Journal of Muscle Research and Cell Motility (2005) 26:427–434
DOI 10.1007/s10974-005-9025-6

Complete human gene structure of obscurin: implications for isoform generation by differential splicing

ATSUSHI FUKUZAWA, SERAPHINA IDOWU and MATHIAS GAUTEL*
The Randall Division of Cell and Molecular Biophysics and The Cardiovascular Division, New Hunt's House, King's College London, SE1 1UL, London, United Kingdom

Abstract

The complete gene giant muscle protein obscurin, a modular protein composed largely of tandem Ig-domains, GDP/GTP exchange factor domains (GEF) for small G-proteins, and differentially spliced kinase domains, was analysed. The splice donor and acceptor sites of the 117 exons give important clues for potential splice pathways. The fusion of the conventional obscurin A, containing only the GEF domain, and obscurin B, fusing into the 3′ kinase exons, was experimentally confirmed and analysed. The linker between the two kinases contains multiple predicted phosphorylation sites, as well as a predicted NFX zinc finger domain. Both kinases show only weak homology to either myosin light chain kinases or other giant muscle protein kinases, suggesting that they are functionally distinct.

Introduction

The giant modular muscle protein obscurin was first identified as a ligand of the Z-disk titin/connectin domains Z9 and Z10 (Young and Gautel, 1999). A complete cardiac cDNA contig of 20 kb revealed a modular architecture composed of intracellular immunoglobulin and fibronectin type 3 domains, combined with several signalling domains: a constitutively calmodulin-binding IQ motif and a predicted GDP/GTP exchange factor (GEF) domain for small GTPases (Young *et al*., 2001), containing a src homology 3 (SH3), dbl homology (DH) and pleckstrin homology (PH) domains. Differential splicing occurs within the Ig-domain encoding exons of the gene, which contains additional 10 exons encoding Ig domains (Young *et al*., 2001). Additional splice variants were reported to arise by internal deletion of exons (Young *et al*., 2001) or by fusion of the 3′ end of the mRNA to exons located up to 20 kb downstream of the 3′ untranslated region of 'conventional' obscurin (Russell *et al*., 2002). These additional 3′ exons were reported to encode Ig and Fn3 domains as well as two predicted serine-threonine protein kinase domains of the myosin-light chain kinase family. A very similar pattern is found in the *C. elegans* analogue unc-89 (Small *et al*., 2004).

Surprisingly, obscurin was also found to interact with domains of a titin/connectin Z-disk/I-band splice variant, novex-3 (Bang *et al*., 2001). The C-terminal part of titin novex-3 interacts with the same obscurin immunoglobulin domains, Ob58 and 59 (revised nomenclature), that also form the binding site for titin Z9Z10, without showing significant sequence similarity. Both interactions are in agreement with the Z-disk targeting observed for the isolated titin-binding domains of obscurin in cardiac muscle cells (Young *et al*., 2001). It came as a surprise, therefore, that antibodies against epitopes along the entire length of obscurin detect the protein at the M-band of cardiomyocytes (Young *et al*., 2001), over 1 μm away from its assumed localisation. Although obscurin can also sometimes localise to the Z-disk, as well as to the Z-disk and M-band simultaneously (Borisov *et al*., 2004; Kontrogianni-Konstantopoulos *et al*., 2003; Young *et al*., 2001), the sheer distance of more than 1 μm (over five times the molecule's predicted maximal extension) between the expected and observed positions defies established views of how giant proteins should to behave. Recently, the cytoplasmic domain of the small sarcoplasmic reticulum (SR) transmembrane protein ankyrin 1.5 was reported to interact with the non-modular C-terminus of 'conventional' obscurin (in the following called obscurin A) (Bagnato *et al*., 2003; Kontrogianni-Konstantopoulos *et al*., 2003). Small ankyrin 1.5 is generally localised to the M-band associated SR, suggesting obscurin can serve as a lateral link between the SR and the myofibril. Since this region is spliced out in the kinase domain-containing obscurin isoforms (obscurin B), at least some of the M-band associated functions of the protein appear to be different. It is presently unclear whether the GEF domain has similar functions at the M-band or Z-disk, or whether the dual localisation of obscurin reflects also a dual function of its signalling domains.

* To whom correspondence should be addressed. E-mail: mathias.gautel@kcl.ac.uk

Despite its size of around 800 kDa, obscurin expression appears to be dynamically regulated during pathological or physiological adaptation of heart muscle (Borisov *et al*., 2003), this maybe less surprising for a protein with more than one localisation.

