2011-08

Gene structure and splicing in schistosomes

Journal of Proteomics, Amsterdam : Elsevier, v. 74, n. 9, p. 1515-1518, Aug. 2011
http://www.producao.usp.br/handle/BDPI/49645

JOURNAL OF PROTEOMICS 74 (2011) 1515–1518

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Review

Gene structure and splicing in schistosomes

Sergio Verjovski-Almeida[a,*], *Ricardo DeMarco*[b]

[a]*Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, 05508-900 São Paulo, SP, Brazil*
[b]*Departamento de Física e Informática, Instituto de Física de São Carlos, Universidade de São Paulo, 13560-970 São Carlos, SP, Brazil*

ARTICLE INFO

Available online 1 April 2011

Keywords:
Alternative splicing
Protein variability
Micro-exon genes
Schistosoma mansoni
Secreted proteins
Host–parasite interface

ABSTRACT

Schistosomes are blood dwelling platyhelminths with a complex life cycle and persist in the definitive host during decades, indicating that they are very successful parasites. The challenge of infecting two hosts from different evolutionary branches, namely an invertebrate snail and a vertebrate, suggests that Schistosomes must display a very sophisticated genetic program to circumvent all the barriers imposed by the hosts' immune systems. Recent large-scale genome and transcriptome data from Schistosomes are facilitating the analysis of gene structure and splicing. Studying the structure of genes coding for secreted proteins is of particular interest since these proteins mediate processes in the host–parasite interface. Recent description of Micro-Exon Genes (MEGs), polymorphic mucin genes (SmPoMucs) and venom allergen-like (SmVALs) proteins with unusual gene structure apparently oriented towards generation of protein variability through alternate splicing, and the presence of multiple copies of these genes, indicate that the parasite developed a sophisticated system to interact with its hosts. This opens up opportunities for further studies with the use of proteomic techniques to better characterize the protein variability created by these systems and their role in parasite survival. Complete description of the functions of these variable proteins will greatly contribute to our understanding of host–parasite interactions.

Contents

* *Corresponding author*. Tel.: +55 11 3091 2173x203; fax: +55 11 3091 2173x202.
E-mail address: verjo@iq.usp.br (S. Verjovski-Almeida).

1874-3919/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jprot.2011.03.022

1. Study of the schistosome genome and transcriptome

After finishing of the human genome sequencing ~24,000 protein-coding genes were detected [1], representing just ~1.8-times the number of genes from simpler organisms such as fly and worm. In contrast, it was verified that on average each human gene produced ~1.5 transcripts as a result of alternate splicing, originating a larger number of proteins per gene in comparison to those of simpler organisms [2]. This highlighted the importance of gene structure and its creative usage by alternate splicing as a mechanism to generate diversity/polymorphism in living organisms. In fact, alternative splicing is ubiquitous in eukaryotes, with a higher frequency in vertebrates in relation to invertebrates [3].

Study of gene structure and alternative splicing in schistosomes have been limited by the lack of a large collection of genomic sequences. A pioneering effort of construction of a *Schistosoma mansoni* bacterial artificial chromosome library has been made in the early 00's [4], but the limited number of sequences generated from this library remained insufficient for a large scale evaluation of schistosome gene structure. Transcript sequence data has also been produced through several EST sequencing efforts [5–10] in the late 90's and early 00's which were able to produce few tens of thousands ESTs. Despite these difficulties some interesting data about peculiarities in schistosome gene structure has been published. This includes the description of very small introns ranging from 31–42 bp in the hypoxanthine–guanine phosphoribosyltransferase gene [11] and in SmIMP25 Integral Membrane-Protein [12]. Also of interest, the existence of trans-spliced genes [13] in *S. mansoni* has been described. Trans-spliced events detected in *S. mansoni* resulted in the addition of a spliced leader in a small fraction of the parasite's RNAs. No bias towards trans-splicing of any particular class of gene or in any parasite tissue was detected [13].

