Functional protein networks unifying limb girdle muscular dystrophy

Morrée, A. de

Chapter 5
Self-regulated alternative splicing at the AHNAK locus

Antoine de Morrée[1], Marjolein Droog[1], Ilona Bisschop[1], Antonietta Impagliazzo[1], Rune R Frants[1], Rinse Klooster[1], Silvère M van der Maarel[1]

[1] Department of human genetics, Leiden University Medical Center, Leiden, The Netherlands

-Chapter 5 has been submitted for publication-

Abstract

AHNAK is an ubiquitously expressed 700 kDa protein involved in cyto-architecture and calcium signaling. It is secondarily reduced in muscle of Dysferlinopathy patients and accumulates in muscle of Calpainopathy patients, both affected by a different form of muscular dystrophy. AHNAK directly interacts with Dysferlin. This interaction is lost upon cleavage of AHNAK by the noble protease CAPN3, explaining the molecular observations in muscle of these patients. Currently, little is known of AHNAK regulation. Here we describe the self-regulation of multiple mRNA transcripts emanating from the *AHNAK* locus in muscle cells. We show that the *AHNAK* gene consists of a giant 17kb exon flanked on both sides by multiple small exons, which together form a continuous open reading frame. This genetic structure is shared by two other genes, AHNAK2 and Periaxin, which all share a common ancestor. Two major AHNAK transcripts are differentially expressed during muscle differentiation that encode for a small (17kDa) and a large (700kDa) protein isoform. We show that the small AHNAK isoform traffics between the cytoplasm and the nucleus to regulate splicing of mRNA transcripts derived from its own locus. Thus AHNAK constitutes a novel mechanism in post-transcriptional control of gene expression.

Author summary

AHNAK is a 700 kDa giant protein. It directly interacts with Dysferlin and CAPN3, two proteins that when mutated individually can cause Limb Girdle Muscular Dystrophy. AHNAK is thought to be important for normal muscle physiology as well, but it is not clear how AHNAK is regulated. According to literature the *AHNAK* gene consists of a single giant exon. We investigated the *AHNAK* gene and its expression in skeletal muscle and identified a novel mechanism of self-regulation through alternative mRNA splicing. We show that the *AHNAK* gene is part of a larger family that includes AHNAK2 and Periaxin, and is characterized by a giant exon that is flanked by multiple small exons. These exons allow for alternative splicing. AHNAK, like Periaxin, expresses two isoforms, a large 700 kDa and small 17 kDa protein. These proteins interact in the cytoplasm, but the small AHNAK is also present in the nucleus. During muscle differentiation the small AHNAK is strongly increased thereby establishing a positive feed back loop to regulate mRNA splicing of its own locus.

Introduction

AHNAK (*AHNAK nucleoprotein*, MIM*103390) is a 700 kDa giant protein which is ubiquitously expressed in vertebrates. It was first described in neuroblastoma cells [227] but has since been found expressed in many other cell types and is common in lining epithelium [89,90]. AHNAK has been observed at many different sub-cellular locations, depending on cell type and condition. It is found in the cell nucleus (neuroblastoma) [226,239] or cytosol (MCF7 and HeLa) [19,91] of cells growing at low confluency but can rapidly translocate to the plasma membrane upon the formation of cell-cell contacts [19,239]. Nuclear transfer is dependent on a C-terminal Protein Kinase B (PKB) phosphorylation site within a nuclear localization signal at the C-terminus [239]. The mouse orthologue of AHNAK, Desmoyokin, was first described as a desmosome-associated protein [105]. In addition AHNAK was found as a protein component of enlargosomes; a secretory vesicle implicated in calcium-sensitive plasma membrane enlargement of PC12-27 cells [52,53].

The AHNAK protein has a tripartite domain structure, consisting of a relatively short N-terminus (251 amino acids), and a long C-terminus (1002 amino acids) separated by a large body of repeats (43 units, the standard unit being 128 amino acids long) [227]. Repeat proteins are not uncommon and many are hypothesized to play a role in mechanical strengthening (eg Dystrophin) and to function as scaffolding proteins. The exact function of AHNAK or its respective protein domains is still under study.

Recently an AHNAK knock-out mouse was described [146,148]. This mouse has no overt phenotype leading to the discovery of the highly similar AHNAK2 (*AHNAK nucleoprotein 2*, MIM*608570) protein which is probably largely redundant to AHNAK [146]. AHNAK2 also has a tripartite domain structure, consisting of an N- and C-terminus separated by repeat units. These repeat units are highly homologous to those of AHNAK, and antibodies recognizing the repeats can react with both proteins [146]. The N- and C-terminal sequences are more divergent. Recent studies in the AHNAK knock-out mouse revealed a mild phenotype in T-cells [176]. AHNAK was found to participate in L-Type Voltage-Gated Calcium Channel (LVGCC) trafficking and thus to be of importance in the second phase of calcium-dependent T-cell activation [175,176].

AHNAK is implicated in cardiomyopathy [100,117]. The C-terminus of AHNAK has been shown to directly interact with the β2 subunit of the LVGCC. This interaction can influence the activity of the channel [4,100,117]. Two LVGCC binding sites are located in AHNAK, and they appear to have opposing effects [117]. Interestingly, one SNP in the *AHNAK* locus was described that severely impaired the LVGCC modulating capacity of AHNAK suggesting its importance in

cardiac physiology. In a genetic study with cardiac arythmia patients, however, this SNP showed no correlation with phenotype [117].

In addition, AHNAK is implicated in Limb-Girdle Muscular Dystrophy (LGMD) [118,119]. We have previously shown that AHNAK interacts with the skeletal muscle proteins Dysferlin (*DYSF*, MIM*603009) and Calpain 3 (*CAPN3*, MIM*114240) [118,119]. Dysferlin is essential for calcium-dependent membrane repair in muscle, and mutations in Dysferlin cause a spectrum of muscular dystrophies including LGMD2B (MIM#253601). The C-terminus of AHNAK directly interacts with the C2A domain of Dysferlin [119], and AHNAK redistributes to cytoplasm together with Dysferlin in regenerating rat muscle [119]. The skeletal muscle protease CAPN3 is part of the same protein complex. Mutations in CAPN3 cause a similar muscular dystrophy phenotype as Dysferlin (LGMD2A, MIM#253600). CAPN3 can directly interact with AHNAK at multiple sites, and cleave it [118]. This cleavage strongly reduces the binding affinity of AHNAK for Dysferlin [118]. Moreover, in patient muscle cryosections absence of Dysferlin causes a secondary reduction in AHNAK, while loss of CAPN3 results in an increase of AHNAK at the muscle membrane. Overall, data suggests that AHNAK is of importance for muscle membrane integrity [118].

Little is known about the *AHNAK* gene locus and its transcriptional regulation. AHNAK has been considered to be a 17kb large intron-less gene [226,227]. However, this contrasts with genome databases such as Ensembl where multiple transcripts are predicted. In addition, AHNAK transcription in muscle has not been explored yet. Hence, we were interested in the transcriptional regulation of AHNAK. In this investigation we show that the *AHNAK* gene consists of a giant exon flanked by multiple small exons. Moreover, these exons allow for alternative splicing and result in a long and a short protein isoform that are differentially regulated in muscle. This genetic structure is evolutionary conserved and shared by two other genes, *AHNAK2* and Periaxin (*PRX*, MIM*605725), which our data indicate to be phylogenetically related. Moreover, the small and large isoform together constitute a novel mechanism of genetic self-regulation. We show that S- and L-AHNAK heterodimerize in the cytosol and that S-AHNAK can translocate into the nucleus. There it is physically in the spliceosome and regulates splicing of AHNAK mRNA. Parallels with Periaxin suggest that this regulatory mechanism is common for the *AHNAK* gene family.