The name obscurin thus still befits this elusive protein which is hard to see, hard to understand and hard to analyse. So far, the powers of genetic intervention in nematode models have not completely clarified the situation, although important insight was made recently. The nematode *C. elegans* contains a possible orthologue of obscurin, the unc-89 gene product. Unc-89 shares many features with obscurin, especially the SH3-DH-PH GEF domain trio, the overall type of constituent domains, including fibronectin domains adjacent to the differentially spliced kinase domains, and the localisation to the M-band in nematode obliquely striated muscle (Benian *et al*., 1996; Small *et al*., 2004). The domain order in unc-89 is rather different than that in obscurin, with the GEF-domain at the N-terminus and up to 45 KSP-phosphorylation repeats in the nematode protein. Unc-89 has so far not been reported at the Z-bodies of nematode muscles; the discovery of the unc-89 kinases, however, suggests a much higher degree of functional homology between obscurin and unc-89 as previously suspected. A mutation in the *unc*-89 gene leads to loss of M-line assembly and disordered A-bands (Benian *et al*., 1996); a potentially similar phenotype on myosin assembly was also observed when part of the obscurin A C-terminus was transfected into skeletal myotubes (Kontrogianni-Konstantopoulos *et al*., 2004). However, since no direct myosin binding domains have been identified in this fragment, this phenotype may be due to a more indirect effect, possibly by compromising calcium handling of the SR, which is required for proper myosin filament assembly (Ferrari *et al*., 1998).

Since obscurin is localised to the periphery of the sarcomere in skeletal (Kontrogianni-Konstantopoulos *et al*., 2003) and cardiac muscle (M. Gautel unpublished observations), the protein is clearly not subject to the same steric and stoichiometric restraints as a strictly intra-sarcomeric protein. Knowledge of the possible domain compositions of obscurin isoforms, and their analysis during development or in diseased muscle requires the availability of the complete gene structure. Here, we present an analysis of the complete obscurin gene and the splice pathways leading to isoforms in the tandem Ig regions, and the kinase domains.

Materials and methods

The complete obscurin gene was identified by BLAST searches of the November 2004 complete human genome release. A 190 kb fragment, covering the previously described genomic region of obscurin as well as the 3′ kinase exons was analysed using GENSCAN (Burge and Karlin, 1997), as well as by homology searches for immunoglobulin and fibronectin-like domains. Domain composition was further analysed by SMART (Letunic *et al*., 2004). Promoter sites were analysed using TFSEARCH (www.cbrc.jp/research/db/TFSEARCH).

The following previously released genomic sequences were analysed and fused using the latest human genome release:

AJ314896 Homo sapiens partial OBSCN gene for obscurin, exons 1 and 2
AJ314898 Homo sapiens partial OBSCN gene for obscurin, exons 5–11
AJ314901 Homo sapiens partial OBSCN gene for obscurin, exons B and 17
AJ314900 Homo sapiens partial OBSCN gene for obscurin, exons a1–16
AJ314903 Homo sapiens partial OBSCN gene for obscurin, exons C2–22
AJ314904 Homo sapiens partial OBSCN gene for obscurin, exons 23–30
AJ314905 Homo sapiens partial OBSCN gene for obscurin, exons 31–56
AJ314906 Homo sapiens partial OBSCN gene for obscurin, exons 57–68
AJ314907 Homo sapiens partial OBSCN gene for obscurin, exons 69–72
AJ314908 Homo sapiens partial OBSCN gene for obscurin, exons 73–87

Reverse transcription polymerase chain reactions (RT-PCR) were performed with the following primers: 5′ TTTCTCGAGC CCTGTGTGGCGGCCCCC against Ob66 in obscurin A, and 3′ against predicted kinase 1 AGTGTCCCTGCAGAGGTGCCGGGCTAC, 5′ CCAGGCTTGCTTCCTTCC and 3′ CTATTGGCGCGTCTGCAGCGACC for Ob69 and kinase 2, using total human brain, cardiac and skeletal muscle cDNA as templates. RT-PCR products were subcloned by TA cloning and analysed by cDNA sequencing. The cDNA sequence was deposited in the EMBL database, accession number AM231061.

Analysis of the complete obscurin gene

The complete obscurin gene OBSCN on chromosome 1q42 was identified by homology searches of the complete human genome, release November 2004. The retrieved 190 kb sequence fragment covered all our previously released partial sequences of the obscurin gene (see Materials and Methods). These were separated by 8 gaps of 12–1571 base pairs. Analysis of the short gap regions revealed that they were intronic sequences. The two longest of these gaps, between entries AJ314896 and AJ314898, and AJ314900 and AJ314903 respectively, were found to contain exons encoding three complete immunoglobulin domains. The coding sequences were analysed for obscurin

