

HHS Public Access

Author manuscript *Bioinformatics.* Author manuscript; available in PMC 2016 October 17.

Published in final edited form as:

Bioinformatics. 2007 October 1; 23(19): 2518-2521. doi:10.1093/bioinformatics/btm384.

FIST: a sensory domain for diverse signal transduction pathways in prokaryotes and ubiquitin signaling in eukaryotes

Kirill Borziak and Igor B. Zhulin

Joint Institute for Computational Sciences, The University of Tennessee – Oak Ridge National Laboratory, Oak Ridge, TN 37831-6173, USA

Igor B. Zhulin: joulineib@ornl.gov

Abstract

Motivation—Sensory domains that are conserved among Bacteria, Archaea and Eucarya are important detectors of common signals detected by living cells. Due to their high sequence divergence, sensory domains are difficult to identify. We systematically look for novel sensory domains using sensitive profile-based searches initiated with regions of signal transduction proteins where no known domains can be identified by current domain models.

Results—Using profile searches followed by multiple sequence alignment, structure prediction and domain architecture analysis, we have identified a novel sensory domain termed FIST, which is present in signal transduction proteins from Bacteria, Archaea and Eucarya. Chromosomal proximity of FIST-encoding genes to those coding for proteins involved in amino acid metabolism and transport suggest that FIST domains bind small ligands, such as amino acids.

1 INTRODUCTION

Signal transduction systems are information processing pathways that link environmental cues to adaptive responses in all living organisms. Major prokaryotic and eukaryotic signal transduction systems are different at the level of their principal components and complexity. Prokaryotic signal transduction pathways consist of simple one- and two-component systems (Ulrich *et al.*, 2005), whereas signal transduction in eukaryotes often involves multi-protein, branched cascades (Citri and Yarden, 2006). However, all living cells react to many similar signals that are present in the environment and inside cells (e.g. small molecules, such as amino acids and carbohydrates), therefore, some level of similarity is expected in their signal input elements. Indeed, several universal input domains have been described in sensory receptors from both prokaryotes and eukaryotes. The most abundant such domain is PAS (named after eukaryotic Per, ARNT and SIM proteins, where it was first described), which is found in diverse signal transduction pathways in organisms ranging from all major prokaryotic clades to humans (Ponting and Aravind, 1997; Zhulin *et al.*, 1997). PAS domains serve as detectors of small molecules, redox potential, oxygen, light and other important parameters (Taylor and Zhulin, 1999). Other input, sensory elements that are

Correspondence to: Igor B. Zhulin, joulineib@ornl.gov.

Conflict of Interest: none declared.

Supplementary information: http://genomics.ornl.gov/research/pubs/fist

found in both branches of life include small-ligand binding GAF (Aravind and Ponting, 1997), Cache (Anantharanam and Aravind, 2000) and CHASE (Anantharanam and Aravind, 2001; Mougel and Zhulin, 2001) domains. On the other hand, many other input domains are limited to bacterial signal transduction, e.g. MHYT (Galperin *et al.*, 2001), NIT (Shu *et al.*, 2003), CHASE2 through CHASE6 (Zhulin *et al.*, 2003) and 4HB_MCP (Ulrich and Zhulin, 2005), etc. Identification of novel sensory domains is a difficult task due to their extreme sequence variation, both in composition and in length. Sensitive similarity search methods, such as Position-Specific-Iterative (PSI) BLAST (Altschul *et al.*, 1997) must be employed to detect relationships between the domain family members, which must be further verified through a careful analysis of a multiple alignment of the domain family.

Recently developed MiST (Microbial Signal Transduction) database (Ulrich and Zhulin, 2007) enables rapid identification of sensory receptors where current domain models implemented in leading domain databases Pfam (Finn *et al.*, 2006) and SMART (Letunic *et al.*, 2006) do not detect any input domains. We carry out systematic similarity searches using protein regions of receptor proteins implicated in signal transduction that contain no identifiable domain. In many cases, such regions do contain known domains; however, current domain models are not sensitive enough to detect them. In other cases, such regions contain potentially new domain. In this report, we describe the identification of such novel domain, which we termed FIST, which is implicated in sensory reception in diverse signal transduction pathways in all three domains of life: Bacteria, Archaea and Eucarya.

2 DOMAIN IDENTIFICATION

During systematic analysis of microbial sensory proteins, we focused on a methyl-accepting chemotaxis protein (MCP) blr4191 from Bradyrhizobium japonicum (gi27352453), which was predicted to be a cytoplasmic chemotaxis receptor. All known chemotaxis receptors contain a sensory domain in their N-terminus followed by a conserved C-terminal signaling domain (Alexander and Zhulin, 2007), which can be identified by a Pfam domain model MCPsignal (accession PF00015). The blr4191 protein contained the full-length C-terminal MCPsignal domain, but lacked any detectable domain in its large (more than 400 amino acid residues) N-terminus. Exhaustive PSI-BLAST searches (E-value cutoff 0.01, compositionbiased statistics on, filter on) against the NCBI nr database (1 March 2006) were initiated with residues 1-449 that include the entire N-terminal region up to the first residue of the MCPsignal domain and continued until no new homologs were identified. Duplicate sequences were excluded from analysis. Profile searches yielded the list of domain family, which we termed FIST (F-box and intracellular signal transduction proteins), containing 176 proteins. Of these, 155 were full length matches, and 15 matched the C-terminal half of the domain (FIST_C) only. This search also revealed an overlap of the newly proposed domain with the Pfam domain of unknown function DUF1745 (also found in the COG database as COG3287, 'uncharacterized conserved protein'), which was detected in more than a half of proteins identified by PSI-BLAST searches.

Sequences that matched only the FIST_C subdomain all belong to F-box-containing eukaryotic proteins. When the region of these proteins that did not match the N-terminal half of the FIST domain (FIST_N) was subjected to a PSI-BLAST search, the only profile

matches returned were from the same subfamily of F-box proteins. However, searches initiated with FIST_C from these proteins returned most of full-length FIST domain proteins. Based on results of PSI-BLAST searches and predicted secondary structure, we built domain models from multiple alignments for FIST_N and FIST_C, as well as for the entire FIST domain, and carried out searches using the HMMER program to retrieve all homologs. Resulting sets contained 253 proteins with full-length FIST, 4 proteins with FIST_N only and 30 proteins with FIST_C only (nr database, 1 May 2007). Representative and complete multiple alignments of subdomains are provided as Figure 1 and Supplementary Material, correspondingly.

3 DOMAIN FEATURES AND ARCHITECTURE

Multiple alignment of full-length FIST domain sequences was built using ClustalX (Thompson et al., 1997) with default parameters and edited with the VISSA technique (Ulrich and Zhulin, 2005), which utilizes PSIPRED (McGuffin et al., 2000) to guide the editing process using predicted secondary structure. The final fold profile consists of 20 beta strands and 7 alpha helices. Multiple alignment of the FIST domain family revealed several highly conserved residues. The most conspicuous motif of the FIST domain family, which contains highly conserved cysteine and arginine, is found in the beta-strand 19 and the following loop (Fig. 1B). It may represent a site for ligand binding or protein-protein interactions. Using SMART (Letunic et al., 2006) and Pfam (Finn et al., 2006) domain architecture tools, we determined that in more than 70% of analyzed sequences the FIST domain comprises a single-domain protein; in the remaining sequences it is present in association with well-known signal transduction domains (Fig. 2) including the sensory PAS domain, GGDEF and EAL domains involved in the turnover of cyclic di-GMP, the MCP signaling domain, HisKA and HATPase_c domains of sensor histidine kinases, the DNAbinding helix-turn-helix TrmB domain and F-box domains of the ubiquitin signaling pathway. The search for transmembrane regions and signal peptides using Phobius (Kall et al., 2004) showed that most FIST-containing proteins analyzed in this study are predicted to be intracellular; however, signal peptides were predicted for a few stand-alone FIST-domain proteins indicating that they are likely to be extracellular.