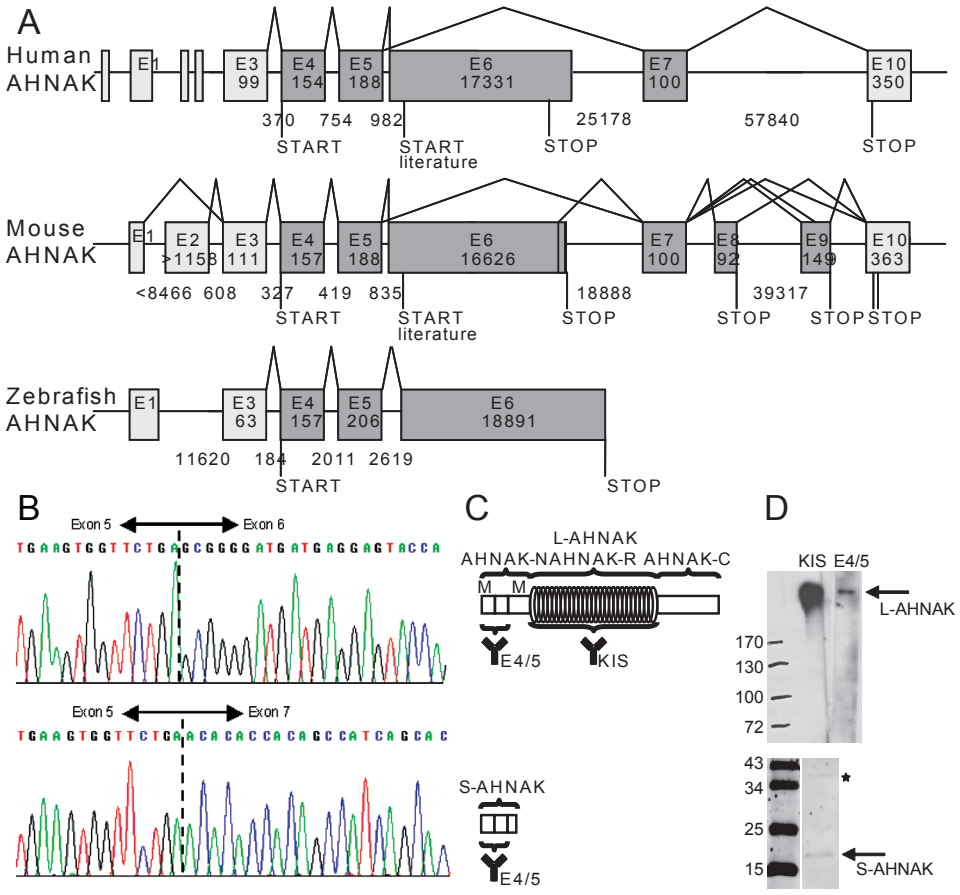

Figure 1: Transcripts from the AHNAK gene in man, mouse and zebrafish. A) Schematic representation of the AHNAK gene in man, mouse and zebrafish. Exons are boxed (E) and numbered. Dark grey boxes form the open reading frame. The length in nucleotides is given. Horizontal connecting lines represent introns, bridging lines are confirmed exon-exon boundaries. Start and stop codons are marked, and for mouse and man the start codon mentioned in literature is also given. B) Exemplary sequence reads for Exon5-Exon6 and Exon5-Exon7 boundaries in man. C) Schematic representation of L- and S-AHNAK proteins, and the available antibodies. L-AHNAK consists of three domains, an N-terminus (N), a body of repeat units (R), and a C-terminus (C). D) Western blot probed with E4/5 antibody and the published KIS antibody. Both antibodies detect a 700 kDa band (upper panel). E4/5 also detects a 17 kDa (lower panel) band. The asterisk marks an S-AHNAK dimer.

Results

The AHNAK gene encodes for a long and a short isoform

To reconstruct the *AHNAK* locus we retrieved sequence information from the Ensembl genome browser. We used the NCBI computer algorithms AceView and Blast-n, which link to databases of reported sequence clones such as ESTs, or expression clones (the repeat containing *AHNAK* gene is in many genomes poorly annotated or incomplete). Thus we could assemble the *AHNAK* gene in man, mouse and zebrafish (Figure 1A). In man, the *AHNAK* locus spans >113 kb. It consists of a giant 17kb exon (E6 in Figure 1A), which encodes for the full open reading frame encoding for the AHNAK repeat protein described in literature [227]. Interestingly, this giant exon is flanked on both sides by multiple small exons (each ~150bp). The two smaller exons upstream of the giant exon (E4 and E5) preserve the open reading frame and relocate the translation start codon 750bp (250 amino acids) upstream of the reported start codon. Moreover, alternative transcripts have been submitted to these databases that exclude the giant exon, in exchange for a small exon (E7) downstream of the giant exon E6. Such transcripts would preserve the open reading frame of exons E4 and E5, and result in a small AHNAK isoform. Thus, the *AHNAK* gene yields two putative isoforms: a large transcript with an 18kb open reading frame and a small transcript with a 0.5kb open reading frame. This genetic structure is conserved in rodents (Figure 1A), canine, and bovine species. In rodents, three alternative small AHNAK isoforms are observed, with facultative exons E8 and E9 within the open reading frame. In zebrafish (*Danio rerio*) the giant exon and the smaller upstream exons are also found, but exons downstream of the giant exon could not be identified (Figure 1A). Summarizing we conclude that the *AHNAK* gene consists of multiple evolutionary conserved exons that might allow for alternative splicing.

To challenge this in silico analysis we designed exon-exon boundary spanning primers to confirm the existence of the different AHNAK transcripts in man, mouse and zebrafish. We generated cDNA from human and mouse skeletal muscle tissue, and zebrafish embryo's, and amplified the predicted products by PCR. Direct sequencing confirmed the expression of all splice variants depicted in Figure 1A, and exemplary sequence outputs for the exon-exon boundaries between E5-E6 (giant transcript) and E5-E7 (small transcript) are shown in Figure 1B for human AHNAK. Clone IDs and splice sites for mouse and human AHNAK are included in supplemental Table S1. With 3' prime RACE experiments we could confirm the two poly-adenylation sequences in the *AHNAK* gene (Figure 1A and Table S1). Finally, CAGE data [112] confirmed that the major transcription initiation site is at Exon 1.

The giant exon encodes for a tripartite protein with a large number of repeats (Figure 1C). The additional two exons would add another 250 amino acids to the

N-terminus of the protein. This novel N-terminal peptide sequence encodes for a predicted PDZ domain and contains a putative Nuclear Export Signal (NES). The short isoform shares this NES and PDZ domain (Figure S1). To confirm that the exons 4 and 5 encoded sequence is translated into protein we analyzed human protein extracts on western blot (Figure 1D). We could detect a 700 kDa band with the well established polyclonal KIS antibody (recognizing the repeat units encoded by the giant exon) as was shown previously [89]. A protein band of similar size was detected with an commercial monoclonal AHNAK antibody raised against the N-terminal 100 amino acids encoded by the small exons E4 and E5 (Figure 1D). This strongly indicates that the open reading frame of the giant AHNAK protein includes the two small exons upstream of the giant exon, as was suggested by the gene expression data. Moreover, with the same N-terminal antibody we could also detect a 17kDa protein band, which corresponds to the predicted size of the putative small AHNAK isoform. We conclude that the AHNAK gene gives rise to at least two different transcripts that each contain an open reading frame and encode for a large AHNAK protein of 700kDa (L-AHNAK) and a small AHNAK protein of 17kDa (S-AHNAK), respectively (Figure 1C).

Three AHNAK-like genes exist in mammals

AHNAK2 is highly similar to AHNAK. We therefore wondered whether other AHNAK-like genes share a similar genetic structure. We retrieved AHNAK-like sequences using the NCBI tools psi-blast and blast-p. We consistently retrieved two significant hits in man and mouse: AHNAK2 and Periaxin. Both genes encode for proteins that have previously been suggested to be similar to AHNAK [146,220]. When we assembled and in silico annotated the sequence information on these genes, we observed that, like AHNAK, both genes contain a giant exon that is flanked upstream by multiple small exons (Figure S2). The small exons form a continuous open reading frame with the giant exon, and a large protein isoform has been described in literature for both genes [92,146,220]. Similar to *AHNAK*, both *AHNAK2* and *PRX* encode for a tripartite protein structure. AHNAK2 consists of repeat units (24*165 amino acids) flanked by a N- and C-terminus [146]. The AHNAK2 protein shares 29% sequence identity with L-AHNAK, and is most similar at the repeat units. At the N-terminus it has a PDZ domain, as we identified in AHNAK. PRX contains a relatively small array of repeats (12x 26 amino acids) [92] and is in total smaller than L-AHNAK and AHNAK2. The translated protein sequence shares 42% sequence similarity and 24% identity with both AHNAK and AHNAK2. Intriguingly, also the *PRX* gene expresses a small isoform [79]. In this transcript the open reading frame is shifted through intron retention, resulting in the exclusion of the giant exon [79]. The corresponding small protein isoform

has been reported in literature but its function remains elusive [79]. For AHNAK2 no support for a small isoform (transcript or protein) could be found (Figure S2). However, taken together we conclude that man has three AHNAK-like genes that have a shared genetic structure suggesting a common regulatory mechanism.

