Volume 262, number 2, 185–188 FEBS 08232 March 1990

A second consensus sequence of ATP-requiring proteins resides in the 21-kDa C-terminal segment of myosin subfragment 1

Morris Burke*, Kallikat N. Rajasekharan*, Shinsaku Maruta° and Mitsuo Ikebe°

*Department of Biology, Case Western Reserve University and °Department of Physiology and Biophysics, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA

Received 22 December 1989; revised version received 30 January 1990

Previous comparisons of sequence homologies of ATP-requiring enzymes have defined three consensus sequences which appear to be involved in the binding of the nucleotide. One of these was identified in the N-terminal 27-kDa segment of the myosin heavy chain but the other two sequences have not hitherto been located in myosin. The present paper proposes that one of these other two consensus sequences is in the 21-kDa C-terminal portion of S1 and that it may contribute to the ATP binding domain.

Myosin; ATP-requiring enzyme; ATP-binding domain

1. INTRODUCTION

With the advent of abundant protein sequence information through DNA sequencing, correlations in the sequences of proteins with related functions have been useful in identifying putative functional regions. This information when used in conjunction with known three-dimensional structures provides good clues to the potential functional sequences in related proteins of yet undefined structure ([1] and references therein). This has been successfully applied to identify the glycine-rich loop in the 27-kDa N-terminal segment of myosin S1 as being part of the nucleotide binding region (sequence A of reference [1]). This assignment has recently been confirmed by the observation that Ser^{180} in this consensus sequence undergoes photo-oxidation by Vi trapped stoichiometrically in S1 with MgADP [2].

Another consensus sequence of highly conserved homology has been identified in F1 ATP synthases and in mammalian cytosolic adenylate kinase (sequence B in [1]) and is shown in table 1. It consists of EX_2RX_m GX_n followed by a five residue segment comprised of Ile or Leu at the first two positions, Tyr or Phe at the third position, an uncharged or Asp in the fourth position and terminated by an aspartate. The subscripts *m* and *n* represent the number of variable residues X between the bordering conserved ones. For F1 ATPases and mammalian cytosolic adenylate kinase, *m* is usually 2 or 3 and *n* is 3 or 4. It has been suggested that the terminal aspartate of this sequence plays a role in the binding of the divalent cation associated with the nucleotide substrate, thereby placing it at the active site of these other ATPases and nucleotide binding proteins. In accord with this view, the terminal five residue segment in the case of mammalian cytosolic adenylate kinase has been shown (together with sequence A) to flank the Mg triphosphate group on the basis of detailed NMR analyses [3,4]. Sequence B was suggested to lie in myosin in the 27-kDa segment adjacent to and on the N-terminal side of sequence A but the homology with myosin is low [1]. Furthermore, the suggested location places B before and close to A, which is opposite to what is observed with the non-myosin ATPases. The present paper raises the possibility that sequence B may reside in the 21-kDa C-terminal segment of S1.

Correspondence address: M. Ikebe, Department of Physiology and Biophysics, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA

2. MATERIALS AND METHODS

Computer searches for frequency of specified sequences in the SWISS PROT database and for secondary structure predictions were done using the PESEARCH and GARNIER programmes respectively of PC/GENE (IntelliGenetics, Inc., Mountain View, CA).

3. RESULTS AND DISCUSSION

A visual search of the rabbit skeletal myosin S1 sequence indicated that a pentapeptide beginning at Ile^{715} showed high homology to the conserved, terminal five residue segment of consensus sequence B. The sequence immediately upstream from this pentapeptide was next examined for homology with the remainder of consensus sequence B by adjusting the spacing to obtain best matches between the assigned conserved residues. There appeared to be a good visual match in which consensus sequence B could be placed in myosin by allow-

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

Table 1

Alignment of homologous sequences in F1 ATP synthases and mammalian cytosolic adenylate kinase and in the proposed location in the 21-kDa segment of myosin S1

