Isolation and characterization of different C-terminal fragments of dystrophin expressed in *Escherichia coli*

Rachel E. MILNER,* Jody BUSAAN* and Marek MICHALAK*†

* Cardiovascular Disease Research Group, Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2S2

Dystrophin, the protein product of the Duchenne muscular dystrophy gene, is thought to belong to a family of membrane cytoskeletal proteins. Based on its deduced amino-acid sequence, it is postulated to have several distinct structural domains; an *N*-terminal region; a central, rod-shaped, domain; and a *C*-terminal domain [Koenig, Monaco & Kunkel (1988) *Cell* 53, 219–228]. The *C*-terminal domain is further divided into two regions; the first has some sequence similarity to slime mould α -actinin, and is rich in cysteine residues; this is followed by the *C*-terminal amino-acid sequence that is unique to dystrophin. Dystrophin is very difficult to purify in quantities sufficient for detailed studies of the structure/function relationships within the molecule. Therefore, in this study, we have expressed selected fragments of the *C*-terminal region of dystrophin, as fusion proteins, in *Escherichia coli*. Importantly, we describe the first successful purification, from *E. coli* lysates, of large quantities of fragments of dystrophin in a soluble form. The first fragment, termed CT-1, encodes the *C*-terminal 201 amino acids of the protein; the second, termed CT-2, spans the cysteine-rich region of the *C*-terminal domain. These fusion proteins were identified by their mobility in SDS/PAGE, by their interaction with appropriate affinity columns and by their reactivity with anti-dystrophin antibodies. The fragment CT-2, which spans a region containing putative EF-hand-like sequences, was found to bind Ca^{2+} in $^{45}\text{Ca}^{2+}$ overlay experiments. In addition, we have discovered that the fragment CT-1, but not fragment CT-2, interacts specifically with the *E. coli* *DnaK* gene product [analogue of heat shock protein 70 (hsp70)]. This interaction is disrupted, *in vitro*, by the addition of ATP. Our results indicate that the two *C*-terminal fragments of dystrophin have differing biophysical properties, indicating that they may play distinct roles in the function of the protein.

INTRODUCTION

Duchenne muscular dystrophy (DMD) is a fatal, X-linked, degenerative disorder of muscle, that is characterized by early onset of proximal muscle weakness followed by severe muscular atrophy [1]. Respiratory or cardiac failure, and death, usually result by the end of the third decade. The biochemical mechanism(s) behind the pathology of DMD is not yet elucidated. In recent years, however, significant advances in the understanding of the disease have been made with the application of molecular genetic approaches [2]. In particular, the complete cDNA of the DMD gene has been isolated and dystrophin, the protein product of the gene, has been identified [3,4]. Importantly, it has also been clearly demonstrated that dystrophin is absent in muscle from DMD patients [4–8].

Historically, surface membranes were implicated in having a role in the pathology of DMD (for review see ref. [9]). This suggestion has been supported by the recent observation that dystrophin is localized to the cytoplasmic face of the muscle plasma membrane [4–11]. The amino-acid sequence of dystrophin, which was deduced from cDNA to the DMD gene [12], suggests that dystrophin shares some structural similarity with a family of membrane cytoskeletal proteins. For example, the *N*-terminal region of dystrophin shares a significant degree of sequence similarity with the actin-binding region of α -actinin. Further, on the basis of sequence similarities with the triple-helical repeat domain identified in spectrin [13], the central portion of the dystrophin molecule is postulated to be a long, rod-shaped domain formed from repeating triple-helical seg-

ments. The *C*-terminal region of skeletal muscle dystrophin contains a region (amino acid residues 3115–3269) with significant sequence similarity to the *C*-terminal region of the slime mould α -actinin [12]. This is followed by 420 amino acid residues which, when originally sequenced, showed no identity to any other previously identified proteins [12]. Interestingly, a 'dystrophin-related protein' that is encoded on chromosome 6 has subsequently been identified, the *C*-terminal region of which has significant sequence similarity with the *C*-terminal region of dystrophin [14,15]. The structural organization of dystrophin described above, that was postulated on the basis of its deduced amino-acid sequence [12], has been supported by protease mapping studies [16] and by electron microscopy [17,18]. This structural evidence has led to the suggestion that dystrophin may play a role in the subsarcolemmal cytoskeleton, for example in maintenance of the flexibility and integrity of the plasma membrane during muscle contraction [16].

In a series of elegant biochemical studies Campbell's group has further supported the hypothesis that dystrophin may have a function in the membrane cytoskeleton. First, the protein has been shown to be a major component of the subsarcolemmal cytoskeleton [19], and secondly, it has been shown to be associated with a complex of integral membrane glycoproteins [20,21] that is specifically enriched in the sarcolemma [22]. The association of dystrophin with sarcolemmal glycoproteins has also been demonstrated in cross-linking experiments [23]. As yet there is no evidence that directly identifies the region of dystrophin responsible for the dystrophin-membrane interactions. Ervasti & Campbell [21] have proposed that the *C*-terminal region of

Abbreviations used: GST, glutathione *S*-transferase; EGTA, ethylene glycol bis(β -aminoethyl ether)-*NNN'*-tetra-acetic acid; hsp, heat shock protein; PBS, phosphate-buffered saline; DMD, Duchenne muscular dystrophy; IPTG, isopropyl β -D-thiogalactopyranoside; LB, Luria broth; KLH, keyhole limpet haemocyanin.

† To whom correspondence should be addressed.

dystrophin may be involved in its interaction with the sarcolemmal glycoproteins. Functionally, this region of dystrophin must be particularly interesting, since it undergoes tissue-specific alternative splicing [24] and also has been shown recently to be highly conserved between different species [25].