A common ancestor for AHNAK-like genes

We were interested in the evolutionary conservation of the AHNAK protein family. Therefore we identified homologous protein sequences in other species where possible. As mentioned before we could detect three AHNAK-like genes in vertebrates, but none in non-vertebrates, indicating that the AHNAKs are a vertebrate invention. Zebrafish has five AHNAK-like genes, but this species is known for its genetic duplications. Given the enormous size of the proteins we started our analyses with a small manageable part. We observed that all three AHNAK-like proteins (AHNAK, AHNAK2 and Periaxin) have a conserved N-terminal PDZ domain, which we used to create an alignment with the webtool T-Coffee (Zebrafish AHNAK4 and 5 do not contain a distinct PDZ domain and were excluded from this analysis). We manually adjusted the alignment (Figure S3A) and generated a maximum likelihood tree (Figure S3B) [95,96]. This shows that AHNAK, AHNAK2 and PRX diverged very early, and that L-PRX and AHNAK2 are more similar to each other than L-AHNAK. The only outliers are L-PRX and AHNAK2 from zebrafish, with the PDZ domain of L-PRX being more similar to that of other AHNAK2 proteins. Zebrafish L-PRX is distant from both the L-PRX and AHNAK2 groups. We next repeated the maximum likelihood analysis with an alignment of the complete protein sequences, which were annotated with the computer algorithm GBlock to reduce size (Figure S3C) [39,241]. This resulted in a similar tree construction, again indicating that L-PRX and AHNAK2 are more alike than L-AHNAK. Zebrafish AHNAK4 and 5 are clear outliers. In this analysis Zebrafish L-PRX and AHNAK2 are close together, and in between the AHNAK2 and L-PRX groups. We used a Bayesian analysis [123] to construct a rooted tree (Figure S3D). L-AHNAK diverged first, followed by AHNAK2 and finally L-PRX. This indicates that L-AHNAK, AHNAK2 and PRX are derived from a common ancestor.

The AHNAK isoforms are differentially regulated during myogenic differentiation

We next investigated whether we could detect alternative splicing of the *AHNAK* locus. As we are primarily interested in the role of AHNAK in skeletal muscle, we differentiated IM2 mouse myoblasts and isolated RNA at different time points during differentiation. A quantitative PCR showed that the large isoform of AHNAK (L-AHNAK) is expressed at high levels throughout differentiation (Figure 2). The

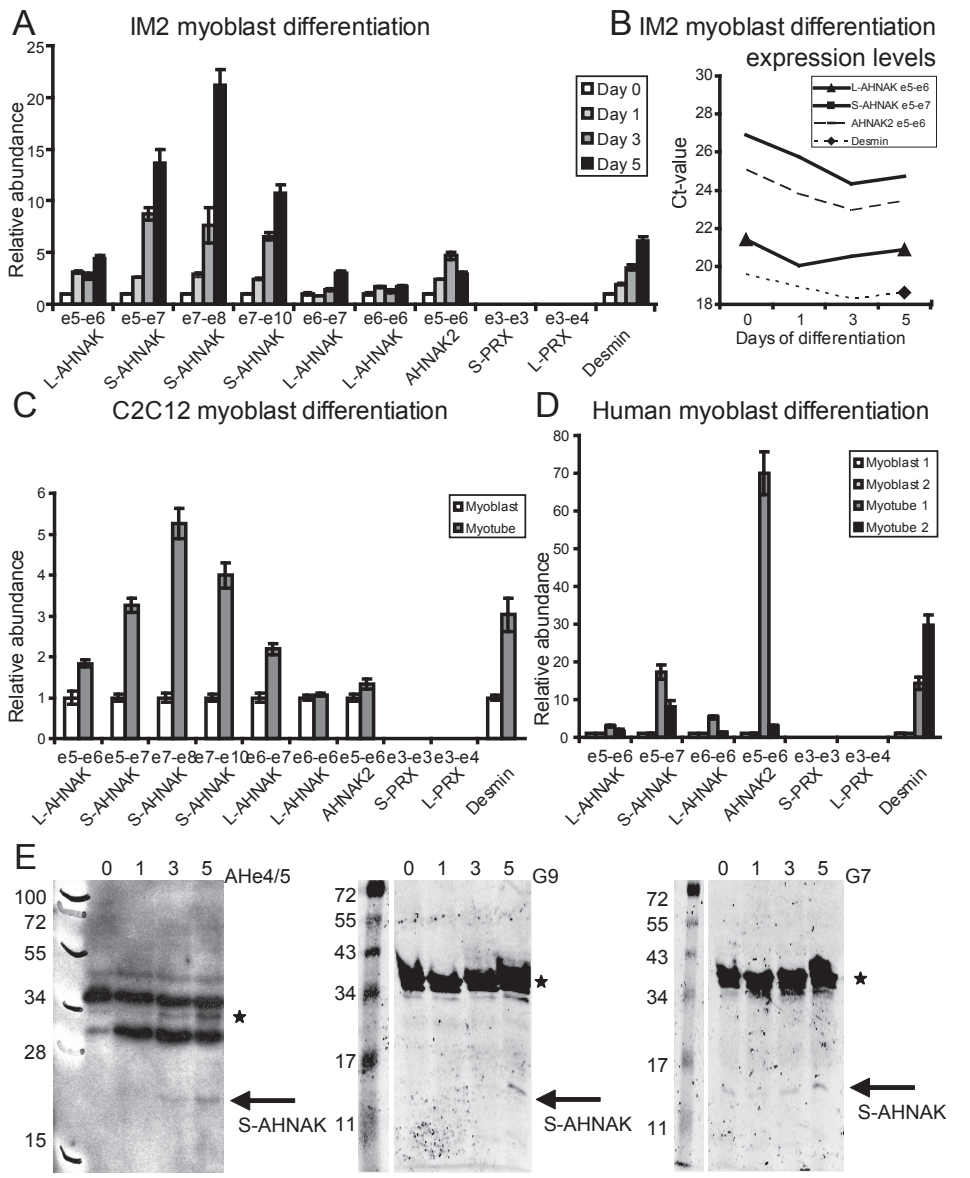

Figure 2: AHNAK is differentially spliced during myogenic differentiation. A) IM2 cells were differentiated and mRNA levels at day 0, 1, 3 and 5 of differentiation were measured by quantitative RT-PCR. GAPDH was used as reference gene. Error bars reflect relative standard deviations of three experiments. S-AHNAK mRNA expression is strongly increased during differentiation. B) Ct values were plotted, showing that even after differentiation S-AHNAK mRNA expression levels are ~20 fold lower than L-AHNAK. AHNAK2 mRNA levels are intermediate. C) C2C12 myoblasts were differentiated and relative mRNA levels were measured before and

small isoform (S-AHNAK) however, strongly increases during differentiation. Nevertheless, even in differentiated myotubes L-AHNAK mRNA expression is still 10 fold higher than S-AHNAK (Figure 2B). AHNAK2 was similarly increased during myoblast differentiation. We could not detect *PRX* expression in muscle cells, consistent with a previous report that *PRX* expression is restricted to the peripheral nervous system [92]. We repeated this experiment in C2C12 mouse myoblasts (Figure 2C) and human primary myoblasts (Figure 2C), and in all cases observed the same pattern: upregulation of S-AHNAK while expression levels of L-AHNAK remain constant. We could confirm our observations on western blot with the commercial antibody against the exon4/5 encoded sequence: an upregulation of a 17 kDa S-AHNAK band during muscle differentiation (Figure 2E). Although we were unable to achieve isoform specific knock-down we successfully selected two heavy chain antibody fragments that can specifically recognize S-AHNAK on western blot and in ELISA (Figure S4), and we could detect the 17 kDa band with these three independent antibodies, indicating its identity as S-AHNAK (Figure 2E). We therefore conclude that the S-AHNAK protein expression is increased during skeletal muscle myoblast differentiation.

