Acta Biochimica Polonica

Vol. 47 No. 4/2000

1019–1026

QUARTERLY

Motifs of the caldesmon family✪

Edward A. Czuryło✉

Nencki Institute of Experimental Biology, Department of Muscle Biochemistry, Warszawa, Poland

Received: 20 September, 2000; revised: 2 November, 2000; accepted: 15 November, 2000

Key words: caldesmon, multiple sequence alignment, conserved sequences, binding sites, secondary structure prediction

Seven highly conserved regions were found in caldesmon molecules from various sources using the multiple sequence alignment method. Their localization coincides with regions where the binding sites to other proteins were postulated. Less conserved and highly divergent regions of the sequences are described as well. These results could refine the planning of caldesmon gene manipulations and accelerate the precise localization of binding sites in the caldesmon molecule and, as a consequence, this could help to elucidate its function in smooth muscle contraction.

Caldesmon (CaD) was first isolated from chicken gizzard almost twenty years ago (Sobue *et al.,* 1981) as a protein that binds calmodulin in a Ca^{2+}-dependent manner and interacts with F-actin in a Ca^{2+}-independent manner. The colocalization of CaD with actin filaments in the cell (Lehman *et al.,* 1989), induction of G-actin polymerization in solutions (Gałązkiewicz *et al.,* 1985), inhibition of actomyosin ATPase activity and its abolition by the Ca^{2+}-calmodulin complex (see reviews Chalovich & Pfitzer, 1997; Chalovich *et al.,* 1998; Huber, 1997; Marston *et al.,* 1998; Marston & Huber, 1996; Sobue & Sellers, 1991) designate CaD as the thin filament regulatory protein in smooth muscle, the search for which lasted for decades.

The CaD family consists of two subfamilies: smooth muscle CaDs of a higher molecular mass (called H-CaDs) and CaDs with a lower molecular mass (L-CaDs) found in both muscle and non-muscle cells (Sobue & Sellers, 1991). In chicken (Haruna *et al.,* 1993) and human (Hayashi *et al.,* 1992), the alternative splicing of a single caldesmon gene produces different protein isoforms, three in chicken

✪Supported by a grant from the State Committee for Scientific Research to the Nencki Institute.

✉Corresponding address: Department of Muscle Biochemistry, Nencki Institute of Experimental Biology, L. Pasteura 3, 02-093 Warszawa, Poland; fax: (48 22) 822 5342, tel.: (48 22) 659 8571; e-mail: eczurylo@nencki.gov.pl

Abbreviations: CaD, caldesmon; PHD, the computer algorithm Profile network prediction HeiDelberg.

and five in human cells. For other organisms, several complete or partial sequences, predominantly of non-muscle CaDs, have been established.

Most research was carried out on H-CaD isoforms, which are unique proteins, 771- or 793-residues long (for chicken and human isoforms, respectively), containing an about 250-residues long "spacer" insertion beginning around position 200 and absent in L-CaDs. All CaDs are rich in ionisable amino-acid residues (202 acidic and 188 basic for chicken, 191 acidic and 206 basic for human H-CaDs). Thus, on average, every other residue carries a charge causing CaD to interact with other biomolecules, including many proteins, so readily that it is hard to distinguish between specific and nonspecific interactions (Czuryło *et al.,* 1991; Czuryło *et al.,* 1997b; Marston *et al.,* 1998; Marston & Huber, 1996). The large number of charged residues prevents the CaD molecule from folding into a globule and forces its extended conformation (Czuryło *et al.,* 1997a; Czuryło *et al.,* 1993; Mabuchi & Wang, 1991). Chicken H-CaD contains more than twice the number of polar, exposed residues (594) as expected for a globular protein of the same molecular mass (274). This facilitates a structure with the largest surface possible: a rod 64 nm long and 2 nm in diameter in which small globules at the ends and a kink in the middle cannot be excluded (Czuryło *et al.,* 1997a). The most effective secondary structure element for this kind of molecule is a helix, the content of which reaches 51% in H-CaD while β-strand comprises only 9% (Czuryło *et al.,* 1993). The 187 residues of the chicken H-CaD spacer form 52 full helical turns, which is probably one of the longest known single helices in proteins. The extra energy necessary to keep it may be supplied by salt bridges between positively and negatively charged side chains in positions i and i+4 (Wang & Wang, 1996).

