Available online at www.sciencedirect.com
ScienceDirect
Genomics 89 (2007) 224–236
GENOMICS
www.elsevier.com/locate/ygeno

The sarcomeric myosin heavy chain gene family in the dog: Analysis of isoform diversity and comparison with other mammalian species

Lisa Maccatrozzo [a], Francesca Caliaro [a], Luana Toniolo [b], Marco Patruno [a], Carlo Reggiani [b,*], Francesco Mascarello [a]

[a] Department of Experimental Veterinary Sciences, University of Padova, 35131 Padua, Italy
[b] Department of Anatomy and Physiology, University of Padova, 35131 Padua, Italy

Received 7 July 2006; accepted 2 August 2006
Available online 20 September 2006

Abstract

Sarcomeric myosin heavy chains (MyHC) are the major contractile proteins of cardiac and skeletal muscles and belong to class II MyHC. In this study the sequences of nine sarcomeric MyHC isoforms were obtained by combining assembled contigs of the dog genome draft available in the NCBI database. With this information available the dog becomes the second species, after human, for which the sequences of all members of the sarcomeric MyHC gene family are identified. The newly determined sequences of canine MyHC isoforms were aligned with their orthologs in mammals, forming a set of 38 isoforms, to search for the molecular features that determine the structural and functional specificity of each type of isoform. In this way the structural motifs that allow identification of each isoform and are likely determinants of functional properties were identified in six specific regions (surface loop 1, loop 2, loop 3, converter, MLC binding region, and S2 proximal segment).

Keywords: Dog; Genome; Sarcomere; Myosin heavy chain; Loops; Muscle contraction

The interaction between myosin and actin is the molecular basis of muscle contraction and ATP hydrolyzed by myosin is the energy source for mechanical power output. The myosin molecules incorporated in the myofibrils of striated muscles belong to class II [1] and are often identified as "sarcomeric myosins." Sarcomeric myosins are heterohexamers consisting of two myosin heavy chains (MyHC) and two pairs of two distinct myosin light chains (MyLC; essential and regulatory light chains or ELC and RLC). MyHC isoforms determine the contractile properties of the myosin molecule and are considered the molecular markers of the fiber type [2]. For this reason the fiber-type names are generally used for sarcomeric myosin isoforms.

In all mammalian species examined until now nine distinct MyHC isoforms (paralogous isoforms) are expressed in striated cardiac and skeletal muscles and incorporated in the thick filaments. The nine MyHC isoforms are coded by distinct genes [1] and can be grouped in three subfamilies [3,4] based on their chromosomal localization. In the human genome, a first subfamily of genes is localized on chromosome 17 and includes three fast isoforms expressed in adult fast skeletal muscle fibers (called 2A, 2X, and 2B, coded by the genes MYH2, MYH1, and MYH4, respectively), two developmental or transitional isoforms expressed during the early stage of muscle development (embryonic; gene MYH3) or during the neonatal period (neonatal or perinatal; gene MYH8), and a sixth isoform expressed only in extraocular muscles (EO; gene MYH13). A second subfamily, localized on chromosome 14, includes the two isoforms specific for cardiac muscle, the slow MyHC isoform (also called β or type 1, coded by the gene MYH7), expressed in ventricular myocardium and in slow skeletal muscle fibers, and the α-cardiac (coded by gene MYH 6) expressed in atrial myocardium and in some specialized skeletal muscles such as the masseter muscle. Finally, the third subfamily is composed of one member, the masticatory isoform (2M or M; gene MYH16), the gene being localized on human chromosome 7 [5] and the protein present only in jaw muscles of carnivores and some primates [6], but not in humans [7]. The

* Corresponding author. Fax: +390498275301.
E-mail address: carlo.reggiani@unipd.it (C. Reggiani).

0888-7543/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2006.08.004

proposed extension of the sarcomeric MyHC family to two other genes (MYH14 and MYH15 [8]) still waits to be confirmed by expression studies.

A large body of evidence has been accumulated on the functional diversity among MyHC isoforms. In all animal species investigated until now the sliding velocity of myosin filaments measured in in vitro motility assays or in muscle fibers increases in the order β/1 or slow, α-cardiac, 2A, 2X, and 2B [9,10]. Less information is available on the developmental isoforms embryonic and perinatal (hereafter called Emb and Peri) and on the specialized isoforms EO and 2M. There are, however, indications that Emb and Peri are associated with low values of sliding filament speed, whereas high values of movement speed are generated by EO and 2M [11,12]. ATPase activity follows a similar trend, increasing from β/1 to α-cardiac and from 2A to 2X and 2B [13]. MyHC 2M has been shown to possess very high enzymatic activity [14]. The variations of the functional properties likely find their basis in the diversity in the primary structure of the MyHC isoforms. Three surface loops (loop 1, loop 2, and loop 3) have been suggested by previous comparative studies as possible determinants of the structural and functional features of various isoforms [15,16]. Mutation studies have given support to those hypotheses [17,18]. The converter domain has been shown to determine functional properties of MyHC in insect muscles [19], and other two potentially relevant regions are the light-chain-binding α-helix and the proximal part of the S2 subfragment where binding sites for MyBP-C are localized.

Among mammalian species, the nucleotide and amino acid sequences of all nine sarcomeric MyHC genes are known in humans (complete sequences for type β/1, 2A, 2X, 2B, EO, Emb, Peri, and α-cardiac and partial sequence for 2M), whereas the sequences of the four adult skeletal muscle MyHC are available for the pig (type β/1, 2A, 2X, and 2B) and cattle (type β/1, 2A, 2X, and 2B), and only three adult skeletal muscle MyHC (type β/1, 2A, and 2X) have been sequenced for the horse. More scattered data are available for laboratory animals such as the mouse (type β/1, 2A, 2X, and α-cardiac), the rat (type β/1, α-cardiac, and Emb), and the rabbit (type 2X and EO). The MyHC 2M sequence of the cat has been reported by Hoh and co-workers [5].

The reconstruction of the phylogenetic relations among the known MyHC isoforms based on a tree of 18 sarcomeric isoforms including scallop and several vertebrate (humans, chicken, quail, rabbit, rat) isoforms [4] and on a tree of all class II human MyHC isoforms [8] has been recently published. The analysis has revealed the presence in mammalian genomes of three groups, 2M, fast/developmental (2A, 2X, 2B, Peri, Emb, and EO), and slow/cardiac (α and β/1), which coincide with the three subfamilies with distinct chromosomal localizations mentioned above. Although all MyHC derive from a common ancestor gene, masticatory MyHC has diverged considerably from fast/developmental and slow/cardiac isoforms after gene duplications and is now distantly related to the other mammalian fast/developmental and slow/cardiac subclasses [4].

The first aim of this study was to identify the sequences of all sarcomeric MyHC isoforms in the recently published canine genome [20]. Several muscle diseases are of interest for veterinarian medicine, and dog muscle adaptations in relation to exercise are often studied. A precise analysis of skeletal muscles in physiological and pathological conditions requires correct and reliable fiber-type identification and this is based on MyHC isoform expression (see above and [2]). The identification of the sequences of MyHC isoforms is an essential step to develop reliable tools to study their expression. The identification of the 9 canine MyHC isoforms brings to 38 the number of sarcomeric MyHC sequences determined in various mammalian species. This offers an unprecedented possibility of carrying out extensive comparisons between paralogs and orthologs to define which molecular features identify each specific type of isoform. Thus, the second aim of this study was to exploit the potential of information contained in the MyHC sequences, to understand better the basis of the structural diversity of sarcomeric myosin isoforms.

Results and discussion

The nucleotide sequences of nine MyHC isoforms, i.e., 2A, 2B, 2X, Emb, Peri, EO, β/1, α-cardiac, and 2M, were identified in the draft of the dog genome available at the NCBI Dog Genome Resources Web site. Exon and intron sequences were generated for each isoform and the putative amino acid (aa) sequences were deduced, as described under Methods. The translation predicted an open reading frame of 1940 aa for 2A, EO, and Emb MyHC; 1939 aa for 2B, 2X, and Peri MyHC; and 1935 aa for MyHC β/1. Only partial putative proteins, composed of 1589 and 1880 aa, respectively, could be predicted for α-cardiac and 2M MyHC. It must be emphasized that, although it is relatively easy to recognize the MyHC-like sequences in genomic DNA, it is more difficult to predict and assemble all exons, generally about 40, that are present in a full-length MyHC transcript.

All novel MyHC sequences were deposited with GenBank and are also available in Supplementary Fig. 1, in which all the sequences of mammalian MyHC isoforms considered in this study are aligned. The accession numbers of the dog MyHC isoforms are reported in Table 1, together with the accession numbers of all other isoforms utilized for the comparative analysis (see below). The availability of these data makes the dog the second best known species, after human, with regard to the MyHC isoform sequences.

