Mutations in Calpain 3 Associated with Limb Girdle Muscular Dystrophy: Analysis by Molecular Modeling and by Mutation in m-Calpain

Zongchao Jia,* Vitali Petrounevitch,* Andrew Wong,* Tudor Moldoveanu,* Peter L. Davies,* John S. Elce,* and Jacques S. Beckmann[†]

*Department of Biochemistry, Queen's University and The Protein Engineering Network of Centres of Excellence, Kingston, Ontario K7L 3N6, Canada; and [†]Department of Molecular Genetics, Weizmann Institute of Science, 76100 Rehovot, Israel

ABSTRACT   Limb-girdle muscular dystrophy type 2A (LGMD2A) is an autosomal recessive disorder characterized by selective atrophy of the proximal limb muscles. Its occurrence is correlated, in a large number of patients, with defects in the human *CAPN3* gene, a gene that encodes the skeletal muscle-specific member of the calpain family, calpain 3 (or p94). Because calpain 3 is difficult to study due to its rapid autolysis, we have developed a molecular model of calpain 3 based on the recently reported crystal structures of m-calpain and on the high-sequence homology between p94 and m-calpain (47% sequence identity). On the basis of this model, it was possible to explain many LGMD2A point mutations in terms of calpain 3 inactivation, supporting the idea that loss of calpain 3 activity is responsible for the disease. The majority of the LGMD2A mutations appear to affect domain/domain interaction, which may be critical in the assembly and the activation of the multi-domain calpain 3. In particular, we suggest that the flexibility of protease domain I in calpain 3 may play a critical role in the functionality of calpain 3. In support of the model, some clinically observed calpain 3 mutations were generated and analyzed in recombinant m-calpain. Mutations of residues forming intramolecular domain contacts caused the expected loss of activity, but mutations of some surface residues had no effect on activity, implying that these residues in calpain 3 may interact in vivo with other target molecules. These results contribute to an understanding of structure-function relationships and of pathogenesis in calpain 3.

INTRODUCTION

The calpains are a family of intracellular cysteine proteases (Sorimachi et al., 1997; Ono et al., 1999; Carafoli et al., 1998; Belcastro et al., 1998). The two calpains first discovered, μ- and m-calpain, are ubiquitously expressed in mammalian tissues. These enzymes are strictly Ca^{2+} dependent, and are thought to be involved in many essential cellular functions involving Ca^{2+} signaling. They consist of a large or catalytic subunit (80 kDa, encoded by the closely related genes *CAPN1* and *CAPN2*, respectively), and a common regulatory subunit (28 kDa, *CAPN4*). The large subunits are divided into four structural domains (I to IV) and the small subunit contains domains V and VI. Crystal structures of m-calpain in the absence of Ca^{2+} have recently been described (Hosfield et al., 1999, Strobl et al., 2000).

Calpain 3, also commonly known as p94, is a muscle-specific form of calpain that appears to consist only of a 94-kDa catalytic subunit without a small subunit, requires little or no Ca^{2+} for activity, and is characterized by rapid and complete autolysis, with a half-life of <10 min (Sorimachi et al., 1993; Kinbara et al., 1998). The amino acid sequences of μ-calpain, m-calpain, and calpain 3 are ~50% identical, and well-aligned throughout the whole sequence, but calpain 3 possesses three additional segments, known as NS at the N terminus, IS1 in domain II, and IS2 between domains III and IV.

Mutations in *CAPN3* have been shown to be responsible for limb-girdle muscular dystrophy type 2A (LGMD2A), or calpainopathy (Richard et al., 1995, 1999), and more than 100 independent pathogenic mutations have been reported (Richard et al., 1997, 1999; Fardeau et al., 1996; Chou et al., 1999; Minami et al., 1999). It is assumed that all of these mutations abolish, or at least reduce, the function of calpain 3, but this has not been demonstrated in every case (Ono et al., 1999). Some of the mutations will obviously abolish protein function, such as nonsense, insertion/deletion, or splice-site mutations, but approximately half are missense mutations, single amino acid substitutions, the effect of which on calpain 3 activity is not immediately predictable. For several of these missense mutations, it was shown that the mutant calpain 3 was no longer able to hydrolyse fodrin, but in some cases was still able to autolyse (Ono et al., 1998; Sorimachi et al., 2000).

Because direct quantitative analysis of calpain 3 function is difficult due to its rapid autolysis and, consequently, extremely short half-life, we have attempted to predict the effects of single amino acid substitutions in calpain 3 using two alternative approaches. We have constructed a molecular model of calpain 3 using m-calpain as a template, making the assumption, based on the high level of sequence identity, that these structures should be very similar, despite the presence in calpain 3 of additional sequences and its apparent lack of a small subunit. The clinically significant mutations are scattered along the whole sequence, including some within the IS1 and IS2 regions, but most occur in regions that can be modeled onto m-calpain. The modeling is also of great interest because many of the clinically observed mutations with pathogenic consequences in-

volve amino acid residues that are highly conserved throughout all mammalian calpains (Richard et al., 1999). To provide experimental support for the modeling studies and to assess directly mutation effects on enzyme activity, several mutations that appear to be pathogenic in calpain 3 have been generated in the corresponding positions in m-calpain.

