Title Polymorphism of sheep MHC Class IIb gene *TAPASIN*

Authors N. Siva Subramaniam, E. F. Morgan, C. Y. Lee, J.D. Wetherall and D.M. Groth

Western Australian Biomedical Research Institute (WABRI) & Centre for Health Innovation Research Institute, School of Biomedical Sciences, Curtin University of Technology, Perth, Western Australia, 6845

Correspondence Associate Professor David Groth, School of Biomedical Sciences, Curtin University, GPO Box U1987, Perth, Western Australia, 6845.
Fax: 61892662342
Email: D.Groth@curtin.edu.au

Abstract

The Major Histocompatibility Complex (MHC) is one of the most gene dense regions in the genome and studies in several species have shown significant associations between the MHC and disease. The endoplasmic reticular glycoprotein, tapasin, is involved in the MHC class I antigen presentation pathway. Sheep *TAPASIN* is located in the class IIb region of the MHC. Sheep *TAPASIN* was subcloned from BAC and cosmid genomic clones and DNA sequenced. *TAPASIN* is 9549 bp in length and encodes a protein of 447 amino acids. The structure of sheep *TAPASIN* was similar to other mammals and consisted of eight exons with a distinctively larger intron between exon three and four. Sheep *TAPASIN* gene had high sequence identity with other mammalian *TAPASINs*. The *TAPASIN* gene sequence is conserved across many mammalian species and is possibly maintained through purifying selection with the average ratio of ds/dn of 3.9. Twenty-six SNPs in sheep *TAPASIN* were identified.

Keywords *TAPASIN*, TAP-binding protein, MHC, sheep, SNP, polymorphism

Tapasin is an MHC class IIb region encoded (*TAPASIN*) protein involved in the classical class I antigen presentation pathway. Tapasin, also known as transporter associated with antigen processing binding protein (TAPBP), is an endoplasmic reticular glycoprotein (Garbi et al. 2003). In the class I presentation pathway, tapasin has several important functions such as the recruitment of transporters associated with antigen processing protein (TAP) in the endoplasmic reticulum (ER) (Lehner et al. 1998), stabilisation of TAP and class I molecule interactions required for optimal peptide loading (Ortmann et al. 1997), assembly of the class I heavy chain (Garbi et al. 2000), peptide selection in the peptide loading complex (PLC) (Garbi et al. 2000;

Howarth et al. 2004) and retention of the class I molecules in the endoplasmic reticulum (Grandea III *et al.* 1995; Grandea III *et al.* 2000; Grandea III & Van Kaer 2001). Mutations within the *TAPASIN* gene can significantly disrupt the functional role of tapasin in the MHC class I presentation pathway (Copeman et al. 1998). The "loss of function" phenotype results in a decrease of antigen presentation at the cell surface (Copeman et al. 1998). The *TAPASIN* gene in sheep is located within the class IIb region of the sheep MHC on chromosome 20 (Mahdy et al. 1989). Unlike humans and most other mammals, the class II region in sheep is split into IIa and IIb regions separated by a non-MHC region (Liu et al. 2006).

Several studies have shown direct or indirect associations between the sheep MHC and Quantitative Trait Loci (QTL) for resistance to disease. In Scottish blackface sheep naturally infested with *Ostertagia circumcincta*, certain MHC class II antigens were associated with 98% lower egg count (Schwaiger et al. 1995). A recent study in Scottish blackface sheep aimed at identifying QTLs for a variety of parasite resistance indicators showed that there was a significant chromosome wide QTL located within the class IIb region (Davies et al. 2006). Significant associations between MHC and intestinal nematodes have been reported in several other studies (Outteridge et al. 1996; Paterson et al. 1998; van Haeringen et al. 1999). Recently, microarray analysis of the response of Perendale sheep to nematodes showed that the more resistant lambs had higher expression of the MHC class II genes (Diez-Tascon et al. 2005). In Rhonschaf sheep, significant associations between faecal egg counts (FEC) and the markers *OarCp73*, *DYMS1* and *BM1815* was observed. The *DYA* gene located within the class IIb subregion of the MHC is closely linked to the microsatellite *DYMS* and

is a possible candidate gene for conferring resistance to *Haemonchus contortus* in these sheep (Charon 2004).

This report describes the structure, DNA sequence and single nucleotide polymorphisms (SNP) of *TAPASIN*, which is located within the sheep MHC Class IIb gene region on chromosome 20. An analysis of the ratio of synonymous to non-synonymous substitutions in the coding region of this gene across several mammalian species was also performed. This study is part of an ongoing development of a haplotypic analysis of the sheep MHC class II region. Information from this study will assist in understanding the important role genes within the MHC play in immunity to disease.