The unique shape of the H-CaD molecule affects the results of secondary structure predictions based mostly on the information from globular proteins where secondary structure elements interact with one another, which is impossible in CaD. Prediction results obtained by the eight most recent methods, publicly available on the NPS@ server (Combet *et al.,* 2000) (http://pbil.ibcp.fr), are close enough to the experimental data (Czuryło *et al.,* 1993) only for the PHD method based on the neural network (Rost *et al.,* 1994). Therefore, for the secondary structure prediction of CaDs (Czuryło *et al.,* 1993), the ALB algorithm was used. This algorithm was selected because it was the only one which allows to decide, whether to take into account the interactions between the secondary structure elements, or to ignore them.

Up to now, no exhaustive analysis of multiple sequence alignments of various caldesmons was carried out.

METHODS

The sequences of caldesmon and its fragments were retrieved from publicly available protein and gene databases. The replicates were manually rejected. Multiple sequence alignment was performed with CLUSTAL X (Higgins *et al.,* 1996; Thompson *et al.,* 1994) with manual editing (especially for fragment sequences) with BIOEDIT (Hall, 1999). The latter program was used to produce figures colored according to the CLUSTAL X protein scheme. Sequences of fragments shorter than two motifs were removed from the final figures. The databases and selected records are listed in Fig. 1.

RESULTS AND DISCUSSION

Analysis of the human H-CaD sequence revealed that its spacer contains a previously undetected elevenfold repeat of 16-residues, with the consensus motive EERqRiKxExEEKrAA; here, the small letters denote deletions or non-conservative replacements and x

stands for any residue (Fig. 1a). A revision of the chicken H-CaD sequence revealed that its spacer also contains eleven, not ten (Hayashi *et al.,* 1989; Wang *et al.,* 1991a), 15-residue repeats, with the general motif EEEKKAAEER(ER)AKA in five of which the residues in parentheses are deleted (Fig. 1a). Notice that the conservation among human CaD repeats is lower than for the chicken protein.

The common properties of CaDs are a result of the conserved regions that probably form crucial structural and functional sites. Two common strongly conservative (over 85% identity) or sometimes fully identical motifs, long enough to be functional, can be distinguished in the N-terminal part of all known CaDs sequences (Fig. 1b). We refer to them as the N1 and N2 motifs. The N1 motif starting from position 27 of both human and chicken H-CaD is identical to the 27 residues of the IK29C peptide (Lee *et al.,* 2000; Li *et al.,* 2000), proposed as the binding site for myosin (Wang *et al.,* 1997). The IK29C peptide has two extra residues at its N-terminus and starts from position 25 of human H-CaD. There is no evidence for the importance of these two extra residues for myosin binding. In the middle part of the N1 motif, a strong helical structure encompassing almost half of the residues was predicted (Czuryło *et al.,* 1993); 2/3 of its residues are charged. In all CaD sequences the N1 motifs exhibit a complete identity. In the N2 motif, clusters of charged residues (20) alternate with those of hydrophobic amino acids (11). Helical regions were predicted for both ends of the N2 motif while no structure was predicted for its middle part (Czuryło *et al.,* 1993) (Fig. 1b). No function can yet be assigned to this motif though it might form the N-terminal calmodulin binding site (Lee *et al.,* 2000; Wang, 1988) or the N-terminal tropomyosin binding site (Marston *et al.,* 1998; Marston & Huber, 1996).

The C-terminal part harbors five motifs (C1 to C5, respectively) with highly conservative or even identical sequences (Fig. 1c) arguing for the existence of conserved, fundamental functions. C1 represents the most charged part of the molecule (104 residues, including 61 charged). A helical region was strongly predicted for the N-terminal half of C1 and a β-strand region for its C-terminal part (Czuryło *et al.,* 1993). No function is yet finally assigned to C1, though the tropomyosin binding site between residues 523 and 580 (Lee *et al.,* 2000) and a low-affinity actin binding site somewhere between residues 491 and 584 (Leszyk *et al.,* 1989; Mornet *et al.,* 1988) are suggested (Fig. 2).