A direct correlation of clinical observations with the effects of mutation on the activity of the enzyme is not expected to be possible in every case. Each of the many distinct calpain 3 mutations has been observed only in one or a few patients, and the mutations usually occur in compound heterozygotes (i.e., where the other calpain 3 allele has a different mutation). Nonetheless, we have attempted to interpret the mutations with the assumption that a reduction or complete loss of calpain activity is a cause of LGMD2A.

MATERIALS AND METHODS

Modeling and structural analysis

The initial backbone threading was achieved using comparative protein modeling methods as implemented in the automated protein modeling server Swiss-Model (Peitsch, 1996). The protein sequence of calpain 3 was mapped to the coordinates of the rat m-calpain large subunit (Hosfield et al., 1999). The resulting model was visually inspected and manually corrected for any obvious defects, and then energy-minimized using program package Sybyl (Tripos, St. Louis, MO). The process of extensive energy minimization was allowed to continue until it reached convergence. For structural analysis of calpain 3 mutations, the amino acids with mutations that are implicated in LGMD2A were converted in the model to their mutated forms, and the mutated calpain 3 models were then briefly energy-minimized again ($\sim$200 cycles). All resulting mutant models were visually checked and analyzed. They were superimposed with the known crystal structures of rat and human m-calpains (Hosfield et al., 1999; Strobl et al., 2000). Model quality was assessed using Procheck (Laskowski et al., 1993) and figures were prepared with Molscript (Kraulis, 1991).

Site-directed mutagenesis and activity measurement

Several point mutations were made in m-calpain, to correspond to calpain 3 mutations reported in LGMD2A patients. Site-directed mutagenesis was performed with antisense primers on single-stranded DNA derived from pET-24-m-80k-CHis$_6$ (Elce et al., 1995). In each case, the presence of the mutation was confirmed by restriction digestion and DNA sequencing.

Expression, purification, and activity assay of mutant m-calpains were performed as previously described (Elce et al., 1995; Dutt et al., 2000).

RESULTS AND DISCUSSION

An alignment of the amino acid sequences of the rat m-calpain large subunit and human calpain 3 is shown in Fig. 1. To avoid confusion in the numbering of residues, all positions are described in terms of the human calpain 3 amino acid sequence. Where necessary in the text, rat m-calpain amino acid residue numbers are given in italics.

Model of calpain 3

As expected, the modeled calpain 3 structure is very similar to that of m-calpain (Fig. 2) in the absence of Ca^{2+}. The r.m.s. deviations for all corresponding Cα between calpain 3 and m-calpain is 1.33 Å. The model is unavoidably less certain in the regions of IS1 and IS2 sequences, which represent insertions of $\sim$60 and $\sim$64 residues, respectively, when compared with the large subunit of m-calpain. Due to lack of homologous structural template, the NS region was not modeled. The insertions are clearly important for two reasons. In the first place, clinically deleterious mutations, P319L in IS1, and S606L and Q638P in IS2, have been observed in these regions. Secondly, these regions are subject to alternative splicing, so that mutations therein might affect some or all of the possible isoforms. The work described here was restricted to residues that can be mapped onto m-calpain, and therefore refers only to the predominant mature form of calpain 3 in skeletal muscle.

In the model, the two inserted sequences, IS1 and IS2, protrude out of the globular core structure, and can largely be accommodated without disruption of the overall fold. Most of IS1 (G269-G329) modeled as a two-stranded β-sheet and a coil. However, residues D267-M272, at the beginning of IS1, form a loop in the calpain 3 model, oriented into the protein toward domain I, that is expected to affect the activity of the enzyme. The IS2 region (S589-Q653) was resolved in the model as two separate α-helices. These helices are in the middle of the transducer arm (*D514-F538*), which, in m-calpain, was proposed to communicate Ca^{2+}-induced changes in domain IV to the rest of the molecule. Introduction of IS2 into this arm may have the effect similar to Ca^{2+}-induced change to this part of the structure and hence contribute to the very low Ca^{2+} requirement for the activation of calpain 3.