By BLASTN analyses we could also localize each MyHC gene in the dog genome. Three distinct localizations were found: one cluster on chromosome 5 contains the fast isoforms (2A, 2X, 2B, Peri, Emb, EO), a second cluster on chromosome 8 includes β/1 and α-cardiac MyHC, while MyHC 2M stands alone on chromosome 6 (Fig. 1). Such cluster organization is similar to that previously described for other mammalian species [3,23].

General comparison between canine MyHC isoforms and their orthologs in other mammalian species

BLAST searches and sequence comparison revealed the expected high levels of amino acid similarity between dog

Table 1
Accession numbers of the mammalian sarcomeric MyHC isoforms available on the NCBI database

Isoform	NCBI Accession No.
Canis 2X	DQ227281
Canis 2B	DQ227282
Canis 2A	DQ227280
Canis Peri	DQ227283
Canis Emb	DQ227279
Canis EO	DQ227284
Canis α	DQ227286
Canis β/1	DQ227285
Canis 2M	DQ227287
Homo 2X	AF111785
Homo 2B	AF111783
Homo 2A	AF111784
Homo Peri	M36769
Homo Emb	X13988
Homo EO	AF111782
Homo α	D00943
Homo β/1	X52889
Rattus 2B	XP_340819
Rattus 2X	XP_213345
Rattus Emb	X04267
Rattus α	X15938
Rattus β/1	X15939
Mus 2A	AAH08538
Mus 2X	XP_137643
Mus α	NM_010856
Mus β/1	AY056464
Sus 2A	AB025260
Sus 2B	AB025261
Sus 2X	AB025262
Sus β/1	AB053226
Bos 2A	AB059398
Bos 2X	AB059399
Bos β/1	AB059400
Equus 2A	AB088365
Equus 2X	AB088366
Equus β/1	AB088367
Oryctolagus 2X	U32574
Oryctolagus EO	AF212148
Felis 2M	U51472
Scallop MHC	X55714

Scallop MyHC was used as outgroup.

MyHC isoforms and the orthologous proteins in other mammals. Sequences of 38 MyHC isoforms belonging to eight mammalian species (human, dog, pig, cattle, horse, mouse, rat, rabbit, cat) and scallop MyHC were aligned using DiAlign software. The proportion of nonidentical amino acids, determined in paired comparison between all isoforms, is reported in Supplementary Fig. 2. As expected [9,24], the comparison between orthologs revealed the highest similarity (e.g., more than 99% identity between rat and mouse β/1 MyHC), whereas the comparison between paralogs showed greater diversity.

Canine MyHC isoforms were highly similar to the orthologous isoforms of other species. As can be seen in Table 2, the identity was 95% for MyHC 2B, 96-97% for 2X, and 97-98% for type 2A and type β/1. Among developmental isoforms the percentage of identical amino acids was higher for MyHC Emb

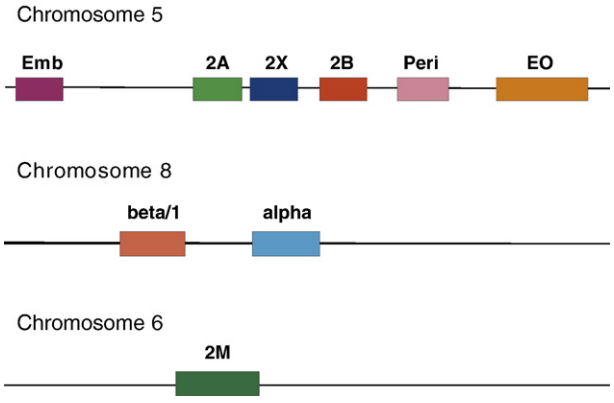

Fig. 1. Chromosomal localization of the sarcomeric MyHC isoforms in the canine genome.

(97-98%) than for Peri (96%), whereas among isoforms expressed in craniofacial muscles the identity was 98% for MyHC 2M (only 1880 aa available) and 94-95% for MyHC EO.

The general rule that orthologs show higher identity than paralogs finds few exceptions. For example, the porcine MyHC 2B is closer to MyHC 2X than to MyHC 2B of other species (see also below, phylogenetic trees and Supplementary Fig. 2). The diversity between paralogs (Table 2 and Supplementary Fig. 2) is

Table 2
Identity at protein level between canine sarcomeric MyHC isoforms and their orthologs in other mammalian species

MyHC	Ortholog	% Identity
(a) *Canis* 2B	Sus	95
	Rattus	95
	Homo	95
(b) *Canis* 2X	*Sus*	97
	Bos	96
	Mus	96
	Equus	96
	Homo	96
	Oryctolagus	97
	Rattus	97
(c) *Canis* 2A	*Sus*	98
	Bos	97
	Equus	98
	Homo	97
(d) *Canis* EO	*Homo*	94
	Oryctolagus	95
(e) *Canis* Emb	*Homo*	98
	Rattus	97
Canis Peri	*Homo*	96
(f) *Canis* β/1	*Sus*	98
	Bos	97
	Mus	98
	Equus	97
	Homo	98
	Rattus	98
(g) *Canis* 2M	*Felis*	98
(h)[a] *Rattus* β/1	*Mus* β/1	99
Canis 2M	*Homo* Peri	57
Canis 2M	*Mus* β/1	60
Scallop	*Homo* EO	51
Scallop	*Canis* β/1	54

[a] The highest and the lowest examples of identity among mammalian sarcomeric isoforms are given.

emphasized by low values of identity, e.g., 75-76% if α-cardiac isoforms are compared with fast/developmental, and becomes particularly evident for MyHC 2M. For example, dog MyHC 2M shows identity values between 57 and 60% with paralogous dog MyHC isoforms, just slightly higher than those of scallop MyHC, which is used as the outgroup (Table 2).

MyHC isoform comparison based on phylogenetic trees

The reconstruction of phylogenetic trees was used as a simple and effective approach to show sequence relationships graphically. Three distinct phylogenetic trees were generated from multiple alignments of amino acid sequences taking into account the whole molecule, head, and rod (see Fig. 2). The junction between head and rod was set at proline 842 at the end of the MyLC binding region (see below). We had to face a few limitations: two isoforms, canine MyHC α-cardiac and murine MyHC 2A, could be not used for the whole molecule comparison as the sequences are not complete. Canine α-cardiac lacked of a portion of the head and was used only for rod analysis, whereas murine 2A, lacking of a portion of the rod, was used only for head analysis. Human MyHC 2M could not be used either, as only a fragment corresponding to the tail is available. Interestingly, the alignment of the tails of the three MyHC 2M orthologs (human, canine, and feline, not shown) revealed that the human isoform has a large deletion of 29 aa starting at 1007 in addition to the mutation at the C terminus described by Stedman and co-workers [7].

Despite these limitations, the number of sarcomeric MyHC isoforms analyzed here is higher than in any previous study (compare, for example, with Qin et al. [4]). With regard to the newly sequenced canine MyHC isoforms, the phylogenetic trees confirmed their positions within each specific group of orthologous MyHC isoforms. From a more general point of view the phylogenetic trees confirmed the classification into three subfamilies, i.e., fast/developmental, cardiac, and masticatory, previously described [4]. Some details deserve comments. MyHC EO isoforms form a well-separated group within the fast/developmental subfamily when both the whole molecule and the head are considered. When the tail sequences are compared, however, the MyHC EO isoforms are organized in a completely separate cluster, suggesting a distinct evolution of the tail. Emb, Peri, 2A, 2X, and 2B isoforms form groups reliably separated (reliability 97-100%) when the whole sequences are compared. The difference becomes less defined when only the head or the tail sequences are compared. In particular, porcine MyHC 2B cannot be separated from the 2X isoforms in the tree of the head domain (Fig. 2B).

Sequence comparison in specific MyHC regions

As stated in the introduction, the availability of the sequences of the orthologous sarcomeric MyHC isoforms in several mammalian species offers a unique chance to identify the structural details that characterize each group of orthologs. To this aim, comparisons were carried out for the regions that are considered relevant to determine the functional diversity among isoforms in view of their greater variability and the effects of mutation experiments (see above, introduction). Previous studies [9,16,24] have shown that the surface loops are very diverse among myosin paralogs but almost identical among myosin orthologs. It should be thus possible to identify which motifs are typical of each group of orthologs and represent a kind of "molecular signature" of each group. In this study, the sequences of 38 MyHC isoforms (4 fast 2B, 8 fast 2X, 6 fast 2A, 2 Peri, 3 Emb, 3 EO, 3 (or 4) α-cardiac, 7 β/1, and 2 2M isoforms) were aligned to compare the three superficial loops indicated as loop 1 at the junction between the 25 K and the 50 K domains, loop 2 at the junction between the 50 K and the 20 K domains, and loop 3, also called the secondary actin-binding loop. Further comparisons were carried out between the sequences of the converter domain, the MyLC binding region, and the proximal region of the S2 segment.