The present study was designed to produce two specific fragments of the C-terminal region of skeletal muscle dystrophin. One fragment, designated CT-2, corresponds to the cysteine-rich region; the other fragment, designated CT-1, corresponds to the last 201 amino-acid residues of dystrophin. Utilizing two different plasmid vectors we have expressed each fragment as a fusion protein with both the *Schistosoma japonicum* enzyme glutathione-S-transferase (GST) [26] and with staphylococcal protein A [27]. We report here purification of both the CT-1 and CT-2 fragments of the C-terminal region of skeletal muscle dystrophin in large quantities. We showed that the two fragments have differing biophysical properties, and that the fragment encoding the C-terminal 201 amino-acid residues of dystrophin (CT-1) interacts with the *Escherichia coli* DnaK chaperone, a protein that has strong similarity to the mammalian heat shock protein (hsp)70 proteins (for review see ref. [28]).

EXPERIMENTAL

Materials

Triton X-100, Tween 20 and ampicillin were purchased from Sigma. $^{45}\text{CaCl}_2$ was obtained from New England Nuclear.

Fig. 1. Diagrammatic representation of the fragments of dystrophin expressed in *E. coli*

The model of the dystrophin molecule is modified from Koenig *et al.* [12]. The C-terminal region of dystrophin was divided into two specific fragments, termed CT-1 and CT-2. These fragments were amplified by PCR and subcloned into unique restriction sites in the pGEX-3X and pRIT2T plasmids, as described in the Experimental section. The locations of the PCR primers a, b, c and d are shown (boxed). The amino-acid sequence of the synthetic peptide used for the production of antibodies is shown (underlined). The residue numbers in (a) correspond to the residue numbers in Koenig *et al.* [12].

Peroxidase-conjugated goat anti-(rabbit IgG) antibody was from Boehringer-Mannheim. Nitrocellulose and Immobilon (polyvinylidene difluoride) membrane filters were from Schleicher and Schuell and Millipore respectively. SDS/PAGE reagents and molecular-mass markers were from Bio-Rad. Restriction endonuclease and DNA-modifying enzymes were obtained from Boehringer-Mannheim, Bethesda Research Laboratories and Bio/Can Scientific. Isopropyl β -D-thiogalactopyranoside (IPTG) was from Boehringer-Mannheim. The plasmids pGEX-3X and pRIT2T, *E. coli* N4830-1 cells, the Mono Q 5/5 and 10/10 f.p.l.c. columns, glutathione-Sepharose 4B and IgG-Sepharose 6 FF were all from Pharmacia. Competent cells of *E. coli* DH5 α were obtained from Bethesda Research Laboratories. *E. coli* BNN103 were a generous gift of Dr. J. H. Weiner (University of Alberta, Alberta, Canada). All other chemicals used were of the highest grade commercially available.

Construction of plasmids encoding different fusion proteins

For the expression and isolation of recombinant proteins two systems were employed. First, a GST fusion system was used, which consists of the plasmid pGEX-3X [26]. This plasmid encodes GST followed by a recognition site for factor Xa cleavage, a series of unique cloning sites (*Eco*R1, *Sma*I and *Bam*H1) and a stop codon. We also used the protein A gene fusion vector pRIT2T which encodes a truncated form of the staphylococcal protein A gene immediately upstream of the multiple cloning site. In this study we expressed two distinct fragments of the C-terminal region of skeletal muscle dystrophin: CT-1, which encodes amino-acid residues 3485–3685; and CT-2, which encodes amino-acid residues 3107–3400 (Fig. 1).

We used the PCR to synthesize cDNA encoding the fragments CT-1 and CT-2. As a template we used a cDNA clone spanning the region 7800–13900 bp from the 5' end of dystrophin cDNA. This clone is called cDMD 9-14 (American Type Culture Collection, depositor Dr. L. M. Kunkel), and its nucleotide sequence is described in Koenig *et al.* [3]. The following oligodeoxynucleotides with 5' flanking *Eco*R1 restriction sites were synthesized and used as primers:

- primer a 5'-TGGAATTCGCTGAGCCAGCCTCGTAGTCCT-3';
- primer b 5'-TGGAATTCACCTAAGGACTCCATCGCTCTGC-3';
- primer c 5'-AGGAATTCCAGGACTGCCATGAAATCCCA-3';
- primer d 5'-AGGAATTCGGATGACAGTCTGCACTGGCAGGTA-3'.

The nucleotide sequence of primer a corresponds to the nucleotides 10661–10681 of dystrophin cDNA, and encodes amino-acid residues 3485–3491 of the protein [3]. This primer was used as a 5' (forward) primer for the synthesis of CT-1 (Fig. 1). Primer b, which was used as the 3' (reverse) primer for the synthesis of CT-1, corresponds to nucleotides 11294–11317 which are within the 3' non-coding region of the dystrophin cDNA (Fig. 1). Primer c encodes the nucleotide sequence 9527–9547 (amino-acid residues 3107–3113) and primer d nucleotides 10388–10408 (amino-acid residues 3394–3400) [3]. Primers c and d were used as the 5' (forward) and 3' (reverse) primers respectively, for the synthesis of CT-2 (Fig. 1). PCRs were carried out in a buffer containing 50 mM-KCl, 10 mM-Tris, pH 8.3, 0.01% gelatin, 200 μ M of each dNTP, 1.5 mM-MgCl $_2$, 2.5 units of Taq I polymerase, 50 ng of the DNA template and 200 ng of each of the appropriate primers. PCR products were purified by PAGE, cut with *Eco*R1 restriction endonuclease and ligated into the

*Eco*R1 restriction site of a phosphatase-treated vector, either pGEX-3X or pRIT2T. Plasmids were then used for transformation of two strains of *E. coli*; first DH5 α and then either BNN103 (GST fusion proteins) or N4830-1 (protein A fusion proteins) for the expression experiments.