In m-calpain, it was immediately obvious from the x-ray structures that domains I and II, although they already have many contacts, must move closer together after Ca^{2+}-induced activation of the enzyme to assemble the active site. Movement of domain I is thought to be facilitated by the autolysis of the N terminus, and movements of both domains I and II, (involving or mediated by domain III), are thought to occur in response to Ca^{2+}-induced conformational changes in domains IV and VI (Hosfield et al., 1999). Although LGMD2A-associated mutations are observed throughout the sequence, many mutations are clustered in only three areas, all of which might affect domain movement (Fig. 3). The three clusters are located at K207-G234, E435-R448, and S479-I502. The main contact of domain I with domain III involves a group of α-helices in domain I containing the first cluster of mutations, K207-G234, that are in close proximity to a variety of distinct regions in domain III including a loop which contains the third cluster of mutations, S479-I502. It is likely that changes in this interaction will affect the flexibility of domain I, and there-
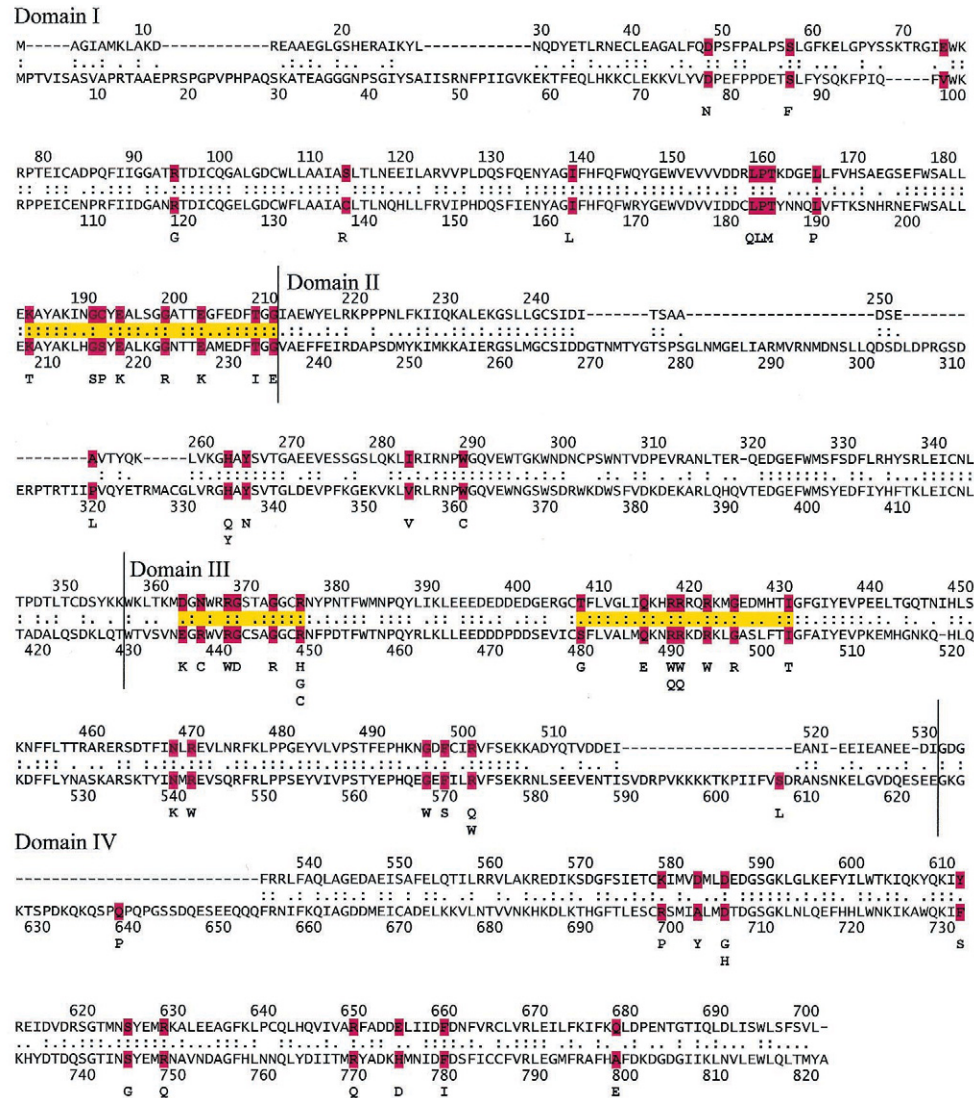
Domain I

FIGURE 1   The primary sequence alignment of large subunit of m-calpain (*upper sequence*) and calpain 3 (*lower sequence*) shows a high degree of homology between these two proteins (47% identity). The amino acids implicated in LGMD2A are seen throughout the sequence of calpain 3 (*purple*). The regions where mutations affect domain-domain interaction and assembly are grouped into three distinct clusters (*yellow bar*). Each mutation residue is highlighted in bold letters, below the corresponding amino acids.

fore affect assembly of the active site. The second cluster of mutations, E435-R448, is located on a loop of domain III that has interactions with domain IV, and is expected to be involved in the communication of Ca^{2+}-induced conformational change throughout these domains.

