A novel domain suggests a ciliary function for ASPM, a brain size determining gene.

Chris P. Ponting

MRC Functional Genetics Unit
University of Oxford
Department of Human Anatomy and Genetics
South Parks Road
Oxford OX1 3QX
Email: chris.ponting@anat.ox.ac.uk
Tel.: +44 1865 285855

Running Title: Novel Domains in Microcephaly and Ciliary Proteins

Bioinformatics: Discovery Note.

Abstract

The N-terminal domain of abnormal spindle-like microcephaly-associated protein (ASPM) is identified as a member of a novel family of ASH (ASPM, SPD-2, Hydin) domains. These domains are present in proteins associated with cilia, flagella, the centrosome and the Golgi complex, and in Hydin and OCRL whose deficiencies are associated with hydrocephalus and Lowe oculocerebrorenal syndrome, respectively. Genes encoding ASH domains thus represent good candidates for primary ciliary dyskinesias. ASPM has been proposed to function in neurogenesis and to be a major determinant of cerebral cortical size in humans. Support for this hypothesis stems from associations between mutations in ASPM and primary microcephaly, and from the rapid evolution of ASPM during recent hominid evolution. The identification of the ASH domain family instead indicates possible roles for ASPM in sperm flagellar or in ependymal cells’ cilia. ASPM’s rapid evolution may thus reflect selective pressures on ciliary function, rather than pressures on mitosis during neurogenesis.
Two observations have been cited as evidence that sequence variation within one particular gene, *ASPM*, has contributed to the dramatic expansion of brain size during primate, and in particular hominid, evolution (Ponting & Jackson, 2005; Gilbert et al., 2005; Woods et al., 2005). The first of these is that truncation mutations in human *ASPM* are the most common cause of primary microcephaly, a neurodevelopmental disorder resulting in substantial reduction in brain size and in mental retardation (Bond et al., 2002; Bond et al., 2003; Kumar et al., 2004). The second is that nucleotide substitutions in *ASPM* exhibit the signature of positive selection both between species and within the human population (Zhang, 2003; Evans et al., 2004; Kouprina et al., 2004; Mekel-Bobrov et al., 2005). Nevertheless, as *ASPM* is broadly expressed in many tissues (Kouprina et al., 2005) it remains possible that its evolution is causally unconnected with the expansion of the cerebral cortex in great apes (Ponting & Jackson, 2005). It is thus important that the full repertoire of molecular functions of *ASPM* be considered before its adaptive influence on brain size is accepted.

Mutations in a second gene, *Hydin*, in *hy3* mice result in a lethal hydrocephalus (Robinson et al., 2002; Davy & Robinson, 2003), the accumulation of cerebrospinal fluid within the brain’s ventricular system. The corresponding chromosomal region in the human genome (HSA16q22) is also associated with a congenital hydrocephalus (Callen et al., 1990). The expression of mouse *Hydin* in ciliated ependymal cells, the bronchi of the lung, spermatocytes and the lining of the oviduct, is consistent with a function within cilia and flagella (Davy & Robinson, 2003). Indeed, an orthologue of mammalian Hydin has been found within purified flagella from *Chlamydomonas reinhardtii* (Pazour et al., 2005). Nevertheless, here, as with *ASPM*, the link between molecule and pathology is obscure, particularly because *Hydin* sequence hitherto has not yielded clues as to its function.

Here, I show that the N-terminal domain of *ASPM* and 33 domains of Hydin are homologous. This domain family is represented among both eukaryotes and bacteria, and is distantly related to the immunoglobulin fold domain of the major sperm proteins in nematodes. Available evidence indicates that these ASH (*ASPM*, *SPD-2*, *Hydin*) domains in eukaryotic proteins associate with microtubules.
Sequence analysis

Significant sequence similarities among ASH domains are apparent by searching current sequence databases using a variety of sequences as queries. For example, a PSI-BLAST (Altschul et al., 1997) search of the non-redundant protein sequence database (nr; ftp://ftp.ncbi.nih.gov/blast/db/) with the N-terminal 200 amino acids of human ASPM detects significantly similar ($E < 0.002$) domains in RW1 orthologues (by iterations 2-4), Hydin (iteration 5), DLEC1 (iteration 5), Caenorhabditis elegans SPD-2 (iteration 6), human CXorf22 (iteration 7) and bacterial adhesins (iteration 5).