immunoglobulin and fibronectin-like domains by homology searches. The first gap contained the C-terminus of the Ig-domain Ob2 (nomenclature as in Young *et al*., 2001), as well as the entire domain Ob3 with the non-modular linker 1 contained within a single exon, and domain Ob4 on a separate exon. The second gap contained domains Ob18, although the sequence quality in this region was poor and the 3′ splice site could not be unambiguously assigned. Ig domain Oba1, which replaces Ob10 and Ob12–Ob15 in a cardiac splice variant (Young *et al*., 2001) was resequenced and found to be identical to Ob11. Since no further Ig or fibronectin domains were detected in addition to those published previously (Young *et al*., 2001), and no additional exons encoding non-modular sequences were predicted by GENSCAN even with non-stringent parameters, a final classification of obscurin domains can now be proposed. A total of 65 Ig domains, and 2 fibronectin domains are found in obscurin A. Together with the 5′ and 3′ untranslated regions, the non-modular protein sequences and the SH3, DH, PH and IQ signalling-domains, they are encoded on 91 exons. For future reference, we propose to number all Ig and Fn3 domains in obscurin A consecutively from 1 to 67 based on their order within the gene. The previous nomenclature by (Young *et al*., 2001) and the resulting, gene-structure based nomenclature are listed in Table 1. The tandem Ig and Fn3 domains are generally encoded contiguously on single exons, with the exceptions of Ob28, Ob57, Ob64, Ob65, Ob66 and Ob67, which are split over two exons. Analysis of the splice sites reveals that almost all Ig- and Fn3-encoding exons follow the conventional pattern (Shapiro and Senapathy, 1987) and code in frame 3 (Table 1). This allows their successive concatenation, as the 3′ splice donor site of any of the tandem Ig domains can fuse to the 5′ acceptor site of any other downstream domain (Table 1, Figure 1). Exceptions of this rule are the splice sites between the N- and C-terminal parts of the split domains, Ob28 and Ob57, which link in frame 1. Although there is a theoretical possibility that the N-terminal half of Ob28 could be spliced to the C-terminal half of Ob57, the lengths of the two C-terminal domain fragments are very different (51 amino acids in exon 28 for Ob28, vs. 25 in exon 59 for Ob57), and further analysis of the putative splice product suggested that the encoded hybrid domain would be unlikely to be properly folded (Table 1).

Obscurin kinases

Several publications have reported the existence of differentially spliced obscurin isoforms containing two serine-threonine kinase domains (Bang *et al*., 2001; Russell *et al*., 2002). However, since no sequences are available in the public database to corroborate these reports, we combined gene analysis with RT-PCR approaches to identify the fusion of obscurin A to the kinase-encoding exons. Analysis of the complete gene revealed a predicted gene starting 1325 bp downstream of the 3′-UTR of obscurin A, with 32 putative exons (of which 26 could be experimentally confirmed) over 17 kb. This sequence encodes two serine-threonine kinase domains, as well as two Ig and one Fn3 domain. The predicted cDNA from these exons is essentially identical to the human brain expressed cDNA sequence KIAA1639 (Nagase *et al*., 2000), although the latter is incomplete and starts with a truncated kinase 1 domain. To assess the potential splicing of obscurin to the kinases in muscle we used a primer pair at the 5′ end of Ob66 (former Ob56) and the 3′ end of kinase 1. RT-PCR in skeletal and cardiac muscle cDNA libraries yielded a fragment of 2.3 kb, which was analysed by DNA sequencing. The analysis of the encoded peptide revealed the existence of three Ig domains, the N-terminal of which are Ob66 and Ob67 (Figure 2a). After the short stretch of linker encoded together with Ob67 on exon 89, a 136 amino acid non-modular sequence is followed by an Ig-domain and a serine-threonine catalytic kinase domain (Figure 2a), henceforth referred to as kinase 1. Fusion of exon 89 to the first of the kinase linker exons, exon 92 (numbered consecutively from the last obscurin A exons) leads to skipping of the ankyrin-binding non-modular C-terminus. The linker/Ig/kinase 1 region is encoded on 15 exons, the last five of which (exon 102 onwards) are identical to the 5′ end of KIAA1639. Analysis of the exon usage of the experimentally verified linking sequence showed that two of the predicted exons between Ob67 and kinase 1 are not represented in this cDNA; it is not clear whether they are expressed at all. Two further exons are predicted within the catalytic domain of kinase 1; their splicing into the kinase sequence would lead to a 38 amino acid insertion in the ATP-binding lobe, and a 29 amino acid insertion after the catalytic base, leading to a vastly expanded activation loop. RT-PCR yielded no indication that these putative exons are actually expressed.

The regulatory domain of kinase 1 is found on exon 106, which contains a contiguous 2 kb sequence, encoding mostly non-modular sequence as well as a NFX zinc finger domain. This linker sequence is followed by four exons encoding one Ig and Fn3 domain, followed by the MLCK-like kinase 2 domain. The last exon of kinase 2 (exon 117) encodes mostly the regulatory domain of kinase 2, which is followed by a stop codon in the same exon. The same stop codon is found in KIAA1639. RT-PCR in brain, cardiac and skeletal muscle cDNAs with primers against the Ig and the kinase domain confirmed this arrangement and the stop codon after kinase 2. Two further predicted exons 3′ of exon 117 may contribute to the 3′ UTR of the obscurin-kinase mRNA.