S-AHNAK can oligomerize *in vitro*

We noticed on the western blots that a second band of approximately twice the size of S-AHNAK (34 kDa) was detected with all three antibodies. This suggested that S-AHNAK might oligomerize. To find support for S-AHNAK oligomerization we performed a GST pull-down experiment. We produced and purified recombinant GST and GST-fused S-AHNAK and used these proteins to pull down T7-tagged S-AHNAK, or T7-tagged L-AHNAK-R (4 repeat units, Figure 1C) as negative control. GST-fused S-AHNAK could efficiently pull down S-AHNAK, but not L-AHNAK-R, suggesting that S-AHNAK can self-interact (Figure 3). Because 75% of the S-AHNAK sequence is conserved in L-AHNAK, including the PDZ interaction domain, we next tested whether S-AHNAK could also bind L-AHNAK. We produced and purified T7-tagged L-AHNAK-N (amino acids 2-500) and repeated the pull-down experiment. GST-fused S-AHNAK could pull-down L-AHNAK-N, unlike GST (Figure

after differentiation. Exons in the S-AHNAK transcript are upregulated with differentiation. D) The experiment was repeated in two human primary skeletal muscle myoblast cell lines, with comparable results. E) IM2 protein extracts were analyzed on western blot with E4/5 antibody or S-AHNAK specific heavy chain antibody fragments G9 and G7 to detect the 17 kDa S-AHNAK protein, which is similarly increased with differentiation. The asterisk denotes an S-AHNAK dimer recognized by all three independent antibodies.

3). This indicates that S-AHNAK can also directly interact with the N terminus of L-AHNAK. To further investigate the nature of the AHNAK self-interaction we performed size-exclusion chromatography of purified recombinant S-AHNAK. 90% Of S-AHNAK runs as a complex of 75 kDa (Figure S5), which corresponds in size to a homotetramer. This interaction is independent of disulfide bonds as replacement of the single conserved cysteine in S-AHNAK did not change the chromatography

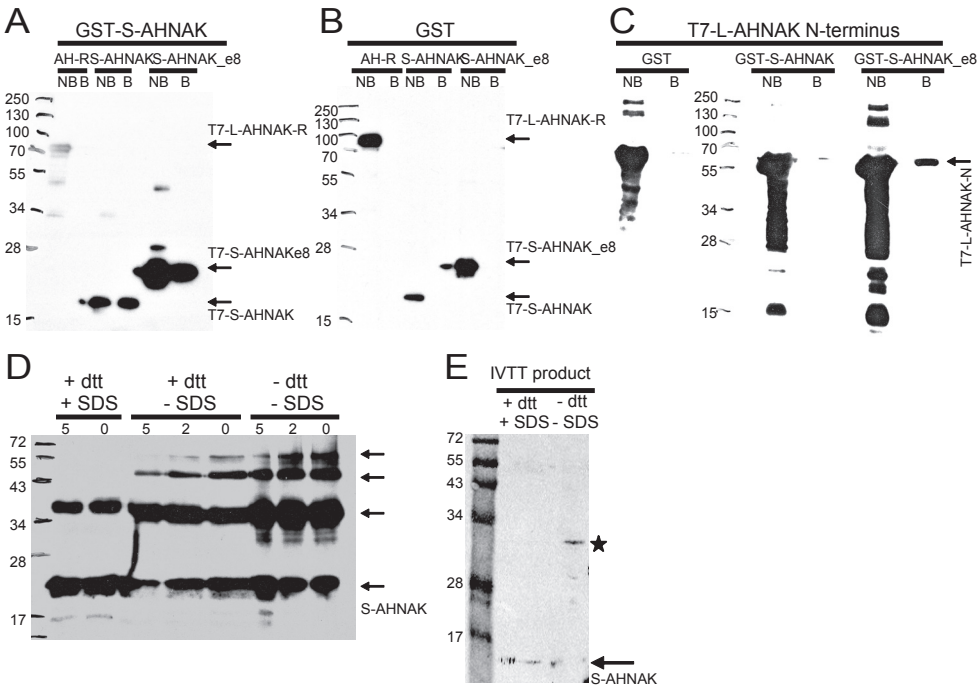

Figure 3: Direct interaction between AHNAK protein isoforms. A+B) GST-S-AHNAK can pull down T7-S-AHNAK and T7-S-AHNAK_e8, but not T7-L-AHNAK-R (A), while unfused GST cannot pull down any of the three proteins (B). C) Contrary to unfused GST, GST-S-AHNAK can also pull down T7-L-AHNAK-N. In all panels bound (B) and non-bound (NB) fractions are shown, detected with a T7 specific antibody. The arrows indicate the protein bands. D) extracts from the size-exclusion chromatography experiment were dissolved in sample buffer containing DTT and SDS (denaturing), SDS only, or neither component (semi-native), and boiled for 5, 2 or 0 minutes. Samples were separated on SDS-PAGE gels and analyzed by western blot with a T7 specific antibody. Arrows denote the monomeric and multimeric S-AHNAK protein bands. E) The S-AHNAK protein used in the VHH selections, which was produced by *in vitro* transcription-translation was dissolved in sample buffer containing SDS and DTT, or neither, and analyzed on western blot with a HIS6 specific antibody. The asterisk denotes the S-AHNAK dimer.

pattern (not shown). We next tested the complex of S-AHNAK and L-AHNAK-N (Figure S5) by mixing equimolar amounts of both proteins. This protein mixture runs as a single complex of 65kDa (Figure S5), which corresponds in size to a heterodimer of a single S-AHNAK molecule together with a single L-AHNAK-N molecule. When we analyzed the homotetrameric and heterodimeric protein complexes on western blot we could detect four distinct S-AHNAK protein bands under semi-native conditions in the absence of SDS and DTT in the sample buffer, which correspond in size to one to four copies of S-AHNAK (Figure 3D). By adding SDS and DTT and boiling, we could disrupt the oligomerization resulting in a double band corresponding to monomeric and dimeric S-AHNAK. When we repeated this for *in vitro* translated S-AHNAK we also observed a 34 kDa band under semi-native conditions (Figure 3e). We conclude that S-AHNAK can oligomerize with itself and with L-AHNAK *in vitro*. In addition this data indicates that the 34 kDa band detected in Figure 2 might correspond to multimeric S-AHNAK.

S-AHNAK colocalizes with L-AHNAK and with nuclear spliceosomal proteins

We proceeded with investigating the cellular localization of both AHNAK isoforms. We used the KIS antibody to detect endogenous L-AHNAK. In addition, because of the low expression levels of endogenous S-AHNAK below the detection limit of our antibodies, we used VSV-tagged S-AHNAK. S-AHNAK-VSV was transiently expressed in U2OS cells and human primary myoblasts and protein localization was studied by indirect immunofluorescent microscopy. As previously reported, endogenous L-AHNAK localizes to the cytosol and is found at the cell membrane, cell-cell contacts, on vesicles and at the Actin cytoskeleton (Figure 4A). S-AHNAK-VSV was also found in the cytosol, where it strongly colocalizes with L-AHNAK, consistent with the observation that recombinant L-AHNAK and S-AHNAK can directly interact. However, unlike L-AHNAK, S-AHNAK was also found in nuclear speckles. Further co-immunostaining experiments showed that nuclear S-AHNAK colocalizes with SC35 (Figure 4B) and PABPN1 (not shown) in speckles, which are distinct structures enriched for spliceosomal and RNA processing proteins. When we expressed recombinant S-AHNAK-VSV in human primary myoblasts a similar staining pattern was observed (Figure 4B).