These sequences were multiply aligned, guided by the PSI-BLAST search results, with additional ASH domains being identified in subsequent HMMER (http://hmmer.wustl.edu) and PSI-BLAST searches. Human gene sequences were often extended using gene predictions and transcript evidence available from the UCSC genome browser (http://genome.cse.ucsc.edu/). In two cases, previous gene predictions (CXorf22 and RP13-11B7.1; LOC255101 and BAC86878) were amalgamated into single loci. An ASH domain in Lowe’s oculocerebrorenal syndrome protein (OCRL-1) was apparent from a PSI-BLAST search of nr using amino acids 540-731 as query. This detects significant similarity ($E < 0.002$) between this region of OCRL-1 and ASH domains in DLEC1 and Hydin within two iterations.

A multiple alignment and secondary structure prediction (Rost & Sander, 1993) of these domain sequences (Figure 1) reveal an 8 β-strand domain with a single highly conserved asparagine. These domains appear to be distantly related to major sperm protein (MSP) domains (Bullock et al., 1996): a PSI-BLAST search of nr using the third ASH domain in mouse Hydin (amino acids 244-338) yields a significant alignment ($E = 2 \times 10^{-3}$) with the MSP domain in human vesicle-associated membrane protein-associated protein A (VAMP-A) in 3 iterations. ASH domains are also readily detectable in bacterial proteins. For example, 16 ASH domains in a Rubrivivax gelatinosus hypothetical protein (GenInfo identifier [gi]: 47572664), as well as the mouse ASPM ASH domain, are detected in a PSI-BLAST search of nr using Bdellovibrio bacteriovorus Bd2309 (gi: 42523760) within 3 iterations.
**Ciliary, centrosomal and Golgi proteins**

In all, 90 ASH domains were found in 13 human proteins (Figure 1). These proteins have been identified within three intracellular compartments, all centred around the centrosome. Firstly, orthologues of Hydin, CXorf22 (FAP47), KIAA1751 (FAP74), LOC255101 (FAP65) and PF6 have all been found in purified *Chlamydomonas* flagella (Pazour et al., 2005). The expression of mouse Cod106, a single ASH domain protein, in sensory hair cells (Reisinger et al, 2005) suggests its involvement in the organisation of these cells’ cilia, and thus in the mechanotransduction of sound.

Secondly, ASPM and Cep192 are centrosomal proteins (do Carmo Avides & Glover, 1999; Andersen et al., 2003), as is SPD-2, a *C. elegans* protein (Pelletier et al., 2004) with a single ASH domain. The centrosome is a microtubule organising centre that is essential for the regulation of cell division in meiotic and mitotic cells. However, it also contributes centrioles to the basal bodies of cilia. This is underlined by the presence of proteins, such as NPHP1, NPHP2, BBS4 and BBS8 which are mutated in ciliary diseases, in both basal bodies and centrosomes (Hildebrandt & Otto, 2005).

Thirdly, the inositol polyphosphate 5-phosphatase OCRL-1, which contains a single ASH domain between its phosphatase and RhoGAP domains, is localised to the Golgi (Olivos-Glander et al., 1995). The Golgi apparatus occurs in close proximity to the cilium as it emerges from the basal body (Poole et al., 1997) and centrosomal proteins appear to assist in organising the Golgi apparatus (Takatsuki et al., 2002; Rios et al., 2004). Mutations in the *OCRL* gene result both in Lowe oculocerebrorenal syndrome, a rare X-linked disorder characterized by mental retardation, congenital cataracts, and renal Fanconi Syndrome, and in Dent disease 2, a renal proximal tubulopathy (reviewed in Lowe, 2005). None of these mutations are found to be present in the OCRL-1 ASH domain. Nevertheless, the finding that OCRL-1 possesses an ASH domain typical of ciliary proteins suggests that detailed studies of the potential roles of OCRL-1 in ciliary function are warranted.

**Molecular function**
ASH domains may possess a microtubule-binding function. This prediction arises from asp, the *Drosophila melanogaster* orthologue of ASPM, binding microtubules within a 512 amino acid region that encompasses its ASH domain (Saunders et al., 1997), and also observations that SPD-2 and PF6 bind centrioles and axonemes (bundles of microtubules and other proteins that form the core of each cilium), respectively (Kemp et al., 2004; Rupp et al., 2001).

ASH domain-containing proteins of unknown function, such as RW1, DLEC1, KIAA0922, KIAA1751 and LOC255101/BAC86878, should now be investigated for their potential roles in ciliary, centrosomal and Golgi function, particularly with respect to microtubule-binding. Genes encoding ASH domains should also be prioritised for investigation as to whether they are mutated in diseases of cilia, flagella or centrosomes, such as Bardet-Biedl syndrome, hydrocephalus and cystic kidney disease (Hildebrandt & Otto, 2005).