4 BIOLOGICAL FUNCTION

FIST represents a new sensory (input) domain in signal transduction. As many other sensory domains, it is found either as a single-domain protein or exclusively in a combination with other domains that transmit signals from the sensory domain down the regulatory pathway, namely transducers (MCPsignal, HATPase_c domains) and output domains (GGDEF, EAL, TrmB). The FIST domain was found in four major classes of prokaryotic signal transduction: methyl-accepting chemotaxis proteins, sensor histidine kinases, di-guanylate cyclases and diesterases, and transcriptional regulators.

Further evidence suggesting FIST involvement in signal transduction comes from the analysis of the genome context (Overbeek *et al.*, 1999) of FIST-encoding genes. Using the MIST database (Ulrich and Zhulin, 2007), we have determined that in more than 30% cases, adjacent (immediately upstream or downstream) to the FIST-encoding gene there is a gene

coding for a known signal transduction protein. Overall, signal transduction genes in 198 prokaryotic genomes that contain FIST domains comprise <7% of the total number of genes (http://genomics.ornl.gov/mist). Thus there is a significant enrichment in signal transduction genes next to FIST-encoding genes. As a sensory domain, FIST is predicted to bind small molecule ligands. In eukaryotes, FIST is found in F-box proteins that are involved in ubiquitin mediated degradation of regulatory proteins, a frequent means of controlling progression through signaling pathways (Hershko and Ciechanover, 1998). Interestingly, genes encoding FIST-containing proteins are also often found next to genes coding for enzymes involved in amino acid metabolism (peptidases, aspartamine synthase) and amino acid transporters. In plants, F-box proteins are known to bind a plant hormone auxin, which is a derivative of tryptophan (Parry and Estelle, 2006). This further suggests that small molecule ligands detected by FIST might be amino acids and their derivatives. Detecting amino acid concentration both inside and outside the cell is important for many physiological processes, and it is likely that FIST-containing sensors play a significant role in converting these signals into changes in transcription, metabolism, development and behavior.

FIST domains are found in 16 phyla (Supplementary Table 1) representing all three kingdoms of life. This broad phyletic distribution suggests the ancient origin of this domain and further underscores its importance as a ubiquitous sensor module in diverse signal transduction pathways.

Acknowledgments

We thank Luke Ulrich for expert technical assistance and two anonymous reviewers for excellent suggestions. This work was supported in part by a grant GM072285 from the National Institutes of Health to I.B.Z.

References

- Alexander RP, Zhulin IB. Evolutionary genomics reveals conserved structural determinants of signaling and adaptation in microbial chemoreceptors. Proc Natl Acad Sci USA. 2007; 104:2885– 2890. [PubMed: 17299051]
- Altschul SF, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389–3402. [PubMed: 9254694]
- Anantharaman V, Aravind L. Cache a signaling domain common to animal Ca(2+)-channel subunits and a class of prokaryotic chemotaxis receptors. Trends Biochem Sci. 2000; 25:535–537. [PubMed: 11084361]
- Anantharaman V, Aravind L. The CHASE domain: a predicted ligand-binding module in plant cytokinin receptors and other eukaryotic and bacterial receptors. Trends Biochem Sci. 2001; 26:579–582. [PubMed: 11590000]
- Aravind L, Ponting CP. The GAF domain: an evolutionary link between diverse phototransducing proteins. Trends Biochem Sci. 1997; 22:458–459. [PubMed: 9433123]
- Citri A, Yarden Y. EGF-ERBB signalling: towards the systems level. Nat Rev Mol Cell Biol. 2006; 7:505–516. [PubMed: 16829981]
- Finn RD, et al. Pfam: clans, web tools and services. Nucleic Acids Res. 2006; 34:D247–D251. [PubMed: 16381856]
- Galperin MY, et al. MHYT, a new integral membrane sensor domain. FEMS Microbiol Lett. 2001; 205:17–23. [PubMed: 11728710]
- Hershko A, Ciechanover A. The ubiquitin system. Annu Rev Bbiochem. 1998; 67:425-479.

- Kall L, et al. A combined transmembrane topology and signal peptide prediction method. J Mol Biol. 2004; 338:1027–1036. [PubMed: 15111065]
- Letunic I, et al. SMART 5:domains in the context of genomes and networks. Nucleic Acids Res. 2006; 34:D257–D260. [PubMed: 16381859]
- McGuffin LJ, et al. The PSIPRED protein structure prediction server. Bioinformatics. 2000; 16:404–405. [PubMed: 10869041]
- Mougel C, Zhulin IB. CHASE: an extracellular sensing domain common to transmembrane receptors from prokaryotes, lower eukaryotes and plants. Trends Biochem Sci. 2001; 26:582–584. [PubMed: 11590001]
- Overbeek R, et al. The use of gene clusters to infer functional coupling. Proc Natl Acad Sci USA. 1999; 96:2896–2901. [PubMed: 10077608]
- Parry G, Estelle M. Auxin receptors: a new role for F-box proteins. Curr Opin Cell Biol. 2006; 18:152–156. [PubMed: 16488128]
- Ponting CP, Aravind L. PAS: a multifunctional domain family comes to light. Curr Biol. 1997; 7:R674–R677. [PubMed: 9382818]
- Shu CJ, et al. The NIT domain: a predicted nitrate-responsive module in bacterial sensory receptors. Trends Biochem Sci. 2003; 28:121–124. [PubMed: 12633990]
- Taylor BL, Zhulin IB. PAS domains: internal sensors of oxygen, redox potential, and light. Microbiol Mol Biol Rev. 1999; 63:479–506. [PubMed: 10357859]
- Thompson JD, et al. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 1997; 25:4876–4882. [PubMed: 9396791]
- Ulrich LE, et al. One-component systems dominate signal transduction in prokaryotes. Trends Microbiol. 2005; 13:52–56. [PubMed: 15680762]
- Ulrich LE, Zhulin IB. Four-helix bundle: a ubiquitous sensory module in prokaryotic signal transduction. Bioinformatics. 2005; 21(Suppl 3):iii45–iii48. [PubMed: 16306392]
- Ulrich LE, Zhulin IB. MiST: a microbial signal transduction database. Nucleic Acids Res. 2007; 35:D386–D390. [PubMed: 17135192]
- Zhulin IB, et al. PAS domain S-boxes in Archaea, Bacteria and sensors for oxygen and redox. Trends Biochem Sci. 1997; 22:331–333. [PubMed: 9301332]
- Zhulin IB, et al. Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in bacteria and archaea. J Bacteriol. 2003; 185:285–294. [PubMed: 12486065]