Amino acid numbering can create some difficulties as differences appear between MyHC isoforms a few residues after the methionine at the N terminus (M1). Thus, to avoid ambiguity the sequence of human MyHC 2B was chosen as a reference in residue numbering and it is specified in the text when other references were used.

Loop 1

The alignments of the residues of loop 1 are shown in Fig. 3A. The loop starts with a negatively charged residue, E204, in human MyHC 2B. If the end is placed at G217, the length varies between 14 aa in 2B, 2X, 2A, and Peri isoforms; 13 aa in EO and α-cardiac isoforms; 12 in Emb and β/1; and only 8 aa in MyHC 2M. A common motif (**E***KKK***EE**), characterized by 1 negatively charged residue (E204) followed by 3 positively charged residues (K) and 2 negatively charged (E208 and E209), can be recognized at the N/5′ side of the loop, whereas a single positively charged residue (K214) marks the C/3′ side of the loop (motif *K*MQG, see scheme in Fig. 3B). The central part of the loop is composed of 4 residues in the three adult fast isoforms (2B, 2X, 2A) and in MyHC-Peri. Glycine is the last of the 4 residues in virtually all isoforms, except Emb and α-cardiac, whereas a proline is consistently present in 2B and EO MyHC isoforms and occasionally in 2A, 2X, and developmental isoforms. MyHC EO is closely similar to fast adult isoforms but 1 of the 2 acid residues is missing (209) and proline is accompanied by the hydrophilic glutamine (Q). MyHC Emb shows further diversity as, in addition to the lack of the second acidic residue at 209, the C/3′-side motif is identical to cardiac and 2M isoforms (*K*G). An additional basic residue K is present in the central and more variable part of loop1, at 212. Thus, in this isoform the electrostatic balance is clearly shifted toward the positive charges (5:2).

The loop 1 sequences of the two "cardiac" MyHC isoforms are rather similar to each other. At the N end of the loop, a motif similar to that present in fast isoforms is conserved (**D***R*x*KK***D**), with the three positively charged residues separated by a residue in position 205 and preceded and followed by a negative charge (Fig. 3B). The C/3′-side motif is characterized by the positive residue K (the *K*G

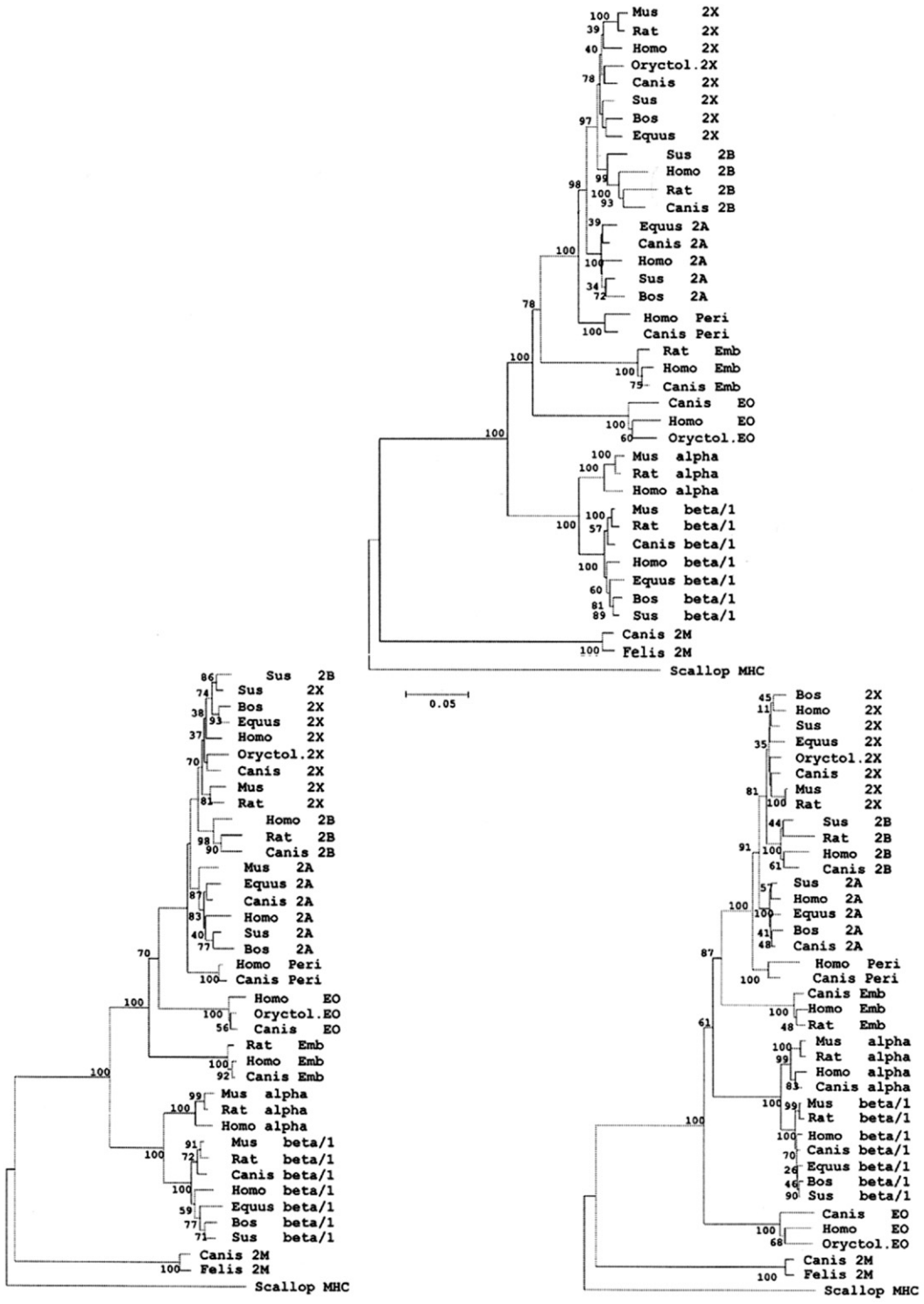

Fig. 2. Phylogenetic trees for (top) the whole MyHC molecule, (bottom left) the head domain, and (bottom right) the tail domain. Numbers are bootstrapping values, the bars indicate phylogenetic distance.

motif), whereas the central part is variable (from five to four residues, respectively, in α and β/1 isoforms) containing hydrophobic amino acids and in several animal species also proline.

Loop 1 of MyHC 2M (Fig. 3B) is markedly different from any other isoform, in the first place because it is much shorter, with six residues missing. In addition, the initial acidic residue (corresponding to 205 in MyHC 2B) is missing, two serine residues replace the lysines followed by a negatively charged D (at 209 in 2M), the central part is missing, but a positively charged amino acid (K212) is still present at the C side of the loop, forming a KG motif. Since MyHC 2M is considered the oldest isoform from the phylogenetic point of view [4] we can hypothesize that through phylogenesis the original sequence *K*QSS**D**G*K*G was modified by insertion of residues particularly

at the N side of the loop. Compared with canine and feline orthologs, the human MyHC 2M (sequence from [8]) shows an additional positively charged K to replace G211 and a doublet TT to replace S208 and S209 (numbering in the MyHC 2M sequence).

Length, flexibility, amount, and distribution of charges are likely essential for the function of the loop 1, which, in view of its position at the opening of the catalytic site, has been related to control of ADP release [15]. Actually, striking correlations between ADP release rate and loop 1 size have been demonstrated comparing smooth muscle MyHC isoforms [25] and studying chimeric MyHC with variable loop 1 [18]. Reduction of flexibility of the loop and reversal (but not removal) of its net positive charge have been shown to result in slowing the ADP release rate [18]. The functional importance of the positively charged residues in loop1 has been attributed to a possible interaction with other charged residues located in front of loop 1 at the opening of the nucleotide binding site [26]. The present comparison between sarcomeric MyHC isoforms shows that the loop 1 length increases from 12 residues in β/1 to 14 residues in MyHC 2B. Such increase in size might be relevant for the filament sliding velocity and ADP release rate, which also increase progressively from slow to 2B myosin [27]. If positive charges are relevant [18], the presence of three negative charges should, however, set a limit to high values of ADP release rate in fast MyHC isoforms compared to cardiac MyHC isoforms, which have only two negative charges. The few available data on functional properties of 2M myosin suggest that MyHC 2M should be closer to a fast than to a slow isoform [12]. To be consistent with the interpretation given above, we can hypothesize that the effect of the small size of loop 1 (only 8 residues) might be compensated for by the lack of proline and the presence of only one negative charge.