The identification of the ASH domain family in Hydin and other ciliary and centrosomal proteins serves to re-emphasise that loss of ciliary function is a key feature of hydrocephalus. In mouse, the ASH domain and axoneme central apparatus protein PF6 is a known binding partner of sperm-associated antigen 6, whose mutation also results in hydrocephalus (Sapiro et al., 2002; Zhang et al., 2005). Although a ciliary localisation of Hydin is expected, its contribution to ensuring the proper circulation or retention of cerebrospinal fluid remains unknown. Nevertheless, a central domain of Hydin could be readily identified as a hitherto unrecognised adenylate kinase homologue, similar to that found in mammalian KLP2 and *Chlamydomonas* Cpc1 (Figure 2). This suggests that this domain in Hydin may catalyse the local increase of ATP which would fuel the activity of dynein, a microtubule motor required for the beating of eukaryotic cilia/flagella.

*Evolution of ASPM*

Prior sequence-based evidence for positive selection on *ASPM* is persuasive, as is evidence that truncating mutations of this gene result in microcephaly. Nevertheless, because *ASPM* is expressed widely in many different tissues (Kouprina et al., 2005) it remains plausible that any one of its functions in these tissues might have been the
subject of adaptive evolution rather than its proposed centrosomal participation in neurogenesis (Bond et al., 2002; 2003; 2005).

Analysis of primate ASPM sequences shows that human and gorilla, but not chimpanzee, lineages exhibit significant and pronounced levels of adaptive evolution (Kouprina et al., 2004). Thus, accelerated evolution of ASPM began well before the three-fold brain size increase separating early hominids (australopithicines, approximately 3 million years ago) from modern humans (Wood and Collard, 1999). Moreover, gorilla and chimpanzee lineages, within which brain sizes have held relatively constant, have experienced adaptive and non-adaptive evolution respectively. If ASPM evolution did lead to brain size increases, it would appear that this first occurred approximately 7-8 million years ago, prior to the last ancestor of gorillas, chimpanzees and humans, rather than during more recent hominin brain enlargement.

A ciliary function for ASPM?

The identification of ASH domains in ASPM, centrosomal and ciliary proteins indicates that ASPM may possess roles not only in mitotic spindle regulation, but also in ciliary and flagellar function. This would be consistent with the domain architecture of ASPM since it shares all of its domain types with numerous ciliary proteins. ASPM contains an ASH domain, as do ciliary proteins Hydin, PF6, FAP47, FAP74 and FAP65; it contains CH domains, in common with ciliary proteins FAP47, Cpc1 and KPL2; and, it contains calmodulin-interacting IQ domains, similar to those in nephrocystin-5 and inversin, whose mutations are associated with ciliary dyskinesias. As described above, the known centrosomal localisation of ASPM would accord with a role within the basal bodies (modified centrioles) of cilia and flagella.

A putative ciliary and flagellar function for ASPM provides two further hypotheses that might account for the adaptive evolution of ASPM. The first is that ASPM possesses a role in sperm flagellar function, which is consistent with its expression in male germ cells entering spermatogenesis (Lüers er al., 2002). This would account
for its unusually rapid evolution because other such proteins involved in male reproductive function have evolved adaptively (Nurminsky et al., 1998).

The second hypothesis is that ASPM assists not in neurogenesis, as proposed previously (Bond et al., 2005), but instead in cilia-mediated neuronal migration. A recent study has shown that motile cilia in brain ependymal cells direct the course of cerebrospinal fluid thereby guiding young migrating neurons (Sawamoto et al., 2006). Loss of ASPM function in cilia thus might account for reduced brain size in primary microcephaly due to decreased cerebrospinal fluid flow. ASPM molecular function in primates’ ependymal cilia might represent the genetic substrate upon which past adaptive evolution has acted.

Summary

The identification of ASH domains in a diverse set of proteins should now provoke renewed interest in understanding the structure, function and evolution of genes that underlie primary microcephaly, hydrocephalus and Lowe oculocerebrorenal syndrome. In particular, it will be of interest to investigate whether these disorders are primary ciliary dyskinesias and to further understand the role of ASH domains in centrosome, ciliary and flagellar function.

Acknowledgements

I am grateful to Prof Philip Beales (Institute of Child Health, University College London) and Drs Caleb Webber and Leo Goodstadt (MRC Functional Genetics Unit, Oxford) for helpful comments.
Figure Legends.