Expression and isolation of recombinant proteins

GST fusion proteins were expressed in BNN103 *E. coli* host cells that were grown in Luria broth (LB) medium containing ampicillin (50 μ g/ml). Cultures were grown to mid-log phase (A_{600} 0.6–1.0) and then expression of the fusion proteins was induced with IPTG (0.1 mM). Cells were incubated for a further 4 h, and were then centrifuged at 3500 *g* for 15 min. The pellet was resuspended in phosphate-buffered saline (PBS) containing 0.1 % Triton X-100 and the cells were lysed using a French press set at 6.9 MPa. The lysates were generally centrifuged at 10000 *g* for 10 min or, where specifically stated, at 100000 *g* for 45 min. The pellet and supernatant fractions were separated, frozen and stored at -70°C until required for further use.

Protein A fusion proteins were expressed in *E. coli* N4830-1, a strain that contains the temperature-sensitive λ cI₈₅₇ repressor. Cells were grown to early stationary phase (A_{600} 0.8–2.0) in LB medium containing ampicillin (50 μ g/ml), at 30°C . The temperature of the medium was rapidly shifted to 42°C and the cells then incubated for 90 min at that temperature. Cells were harvested by centrifugation at 5000 *g* for 5 min, resuspended in 50 mM-Tris, pH 7.6, 150 mM-NaCl, 0.05 % Tween 20 and lysed using a French press, as described above. The lysate was then centrifuged at 20000 *g* for 20 min or, where specified, at 100000 *g* for 45 min.

Purification of GST and protein A fusion proteins

GST and GST-fusion proteins were purified by one-step glutathione–Sephacrose 4B affinity chromatography (column size: 2 cm $\times$ 12 cm; 40 ml bed volume). Samples of *E. coli* extracts (2 mg/ml) were applied to a glutathione–Sephacrose 4B column that had been equilibrated with PBS containing 1 % (v/v) Triton X-100. The first flow-through was reapplied on to the column, which was then washed with 400 ml of PBS. Fusion proteins were eluted from the affinity column with a buffer containing 5 mM-glutathione and 50 mM-Tris/HCl, pH 8.0. Protein A fusion proteins were purified from the supernatant using IgG–Sephacrose 6 FF affinity chromatography (column size: 2 cm $\times$ 15 cm; approx. 50 ml bed volume). The column was equilibrated with 50 mM-Tris, pH 7.6, 150 mM-NaCl, 0.05 % Tween 20 before use. Samples of *E. coli* extracts were applied and the column was then washed with 10 bed volumes of the equilibration buffer, followed by two bed volumes of 5 mM-ammonium acetate, pH 5.0. The sample was eluted with a solution of 0.5 M-acetic acid, pH 3.4. In both cases, over 90 % of the fusion protein located in the soluble fraction of the *E. coli* lysate could be purified to near homogeneity using these one-step column chromatography procedures.

Fractions containing fusion protein were concentrated using an Amicon concentrator and were then dialysed against 50 mM-Tris/HCl, pH 7.1, containing 1 mM-EDTA. Dialysed samples were then subject to Mono Q f.p.l.c., as described previously [29]. Essentially, the samples were loaded on to a Mono Q column that had been equilibrated with the dialysis buffer. The column was washed with 50 mM-Tris/HCl, pH 7.1, containing 1 mM-EDTA and then bound proteins were eluted with a linear salt gradient (0–750 mM-NaCl). The gradient was applied at a flow rate of 1 ml/min, over 30 min.

The plasmid pGEX-3X used in this study encodes GST followed by the factor Xa cleavage site Ile-Gly-Glu-Arg [26]. In

some experiments the purified GST–CT-1 fusion protein was digested with factor Xa to remove GST [29a]. Although the factor Xa recognition site is not found in the C-terminal region of dystrophin, incubation of GST–CT-1 fusion protein with factor Xa led to a complete degradation of the CT-1 (results not shown). Therefore, the GST fusion protein was used in this study.

Production of antibodies

Antibodies were raised against a synthetic peptide encoding part of the C-terminal region of dystrophin (Fig. 1 underlined segment) and also against the protein A–CT-1 fusion protein. The peptide NH₂-Pro-Ser-Ser-Arg-Gly-Asn-Thr-Pro-Gly-Lys-Pro-COOH, which encodes amino-acid residues 3668–3679 (Fig. 1) of skeletal muscle dystrophin, was chemically synthesized on a peptide synthesizer (Model 430A; Applied Biosystems, Foster City, CA, U.S.A.) and then coupled to keyhole limpet haemocyanin (KLH) by workers at the Alberta Peptide Institute, University of Alberta. New Zealand rabbits were immunized by subcutaneous injection of 0.5 mg of the KLH-coupled synthetic peptide emulsified in Freund's complete adjuvant. After 2 weeks the immunization was repeated with 0.5 mg of the peptide emulsified in Freund's incomplete adjuvant. Sera were screened 2–3 days later against proteins that had been electrophoretically transferred to nitrocellulose membranes, as described by Towbin *et al.* [30]. Antibodies were also raised against the purified protein A–CT-1 fusion protein of dystrophin (Fig. 1). The protein A–CT-1 fusion protein was purified by IgG affinity chromatography followed by Mono Q f.p.l.c. The pure fusion protein was then injected, and antibodies raised, as described above.