<pre>EEEEEE</pre>		
<pre>Hma 3 4 EPDFQ0TCSP-ATDYDAVLWGARSVVGSD-TEINCCSSGETTETG-SGNGT Whu 3 2 TFLAVLLGSR.SHTDQAVDLAAVQASVEPAALGCVAQGHELESPFA Chei 2 NLVQVFSGICNKEFINEVIRUKLLKLPQCKINQATSSGEILMO-DIFEREC Syn 6 3 RPUGFLVATEAPALNLSSNLTFLRETTKI-EHNGGVAPGLCVEDAFIRESCAV Mar 3 2 GONFGFLVATEAPALNLSSNLTFLRETTKI-EHNGGVAPGLCVEDAFIRESCAV Atha 3 3 RDCPVIANTCC-GNMEETITLITEVCSN-ACCHSSESKOIFNO-OPDTKGV Atha 3 3 PDCPVIANTCC-GNMEETITLITEVCSN-ACCHSSESKOIFNO-OPDTKGV Atha 3 7 PDCVIANTCC-GNMEETITLITEVCSN-ACCHSSESKOIFNO-OPDTKGV Atha 3 7 PDCVIANTCC-GNMEETITLITEVCSN-ACCHSSESKOIFNO-OPDTKGV Atha 3 7 PDCVIANTCC-GNMEETITLITEVCSN-ACCHSSESKOIFNO-OPDTKGV Atha 3 RDCPVIANTCC-GNMEETITLITEVCSN-ACCHSSESKOIFNO-OPDTKGV EEEEEE. EEEEEEEEE. HHHHHHHHHH WU VAVWLASCOPPAETFHLDFVRTSGALITGYR-6DLHLLDDP GNMag AAN(GLE)DDQFRVFICADAPKAKUH</pre>	Α	EEEEE
<pre>Mutub 32 TPALAVLLGSR-SHTDQAVDLLAAVQASVEPAALT CVAQCIVACRHELENEPA Syne 63 FRNLGILEVQFSGINNEFINKURLIKLIPC-CKIVATSSGETUCADAEIRDEC Syne 63 FRNLGILEVQSAPASETIRVLPLISELLEV DVLICGGGGIVGGHEIEEGFA Mari 23 QPDLIOFYANTD-IIETEDWANLLTHYCPT-ACREMSSEKQIFNO-OPDRKSV Atha 93 RPQFVIANTD-IIETEDWANLLTHYCPT-ACREMSSEKQIFNO-OPDRKSV Atha 93 RPQFVIANTD-IIETEDWANLLTHYCPT-ACREMSSEKQIFNO-OPDRKSV Atha 93 RPQFVIANTD-IIETEDWANLLTHYCPT-ACREMSSEKQIFNO-OPDRKSV Atha 93 RPQFVIANTDIETETDAVALLTHYCPT-ACREMSSEKQIFNO-OPDRKSV Atha 93 RPQFVIANTDIETETTAIIERTYLSELSEKQIFNO-OPDRKSV Atha 93 RPQFVIANTDIETETTAIIERTYLSELSEKQIFNO-OPDRKSV Atha 93 RPQFVIANTDIETETTAIIENTYLSELSEKQIFNO-OPDRKSV Atha 93 RPQFVIANTDIETETTAIIENTYLSELSEKQIFNO-OPDRKSV KUTGVVSSDEMEFFSSISTELSADPERCLFEAVHDLP-12RTIINLODG Cbei IISISVFEKTIIKSILVKDNICDFSKGVKIASLL-5KVIISFOBS Yne LSISLAVLPDPDFFFHLPORTESGALTTYK-6CPACILLEDP Cbei IISISVFEDDFFFHLINCAPICKQOLEDGLAPFSKHDLV-7FCUGUVGGD AAMIGHLEDDGFEVFLFANDLEDGLAPGANLEHDECDAADMAKHHFCPACILLEDP AMAI QLESTRA</pre>	Hmar	
Cbei 28 NILVQVFSGICNKEFINEVIRULKLLEPGCKIV&ATSSGEILMO-DIFEREEC Nmag 36 GANRGELVATEAPALNISSNLTFLEETTHI-DEHW GVAPGLCVEDAEIREEGA Mmag 36 GANRGELVATEAPALNISSNLTFLEETTHI-DEHW GVAPGLCVEDAEIREEGA Mmar 32 GOPLIOFYANTO-ITETERDWAILTHYCEN-ACCHMSSEKCIPHO-OPOTKGY Atha 93 RPGPVIAHITCGNMEETITLITERVGSRVPIIVSVTGILGKEA-CNDKAGE 80% .t.hhhhans.p.thht.htl.hhlsosset.lhth.tps. EEREEEE. EREEEEEEE. HHHHHHHHHHH WU VAVWSSOUPAETFHLDFVRTOSGALITGYE-6DLHLLDDP LSLSIAVLPDVALHPFVIRGNQLPDLDAPGAMULVG-7HFLLLAGGE Nmag AAMIGHLPPDGFRVFILSIVGDAADWAKHFFPCGULVGD Mmag AAMIGHLPPDGFRVFILSIVGDAADWAKHFFPCGULVGD Mmag AAMIGHLPPDGFRVFILSIVGDAADWAKHFFPCGULVGD Atha VILHSTSDD-19-MKVDIIPUQAKGEGAABMEDKYVDIINVNS-8PACILLPAE thus VSFPSLIJERLTGLPG-TTVVGCSAGD-DLOLGCTHVFRDD-RVETDAVULAL Atha NAGIGMKVTBLFGVLDEG-TTVVGCSAGG-CGRRGDTLFEBVNB-6PAGUVAGD SSR-1SELJGGLDA-YPGAVKWGLASGG-RGFRGNALFLDA-ELYREGTUGUGWH Hub VSFPSLIJERLINTDLPGTVVGCSAGG-CGRRGDTLFLDA-ELYREGTUGUGWH Hub SSR-1SELJGGLDA-YPGAVKWGLASGG-RGFRGNALFLDA-ELYREGTUGUAL STM SSR-1SELJGGLDA-YPGAVKWGLASGG-RGFRGNALFLDA-ELYREGTUGUAL ALD THARTPVLHKDYVNGLGCGFRI-GBVVGGIGSFLHKKSEPENVULQUDUDUGUAULAAVAA SSR-1SELJGGLDA-YPGAVKWGLASGG-GGRRGNALFLDA-ELYREGTUGUAL HUT SSSSLFSQUNTHKTKRCGDIGGLHSSNGTWVLQDDLUGIAULAFA ADTATEPVLIKEDVTUNGHEEI-SEANTVFWAD-GSTVYELGGPP 234 S5377251 Mmag GSGIEVTCMTQGCSPL-GEVHVWGRAS-GSTVYELGGPP 234 S53755 B	Mtub	32 TPALAVLLGSR-SHTDQAVDLLAAVQASVEPAALICCVAQGIVAGRHELENEPA
<pre>Syne 6: PRILGLEVSAAFASETIVLPLISELLEV -DVLNCGCGGGIVGGGHEEEGPA Mari 2: OPDLIGYXANTD-IIETROWNLLTHYCEN-ACRIGMSEKOIFNO-OPDTKGV Mari 2: OPDLIGYXANTD-IIETROWNLLTHYCEN-ACRIGMSEKOIFNO-OPDTKGV Atha 9: PCQFVIANTD-IIETROWNLLTHYCEN-ACRIGMSEKOIFNO-OPDTKGV st. hhhhahsp. thth.htt.hhlgsost.lhthtps. EEEEEEPEEEEEEEHHHHHHHHHHHEEEEE UTVGVVSSOBMEFFSSISTEISADPERCLFEAVHDLP-12FTIINLEDG Syne LSISIAVLPDVALHEPTIRONOLDDLDAPSAWHDLV-12FTIINLEDG Syne LSISIAVLPDVALHEPTIRONOLDDLDAPSAWHDLV-5KVIISFGDS Syne LSISIAVLPDOPENTFHLDFYRTCSKGVKIASLL-5KVIISFGDS Syne LSISIAVLPDOPENTFHLDFYRTCBADMAKHHPCVGLVHGD Mari TLICSYFNATTLHKNGLVKNNILDDLDAPSAWHDLV-5FVIISFGDS Na AMIGHLPDOPENTFLEND-SADPERCHLPFAVHDIRON</pre>		