To confirm that the endogenous S-AHNAK protein shows a similar subcellular localization we performed protein fractionation experiments of IM2 myoblasts and differentiated myotubes followed by western blot analysis of S-AHNAK. Figure 5A shows that S-AHNAK is in the cytosolic and the soluble nuclear fraction in myotubes. On the contrary, the 34 kDa band is exclusively present in the nucleus. This may suggest that monomeric S-AHNAK is cytosolic and the oligomer is

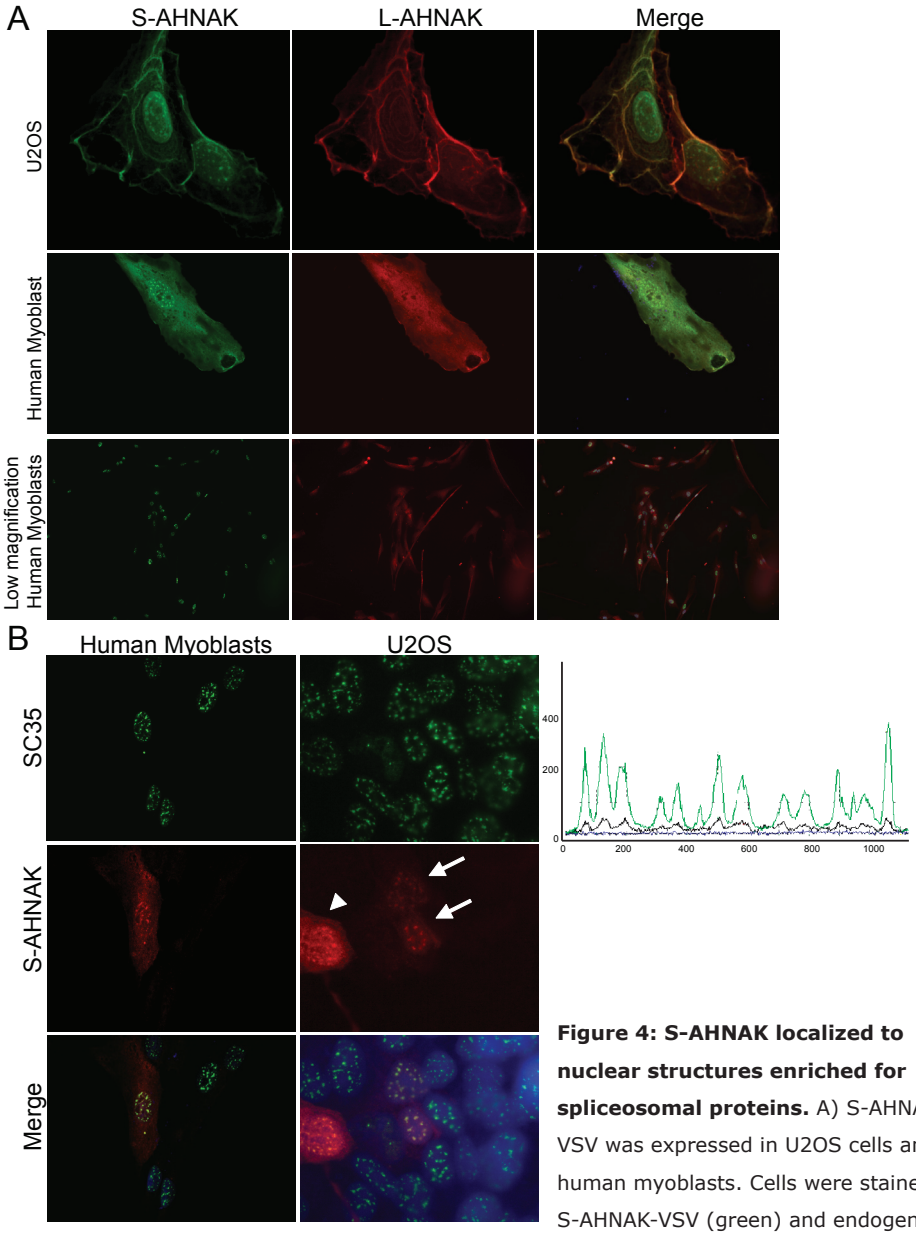

Figure 4: S-AHNAK localized to nuclear structures enriched for spliceosomal proteins. A) S-AHNAK-VSV was expressed in U2OS cells and human myoblasts. Cells were stained for S-AHNAK-VSV (green) and endogenous L-AHNAK (red). S-AHNAK and L-AHNAK colocalize outside the nucleus. S-AHNAK is also observed in nuclear structures. B) Nuclear S-AHNAK-VSV shows strong colocalization with endogenous SC-35, in both U2OS and human myoblast cells. An intensity histogram of the merged immunofluoresence in U2OS cells (denoted by the two arrows) is shown on the right demonstrating their co-occurence. The arrow head denotes a cell with high S-AHNAK overexpression, which was not analyzed.

localized to the nucleus. Support for this prediction is found in the pull-down experiments which show a stronger interaction between S-AHNAK molecules, than between the two isoforms (Figure 3). Correct fractionation was confirmed by western blot analysis of nuclear Histone H2 and cytosolic CAPN3 (Figure 5A). Moreover, similar fractionation experiments on S-AHNAK-VSV transfected U2OS cells confirm its presence in the nuclear fraction, both as a 17 kDa and a 34 kDa band (data not shown).

To verify that endogenous S-AHNAK localizes to SC-35 containing speckles we performed immunoprecipitation experiments on IM2 myoblasts and myotubes with the monovalent heavy chain antibody fragment VHH A2. We could specifically immunoprecipitate S-AHNAK from both myoblasts and myotubes (Figure 5B). More S-AHNAK is found in the myotube IP, consistent with the increased expression of S-AHNAK during differentiation. We probed the IP fractions on western blot for SC35, which colocalizes with recombinant S-AHNAK, and the spliceosomal hnRNP

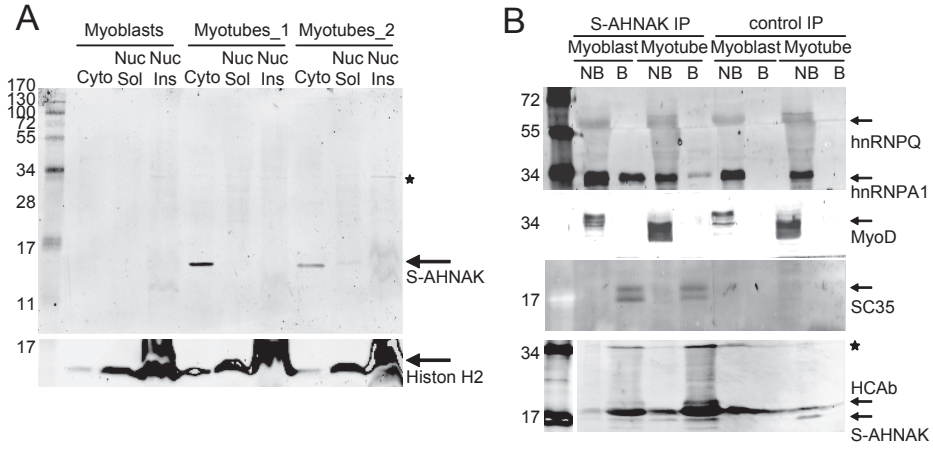

Figure 5: Endogenous S-AHNAK is in nuclear structures enriched for spliceosomal proteins. A) IM2 myoblasts and myotubes (in duplo) were fractionated and cytosolic, and nuclear fractions analyzed on western blot for S-AHNAK protein levels. There is no detectable S-AHNAK in myoblasts. In myotubes most S-AHNAK is in the cytosolic fraction, yet a minor amount is in the soluble nuclear fraction. Interestingly, the 34 kDa S-AHNAK dimer band is exclusively in the unsoluble nuclear fraction (*). Fractionation was confirmed by analyzing Histone H2 levels on western blot. B) S-AHNAK was immunoprecipitated from protein extracts from human primary skeletal muscle myoblasts and myotubes. Bound (B) and non-bound (NB) fractions were analyzed on western blot for S-AHNAK, the spliceosomal proteins SC-35, hnRNP-A1 and hnRNP-Q and the transcription factor MyoD showing that SC-35 and hnRNPA1 coimmunoprecipitate with S-AHNAK.

proteins A1, and Q. SC35 and hnRNP A1 co-immunoprecipitate with S-AHNAK, while hnRNP Q does not. Moreover, the nuclear transcription factor MyoD does not co-immunoprecipitate with S-AHNAK, confirming the specificity of the other interactions. This indicates that, like the recombinant S-AHNAK, endogenous S-AHNAK is physically present in nuclear speckles enriched for spliceosomal proteins.