SDS/PAGE immunoblotting and analysis of $^{45}\text{Ca}^{2+}$ binding

SDS/PAGE was carried out on 7 %, 10 % or 12.5 % polyacrylamide gels, as described by Laemmli [31]. After electrophoresis, the proteins were either stained with Coomassie Blue or transferred electrophoretically on to nitrocellulose membranes [30]. Immunoblotting was carried out as described previously [32] and antibody binding was detected using appropriate peroxidase-conjugated second antibodies and a standard peroxidase colour-development reaction. For analysis of $^{45}\text{Ca}^{2+}$ binding the membranes were incubated with $^{45}\text{Ca}^{2+}$ as described by Maruyama *et al.* [32a] and exposed to Kodak X-ray film. Standards were Bio-Rad high- and low-range molecular mass marker proteins; myosin (200000), β -galactosidase (116250), phosphorylase *b* (97400), BSA (66200), ovalbumin (42700), bovine carbonic anhydrase (31000), soybean trypsin inhibitor (21500) and lysozyme (14400). Bio-Rad high- and low-range prestained molecular mass markers were also used; myosin (205000), β -galactosidase (116500), phosphorylase *b* (106000), BSA (80000), ovalbumin (49000), carbonic anhydrase (32000), soybean trypsin inhibitor (27000) and lysozyme (17000).

Miscellaneous

All recombinant techniques were carried out according to standard protocols [33]. Sarcolemmal vesicles were isolated by sucrose floatation according to the method described by Jones [34] in the presence of protease inhibitors [32]. Protein was determined by the method of Lowry *et al.* [35] or Bradford [36]. N-terminal amino-acid-sequence analysis was carried out after proteins had been electroblotted to Immobilon (polyvinylidene difluoride) membranes [37]. The automated analyses were performed with an Applied Biosystems Model 470A gas-liquid-phase protein sequencer connected on-line to an Applied Biosystems Model 120A h.p.l.c., using current protocols of Applied Biosystems. All sequencer chemicals were from Applied Biosystems, Foster City, CA, U.S.A.

RESULTS

Expression of the recombinant C-terminal (CT) region of dystrophin

Fig. 1(a) shows the amino-acid sequence of the C-terminal region of dystrophin (a) and includes a diagrammatic representation of the fragments of this region that have been expressed, in this study, as fusion proteins (Fig. 1b). Fig. 1(b)

Fig. 2. Immunological identification of CT-1 and CT-2 fusion proteins in lysates of *E. coli*

Cardiac sarcolemmal proteins and *E. coli* lysates were separated by SDS/PAGE, electrophoretically transferred to nitrocellulose membranes, and reacted with anti-dystrophin antibodies as described in the Experimental section. (a) Cardiac sarcolemmal proteins screened with anti-(dystrophin synthetic peptide) antibodies. (b) Lane 1, *E. coli* lysate, containing GST-CT-1 fusion protein, screened with anti-dystrophin serum; lane 2, *E. coli* lysate, containing recombinant GST, screened with anti-dystrophin antibodies. (c) Lane 1, *E. coli* lysate, containing protein A-CT-1 fusion protein, screened with anti-dystrophin antibodies; lane 2, *E. coli* lysate, containing protein A-CT-1 fusion protein, screened with peroxidase-linked goat IgG; lane 3, *E. coli* lysate, containing recombinant protein A, screened with peroxidase-linked goat IgG. (d) Different fractions of an *E. coli* lysate containing the protein A-CT-2 fusion protein were immunoblotted with peroxidase-linked goat IgG. Lane 1, 20000 *g* pellet; lane 2, 20000 *g* supernatant; lane 3, 100000 *g* supernatant. The positions of molecular-mass marker proteins are indicated.

includes a schematic representation of the different structural domains of dystrophin, as proposed by Koenig *et al.* [12]. We have designed, expressed and purified two C-terminal fragments, termed CT-2 and CT-1. The first fragment (CT-1) spans the C-terminal 201 amino-acid residues of skeletal muscle dystrophin and the second fragment (CT-2) spans the cysteine-rich region, encompassing the 150 amino-acid residues with significant sequence similarity to slime mould α -actinin. These fragments of dystrophin were expressed in *E. coli*, purified and characterized. The precise sequences of the primers used in the PCR are given in the Experimental section, and the amino-acid sequences of CT-1 and CT-2, are given in Fig. 1.

Upon lysis of the cells using a French press the recombinant GST-CT-1 fusion protein (approx. 49 kDa) was found to be a major component of the *E. coli* soluble extract (results not shown). The protein A-CT-1 fusion protein (approx. 50 kDa) was also highly soluble (results not shown). The solubility of both of the CT-1 dystrophin fusion proteins suggests that they might have been correctly folded during synthesis in *E. coli*.

The CT-1 fusion proteins were identified by Western blotting, using an antiserum raised against a synthetic peptide that corresponds to the dystrophin amino-acid sequence Pro-Ser-Ser-Arg-Gly-Arg-Asn-Thr-Pro-Lys-Pro (residues 3668–3679). These anti-(synthetic peptide) antibodies detect dystrophin in isolated cardiac sarcolemma membranes (Fig. 2a). The same antibodies recognized the GST-CT-1 fusion protein (Fig. 2b, lane 1) and the protein A-CT-1 fusion protein (Fig. 2c, lane 1) in immunoblots of soluble extracts from cells expressing these proteins. Since the fragment CT-1 spans amino-acid residues 3485–3685, its cross-reactivity with antibodies against the synthetic peptide (residues 3668–3679) confirms its correct transcription. Purified recombinant GST does not cross-react with the anti-dystrophin antibodies (Fig. 2b, lane 2). The protein A-CT-1 fusion protein could also be detected in the supernatants by immunoblotting with a peroxidase-linked goat (anti-rabbit IgG) antibody. The goat IgG binds to the protein A portion of the protein A-CT-1 fusion protein (Fig. 2c, lane 2) and also to purified recombinant protein A (Fig. 2c, lane 3). These observations confirm the identity of the fusion proteins used in this study.