Genomic DNA was isolated from the leukocytes from individual merino sheep using a Qiagen tissue DNA isolation kit according to manufacturer's instructions. DNA was suspended in TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0) and stored at -20^{o}C until required. *TAPASIN* degenerative PCR primers (5'CTGYCTTGYRTCCCACTTCT3'and 5'CCAGGGTGACCTCAGCRCTG 3') were identified from the alignment of human and mouse *TAPASIN* gene sequences. These primers amplify a small 259bp DNA fragment of the sheep TAPASIN gene (Qin 2009) that could be sequenced and from which overgo primers (Gustafson et al. 2003) were designed (5'TGCAGAGAGGCTTACAGAGCCATC3' and 5'CCTGAGACATCACTCAGATGGCTC3'). The PCR reaction mix (50 ul) comprised: 100 ng sheep genomic DNA, 1 X PCR buffer (Invitrogen), 1.5 mM $MgCl_2$, 200 uM dNTP, 5 pmol of each primer (Geneworks), 0.2 mg/ml BSA (Roche) and 2.5U Platnium Taq polymerase (Invitrogen). The Overgo primers were radioactively labelled using overgo technology (Gustafson et al., 2003) and hybridised to BAC (CHORI 243) genomic DNA library filters in a buffered solution containing

20 × SSPE, 1%BSA, 7%SDS and 0.5M EDTA. The filters were incubated at 53^{o}C overnight, and washed in 1xSSC buffer containing 0.1% SDS. The filters were sealed in plastic and exposed to Kodak x-ray film with an intensifying screens at -80°C for 2 days prior to development of the film. DNA from the CHORI clone 461K3 clone was isolated using Qiagen large construct kit according to the manufacturer's protocol and digested with Pst I and cloned into pGEM5f. Random clones were selected and sequenced using M13 primers by Macrogen (Korea). Clones containing *TAPASIN* sequences were subjected to further sequencing.

Multiple pass DNA sequencing resulted in approximately 11 Kbp of quality sequence, with *TAPASIN* identified as being 9549 bp in length and encoding a predicted protein of 447 amino acids. The overall structure of the sheep *TAPASIN* gene (Genbank accession EU814901) was similar to that reported for several other mammals, cattle (NW_001494145), chimpanzee (NW_001236523), dog (NW_876254), human (NW_923073) and mouse (NW_001030615). A consensus genetic structure for sheep *TAPASIN* was determined from analysis of the predictions from GENSCAN (Burge & Karlin 1997) Burge and Karlin, 1997), TWINSCAN (Korf et al. 2001) and FGENESH (http://www.softberry.com). Sheep *TAPASIN* was found to comprise eight exons, with a distinctively larger intron of 5526 bp found between exons three and four.

Clustal W (Thompson et al. 1994) multiple sequence alignment was performed with *TAPASIN* from sheep, cattle, human, rat, mouse, horse, chimpanzee, macaca, dog and zebrafish. This analysis showed that there was significant amino acid sequence conservation (Figure 1). The various features of tapasin including signal peptide, C and V immunoglobulin domains, conserved cysteine residues, motifs 3 and 4, transmembrane region and cytoplasmic tail could all be identified and are also shown

in Figure 1. These features were identical to those shown previously for rat, Atlantic salmon and several other mammals (Ortmann *et al.* 1997; Deverson *et al.* 2001; Jorgensen *et al.* 2007). The average DNA sequence identity between the coding sequence of sheep *TAPASIN* gene and other organisms was 85.3% whereas the amino acid sequence identity was slightly lower at 83.6%. BLAST sequence alignment showed higher DNA sequence identity with cattle *TAPASIN* (95%) when compared to a similar analysis with human *TAPASIN* (84%). A bootstrapped Neighbour-Joining phylogenetic tree was constructed using the Clustal W multiple sequence alignment. 100 percent of the trees obtained placed the sheep and cattle tapasin in a separate clade from the primate tapasins (1000 bootstraps), indicating, as expected, that sheep tapasin is more closely related to cattle tapasin than to the primate tapasins (Sup. Figure 1). Relative to the cattle amino acid sequence, there are amino acid differences in the IgV domain (positions 170 and 171). However the observed change at position 170 appears to predate the generation of the ungulate lineage as it is also present in primates, horses and zebrafish. The variation position 171 however appears to be unique to sheep, at least in this small cohort of species. In the TPN motif 3 domain, there is a deleted amino acid in cattle relative to the other species. This position in sheep is the identical to horse *TAPASIN* implying a three base pair deletion has occurred in cattle. Unique amino acid changes occur at position 4 in the leader sequence and position 323 in the Ig C domain.