Calpain 3-related mutations expressed in m-calpain

For logistical reasons, only six calpain 3 mutations were directly analyzed by mutagenesis in m-calpain. These six m-calpain mutants, each incorporating a single amino acid substitution corresponding to a known pathogenic mutation in calpain 3, were created by mutagenesis, expressed in *Escherichia coli*, purified, and their specific activities mea-

sured (Table 1). All mutants were well expressed as soluble heterodimers, which we generally take as an indication that their folding proceeded normally. Of these, four mutant enzymes had very low, but not zero, specific activities, consistent with a pathogenic role in LGMD2A; but two, R448H and D705G, were fully active, so that their apparent pathogenicity is not immediately understood. The effects of these m-calpain mutations will be discussed together with the calpain 3 model in the following sections.

Domain I

There are 18 mutations within this domain (Fig. 4B), and, except for S86F, they are mostly associated with a relatively mild course of the disease. The mutations D77N and S86F
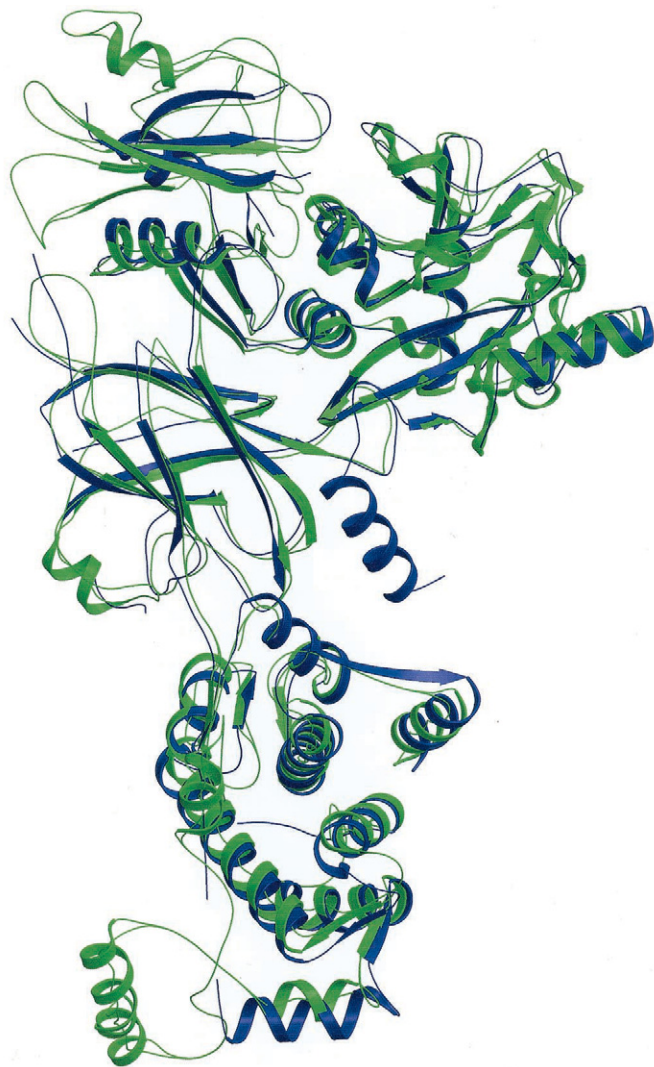
FIGURE 2 Superimposition of calpain 3 model (*green*) and the crystal structure of m-calpain (*blue*) by Cα alignment. The modeled structure of calpain 3 displays a high similarity to m-calpain with the exception of two insertion regions, IS1 and IS2.
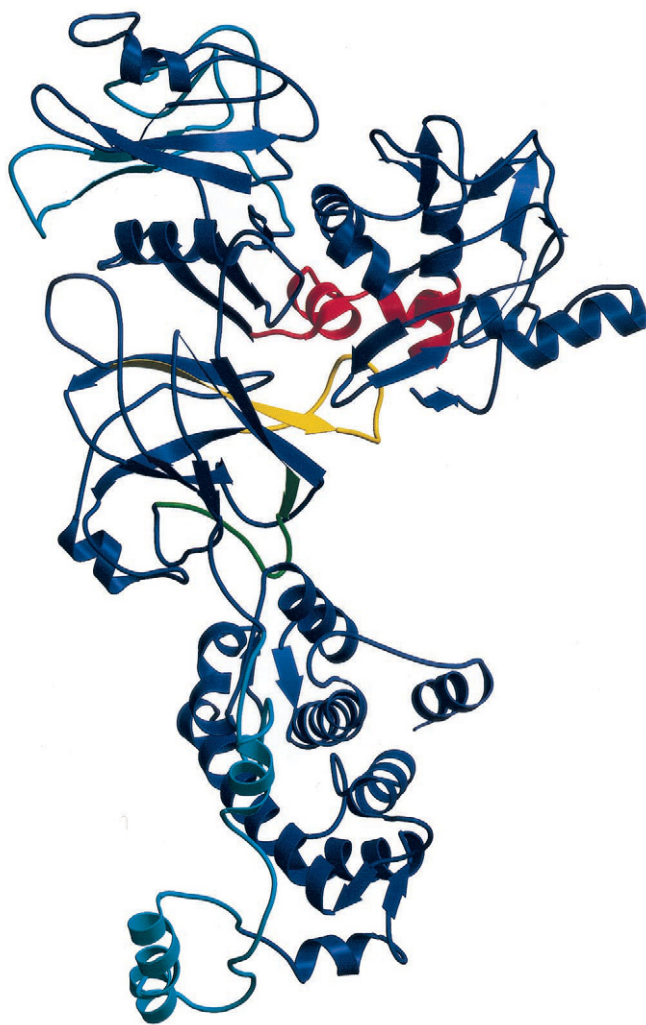
FIGURE 3 Three mutation clusters identified in the primary structure of calpain 3 are shown as colored regions in the corresponding model. The mutation clusters in domain I (*red*) and domain III (*yellow*) share a large interface and are expected to reduce the rigidity of domain I to facilitate calpain 3 inactivation. The third mutation cluster in domain III (*green*) may have an effect on domain III-domain IV interaction. The two insertion regions, IS1 and IS2, are located in domain II and in the linker region of domain III, respectively (*light blue*).