Analysis of the intronic sequence between exon 91 and the first kinase exon, exon 92, for putative promoter sites showed the existence of multiple consensus

Table 1. Exons of the obscurin gene and their encoded domains

Exon Number	Previous nomenclature	New nomenclature	5′ splice	3′ splice
1	5′ UTR	5′ UTR		AGAG/GT
2	Ob1, Ob2, non mod 1, Ob3	Ob1, Ob2, non mod 1, Ob3	AG/GTCCC	CGCG/GT
3	Ob4	Ob4	AG/AGCCC	AAGG/CA
4	Ob5	Ob5	AG/GAAAC	TCGG/GT
5	Ob6	Ob6	AG/CCCCC	CTGG/GT
6	Ob7	Ob7	AG/CCCCC	CGGG/GT
7	Ob8	Ob8	AG/CGGCC	CAGG/GT
8	Ob9	Ob9	AG/GCCTC	TCTG/GT
9	ObA	Ob10	AG/AGCCC	GCAG/GT
10	Ob10	Ob11	AG/AGCCC	AAAG/GT
11	Ob11/Oba1	Ob12	AG/AGCCC	ACAG/GT
12	Ob12	Ob13	AG/AGCCC	ACAG/GT
13	Ob13	Ob14	AG/AGCCC	ACAG/GT
14	Ob14	Ob15	AG/AGCCC	TCAG/GT
15	Ob15	Ob16	AG/AGCCC	GCAG/GT
16	Ob16	Ob17	AG/AGCCC	CCAG/GT
17	ObB	Ob18	AG/AGCCC	GCAG/GT
18	Ob17	Ob19	AG/AGCCC	GCAG/GT
19	Ob18	Ob20	AG/CCCAA	Unclear
20	ObC	Ob21	AG/GTGGT	GCTG/GT
21	ObD	Ob22	AG/AGCCC	GCAG/GT
22	Ob19	Ob23	AG/AGCTG	GAAG/GT
23	Ob20	Ob24	AG/AGGTG	AGTG/GT
24	Ob21	Ob25	AG/TGCCC	CGAG/GT
25	Ob22	Ob26	AG/AGGCG	GAGA/GT
26	Ob23	Ob27	AG/TGCGC	ACCG/GT
27	Ob24 part 1	Ob28 part 1	AG/AGCTT	GAAG/GT
28	Ob24 part 2	Ob28 part 2	AG/GACGG	CAAG/GT
29	non mod 2	non mod 2	AG/GAAGG	AAGG/GT
30	Ob25	Ob29	AG/AGCTG	GAAG/GT
31	Ob26	Ob30	AG/AGCAA	TCAG/GT
32	Ob27	Ob31	AG/ACATC	GAAG/GT
33	Ob28	Ob32	AG/CGCGG	CGAG/GT
34	Ob29	Ob33	AG/TGAAG	CACG/GT
35	Ob30	Ob34	AG/ATCTG	AGAG/GT
36	Ob31	Ob35	AG/AGAGG	GAAG/GT
37	Ob32	Ob36	AG/AAAAA	CAAG/GT
38	Ob33	Ob37	AG/GCCGG	ACTG/GT
39	Ob34	Ob38	AG/AGAAG	CATG/GT
40	Ob35	Ob39	AG/CGCGG	ACTG/GT
41	Ob36	Ob40	AG/CCCTG	AGGC/GT
42	Ob37	Ob41	AG/CCATG	AAGG/GT
43	Ob38	Ob42	AG/CCCTG	AGGG/GT
44	Ob39	Ob43	AG/CTCTG	AGGG/GT
45	Ob40	Ob44	AG/CTCTA	AGGG/GT
46	Ob41	Ob45	AG/CCATG	AGGG/GT
47	Ob42	Ob46	AG/CCCTG	AGGG/GT
48	ObE	Ob47	AG/CCCTG	AGGG/GT
49	ObF	Ob48	AG/CCCTG	AGGG/GT
50	ObG	Ob49	AG/CCCTG	AGGG/GT
51	ObH	Ob50	AG/CCCTG	AGGG/GT
52	ObI	Ob51	AG/CCCTG	AGGG/GT
53	ObJ	Ob52	AG/AGGCC	AAGG/GT
54	Ob43	Ob53	AG/CCCCA	ACAG/GT
55	Ob44	Ob54	AG/CTGCG	CGTG/GT
56	Ob45	Ob55	AG/TCCCC	ACAG/GT
57	Ob46	Ob56	AG/AGCCT	AGAG/GT
58	Ob47 part 1	Ob57 part 1	AG/CTCCT	CAAG/GT
59	Ob47 part 2	Ob57 part 2	AG/GTGAC	ACAG/GT
60	Ob48	Ob58	AG/CCAAG	CGAG/GT
61	Ob49	Ob59	AG/GCTGG	CTGG/GT
62	Ob50	Ob60	AG/GCCTC	CTTG/GT
63	Ob51	Ob61	AG/CAGAG	CAGG/GT
64	non mod 3 part 1	non mod 3 part 1	AG/TGGCA	CAAG/GT
65	non mod 3 part 2, ob 52, non mod 4	non mod 3 part 2, ob 62, non mod 4	AG/GTTGG	GCTG/GT
66	Ob53	Ob63	AG/CCCCA	GACT/GT

Table 1. (Contd.)

