Molecular & Biochemical Parasitology 183 (2012) 115–121

Contents lists available at SciVerse ScienceDirect

Molecular & Biochemical Parasitology

Characterization of the *eg95* gene family in the G6 genotype of *Echinococcus granulosus*☆

Cristian A. Alvarez Rojas[a,*], Charles G. Gauci[a], Matthew J. Nolan[a], Majid Fasihi Harandi[b], Marshall W. Lightowlers[a]

[a] *Faculty of Veterinary Science, The University of Melbourne, 250 Princes Highway, Werribee, Victoria 3030, Australia*
[b] *Department of Medical Parasitology, School of Medicine, Kerman University of Medical Sciences, Kerman, Iran*

ARTICLE INFO

Article history:
Received 5 January 2012
Received in revised form 7 February 2012
Accepted 8 February 2012
Available online 18 February 2012

Keywords:
Echinococcus granulosus
eg95
Vaccine
Genotype
Gene family

ABSTRACT

Cystic echinococcosis in humans and livestock animals is caused by infection with the cestode parasite *Echinococcus granulosus*. A number of genotypes of the parasite (designated G1–G10) are known to exist, with the genotype cluster G1–G3 and genotype G6 being responsible for the majority of humans infections. A recombinant vaccine has been developed for use in livestock to prevent infection with *E. granulosus*. The vaccine is based on the antigen EG95 which is expressed in the early larval stage (oncosphere) of the parasite. The EG95 antigen was originally cloned from the G1 genotype of *E. granulosus* and the protein has been found to be encoded by members of a small family of related genes in this genotype. Reliable information has not been available about the likely efficacy of the EG95 vaccine against genotypes other than G1. In this study, genomic DNA cloning techniques were used to characterize seven *eg95*-related gene fragments from the G6 genotype of *E. granulosus*. Three proteins appear to be encoded by these genes. Considerable differences were found between the EG95 related proteins from the G6 genotype compared with the EG95 protein from the G1 genotype. These differences suggest that the EG95-related proteins from the G6 genotype may have different antigenic epitopes compared with the current vaccine antigen. Data presented in this study have implications for future vaccine design and provide the information that would enable a G6 genotype-specific vaccine to be developed against *E. granulosus*, should this be considered a desirable addition to the available tools for control of cystic echinococcosis transmission.

1. Introduction

Echinococcus granulosus causes cystic echinococcosis in animals and humans. The parasite has a worldwide distribution with the highest prevalence occurring in parts of Eurasia, north and east Africa, and South America [1]. The disease has re-emerged in some areas and is still present in many areas where control programs have been implemented [2]. Genetic variability has long been recognized in *E. granulosus* with a number of genotypes described and designated G1–G10 [3–5]. The taxonomy of the genus has been reviewed and several genotypes of what were previously considered to be variants of *E. granulosus* are now considered to be regarded as distinct species [6,7]. The G1–G3 cluster genotype is the most common and is responsible for most human infection [8–12]. The G6 genotype is also an important aetiological agent of human cases in specific areas [13–15].

A vaccine has been developed for use against *E. granulosus* infections in livestock intermediate hosts so as to reduce transmission of the parasite and indirectly reduce the incidence of infection in humans [16]. The vaccine utilizes a recombinant protein, designated EG95, which is uniquely expressed in the parasite's oncosphere life cycle stage [17]. EG95 has been found to induce high levels of protection (96–100%) in experimental vaccine trials undertaken in sheep and other intermediate hosts in a number of countries against challenge infections with *E. granulosus* either known or believed to be of the G1 genotype [18–20]. Investigations by Chow et al. [21] found that the EG95 antigen was encoded by members of a family of genes in the G1 genotype, with four genes encoding the same EG95 protein antigen, while two other genes were found to encode related proteins and another was predicted to be a pseudogene. Currently an *E. granulosus* G1 genome sequencing program is being undertaken at the Wellcome Trust Sanger Institute led by Matt Berriman in collabo-ration with Cecilia Fernandez (Universidad de la Republica, Uruguay), however the available dataset (December 2011) [22] appear to be incomplete

Abbreviation: FnIII, Fibronectin III domain.

☆ *Note*: Nucleotide sequence data reported in this paper are available as GenBank ID: JQ285934, JQ285935, JQ285936, JQ285937, JQ285938, JQ285939 and JQ285940.

* Corresponding author. Tel.: +61 9731 2292; fax: +61 9731 2366.
E-mail address: cristianalvarezrojas@yahoo.com (C.A. Alvarez Rojas).

0166-6851/$ – see front matter © 2012 Published by Elsevier B.V.
doi:10.1016/j.molbiopara.2012.02.005

and do not yet provide accurate information about the number of *eg95*-related genes in the genome [23]. It was recognized from the earliest times in the development of the EG95 vaccine that there was a need to characterize EG95-related proteins in different isolates and genotypes of *E. granulosus* in order to assess the vaccine's potential for protection against different *E. granulosus* strains [24]. Subsequently, investigations into the variability of *eg95*-related genes have been undertaken by a number of groups using PCR-based strategies with non gene-specific primers [25,26]. One of these studies in particular revealed a high degree of variability in the *eg95* gene family members [25]. However the reliability of the data obtained is unclear because some of the amplified products could be the results of a number of different artifacts that are known to occur when PCR is used with non-gene specific primers [27]. A preliminary study of *eg95*-related genes in *E. granulosus* G6/G7 was performed by Chow et al. [28] using gene specific primers in PCR. A single *eg95*-related gene was identified from both G6 and G7 genotypes, showing substantial nucleotide variability with the *eg95* gene family members from G1. In the absence of complete and reliable data about *eg95*-related genes in the G6 genotype of *E. granulosus*, full characterization of *eg95*-related genes from this genotype was undertaken in this study.