Recent sequencing of the genome from two different species of Schistosomes, *S. mansoni* [14] and *S. japonicum* [15], in conjunction with gene expression information obtained by large scale transcriptome sequencing of *S. mansoni* [16] and *S. japonicum* [17], provided for the first time the possibility to describe the structure of most genes from the two parasites. It is important to note, however, that both genome and transcriptome data in these projects are partial and several genes are still incompletely described or wrongly predicted. With the recent advance of sequencing technologies [18], it is expected that we will soon have a considerable increase in the amount of sequencing data for Schistosomes, permitting a more accurate description of their genes.

2. Alternative splicing of transcripts coding for proteins regulating transcription or splicing

With the use of genomic data, some studies correlated alternative splicing with different aspects of Schistosome biology. This has permitted researchers to obtain hints of the molecular mechanisms underneath important processes. Alternative splicing in transcripts coding for proteins involved in control of transcription and splicing is a subject of particular interest, since a change in these processes may impact several transcripts, thus producing a cascade effect that may be responsible for drastic changes in the parasite.

Using the genome sequence and transcript data from semi-quantitative RT-PCR, a gender biased differential alternative splicing of the SmCA150 gene in *S. mansoni* was described [19]. CA150 is a transcriptional cofactor that has been shown in model organisms to interact with the spliceosome and influence the alternative splicing of genes in vitro. The observed alternative splicing of SmCA150 could create a cascade effect that may be responsible for the differences in splicing of several other genes between male and female [19].

Other interesting works have studied the generation of alternatively spliced messages for transcription factor genes, namely a heat-shock transcription factor of *S. mansoni* [20] and a homologue of epidermal growth factor receptor [21]. In the first case, it was possible to detect an alternative splicing that produced a truncated form that was more prevalent in cercariae than in adult worm. This was correlated with the higher expression of HSP-70 in adults in comparison to cercariae [20]. In the latter case, an alternative promoter site was detected generating two different transcripts, one with an additional exon. However, such differences in sequence did not generate a difference in the coding region, and further studies are required to evaluate the impact of the different UTRs on stability of the messages [21].

3. Structure and splicing of genes coding for secreted proteins

Interestingly, some recent publications focused on the structures of *S. mansoni* genes coding for protein products that are secreted by the parasite. As we will review below, some of these works describe quite unusual gene structures, suggesting that these parasites must have developed sophisticated genetic systems to interact with their hosts.

One remarkable example of a unique gene structure described with information from genome and transcriptome data from *S. mansoni* is the Micro-Exon Genes (MEGs) [22]. MEGs are genes that have the majority of their coding region composed of very small (≤36 bp) and symmetrical exons. The design of this system permits alternate splicing of these exons in a manner that the addition/removal of exons will not disrupt the reading frame because of the symmetry of the exons and will generate transcripts coding for proteins with difference in a few amino acids [14,22]. Some MEGs display several copies in the genome, which increases the potential variability.

Micro-exons have been detected in a number of model organisms (*Arabdopsis thaliana*, *Caenorhabditis elegans* and *Drosophila melanogaster*) and in humans [23] however, all examples involve one single micro-exon per gene [23]. No genes with several micro-exons organized in tandem were ever described outside the Schistosoma genera. This implies that such MEGs have been recently created and may represent an adaptation to the parasitic lifestyles of Schistosomes.

MEGs code for proteins with signal peptide and several of their protein products have been detected in secretions or immuno-localized in schistosome glands [22]. Moreover, their

transcription is increased in the intra-mammalian stages of their life cycle [22]. This data indicates that such proteins are being directly exposed to the host and its immune system.

