

S. acutus (ACB06751): pI=3.95, no putative glycosylation site

MKLFAAVVLVTCALASVQGRTLQQAAAAPTTVAQALAQTPSLSTLNAAVQAAGIDIPADAAW***TIF***
APKNEAFSDDDDIREETGLTAQQLQPENKQALTQLLQYHIVPSGALRAAQLQDGGQQLTTALAGAA
PLKVDIEDEKVEIEVPDAGNDDDGDDADVEQADMTVG***NSVIHIVDEVLI***PASLRGSRQPARG

C. vacuolata (CvacAstaP1 MZ063685): pI=6.12, 10 putative glycosylation sites

MRTLALFVLACTIGAAYGRELLQAAPAAASYNTLAQAVAAANASAP***NLS***SILLAAVKAA***NVS***GALGP
NTTW***TILAPT******NQAF***TSRL***NES***LGITPQQLQPANCKTLVEVLSYHVIPSGAVYSS***NLT***NGQEVPT
ALAGADPLTVAIN***NRT***VTFEGAEN***NAT***VTVANIRAG***RSVIHIVINDVLL***PAGVG***NSSNTT***AEGEST
APGAAGSRDVAAGSAVGGGSTSAPAPAAAKGAAGAAQVPAVLLASMLLIGMLF

C. vacuolata (CvacAstaP2 MZ063686): pI=10.3, 2 putative glycosylation sites

MARNAIIVFLAIVCLVAGATA***DKADPAKKKSAPSYPSLS***DAVKAAATAPQLTTLAAINAAGVASA
LKPDTAW***TILAPT******NKAF***ETRLAKL***NVT***ADALLKNKDLLVKVLSYHVIPSGAVYSSQLKDKQTVAT
ALK***NAT***LTVKIYGKNVAFKGPVNKAKVVADIKAG***KSVVHVIDDVLL***VPPGVVSPEVAQKWKAIEYK
SKKDASMKKMDKP

Crl. variabilis (EFN56922): pI=4.15, 1 putative glycosylation site

MARHSAQSYDNIGELLDDTPKLSILKAALEAAGLDELASDDALEA***TLFAPT******DKAF***ADAITAL***NLT***
AEEVLGDKDLLTTVLSYHVIPGVAVEAAALEEGQILQITILEGAAGQLKYYTFKGLPRLRTTADQV
ARITETDISAG***KAIVHLIDAVLI***PGDEFFSAPAPGPSP

Cmy. reinhardtii (XP001700577): pI=6.33, 15 putative glycosylation sites

MARWTRPGASTILVASCLLLAVAGSSAQ***Q***TI***LQAIQSANLT***LLAAGLQVADLATT***LNST***TG***PNITLF***
APSDALLAMTAALNTTPNALLTLGPKVAPVFTYHAITTPLLAADIP***TG***DT***PYATLNTGLNLT***TVSK
AA***NGS***VVVRA***LGSDANVIRTDITAGGAVVHVIDSVLLPFFLSV***ASAAT***TRTPQLSSLLSAVAS***NGLA
PALSNPNL***NVTVFAPVNDAF******NAS***SAYLTSV***NAS***ITDVLT***YHVATSRVLNG***SEVTAS***PITLTLNSR***
ANLTTVARDGNNVVVTPVGATAATVLAANIPVGLDLTTGGP***RTFVHLVNAVLL***PFYTTVANAAERAG
LTTLVAAVAASDPAFLAAVTDPTFRG***TILAPSNA******AF***AATLASL***NVT***AAQLLADKDNLRRI***LNAHII***
PNAAVFSNQLSN***QTVNTT***GGALLTVSIGNIVRFNAAKSSAKVVAADV***SIGNGRAVVHVIDRVLL***
PADLTLPTPYVTVAGAAQAANLSTLLAAVTASDASFLTTLTDPDFNGTVLAPTPLSDCHVALLMPN
AAFTAALTALGV***NAT***QLLADKDNLRRI***LNAHII***TPGAVRSSQLVNNQNITTLDGVLTVR***INGT***GVF
FVAAKSTAKVVTPDVIAG***NAVVHTIDFVLL***PASVTLTTGGSGGGGGSGGG***GAAGMASPSVLTLLSSA***
LAFLALFAFGRF

Figure S1. The deduced amino acid sequences of the AstaP orthologs. N-terminal signal peptides are highlighted by yellow. H1 and H2 motifs of the fasciclin domains are shown in boldface italics. Potential sites for N-linked glycosylation are shown in red fonts.

