp. Parasitol. 75, 72–86.

Parsons, J.C., Bowman, D.D., Grieve, R.B., 1986. Tissue localization of excretory-secretory antigens of larval *Toxocara canis* in acute and chronic murine toxocariasis. Am. J. Trop. Med. Hyg. 35, 974–981.

Perez-Vilar, J., Hill, R.L., 1999. The structure and assembly of secreted mucins. J. Biol. Chem. 274, 31751–31754.

Smith, J.D., Chitnis, C.E., Craig, A.G., Roberts, D.J., Hudson-Taylor, D.E., Peterson, D.S., Pinches, R., Newbold, C.I., Miller, L.H., 1995. Switches in expression of *Plasmodum falciparum* var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell 82, 101–110.

Tetteh, K.K.A., Loukas, A., Tripp, C., Maizels, R.M., 1999. Identification of abundantly-expressed novel and conserved genes from infective stage larvae of *Toxocara canis* by an expressed sequence tag strategy. Infect. Immun. 67, 4771–4779.