**e-sarcoglycan, a Broadly Expressed Homologue of the Gene Mutated in Limb-Girdle Muscular Dystrophy 2D**

(Received for publication, September 11, 1997, and in revised form, October 17, 1997)

Audrey J. Ettinger, Guoping Feng, and Joshua R. Sanes†

From the Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, Missouri 63110

The sarcoglycans are transmembrane components of the dystrophin-glycoprotein complex, which links the cytoskeleton to the extracellular matrix in adult muscle fibers. Mutations in all four known sarcoglycan genes (α, β, γ, and δ) have been found in humans with limb-girdle muscular dystrophy. We have identified a novel protein, e-sarcoglycan, that shares 44% amino acid identity with α-sarcoglycan (adhalin). We show that e-sarcoglycan is a membrane-associated glycoprotein and document its expression by Northern blotting, immunoblotting, and immunofluorescence. In contrast to α-δ sarcoglycans, which are expressed predominantly or exclusively in striated muscle, e-sarcoglycan is broadly distributed in muscle and nonmuscle cells of both embryos and adults. These results raise the possibility that sarcoglycan-containing complexes mediate membrane-matrix interactions in many cell types.

A group of proteins called the dystrophin-glycoprotein complex (DGC) is critical to the stability of muscle fiber membranes (1, 2). Components of the DGC include several cytoplasmic proteins (for example dystrophin and the syntrophins) and two subcomplexes (3) of transmembrane glycoproteins, the dystroglycans (α and β; Ref. 4), and the sarcoglycans (α, β, γ, and δ; Refs. 5–11). The best characterized component of the DGC, dystrophin, is a 427-kDa rod-shaped protein that binds cytoskeletal actin and is associated with the cytoplasmic face of the membrane (12, 13). Loss of dystrophin leads to Duchenne muscular dystrophy, an inevitably fatal wasting of skeletal and cardiac muscle (14). Dystroglycans bind dystrophin intracellularly and laminin extracellularly, thus forming a critical link between the extracellular matrix and the cytoskeleton (15). The sarcoglycans are less well characterized, but three lines of evidence suggest that they may also be involved in membrane-matrix interactions. First, mutations in all four sarcoglycan genes have been found in patients with limb-girdle muscular dystrophy (6–8, 10, 16–21). Second, treatment of cultured muscle cells with antisense oligonucleotides to α-sarcoglycan inhibits their adhesion to substrata (22). Third, the laminin isoform composition of muscle fiber basal lamina is altered in patients lacking α-sarcoglycan (23, 24). Thus, an attractive hypothesis is that the sarcoglycans and dystroglycans mediate related interactions of muscle cells with basal lamina, both of which are crucial for membrane integrity.

Recently, homologues of dystrophin and components of the DGC have been found in numerous nonmuscle tissues. For example, two homologues of dystrophin, utrophin and dystrophin-related protein 2, are expressed at higher levels in some nonmuscle tissues than in muscle (25, 26). Moreover, whereas 427-kDa dystrophin is largely confined to muscle, smaller products of the same gene, generated from alternative promoters and by alternative splicing, are widely distributed (12, 27). Likewise, cytoplasmic proteins associated with the DGC such as dystrobrevin and the syntrophins and transmembrane components such as dystroglycans are present in muscle and nonmuscle tissues alike (28–31). Roles of the DGC in nonmuscle tissues are still largely unexplored, but dystroglycan has been implicated in early embryonic development (32), formation of renal epithelia (33), and adhesion of Schwann cells to basal laminae (34).

In contrast to the broad distribution of dystrophins, syntrophins, dystrobrevin, and dystroglycans, the known sarcoglycan genes are all expressed at higher (β) or far higher (α, γ, and δ) levels in muscle than in nonmuscle tissues (5–11, 16). This distribution suggested that the sarcoglycans mediate muscle-specific functions. However, it also seemed possible that nonmuscle cells expressed homologues of sarcoglycans, as noted above for homologues of dystrophin. Such homologues would be of clinical as well as of cell biological interest; forced up-regulation of utrophin is an attractive approach for ameliorating the symptoms of dystrophin deficiency (35–37), and the same strategy might be applied to patients with sarcoglycan deficiencies. Accordingly, we decided to search for broadly expressed homologues of sarcoglycans. We report here the molecular cloning and characterization of e-sarcoglycan, a homologue of α-sarcoglycan (adhalin).