S-AHNAK regulates mRNA splicing

Given that S-AHNAK colocalizes with spliceosomal proteins we hypothesized that it may be involved in mRNA splicing. To test this we expressed mouse recombinant S-AHNAK in human U2OS cells and human primary myoblasts and myotubes. After 24 hours we measured mRNA levels for the *AHNAK* locus by quantitative PCR using species-specific primers. As shown in Figure 6 the relative amount of endogenous S-AHNAK transcript increases with increasing expression levels of recombinant S-AHNAK, while the L-AHNAK transcript is reduced. This indicates that S-AHNAK can regulate splicing of mRNA molecules derived from its own locus.

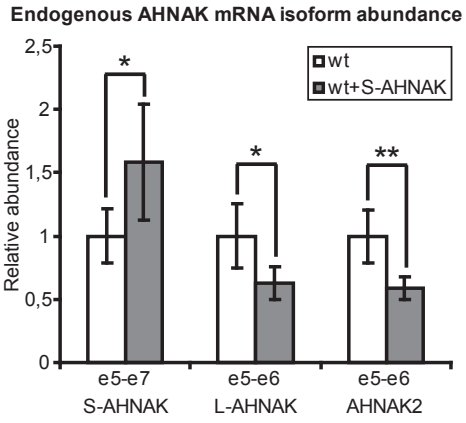

Figure 6: Recombinant S-AHNAK affects splicing of endogenous AHNAK transcripts. Human cells (Myoblast, myotubes and U2OS) were transiently transfected with mouse S-AHNAK, and endogenous AHNAK mRNA levels were analyzed by quantitative RT-PCR with species-specific primers. GAPDH was used as a reference gene. S-AHNAK increases its own mRNA product, and simultaneously decreases the L-AHNAK mRNA isoform and AHNAK2. Error bars reflect relative standard deviations of 4 independent experiments (each measured in triplo) conducted in human U2OS, myoblast and myotube (2x) cells. $P<0.05$ (*), or $p<0.005$ (**) according to a paired t-test with equal variance.

Discussion

We have investigated AHNAK evolution, transcription, and function. Using alignment tools and phylogenetic tree constructions we observed that vertebrate genomes contain three AHNAK-like genes, coding for AHNAK, AHNAK2 and PRX. All AHNAK-like genes have a similar structure with a giant exon flanked by multiple smaller exons. To our knowledge the AHNAKs are the only family with such gene structure. For *AHNAK* and *PRX* differential splicing is observed, resulting in two major protein isoforms that differ considerably in molecular mass. We propose that PRX is a third AHNAK family member.

The AHNAK gene has a conserved exon structure that consists of 10 exons in rodents, which differs from previous reports. We could confirm that the major transcription initiation site in muscle lies at Exon 1 in agreement with Ensembl. However, we could also identify expression of the previously unidentified Exon 2. It remains to be shown whether this functions as an alternative first exon. In addition we identified two alternative 3' untranslated sequences that contain a validated poly-adenylation sequence, and terminate the L-AHNAK and S-AHNAK transcripts. In mouse and rat several additional S-AHNAK transcript variants were identified, both in cultured cells, as in tissue (Figures 1A and 3A). No evidence for similar additional exons was found in other families, indicating this is specific for rodents. The functional consequences of each of these variants are currently unknown. The three variants of mouse S-AHNAK showed similar localization in cells (not shown). In addition, we identified several alternative 3' sequences for the L-AHNAK transcript in mouse, suggesting a specific function of these sequences. However, the expression levels of these isoforms are very low in muscle cells (Figure 3C). Therefore we believe it is unlikely that these isoforms play a major role in AHNAK function in muscle.

Given the size and the repeat nature of the three giant AHNAK-like proteins it is likely that they perform a structural role in the cells. In myelating Schwann cells L-AHNAK is involved in laminin dependent cell adhesion [220], and in epithelial cells it is involved in calcium and e-cadherin dependent cell-cell contacts [19]. Down regulation of AHNAK reduces cell-cell or cell-matrix adhesion, and impairs remodeling of the cortical Actin cytoskeleton [19,220]. AHNAK associates with filamentous Actin in epithelial, Schwann and cardiac cells. Moreover, L-AHNAK is important for pseudopod formation in metastasizing tumor cells [222]. PRX is highly expressed in the peripheral nervous system and the L-PRX protein is important for axon remodeling [59]. This suggests functional commonalities between the two proteins with regard to their structural roles, which has been noted earlier by Salim *et al* [220], stating that L-AHNAK is highly expressed in cells that suffer from a high susceptibility to membrane injuries [220]. Finally, it

has been suggested that AHNAK2 is functionally redundant to L-AHNAK.

By showing that the S-AHNAK isoform can regulate splicing of its own mRNA we uncovered a novel mechanism of post-transcriptional control of gene expression. We observed that during muscle differentiation S-AHNAK mRNA and protein levels strongly increase, while L-AHNAK levels remain constant. This indicates that differentiating myotubes have increasing need for S-AHNAK function. Based on our data we suggest the following model (Figure 7): S-AHNAK is docked on the N-terminus of L-AHNAK, which ensures its cytoplasmic retention at the cortical cytoskeleton. In skeletal muscle, L-AHNAK is recruited to the cell membrane by Dysferlin. Upon a yet to be identified trigger the AHNAK heterodimer is dissociated. S-AHNAK subsequently translocates into the nucleus. There it forms a homotetrameric complex that associates with the spliceosome possibly to match mRNA modifications to external cues. It affects mRNA splicing of AHNAK family members by increasing the levels of its own transcript, and simultaneously decreases the levels of the giant L-AHNAK and AHNAK2 transcripts. In addition to regulating the AHNAK gene loci, S-AHNAK likely also targets other loci but this remains to be investigated. However, in addition to its nuclear function, S-AHNAK needs to be restored in the cytosol. The fact that S-AHNAK increases the levels of its own mRNA is in agreement with a rapid restoration of cytosolic S-AHNAK protein.

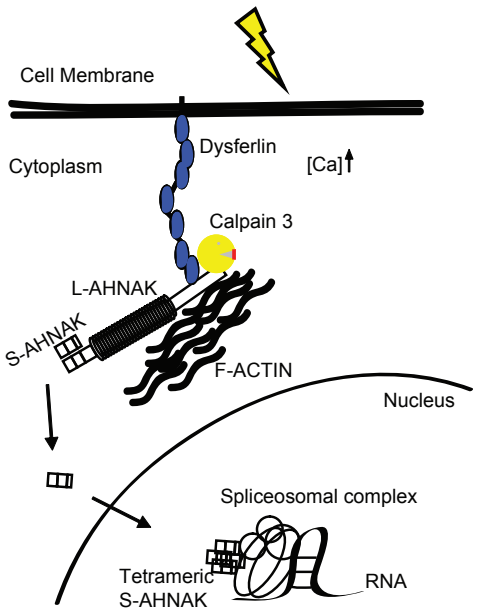

Figure 7: a model for S-AHNAK function. S-AHNAK is docked onto L-AHNAK at the cell periphery. Upon a structural trigger free intracellular calcium increases and the heterodimer is dislodged. S-AHNAK will translocate into the nucleus. There it localizes as a homotetramer to spliceosomal structures and influences mRNA processing.

It is tempting to speculate that this mechanism of self regulation is conserved in the other AHNAK family members. Intriguingly, a small PRX isoform was described. S-PRX has been reported to reside both in the cytosol and in the cell nucleus [79], while L-PRX is only cytosolic. It will be interesting to see whether this nuclear localization conforms to spliceosome-like speckles, and whether S-PRX affects mRNA levels of the AHNAKs as well.

It is intriguing that S-AHNAK largely colocalizes with L-AHNAK in the cytosol. Based on the size exclusion chromatography data, which showed that L- and S-AHNAK can form a heterodimer, we suggest that the cytosolic AHNAK isoforms form a complex. GST pull-down experiments showed that PRX can also self-interact [79,224]. The authors showed that the interaction is dependent on the PDZ domain. Moreover, by mutating two conserved residues within the 90 amino acid long PDZ domain, they could abolish the self-interaction. As S- and L-AHNAK also contain a PDZ domain we hypothesize that this domain underlies AHNAK self-interaction. Unfortunately, the essential residues identified in the PRX PDZ domain (GLGF-motif) are not conserved in AHNAK, leaving open the question which residues are important for oligomerization.