may influence the positioning of the N terminus of the enzyme; S86F clearly introduces a major steric disturbance, and clinically is one of the most severe missense mutations. The residues R118 and C137 are very close to portions of domain II, and their mutations, R118G and C137R, introduce major changes in size and charge, which would disrupt domain I/domain II interaction in the resting state, and would also affect assembly of the active site during enzyme activation. The m-calpain mutation, *R94G*, corresponding to R118G, was expressed in *E. coli*, and had only 5% of wild-type activity (Table 1). The effects of several other mutations in domain I are more difficult to explain. For instance, the mutations V98I and I162L both exchange two very similar hydrophobic residues, without any obvious effects on the structure. It is possible that these residues

Table 1 Mutations in rat m-calpain

Actual mutation in rat m-Calpain	Corresponding mutation in calpain 3	Expression level (wild type, %)*	Specific activity (wild type, %)[†]
R94G	R118G	100	5
E193K	E217K	100	5
G198R	G222R	100	10
E202K	E226K	100	5
R375H	R448H	100	100
D585G	D705G	100	100

*Expression levels of the mutated 80-kDa m-calpain subunits were estimated by immunoblotting and expressed as a percentage of wild-type m-calpain.
[†]Specific activities of the purified enzymes were measured as described (Dutt et al., 2000) and expressed as a percentage of wild-type m-calpain. Each purification was performed only once.
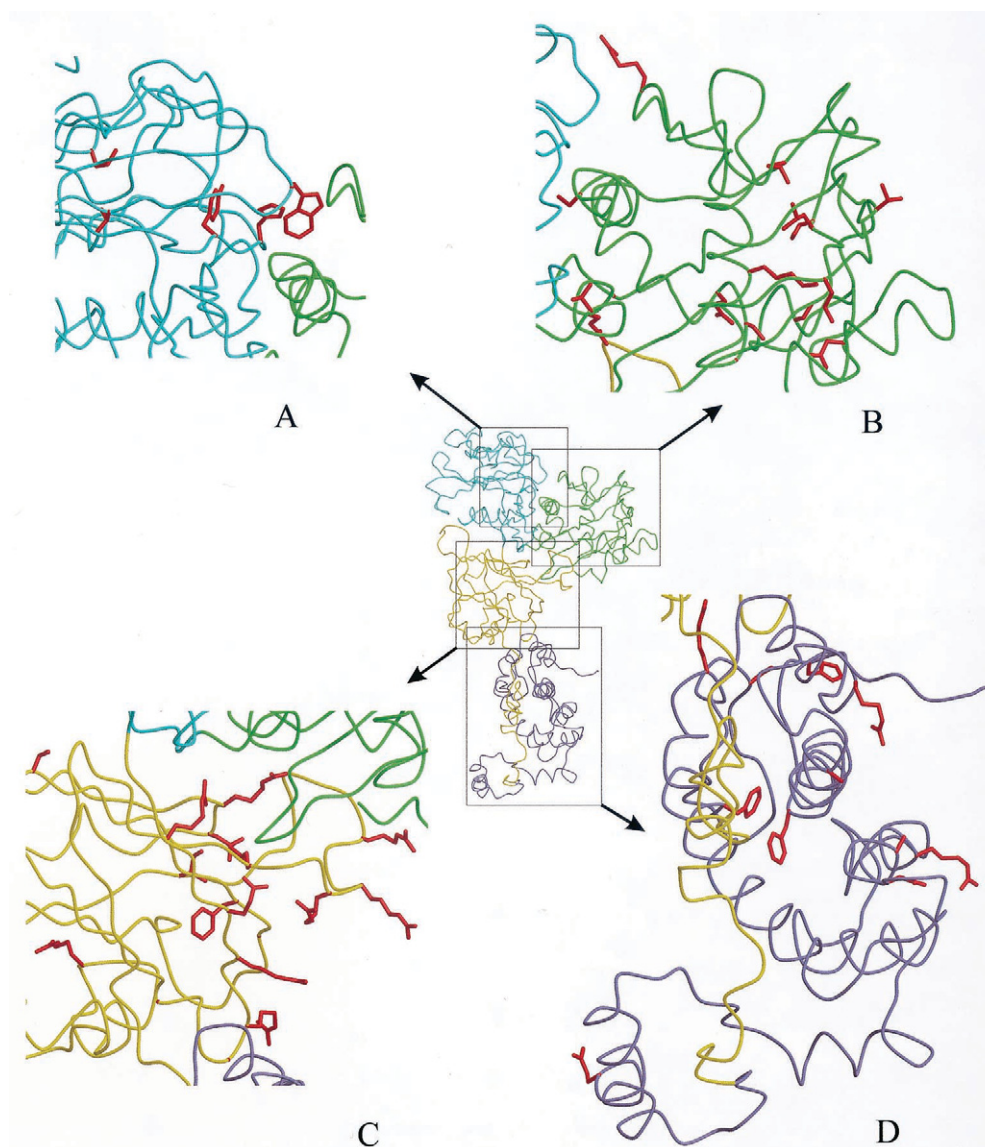
FIGURE 4   Clustered mutation sites in the calpain 3 model. *A, B, C,* and *D* are the regions of domains II, I, III, and IV, respectively. The majority of mutations implicated in the development of LGMD2A are located at the domain interfaces with an especially high occurrence at the domain I-domain III contact region.

