Biotechnology Advances 27 (2009) 376–388

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

A portrait of the "SCP/TAPS" proteins of eukaryotes — Developing a framework for fundamental research and biotechnological outcomes

C. Cantacessi [a], B.E. Campbell [a], A. Visser [b], P. Geldhof [b], M.J. Nolan [a], A.J. Nisbet [c], J.B. Matthews [c], A. Loukas [d], A. Hofmann [e], D. Otranto [f], P.W. Sternberg [g], R.B. Gasser [a,*]

[a] *Department of Veterinary Science, The University of Melbourne, Australia*
[b] *Laboratory for Parasitology, The University of Gent, Belgium*
[c] *Moredun Research Institute, Scotland, United Kingdom*
[d] *Division of Infectious Diseases, Queensland Institute of Medical Research, Australia*
[e] *Eskitis Institute for Cell and Molecular Therapies, Griffith University, Australia*
[f] *Department of Veterinary Public Health, Faculty of Veterinary Medicine, Bari, Italy*
[g] *Biology Division, California Institute of Technology, Pasadena, California, USA*

ARTICLE INFO

Article history:
Received 10 December 2008
Received in revised form 5 February 2009
Accepted 11 February 2009
Available online 21 February 2009

Keywords:
Sperm coating protein-extracellular domain
SCP/TAPS
Activation-associated secreted proteins
Hookworms
Classification
Nomenclature
Bayesian inference
Venom allergen
Pathogenesis-related proteins

ABSTRACT

A wide range of proteins belonging to the SCP/TAPS "family" has been described for various eukaryotic organisms, including plants and animals (vertebrates and invertebrates, such as helminths). Although SCP/TAPS proteins have been proposed to play key roles in a number of fundamental biological processes, such as host–pathogen interactions and defence mechanisms, there is a paucity of information on their genetic relationships, structures and functions, and there is no standardised nomenclature for these proteins. A detailed analysis of the relationships of members of the SCP/TAPS family of proteins, based on key protein signatures, could provide a foundation for investigating these areas. In this article, we review the current state of knowledge of key SCP/TAPS proteins of eukaryotes, with an emphasis on those from parasitic helminths, and undertake a comprehensive, systematic phylogenetic analysis of currently available full-length protein sequence data (considering characteristic protein signatures or motifs) to infer relationships and provide a framework (based on statistical support) for the naming of these proteins. This framework is intended to guide genomic and molecular biological explorations of key SCP/TAPS molecules associated with infectious diseases of plants and animals. In particular, fundamental investigations of these molecules in parasites and the integration of structural and functional data could lead to new and innovative approaches for the control of parasitic diseases, with important biotechnological outcomes.

Contents

* Corresponding author. Tel.: +61 3 97312000; fax: +61 3 97312366.
E-mail address: robinbg@unimelb.edu.au (R.B. Gasser).

0734-9750/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.biotechadv.2009.02.005

1. Introduction

Pathogens have evolved a range of strategies to invade their hosts, while hosts have developed immune and other defence mechanisms against pathogens (reviewed by Schmid-Hempel, 2008). Numerous host–pathogen interactions involve molecular processes, in which proteins and other molecules from a pathogen target host proteins to initiate or maintain the infection, stimulate or evade the host's immune response and/or cause disease (reviewed by Schmid-Hempel, 2008). Various groups of proteins have been proposed to play major biological roles in the host–pathogen interplay (e.g., Vermeire et al., 2008). Among these molecules are the sperm-coating protein (SCP)-like extracellular proteins, also called SCP/Tpx-1/Ag5/PR-1/Sc7 (SCP/TAPS; Pfam accession number no. PF00188). SCP/TAPS family members have been identified in various eukaryotes and belong to the cysteine-rich secretory protein (CRISP) "superfamily" (Chalmers et al., 2008). SCP/TAPS proteins include rodent sperm-coating glycoproteins (or acidic epididymal glycoproteins, proposed to be involved in sperm maturation during its passage through the epididymis) (Jalkanen et al., 2005), mammalian testis-specific protein (Tpx-1) (Kasahara et al., 1989), glioma pathogenesis-related protein (Murphy et al., 1995; Yamakawa et al., 1998; Rosenzweig et al., 2006), venom allergen 5 from vespid wasps and the venom allergen 3 from fire ants, which mediate allergic reactions to the bites by some insects of the order Hymenoptera (Lu et al., 1993) as well as plant pathogenesis proteins (PRPs) of the PR-1 "subfamily" which are synthesized in response to infections with pathogens or other stress-inducing factors (reviewed by van Loon et al., 2006).

Data for a range of eukaryotes suggest that all SCP/TAPS molecules share a common primary structure; the signal peptide is followed by the SCP-extracellular domain (InterProScan: IPR014044), which is proposed to act as a Ca^{2+}-chelator in various signalling processes (Fernández et al., 1997). In the SCP/TAPS proteins of yeast, the SCP-domain is variable in length, polymorphic and flanked by a threonine-rich region. The C-terminus of family members representing mammals and reptile venoms harbours the CRISP-domain, typified by ten conserved cysteine residues. In invertebrates, such as parasitic helminths (including roundworms and flatworms), this portion of the protein is smaller and contains only four to six cysteine residues (Yatsuda et al., 2002).