Another protein fragment was expressed in this study, and was termed CT-2. The amino-acid residues encoded in this fragment span the majority of the cysteine-rich region and include the sequence (approx. 150 amino acids) with reported homology to the slime mould α -actinin [12] (see Fig. 1). CT-2, like CT-1, was expressed as both GST and protein A fusion proteins. In comparison with the marked solubility of the CT-1 fusion proteins, we found that the CT-2 fusion proteins were much less soluble. In both cases a major protein band of molecular mass 61/62 kDa was apparent in the pellet fraction of the lysate, but not in the supernatant fraction (results not shown). Insolubility is a documented problem for a wide variety of proteins expressed in *E. coli*, many of which are frequently only recovered in insoluble inclusion bodies [38, 39]. In this case, however, a portion of the GST-CT-2 and protein A-CT-2 fusion proteins was present in the soluble fraction since we were able to purify quantities of both from 100000 *g* supernatants of the *E. coli* extracts, using affinity chromatography (see below). In addition, on immunoblots performed with a peroxidase-linked goat IgG, the protein A-CT-2 fusion protein could be detected in the supernatant fraction of extracts centrifuged at 20000 *g* (Fig. 2d, lane 2) and at 100000 *g* (Fig. 2d, lane 3). The relative solubility of the two different fragments of the C-terminal region of dystrophin, CT-1 and CT-2, was estimated assuming equivalent levels of expression of the fusion proteins (this assumption is supported by SDS/PAGE analysis). The amount of GST-CT-1 obtained after glutathione-Sepharose affinity chromatography

Fig. 3. Purification of dystrophin fusion proteins

Fusion proteins were purified from *E. coli* lysates by affinity chromatography, as described in the Experimental section. Proteins eluted from the affinity columns were analysed by SDS/PAGE and Coomassie Blue staining. Lane 1, GST-CT-1 fusion protein; lane 2, GST-CT-2 fusion protein; lane 3, protein A-CT-1 fusion protein; lane 4, protein A-CT-2 fusion protein. A 70 kDa polypeptide which co-elutes with the CT-1 fusion proteins is indicated by the arrow. The positions of molecular-mass marker proteins are shown.

Fig. 4. Ca^{2+} -binding to GST-CT-2 fusion proteins

Affinity-purified GST-CT-2 fusion protein and recombinant GST (6 μg) were subjected to SDS/PAGE, transferred to nitrocellulose filters, and incubated with $^{45}\text{Ca}^{2+}$ as described in the Experimental section. $^{45}\text{Ca}^{2+}$ binding was determined by autoradiography: lane 1, recombinant GST; lane 2, GST-CT-2 fusion protein. The positions of Bio-Rad prestained molecular-mass marker proteins is indicated.

was approx. 15 % of the total protein in the initial extract. This compares with a yield of approx. 3 % for GST-CT-2.

Purification and characterization of CT-1 and CT-2, the recombinant C-terminal fragments of dystrophin

The GST and the protein A fusion proteins were purified from *E. coli* extracts by one-step affinity chromatography on glutathione-Sepharose 4B and IgG-Sepharose 6 FF columns respectively. Fig. 3 shows SDS/PAGE of proteins eluted from the columns. We noted that at this stage of purification both of the GST fusion proteins co-purified with some recombinant GST (Fig. 3, lanes 1 and 2). In addition, the GST-CT-1 and protein A-CT-1 fusion proteins co-purified with a number of lower molecular mass contaminants (Fig. 3, lanes 1 and 3). These cross-reacted with anti-dystrophin antibodies (results not shown)

indicating that they may be degradation products of the fusion proteins. We also found that significant quantities of a polypeptide of approx. 70 kDa consistently co-eluted with both the GST-CT-1 and the protein A-CT-1 fusion proteins (Fig. 3, lanes 1 and 3). This band was not observed during the affinity purification of CT-2 fusion proteins (Fig. 3, lanes 2 and 4).

Since the portion of the C-terminus of dystrophin encoded in fragment CT-2 shares some sequence similarity with the C-terminus of slime mould α -actinin, a region containing putative EF-hand-like sequences [12], we investigated the potential Ca^{2+} -binding behaviour of the fusion proteins using a $^{45}\text{Ca}^{2+}$ overlay technique (Fig. 4). We have established recently that recombinant GST does not bind significant amounts of $^{45}\text{Ca}^{2+}$, either during equilibrium dialysis or in a $^{45}\text{Ca}^{2+}$ overlay [29a]; the GST fusion proteins were used in these experiments. Fig. 4 (lane 2) shows that the GST-CT-2 fusion protein bound $^{45}\text{Ca}^{2+}$ in the overlay. Densitometric scanning of the autoradiographs revealed that the binding of Ca^{2+} to the GST-CT-2 fusion protein was 7–10-fold greater than that to the GST alone (Fig. 4, lane 1).

The recombinant CT-1 fragment of dystrophin is associated with the *E. coli* DnaK gene product

To purify the CT-1 and CT-2 fusion proteins further they were subjected to f.p.l.c. on a Mono Q column. During the course of this work we found that the CT-1 fragment of dystrophin co-purified with the *E. coli* DnaK gene product, a member of the hsp70 family of proteins (see below). Fig. 5 illustrates the resolution of the GST-CT-1 fusion protein on the Mono Q column and shows the results of SDS/PAGE analysis of proteins in the eluted fractions. Two protein peaks were observed (Fig. 5a). The first protein peak (fractions 7–9), at approx. 200 mM-NaCl, consisted of purified GST-CT-1 (Fig. 5b, lanes 2–4). The lower molecular mass polypeptides apparent in small quantities in these fractions are probably degradation products of the GST-CT-1 fusion protein, since they are all detected in Western blots with our anti-dystrophin antiserum (results not shown). The second peak (fractions 18–20), at approx. 450 mM-NaCl, consisted of a portion of the GST-CT-1 fusion protein that co-eluted with the 70 kDa polypeptide (Fig. 5b, lanes 6–8). The intensity of Coomassie Blue staining for the two different polypeptides suggested that they were present in approx. equal quantities. The identity of the GST-CT-1 fusion protein in these fractions was confirmed by immunoblotting (Fig. 5c).