affect either folding of the enzyme, or its interactions with cellular targets.

Eight of the domain I mutations are found, however, in the cluster K207-G234 (Fig. 3), which makes many contacts with a loop in domain III, and this loop itself contains the third cluster of mutations (S479-I502). The E217K, G222R, and E226K mutations were bacterially expressed in the corresponding positions in m-calpain, and all had very weak activity (Table 1). This is consistent with the model because E217, for example, is close to K494 in domain III, and E226 and E229 are close to R541 (which does not belong to the third cluster); these mutations in calpain 3 would be expected to disrupt salt links between domains I and III. In agreement with those observations, it may be noted that the

R541W mutation in domain III is also associated with the disease. G222 is close to L495, so the mutation G222R is likely to disturb the local interactions.

Domain II

This domain contains relatively few pathogenic mutations (Fig. 4 A), but most of them are readily interpretable. H334QY abolishes the histidine residue of the catalytic triad, which will drastically reduce activity. Interestingly, the patient carrying this H334Q mutation is a compound heterozygote with a null mutation, so that any calpain 3 activity must derive from the H334Q allele. This patient
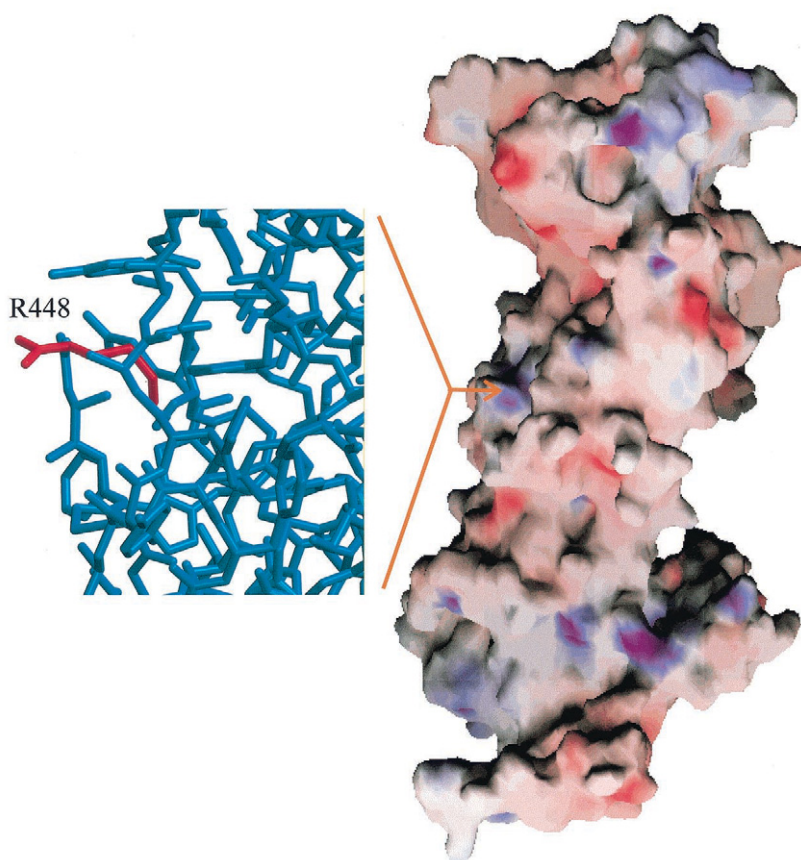
FIGURE 5 The position of R448 in domain III of the calpain 3 model. This residue is implicated in LGMD2A but is remote from the neighboring domains. R448 is thought to be potentially involved in the interaction with other partners.