Exon Number	Previous nomenclature	New nomenclature	5′ splice	3′ splice
67	non mod 5 part 1	non mod 5 part 1	AG/TGACA	TCCAG/GT
68	non mod 5 part 2, Ob54 part 1	non mod 5 part 2, Ob64 part 1	AG/CCGGG	AAAG/GT
69	Ob54 part 2	Ob64 part 2	AG/GCTAC	CGAG/GT
70	non mod 6 part 1	non mod 6 part 1	AG/GCCCT	TCAG/GT
71	non mod 6 part 2, Ob55 part 1	non mod 6 part 2, Ob65 part 1	AG/ATGTG	GAAG/GT
72	Ob55 part 2	Ob65 part 2	AG/GACGC	GGCC/GT
73	non mod 7 part 1	non mod 7 part 1	AG/TGCCT	CAGGG/GT
74	non mod 7 part 2	non mod 7 part 2	AG/GACCA	AAAG/GT
75	non mod 7 part 3	non mod 7 part 3	AG/AAGAG	CAGG/GT
76	non mod 7 part 4	non mod 7 part 4	AG/CCAAG	CGAG/GT
77	SH3	SH3	AG/ATCTT	CAAG/GT
78	DH part 1	DH part 1	AG/CTGTC	CTGAG/GT
79	DH part 2	DH part 2	AG/CTCTG	AGCAG/GT
80	DH part 3	DH part 3	AG/CTTCCT	CAAG/GT
81	DH part 4	DH part 4	AG/AAATA	GAAG/GT
82	DH part 5, PH part 1	DH part 5, PH part 1	AG/GAGCT	CCAG/GT
83	PH part 2	PH part 2	AG/GGCCA	GAAG/GT
84	PH part 3	PH part 3	AG/CTGAG	TGGC/GT
85	Ob56 part 1	Ob66 part 1	AG/GGCCC	AAAG/GT
86	Ob56 part 2	Ob66 part 2	AG/ATGGG	CAAG/GT
87	Ob57 part 1	Ob67 part 1	AG/TCCCA	ATCAG/GT
88	Ob57 part 2	Ob67 part 2	AG/GTGGT	TGAG/GT
89	Ob57 part 3, non mod COOH part 1	Ob67 part 3, non mod COOH part 1	AG/CTGGT	GAAG/GT
90	non mod COOH part 2	non mod COOH part 2	AG/TCACT	ATCAT/GT
91	non mod COOH part 3	non mod COOH part 3, 3′ UTR	AG/AGAAG	///////
92		kinase linker part 1	AG/ACACC	ACCTG/GT
93		kinase linker part 2	AG/CAGCC	ATTCG/GT
94		kinase linker part 3	AG/GATGG	CGCAC/GT
95		kinase 1 Ob68 part 1	AG/GGCCC	ACAAG/GT
96		kinase 1 Ob68 part 2	AG/GACAG	TGGGG/GT
97		kinase 1 Ob68 part 3	AG/GGGAC	GGAAG/GT
98		kinase 1 catalytic domain part 1	AG/GGGCG	GAGCT/GT
99		kinase 1 catalytic domain part 2	AG/GTGCT	CCGAG/GT
100		kinase 1 catalytic domain part 3	AG/GTCAA	TAAAG/GT
101		kinase 1 catalytic domain part 4	AG/CCCTC	ATTTG/GT
102		kinase 1 catalytic domain part 5	AG/GGCCA	CTCAG/GT
103		kinase 1 catalytic domain part 6	AG/CCTGA	CCTCA/GT
104		kinase 1 catalytic domain part 7	AG/GGCCC	TCCTG/GT
105		kinase 1 catalytic domain part 8	AG/AAATC	GGCAG/GT
106		inter-kinase linker, ZnF	AG/CGTTC	GGAAG/GT
107		inter-kinase linker, part 2	AG/GTCTG	CCGAG/GT
108		kinase 2 Ob69 part 1	AG/CGCCG	CAAAG/GT
109		kinase 2 Ob69 part 2	AG/ACGGA	GGCAG/GT
110		kinase 2 Ob70 part 1	AG/AGCGC	AGAAG/GT
111		kinase 2 Ob70 part 2	AG/GCGGC	CCTGG/GT
112		kinase 2 catalytic domain part 1	AG/CCTCT	CAGAG/GT
113		kinase 2 catalytic domain part 2	AG/GGGCC	GAGAG/GT
114		kinase 2 catalytic domain part 3	AG/GGCCT	CATGG/GT
115		kinase 2 catalytic domain part 4	AG/CTCCA	ATCAT/GT
116		kinase 2 catalytic domain part 5	AG/GCTGA	CCCTG/GT
117		kinase 2 catalytic domain part 6, STOP	AG/GGGCC	////////
118		predicted exon		
119		predicted exon		