RT-PCR experiments confirmed that MEGs produce alternatively spliced transcripts. 2D electrophoresis of schistosomula and egg secretions revealed several spots representing proteins of different PI and molecular weight, and mass spectrometry of the tryptic digestion from these spots identified them as products of MEG-2 and MEG-3 [22]. Although a 2D electrophoresis clearly shows very distinct proteins, the peptides detected by mass spectrometry were encoded by adjacent micro-exons with no examples of exon skipping. This was interpreted as resulting from the fact that each spot must represent a mixture of several slightly different isoforms and the peptides from adjacent exons must predominate in the pool of tryptic peptides [22]. If this were the case, only the utilization of multiple protein separation methods and proteomic analyses of the resulting tryptic peptides would improve the resolution of the process, and allow the detection of the more rare peptides that would confirm the utilization of alternatively spliced micro-exons for generation of different protein variants.

Another set of genes that code for secreted proteins and display an unusual gene structure is that of polymorphic mucins (SmPoMucs). They are a multi-gene family with all members (~10) displaying a structure with tandem repeats (ranging from 0 to 20 repeats) of one 27 bp exon near its 5' end [24]. Apparently, during their evolution these 4 different subfamilies suffered a recombination between themselves allowing mosaic genes creation and contributing to a higher variability. Events of alternative splicing and trans-splicing between transcripts of two subfamilies contributed to generate an even higher variability of messages, creating a scenario defined by the authors as a "controlled chaos" [24]. 2D electrophoresis experiments with sporocytes protein extract showed groupings of spots with slightly different isoelectric point and molecular weight, identified as SmPoMucs by mass spectrometry [24,25], indicating that the variability observed in the transcripts resulted in the production of different proteins, as expected. In addition, these SmPoMucs are glycosylated in their O-Glycosylation sites and the alterations in their sequences as a result of splicing appear to affect their glycosylation patterns.

It is interesting to note that both MEGs and SmPoMucs have different complex genetic systems to produce protein variation, including the alternative splicing of small exons. However, for SmPoMucs most of the variation due to alternative splicing is restricted to the region containing the tandem repeats, while in MEGs the small exons cover almost the entire coding region, making most of the protein susceptible to variation. Also, MEGs appear to be mainly expressed in the intra-mammalian stages while SmPoMucs are predominant in the stages in contact with the invertebrate host. SmPoMucs have several O-glycosylation sites on their structures, in contrast only a few MEGs display this kind of glycosylation site. The immune systems from vertebrate and invertebrate hosts are very different, making it likely that distinct strategies must be employed to circumvent them. In this respect, Moné et al. [26] have shown that a specific set of highly variable immune receptors (FREP3, Fibrinogen Related Proteins) of the mollusk *Biomphalaria glabrata* formed immune complexes with highly polymorphic and individually variable mucin determinants from its specific trematode parasite *S. mansoni* (the SmPoMucs). Furthermore, recent results obtained by Hanington et al. [27] show that FREP3 plays a central role in the resistance of *B. glabrata* to infection with *S. mansoni,* thus supporting that *S. mansoni* polymorphic secreted proteins, such as SmPoMucs, can play a key role in host–parasite interplay.

Another interesting class of secreted proteins for which alternative splicing appears to have an important role in creating different protein isoforms is the *S. mansoni* venom allergen-like (SmVAL) gene family. This family was first described during investigation of the *S. mansoni* transcriptome due to its similarity to wasp venom allergen [16]. Further studies of such family were conducted by Chalmer et al. [28], who described the organization of such genes in clusters and their expression patterns throughout the parasite life cycle. Although gene structures of most SmVAL are unremarkable, one of its members, namely SmVAL6, has a very unusual structure. The first four gene exons have a usual size (average 115 bp) and code for the SCP/TAPS domain, which is present in all members of this family. The rest of SmVAL6, however, is constituted mostly by several small symmetrical exons [28]. This makes the structure pattern in this portion of SmVAL6 identical to that described for MEGs. In fact, we propose that, since it does not display a similarity to other SmVAL, this portion must represent a late acquisition of SmVAL6. Considering that the structure pattern is virtually identical to those observed in MEGs, the most likely event that generated this gene is a recombination between the ancestral of SmVAL6 that contained only the first four exons with a MEG, creating a composite structure. Furthermore both MEGs and VALs are secreted proteins; it is expected that both portions of the SmVAL6 protein have the potential to interact with the human host, and a possible synergistic action from both proteins may explain why this mosaic gene was fixed in *S. mansoni*. Interestingly, analysis of SmVAL6 transcripts by Chalmers et al. [28] has identified alternative splicing of many small exons within the 3' region, resulting in as many as 35 isoforms. In addition, statistically significant, developmentally regulated, alternative splicing of two of these exons was observed [28]. We propose that experiments such as pull-down immuno-precipitation of SmVAL6 proteins under non-denaturing conditions combined with mass-spectrometry could identify specific and diverse interacting partners of such different protein variants.