Synonymous and non-synonymous nucleotide substitutions were analysed using the SNAP software package (www.hiv.lanl.gov and http://hcv.lanl.gov/content/sequence/SNAP/SNAP.html) (Nei & Gojobori 1986; Korber 2000). This analysis showed that the *TAPASIN* coding sequence is conserved across all the mammalian species studied. The average ratio of Jukes-Cantor corrected

ds/dn was 3.91 (ranging from Jukes-Cantor corrected ds/dn of 2.48 for chimpanzee versus human to a Jukes-Cantor corrected ds/dn of 4.88 for sheep versus human). This high conservation is most likely maintained through purifying selection. Tapasin is an important component of the MHC class I assembly with several domains shown to have critical functions in the assembly process (Rizvi & Raghavan 2009). Therefore, it is not surprising that tapasin displays high amino acid sequence conservation between species. The amino acid sequence comparisons are shown in supplementary Figure 2.

Sequencing was performed on six unrelated merino sheep. Pairwise alignment of DNA sequences using Vector NTI software (Invitrogen) resulted in the direct identification of thirteen SNPs. Furthermore, when BLAST alignment of the consensus sheep sequence obtained in this study was performed against the genome sequence data generated by the International Sheep Genome Consortium (ISGC) (https://isgcdata.agresearch.co.nz/), thirteen additional SNPs were identified. Each SNP identified from the ISGC data occurred either within different breeds of sheep or multiple sheep within one breed. Seven of the 13 SNP (54%) identified in this project by sequence alignment of the 6 unrelated merino sheep were also independently verified through the BLAST alignment with the ISGC data. The estimated variation within the sheep *TAPASIN* gene is approximately one SNP per 367 bp. The SNP density observed for *TAPASIN* was comparable with the genome wide SNP survey (Kijas et al, 2009). All the SNPs identified in this study were located within non-coding regions of the gene. It was observed however, that the DNA sequence generated in this study when subject to BLAST alignment with the ISGC data, the matches in the ISGC data did not completely span the entire gene, suggesting possible

gaps in the ISGC sequence data. Details and position of all the SNPs identified in this study and the ISGC data are shown in Table 1.

The SNPs that have been identified in this study will contribute to an essential framework of SNPs in the MHC class IIb region and thus will be an important resource for the future characterisation and dissection of the sheep MHC haplotypes and their possible role in disease.

Acknowledgements

PhD Studentship support for NSS from WABRI and Curtin University. All animal experiments were performed according to the Australian Code of Practice for the care and use of the animals for scientific purposes. The authors would like to acknowledge the contribution of Dr Brian Dalrymple to the ISGC screening.

References

Burge C & Karlin S (1997) Prediction of complete gene structures in human genomic DNA. Journal of Molecular Biology 268: 78-94

Charon KM (2004) Genes Controlling Resistance to Gastrointestinal Nematodes in Ruminants. Animal Science Papaers and Reports 22: 135 - 9

Copeman J, Bangia NJC & Cresswell CP (1998) Elucidation of the genetic basis of the antigen presentation defects in the mutant cell line .220 reveals polymorphism and alternative splicing of the tapasin gene. European Journal of Immunology 28: 3783-91

Davies G, Stear MJ, Benothman M, Abuagob O, Kerr A, Mitchell S & Bishop SC (2006) Quantitative trait loci associated with parasitic infection in Scottish blackface sheep. Heredity 96: 252

Deverson EV, Powis SJ, Morrice NA, Herberg JA, Trowsdale J & Butcher GW (2001) Rat tapasin: cDNA cloning and identification as a component of the class I MHC assembly complex. Genes Immun 2: 48-51

Diez-Tascon C, Keane OM, Wilson T, Zadissa A, Hyndman DL, Baird DB, McEwan JC & Crawford AM (2005) Microarray analysis of selection lines from outbred populations to identify genes involved with nematode parasite resistance in sheep. Physiol. Genomics 21: 59-69

Garbi N, Tan P, Diehl AD, Chambers BJ, Ljunggren HG, Momburg F & Hammerling GJ (2000) Impaired immune responses and altered peptide repertoire in tapasin-deficient mice. Nature Immunology 1: 234