Previous nomenclature is listed in comparison to the revised nomenclature based on the complete gene structure. The splice site junctions at the 5′ and 3′ end of the exons are shown, with the first and last complete coding triplets of the exon underlined.

sites for GATA-1 and -2, p300, MZF1, HSF2, and AML-1a. Transcription of the obscurin kinases could therefore be driven by transcriptional activity within this short segment, allowing for its expression in non-muscle tissues such as brain.

Discussion

The complete gene encoding the giant striated muscle protein obscurin OBSCN reveals a structure that is largely a close reflection of the modularity of the

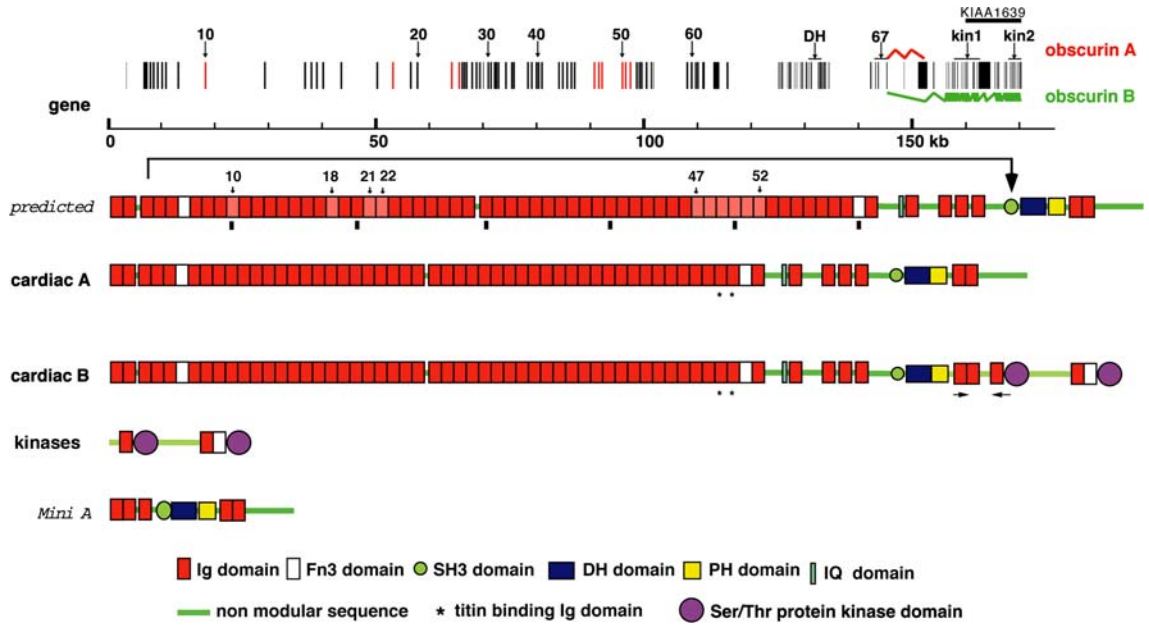

Fig. 1. The schematic structure of the obscurin gene (OBSCN), and the derived protein domain patterns. A total of 67 Ig and Fn3 domains are encoded on 91 exons for obscurin A; differential splicing can lead to generation of obscurin B by the fusion of two kinase domains and a further two Ig and one Fn3 domain. The alternative splice paths leading to obscurin A (red) and obscurin B (green) are highlighted at the schematic exon structure representation. The extent of the human brain obscurin kinase cDNA KIAA1639 is marked. The predicted (italics) complete sequence of obscurin A contains Ig-encoding exons (highlighted in red in the schematic gene structure, and domains shaded in pink with domain numbers on top) not found in the experimentally determined complete cardiac cDNA (cardiac A), which follow the same splice rules as the rest and can thus be spliced in consecutively. Skipping of exons 90 and 91 links Ob67 to the kinase exons (cardiac B). A putative mini-obscurin could be generated by fusion of exons 2 and 77 (mini A) as indicated by the long arrow. The separately expressed kinase variant is shown below this isoform. The position of every tenth domain in the complete predicted domain pattern is indicated by lines. Arrows denote the positions of the primers used in the RT-PCR for the obscurin-kinase linking region.