<pre>Maag 36 GANFGFLYATEAPALNLSSNLTFLRETTH: EHNG GVAPCLCVEDAEIDEGAV Atha 39 RPGPUIAWITCGRMEETLTLITERVGSNLTFLSWIGGNSSECIPNO-OPOTKSV Atha 39 RPGPUIAWITCGRMEETLTLITERVGSNVEIUSWUTGILGKEA-CNDRAGE 80% .t.hhhhahsp.tht.ht.hhlGsosst.lhth.tps. EEEEEEEEEEEEEEHHHHHHHHHHEEEEEE Haar VUTGVSSOUPARTFHLDFVRTGSGALITGYR-6DHLLLDDF Coli ISISLAVLPDVALHPFVLGNQLPULDAPFSANTDLVG-7HLLLADG Syne LSISLAVLPDVALHPFVLGNQLPULDAPFSANTDLVG-7HLLLADG Mari ILICSYFRATTLHKSILVKDNICDFSGVKIASLSCVUISGD Mari ILICSYFRATTLHKSILVKDNILNGSCHTLFPAYME-6PACIULADA Atha VRLHSTSDD-19-MKVDIIPVIGNGCLNGSCHTLFPAYME-6PACIULADA Atha VRLHSTSDD-19-MKVDIIPVIGNGCLNGSCHTLFPAYME-6PACIULADA Atha VRLHSTSDD-19-MKVDIIPVIGNGCON-DLGLQCTNVFRND-RVETDAVULTI Nucl Syne SSR ISELLQCUDFAYFOAVKYQGSAGBAEMCDFYVDIEN-PVITTSDAVULTI SIM-GBELLQCUDFAYFOAVKYQGIASGLSSNGTWVITFE-OVLKSNNVAAVLA Syne SSR ISELLQCUDFAYFOAVKYQGIGSFLHKKGBERNVQLQUDLQCHAVUAFA ALD FTARTFPVILKELDYNNGGCRFI-CBCNGCTELEDER-DVLTSGUUGUEL LSISSSLFSQUNTKKTKRCGUSGLSSNGTWVIDYDCLQCHAVUAFA ALD FTARTFPVILKELDYNNNSSFNCIFI-GKEHTINGR-GSTVFELDGRP 234 55377251 Mmag GSGIEVITNNNGREFI-SEAMTVMERAE-GSTVFELDGRP 234 55377251 Mmag GSGIEVITNNNSSFNCIFI-GKEHTINGR-GSTVFELDGRP 234 55377251 Mmag GSGIEVITNNNSSFNCIFI-GKEHTINGRGSTVFELDGRP 234 55377251 Mmag GSGIEVITNNNSSFNCIFI-GKEHTINGRGSTVFELDGRP 234 55377251 Mmag GSGIEVITNNNSSFNCIFI-GKEHTINGRCSTVFELDGRP 234 55377251 Mmag GSGIEVITNNNSSFNCIFI-GKEHTINGRCSTVFELDGRP 234 55377251 Mmag GSGIEVITNNNSSFNCIFI-GKEHTINGRSGTVFELDGRP 234 55377251 Mmag GSGIEVITNNNSSFNCIFI-GKEHTINGRNINVSGOES 319 26451740 80% .t.t.Lhthht.hhthph.s.hhlTtsp.thlhpltF FMT GGAHSVSVVGGCRFI-GEVITVGENANNNYSGESC 319 26451740 80% .t.t.hhhthHHHHHHHHHHHHHHHHHHHHHHHHHHHHH</pre>	Syne	63 RPNLGILFVSAAFASEYIRVLPLLSELLEV-DVLIGCSGGGIVGGGHEIEEGPA
<pre>Mari 23 QPDLIGFYANTD-IIETEDWVMLITHYCEN-ACREWSEKQIFNO-OPDYKSV Acha 93 RPOYIANTC-IMMERTIFIETEWGSRVPIISVUTGILGKEA-CNDKAGE 80% .t.hhhhahsp.tht.htt.hhlGsostlht.c.thtps. EEEEEEEEEEEEEEHHHHHHHHHH EEEEECBMEETFSIJTESADPERCLFEAVHDLP-12RTINLEDG Chei IISISVFEKTIIKSILVKDNICDFSKGWKIASLL-5KVIISFOBS yme LSUSLAUPEDUPAETFHLDPYRTCSGALITGYF-G-DLHLLADGF Maag AAMIGHLBPDOFEVFICDFSKGWKIASLL-5KVIISFOBS AAMIGHLBPDOFEVFICDFAKGWKIASLL-5KVIISFOBS AAMIGHLBDDOFEVFICDFAKGWKIASLL-5KVIISFOBS AAMIGHLBTLIHNCAPIGKDQLDDLDAPGKHTUF-7PCUGLVKGD Maag AAMIGHLBTLIHNCAPIGKDQLDDLDAPGKHTUF-7FCUGLVKGD Mari LICSYFNATLIHNCAPIGKDQLDULDAFFSMIDLVG-7FKLLLADGF May RAMIGHLBPDOFEVFIGDAADMKHHFPAGIVLADD HAV VRLHSTSDD-19-MKVDIIPVIQAKGESGABMEDKYVMDIRNNMS-8PAGLILEAE 80% 1.hhhhh.h.h.h.t.t.h.h.hhhhh.sh REHENTOLGGTTVVGVVSG-GRRCDTLFERD-VUTSGUGVRL Chei SIN-GEELLAGINS-IND-HVUXAGIAGK-SDPAYETYVFTEE-GVEKMAVAAVAL IISSSSLFSQVNTHGGATVVGOUSG-GRRCDTLFIRRD-DVITSGUGVGLAV Maag RPD-PDVFKAVTDLARAGFLAGCLKAASGPAADLAG-IATDGGLAGVLL III LSSSSLFSQVNTVAGGLAGG-GGSUVELDGES-LVKRAVAAVAL III LSSSSLFSQVNTVAGGLAGG-GUGSGLBSNGTVUVQOULDEEHIVVAFA Atha DTHATEVULKKLDYAMPA-ETVIVGGUGSF-LKSNGTVVVQUDLDELAVIAVAFA Atha DTHATEVULKKLDYAMPA-ETVIVGGUGSF-LKSNGTVVLQUDLLGEHAVUVAFA Atha C</pre>	Mmag	36 GANFGFLYATEAFALNLSSMLTFLRETTRI-EHWVGGVAPGLCVEDAEIRDEGAV
Atha 93 RPOPVIANITCGMMEETITLITERVGSRVEIIVSVVTGILGKEA-CNDRAGE 80% .t.hhhhahsp.tht.ht.hhlGsosst.lhth.tps. EEEEEE. EEEEEEEEE. HHHHHHHHHH. EEEEEE. EEEEEEEE. EEEEEEEEEE. HHHHHHHHHH. EEEEEE. Mund VAUWIASOPPATFHLDFVRTGSGALTGYR-GDHLLLDDGF Mund VAUWIAS		