The C2 motif (17 residues, including 8 hydrophobic) prefers a β-strand conformation (Czuryło *et al.,* 1993). Together with the hydrophobic C3 motif, it creates the most hydrophobic region in the whole CaD molecule (Fig. 1c). The C3 motif exhibits a complete identity in all CaD sequences. A short helix is expected after the proline-rich region in its N-terminus (Czuryło *et al.,* 1993). Both the C2 and C3 motifs are potential actin binding site(s). Chalovich and co-authors (1992) propose a single site extending between residues L612(653) and F673(722), including the entire C2 and most of the C3 motif, while Marston and co-authors (Marston *et al.,* 1998; Marston & Huber, 1996) suggest the existence of two independent sites for residues G620(662)–Y640(682) and G666(708)–G684(726), which nicely coincides with both motifs (Fig. 2). In addition, the last nine residues of the motif C3 are generally accepted to belong to calmodulin binding site A (Lee *et al.,* 2000; Marston *et al.,* 1998; Marston & Huber, 1996; Wang *et al.,* 1991b).

The C4 motif has 45 residues, only 10 of which are charged. Its N-terminus is glycine-rich and the C-terminus proline-rich. In its first half there are two short helices separated by short β-strands, while the second half contains mostly turns and loops, and only the few last residues tend to form a helix extending through most of the C5 motif (Czuryło *et al.,* 1993). The C5 motif is 12 residues long, half of which are charged (Fig. 1c). The middle part of the C4 motif coincides with the generally