The protein A-CT-1 fusion protein was subjected to Mono Q f.p.l.c. under conditions identical to those used for the GST-CT-1 fusion protein: a similar elution profile, with two peaks, was obtained (results not shown). The first peak, which eluted from the column at approx. 200 mM-NaCl, contained the protein A-CT-1 fusion protein. The second peak, which eluted at approx. 350 mM-NaCl, contained both the fusion protein and the approx. 70 kDa polypeptide.

The 70 kDa protein which co-eluted with both GST-CT-1 and protein A-CT-1 fusion proteins was subjected to N-terminal amino-acid-sequence analysis after transfer to Immobilon membranes [37]. The following sequence was obtained: NH_2 -Gly-Lys-Ile-Ile-Gly-Ile-Asp-Leu-Gly-Thr-Thr-Asn-Ser-Xaa-Val-Ala-Ile-Met. This sequence is identical to the N-terminal sequence of the *E. coli* DnaK gene product [40] that is known to be the equivalent of the mammalian hsp70 [28,41]. Our results indicate that the fragment, CT-1, of the C-terminal region of dystrophin can be tightly associated with the bacterial DnaK protein.

Characterization of the interaction between recombinant dystrophin (CT-1) and the *E. coli* DnaK protein

The *E. coli* DnaK chaperone has been shown to change conformation upon ATP hydrolysis, resulting in its dissociation

Fig. 5. F.p.l.c. of the affinity-purified GST-CT-1 fusion protein

The GST-CT-1 fusion protein was purified from *E. coli* lysates by affinity chromatography, and was then subject to Mono Q f.p.l.c. as described in the Experimental section. (a) Approx. 20 mg of affinity-purified GST-CT-1 was applied to an f.p.l.c. Mono Q column, followed by elution with a linear NaCl gradient (0–750 mM). Fractions (1 ml) were collected and the absorption at 280 nm was monitored. (b) The protein composition of the fractions was analysed by SDS/PAGE and Coomassie Blue staining: lane 1, the affinity-purified GST-CT-1 applied to the column; lanes 2–5, fractions 7–10; lanes 6–8, fractions 18–20. (c) Proteins in the fractions were separated by SDS/PAGE, transferred electrophoretically to nitrocellulose membranes, and then screened with anti-dystrophin antibodies: lanes 9–10, fractions 18–20. The positions of molecular-mass marker proteins are shown.

from bound target proteins [42]. Therefore, in order to characterize further the interaction between the *E. coli* DnaK protein and the CT-1 fusion proteins, we determined the effect of ATP on their association. This was carried out using glutathione-Sepharose 4B affinity chromatography. The soluble *E. coli* lysate containing the GST-CT-1 and the DnaK proteins was applied to a glutathione-Sepharose 4B affinity column. After thorough washing, 3 mM-ATP was applied to the column (in 50 mM-Tris/HCl, pH 8.0, containing 3 mM-MgCl₂) and a single polypeptide band eluted. This band had a mobility corresponding to the SDS/PAGE mobility of the *E. coli* DnaK protein (Fig. 6, lane 1). The identity of this protein band was confirmed by immunoblotting (results not shown). Finally, the remaining protein bound to the affinity column was eluted with glutathione (see the Experimental section). The fractions contained the GST-CT-1 fusion protein, which, notably, was not associated with any of the DnaK protein (Fig. 6, lane 2). Identical results were obtained when the DnaK-GST-CT-1 complex (purified by Mono Q f.p.l.c.) was applied to a glutathione-Sepharose affinity column.

Fig. 6. ATP-dependent dissociation of the DnaK protein from the CT-1 fusion proteins

Affinity chromatography and SDS/PAGE were carried out as described in the Experimental section. *E. coli* lysate containing GST-CT-1 was applied to a glutathione-Sepharose 4B affinity column. Bound proteins were eluted with 3 mM-ATP, 3 mM-MgCl₂ (lane 1), followed by 5 mM-glutathione (lane 2). Fractions were analysed by SDS/PAGE. The arrowhead indicates the recombinant GST-CT-1. The positions of molecular-mass marker proteins are indicated.

In the absence of ATP the entire complex bound to the column; the *E. coli* DnaK protein was then selectively dissociated from the complex in the presence of 3 mM-ATP (results not shown).

DISCUSSION

In this study we describe the expression of fragments of the C-terminal region of dystrophin, in *E. coli*. We also describe the purification of these fragments from the *E. coli* lysate. We have shown that fragment CT-2 of the C-terminal region of dystrophin binds Ca²⁺, in Ca²⁺-overlay experiments, while fragment CT-1 specifically interacts with the *E. coli* DnaK gene product, a protein that is the *E. coli* equivalent of the mammalian hsp70 proteins [40].

Recently, a protocol has been published which describes the first successful purification of dystrophin, from skeletal-muscle membranes [43]. This is an important development; however, approx. 1.2–1.5 g of crude muscle membranes are required to purify only 30 µg of dystrophin [43]. The low yield, coupled with the extremely low natural abundance of dystrophin (the protein represents only 0.002 % of total muscle proteins) [4], makes the isolation of even moderate quantities of native dystrophin an almost impossible task. For this reason, we adopted the alternative approach of expressing fragments of dystrophin as fusion proteins. The experiments described in this paper constitute the only currently available alternative for obtaining quantities of protein sufficient to allow studies on the structure and function of dystrophin.