Interesting in this respect is that the oligomerization domain in PRX is flanked by a sequence that regulates subcellular localization [223]. It is a common feature of proteins that translocate between cytosolic and nuclear compartments that these sequences are regulated by post-translational modifications (such as phosphorylation, as was shown for L-AHNAK), or protein-protein interactions. In a recent study subcellular localization of the protein PREP was shown to be dependent on interactions with the cytoskeleton, which enabled cytoplasmic retention, and a NES which ensured nuclear export [103]. Disrupting the cytoskeleton enabled transport into the nucleus. To allow for nuclear localization, additional masking of the NES is required. We propose that tetramerization of S-AHNAK masks its predicted NES sequence. In all, this suggests that the subcellular localization of S-AHNAK reflects a common cellular mechanism.

Mutations in the *PRX* gene cause a form of Charcot-Marie-Tooth disease (CMT1B, MIM#118200) [137]. L-AHNAK has been implicated in Limb-Girdle Muscular Dystrophy as it directly interacts with Dysferlin and CAPN3. These observations underline the relevance of the AHNAK family in both physiology and pathophysiology. It was previously shown that L-AHNAK is proteolytically modified by CAPN3, and has a cleavage site in the N-terminus, located 100 aa into the exon 6 encoded part [118]. This would result in a protein fragment similar to S-AHNAK. Possibly, this fragment might share properties of S-AHNAK such as oligomerization.

Summarizing we showed that the *AHNAK* gene forms a family with *AHNAK2*

and *PRX* and differentially expresses multiple transcripts and protein species. S-AHNAK regulates splicing of AHNAK mRNA transcript and constitutes a novel mechanism of post-transcriptional self-regulation.

Methods

Cell culture

Mouse IM2 myoblast were maintained under permissive conditions at 33°C and 10% CO2 in DMEM (61965, GIBCO) supplemented with L-Glutamine, 1% Pen/Strep (Gibco-BRL), 20% FCS, IFN-γ, and Chick embryo extract. For differentiation cells were grown to 70% confluency and switched to DMEM (61965) supplemented with 10% HS, L-Glutamine, and 1% Pen/Strep. Mouse C2C12 cells were maintained in DMEM (1181) supplemented with L-Glutamine, 1% Pen/Strep, 1% glucose and 10% FCS at 37°C. For differentiation cells were switched to medium with 2% HS. Human primary myoblasts were grown in F-10 medium (31550), 1% Pen/Strep, 20% FCS, 10ng/ml bFGF, 1uM Dexamethazone. Differentiation was achieved in DMEM (41966), 2% HS. U2OS cells were maintained in Dulbecco's modified Eagle's medium (Gibco-BRL) supplemented with 10% FCS, 1% Pen/Strep.

Antibodies

The following antibodies were used in this study: RaAHNAK (KIS 1;5,000), MaVSV (P5D4), MaE4/5 (MaAHNAK, Abnova), MaMYC (9E10, Roche), RaMYC (Cell Signalling 1;5,000), MaHistone2, MaSC35, GaSC35 (Santa Cruz), 3F5 (VHH against PABPN1), 3A (VHH against β-Amyloid), MahnRNP-Q, RahnRNP-A1, RaMyoD, MaM13 and MaM13hrp (both 1;5,000 Novagen) GaMousealexa488 (Molecular Probes), GaRabbitalexa594 (Molecular Probes) at 1;1,000 and 1;2,000, respectively. GaRabbitIRDye800 and GaMouseIRDye680 (Westburg) were used at 1;5,000 for western blotting.

RNA extraction and cDNA synthesis

RNA was extracted from homogenized tissue or cultured cells with a kit (Machery-Nagel) according to the manufacturer’s protocol. 1µg RNA was used as input for a cDNA synthesis reaction (Fermentas) according to the manufacturer’s protocol. Genomic DNA was digested on the column and cDNA was purified with a gel-extraction kit (Machery-Nagel).

Polymerase Chain Reaction (PCR)

All primers were designed with the webtool Primer3 (frodo.wi.mit.edu/primer3/), with mispriming against human or rodent databases. End-point PCR reactions were

performed with Phusion polymerase in HF buffer according to the manufacturer's protocol. Quantitative PCR reactions were performed with SYBRgreen, in 15µl reaction volume with 3ng cDNA input. Primer sequences are depicted in Table S1, All primers had comparable efficiencies between 95% and 105%. All measurements were performed in triplo.

Relative mRNA expression levels were calculated using GAPDH as reference gene, with the method of Pfaffl [200]. Briefly, expression differences were determined with the following formula: 2^dCt and the squareroot of (sdGAPDH^2+sdGOI^2), where GOI means Gene Of Interest. The standardized expression (relative dCt) was calculated by: 2^dCt*sddCt*LN(2)/N technical replicates. Finally, for relative expression levels, one value was set to 1 (100%). To calculate statistical significance in Figure 6, relative ct values of 4 biologically independent experiments were averaged. The square root of the added individually squared standard deviations was plotted as error bar in Figure 8. Finally, a paired t-test with equal variance was performed.

Cloning

The open reading frames of S-AHNAK and L-AHNAK N-terminus were amplified from human and mouse skeletal muscle cDNA and cloned into pET28 (prokaryotic) and pSG8 (eukaryotic) expression vectors with unique restriction sites for BamHI and XhoI. In addition mouse S-AHNAK was cloned into pIVEX1.3 for *in vitro* transcription-translation. Exon-exon boundary spanning amplicons were separated on agarose gels, purified and cloned in pTOPO-blunt vector (Invitrogen). All vector cloned sequences were confirmed by direct sequencing (LGTC, Leiden, the Netherlands).

Transfection

Cells were plated and grown to 50% subconfluence. U2OS cells were transfected with FuGENE-6 (Roche Applied Science) and C2C12, IM2 and immortalized myoblasts were transfected with Lipofectamine PLUS (Invitrogen), both according to manufacturers protocol. 24h Post-transfection cells were washed and harvested for further analysis.

Recombinant protein production

pET28 expression constructs were transformed into BL21 (DE3)-RIL E*scherichia coli* (Stratagene) cells by electroporation. Bacterial cells were grown to log phase and recombinant protein production was initiated by the addition of IPTG (Fermentas, Germany) to a final concentration of 1 mM. After 2.5 h induction at 37ºC, the cells were collected by centrifugation at 3,000 g, and lysed in lysis

buffer 1 (50 mM Tris, pH 7.4, 1 mM EDTA, 1.5mg/ml lysozyme, 0.15 M NaCl, 1% Triton X-100, plus 1× protease inhibitor cocktail (Roche Molecular Biochemicals, Basel, Switzerland)). All proteins were in the resulting soluble fraction except AHNAK N-terminus, which was extracted by 5 min sonication on ice. Lysates were spun down at 13,000 g for 30 min at 4ºC, and the proteins purified via IMAC with a Profinia purificator (BioRad). Concentration was obtained with absorption spectroscopy, and protein stability analyzed on SDS-PAGE gels.

An *in vitro* translation was performed with S-AHNAK in pIVEX vector using the wheat-germ *in vitro* transcription translation kit according to the manufacturer's protocol. HIS6-tagged proteins were subsequently purified using TALON resin (Clontech). In short TALON resin was pre-equilibrated with PBS and added to the protein mixture. Samples were incubated for 2h, 4°C, and subsequently spun down. After 3 wash steps (5 min) the TALON resin was loaded onto a column and subjected to a 300x bed volume wash. Proteins were eluted with 200mM imidazole, and dialyzed O/N at 4°C to PBS. Concentration and purity were analyzed on SDS-PAGE gels.

GST pulldown

Equal amounts of purified GST fusion proteins were immobilized to preequilibrated Gluthione-Sepharose 4B (Amersham, Sweden), incubated at room temperature for 30 minutes, and then centrifuged at 500g for 5 minutes. The supernatant was removed and the Glutathione-Sepharose was washed three times with lysis buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 0.2% Triton X-100). Purified T7 fusion protein was added and incubated for 2h 4°C. The sepharose was washed 5x (3x short, 2x >20 min) and bound proteins eluted by boiling in sample buffer. Unfused GST served as a negative control.