<pre>80% .t.hhhhahsp.thht.htt.hhlGsosst.lhth.tps. 80% .t.hhhhahsp.tht.htt.hhlGsosst.lhth.tps. EEEEEEEEEEEEEEHHHHHHHHHH WTVGVVSSDGMEFFSSISTELSADPERCLFEAVHDLP-12RTINLEDG FMAGE SADPERCFLPFVRCDSKGWITASLL-5KUIISPOB Syme LSLSLAVLPDVALHPYLRGNQLDDLDAPPSAWIDLVG-7HFLLADGF Maari TLICSYFNATTLHHCAPIGKDQLDDLDAPPSAWIDLVG-7HFLLADGF Maari TLICSYFNATTLHHCAPIGKDQLDDLDAPPSAWIDLVG-8PRCJUKADD Mari TLICSYFNATTLHHCAPIGKDQLDQLDAPPSAWIDLVG-8PRCJUKADD Nati VELKSTSDD-19-MKVDIFVIQAKESGSABMEDKFVMDIRNYMS-8PRCJULFAB 80% l.hhhhh.h.h.h.t.t.h.h.h. MAGIGNKVTKLTEQVLDDEF TFVVGCSAGD-DLQLQCHVFRND-RVSTDAVVILI NSSSSIFSQVTHDEGTTVVGCSAGDDLQLQCHVFRND-RVSTDAVVILI SIN-GEELINGINS-INDHVUXGCIAGK-SDPAYETVYFTE-GVEKNAVAAVAL Mari LSISSSIFSQVTHFDGATVVGCJAGGERGONALFLDA-ELYREGTGLAA A:-AEOALDVTNHCHGEI-SEMTVVGCJAGG-RGCNALFLDA-ELYREGTGLAA A:-SSSSIFSQVTHFGAVVIGALSG-RGFORMIFLDA-ELYREGTGLAGA Mari SISSIFSQVTHGAVVIGALSG-GTVIELGGRP 234 55377251 Mari LSISSSIFSQVTHGAVVIGALGGL-GSTVIELGGRP 234 55377251 Mari S-DVEKAVTOLABACFINGULASGTVIELGGRP 234 55377251 Mmag GSGIEVITMNRSSNCIFI-GKEHILVER-DNIVKKIGTIS 217 82748819 NFSGNVVLDAVVAGCFI-GDEVIVGRAB-GSTVIELGGRP 234 55377251 Mmag GSGIEVITMNRSSNCIFI-GKEHILVER-DNIVKKIGTIS 217 82748819 NFS -S-CNVUDAVVAGCFI-GEVIVGRAB-GSTVIELGGRP 220 81255484 K-GEFLNVTNNRSSNCIFI-GKEHILVER-DNIVKKIGTIS 217 8274819 NFS -S-CNVUDAVVAGCFI-GEVIVGRAB-GSTVIELGGRP 221 8215598 B .HHHHHHHHH Mari 235 -AFEAMPDAIRED-14-GEEDIVLCAANVNNSLGPSC 319 26451740 Nmag GSGIEVITMMTGGCSL-GEVIVWERANVNNSLGPSC 319 26451740 Nmag GSGIEVITMOHOGCSL-GEVIVWERANVNNSLGPSC 319 26451740 Nmag 122 -LUVLEEVGUACCHEDI-SKNKVRAMNVNSLGPSC 319 26451740 Nmag 123 -ASNYLQUVSTFDMILLGGQVDLSSLTSE-10-IDASGVVGLSFGHRI HMAT 215 -APGENRDAIRED-14-GEEDIVLLGRYELGIESE-22-TTGNINKWGHTTD MATI 215 -APGENRDAIRED-14-GEEDIVLLGRYELGIESE-22-TTGNINKWGHTTD MATI 215 -APGENRDAIRED-14-GEEDIVLLGRYELGIESE-20-TGNINKWGHTD NAGIGIMSELCPDAVTINSNSQVYAVANAANANANNELEGGSVAGGFVTTA MATI 25 -AQ</pre>		
<pre>Hmar VTVGVVSSDSMRFFSSISTELSADFERCLFEAVHDLP-12-RTIINLHDD Kub VAVWLASSGPAETHILFVFNTGSGALTGXRCDVILLPD Cbei IISISVEKTIIKSIUKDNICDGSKGVKIASLL-5-KVIISFDDS Syne LSLSLAVLPDVALHPFYLRGNQLPDLDAPPSNKIDLGTKUISFDG Mari TLICSYFNATTLIHNCAPIGKQCLNGSCHTLFEVNE6-PAGIVLADG Mari TLICSYFNATTLIHNCAPIGKQCLNGSCHTLFEVNE6-PAGIVLADG Mari TLICSYFNATTLIHNCAPIGKQCLNGSCHTLFEVNE6-PAGIVLADG Mari LSISTSDD-19-WCVGVIGC-GREKVFWDININYSS-8PACILIFAE 80% 1.hhhhh.h.h.ht.h.h.h.h.hhhhh.shHHHHHHHHHHH.EEEEEEEE.EEE.EEE.EEE.</pre>	80%	
Cbei IISISVFEKTIIKSIUKDNICDFSKGYKIASLL-5KVIISFCDS Mmag AMIGHLPPDQFRVFICJDADAPSAMIDUG-7HLLADGG Mmag AMIGHLPPDQFRVFIGDAADWAKRHFPCVGLVHGD Mari TLICSYFNATTLIHNCAPIGKDQLMQSCHTLFEPVNE-6-PAGIVLADD AAMIGHLPPDQFRVFIGDAADWAKRHFPCVGLVHGD Mari LICSYFNATTLIHNCAPIGKDQLMQSCHTLFEPVNE-6PAGIVLADD ALA VRLHSTSDD-19-WKUDI PV1CAKGESGABMEKFVMDIRNYNS-8PACILIEAE 80% 1.hhhhh.h.ht.h.h.hb.hhhh.sh HHHHHHHHHHEEEEEEEEEE.EEEEE.EEEEEE		VTVGVVSSDSMRFFSSISTELSADPERCLFEAVHDLP-12RTIINLHDG
<pre>Syne LsLSLAVLPDVALMPFYLRGNQLPDLDAPPSAWIDLVG7HFLLLADGF Mmag AAMIGHLPPDQFRVFI TLICSYFNATTLIHNCAPIGKDQLDNQSCHTLFEPVNE6PAGIVLADG Atha VRLHSTSDD-19-MKVDIIVUQAKGESGAEMEDKFVMDIRNYMS8PACLILFAB 80% 1.hhhh.sh.hh.h.m.t.hhhh.sh HHHHHHHHHHEEEEEEEE.EEEEEEEEEEE EEEEE.EE.</pre>	Mtub	VAVWLASGPPAETFHLDFVRTGSGALITGYR6DLHLLLPDP
<pre>AMMIGHLPPDQFVPIGDADWAKRHFCVGUHHGD Mari TLIGYPNAPAGUILAD Atha VRLHSTSDD-19-MKVDIIPVIQAKGESGAEMEDKFVMDIRNYMS8PAGUILAD Atha VRLHSTSDD-19-MKVDIIPVIQAKGESGAEMEDKFVMDIRNYMS8PAGUILAD Atha VRLHSTSDD-19-MKVDIIPVIQAKGESGAEMEDKFVMDIRNYMS8PAGUILIFAE 80% 1.hhhhhh.h.ht.h.h.h.h.h.h.h.h.h.h.</pre>	Cbei	IISISVFEKTIIKSILVKDNICDFSKGVKIASLL5KVIISFGDS