In this study we used fusion protein systems of GST [26] and protein A [27]. Both systems offer the advantages of high-level expression combined with procedures for simple, one-step, affinity purification under mild conditions. These features have enabled us to purify large quantities of fusion proteins carrying fragments of dystrophin. Another advantage of fusion-protein systems is that they allow the expression of discrete structural domains of a protein. In this study, we cloned the specific fragments as follows: the first fragment (CT-1) spans the C-terminal 201 amino-acid residues of skeletal muscle dystrophin and the second fragment (CT-2) spans the cysteine-rich region.

The fusion proteins used in this study were identified on the

basis of their purification from bacterial lysates by appropriate affinity chromatography, by their mobility in SDS/PAGE, and by their reactivity with specific antibodies or with peroxidase-labelled IgG. Significantly, antibodies raised against a synthetic peptide from the C-terminal region of dystrophin cross-react with the CT-1 fusion protein. In turn, antibodies raised against the CT-1 fragment cross-react with dystrophin in isolated sarcolemmal vesicles, further confirming the identity of the fragment (R. E. Milner, J. Busaan & M. Michalak, unpublished work). The fusion proteins were also characterized, in this study, by their interactions with a Mono Q column, and by their Ca^{2+} -binding properties in a $^{45}\text{Ca}^{2+}$ -overlay experiment. At pH 7.6, and in the absence of salt, all fusion proteins expressed in this study bound to a Mono Q f.p.l.c. column. Interestingly, both GST-CT-1 and protein A-CT-1 fusion proteins eluted from the Mono Q column at approx. 200 mM-NaCl, while both GST-CT-2 and protein A-CT-2 fusion proteins eluted at approx. 450 mM-NaCl (R. E. Milner, J. Busaan & M. Michalak, unpublished work). Since neither recombinant protein A nor recombinant GST bind to the Mono Q column under the conditions used in this study (results not shown) these results indicate that the interaction between the fusion proteins and the Mono Q matrix is mainly via the dystrophin fragments and not the fusion tails. Since the predicted pI values for CT-2 and CT-1 are approx. 9.7 and approx. 4.8 respectively, it is surprising that CT-2 exhibits the stronger interaction with the anion-exchange resin. However, both fragments contain similar numbers of acidic residues (33 mol% for CT-1 compared with 28 mol% for CT-2) and it is possible, therefore, that CT-2 might fold with its acidic residues exposed on the surface, while those in CT-1 might be buried within the molecule.

During this study, we noted additional significant differences between the two C-terminal fragments of dystrophin. We discovered that the fragment CT-1 interacts strongly with the *E. coli* DnaK protein [40]. The interaction was demonstrated by co-chromatography of the two polypeptides during both glutathione-Sepharose 4B affinity chromatography and Mono Q f.p.l.c. These interactions are specific to the CT-1 fragment since GST-CT-2, recombinant GST and a variety of different fusion proteins expressed in our laboratory [29a] have never shown any detectable interaction with the DnaK protein. The possible functional significance of this interaction is, as yet, an intriguing question. The *E. coli* DnaK protein is the equivalent of the mammalian hsp70 chaperones [40]. It has been suggested that the action of these proteins is closely related to their ability to bind and hydrolyse ATP [28,44]. Recently, hsp70 has been demonstrated to change conformation upon hydrolysis of ATP, triggering its dissociation from bound target proteins [42]. In accordance with this, we found that the interaction between CT-1 and the *E. coli* DnaK protein is specifically disrupted by the addition of ATP to the medium.

The *E. coli* DnaK protein is the only member of the hsp70 family expressed in prokaryotes [40]. The heat-shock proteins are expressed constitutively or in response to a variety of stimuli that range from pathophysiological to normal stress conditions (for review see [28]). Molecular chaperones, such as the DnaK gene product, are believed to bind to incorrectly folded proteins, thus preventing their aggregation and precipitation. The CT-1 fragment may not fold correctly when expressed in *E. coli* and be 'rescued' by interacting with the DnaK gene product. It is tempting to speculate, however, that interactions between DnaK, an analogue of mammalian hsp70, and the C-terminal region of dystrophin might be important during synthesis and/or trafficking of this very large protein.

In conclusion, we have expressed and purified fragments of the C-terminal region of dystrophin in a soluble form. This de-

velopment provides an important means by which the structure/function relationships within the molecule can be investigated in detail for the first time. In addition, we have discovered that this region of dystrophin interacts with the *E. coli* DnaK protein.

We thank Dr. B. Hodges for valuable advice in designing the synthetic peptide used for the production of antibodies. We thank Ms. K. Burns for suggestions concerning the terminology used for the fusion proteins. The superb technical help of Koji Shum is also greatly appreciated. This work was supported by the Muscular Dystrophy Association of Canada and the Alberta Heritage Foundation for Medical Research. R.E.M. is an Alberta Heritage Foundation for Medical Research Postdoctoral Fellow. M.M. is a Medical Research Council of Canada Scientist and a scholar of the Alberta Heritage Foundation for Medical Research.