Figure 1
Multiple sequence alignment of representative ASH domains among over 900 such domains currently identifiable in nr using HMMER and an \( E \)-value upper-threshold of 0.1. The alignment is presented using CHROMA (Goodstadt & Ponting, 2001) and an 80% consensus of the eukaryotic sequences. The predicted (Rost and Sander, 1993) secondary structure (E=\( \beta \)-strand) is shown beneath the alignment, as is the alignment with the MSP domain from human VAP33 and its known secondary structure. Amino acids excised from the alignment are indicated in parentheses. Consensus abbreviations (amino acids): a, aromatic (FWHY, blue lettering on a dark yellow background); b, big (EFHIKLMQRWY, blue on light yellow); h, hydrophobic (ACFGHILMTVWY, black on dark yellow); l, aliphatic (ILV, grey on dark yellow); p, polar (CDEHKNQRST, blue on white); and s, small (ACDGNPSTV, dark green on white). A multiple sequence alignment of human ASH domains is available via http://www.mrcfgu.ox.ac.uk/ponting/ASH.aln. The following sequences are shown: human (Homo sapiens, Hs) ASPM (GenInfo identifier [gi]: 55976785) amino acids 48-141; Drosophila melanogaster (Dm) Asp (gi: 45549231) amino acids 36-131; mouse (Mus musculus, Mm) Hydin (gi: 46361984) amino acids 616-740, 1219-1423, 3091-3197 and 5056-5154; human CXorf22 (gi: 20380780, amino acids 12-111 and 735-857); human centrosomal protein 192kDa (Cep192; gi: 50811889) amino acids 781-884, 1174-1284 and 1848-1938; human deleted in lung and esophageal cancer 1 (DLEC1; gi: 4826696) amino acids 377-473, 864-964 and 1640-1744; a human open reading frame (ORF1; LOC255101; gi: 34533999, amino acids 149-240 and 781-872; gi: 34916026, amino acids 733-827); Caenorhabditis elegans (Ce) SPD-2 (SPindle Defective 2; gi: 17507089) amino acids 482-574; human RW1 (gi: 12230553) amino acids 35-128, 472-587 and 756-895; a human ORF (ORF2; KIAA0922; gi: 35038564) amino acids 443-576 and 601-703), Myc-binding protein-associated protein (AMAP-1; gi: 20149700) amino acids 432-538; human projection protein PF6 (gi: 46240864) amino acids 2081-2180; a human ORF (ORF3; KIAA1751) gi: 12698047 (amino acids 462-560, mRNA DN831193) and gi:51339036 (amino acids 16-142); Oculocerebrorenal syndrome of Lowe phosphatidylinositol polyphosphate 5-phosphatase, isoform b (gi: 13325070) amino acids 570-676; type II inositol polyphosphate 5-phosphatase, isoform b (gi: 1352493) amino acids 624-729; Rubrivivax gelatinosus PM1 (Rg) ORF (gi: 47572664) amino acids 126-221; Bdellovibrio bacteriovorus (Bb) Bb2309 (gi: 42523760) amino acids 43-134; and, Aquifex aeolicus (Aa) serine protease (SP; gi: 15606951) amino acids 415-509.

Figure 2. Multiple sequence alignment of adenylate kinase domain homologues in Hydin, Chlamydomonas reinhardtii Cpc1, human KPL2, and Escherichia coli adenylate kinase (adk) presented using CHROMA (Goodstadt & Ponting, 2001). A Conserved Domain Database search (Marchler-Bauer et al., 2005) with amino acids 2061-2311 of mouse Hydin yielded a significant alignment (\( E = 2 \times 10^{-3} \)) to cd01428 which represents the adenylate kinase domain family. Sequences and abbreviations: Hydin_Hs, Homo sapiens Hydin (GenInfo identifier [gi]: 14042342); Hydin_Bt, Bos
taurus Hydin (gi:76673899); Hydin_Mm, *Mus musculus* Hydin (gi:46361984); Hydin_Gg, *Gallus gallus* Hydin (gi: 50753651); Hydin_Xt, *Xenopus tropicalis* Hydin (gi: 58621133); Hydin_Dr, *Danio rerio* Hydin (gi: 68390275); Hydin_Sp, *Strongylocentrotus purpuratus* Hydin (gi: 72013378); Hydin_Ci, *Ciona intestinalis* Hydin (gi: 19510801 and 16884471); Cpc1_Cr, *Chlamydomonas reinhardtii* central pair complex 1 (gi: 48249492); KPL2_Hs, *Homo sapiens* KPL2 (gi: 46411154); and adk_Ec, *Escherichia coli* adenylate kinase (gi: 26246489). Abbreviations are as for Figure 1, except: t, tiny (AGS, light green on white); -, negatively-charged (DE, red on white); and, *, Ser/Thr (ST, light blue on white).
REFERENCES


<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence</th>
<th>Consensus/80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydin_Hs</td>
<td>ALTHALSXSANAVSLAKYNYACSSILSLAVSN (128) VLEDELVQLHASKQLS----DYYRWWFDL</td>
<td>.111bGsbPtGK*sst..lt+bYshahLs1Dslb-Altss</td>
</tr>
</tbody>
</table>