<pre>Mari TLICSYFNATTLIHNCAPIGKDQLNQSCHTLPEPVNE6PAGIVLADD Atha VLHKSTEDD-19-MKVDIIFVIGKAGESGAEMEDKYVDITTVE-6PAGIVLADD Atha VLHKSTEDD-19-MKVDIIFVIGKAGESGAEMEDKYVDITVTWS-0-8-PAGLILFAB 80% 1.hhhhhhh.h.ht.h.h.h.h.h.hhhh,shHHHHHHHHHHHEEEEEEEEEEEEEEEEEEBEEEEE Hmar MAGIGNKVTRLTEQYLDDEE-TPVVGSAGD-DLQLQQTHVFRND-RVETDAVVLTLI YSFPSNLIERLNTDLPGTTVVGSAGD-OLQLQQTHVFRND-RVETDAVVLTLI Syne SSR-ISELLQGLDFAYPGAVKVGGLASGG-RGPRGNALFLDA-ELYREGTVGLAL Mari CEELLEGINAS.IND-HVUVAGGIAGKSDPAYETYVTEE-GVENAVAAVAL Mari LSISSSLFSQVNTHKKFCGLGGLHSSNQTWVLYQDQLAGHAVLAGIAFA Atha DTHATEPVLHKLDYAMPA-ETVIVGQJIGEFLHKRGNEPRNVQLQKDDIRVLAGLIFA 80% t.t.thsth.hshhGG.uus.tttthhhhtt.thhtpshhhhhhEEEEEEEEEEE Hmar AAEDALPVTVNGHEPI-SEAMTVGRAE-GSTYYELDGRP 234 55377251 Mtub PGAHSVSVVSQCCPFI-GEPYIVGAD-GAVITELGGRP 220 81255484 Cbei K-GEFLNVNNRSSNPCTFI-GKHENTUVE-NONUKKIGTIS 217 82748819 Syne SGNVVLDAVVAQGCRFI-GDPLRVEAE-GNVILSLEGRP 267 8666541 Mmag GSGIEVITSMTGGSPJ-GEVHTVGRAP-ONILQCIDHP 214 8718909 Atha R-8-ERIQFDTAISNOMSSVDLRYKAANVNVSLGPSC 319 26451740 80%th.h.hhtthpsh.s.hhlTtsp.tphhhplstp. BHHHHHHHHHPCDEVLVSRGLGIGIVUD7-QGDFLIRGLGAP-CD 214 -PLKRLKEIVLGMAPDEQELVSRGLGIGIVUD7-QGDFLIRGLGAP-CD 224 -PLKUREMKTLGMNDUCKKIGNEPLAKKYNKYSLGPSC 319 26451740 80%th.h.hhtthpsh.s.hhlTtsp.tphhhplstp. BHHHHHHHHHEEEEHHHHHHHHHHHHHH</pre>	Syne	LSLSLAVLPDVALHPFYLRGNQLPDLDAPPSAWIDLVG7HFLLLADGF
Atha VRLHSTSDD-19-MKVDIIPVIQAKGESGAEMEDKFVMDIRNYMG8PACLILFAE 80% 1.hhhhhh.h.hEEEEEEEE	Mmag	AAMIGHLPPDQFRVF1GDAADWAKRHFPCVGLVHGD
<pre>80% 1.hhhhhhhh.ht.t.h.h.h.h.h.hhhh.sh HHHHHHHHHHEEEEEEEEEEEEEEEEEEE</pre>	Marı	TLICSYFNATTLIHNCAPIGKDQLNQSCHTLFEPVNE6PAGIVLADD
<pre>MAGIGNKVTRLTEQYLDDEZ-TPVVGSAGDDLQLQQTHVFRND-RVETDAVULTLI MAGIGNKVTRLTEQYLDDEZ-TPVVGSAGDDLQLQQTHVFRND-RVETDAVULTLI Chei SIM-GEELLNGINS-IND-HVLVAGGLASGSDPAYETYVFTEE-GVEKNAVAAVAL SYNE SSR-ISELLQGIDFAYEOAVKVGGLASG-SDPAYETYVFTEE-GVEKNAVAAVAL SYNE SSR-ISELLQGIDFAYEOAVKVGGLASG-SDPAYETYVFTEE-GVEKNAVAAVAL SYNE SSR-ISELLQGIDFAYEOAVKVGGLASG-SDPAYETYVFTEE-GVEKNAVAAVAL Mari LSISSSLFSQVNTHKTKFGGLSGLHSSNQTWVLYQQLLQEHAVLVAFA Atha DTHATEVVLKKLDYAMPAETVIVGGIGEFLHKRONEPRNVQLQKDDIRVLAGUIFA 80% .t.t.lhtsht.hsh.hhGG.uus.tttthhhhtt.thhtpshhhhh EEEEEEEE.EEEEEEEE Hmar AAEDALPVTVNHGHEPI-SEAMTVMEARGXVTYELDGRP 234 55377251 Mtub PGARVSUVGGCCPI-GEPTIVGAD-GAVTELGGRP 220 81255484 Chei KGEFLNVNNGSSFNCIPI-GKEHEINUVE-DNIVKKIGTIS 217 82748819 Syne SGIVUTINDGCSPI-GKEHEINUVE-ONIVKKIGTIS 217 82748819 Syne SSGIEVLTGMTQCSPI-GKHTWIESAGXVILLSLEGRP 267 86606541 marg GSGIEVLTGMTQCSPI-GVHTWEAEGXVILLSLEGRP 267 86606541 80%th.h.hh.tthpsh.s.hhlTtsptphhlplstp.</pre> B HHHHHHHHHHHHHHHH.EEEEPEEEE.EEEE. Hmar 235 -AFEAWRDAIRED-14-GSEDLVMLLGRYELGIESE-22-TTGYNIRWPGHTTD Mtub 221 -PLKRLKEIVLGMAPDEQUSNGLQIGIVUD7-QGDFLIRGLGADP Chei 218 -VKEFYNKIGNNDDQILWSRGLQIGISE-22-TTGYNIRWPGHTTD Mari 215 -APEAWRDAIRED-14-GSEDLVMLLGRYELGIESE-22-TTGYNIRWPGHTTD Mari 215 -APEAWRDAIRED-14-GSEDLVMLLGRYELGIESE-22-TTGYNIRWPGHTTD Mari 215 -AQUVFNYLGNGSP5DQRLARQALFIGLLMD7-SOFDJLIVLGIDP Mari 215 -AQUVFNYLGNGSP5DQRLARQALFIGLMDD-SHDYQWTILGIP Mari 215 -AQUVFNYLGNGSP5DQRLARQALFIGLWDNIXJSFHHIKKINA- 222 -ASNYLQQVYSTSDMNIILAGQVONLSSLTSE-10-IDASGVUSLSFSHERI 30% .sht.htphlth.slhhh.tt.hhps.thp. EEEEEEEEEE. HHHHHHHHHHHHHHHHHHHH		
<pre>MaGIGNKVTRLTEQYLDDEE - TPVVGGVAGD - DLQLQQTHVFRND - RVETDAVUTLL Mub YSFPSNLLIELLNTDLPG TVVVGGVAGG - GRRGORDTLFRDE - DUTXSGLVGVRL Cbei SIM-GEELLNGINS - IND - HVLVAGGIAGG - SDPAYETYVFTEE - GVEKNAVAAVAL SSR-ISELLQGLDFAYPG AVKVGGLAGG - SDPAYETYVFTEE - GVEKNAVAAVAL SSR-ISELLQGLDFAYPG AVKVGGLAGG - SDPAYETYVFTEE - GVEKNAVAAVAL Mari LSISSSLFSQVNTHKT FKGGLSG LHSSNQTVVLVQDQLLQEHAVLVAFA Atha DTHATEPVLHKLDYAMPA - ETVIVGQIGEFLHKSGNEPRNVQLQKDDIRVLAGLIFA 80% .t. thist.hsh.hnGG.uus.ttetthhnhtt.thhtpshhhhh EEEEEEEE. EEEEEEEEEEEEEE. Hmar A AEDALPVTVNHGHEPI - SEMETVVRAE - GSTVYELDGRP 214 55377251 Mtub PGAHSVSVVSQGCRPI - GEPTIVGAD - GAVITELGGRP 220 81255484 K - GEFLMVNNRSSFNCIPI - GKEHEILVE - DNIVKKIGTIS 217 82748819 Syne S GNVLDAVAQGCPI - GPPIVIGAD - GAVITELGGRP 212 23015999 Mari S DKLVIENDAFVDAGIDI - GKEHEILVE - DNIVKKIGTIS 217 82748819 Syne S OKUVLENDAFVDAGIDI - GKHKWIMAQ - DNILQQIDHLP 214 87118909 Atha R - 8 ERIQPTTAISNMSS VUDIXKAA NVNSLGSC 319 26451740 80%th.h.htthpsh.s.hhlTsp.tphlhplstp.