**EXPERIMENTAL PROCEDURES**

Isolation and Sequencing of cDNAs—The rabbit α-sarcoglycan sequence was used to search the data base of expressed sequence tags (ESTs) using the BLAST program (38) at the National Center for Biotechnology Information. EST 273055 was obtained from Research Genetics (Huntsville, AL) and used to screen an adult mouse lung cDNA library (Stratagene, La Jolla, CA). A clone obtained in this way was used to reprobe the same library. For screening, inserts were labeled with α-[32P]dUTTP using a random primed labeling kit (Boehringer Mannheim). All sequences were determined from both strands using the dideoxy terminator method and an automated sequencer.

Northern Analysis—Total RNA was prepared by the acid guanidinium-phenol procedure (39). Poly(A)+ RNA was isolated by one passage over an oligo(dT)-cellulose column. 10 μg of poly(A)+ RNA was loaded in each lane of an 0.8% agarose gel containing 6.3% formaldehyde and 1 × MOPS buffer (40). The gel was blotted onto a nylon membrane and fixed by UV cross-linking. The membrane was probed overnight with the human EST sequence at 42 °C in 35% formamide buffer. The blot was
then washed with 1× SSC/0.1% SDS at room temperature, 45 °C, and 60 °C.

Generation of Antibodies—An EcoRI fragment of mouse e-sarcoglycan extending from nucleotides 437 to 1235 (see Fig. 1) was subcloned into pET23a (Novagen) and introduced into DE3 bacteria. Cultures were grown until late log stage and then induced with isopropyl thio-galactopyranoside for 1–2 h. Inclusion bodies were isolated by cell lysis and centrifugation, and fusion protein was purified by SDS-polyacrylamide gel electrophoresis. Two rabbits were injected with gel slices containing the fusion protein (Animal Pharm Services, Healdsburg, CA). Antibodies were affinity purified from the resulting antisera by adsorption to and elution from fusion protein immobilized on nitrocellulose (40).

Immunoblotting—Mice were perfused through the heart with saline, and then tissues were dissected and homogenized in ice-cold buffer containing 40 mM Tris, 150 mM NaCl, 10 mM EDTA, 2 mM EGTA, 1 mg/ml soybean trypsin inhibitor, and 0.5 mM Pefabloc (Boehringer Mannheim). The homogenates were centrifuged at 100,000 g for 45 min. In some experiments, aliquots of the pellet were subjected to deglycosylation with the enzyme PNGase F (New England Biolabs) or were extracted with 1% Triton X-100 and then centrifuged again at 16,000 g for 15 min. Subsequently, 20-μg aliquots of protein were boiled in SDS-containing buffer with dithiothreitol, separated by SDS-polyacrylamide gel electrophoresis, and electrophoretically transferred from gels to nitrocellulose filters. The filters were blocked with 5% nonfat dry milk/0.5% Tween 20 in Tris-saline and then incubated overnight with either affinity purified rabbit antibody to e-sarcoglycan or monoclonal antibody NCL-50DAG to α-sarcoglycan (Novocastra, Newcastle-upon-Tyne, UK). Blots were then incubated for 1 h with appropriate alkaline-phosphatase-conjugated secondary antibodies and stained with 5-bromo-4-chloro-3-indoyl phosphate and nitro blue tetrazolium (Sigma-Aldrich).

Immunofluorescence—Tissues from adult rats and mice and whole embryonic day (E) 13.5 mice were frozen in liquid nitrogen-cooled isopentane and sectioned at 8–12 μm in a cryostat. Sections were incubated successively with 1% bovine serum albumin in phosphate-buffered saline; a mixture of affinity purified rabbit anti-e-sarcoglycan and mouse anti-α-sarcoglycan antibodies (see above) in bovine serum albumin/phosphate-buffered saline; and a mixture of fluorescein-conjugated secondary antibodies. The primary and secondary antibodies were detected by use of rhodamine- or fluorescein-conjugated secondary antibodies (Sigma-Aldrich).

Fig. 1—continued

**FIG. 1.** Sequences of α- and e-sarcoglycans. A, predicted domain structures of mouse and human α- and e-sarcoglycans. Identities at the amino acid level are indicated separately for the signal sequences and extracellular, transmembrane, and intracellular domains of the mouse proteins. B, nucleotide and deduced amino acid sequences of mouse e-sarcoglycan and partial amino acid sequence of human e-sarcoglycan. Identities between the human and mouse sequences are indicated by dashes. Predicted signal sequence and transmembrane domain are underlined. Consensus sites for N-glycosylation (N), phosphorylation by casein kinase II and/or by protein kinase C (S/T) and spaced extracellular cysteines (C) are in bold letters. C, amino acid sequence comparison of mouse α- and e-sarcoglycans. Features are indicated as in B. The mouse and human e-sarcoglycan sequences are available from GenBank™ under accession numbers AF031919 and AF031920, respectively. The mouse α-sarcoglycan sequence is from Ref. 11.