Non Myosin ATPases	
Bovine ATPase β	V A E Y F R - D Q E G Q D - V L L F I D N I
E. Coli ATPase β	M A E K F R - D - E G R D - V L L F V D N I
E. Coli ATPase α	M G E Y F R - D R - G E D A L I I Y D D L S
R. Blastica ATPase α [29]	I G E F F R - D N - G R H A L M I Y D D L S
Adenylate kinase	G E E F E R - - K I G Q P T L L L Y V D A G
Myosin	
Vertebrates	
Chicken gizzard [30]	R C N G V L E G I R I C R Q G F P N R I V F Q E F R
Chicken skeletal [31]	R C N G V L E G I R I C R K G F P S R V L Y A D F K
Rat skeletal embry [32]	R C N G V L E G I R I C R K G F P N R I L Y G D F K
Rabbit skeletal[b]	R C N G V L E G I R I C R K G F P S R I L Y A D F K
Rabbit cardiac [33]	R C N G V L E G I R I C R K G F P N R I L Y G D F R
Human skeletal embryo [34]	R C N G V L E G I R I C R K G F P N R I L Y G D F K
Chicken intest. epit [35]	R C N G V L E G I R I C R Q G F P N R V V F Q E F R
	SH2(697) SH1(707)
Invertebrates	
D. discoideum [36]	R C N G V L E G I R I T R K G F P N R I I Y A D F V
Yeast [37]	R C N G V L E G I R L A R E G Y P N R I A F Q E F F
Acanthamoeba II [38]	K C N G V L E G I R I A R R G W P N R L K Y N E F K
Nematode [39]	T C N G V L E G I R I C R K G F P N R T L H P D F V
D. melanogaster [40]	T C N G V L E G I R I C R K G F P N R M M Y P D F K
Myosin I	
Acanthamoeba IB [5]	K Y L G L L E N V R I R R A G Y A Y R Q S Y - D K F
D. discoidium I [41]	Q Y L G L L E N V R V R R A G F A Y R N T F - D K V
Acanthamoeba I [41]	Q Y L G L L E N V R V R R A G F A Y R A E F - D R L

[a] Sequences from Walker et al. [1]

[b] Elzinga (personal communication)

Sequence B has the motif: $EXXRX_mGX_n$ followed by a five residue segment of (hydrophobic, aliphatic)-(hydrophobic, alipathic)-(Phe,Tyr)-X-(Asp,Glu). The conserved residues and the five residue terminal segment are shown boxed. Blank spaces were inserted for best fit.

ing *m* to be 4, corresponding to a single addition in the spacing between the assigned conserved arginine and glycine residues.

The proposed location for sequence B in rabbit myosin S1 is in the 21-kDa C-terminal segment, beginning at Glu^{702} (five residues downstream from SH2) and ending at Asp^{719}. In this location it is highly conserved in myosins from sources as diverse as *Dictyostelium discoideum* and human embryonic skeletal muscle; the homology with the consensus sequence of the F1 ATPases and mammalian, cytosolic adenylate kinase is also apparent from table 1. The substitutions for myosin from avian gizzard and intestinal epithelial cells and from yeast are quite conservative, but there is lower homology with the myosins of the invertebrates nematode and *Acanthamoeba* II. The homology of the terminal five residue segment is only partially conserved in the type I myosins of *Acanthamoeba* and *Dictyostelium discoideum*. This latter class of protein differs in a number of important aspects from conventional myosins, and has apparently been formed by fusion of a myosin-related gene and that of an unknown protein [5].

Searches of the protein data base (see section 2) for patterns represented by $EX_2RX_mGX_n(IL)X(FY)X(DE)$, where the subscripts *m* and *n* refer to the spacings of variable residues between the conserved ones in the proteins listed in table 1, were next examined. The pattern was made more general than necessary by allowing the second position of the terminal pentapeptide to be variable, since it is usually an aliphatic, hydrophobic residue. The type I class of myosin was not used to set this pattern since, as mentioned above, this represents an unusual 'hybrid' class [5]. The results of this search are presented in table 2. A total of 23 proteins were scored, all of which were ATP-requiring and all but one was either an ATP synthase, adenylate kinase or myosin. The exception was phosphoribosyl ATP-pyrophosphohydrolase [6]. The results of *m*/*n* combinations of 3/4 or 4/3, not occurring within any protein in table 1, scored no protein in the database. Substitution of these patterns at the first position of the terminal pentapeptide by valine or methionine (alternate hydrophobic, aliphatic residues) scored two proteins both of which were F1 ATPases. It would appear, therefore, that the proposed consensus sequence is fairly specific for myosin and some ATP-requiring proteins.