2. Materials and methods

2.1. Extraction of parasite nucleic acids and Southern blot experiments

Fresh protoscoleces were collected from individual hydatid cysts from naturally infected camels in slaughterhouses in Iran. DNA was extracted as previously described [29]. Following phenol/chloroform extraction, total nucleic acids were precipitated with isopropanol and resuspended in sterile distilled water and stored at −20 °C. The parasite genotype was determined using PCR with genomic DNA as template and amplifying 366 bp from the *cox1* gene [3]. Genomic DNA was digested with *Eco*RI or *Xho*I, passively transferred to positively charged membrane [30] and hybridized with DIG-labelled probes. Three probes were generated by PCR using the *eg95* cDNA as template (entire *eg95* cDNA, 5′ and 3′ fragments of *eg95*) using primers previously described [21]. The probes were labelled with digoxigenin for chemiluminescent detection according to the manufacturer's protocol using DIG High Prime kit (Roche). Hybridization washes were performed at 65 °C in 0.5 × SSC (75 mM NaCl, 0.75 mM sodium citrate).

2.2. Genomic DNA library construction and screenings

Genomic DNA from protoscoleces derived from a single individual cyst confirmed to be G6 genotype was digested with either *Eco*RI or *Xho*I. Size selected DNA fragments ≤10 kb were ligated into λZAP II (Agilent Technologies). For DNA fragments greater than 10 kb, λDASH and FIX (Agilent Technologies) were used. Vectors were digested independently, ends phosphatased using Thermosensitive Alkaline Phosphatase (Promega) and subsequently vector arms were purified from agarose gel using β-agarase1 (New England Biolabs) after fragmentation. Each library was packaged with Gigapack III Gold (Agilent Technologies). Screenings, by DNA hybridization, were performed in a bacterial host recommended for each λvector using the *eg95* cDNA full length probe. Positive clones underwent both secondary and tertiary screenings to ensure clonality. For λZAP, the positive inserts were excised using the ExAssist interference-resistance Helper phage (Agilent Technologies). λDNA was extracted using conventional methods for FIX and DASH vectors [31].

2.3. Nomenclature of eg95-related genes

The nomenclature used to date for genes in the *eg95* gene family from the G1 genotype of *E. granulosus* was created by Chow et al. [21] and includes *eg95-1* (GenBank ID: AF134378), *eg95-2* (AF199351, AF199352), *eg95-3* (AF199353), *eg95-4* (AF199349), *eg95-5* (AF199350), *eg95-6* (AF199347) and *eg95-7* (AF199348). In assigning a nomenclature for *eg95*-related genes in other genotypes of *E. granulosus*, it became necessary to specify the *E. granulosus* genotype of origin as part of the nomenclature. Hence the nomenclature adopted for the genes from the G1 genotype was modified from that previously described [21], so that *eg95-1* is referred to as *eg95-1G1, eg95-2* as *eg95-2G1* and so on for the other gene family members. Nomenclature for the *eg95*-related genes in other genotypes adopts a similar convention, *eg95-xG6* with *x* being a number determined by relative homology to *eg95-1G1* in the sequence alignment; the nomenclature does not imply relatedness between, for example, *eg95-3G1* and *eg95-3G6*. The proteins encoded by, or predicted to be encoded by the genes, have been designated EG95-1G1, *etc*.

2.4. DNA sequence analysis

Sequencing reactions were performed using dideoxy chain termination method using the Applied Biosystems ABI PRISM sequencing system and BIG DYE V3.1 terminator cycle sequencing kit. Sequences obtained were analyzed using the software Accelrys® Gene 2.5 (Accelrys). Sequencing was performed in both strands using vector primers and designing additional primers for primer walking.

2.5. Phylogenetic analysis

Phylogenetic analysis of sequence data was conducted by Bayesian inference (BI) using Monte Carlo Markov Chain (MCMC) analysis in MrBayes 3.1.2 [32]. For the phylogenetic analysis of sequence data we employed the general time-reversible model (GTR), allowing for a 'site-specific rates model', with the 'rates' varying according to a defined partition (*i.e.* codon position) (prset ratepr = variable). Posterior probabilities (pp) were calculated via 1,000,000 generations, utilizing four simultaneous tree-building chains, with every 100th tree being saved. At this point, the standard deviation of split frequencies was <0.01, and the potential scale reduction factor (PSRF) approached one. A 50% majority rule consensus tree for each analysis was constructed based on the final 75% of trees generated by BI.

3. Results

3.1. Southern blot

Genomic DNA identified as belonging to the G6 strain of *E. granulosus* and digested with *Eco*RI showed a hybridization pattern consisting of six bands (designated I to VI, Fig. 1). The bands ranged from 11 to 1.2 kb in size and hybridized with the full length *eg95* cDNA probe. Hybridization of the digested G6 genomic DNA with the 5′ *eg*95 cDNA probe showed a pattern containing bands II, III, V and VI (Fig. 1, panel B). Hybridization with the 3′ probe showed a pattern with bands I and IV as shown in Fig. 1, panel C, suggesting that *eg95*-related genes in the G6 genotype contain an *Eco*RI restriction site and that the *eg95* gene family in G6 contains at least 4 members.

Genomic DNA digested with *Xho*I showed a hybridization pattern consisting of two bands of approximately 15 and 7 kb after hybridization with the full length *eg95* cDNA probe (Fig. 1D). A similar pattern was observed with the 5′ and 3′ *eg95* cDNA probes