</pre> B HHHHHHHHHHHHHHHHHEEEE EEEEE. EEE Hmar 235 - AFEAWRDAIRED 14 - GSEDLVMLLGRYELGIESE - 2 - TTGYNIRWPGHTTD Mub 221 - PLHELREIVLGM SPSDQLVARGIQIIVD - 7 - GODFLIRVILGDP Che 218 - VKEFYNKJCTM NNDQILKRGNQIIVD - 7 - GODFLIRVILGDP ADVULEVEGEL SPSDQLVARGAPIHVGLPAEGD SHD VRTLIGIDP Mmag 212 - ALDVLKEEVGEL LARDLRIAGYHVGLPAEGD SHD VRTLIGIDP Mmag 212 - ALDVLKEEVGEL STDIKVNNLFALKYKTS HKETHSVPLSFAE- Hsap 232 - ASNYLQUVSTFSD - MNIILAGGVVDLRSLTFE - 10 - IDASGVVGLSFSGHRI 30%t.th.ph.t	80%	1.nnnnn.nn.ntnn
Hmar AAEDALPYTVNHGHEPI-SEAMYVERAEGSTVYELDGRP 234 55377251 Mtub PAEDALPYTVNHGHEPI-SEAMYVERAEGNVIELDGRP 234 55377251 Mtub SGANYVSVSQGCRPI-GEPYIVGADGAVITELGGRP 220 81255484 Chei KGEFLAVINNRSSFNCIPI-GEPYIVGADGAVITELGGRP 220 81255484 Syne SCRNVULDAVVAQCCRPI-GPLEVERAECNVILSLEGRP 267 8666541 Mmag GSGIEVLTGMTQCSPL-GEVHTVBESWQGVVMALGGRP 212 23015999 Aris SDKLVIENDAPVDSIDI-GKNKNVMAQDNILQQIDHLP 214 8718909 Aris sth.hhh.tthpsh.shhlttsptphlhplstp. 26451740 So% th.hhh.tthpsh.shhlttsptphlhplstp. 26451740 B th.hhh.tthpsh.shhlttsptphlhplstp. 26451740 Mtub PLAVLQLAELNPDOGLUSGLQIGIVDV-7-QGDFLIRGLLGAPP 216 Chei 218 -AFEAMRDAIRED-14-GSEDLWALLGRYELGIESE-22-TTGYNIRWPGHTD Mtub TULHELREIULGANDEOGLUSGLQIGIUDV-7-QGDFLIRGLLGAPP- Mmag (2) -ALDVLKEEVGELLARDLRRIAGYIHVGLPAEGDO-SHDYQVRVILGIDP Mmag (2) -ALDVLKEEVGELLARDLRRIAGYIHVGLPAEGDD-SHDYQVRUTULSIGIDP Mari .sht.htphlthuslhht.tt.hhps.t.hph.tthps.tsFAE- Hsap 232 -ASNYLQVVSTSDMNIILAGGQVVNLS	Hmar Mtub Cbei Syne Mmag Mari Atha 80%	MAGIGNKVTRLTEQYLDDEE - TPVVGGSAGD - DLQLQQTHVFRND - RVETDAVULTU YSFPSNLLIERLNTDLPG TTVVGGVVSG - GRRGDTRLFRDR-DVLTSGLVGVRL SIM - GEELLNGINS - IND - HVLVAGGIAGK - SDPAYETYVFTEE - GVEKNAVAAVAL SSR - ISELLQGLDFAYPG AVKVGGLASGG - RGPRGNALFLLDA - ELYREGTVGLAL PRD - PDVFKAVTDLAAEA GFLAGGLMSA SOPAAQLAG - LATDGGLAGVLL LSISSSSFSQNTIKKT FKCGGLSG HSSNQTWVLYODQLLQEHAVLVAFA DTHATEPVLHKLDYAMPA ETVIVGQ
Hmar AAEDALPYTVNHGHEPI-SEAMYVERAEGSTVYELDGRP 234 55377251 Mtub PAEDALPYTVNHGHEPI-SEAMYVERAEGNVIELDGRP 234 55377251 Mtub SGANYVSVSQGCRPI-GEPYIVGADGAVITELGGRP 220 81255484 Chei KGEFLAVINNRSSFNCIPI-GEPYIVGADGAVITELGGRP 220 81255484 Syne SCRNVULDAVVAQCCRPI-GPLEVERAECNVILSLEGRP 267 8666541 Mmag GSGIEVLTGMTQCSPL-GEVHTVBESWQGVVMALGGRP 212 23015999 Aris SDKLVIENDAPVDSIDI-GKNKNVMAQDNILQQIDHLP 214 8718909 Aris sth.hhh.tthpsh.shhlttsptphlhplstp. 26451740 So% th.hhh.tthpsh.shhlttsptphlhplstp. 26451740 B th.hhh.tthpsh.shhlttsptphlhplstp. 26451740 Mtub PLAVLQLAELNPDOGLUSGLQIGIVDV-7-QGDFLIRGLLGAPP 216 Chei 218 -AFEAMRDAIRED-14-GSEDLWALLGRYELGIESE-22-TTGYNIRWPGHTD Mtub TULHELREIULGANDEOGLUSGLQIGIUDV-7-QGDFLIRGLLGAPP- Mmag (2) -ALDVLKEEVGELLARDLRRIAGYIHVGLPAEGDO-SHDYQVRVILGIDP Mmag (2) -ALDVLKEEVGELLARDLRRIAGYIHVGLPAEGDD-SHDYQVRUTULSIGIDP Mari .sht.htphlthuslhht.tt.hhps.t.hph.tthps.tsFAE- Hsap 232 -ASNYLQVVSTSDMNIILAGGQVVNLS		EFFEFFE EF FFFFFE FFFFF
<pre>Mtub PGABSVSVVSQGCRPI-GEPITVGADGAVITELGGRP 220 8125544 Cbei KGEFLNVNNRSSPNCIPI-GEPITVGADGAVITELGGRP 220 8125544 Cbei KGEFLNVNNRSSPNCIPI-GEPITVGEDONVILSLEGRP 267 86606541 Mmag GSGIEVITGMTQGCSPI-GEVHTVGESWQVVVALDGRP 212 23015999 Mari SOKLVIENDAPVDSIDI-GKVMTVGEWNILQUIDHLP 214 87118909 Atha R-8-ERIQFDTAISNGMSSVDLRYKAANVNVSLGPSC 319 26451740 80%th.hhh.tthpsh.shhlTtsptphlhplstp. 8</pre>	Hmar	AAEDALPVTVNHGHEPI-SEAMTVWRAEGSTVYELDGRP 234 55377251