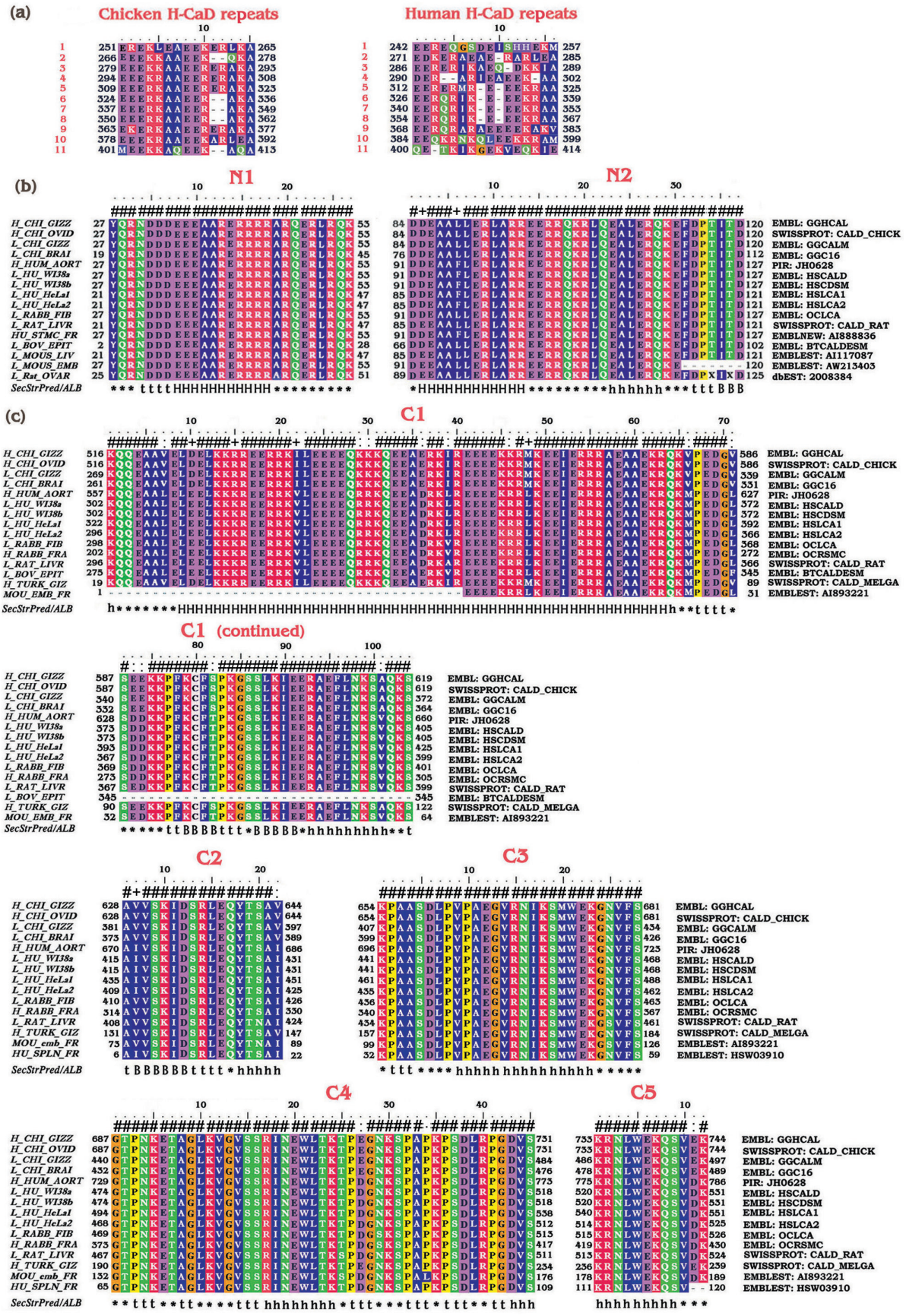

Figure 1. Multiple alignment of the conserved motifs of caldesmons found in the publicly available sequence databases. Legend to Fig. 1. continued on the next page.

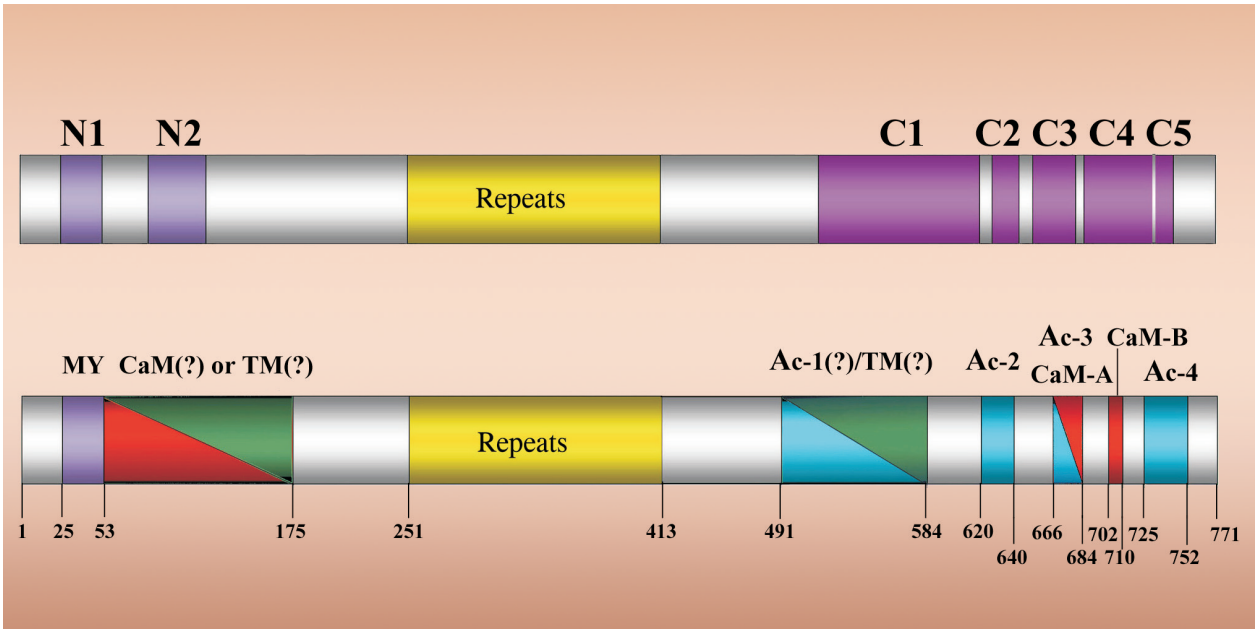

Figure 2. The comparison of the distribution of the H-isoform of chicken caldesmon motifs (a) with its binding sites (b) for myosin (MY), calmodulin (CaM), tropomyosin (TM) and actin (Ac).