The proposed assignment of sequence B in myosin may have interesting implications regarding the folding

Table 2

Results of searches of the SWISSPROT database for sequences of the pattern: $EX_2RX_mGX_n(IL)X(FY)X(DE)$ where $m = 2$, 3, or 4 and $n =$ 3 or 4. Total sequences in database are 10008

m	n	Proteins scored				
		Total	Myosin	F1-ATPases	Adenylate kinase	Others
2	3	1	0	1	0	0
2	4	12	0	9	2	1[a]
3	3	6	0	6	0	0
4	4	4	4	0	0	0
		23	4	16	2	1

[a] Phosphoribosyl-ATP pyrophosphohydrolase [6]

of the S1 heavy chain to generate the ATP binding domain. It is conceivable that sequence A and the terminal five residues of sequence B may have similar three dimensional relationships in myosin as they exhibit in adenylate kinase, where they both flank the Mg triphosphate group of the ATP [3,4]. Support for a close arrangement of the two consensus sequences in S1 comes from recent probing of the folded structure of S1 by cross-linking. These studies have established that SH1, which resides in the proposed sequence, is close to the 27-kDa segment [7,8], while SH2 has been shown to be within 0.3 to 0.45 nm of Lys^{184} or Lys^{189}, at or adjacent to sequence A, respectively [9,10]. Additional evidence that regions of the 21-kDa segment may be at, or close to, the ATP site comes from recent studies [11] in which a photoaffinity analogue, 3′-*O*-(*N*-methylanthraniloyl)-8-azidoadenosine 5′-triphosphate, labeled avian skeletal S1 in the 21-kDa segment at a 50 residue peptide starting at Leu^{660} and containing SH1 and SH2. We have also recently found that the bifunctional photoreactive analogue, 3′-*O*-(azidonitrobenzoyl)-8-azidoadenosine 5′-triphosphate, cross-links the 21-kDa to the 27-kDa segment in rabbit skeletal S1 [12].

The proximity of elements of the two consensus sequences is also predicted by the recent 3D-lattice model of Botts et al. [13], which is based primarily on fluorescence energy transfer measurements. Such an arrangement of the two sequences would imply that regions of both the 27-kDa and 21-kDa segments may contribute to the ATPase site in myosin S1, and that these two segments, which are well separated in the primary sequence, may not be separate domains as has been suggested from limited proteolysis studies [14,15]. Recent studies on the progressive unfolding of S1 indicate that the S1 structure may be comprised of a stable domain consisting of the interacting 21-kDa, 27-kDa, and light chain, and a less stable middle 50-kDa segment [16,17]. Secondary structure predictions ([18] and see section 2) based on the sequences for adenylate kinase and *Dictyostelium discoideum* myosin, indicate interesting common features. In particular, these predictions yield ordered structures (α/β or β/β respectively) separated by a turn near the conserved glycine followed by two or three residues in a random structure.

The relationship of SH1 and SH2 to sequence B would also be consistent with other well characterized observations accompanying their modification such as changes in the ATPase properties [19], trapping of Mg nucleotide and MgPPi when they are covalently bridged [20], changes in their separation on addition of MgADP [21–23] and changes in their reactivities induced by binding of nucleotide [24,25]. It is of interest to note that modification of the equivalent thiols in the 25-kDa C-terminal segment of smooth myosin S1 abolishes the formation of the 10S conformation [26] usually associated with trapping of MgATP [27]. Furthermore, this location for sequence B would also be at a recently defined actin contact near SH1 [28], consistent with the well known communication between the ATPase and actin sites.

The proposed location and contribution of sequence B in myosin S1 to the formation of the ATP site can eventually be subjected to experimental testing by mutagenesis or other approaches. For example, deletion of residues in the five residue segment should have profound effects on the ATPase properties of myosin. Alternatively, replacement of the common aspartate (or glutamate) by a lysine should produce dramatic effects on the cation dependent ATPase if it is involved in coordination to Mg nucleotide.

Acknowledgements: We wish to thank Dr Marshall Elzinga for kindly providing us with the complete sequence of rabbit skeletal myosin heavy chain. This work was supported by NIH Grants (M.B. and M.I.).

REFERENCES

[1] Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) EMBO J. 1, 945–951.
[2] Cremo, C.R., Grammer, J.C. and Yount, R.G. (1989) J. Biol. Chem. 264, 6608–6611.
[3] Fry, D.C., Kuby, S.A. and Mildvan, A.S. (1985) Biochemistry 24, 4680–4694.
[4] Fry, D.C., Kuby, S.A. and Mildvan, A.S. (1986) Proc. Natl. Acad. Sci. USA 83, 907–911.
[5] Jung, G., Korn, E.D. and Hammer, J.A., iii (1987) Proc. Natl. Acad. Sci. USA 84, 6720–6724.
[6] Donahue, T.F., Farabaugh, P.J. and Fink, G.R. (1982) Gene 18, 47–59.
[7] Lu, R.C., Moo, L. and Wong, A.G. (1986) Proc. Natl. Acad. Sci. USA 83, 6392–6396.
[8] Sutoh, K. and Lu, R.C. (1987) Biochemistry 26, 4511–4516.
[9] Hiratsuka, T. (1987) Biochemistry 26, 3168–3173.
[10] Sutoh, K. and Hiratsuka, T. (1988) Biochemistry 27, 2964–2969.
[11] Maruta, S., Miyanishi, T. and Matsuda, G. (1989) Eur. J. Biochem. 184, 213–221.
[12] Maruta, S., Burke, M. and Ikebe, M. (1990) submitted.