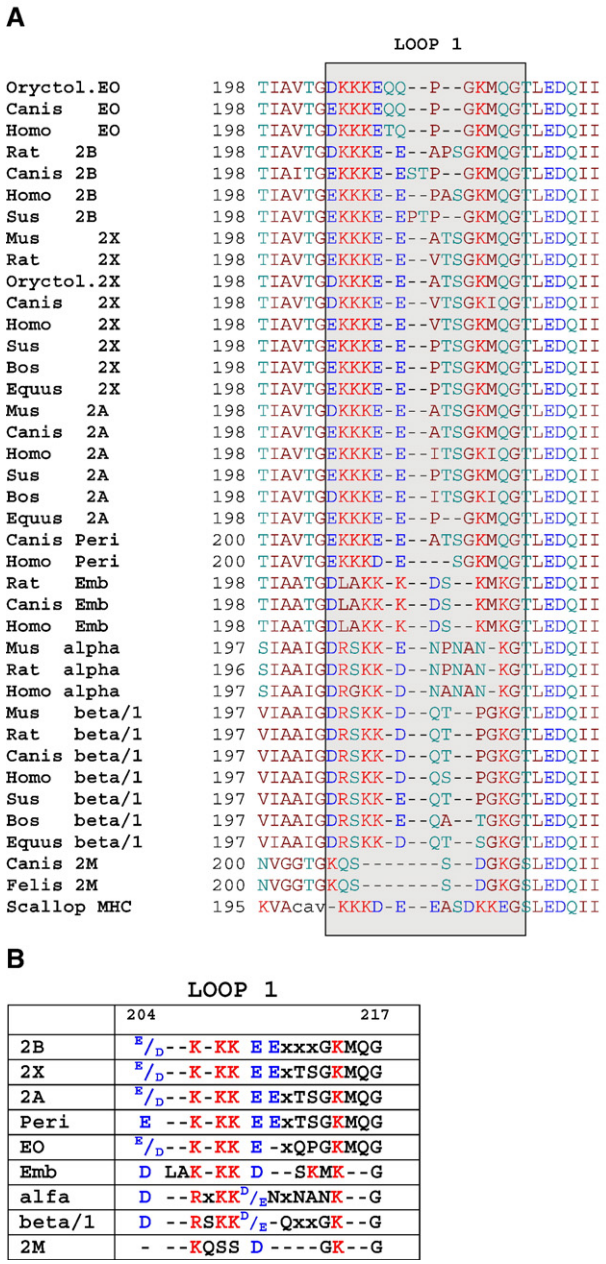

Fig. 3. Loop 1 structure. 38 sequences are aligned in panel A and the scheme of the specific motif is shown in panel B. The color code is red, basic amino acids (H, K, R); dark red, nonpolar amino acids (A, C, G, I, L, M, P, V); teal, uncharged polar amino acids (N, Q, S, T); blue, acidic amino acids (D, E); dark yellow, aromatic amino acids (Y, W, F). Please note that (i) only uppercase letters are considered to be aligned, (ii) the residue numbers refer to human MyHC 2B, (iii) x denotes variable residues, (iv) P is variable in 2X and 2A, and (v) equine 2A does not show the TS doublet present in all 2A and 2X isoforms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Loop 3

Loop 3 starts at K568 in fast 2B MyHC isoforms and is characterized by the marked predominance of the positively charged residues and the presence of proline (see Fig. 4). A motif with 4 lysines and 1 glutamate, *KPK*xx*KGK*(x)x**E**A*H,* is present in all MyHC isoforms of the cluster of fast/developmental isoforms and can be found also in MyHC 2M. The loop size is constant in all isoforms, with the exception of MyHC 2M, which has an additional residue (G573 for the MyHC 2M). No information is available about loop 3 in human MyHC 2M. The two cardiac MyHC isoforms show partial replacement of the lysines with arginine, with no changes in charge. Whereas the first proline at 569 is present in all isoforms (see Fig. 4B), the second proline (P571) is present only in 2B, 2X, and EO and the third proline is present only in a few isoforms (one of four 2B, two of eight 2X, one of six 2A, and four of seven β/1, see Fig. 4B). Thus, loop 3 is highly conserved between MyHC isoforms expressed in mammalian striated muscles. Apparently, the size, the positive charge, and the flexibility make loop 3 of sarcomeric MyHC optimal for interaction with actin, more precisely with residues 1-28 of a second actin molecule [28]. The number and the locations of the proline residues are the only substantial diversity among isoforms. In particular, the change in loop flexibility caused by P571 present only in the isoforms that generate a higher speed of filament sliding may have some functional relevance. It is furthermore remarkable that scallop loop 3 does not align with the corresponding mammalian sarcomeric peptide. The phylogenetic tree relative to loop 3 (Supplementary Fig. 4) shows that the separations

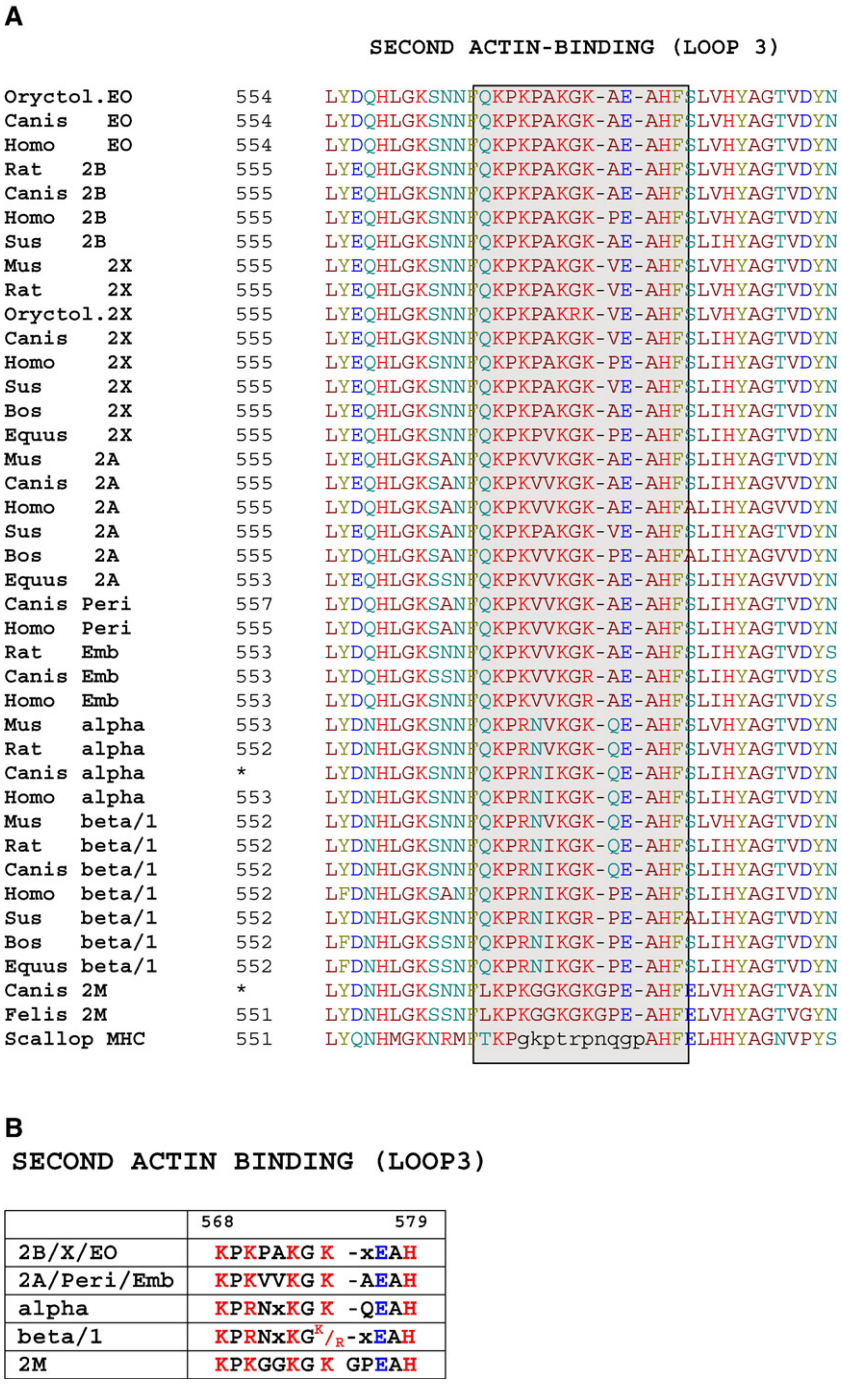

Fig. 4. Structure of the second actin-binding loop. 38 sequences are aligned in panel A and the scheme of the specific motif is shown in panel B. The color code is red, basic amino acids (H, K, R); dark red, nonpolar amino acids (A, C, G, I, L, M, P, V); teal, uncharged polar amino acids (N, Q, S, T); blue, acidic amino acids (D, E); dark yellow, aromatic amino acids (Y, W, F). Please note that (i) only uppercase letters are considered to be aligned, (ii) the residue numbers in B refer to human MyHC 2B, and (iii) x denotes variable residues. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

between MyHC 2M, cardiac isoforms, and the fast/developmental group are significant.