In (b), position numbers are stated for the beginnings and ends of the binding sites, while for the beginnings and ends of the motifs see the first sequence of the alignments in Fig. 1b, c.

accepted calmodulin binding site B (Marston *et al.,* 1998), while its C-terminus and probably a part, or even entire, C5 motif contain the fourth, most C-terminal actin binding site (residues L725–V752) (Bartegi *et al.,* 1990; Fraser *et al.,* 1997; Gao *et al.,* 1999; Huber *et al.,* 1998; Marston *et al.,* 1998; Mezgueldi *et al.,* 1994).

Besides seven highly conserved (with over 85% identity) motifs, the CaD molecule contains some less conserved regions, e.g. the sequence between the N1 and N2 motifs or be-

Figure 1. continued

(a) The alignment of chicken and human H-CaD repeats. (b) The conserved motifs found in the N-terminal part of caldesmons. (c) The conserved motifs found in the C-terminal part of caldesmons. The CaD isoform and abbreviated name of the organism and the organ from which the protein was prepared are listed on the left of the alignment, while the name of the database and record in it are given on the right of the alignment. The abbreviations are: H_CHI_GIZZ, chicken gizzard H-CaD; H_CHI_OVID, chicken oviduct H-CaD; L_CHI_GIZZ, chicken gizzard L-CaD; L_CHI_BRAI, chicken brain L-CaD; H_HUM_AORT, human aorta H-CaD; L_HU_WI38a, human aorta L-CaD I; L_HU_WI38b, human aorta L-CaD II; L_HU_HeLa1, human HeLa cell L-CaD I; L_HU_HeLa2, human HeLa cell L-CaD II; L_RABB_FIB, rabbit fibroblast L-CaD; L_RAT_LIVE, rat liver L-CaD; HU_STMC_FR, fragment of human stomach CaD (unknown isoform); L_BOV_EPIT, fragment of adult bovine oviduct epithelium L-CaD; L_MOUS_LIV, fragment of mouse liver mlia L-CaD; L_MOUS_EMB, fragment of mouse embryo mewa L-CaD; L_RAT_OVAR, fragment of rat ovary L-CaD; H_RABB_FRA, fragment of rabbit smooth muscle H-CaD; H_TURK_GIZ, fragment of turkey gizzard H-CaD; MOU_EMB_FR, fragment of mouse total embryo CaD (unknown isoform) and HU_SPLN_FR, fragment of human fetal liver and spleen CaD (unknown isoform). Sequences of protein fragments were included in the alignment procedure if their length exceeded two motifs. Secondary structure prediction (Czuryło *et al.,* 1993) shows helices as H/h (with high and average probability, respectively), β-strands as B, loops and β-turns as t and other structures as *. Above the sequences, the positions in question were marked with: (#), when residues were identical; (+), when a residue was replaced by the same class of residue which resulted in small surface and structural alterations or (:), when surface and structural effects were pronounced, and (·), in the case of replacement by another class residue. The residues are colored according to the CLUSTAL X protein scheme.

tween P458(475) and V473(F490) for the chicken (human) protein. These parts of the proteins could be considered as motifs of lower homology (with about 60% identity on long enough segments of sequences).

There are also regions of major differences, for example, the region between G496(S511) and E511(552) where 26 deletions are found in avian CaDs compared to mammalian CaDs. An instance of such a difference is a characteristic 18 residues long sequence, (534)KRLEELRRRRGETESEEF(551), just before the C1 motif of all CaD isoforms, except in avian where it is replaced by a single S510 residue. Also the 10 residue long (81)TTTTNTQVEG(90) sequence in human CaDs is replaced by (81)RST(83) in chicken CaDs. Generally, the C-termini of the N-terminal part and the N-termini of the C-terminal part of CaD represent less conserved regions of the entire molecule.

As presented schematically in Fig. 2, it may be stated that the distribution of all previously suggested binding sites coincides to a large extent with the proposed CaD motifs. This might be useful in further detailed studies of the CaD interaction with other proteins and directly related to CaD function in smooth muscle contraction.

I am indebted to Dr. Patrick Groves for critical reading of the manuscript and helpful discussion and Dr. Natalia Kulikova for valuable help in preparation of figures.

REFERENCES

Bartegi, A., Fattoum, A., Derancourt, J. & Kassab, R. (1990) Characterization of the carboxyl- terminal 10-kDa cyanogen bromide fragment of caldesmon as an actin-calmodulin-binding region. *J. Biol. Chem.* **265**, 15231–15238.