SCP/TAPS homologues have been identified in a range of invertebrates, particularly arthropods, roundworms (nematodes), flatworms (trematodes), and plants. Despite the roles that SCP/TAPS proteins are proposed to play in fundamental biological processes in various eukaryotes (e.g., van Loon et al., 2006; Gibbs et al., 2008), attempts to define criteria for the unequivocal classification of these molecules are scant and limited to SCP/TAPS proteins of the vinegar fly, *Drosophila melanogaster* (see Kovalick and Griffin, 2005), the blood fluke *Schistosoma mansoni* (see Chalmers et al., 2008) and plants (van Loon, 1990). To date, 17 different "subfamilies" of plant PRPs have been proposed and named in the order in which homologues with defined biological and biochemical properties were discovered (van Loon et al., 2006). In parasitic nematodes, *Ancylostoma*-secreted proteins or activation-associated secreted proteins (ASPs) were first characterized from hookworms and subsequently from related strongylid nematodes (e.g., Visser et al., 2008). As ASPs are abundant in the excretory/secretory (ES) products of the infective third-stage larvae (L3), they are thought to play an important role in the transition from the free-living to the parasitic stage during the invasion of the host (Hawdon et al., 1996, 1999; Moser et al., 2005; Datu et al., 2008). Because of their immunogenic properties, one ASP (namely *Na*-ASP-2) is under investigation as a vaccine candidate against the disease (=necatoriasis) caused by the hookworm *Necator americanus* in humans (Bethony et al., 2005; Loukas et al., 2006; Mendez et al., 2008; Xiao et al., 2008). Attempts to classify ASPs have been based on the number and features of the SCP-domain(s) of these molecules. Thus far, three types of ASPs have been described: 'double domain ASPs' which have two distinct but related SCP-domains, and the 'C-type single domain ASPs' and 'N-type single domain ASPs' which have the highest homology to the C- and N-terminus of the SCP-double domain ASPs, respectively (Geldhof et al., 2003). C-type single domain and double domain ASPs have been identified in a range of nematodes parasitic in animals and plants (Visser et al., 2008). To date, N-type single domain ASPs have been characterized only for *Cooperia punctata*, *Ostertagia ostertagi* and *Teladorsagia circumcincta*, three gastrointestinal trichostrongylid nematodes of ruminants (Yatsuda et al., 2002; Geldhof et al., 2003; Smith et al., 2009).

Usually, new SCP/TAPS gene orthologues or protein homologues identified in parasitic nematodes (based on nucleotide and/or amino acid sequence identities or homologies) are named without consideration of the SCP/TAPS protein group as a whole. Also, names are sometimes assigned to SCP/TAPS gene orthologues/protein homologues for which only partial sequence data are provided or available, which has resulted in an unclear relationship among gene(s), transcript(s) and protein(s). The confusion and inconsistencies in classification are exacerbated, as different authors use different names for the same groups of molecules. Besides their usual designation as ASPs, SCP/TAPS molecules of parasitic nematodes have also been named 'ASP-like' (AL), based on their homology to ASPs, identified first in *Ancylostoma* (Saverwyns et al., 2008), and 'venom allergen-like' (VAL; Frank et al., 1996; Murray et al., 2001; Chalmers et al., 2008), 'venom allergen-homologues' (VAH, Tetteh et al., 1999; Anand et al., 2007) and/or 'venom allergen-proteins' (VAP; Gao et al., 2001), based on the homology with the major allergen in the venom of the yellow jacket wasp (Henriksen et al., 2001). In the free-living nematode *Caenorhabitis elegans*, at least 17 different ASP-related genes have been identified (cf. Zhan et al., 2003) and given various names (e.g., *vap* or *scl* = 'SCP-like'). In *C. elegans*, SCP/TAPS molecules have been shown to be involved in biological aspects such as anti-microbial activity (O'Rourke et al., 2006), normal body formation (Wang and Kim, 2003) and lifespan (Ookuma et al., 2003).

Given that SCP/TAPS proteins are considered by most authors to be of major biological importance in a wide range of eukaryotes (plants, vertebrates and invertebrates), the inconsistencies in classification as well as the lack of structural and functional information for most of them, in the present article, (i) we reviewed the current state of knowledge of these molecules, with an emphasis on those from parasitic helminths, (ii) undertook a comprehensive, systematic phylogenetic analysis of current protein sequence data (considering characteristic protein signatures/motifs) to infer relationships and (iii) proposed a framework (based on statistical support) for the naming of these proteins. This framework is intended to support genomic and molecular biological explorations of key SCP/TAPS molecules associated with infectious diseases in plants and animals. In particular, fundamental investigations of these molecules in parasites, using novel genomic, proteomic and bioinformatic technologies, and the integration of structural and