Loop 2

The loop 2 structure is shown in Fig. 5. In human MyHC 2B, which is taken for residue numbering, loop 2 starts on the N-terminal side at residue G626 after the conserved doublet LF (623 and 624) and ends on the C-terminal side with the conserved peptide GSSF (F648). The length of the loop is thus 23 aa for MyHC 2B and varies between 22 and 26 aa for the isoforms of the cluster on chromosome 5. The length is 23 aa for MyHC α and 22 for MyHC β/1 and reaches the shortest value for MyHC 2M, with only 17 residues. A more detailed comparison inside the cluster of fast/developmental isoforms reveals that in the Emb and Peri isoforms loop 2 is consistently

22 aa long, whereas in EO it is 24 aa long. The interspecies and interisoform variability is higher among fast 2A, 2X, and 2B isoforms. In particular, the sequences of murine MyHC 2A and MyHC 2X show an insertion of 3 residues (SGG) in the central part of the loop, which thus becomes 26 aa long. The size of loop 2 in human MyHC 2A is controversial, with either 25 [3]or 23 residues [8] (shown in Fig. 5A) due to an insertion of 2 aa in the central part of the loop. The uncertainty about loop size (human MyHC 2A) and the lack of alignment (murine MyHC-2A and canine MyHC-2B, see Fig. 5A) suggest that further validation is needed before this variability can be considered of biological/functional relevance.

A

LOOP 2

```
Oryctol.EO      620 SFLFSNY--AGAEAGD----SGGSKKGGKKKGSSFQTVSAVFREN
Canis   EO      620 SFLFSNY--AGAEAGD----SGGSKKGGKKKGSSFQTVSAVFREN
Homo    EO      620 SFLFSNY--AGAETGD----SGGSKKGGKKKGSSFQTVSAVFREN
Rat   2B        621 AFLFSGGQAAE---AE----GGGGKKGGKKKGSSFQTVSALFREN
Canis 2B        621 ALLFAGGQSAE---AE----SGGGKKGGKKKGSSFQTVSALFREN
Homo  2B        621 AFLFSGAQTAE---AE----GGGGKKGGKKKGSSFQTVSALFREN
Sus   2B        621 AFLFAER--QSS---E----EGGTKKGGKKKGSSFQTVSALFREN
Mus     2X      621 AYLFSGA--AAA--AEAeSGGGGGKKGAKKKGSSFQTVSALFREN
Rat     2X      621 AYLFSGA--AAA--AEA---GGGGKKGAKKKGSSFQTVSALFREN
Oryctol.2X      621 AFLFTGTAAAE---AE----G-GGKKGGKKKGSSFQTVSALFREN
Canis   2X      621 ANLFSGATAAE---AE----AGGGKKGGKKKGSSFQTVSALFREN
Homo    2X      621 ALLFVGATGAE---AE----AGGGKKGGKKKGSSFQTVSALFREN
Sus     2X      621 AFLFTGAAGAD---AE----AGGGKKGGKKKGSSFQTVSALFREN
Bos     2X      621 ALLFSGP--AS---GEA---EGGPKKGGKKKGSSFQTVSALFREN
Equus   2X      621 ALLFSGP--ASA---DA---EAGGKKGGKKKGSSFQTVSALFREN
Mus    2A       621 AYLFSGAQTAE---AEA-SSGGAAKKGAKKKGSSFQTVSALFREN
Canis  2A       621 AYLFSGAQTAE---AEA---SGGAKKGGKKKGSSFQTVSALFREN
Homo   2A       621 AQLFSGAQTAE---GE----GGGAKKGGKKKGSSFQTVSALFREN
Sus    2A       621 AFLFSGAQTGE---AE----AGGTKKGGKKKGSSFQTVSALFREN
Bos    2A       621 AFLFSGTPTGD---SEA---SGGTKKGGKKKGSSFQTVSALFREN
Equus  2A       619 ALLFSGAQTAD---AE----AGGVKKGGKKKGSSFQTVSALFREN
Canis Peri      623 ASLFSTY--ASA---EA---DSGAKKGAKKKGSSFQTVSALFRE
Homo  Peri      621 ASLFSTY--ASA---EA---DSSAKKGAKKKGSSFQTVSALFREN
Rat   Emb       619 AHLYATF--ATT---DA---DGGKKKVAKKKGSSFQTVSALFREN
Canis Emb       619 AHLYATF--ATA---DA---DSGKKKVAKKKGSSFQTVSALFREN
Homo  Emb       619 AHLYATF--ATA---DA---DSGKKKVAKKKGSSFQTVSALFREN
Mus   alpha     619 ATLFSTY--ASADTG-----DSGKGKGGKKKGSSFQTVSALHREN
Rat   alpha     618 ATLFSTY--ASADTG-----DSGKGKGGKKKGSSFQTVSALHREN
Canis alpha     *   ATLFSSY--ASAD-G-----DSSKGKGGKKKGSSFQTVSALHREN
Homo  alpha     619 ATLFSSY--ATADTG-----DSGKSKGGKKKGSSFQTVSALHREN
Mus   beta/1    618 SNLFANY--AGADAP------ADKGKGKAKKGSSFQTVSALHREN
Rat   beta/1    618 SNLFANY--AGADAP------VDKGKGKAKKGSSFQTVSALHREN
Canis beta/1    618 SNLFANY--AGADAP------VEKGKGKAKKGSSFQTVSALHREN
Homo  beta/1    618 STLFANY--AGADAP------IEKGKGKAKKGSSFQTVSALHREN
Sus   beta/1    618 SNLFANY--AGADTP------VEKGKGKAKKGSSFQTVSALHREN
Bos   beta/1    618 SSLFANY--AGFDTP------IEKGKGKAKKGSSFQTVSALHREN
Equus beta/1    618 SNLFANY--LGADAP------IEKGKGKAKKGSSFQTVSALHREN
Canis 2M        *   ALLFK-------EEE------APAGSKKQKRGSSFMTVSNFYREQ
Felis 2M        618 ALLFK-------EEE------APAGSKKQKRGSSFMTVSNFYREQ
Scallop MHC     621 LFKAP-------EEP------AGGGKKKKGKSSAFQTISAVHRES
```

B

LOOP 2

	626 631 633 638 648
2B	xxQxx - E A-- Ex-GGxK-KGGK-KKGSSF
2X	Gxxxx(1)$^{E}/_{D}$A-- Ex?xGxK-KGxK-KKGSSF
2A	GxxTx(1)$^{E}/_{D}$A-- Ex?GGxK-KGxK-KKGSSF
Peri	TYASA - E Axx D--SxAK-KGxK-KKGSSF
EO	NYAGA - E xG- Dñ-GGSK-KGGK-KKGSSF
Emb	TFATx - D A-- Dñ-xGKK-KGGK-KKGSSF
alpha	xYAxA - D xG- D---SGKxKGGK-KKGSSF
beta/1	NYxGx - D xPx$^{E}/_{D}$-----KGK-GKAKKGSSF
2M	----- - E --- EE-APAGSK--KQKRGSSF

Fig. 5. Loop 2 structure. 38 sequences are aligned in panel A and the scheme of the specific motif is shown in panel B. A phylogenetic tree of loop 2 is shown in panel C. The color code is red, basic amino acids (H, K, R); dark red, nonpolar amino acids (A, C, G, I, L, M, P, V); teal, uncharged polar amino acids (N, Q, S, T); blue, acidic amino acids (D, E); dark yellow, aromatic amino acids (Y, W, F). Please note that (i) only uppercase letters are considered to be aligned, (ii) the residue numbers refer to human MyHC 2B, and (iii) x denotes variable residues. Some uncertainties that need validation are indicated in the scheme (B). (1) indicates that six residues on the N side of E631 are present only in 2X of *Mus* and *Rattus* and 2A of *Bos*, (?) indicates a residue that needs validation in murine 2A and 2X, and finally porcine 2B shows an additional ER doublet, which also needs validation and makes porcine 2B different from its orthologs (see tree in panel C). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

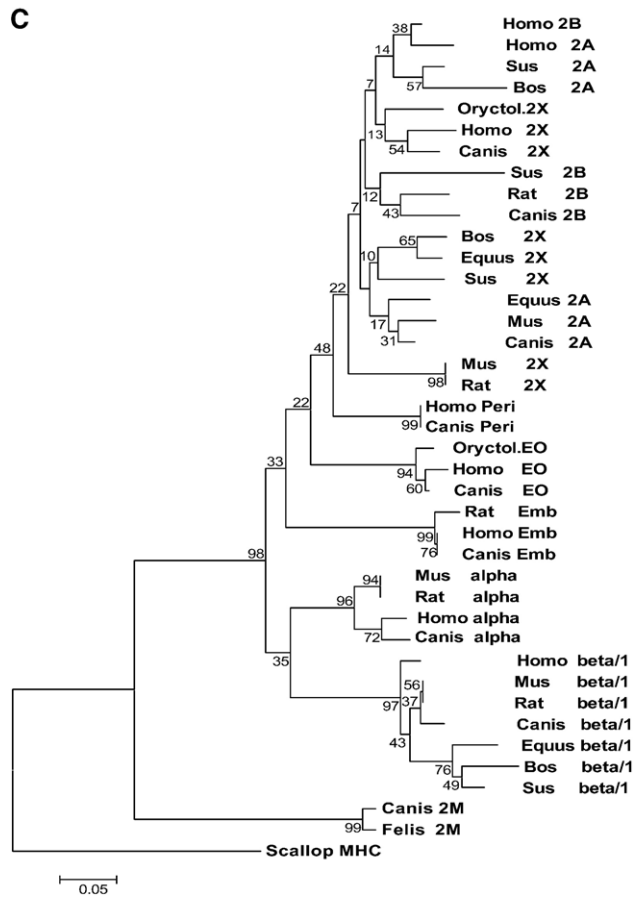

Fig. 5 (*continued*).