Chalovich, J.M., Bryan, J., Benson, C.E. & Velaz, L. (1992) Localization and characterization of a 7.3-kDa region of caldesmon which reversibly inhibits actomyosin ATPase activity. *J. Biol. Chem.* **267**, 16644–16650.

Chalovich, J.M. & Pfitzer, G. (1997) Structure and function of the thin filament proteins of smooth muscle; in *Cellular Aspects of Smooth Muscle Function* (Kao, C.Y. & Carsten, M.E., eds.) pp. 253–287, Cambridge University Press, Cambridge, U.K.

Chalovich, J.M., Sen, A., Resetar, A., Leinweber, B., Fredricksen, R.S., Lu, F. & Chen, Y.D. (1998) Caldesmon: Binding to actin and myosin and effects on elementary steps in the ATPase cycle. *Acta Physiol. Scand.* **164**, 427–435.

Combet, C., Blanchet, C., Geourjon, C. & Deléage, G. (2000) NPS@: Network Protein Sequence Analysis. *Trends Biochem. Sci.* **25**, 147–150.

Czuryło, E.A., Emelyanenko, V.I., Permyakov, E.A. & Dąbrowska, R. (1991) Spectrofluorimetric studies on C-terminal 34 kDa fragment of caldesmon. *Biophys. Chem.* **40**, 181–188.

Czuryło, E.A., Hellweg, T., Eimer, W. & Dąbrowska, R. (1997a) The size and shape of caldesmon and its fragments in solution studied by dynamic light scattering and hydrodynamic model calculations. *Biophys. J.* **72**, 835–842.

Czuryło, E.A., Kulikova, N. & Dąbrowska, R. (1997b) Does calponin interact with caldesmon? *J. Biol. Chem.* **272**, 32067–32070.

Czuryło, E.A., Venyaminov, S. & Dąbrowska, R. (1993) Studies on secondary structure of caldesmon and its C-terminal fragments. *Biochem. J.* **293**, 363–368.

Fraser, I.D., Copeland, O., Bing, W. & Marston, S.B. (1997) The inhibitory complex of smooth muscle caldesmon with actin and tropomyosin involves three interacting segments of the C-terminal domain 4. *Biochemistry* **36**, 5483–5492.

Gałązkiewicz, B., Mossakowska, M., Osińska, H. & Dąbrowska, R. (1985) Polymerization of G-actin by caldesmon. *FEBS Lett.* **184**, 144–149.

Gao, Y., Patchell, V.B., Huber, P.A., Copeland, O., El-Mezgueldi, M., Fattoum, A., Calas, B., Thorsted, P.B., Marston, S.B. & Levine, B.A. (1999) The interface between caldesmon domain 4b and subdomain 1 of actin studied by

nuclear magnetic resonance spectroscopy. *Biochemistry* **38**, 15459–15469.

Hall, T.A. (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98.

Haruna, M., Hayashi, K., Yano, H., Takeuchi, O. & Sobue, K. (1993) Common structural and expressional properties of vertebrate caldesmon genes. *Biochem. Biophys. Res. Commun.* **197**, 145–153.

Hayashi, K., Kanda, K., Kimizuka, F., Kato, I. & Sobue, K. (1989) Primary structure and functional expression of h-caldesmon complementary DNA. *Biochem. Biophys. Res. Commun.* **164**, 503–511.

Hayashi, K., Yano, H., Hashida, T., Takeuchi, R., Takeda, O., Asada, K., Takahashi, E., Kato, I. & Sobue, K. (1992) Genomic structure of the human caldesmon gene. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 12122–12126.

Higgins, D.G., Thompson, J.D. & Gibson, T.J. (1996) Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* **266**, 383–402.

Huber, P.A. (1997) Caldesmon. *Int. J. Biochem. Cell Biol.* **29**, 1047–1051.

Huber, P.A.J., Gao, Y., Fraser, I.D.C., Copeland, O., el-Mezgueldi, M., Slatter, D.A., Keane, N.E., Marston, S.B. & Levine, B.A. (1998) Structure-activity studies of the regulatory interaction of the 10 kilodalton C-terminal fragment of caldesmon with actin and the effect of mutation of caldesmon residues 691–696. *Biochemistry* **37**, 2314–2326.