shows a relatively mild course of disease. The related *H262A* mutation in m-calpain had been made previously, and was inactive (Arthur et al., 1995). The mutation Y336N alters a set of direct and indirect contacts that affect the orientation and electrostatic properties both of H334 and of W360. The mutation W360C is of interest because this tryptophan residue has frequently been studied in cysteine proteases, and may have two roles in calpain. It appears to affect the active-site triad directly in many cysteine proteases, and it may be additionally important in m-calpain in the relative motions of domain I and II (Hosfield et al., 1999). Several patients homozygous for the W360C mutation have been described, and show a slightly more severe disease development than those with H334Q. In m-calpain the corresponding W288Y mutation gave an enzyme that was very weakly active (Arthur et al., 1995).

Domain III

As mentioned earlier, most of the mutations in domain III lie in the region of contact between domain III and domain I (Figs. 3 and 4 C). Four residues in this region are directly involved in salt links with residues in domain I, so that the pathogenic effects of their mutations (E435K, R493W, R541W, R572WQ), can be readily understood in terms of disruption of salt links and of communication between these two domains. The mutations R440W, G441D, and G445R are all located at the interface between domains III and IV, and involve an alteration in charge. These mutations must also disrupt communication between Ca^{2+}-induced alteration in domain IV and the rest of the molecule. The residue R448 is an exception (Fig. 5). It undergoes pathogenic mutations to H, G, and C, but in the model R448 is exposed at the surface and has no interactions with other residues. In m-calpain, the corresponding *R375H* expressed in *E. coli* was fully active in our in vitro assay. For both R448H and D705G (see below) it is assumed that the observed mutations do not affect the intrinsic activity of calpain 3, but must disturb the interaction of calpain 3 with other molecules in vivo.

Domain IV

There are 11 reported pathogenic mutations in this domain (Fig. 4 D), but the model does not offer clear explanations in every case. The mutations Q638P (located in a modeled α-helix that is not present in m-calpain), and R698P (also in an α-helix), will cause serious disruption because helices cannot accommodate a Pro residue. The mutation D705G involves a key Ca^{2+}-chelating residue in *EF-hand* 2, and

might be expected to affect the Ca^{2+} requirement of calpain 3. However the corresponding mutant in m-calpain, *D585G*, was fully active, and we have shown elsewhere that other alterations in *EF-hand* 2 had very little effect on m-calpain (Dutt et al., 2000). Owing to the difficulties of preparing calpain 3, its Ca^{2+} requirement has not been directly measured.

In conclusion, we have noted that most calpain 3 missense mutations are clustered in three areas that appear to affect intramolecular domain interactions and may impair the assembly and activation of this multi-domain protein. In particular, the flexibility of domain I is directly affected in many mutants, and is very likely to impair assembly of the active site and consequently function of calpain 3. The pathogenic effects of many missense mutations in calpain 3 can therefore be readily understood in terms of loss of catalytic activity, and the predictions are supported by expression of the related mutations in m-calpain. A very high proportion of the clinically observed mutations involves an alteration in charge of side chains involved in internal salt links, changes that are expected to be highly disruptive. These results also strongly suggest that the proposed model of calpain 3 is close to correct. There are some mutations, exemplified by V98I, where the mutation is not predicted to have much effect on proteolytic capacity; and there are at least two mutations, R448HGC and D705G, where the mutation has no effect when expressed in m-calpain. The clinical consequences of these latter two mutations seem likely to involve effects in vivo on the interaction of calpain 3 with its substrates or binding targets.

REFERENCES

Arthur, J. S. C., S. Gauthier, and J. S. Elce. 1995. Active site residues in m-calpain: identification by site-directed mutagenesis. *FEBS Lett.* 368: 397–400.

Belcastro, A. N., L. D. Shewchuk, and D. A. Raj. 1998. Exercise-induced muscle injury: a calpain hypothesis. *Mol. Cell. Biochem.* 179:135–145.

Carafoli, E., and M. Molinari. 1998. Calpain: a protease in search of a function? *Biochem. Biophys. Res. Commun.* 247:193–203.

Chou, F. L., C. Angelini, D. Daentl, C. Garcia, C. Greco, I. Hausmanowa-Petrusewicz, A. Fidzianska, H. Wessel, and E. P. Hoffman. 1999. Calpain III mutation analysis of a heterogeneous limb-girdle muscular dystrophy population. *Neurology.* 52:1015–1020.

Dutt, P., J. S. C. Arthur, P. Grochulski, M. Cygler, and J. S. Elce. 2000. Roles of individual EF-hands in the activation of m-calpain by calcium. *Biochem. J.* 348:37–43.

Elce, J. S., C. Hegadorn, S. Gauthier, J. W. Vince, and P. L. Davies. 1995. Recombinant calpain II: improved expression systems and production of

a C105A active-site mutant for crystallography. *Protein Eng.* 8:843–848.