The loop can be divided in three portions: the N-terminal part, the central part, and the C-terminal part. Among the three, the central part shows the highest variability. At the N-terminal side of the loop, five residues without electrical charge followed by a negatively charged doublet of glutamate or aspartate (631 and 633) and generally separated by an alanine residue delineate a motif, xxxxx**D**x**E**. The separation between the two negatively charged residues is given by one residue in 2B, 2X, 2A, Peri, and Emb MyHC, becomes two in EO and α-cardiac, and three in Peri and β/1 (thus xxxxx**D**xPx**E**), in which a proline is present between the two negative charges (see Fig. 5B).

The central part of the loop (between E633 and K638) is formed by a number of residues specific for each isoform without any difference between species, i.e., no residue for MyHC- β/1; two for MyHC α-cardiac; three for Peri, Emb, and EO isoforms; and four for MyHC 2B. Longer sequences can be found in some species for MyHC 2A and MyHC 2X, with exceptions represented by murine 2A and 2X (see above). The sequence of human MyHC 2A is still controversial (see above). Interestingly, a lysine is present in MyHC Emb.

Five positively charged lysine residues mark the C side of the loop. In the fast/developmental MyHC isoform group, the C part is organized according to the motif *KK*xx*KKK*GSSF (where xx are either G/V or G/A), very conserved inside orthologous isoform groups, whereas in MyHC α-cardiac the motif is *K*x*K*xx*KKK*GSSF and in MyHC β/1 *K*x*K*x*K*x*KK*GSSF.

MyHC 2M shows a loop 2 sequence (*K***EEE**APAGS*KK*Q-*KR*GSSFM) perfectly identical in the three orthologs available until now (cat and dog and human, not shown in Fig. 5, from [8]), but completely different from any other isoform. The loop is much shorter (only 17 residues), with 5 residues missing at the N side, shows an additional negatively charged residue, and lacks a positively charged residue at the N side, which thus consists of only 4 positively charged residues. A proline residue is present in the central part, which is composed of 5 residues.

Loop 2 is a major site of interaction between myosin and actin. The positively charged lysines are assumed to interact with the six negatively charged residues close to the actin N terminus during the initial docking step of actomyosin interaction [29,30]. The loop is in proximity to the hydrophobic residues P529, M530, I535, F542, and P543, which contribute to the stereospecific interaction with A141, I341, I345, and L349 on the actin surface [30]. Changes in loop size, charge, and flexibility have been shown to influence ATPase activity with little or no effect on filament sliding velocity [17]. If isoforms are ordered with increasing ATPase activity from MyHC β/1 to α-cardiac, 2A, 2X, and 2B [31,32], one can observe that the size of the central part regularly increases in direct relation to the ATPase activity, whereas the size of the N part and the size of the C part decrease. In other words, the negatively charged residues at the N side of the loop are more compact or close to one another and the distance between the negative N side and the positive C side is greater in the fastest (MyHC 2B) than in the slowest isoform (MyHC β/1). Loop 2 of MyHC 2M is very short, but the central part is composed of five residues; thus it is not shorter than in the isoforms with very high enzymatic activity, such as 2B. The size of the separation between the negatively charged N side and the positive side might explain the high ATPase activity of myosin 2M (see [12]).

In view of the relevance of loop 2 for ATPase activity, it is possible that loop 2 structure might be responsible not only for the diversity of ATPase rate between isoforms but also for the specific sensitivity of ATPase activity to alkali or acid preincubation. In support of this hypothesis comes the fact that acid-labile myosin ATPase is characteristic of fast myosin isoforms, which exhibit groups of two to five consecutive glycine residues. The presence of such glycine chains makes the loop unstable and prone to denaturation in acid environment. The structure of the loop in 2M and β/1 isoforms is made stable not only by the lack of the glycine chains but also by the presence of a proline residue in the central part of the loop.

The phylogenetic tree of the loop 2 (Fig. 5C) confirms the significant separation between the subfamilies fast/developmental, cardiac, and 2M. Inside the fast/developmental subfamily Emb, Peri, and EO appear as separate groups, whereas the diversity among 2A, 2X, and 2B is less significant.

Converter domain

Mutation studies in *Drosophila* [19] and analysis of point mutations in hypertrophic cardiomyopathy [33] have revealed how residue changes in the converter domain influence cross-bridge kinetics. This region was thus comparatively examined in

the sarcomeric mammalian MyHC isoforms. The peptide 699–714, which comprises the two active thiols C699 and C709, shows a sequence (CNGVL**E**GIRICR*K*GFP) completely conserved in all isoforms including MyHC 2M (see Supplementary Fig. 5 and Fig. 6). The peptide 714–726 (S*R*IxYx**D**/**E**F*K*/*R*Q*R*Y) is also rather conserved with minor diversity among the three isoform subfamilies, whereas more important diversity can be seen in the next peptide (726–759). In the fast/developmental group the peptide 726–759 shows the structure *K*/*R*xLNASAI-P**E**GQxI**D**S*K*xAx**E***K*LLxSI**D**x**D***H*/*R*xQ with some differences between MyHC EO, which has an additional negative charge and lacks a positive charge, and the other members of the group. In the two cardiac isoforms the triplet ASA (731–733) is replaced by PAA, thus adding a second proline to that present in all isoforms (P735, see Fig. 6). The diversity becomes even more pronounced when the 726–759 peptide is compared with the corresponding peptide in MyHC 2M. The peptide is one residue shorter, lacks three charged residues (two positive and one negative), and contains an additional proline like the cardiac isoforms. On the whole, a contribution of the differences in the converter domain to the functional specificity of myosin 2M cannot be excluded. The next peptide, 760–780 (Y*K*/*R*F/IG*H*T*K*VFF*K*AGLLG*L*/*T*L**EE**M, see Fig. 6), is very conserved with a few minor residue changes.

MyLC binding regions

The preferential association between MyHC and MyLC isoforms (see for a review [2]) is well known and might find an explanation either in isoform coexpression, and thus in transcriptional regulation, or in a preferential binding between MyHC and MyLC isoforms. To contribute to the answer to this question, the comparison between the binding regions of various MyHC isoforms can be of interest. The long α helix (Supplementary Fig. 6), which contains the binding sites for essential and regulatory MyLC, spans from the methionine (M780) located at the end of the converter and followed by the positively-negatively charged residues *R***DE***K* or *R***DD***K* to the end of the myosin head marked by the proline (P842), which determines a clear turn of direction of the helix. Fig. 7A shows the results of the alignment of the sequences of the 38 MyHC isoforms. A consensus motif (792–802), xQxxx*R*Gxxx*R*, which can match the essential light chain as well as calmodulin and is characterized by hydrophobic and positively charged residues, can be identified in all MyHC isoforms and can be extended to *K*xxQxxx*R*GxLx*R*x**E**xx*K*/*R* if only the 2B, 2X, and Peri isoforms are considered. Diversity can be found outside the consensus motif, as additional positive charges at 787 are present on the N side of the consensus motif in the Emb, β/1, α-cardiac, and 2M isoforms and on the C side at 806 in the 2B, EO, α-cardiac, and 2M isoforms. A third additional positive charge is present inside the consensus motif at 795 in 2M isoforms, in many 2A isoforms, and in some 2X isoforms (see Supplementary Fig. 6). Thus, cardiac isoforms have two additional positive residues flanking the consensus motif, whereas among fast/developmental isoforms an additional positive charge is present at 787 in Emb and 806 in 2B and EO. In MyHC 2M the positive charge at 787 corresponds to the lack of it at 782 (numbering of MyHC 2M).

The RLC binding region starts at the N side with the highly charged peptide **E***RR***D** or **E***RR***E,** with the exception of only Peri, Emb, and 2M, and ends at the C side with the conserved sequence *K*I/V*K*P (see Fig. 7B). A consensus motif, IQxxx *R*xxxxx*K*/*R*, for binding of the regulatory MyLC can be recognized in all MyHC isoforms. Diversity can be found in the fast/developmental family with EO lacking one negative charge at 835. Cardiac and 2M isoforms lack a positively charged residue (830). MyHC 2M is clearly different, showing more positively charged residues and lacking proline 832.