4. Conclusions

Although the importance of alternative splicing in genetic programming of the organisms is being recognized, its study in Schistosomes is still incipient with a few studies on selected genes. Nonetheless, some interesting findings have been reported. The existence of exotic and unique gene structures apparently aimed to generate secreted protein variability is of special interest. The fact that these systems produce proteins that are being secreted to enter in contact with the host and that they are apparently exclusive from Schistosomes make us believe that they essentially represent adaptation to the parasitic life style. This opens up the opportunity of further studies to better characterize protein variation created by these systems with the use of proteomic techniques and other studies

aiming to determine the function of these proteins in the interaction of the parasite with the host.

REFERENCES

[1] Collins FS, Lander ES, Rogers J, Waterston RH, Conso IHGS. Finishing the euchromatic sequence of the human genome. Nature 2004;431:931–45.

[2] Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860–921.

[3] Keren H, Lev-Maor G, Ast G. Alternative splicing and evolution: diversification, exon definition and function. Nat Rev Genet 2010;11:345–55.

[4] Le Paslier MC, Pierce RJ, Merlin F, Hirai H, Wu WJ, Williams DL, et al. Construction and characterization of a *Schistosoma mansoni* bacterial artificial chromosome library. Genomics 2000;65:87–94.

[5] Franco GR, Adams MD, Soares MB, Simpson AJG, Venter JC, Pena SDJ. Identification of new *Schistosoma mansoni* genes by the EST strategy using a directional cDNA library. Gene 1995;152:141–7.

[6] Franco GR, Rabelo EM, Azevedo V, Pena HB, Ortega JM, Santos TM, et al. Evaluation of cDNA libraries from different developmental stages of *Schistosoma mansoni* for production of expressed sequence tags (ESTs). DNA Res 1997;4:231–40.

[7] Santos TM, Johnston DA, Azevedo V, Ridgers IL, Martinez MF, Marotta GB, et al. Analysis of the gene expression profile of *Schistosoma mansoni* cercariae using the expressed sequence tag approach. Mol Biochem Parasitol 1999;103:79–97.

[8] Fung MC, Lau MT, Chen XG. Expressed sequence tag (EST) analysis of a *Schistosoma japonicum* cercariae cDNA library. Acta Trop 2002;82:215–24.

[9] Fietto JLR, DeMarco R, Verjovski-Almeida S. Use of degenerate primers and touchdown PCR for construction of cDNA libraries. Biotechniques 2002;32:1404.

[10] Merrick JM, Osman A, Tsai J, Quackenbush J, LoVerde PT, Lee NH. The *Schistosoma mansoni* gene index: gene discovery and biology by reconstruction and analysis of expressed gene sequences. J Parasitol 2003;89:261–9.

[11] Craig SP, Muralidhar MG, Mckerrow JH, Wang CC. Evidence for a class of very small introns in the gene for hypoxanthine–guanine phosphoribosyltransferase in *Schistosoma mansoni*. Nucleic Acids Res 1989;17:1635–47.

[12] Markovics A, Ram D, Grossman Z, Ziv E, Lantner F, Schechter I. Cloning and characterization of the SmIMP25 Integral Membrane-Protein of the parasitic helminth *Schistosoma mansoni*. Bba Gene Struct Expr 1994;1218:273–82.