encoded protein. Of the 67 Ig and Fn3 domains, 60 are encoded on single exons reflecting their protein domain boundaries. Complex splicing occurs only towards the 3′ end of the gene, where the signalling domains of obscurin are encoded on 8 exons for the SH3-DH-PH trio, and 8 and 6 exons for the two kinase domains, respectively. The splice donor and acceptor sites for the tandem Ig domains, which form the bulk of obscurin domains, are entirely compatible from Ob3 to Ob66. This suggests that the potential for variability in obscurin by differential splicing, and hence the number of potential isoforms, is enormous. Differential splicing might also separate the two titin-binding Ig-domains Ob58 and Ob59, which would abolish their ability to interact with titin. The 10 Ig domains not found in the complete cardiac cDNA (previously denoted A–J) follow that same pattern, indicating that they can be consecutively spliced in a similar manner. The predicted full-length peptide of obscurin A therefore contains the full 67 domains. The 5′ splice site of exon 78 encoding the SH3 domain is also in frame 1. Another possibility is the large-scale skipping of exons in the tandem Ig region. This potential for the generation of smaller isoforms may explain the occasional observation of smaller bands in Western blots and will need to be experimentally addressed.

Kinase 1 as well as kinase 2 shows homology to the intrasterically regulated myosin light chain kinase-like kinases (Wilmanns *et al.*, 2000); however, the degree of sequence homology to *bona fide* MLCKs is quite weak (33% identity with human smooth muscle MLCK for kinase 1 and 26% for kinase 2), suggesting that they are likely to be functionally distinct. In fact, their relation to titin kinase (28% identity for kinase 1 and 27% for kinase 2) is not significantly different, and the conservation between the two kinases with 30% identity suggests that the two have distinct functions. This is further supported by closer inspection of crucial residues involved in kinase activation and regulation. Whereas kinase 1 contains a C-terminal predicted regulatory domain with a sequence likely to function as a calmodulin-binding domain (underlined in Figure 2a), this is not the case for kinase 2. Interestingly, a shared feature between kinase 2 and titin is an unusual arginine residue at position +2 from the catalytic base, aspartate 127 in the titin kinase structure ((Mayans *et al.*, 1998), L*D*LRSENMI in kinase 2; F*D*IRPENII in titin). In almost all serine/threonine kinases including kinase 1, this residue is lysine. In titin, this arginine is involved in a unique autoinhibition mechanism, where a tyrosine from the P+1 loop is coordinated against the catalytic aspartate by this arginine (Mayans *et al.*, 1998). The presence of an arginine in this conspicuous position in kinase 2 (Figure 2b) suggests further that the regulation and hence function of the two obscurin kinases is likely to be different. Until their function has been elucidated, it seems inappropri-

Fig. 2. (a) The experimental amino acid sequence of the differentially spliced junction between Ob67 and the obscurin kinase linker identified by RT-PCR. The tandem Ig domains Ob66 and Ob67 are highlighted in blue with their signature WYK motif underlined; the same motif is underlined in Ob68, the Ig domain preceding the catalytic kinase domain (green). The linker between Ob67 and Ob68 (brown) contains an N-terminal region rich in proline, threonine and serine, suggesting potential phosphorylation sites for SP-directed kinases, whereas the C-terminal region contains several bulky hydrophobic residues (3 tryptophane, 2 tyrosine) and is likely to be structured and potentially involved in protein interactions. The catalytic aspartate in kinase 1 is highlighted in red and the predicted autoinhibitory domain in italics; a putative basic amphipatic sequence might serve as a calmodulin-binding site. (b) Sequence comparison between obscurin kinase 2 (top lines) and titin kinase (bottom lines). Note the unusual arginine residue at +2 from the catalytic aspartate (highlighted in blue) that coordinates an inhibitory tyrosine from the extended activation loop (highlighted in red). Obscurin kinase 2 contains a tyrosine 7 residue N-terminal to that of titin in the activation loop, structural information will be required to assess whether there is an analogous autoinhibition mechanism.

ate to classify them as myosin light chain kinases as suggested (Sutter *et al.*, 2004).

The existence of multiple promoter consensus sites in the 1.3 kb genomic sequence between exons 91 and 92 could explain the expression of this region in non-muscle tissues, as evidenced by the expression in brain of KIAA1639. A similar internal promoter has been found and experimentally analysed in the possible invertebrate analogue from *C. elegans*, unc-89 (Small *et al.*, 2004). In contrast to *C. elegans*, there is as yet no indication for the transcription of isolated obscurin kinases from this putative promoter, as the truncation of kinase 1 in KIAA1639 might also reflect an incomplete cDNA. Our RT-PCR from human brain cDNA showed that kinase 2 is expressed in this tissue (not shown). However, it is unclear which exons would serve as 5′ UTR for translation initiation for the separate obscurin kinase expression, unless the predicted

exons within the kinase linker that are not represented in the muscle isoform obscurin B could serve that purpose. Potentially, the truncation of kinase 1 would also go along with this ectopic expression similar to the inactive, truncated kinases in *C. elegans* (Small *et al*., 2004). The non-muscle expression suggests, at any rate, that the obscurin kinases participate in a more ubiquitous signalling pathway, possibly related to the association of obscurin to derivatives of the endoplasmic reticulum.