In conclusion, the MyLC binding region appears to be very conserved among isoform groups, although minor differences in number and positions of positively charged residues are present. It is possible that such diversity in charged and hydrophobic residues contributes to preferential binding of one MyLC isoform over another.

Proximal S2 segment

Proline 842 represents the transition between the coiled-coil structure and the separated heads. The angle of about 30° generated by the proline residue is well above the angle of 11° compatible with the coiled-coil structure [34]. A direct involvement of the S2 region in the force generation by myosin has been considered in several studies [35,36]. The functional relevance is further suggested by the impact of the mutations of MyBP-C, which binds to this region of the myosin molecule, and by mutations in the S2 fragment that might alter the interaction with the regulatory MyBP-C motif [37]. Some diversity among isoforms can be found in the peptide comprised between P842 and the sequence *K***EL****EE***K**MV (starting at

	699 714 760 780
	CNGVLEGIRICRKGFPYKFGHTK ↓ YK/RF/IGHTKVFFKAGLLGL/TLEEM
2B,2X,2A	715 SRILYADFKQRYKVLNASAIPEGQxIDSKKASEKLLxSIDxDHTQ 759
Peri	SRILYGDFKQRYKVLNASAIPEGQFIDSKKASEKLLASIDIDHTQ
Emb	SRILYGDFKQRYRVLNASAIPEGQFIDSKKACEKLLASIDIDHTQ
EO	SRILYADFKQRYRILNASAIPEGQFIDSKNASEKLLxSIDVDREQ
alpha,beta/1	NRILYGDFRQRYRILNPAAIPEGQFIDSRKGAEKLLxSLDIDHNQ
2M	NRMQYPEFKQRYQVLNPNVIPQG-FVDNKKASELLLGSIDLDVNE

Fig. 6. Scheme showing the primary structure of the converter region in sarcomeric MyHC isoforms. Color code is as in Figs. 3, 4, and 5. x denotes variable residues. Note that the variable part 715–759 is flanked by two peptides (699–714 and 760–780) completely conserved in all isoforms.

A

	780 xQxxxRGxxxR 808
2B	MRD D KLAQLITRTQAxCRGxLMRVEFRKM
2EO	MRD E KLVTLMTRTQAxCRGYLMRVEFKKM
2X	MRD^{D}/$_{E}$KLAQLITRTQAxCRGxLARVEYQKM
2A	MRD^{D}/$_{E}$KLAQxITRTQAxCRGFLARVEYQKM
Peri	MRD E KLxQIITRTQAVCRGFLMRVEYQKM
Emb	MRD^{D}/$_{E}$RLAKLITRTQAVCRGFLMRVEFQKM
alpha	MRD E RLSRIITRIQAQARGQLMRIEFKKx
beta/1	MRD E RLSRIITRIQAQSRGVLxRMEFKKL
2M	MRD^{D}/$_{Q}$RLAKIMTMLQCRLRGFLMRIEFKKM

B

	809 IQxxxRxxxxxK 842
2EO	MERR^{D}/$_{E}$S IFCIQYN I RxFMNVKHWPWMNLFFKIKP
2B, X, A	VERR E S IFCIQYN^{V}/$_{I}$RAFMNVKHWPWMKLYFKIKP
Peri	LQRR E A LFCIQYN^{V}/$_{I}$RAFMNVKHWPWMKLFFKIKP
Emb	VQRR E S IFCIQYN I RAFMNVKHWPWMKLFFKIKP
Cardiac	LERR D^{A}/sLLIIQWN I RAFMGVKNWPWMKLYFKIKP
2M	LERR I G LKVIQRN T RKFLELRFWGWWKLYNKVKP

Fig. 7. Schemes showing the structures of (A) the ELC binding region and (B) the RLC binding region in sarcomeric mammalian MyHC isoforms. Color code as in Figs. 3, 4, and 5. x denotes variable residue.

K875), which is strictly conserved in all isoforms. The fast group of isoforms is rather homogeneous (see Fig. 8), with the exception of residue 861 at which place a negatively charged residue is present in 2B, 2X, and EO but not in 2A, Peri, and Emb. An additional negatively charged residue (at 867) is generally lacking in MyHC 2X. In cardiac isoforms the negative charge at 861 is also absent and the negative charge at 868 is shifted toward the C end. The difference between 2M and all other isoform is very pronounced as the balance between negative and positive charges (10:10 or 11:10) appears definitely shifted toward the negative sign (10:8) and clusters of two or three negative charges appear. Both the stability of the α-helix structure and the binding with MyBP-C might be affected by such diversity.

Conclusions

This study provides the first complete sequences of all canine sarcomeric MyHC isoforms and makes use of the comparison between the newly obtained set of the sequences with sequences already published to identify structural features specific to various groups of MyHC isoforms. The identification of nine sarcomeric MyHC sequences makes the dog the third species best known after humans and mice.

As mentioned in the introduction, the interest in the canine MyHC sequences is high for scientists interested in muscle physiology and physiopathology as well as for veterinarians who study dog diseases and for zoologists interested in the large variability among canine strains. Even sport science is interested in dog muscles, and the availability of the sequences offers a straightforward and reliable way to study muscle specialization and adaptation.

In our view, however, the results obtained from the alignment of 38 MyHC isoforms are not less important. The number of isoforms considered goes beyond that considered in previous investigations and allows the identification of the peptides that form the basis for structural diversity among isoforms of sarcomeric MyHC. With the exception of MyHC 2M, the sarcomeric MyHC form a very homogeneous group with percentage of identical amino acids above 90% when orthologs are compared and above 70% when paralogs are compared. The connection between the structural diversity and the functional diversity is only speculative unless mutation experiments and creation of chimeric myosins are performed. The present results,

	843 861 882
2B	LLKSAETEKEMANMKE^{D}/$_{E}$FEKxKE^{D}/$_{E}$LAKxEAKRKELEEKMV
2X	LLKSAETEKEMANMKE E FEKTKE^{X}/$_{E}$LAKxEAKRKELEEKMV
2A,Peri,Emb	LLKSAETEKEMATMKE E FQKTKD E LAKSEAKRKELEEKMV
2EO	LLKSAEAEKEMATMKE D FERxKE E LARSEARRKELEEKMV
cardiac	LLKSAETEKEMAxMKE E FxRxKD A LEKSEARRKELEEKMV
2M	LLNVARQEEEMKAKEE E LRNAMx K TQELISRVKELEEKMV

Fig. 8. Scheme showing the structure of the proximal part of the S2 segment in mammalian sarcomeric MyHC isoforms. Color code as in Figs. 3, 4, and 5. x denotes variable residue.

however, form in our opinion the essential basis to design such experimental studies on structure-function relations inside the highly specialized family of sarcomeric MyHC isoforms.

Methods

Nucleotide sequences were identified and amino acid sequences were deduced using NCBI databases and tools. The dog genome database (http://www.ncbi.nlm.nih.gov/projects/genome/guide/dog/) contains either fully sequenced contigs (from ~1 kb to several hundreds of kilobases) or assembled supercontigs of several dozen megabases, interrupted by short gaps. The BLAST programs [21] were used for database searching against the most recent genome assembly (CanFam2.0, May 2005).

Human cDNA sequences of MyHC (see Table 1) were the query sequences. After BLASTN searches, the best scoring alignment for each MyHC isoform identified the corresponding dog gene. At the same time, information about the exact chromosomal position in the dog genome was retrieved. The genes have been fully annotated to identify exons, introns, intron-exon boundaries, and the 5′ and 3′ untranslated regions. Intron sequences were removed to obtain uninterrupted coding sequences (CDS).

The assembled CDS for each dog MyHC isoform were translated by means of the BCM Search Launcher Sequence Utilities software (http://searchlauncher.bcm.tmc.edu/seq-util/seq-util.html) to predict the corresponding MyHC protein. Multiple protein alignments were generated by means of the DiAlign Algorithm (http://www.genomatix.de/products/index.html) from all the complete mammals MyHC isoforms available in the NCBI GenBank database (see Table 1) and the dog protein sequences obtained with this procedure. In addition to whole protein alignment, head and rod domains of MyHC were aligned separately and analyzed. The most variable functional sites of the myosin head—loop 1, loop 2, loop 3, the converter domain, MLC binding regions, and the proximal S2 segment—were also aligned separately.

Phylogenetic tree reconstructions were carried out using MEGA software version 3.1 [22]. The tree relationships were reconstructed by the neighbor-joining methods and the genetic distance was calculated by *p* distance. The bootstrap test was performed with 500 replications. The trees were based on the protein sequence alignment of the mammal MyHC isoforms using scallop MyHC as the outgroup. For the phylogenetic study only complete amino acid MyHC sequences were considered, i.e., 37 MyHC sequences from eight distinct species. Two isoforms, murine fast 2A and canine α-cardiac, were only partially available and were only occasionally utilized for comparison.