[13] Davis RE, Hardwick C, Tavernier P, Hodgson S, Singh H. RNA trans-splicing in flatworms. Analysis of trans-spliced mRNAs and genes in the human parasite, *Schistosoma mansoni*. J Biol Chem 1995;270:21813–9.

[14] Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, Cerqueira GC, et al. The genome of the blood fluke *Schistosoma mansoni*. Nature 2009;460:352–8.

[15] Zhou Y, Zheng HJ, Chen YY, Zhang L, Wang K, Guo J, et al. The *Schistosoma japonicum* genome reveals features of host–parasite interplay. Nature 2009;460:345–51.

[16] Verjovski-Almeida S, DeMarco R, Martins EAL, Guimaraes PEM, Ojopi EPB, Paquola ACM, et al. Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*. Nat Genet 2003;35:148–57.

[17] Hu W, Yan Q, Shen DK, Liu F, Zhu ZD, Song HD, et al. Evolutionary and biomedical implications of a *Schistosoma japonicum* complementary DNA resource. Nat Genet 2003;35: 139–47.

[18] Schuster SC. Next-generation sequencing transforms today's biology. Nat Methods 2008;5:16–8.

[19] DeMarco R, Oliveira KC, Venancio TM, Verjovski-Almeida S. Gender biased differential alternative splicing patterns of the transcriptional cofactor CA150 gene in *Schistosoma mansoni*. Mol Biochem Parasit 2006;150:123–31.

[20] Ram D, Ziv E, Lantner F, Lardans V, Schechter I. Stage-specific alternative splicing of the heat-shock transcription factor during the life-cycle of *Schistosoma mansoni*. Parasitology 2004;129:587–96.

[21] Shoemaker CB, Ramachandran H, Landa A, dos Reis MG, Stein LD. Alternative splicing of the *Schistosoma mansoni* gene encoding a homologue of epidermal growth factor receptor. Mol Biochem Parasitol 1992;53:17–32.

[22] DeMarco R, Mathieson W, Manuel SJ, Dillon GP, Curwen RS, Ashton PD, et al. Protein variation in blood-dwelling schistosome worms generated by differential splicing of micro-exon gene transcripts. Genome Res 2010;20:1112–21.

[23] Volfovsky N, Haas BJ, Salzberg SL. Computational discovery of internal micro-exons. Genome Res 2003;13:1216–21.

[24] Roger E, Grunau C, Pierce RJ, Hirai H, Gourbal B, Galinier R, et al. Controlled chaos of polymorphic mucins in a metazoan parasite (*Schistosoma mansoni*) interacting with its invertebrate host (*Biomphalaria glabrata*). Plos Negl Trop D 2008;2:e330.

[25] Roger E, Gourbal B, Grunau C, Pierce RJ, Galinier R, Mitta G. Expression analysis of highly polymorphic mucin proteins (Sm PoMuc) from the parasite *Schistosoma mansoni*. Mol Biochem Parasit 2008;157:217–27.

[26] Mone Y, Gourbal B, Duval D, Du Pasquier L, Kieffer-Jaquinod S, Mitta G. A large repertoire of parasite epitopes matched by a large repertoire of host immune receptors in an invertebrate host/parasite model. PLoS Negl Trop Dis 2010;4:e813.

[27] Hanington PC, Forys MA, Dragoo JW, Zhang SM, Adema CM, Loker ES. Role for a somatically diversified lectin in resistance of an invertebrate to parasite infection. Proc Natl Acad Sci USA 2010;107:21087–92.

[28] Chalmers IW, McArdle AJ, Coulson RMR, Wagner MA, Schmid R, Hirai H, et al. Developmentally regulated expression, alternative splicing and distinct sub-groupings in members of the *Schistosoma mansoni* venom allergen-like (SmVAL) gene family. BMC Genomics 2008;9:89.