[13] Botts, J., Thomason, J.F. and Morales, M.F. (1989) Proc. Natl. Acad. Sci. USA 86, 2204–2208.
[14] Mornet, D., Betrand, R., Pantel, P., Audemard, E. and Kassab, R. (1981) Biochemistry 20, 2110–2120.
[15] Mornet, D., Ue, K. and Morales, M.F. (1984) Proc. Natl. Acad. Sci. USA 81, 736–739.
[16] Setton, A. and Muhlrad, A. (1984) Arch. Biochem. Biophys. 235, 411–417.
[17] Burke, M. and Sivaramakrishnan, M. (1986) J. Biol. Chem. 261, 12330–12336.
[18] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[19] Yamaguchi, M. and Sekine, T. (1966) J. Biochem. 59, 24–33.
[20] Wells, J.A., Knoeber, C., Sheldon, M.C., Werber, M.M. and Yount, R.G. (1980) J. Biol. Chem. 255, 11135–11140.
[21] Burke, M. and Reisler, E. (1977) Biochemistry 16, 5559–5563.
[22] Dalbey, R.E., Weiel, J. and Yount, R.G. (1983) Biochemistry 22, 4696–4706.
[23] Cheung, H.C., Gonsoulin, F. and Garland, F. (1985) Biochim. Biophys. Acta 832, 52–62.
[24] Reisler, E., Burke, M. and Harrington, W.F. (1974) Biochemistry 13, 2014–2022.
[25] Schaub, M.C., Watterson, J.G. and Wasser, P.G. (1975) Hoppe-Seyler's Z. Physiol. Chem. 356, 325–339.
[26] Nath, N., Nag, S. and Seidel, J.C. (1986) Biochemistry 25, 6169–6176.
[27] Cross, R.A., Cross, K.E. and Sobieszek, A. (1986) EMBO J. 5, 2637–2641.
[28] Suzuki, R., Nishi, N., Tokura, S. and Morita, F. (1987) J. Biol. Chem. 262, 11410–11412.
[29] Tybulewicz, V.L.J., Falk, G. and Walker, J.E. (1984) J. Mol. Biol. 179, 185–214.
[30] Onishi, H., Maita, T., Miyanishi, T., Watanabe, S. and Matsuda, G. (1986) J. Biochem. 100, 1433–1447.
[31] Maita, T., Hayashida, M., Tanioka, Y., Komino, Y. and Matsuda, G. (1987) Proc. Natl. Acad. Sci. USA 84, 416–420.
[32] Strehler, E.E., Strehler-Page, M., Perriard, J., Periasamy, M. and Nadal-Ginard, B. (1986) J. Mol. Biol. 190, 291–317.
[33] Kavinsky, C.J., Umeda, P.K., Levin, J.E., Sinha, A.M., Nigro, J.M., Jakovcic, S. and Rabinowitz, M. (1984) J. Biol. Chem. 259, 2775–2781.
[34] Eller, M., Stedman, H.H., Sylvester, J.E., Fertels, S.H., Rubenstein, N.A., Kelly, A.M. and Sarkar, S. (1989) Nucleic Acid Res., in press.
[35] Shohet, R.V., Conti, M.A., Kawamoto, S., Preston, Y.A., Brill, D.A. and Adelstein, R.S. (1989) Proc. Natl. Acad. Sci. USA 86, 7726–7730.
[36] Warrick, H.M., De Lozanne, A., Leinwand, L.A. and Spudich, J.A. (1986) Proc. Natl. Acad. Sci. USA 83, 9433–9437.
[37] Watts, F.Z., Shiels, G. and Orr, E. (1987) EMBO J. 6, 3499–3505.
[38] Hammer, J.A., iii, Bowers, B., Paterson, B.M. and Korn, E.D. (1987) J. Cell Biol. 105, 913–925.
[39] Karn, J., Brenner, S. and Barnett, L. (1983) Proc. Natl. Acad. Sci. USA 80, 4253–4257.
[40] George, E.L., Ober, M.B. and Emerson, C.P., jr (1989) Mol. Cell Biol. 9, 2957–2974.
[41] Jung, G., Saxe, C.L., iii, Kimmel, A.R. and Hammer, J.A., iii (1989) Proc. Natl. Acad. Sci. USA 86, 6186–6190.