Acknowledgment

This work was partially supported by MIUR (Ministero Istruzione Università Ricerca) via PRIN (Progetto di Ricerca di Interesse Nazionale) 2004.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2006.08.004.

References

[1] J.S. Berg, B.C. Powell, S.E. Cheney, A millennium myosin census, Mol. Biol. Cell 12 (2001) 780–794.
[2] S. Schiaffino, C. Reggiani, Molecular diversity of myofibrillar proteins: gene regulation and functional significance, Physiol. Rev. 76 (1996) 371–423.
[3] A. Weiss, S. Schiaffino, L. Leinwand, Comparative sequence analysis of the complete human sarcomeric myosin heavy chain family: implications for functional diversity, J. Mol. Biol. 290 (1999) 61–75.
[4] H. Qin, M.K.H. Hsu, B.J. Morris, J.F.Y. Hoh, A distinct subclass of mammalian striated myosins: structure and molecular evolution of 'superfast' or masticatory myosin heavy chain, J. Mol. Evol. 55 (2002) 544–552.
[5] J.F.Y. Hoh, H. Qin, M. Hsu, J. Yu, H. Eyre, G. Sutherland, Cloning of the full length cat superfast myosin heavy chain cDNA and mapping to its human homologue to chromosome 7q22, J. Muscle Res. Cell Motil. 20 (1999) 80–87.
[6] A. Rowlerson, J. Pope, J. Murray, R.B. Whalen, A.G. Weeds, A novel myosin present in cat jaw-closing muscles, J. Muscle Res. Cell Motil. 2 (1981) 415–438.
[7] H.H. Stedman, B.W. Kozyak, A. Nelson, D.M. Thesier, L.T. Su, D.W. Low, C.R. Bridges, J.B. Shrager, N. Minugh-Purvis, M.A. Mitchell, Myosin gene mutation correlates with anatomical changes in the human lineage, Nature 428 (2004) 415–418.
[8] P.R. Desjardins, J.M. Burkman, J.B. Shrager, L.A. Allmond, H.H. Stedman, Evolutionary implications of three novel members of the human sarcomeric myosin heavy chain gene family, Mol. Biol. Evol. 19 (2002) 375–393.
[9] M.A. Pellegrino, M. Canepari, G. D'Antona, C. Reggiani, R. Bottinelli, Orthologous myosin isoforms and scaling of shortening velocity with body size in mouse, rat, rabbit and human muscles, J. Physiol. 546 (2003) 677–689.
[10] L. Toniolo, M. Patruno, L. Maccatrozzo, M.A. Pellegrino, M. Canepari, R. Rossi, et al., Fast fibres in a large animal: fibre types, contractile properties and MHC expression in pig skeletal muscles, J. Exp. Biol. 207 (2004) 1875–1886.
[11] J.B. Shrager, P.R. Desjardins, J.M. Burkman, S.K. Konig, S.K. Stewart, L. Su, et al., Human skeletal myosin heavy chain genes are tightly linked in the order embryonic-IIa-IId/x-ILb-perinatal-extraocular, J. Muscle Res. Cell Motil. 21 (2000) 345–355.
[12] J.F.Y. Hoh, 'Superfast' or masticatory myosin and the evolution of jaw-closing muscles of vertebrates, J. Exp. Biol. 205 (2002) 2203–2210.
[13] G.J.M. Stienen, J. Kiers, R. Bottinelli, C. Reggiani, Myofibrillar ATPase activity in skinned human skeletal muscle fibres: fibre type and temperature dependence, J. Physiol. 493 (1996) 299–307.
[14] Y. Saeki, C. Kato, M. Satomi, K. Yanagisawa, ATPase activity and tension development in mechanically-skinned feline jaw muscle, Arch. Oral Biol. 32 (1987) 207–210.
[15] J.A. Spudich, How a molecular motor works, Nature 372 (1994) 515–518.
[16] H.V. Goodson, H.M. Warrick, J.A. Spudich, Specialized conservation of surface loops of myosin: evidence that loops are involved in determining functional characteristics, J. Mol. Biol. 287 (1999) 173–185.
[17] T.Q. Uyeda, K.M. Ruppel, J.A. Spudich, Enzymatic activities correlate with chimaeric substitutions at the actin-binding face of myosin, Nature 368 (1994) 567–569.
[18] H.L. Sweeney, S.S. Rosenfeld, F. Brown, L. Faust, J. Smith, J. Xing, et al., Kinetic tuning of myosin via a flexible loop adjacent to nucleotide binding pocket, J. Biol. Chem. 273 (1998) 6262–6270.
[19] D.M. Swank, A.F. Knowles, J.A. Suggs, F. Sarsoza, A. Lee, D.W. Maughan, S.I. Bernstein, The myosin converter domain modulates muscle performance, Nat. Cell Biol. 4 (2002) 312–317.
[20] K. Lindblad-Toh, C.M. Wade, T.S. Mikkelsen, E.K. Karlsson, D.B. Jaffe, M. Kamal, et al., Genome sequence, comparative analysis and haplotype structure of the domestic dog, Nature 438 (2005) 803–819.
[21] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[22] S. Kumar, K. Tamura, I. Jakobsen, M. Nei, MEGA: molecular evolutionary genetics analysis (2001).
[23] V. Mahdavi, A.P. Chambers, B. Nadal-Ginard, Cardiac alpha- and beta-myosin heavy chain genes are organized in tandem, Proc. Natl. Acad. Sci. USA 81 (1984) 2626–2630.
[24] M. Canepari, R. Rossi, M.A. Pellegrino, R. Bottinelli, S. Schiaffino, C. Reggiani, Functional diversity between orthologous myosins with minimal sequence diversity, J. Muscle Res. Cell Motil. 21 (2000) 375–382.
[25] C.A. Kelley, M. Takahashi, J.H. Yu, R.S. Adelstein, An insert of seven amino acids confers functional differences between smooth muscle myosins from the intestines and vasculature, J. Biol. Chem. 268 (1993) 12848–12854.

[26] J. Sant'Ana Pereira, D. Pavlov, M. Nili, M. Greaser, E. Homsher, R.L. Moss, Kinetic differences in cardiac myosins with identical loop 1 sequences, J. Biol. Chem. 276 (2001) 4409–4415.
[27] S. Weiss, R. Rossi, M.A. Pellegrino, R. Bottinelli, M.A. Geeves, Differing ADP release rates from myosin heavy chain isoforms define the shortening velocity of skeletal muscle fibers, J. Biol. Chem. 276 (2001) 45902–45908.
[28] J. Van Dijk, M. Furch, C. Lafont, D.J. Manstein, P. Chaussepied, Functional characterization of the secondary actin binding site of myosin II, Biochemistry 38 (1999) 15078–15085.
[29] I. Rayment, W.R. Rypniewski, K. Schmidt-Base, R. Smith, D.R. Tomchick, M.M. Benning, et al., Three-dimensional structure of myosin subfragment-1: a molecular motor, Science 261 (1993) 50–58.
[30] I. Rayment, H.M. Holden, M. Whittaker, C.B. Yohn, M. Lorenz, K.C. Holmes, R.A. Milligan, Structure of the actin-myosin complex and its implications for muscle contraction, Science 261 (1993) 58–65.
[31] R. Bottinelli, M. Canepari, C. Reggiani, G.J. Stienen, Myofibrillar ATPase activity during isometric contraction and isomyosin composition in rat single skinned muscle fibres, J. Physiol. (London) 481 (1994) 663–675.
[32] N.R. Alpert, C. Brosseau, A. Federico, M. Krenz, J. Robbins, D.M. Warshaw, Molecular mechanics of mouse cardiac myosin isoforms, Am. J. Physiol. Heart Circ. Physiol. 283 (2002) H1446–H1454.
[33] I. Rayment, H.M. Holden, J.R. Sellers, L. Fananapazir, N.D. Epstein, Proc. Natl. Acad. Sci. USA 92 (1995) 3864–3868.
[34] G. Offer, P.J. Knight, The structure of the head-tail junction of the myosin molecule, J. Mol. Biol. 256 (1996) 407–416.
[35] S.S. Margossian, J.W. Krueger, J.R. Sellers, G. Cuda, J.B. Caulfield, P. Norton, H.S. Slayter, Influence of the cardiac myosin hinge region on contractile activity, Proc. Natl. Acad. Sci. USA 88 (1991) 4941–4945.
[36] H. Sugi, T. Kobayashi, T. Gross, K. Noguchi, T. Karr, W.F. Harrington, Contraction characteristics and ATPase activity of skeletal muscle fibers in the presence of antibody to myosin subfragment 2, Proc. Natl. Acad. Sci. USA 89 (1992) 6134–6137.
[37] M. Gruen, M. Gautel, Mutations in b-myosin S2 that cause familial hypertrophic cardiomyopathy (FHC) abolish the interaction with the regulatory domain of myosin-binding protein-C, J. Mol. Biol. 286 (1999) 933–949.