Fardeau, M., D. Hillaire, C. Mignard, N. Feingold, J. Feingold, D. Mignard, B. de Ubeda, H. Collin, F. M. Tome, I. Richard, and J. S. Beckmann. 1996. Juvenile limb-girdle muscular dystrophy. Clinical, histopathological and genetic data from a small community living in the Reunion Island. *Brain.* 199:295–308.

Hosfield, C. M., J. S. Elce, P. L. Davies, and Z. Jia. 1999. Crystal structure of calpain reveals the structural basis for Ca^{2+}-dependent protease activity and a novel mode of enzyme activation. *EMBO J.* 18: 6880–6889.

Kinbara, K., S. Ishiura, S. Tomioka, H. Sorimachi, S. Y. Jeong, S. Amano, H. Kawasaki, B. Kolmerer, S. Kimura, S. Labeit, and K. Suzuki. 1998. Purification of native p94, a muscle-specific calpain and characterization of its autolysis. *Biochem. J.* 335:589–596.

Kraulis, P. J. 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24:946–950.

Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton. 1993. Procheck: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26:283–291.

Minami, N., I. Nishino, O. Kobayashi, K. Ikezoe, Y. Goto, and I. Nonaka. 1999. Mutations of calpain 3 gene in patients with sporadic limb-girdle muscular dystrophy in Japan. *J. Neurol. Sci.* 171:31–37.

Ono, Y., H. Shimada, H. Sorimachi, I. Richard, T. C. Saido, J. S. Beckmann, S. Ishiura, and K. Suzuki. 1998. Functional defects of a muscle-specific calpain, p94, caused by mutations associated with limb-girdle muscular dystrophy type 2A. *J. Biol. Chem.* 273:17073–17078.

Ono, Y., H. Sorimachi, and K. Suzuki. 1999. New aspect of the research on limb-girdle muscular dystrophy 2A: a molecular biologic and biochemical approach to pathology. *Trends Cardiovasc. Med.* 9:114–118.

Peitsch, M. C. 1996. ProMod and Swiss-Model: Internet-based tools for automated comparative protein modeling. *Biochem. Soc. Trans.* 24: 274–279.

Richard, I., L. Brenguier, P. Dincer, C. Roudaut, B. Bady, J. M. Burgunder, R. Chemaly, C. A. Garcia, G. Halaby, C. E. Jackson, D. M. Kurnit, G. Lefranc, C. Legum, J. Loiselet, L. Merlini, A. Nivelon-Chevallier, E. Ollagnon-Roman, G. Restagno, H. Topaloglu, and J. S. Beckmann. 1997. Multiple independent molecular etiology for limb-girdle muscular dystrophy type 2A patients from various geographical origins. *Am. J. Hum. Genet.* 60:1128–1138.

Richard, I., O. Broux, V. Allamand, F. Fougerousse, N. Chiannilkulchai, N. Bourg, L. Brenguier, C. Devaud, P. Pasturaud, and C. Roudaut. 1995. Mutations in the proteolytic enzyme calpain3. cause limb-girdle muscular dystrophy type 2A. *Cell.* 81:27–40.

Richard, I., C. Roudaut, A. Saenz, R. Pogue, J. E. Grimbergen, L. V. Anderson, C. Beley, A. M. Cobo, C. de Diego, B. Eymard, P. Gallano, H. B. Ginjaar, A. Lasa, C. Pollitt, H. Topaloglu, J. A. Urtizberea, M. de Visser, A. van der Kooi, K. Bushby, E. Bakker, A. Lopez de Munain, M. Fardeau, and J. S. Beckmann. 1999. Calpainopathy—a survey of mutations and polymorphisms. *Am. J. Hum. Genet.* 64:1524–1540.

Sorimachi, H., S. Ishiura, and K. Suzuki. 1997. Structure and physiological function of calpains. *Biochem. J.* 328:721–732.

Sorimachi, H., N. Minami, Y. Ono, K. Suzuki, and I. Nonaka. 2000. Limb-girdle muscular dystrophy with calpain 3 (p94) gene mutations (calpainopathy). *Neurosci. News.* 3:20–27.

Sorimachi, H., N. Toyama-Sorimachi, T. C. Saido, H. Kawasaki, H. Sugita, M. Miyasaka, K. Arahata, S. Ishiura, and K. Suzuki. 1993. Muscle-specific calpain, p94, is degraded by autolysis immediately after translation, resulting in disappearance from muscle. *J. Biol. Chem.* 268: 10593–10605.

Strobl, S., C. Fernandez-Catalan, M. Braun, R. Huber, H. Masumoto, K. Nakagawa, A. Irie, H. Sorimachi, G. Bourenkow, H. Bartunik, K. Suzuki, and W. Bode. 2000. The crystal structure of calcium-free human m-calpain suggests an electrostatic switch mechanism for activation by calcium. *Proc. Natl. Acad. Sci. U.S.A.* 97:588–592.