Future analysis will reveal how the transcription and differential splicing of this segment is regulated, and what biological role the two kinases and their large inter-kinase insertion play in muscle and non-muscle tissue. The complete gene structure may now also be used to search for disease-associated mutations in the obscurin gene.

Acknowledgements

AF is supported by the Volkswagen Foundation, SI and MG by an International Appointment Initiative Award of the Medical Research Council of Great Britain.

References

Bagnato P, Barone V, Giacomello E, Rossi D and Sorrentino V (2003) Binding of an ankyrin−1 isoform to obscurin suggests a molecular link between the sarcoplasmic reticulum and myofibrils in striated muscles. *J Cell Biol* **160:** 245–253.

Bang ML, Centner T, Fornoff F, Geach AJ, Gotthardt M, McNabb M, Witt CC, Labeit D, Gregorio CC, Granzier H and Labeit S (2001) The complete gene sequence of titin, expression of an unusual approximately 700-kDa titin isoform, and its interaction with obscurin identify a novel Z-line to I-band linking system. *Circ Res* **89:** 1065–1072.

Benian GM, Tinley TL, Tang X and Borodovsky M (1996) The *Caenorhabditis elegans* gene *unc*-89, required for muscle M-line assembly, encodes a giant modular protein composed of Ig and signal transduction domains. *J Cell Biol* **132:** 835–848.

Borisov AB, Kontrogianni-Konstantopoulos A, Bloch RJ, Westfall MV and Russell MW (2004) Dynamics of obscurin localization during differentiation and remodeling of cardiac myocytes: obscurin as an integrator of myofibrillar structure. *J Histochem Cytochem* **52:** 1117–1127.

Borisov AB, Raeker MO, Kontrogianni-Konstantopoulos A, Yang K, Kurnit DM, Bloch RJ and Russell MW (2003) Rapid response of cardiac obscurin gene cluster to aortic stenosis: differential activation of Rho-GEF and MLCK and involvement in hypertrophic growth. *Biochem Biophys Res Commun* **310:** 910–918.

Burge C and Karlin S (1997) Prediction of complete gene structures in human genomic DNA. *J Mol Biol* **268:** 78–94.

Ferrari MB, Ribbeck K, Hagler DJ and Spitzer NC (1998) A calcium signaling cascade essential for myosin thick filament assembly in *Xenopus* myocytes. *J Cell Biol* **141:** 1349–1356.

Kontrogianni-Konstantopoulos A, Catino DH, Strong JC, Randall WR and Bloch RJ (2004) Obscurin regulates the organization of myosin into A bands. *Am J Physiol Cell Physiol* **287:** C209–C217.

Kontrogianni-Konstantopoulos A, Jones EM, Van Rossum DB and Bloch RJ (2003) Obscurin is a ligand for small ankyrin 1 in skeletal muscle. *Mol Biol Cell* **14:** 1138–1148.

Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP and Bork P (2004) SMART 4.0: towards genomic data integration. *Nucleic Acids Res* **32:** D142–D144.

Mayans O, Van der Ven P, Wilm M, Mues A, Young P, Fürst DO, Wilmanns M and Gautel M (1998) Structural basis of the activation of the titin kinase domain during myofibrillogenesis. *Nature* **395:** 863–869.

Nagase T, Kikuno R, Nakayama M, Hirosawa M and Ohara O (2000) Prediction of the coding sequences of unidentified human genes. XVIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins *in vitro*. *DNA Res* **7:** 273–281.

Russell MW, Raeker MO, Korytkowski KA and Sonneman KJ (2002) Identification, tissue expression and chromosomal localization of human Obscurin-MLCK, a member of the titin and Dbl families of myosin light chain kinases. *Gene* **282:** 237–246.

Shapiro MB and Senapathy P (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* **15:** 7155–7174.

Small TM, Gernert KM, Flaherty DB, Mercer KB, Borodovsky M and Benian GM (2004) Three new isoforms of *Caenorhabditis elegans* UNC−89 containing MLCK-like protein kinase domains. *J Mol Biol* **342:** 91–108.

Sutter SB, Raeker MO, Borisov AB and Russell MW (2004) Orthologous relationship of obscurin and Unc-89: phylogeny of a novel family of tandem myosin light chain kinases. *Dev Genes Evol* **214:** 352–359.

Wilmanns M, Gautel M and Mayans O (2000) Activation of calcium/calmodulin regulated kinases. *Cell Mol Biol (Noisy-le-grand)* **46:** 883–894.

Young P, Ehler E and Gautel M (2001) Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor protein involved in sarcomere assembly. *J Cell Biol* **154:** 123–136.

Young P and Gautel M (1999) Obscurin, a novel giant muscle protein and a ligand of Z-disk titin. *J Muscle Res Cell Motil* **20:** 104.