---

**sequences of α- and e-sarcoglycans.** A, predicted domain structures of mouse and human α- and e-sarcoglycans. Identities at the amino acid level are indicated separately for the signal sequences and extracellular, transmembrane, and intracellular domains of the mouse proteins. B, nucleotide and deduced amino acid sequences of mouse e-sarcoglycan and partial amino acid sequence of human e-sarcoglycan. Identities between the human and mouse sequences are indicated by dashes. Predicted signal sequence and transmembrane domain are underlined. Consensus sites for N-glycosylation (N), phosphorylation by casein kinase II and/or by protein kinase C (S/T) and spaced extracellular cysteines (C) are in bold letters. C, amino acid sequence comparison of mouse α- and e-sarcoglycans. Features are indicated as in B. The mouse and human e-sarcoglycan sequences are available from GenBank™ under accession numbers AF031919 and AF031920, respectively. The mouse α-sarcoglycan sequence is from Ref. 11.
show that similar amounts of RNA were loaded in all lanes. A, top, Northern blot of RNA from the indicated adult mouse tissues and from an E14 embryo, probed with an \( \epsilon \)-sarcoglycan cDNA. A single 1.8-kilobase species is present in all tissues. A probe for the housekeeping gene EF1\( \alpha \) was used to show that similar amounts of RNA were loaded in all lanes. B, immunoblot of protein from the indicated adult mouse tissues, probed with affinity purified antibodies to recombinant \( \epsilon \)-sarcoglycan. A major 45-kDa species is present in all tissues, and brain also contains proteins of \(-47\) and \(-48\) kDa. C, crude extracts from mouse heart and lung, probed with antibodies to \( \epsilon \)- and \( \alpha \)-sarcoglycan. \( \alpha \)-Sarcoglycan is \(-2\) kDa larger than \( \epsilon \)-sarcoglycan in heart. No \( \alpha \)-sarcoglycan is detectable in lung. \( kb \), kilobases; \( kd \), kilodaltons.

RESULTS AND DISCUSSION

Molecular Cloning of \( \epsilon \)-Sarcoglycan—We searched public data bases for sequences related to sarcoglycans and identified one such human EST. This cDNA was obtained and sequenced, revealing an insert of 730 base pairs containing an open reading frame (ORF) of 145 aa, a putative 3'-untranslated region, and a polyadenylyl tail. The ORF was 51% identical to the carboxy-terminal region of human \( \alpha \)-sarcoglycan (Fig. 1A). We call the predicted novel protein \( \epsilon \)-sarcoglycan.

Northern analysis provided evidence for the existence of a related mRNA in several adult mouse tissues, including lung (see below). We therefore screened a mouse lung cDNA library with the EST to identify murine orthologues of the human gene. Clones were used to reprobe the same library until overlapping clones extending 1524 nucleotides were obtained (Fig. 1B). DNA sequencing revealed an ORF of 405 amino acids, extending from a candidate initiator methionine (41) that corresponds in position to the translation initiation sites in mouse, rabbit, hamster, and human \( \alpha \)-sarcoglycan (Fig. 1C). Comparison of the mouse \( \epsilon \)-sarcoglycan sequence to the partial sequence obtained from the human EST shows that this protein exhibits 95% identity between the species. This homology is similar to that between human and mouse \( \alpha \)-sarcoglycan (88% identity over the full length; Ref. 11).

The mouse \( \epsilon \)-sarcoglycan protein is predicted to begin with a 22-aa signal sequence (42), followed by a 262-aa extracellular domain, a single 23-aa hydrophobic transmembrane domain, and a 98-aa intracellular domain (Fig. 1A). The predicted molecular mass of the mature protein is 43.5 kDa. The extracellular domain bears a single consensus sequence for N-glycosylation. The intracellular domain contains three consensus sites for phosphorylation, one each for casein kinase II, protein kinase C, and both. Mouse \( \alpha \)- and \( \epsilon \)-sarcoglycans are 44% identical (62% similar) at the amino acid level. Homology extends over the whole length of the molecules, with particularly conserved regions both extracellularly (21 of 24 consecutive residues at aa 168–191) and extending through the transmembrane domain (31 of 43 consecutive residues at aa 283–325). The potential site for N-glycosylation in \( \epsilon \)-sarcoglycan corresponds to one of two such sites in \( \alpha \)-sarcoglycan, and the spacing of four extracellular cysteines is perfectly conserved between the homologues (Fig. 1C).