REFERENCES

- Moser, H. (1984) *Hum. Genet.* **66**, 17–40
- Worton, R. G. & Thompson, M. W. (1988) *Annu. Rev. Genet.* **22**, 601–629
- Koenig, M., Hoffman, E. P., Bertelson, C. J., Monaco, A. P., Feener, C. & Kunkel, L. M. (1987) *Cell* **50**, 509–517
- Hoffman, E. P., Brown, R. H. & Kunkel, L. M. (1987) *Cell* **51**, 919–928
- Arahata, K., Ishiura, S., Ishiguro, T., Tsukahara, T., Suhara, Y., Eguchi, C., Ishihara, T., Nonaka, I., Ozawa, E. & Sugita, H. (1988) *Nature (London)* **333**, 861–863
- Bonilla, E., Samitt, C. E., Miranda, A. F., Hays, A. P., Salvati, G., Dimauro, S., Kunkel, L. M., Hoffman, E. P. & Rowland, L. P. (1988) *Cell* **54**, 447–452
- Watkins, S. C., Hoffman, E. P., Slayter, H. S. & Kunkel, L. M. (1988) *Nature (London)* **333**, 863–866
- Zubrzycka-Gaarn, E., Bulman, D. E., Karpati, G., Burghes, A. H. M., Belfall, B., Klamint, H. J., Talbot, J., Hodges, R. S., Ray, P. N. & Worton, R. G. (1988) *Nature (London)* **333**, 466–469
- Rowland, L. P. (1980) *Muscle Nerve* **3**, 3–20
- Sugita, H., Arahata, K., Ishiguro, T., Suhara, Y., Tsukahara, T., Ishiura, S., Eguchi, C., Nonaka, I. & Ozawa, E. (1988) *Proc. Jpn. Acad.* **64B**, 37–39
- Salvati, G., Betto, R., Ceoldo, S., Biasia, F., Bonilla, E., Miranda, A. F. & Dimauro, S. (1989) *Biochem. J.* **258**, 837–841
- Koenig, M., Monaco, A. P. & Kunkel, L. M. (1988) *Cell* **53**, 219–228
- Speicher, D. W. & Marchesi, V. T. (1984) *Nature (London)* **311**, 177–180
- Love, R. D., Hill, D. F., Dickson, G., Spurr, N. K., Byth, B. C., Marsden, R. F., Walsh, F. S., Edwards, Y. H. & Davies, K. E. (1989) *Nature (London)* **339**, 55–58
- Khurana, T. S., Hoffman, E. F. & Kunkel, L. M. (1990) *J. Biol. Chem.* **265**, 16717–16720
- Koenig, M. & Kunkel, L. M. (1990) *J. Biol. Chem.* **265**, 4560–4566
- Pons, F., Augier, N., Heilig, R., Leger, J., Mornet, D. & Leger, J. J. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 7851–7855
- Murayama, T., Sato, O., Kimura, S., Shimizu, T., Sawanda, H. & Murayama, K. (1990) *Proc. Jpn. Acad.* **66B**, 96–99
- Ohlendeck, K. & Campbell, K. P. (1991) *FEBS Lett.* **283**, 230–234
- Campbell, K. P. & Kahl, S. D. (1989) *Nature (London)* **338**, 259–262
- Ervasti, J. M. & Campbell, K. P. (1991) *Cell* **66**, 1121–1131
- Ohlendeck, K., Ervasti, J. M., Snook, J. B. & Campbell, K. P. (1991) *J. Cell Biol.* **112**, 135–148
- Yoshida, M. & Ozawa, E. (1990) *J. Biochem. (Tokyo)* **108**, 748–752
- Feener, C. A., Koenig, M. & Kunkel, L. M. (1989) *Nature (London)* **338**, 509–511
- Lemaire, C., Heilig, R. & Mandel, J. L. (1988) *EMBO J.* **7**, 4157–4162
- Smith, D. B. & Johnson, K. S. (1988) *Gene (Amst.)* **67**, 31–40
- Nilsson, B., Abrahmsen, L. & Uhlen, M. (1985) *EMBO J.* **4**, 1075–1080
- Schlesinger, M. J. (1990) *J. Biol. Chem.* **265**, 12111–12114
- Milner, R. E., Baksh, S., Shemanko, C., Carpenter, M. R., Smillie, L., Vance, J. E., Opas, M. & Michalak, M. (1991) *J. Biol. Chem.* **266**, 7155–7165
- 29a. Baksh, S. & Michalak, M. (1991) *J. Biol. Chem.* **266**, 21458–21465
- Towbin, H., Staehelin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4350–4354
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Michalak, M., Fliegel, L. & Wlasichuk, K. (1990) *J. Biol. Chem.* **265**, 5869–5874

- 32a. Maruyama, K., Mikawa, T. & Ebashi, S. (1984) *J. Biochem. (Tokyo)* **95**, 511–519
33. Ausubel, F. M., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1989) *Current Protocols in Molecular Biology*, vols. 1 and 2, Greene Publishing Associates and Wiley-Interscience, New York
34. Jones, L. R. (1988) *Methods Enzymol.* **157**, 85–91
35. Lowry, D. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. (1951) *J. Biol. Chem.* **193**, 265–275
36. Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254
37. Matsudaira, P. (1987) *J. Biol. Chem.* **262**, 10035–10038
38. Marston, F. A. O. (1986) *Biochem. J.* **240**, 1–12
39. Wilkinson, D. L. & Harrison, R. G. (1991) *Bio/Technol.* **9**, 443–446
40. Bardwell, J. C. A. & Craig, E. A. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 848–852
41. Beckman, R. P., Mizzen, L. A. & Welch, W. J. (1990) *Science* **248**, 850–854
42. Liberek, K., Skowrya, D., Zylicz, M., Johnson, C. & Georgopoulos, C. (1991) *J. Biol. Chem.* **266**, 14491–14496
43. Ervasti, J. M., Kahl, S. D. & Campbell, K. P. (1991) *J. Biol. Chem.* **266**, 9161–9165
44. Welch, W. J. & Feramisco, J. R. (1985) *Mol. Cell. Biol.* **5**, 1229–1237

Received 8 May 1992/16 July 1992; accepted 7 August 1992