Lee, Y.H., Gallant, C., Guo, H., Li, Y., Wang, C.A. & Morgan, K.G. (2000) Regulation of vascular smooth muscle tone by N-terminal region of caldesmon. Possible role of tethering actin to myosin. *J. Biol. Chem.* **275**, 3213–3220.

Lehman, W., Craig, R., Lui, J. & Moody, C. (1989) Caldesmon and the structure of smooth muscle thin filaments: Immunolocalization of caldesmon on thin filaments. *J. Muscle Res. Cell Motil.* **10**, 101–112.

Leszyk, J., Mornet, D., Audemard, E. & Collins, J.H. (1989) Amino acid sequence of a 15 kilodalton actin-binding fragment of turkey gizzard caldesmon: Similarity with dystrophin, tropomyosin and the tropomyosin-binding region of troponin T. *Biochem. Biophys. Res. Commun.* **160**, 210–216.

Li, Y., Zhuang, S., Guo, H., Mabuchi, K., Lu, R.C. & Wang, C.A. (2000) The major myosin-binding site of caldesmon resides near its N-terminal extreme. *J. Biol. Chem.* **275**, 10989–10994.

Mabuchi, K. & Wang, C.L. (1991) Electron microscopic studies of chicken gizzard caldesmon and its complex with calmodulin. *J. Muscle Res. Cell. Motil.* **12**, 145–151.

Marston, S. & Huber, P. (1996) Caldesmon; in *Biochemistry of Smooth Muscle Contraction* (Barany, M., ed.) pp. 77–90, Academic Press, Inc., San Diego.

Marston, S., Burton, D., Copeland, O., Fraser, I., Gao, Y., Hodgkinson, J., Huber, P., Levine, B., el-Mezgueldi, M. & Notarianni, G. (1998) Structural interactions between actin, tropomyosin, caldesmon and calcium binding protein and the regulation of smooth muscle thin filaments. *Acta Physiol. Scand.* **164**, 401–414.

Mezgueldi, M., Derancourt, J., Calas, B., Kassab, R. & Fattoum, A. (1994) Precise identification of the regulatory F-actin- and calmodulin-binding sequences in the 10-kDa carboxyl-terminal domain of caldesmon. *J. Biol. Chem.* **269**, 12824–12832.

Mornet, D., Audemard, E. & Derancourt, J. (1988) Identification of a 15 kilodalton actin binding region on gizzard caldesmon probed by chemical cross-linking. *Biochem. Biophys. Res. Commun.* **154**, 564–571.

Rost, B., Sander, C. & Schneider, R. (1994) PHD — an automatic mail server for protein secondary structure prediction. *Comput. Appl. Biosci.* **10**, 53–60.

Sobue, K., Muramoto, Y., Fujita, M. & Kakiuchi, S. (1981) Purification of a calmodulin-binding protein from chicken gizzard that interacts with F-actin. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 5652–5655.

Sobue, K. & Sellers, J.R. (1991) Caldesmon, a novel regulatory protein in smooth muscle

and nonmuscle actomyosin systems. *J. Biol. Chem*. **266**, 12115–12118.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. **22**, 4673–4680.

Wang, C.L. (1988) Photocrosslinking of calmodulin and/or actin to chicken gizzard caldesmon. *Biochem. Biophys. Res. Commun.* **156**, 1033–1038.

Wang, C.L., Chalovich, J.M., Graceffa, P., Lu, R.C., Mabuchi, K. & Stafford, W.F. (1991a) A long helix from the central region of smooth muscle caldesmon. *J. Biol. Chem.* **266**, 13958–13963.

Wang, C.L., Wang, L.W., Xu, S.A., Lu, R.C., Saavedra-Alanis, V. & Bryan, J. (1991b) Localization of the calmodulin- and the actin-binding sites of caldesmon. *J. Biol. Chem*. **266**, 9166–9172.

Wang, E. & Wang, C.L. (1996) (i, i + 4) Ion pairs stabilize helical peptides derived from smooth muscle caldesmon. *Arch. Biochem. Biophys*. **329**, 156–162.

Wang, Z., Jiang, H., Yang, Z.Q. & Chacko, S. (1997) Both N-terminal myosin-binding and C-terminal actin-binding sites on smooth muscle caldesmon are required for caldesmon-mediated inhibition of actin filament velocity. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11899–11904.