No additional relatives of \( \alpha \)- or \( \epsilon \)-sarcoglycan were found in public data bases. Although \( \alpha \), \( \beta \), \( \gamma \), and \( \delta \) sarcoglycans are all glycosylated transmembrane proteins with a long extracellular domain, a short intracellular domain, and a single transmembrane domain, \( \alpha \)-sarcoglycan shows no significant sequence homology to \( \beta \), \( \gamma \), or \( \delta \) sarcoglycans. It is therefore not unexpected that \( \epsilon \) also exhibits no homology to \( \beta \), \( \gamma \), or \( \delta \) sarcoglycans.

Expression of \( \epsilon \)-Sarcoglycan—As a first step in examining the expression of \( \epsilon \)-sarcoglycan, we performed Northern analysis. A single mRNA of \(-1.8\) kilobases was identified in all tissues examined (Fig. 2A). Levels were highest in lung, moderate in brain, heart, and skeletal muscle, and low but detect-
able in kidney and liver. We also detected $\epsilon$-sarcoglycan mRNA in E14 embryos (Fig. 2A). The sizes of $\alpha$- and $\epsilon$-sarcoglycan mRNAs are similar, but their distributions differ markedly: $\alpha$ is largely restricted to striated (cardiac and skeletal) muscle, with very low levels of expression in lung and none in brain (5, 11, 16).

To obtain reagents for studying the $\epsilon$-sarcoglycan protein, a recombinant extracellular fragment was used to generate rabbit antisera, and antibodies were purified using immobilized immunogen. Each of the two independently generated antisera recognized a major band of 45 kDa on Western blots (Fig. 2B and data not shown). A single major immunoreactive species was observed in heart, skeletal muscle, lung, liver, and kidney. In brain, two higher molecular masses bands, of $\sim$47 and 48 kDa were also visible. The distribution of the protein paralleled that of $\epsilon$-sarcoglycan mRNA, being highest in lung and lowest in kidney and liver (Fig. 2A and B).

The homology between $\alpha$- and $\epsilon$-sarcoglycans raised the possibility that antibodies to either protein might recognize both. To assess this possibility, we probed extracts of heart (which contains both $\alpha$ and $\epsilon$) and lung (which contains only $\epsilon$) with our antibodies to $\epsilon$-sarcoglycan and a commercially available monoclonal antibody to $\alpha$-sarcoglycan. In heart, anti-$\alpha$-sarcoglycan stained a band $\sim$2 kDa larger than that recognized by anti-$\epsilon$-sarcoglycan (Fig. 2C, four left-hand lanes). In lung, anti-$\alpha$-sarcoglycan detected no protein (Fig. 2C, two right-hand lanes). Thus, the antibodies to recombinant $\epsilon$-sarcoglycan and a widely used monoclonal antibody to $\alpha$-sarcoglycan are not detectably cross-reactive.

**Characterization of $\epsilon$-Sarcoglycan Protein—**$\alpha$-Sarcoglycan is membrane-associated and glycosylated. We asked whether the same was true for $\epsilon$-sarcoglycan. $\epsilon$-Sarcoglycan from heart (Fig. 3A) and skeletal muscle (data not shown) were insoluble in saline but were solubilized by treatment with Triton X-100 detergent. Thus, $\epsilon$-sarcoglycan is membrane-associated.

To assess glycosylation, tissue homogenates from lung, brain, kidney, and heart were treated with endoglycosidase F and then analyzed by electrophoresis and immunoblotting. Enzymatic digestion decreased the apparent molecular mass of $\epsilon$-sarcoglycan by $\sim$1–2 kDa (Fig. 3B and data not shown). Similar treatment of $\alpha$-sarcoglycan has been reported to shift its apparent molecular mass by $\sim$4 kDa (43). This difference is consistent with the presence of two sites for $N$-linked glycosylation on $\alpha$-sarcoglycan but only one on $\epsilon$-sarcoglycan. Differential glycosylation also provides an explanation for the initially paradoxical observation that the apparent molecular mass of $\alpha$-sarcoglycan is $\sim$2 kDa larger than that of $\epsilon$-sarcoglycan, whereas the predicted molecular mass of $\epsilon$-sarcoglycan is $\sim$2 kDa larger than that of $\alpha$-sarcoglycan.