Garbi N, Tiwari N, Momburg F & Hammerling GJ (2003) A major role for tapasin as a stabilizer of the TAP peptide transporter and consequences for MHC class I expression. Immunology 33: 264-73

Grandea III AG, Androlewicz MJ, Athwal RS, Geraghty DE & Spies T (1995) Dependence of Peptide Binding by MHC Class I Molecules on Their Interaction with TAP. Science 270: 105-8

Grandea III AG, Golovina TN, Hamilton SE, Sriram V, Spies T, Brutkiewicz RR, Harty JT, Eisenlohr LC & Van Kaer L (2000) Impaired Assembly yet Normal Trafficking of MHC Class I Molecules in Tapasin Mutant Mice. Immunity 13: 213-22

Grandea III AG & Van Kaer L (2001) Tapasin: an ER chaperone that controls MHC class I assembly with peptide. Trends in Immunology 22: 194-9

Gustafson AL, Tallmadge RL, Ramlachan N, Miller D, Bird H, Antczak DF, Raudsepp T, Chowdhary BP & Skow LC (2003) An ordered BAC contig map of the equine major histocompatibility complex. Cytogenetic and Genome Research 102: 189

Howarth M, Williams A, Tolstrup AB & Elliott T (2004) Tapasin enhances MHC clss I peptide presentation according to peptide half-life. PNAS 101: 11737-42

Jorgensen SM, Grimholt U & Gjoen T (2007) Cloning and expression analysis of an Atlantic salmon (Salmo salar L.) tapasin gene. Dev Comp Immunol 31: 708-19

Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, et al. (2009) A Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds. PLoS ONE 4(3): e4668. doi:10.1371/journal.pone.0004668

Korber B (2000) HIV Signature and Sequence Variation Analysis. Computational Analysis of HIV Molecular Sequences, Chapter 4, pages 55-72. Allen G. Rodrigo and Gerald H. Learn, eds. Dordrecht, Netherlands: Kluwer Academic Publishers

Korf I, Flicek PD, Duan & Brent MR (2001) "Integrating genomic homology into gene structure prediction". Bioinformatics 17: 140-8

Lehner PJ, Surman MJ & Cresswell P (1998) Soluble Tapasin Restores MHC Class I Expression and Function in the Tapasin-Negative Cell Line .220. Immunity 8: 221-31

Liu H, Liu K, Wang J & Ma RZ (2006) A BAC clone-based physical map of ovine major histocompatibility complex. Genomics 88: 88-95

Mahdy EA, Makinen A, Chowdhary BP, Andersson L, Gustavsson I & (1989) Chromosomal localization of the ovine major histocompatibility complex (OLA) by in situ hybridization. Hereditas 111: 87-90

Nei M & Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3: 418-26

Ortmann B, Copeman J, Lehner PJ, Sadasivan B, Herberg JA, Grandea AG & et al (1997) A critical role for tapasin in the assembly and function of multimeric MHC class I-TAP complexes. Science 277: 1306-9

Outteridge PM, Andersson L, Douch PGC, Green RS, Gwakisa PS, Hohenhaus MA & Mikko S (1996) The PCR typing of MHC-DRB genes in the sheep using

primers for an intronic microsatellite: Application to nematode parasite resistance. Immunol Cell Biol 74: 330-6

Paterson S, Wilson K & Pemberton JM (1998) Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population. Proceedings of the National Academy of Sciences of the United States of America 95: 3714-9

Qin J (2009) Characterisation of the Central Region of the Sheep Major Histocompatibility Complex. Dissertation, Curtin University of Technology, Perth

Rizvi SM & Raghavan M (2009) Mechanisms of Function of Tapasin, a Critical Major Histocompatibility Complex Class I Assembly Factor. Traffic doi:10.1111/j.1600-0854.2009.01025.x

Schwaiger FW, Gostomski D, Stear MJ, Duncan JL, McKellar QA, Epplen JT & Buitkamp J (1995) An ovine Major histocompatibility complex DRB1 allele is associated with low faecal egg counts following natural, predominantly Ostertagia circumcincta infection. International Journal for Parasitology 25: 815-22

Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Research 22: 4673-80

van Haeringen WA, Gwakisa PS, Mikko S, Eythorsdottir E, Holm LE, Olsaker I, Outteridge P & Andersson L (1999) Heterozygosity excess at the cattle DRB locus revealed by large scale genotyping of two closely linked microsatellites. Animal Genetics 30: 169-76