Cell fractionation

Cells were trypsinized and spun down in culture medium. Pellet was washed in PBS, and resuspended in Buffer A (10 mM Tris pH 7,4, 1,5 mM MgCl2, 10 mM KCl, 0,5 mM DTT, 1x protease inhibitor tablet). Cells were put on ice for 10s, and lysed by the addition of NP-40. Lysate were spun down immediately at 8,000g, 4°C 10s. The supernatant contained the cytoplasmic fraction. The pellet was washed 3x in buffer A, and subsequently dissolved in Buffer B (50 mM Tris pH 7,4, 150 mM NaCl, 1% NP40, 1x Protease inhibitor cocktail), and stored at -80°C for 30 min. Samples were thawed on ice, and spun down at 8,000g, 10 min, 4°C. The supernatant contained the soluble nuclear fraction, and the pellet the insoluble nuclear fraction. Fractionation was verified on western blot with antibodies against Histon H2 and CAPN3.

Size-exclusion chromatography

The Superdex 200 column (GE Healthcare) was pre-equilibrated in PBS for 30 min, and calibrated with marker proteins. Purified proteins at 1mg/ml in 200µl were added to the column and flow-through (2.0ml/min) was analyzed at A280 absorption with Unicorn software (4.12).

Immunization and phage display library construction

Two llama's (Utrecht University) were immunized with four boosts of 100ug purified recombinant mouse S-AHNAK. One week after the last boost, plasma blood mononuclear cells were collected and RNA was isolated by phenol/chloroform extraction. cDNA was synthesized from 40ug RNA using the reverse transcriptase kit (Invitrogen), according to the protocol of the manufacturer (Promega). VHH regions were amplified in a nested PCR reaction. The first PCR reaction, that both amplifies VHH as well as VH genes, was performed with primers that annealed at the leader region and the CH2 region. The product was run on a 1.5% agarose gel, and the lower band of the resulting two PCR products, representing the VHH genes, was purified via gel extraction. The second PCR reaction was performed with primers annealing at the framework 1 and CH2 region to introduce a 5-prime SfiI site. The VHH genes were subsequently cloned into a phagemid vector using the restriction enzymes SfiI and BstEII, and transformed into TG1 cells. The size of the library was estimated by plating serial dilutions of the transformation. Phage were rescued as described before [170].

VHH selection

200ng of purified *in vitro* transcription translation S-AHNAK was directly coated onto a maxisorb 96 well plate O/N at 4°C. The plate was washed twice in PBS and blocked for 1 hour with 4% marvel PBS (mPBS). The phage library, pre-incubated in mPBS, was allowed to bind for 2 hours at room temperature. Non-bound phage were removed by extensive washing, 15 times PBST and 2 times PBS. Bound phage were eluted in 100mM triethylamine for 10 minutes at room temperature, and the pH was subsequently neutralized using 1M Tris pH 7.5. Exponentially growing TG1 cells were infected with the output phage. The infection mix was plated on LB agar plates containing 2% D-glucose and ampicillin.

Phage and VHH ELISA

Clonal phage were prepared as described before [170] and tested for their ability to bind 100ng purified protein coated on maxisorb 96 well plates. Positive clones were sequenced and unique clones were further evaluated for S-AHNAK specificity in a VHH ELISA as described before [144]. In short, VHH were produced by adding

IPTG to a final concentration of 1mM to exponentially growing TG1 clones in 2TY medium containing ampicillin and 0.1% D-glucose for four hours. Cells were harvested and lysed as described above and the VHHs were subsequently purified using TALON (Clontech) according to the instructions of the manufacturer. A dilution series of each VHH in mPBS was incubated in antigen coated and blocked maxisorb 96 well plates for 2 hours. Bound VHH was detected using a mouse anti-myc antibody followed by a HRP conjugated rabbit anti-mouse antibody, with extensive washing after each incubation. Bound HRP conjugated antibodies were visualized using OPD as substrate and the reaction was stopped after 10 minutes using 1M H2SO4.

Cloning of bivalent VHH

Bivalent VHH antibody fragments were obtained by cloning the VHHs in tandem into the expression vector pUU-11 as described before [144].

Western blot

Cell lysates were prepared in Llaemli sample buffer and loaded onto SDS-PAGE gels. Proteins were separated and blotted onto PVDF membranes. For L-AHNAK, gels were blotted onto nitrocellulose. Blots were washed, blocked in 4% mPBS for 30 min, and incubated with primary antibody diluted in blocking buffer for 2h at RT. Subsequent washes in PBS-tween were followed with incubation with secondary antibodies for 1h in the dark. MaAHNAK was detected with HRP conjugated secondary antibodies. Blots were washed and scanned with an odyssey scanner (Licor, Lincoln, Nebraska, USA).

Immunostaining

Transiently transfected cells were fixed in formalin solution for 10 min, followed by permeabilization in 0.3% Triton (Sigma) for 5 min. Cells were washed, blocked in 1% BSA (Sigma) for 10 min, and incubated with 1st antibody 2h RT. Cells were subsequently washed 3x and incubated with 2nd antibody for 1h RT. Cells were washed in PBS and mounted onto slides with AquaPolymount supplemented with DAPI).

Immunoprecipitation

Cells were lysed in triton buffer (50mM TrisHCl, pH 7.5, 150mM NaCl, 0.2% Triton X100, 1x protease inhibitor cocktail (Roche)). Cultured cells were prepared freshly by washing in PBS and lysed by scraping on ice in lysis buffer. All homogenates were spun down at maximum speed, 4°C, 20 min. Protein A Sepharose CL-4B (GE Healthcare) was washed 3x in lysisbuffer and used to preclear the homogenates

for 1h, at 4°C tumbling. Sepharose was removed and antibody added (50µg VHH) for O/N incubation at 4°C, tumbling. Thereafter washed Sepharose was added and incubated for 2h, 4°C, tumbling). Homogenates were spun down at 500g and supernatant stored as non-bound fraction. The Sepharose was washed 5x (3x short, 2x long (>20min tumbling at 4°C)). Finally, all fluid was removed and protein eluted by boiling in sample buffer.

Sequence analysis

Sequence data was retrieved from Ensembl (www.ensembl.org) and NCBI databases (AceView, www.ncbi.nlm.nih.gov/IEB/Research/Acembly/). Blast searches were performed with the NCBI web server, specifically with psi-blast and blast-p and tbast-n vs the non-redundant protein or expressed sequence-tag (EST) databases. All retrieved sequence clones are in Table S1. Species for which the sequence data yielded incomplete material were excluded from further analyses, and included chicken and frog. Presumably, the repeat nature of the AHNAK genes precluded complete sequencing of these genomes.

Phylogenetic analysis

Phylogenetic analyses were performed on the complete amino acid sequences, and separately the PDZ domains. The PDZ domains were predicted with the CDD tool at NCBI (www.ncbi.nlm.nih.gov/Structure/cdd). Sequences were aligned using T-Coffee. The alignments were manually checked in GeneDoc (www.nrbsc.org/gfx/genedoc/), and in the case of the PDZ adjusted at the termini (Alignments and adjustments, are in Figure S3 and Table S3). The alignment of the complete amino acids sequences was analyzed with GBlock software [39,241] and reduced accordingly. The amino acid substitution model was estimated using ProtTest [1] and found to LG+G+I (LG substitution model with estimation of gamma distribution shape parameter alpha and the proportion of invariable sites) as the best fitting out of 112 examined models for the PDZ sequences, and the GBlock sequences. A maximum likelyhood tree was constructed with the webtool PhyML [95,96], using 100,000 tries. A Bayesian inference was performed with mrbayes [123]. As mrbayes does not contain the LG model, the second best model, WAG was used, with 10,000,000 generations. Trees were visualized with TreeView.

Acknowledgements

We thank Dr. J. Baudier for sharing the KIS RaAHNAK antibody. We thank Dr. D.J.M Peters for critical reading of the manuscript. We thank Dr. P.A.C. 't Hoen for valuable comments and for the CAGE sequence analysis.