**Cellular Localization of $\epsilon$-Sarcoglycan—**We used a double label immunofluorescence technique to compare the distributions of $\alpha$- and $\epsilon$-sarcoglycan in skeletal muscle. As previously reported (44, 45), $\alpha$-sarcoglycan was associated with skeletal muscle fibers but was absent from other intramuscular cells (Fig. 4A). In contrast, $\epsilon$-sarcoglycan was present at higher levels in blood vessels and nerves than in muscle fibers (Fig. 4A). In intramuscular nerves, $\epsilon$-sarcoglycan was present both in the endoneurium (Schwann cells that ensheathe axons) and in the perineurium (fibroblast-derived cells that ensheathe bundles of axon-Schwann cell units; Fig. 4C). Examination at higher power showed that much of the sarcolemma-associated $\epsilon$-like immunoreactivity was in fact associated with capillaries, whereas $\alpha$-like immunoreactivity was confined to the sarcolemma (Fig. 4, B and B'). There was, however, also clear $\epsilon$
immunoreactivity associated with muscle fiber surfaces. Consistent with this result, we have detected e-sarcoglycan in cultures of a clonal muscle cell line by immunoblotting (data not shown).
e-sarcoglycan immunoreactivity was also widely distributed in nonmuscle tissues. In lung, the protein was associated with both alveoli and bronchioles (Fig. 4D). In kidney, immunoreactivity was particularly intense in glomerular mesangium (Fig. 4E). Sections of brain, heart, thymus, and intestine were also examined, and in each case large subsets of membranes were e-sarcoglycan-positive (data not shown). Of these tissues, α-sarcoglycan was detectable only in heart by the immunofluorescence method we used. Thus, consistent with results from northern and immunoblotting analyses, e-sarcoglycan is far more broadly distributed than is α-sarcoglycan.

To assess the distribution of e-sarcoglycan during development, sections of E13.5 embryos were double-labeled with antibodies to α- and e-sarcoglycan. No α-sarcoglycan was detectable in embryos at this stage (data not shown), consistent with previous reports that this protein appears at a relatively late stage of myogenesis (11, 44, 45). In contrast, e-sarcoglycan was present in blood vessels throughout the body and in heart, intestine, meninges of spinal cord and brain (Fig. 4F), and developing kidney (Fig. 4G). Thus, in embryos, e-sarcoglycan is expressed early and more widely than α-sarcoglycan.

Conclusions—Previously described sarcoglycans, which underlie limb-girdle muscular dystrophy types 2C–F, are predominantly or exclusively expressed in striated muscle. In contrast, e-sarcoglycan is widely expressed in both muscle and nonmuscle cells, and in embryos as well as adults. These results provide the first suggestion that sarcoglycans, like other components of the DGC (15, 32–34), may be important for embryonic development and/or for integrity of nonmuscle tissues. Critical questions, which we are now addressing, are whether e-sarcoglycan is physically associated with other DGC components; whether homologues of β, γ, or δ sarcoglycan exist; and whether levels of α-sarcoglycan are altered in any muscular dystrophies. Currently, there is considerable interest in the possibility that up-regulation of utrophin, the broadly distributed homologue of dystrophin, is a better therapeutic approach to Duchenne muscular dystrophy than replacement of the mutated homologue of dystrophin, is a better therapeutic approach to Duchenne muscular dystrophy than replacement of the mutated homologue of dystrophin, and/or for integrity of nonmuscle tissues. This context, it is interesting that cardiac involvement is common in dystrophinopathies (Duchenne and Becker dystrophy) but uncommon in the sarcoglycanopathies (limb-girdle muscular dystrophy 2C–F; Ref. 46). One intriguing possibility is that the levels of e-sarcoglycan in cardiac muscle, which are higher than those in skeletal muscle, ameliorate the effects of sarcoglycan mutations on the heart. If so, further up-regulation might confer a similar benefit to skeletal muscle.

Acknowledgments—We thank Jeannette Cunningham, Mark Grady, Renate Lewis, Jeff Miner, Bruce Patton, and Ellen Ryan for advice and assistance.