Cbei KGEFLMVINNRSSFNCIPI-GREHEIDLVE-DNIVKKIGTIS 217 82748819 Syne SGRVVLDAVVAQGCRPI-GDPLRVGAEGNVILSLGGRP 267 86606541 Mag GSGIEVLTGMTQGCSPL-GEVHTVDESWQGVVMALDGRP 212 23015999 Mari SDKLVIENDAFVDSIDI-GKKMKVDAMQDNILQQIDHLP 214 87118909 Atha R-8-ERIGPTDISNGMSSVDLRYKAANVNVSLGPSC 319 26451740 80%th.h.thtpsh.shhltsptphlhplstp. B .HHHHHHHHHHHHHHHHHEEEEEEEEEEE Hmar 235 -AFEAWRDAIRED-14-GSEDLVMLLGRYELGIESE-22-TTGYNIRWPGHTTD Mtub 221 -PLHELREIVLGMAPDEQELVSRGLQIGIVVD-7-QGDFLIRGLLGADP Cbei 218 -VKEFYNKYLGTMNNDQILKMKNFPLMVKRNNRYFFTHLKFINA Syne 268 -PLAVLQDLAERLSSDQRLARQALFIGLLMD7-SGDFLIRGLLGADP Haap 212 -ALDVLKEEVGELSTDIKVNLFALKYKTSHKETHSVPLSFAE Haap 232 -ASNYLQQVVSTFSDMNIILAGGQVDNLSSLTSE-10-IDASGVVGLSFSGHRI s.th.htphlth.slhhh.tthhps.thp. EEEEEPEEEEEHHHHHHHHHHHHHHHHHH		PGAHSVSVVSOGCRPI-GEPYIVGGADGAVITELGGRP 220 81255484
Syne SGNVVLDAVAQGCRPI-GPDLKWEAEGNVILSLEGRP 267 8666541 Mmag GSGIEVLTGMTQGCSPL-GEVHTWESWQGVVMALDGRP 212 23015999 Aria SDKLVIENDAFVDSIDI-GKKMEWMAQDNILQQIDHLP 214 87118909 Atha R-8-ERIQFDTAISNGMSSVDLRYKAANVNVSLGPSC 319 26451740 80% th.hhh.tthpsh.shhlttsptphlhplstp. B th.hhh.tthpsh.shhlttsptphlhplstp. B th.hhh.tthpsh.shhlttsptpllkplstp. Cbi 218 .KEFMRDAFKDGLGIGIVD-7-GODFLIRGLIGADP Cbi 218 -VKEFYNKYLGTNNNDQILRMGNKFPLMVKRNNRYFSTHILKFINA Syne 268 -LADVLREVGELLARDLRRIAGGUJGIGIVD-7-GODFLIRVLIGIDP Mmag 212 -ADVLKVEGELLARDLRRIAGGUYHVGLPAGED-SHDVQVRTLGIDP Marg 215 -AQQVNRYLGNGSTDIKVNNLPALKYKTSHKETHSVPLSFAE Hsap 232 -ASNYLQQVVSTFSDMNIILAGGQVDNLSSLTSE-10-IDASGVVGLSFSGHRI so% .sht.htplt hslhhh.tt.hhp.ttp EEEEEEEEEE		KGEFLNVNNRSSFNCIPI-GKEHEINLVEDNIVKKIGTIS 217 82748819
Mmag GSGIEVLTCMTQGCSPL-GEVHTVDESWQGVVMALDGRP 212 22015999 Mari SDKLVIENDAFVDSIDI-GKKMRVQAMQDNILQQIDHLP 214 87118909 Atha R-8-EERIGPTDTAISNGMSSVDLRYKAANVNVSLGFSC 319 26451740 80% th.h.thtpsh.s.hhltsp.tphlhplstp. B th.h.hhttpsh.s.hhltsp.tphlhplstp. B th.h.hhttpsh.s.hhltsp.tphlhplstp. B th.N.Nhttpsh.s.hhltsp.tphlhplstp. Mub 221 PLHRLREIVLGMAPDEQELVSRGLQIGIVVD-7-CGDFLIRGLGADP Ch218 -VKEFYNKJUSTNNNDQILRMGNKFPLMVKKNNRYFSTHILKFINAP Syne 268 PLAVLQDLAERLSPEDQELVSRGLQIGIVVD-7-GODFLIRGLIGADP Mmag 212 -ALDVLKEEVGELSFDDKLARQALFIGLMD7-SGDFLIRVILGIDP Mmag 212 -ALDVLKEEVGELSFDDKLARQALFIGLMD7-SGDFLIRVILGIDP Mari 215 -AQQVFNRYLONGSTDIKVNNLFALKYKTSHKETHSVPLSFAE Haap 232 -ASNYLQVVSTFSDMNIILAGGVVDNLSALTSE-10-IDASGVVGLSFSGHRI S0% .sht.htphlt. .sht.htps.thp. EEEEE .EEEEE eEEEEE Hmar TEGPLDFAVTVSEGTEVVVTHSNKSQVVAVRNAAANNAVNELRGGSVAGGYVTA Mcb1 TGGIGVVGATVGQTVQFURDAAADKDLRAVRAAANAELPGPBVGGLFTGN	Svne	SGNVVLDAVVAOGCRPI-GDPLRVTEAEGNVILSLEGRP 267 86606541
Mari SDKLVIENDAFVDSIDI-GKKMKVLAQDNILQQIDHLP 214 87118909 Atha R-8-ERIQFDTAISNGMSSVDLRYKAANVNVSLGPSC 319 26451740 80%th.hh.tthpsh.shhlTtsptphlhplstp. BEEEEEEEEE Hmar 235 -AFEAMRDAIRED-14-GSEDLVMLLGRYELGIESE-22-TTGYNIRWPGHTTD Mtub 221 -PLHRLREIVLGMAPDEQELVSRGLQIGIVUD-7-QGDFLIRGLLGADP Cbei 218 -VKEFYNKYLGTNNDDQILRMGNKPPLMVKRNNKFSTHILKFINA- Syne 268 -PLAVLQUAERLSPSDQRLARQALFIGLLMD-7-SGDFLIRVLLGADP Cbei 218 -VKEFYNKYLGTNSPSDQRLARQALFIGLLMD-7-SGDFLIRVLIGIDP Mmag 212 -ALDVLKEEVGELLARDLRRIAGYHVGLPAEGDD-SHDYQVRTLIGIDP Mmag 212 -ALDVLKEEVGELSTDIKVMLPALYKKTSHKETHSVPLSFA- Hsap 232 -ASNYLQQVVSTFSDMNIILAGGQVDNLSSLTSE-10-IDASGVVGLSFSGHRI 80% .sht.htphlthslhhh.tt.hhpsthp EEEEEEEEEEHHHHHHHHHHHHHHHHH	Mmaq	GSGIEVLTGMTOGCSPL-GEVHTVTESWOGVVMALDGRP 212 23015999
Atha R-8-ERIQFDTAISNGMSSVDLRTKAANVNVSLGPSC 319 26451740 80%th.hhh.tthpsh.shhlTtsptphlhplstp. Bth.hhh.tthpsh.shhlTtsptphlhplstp. BEEEEEEE Hmar 235 -AFEAWRDAIRED-14-GSEDLVMLLGRYELGIESE-22-TTGYNIRWPGHTTD Cbei 218 -VKEFYNKYLGTNDPDCgLISRGLQGIGIVD-7-GSDFLIRGLLGADP Cbei 218 -VKEFYNKYLGTNNNDQILRMGNKPPLMVKRNNRYFSTHILKFINA Syne 268 -PLAVLQDLAERLSPSDQLLARQALFIGLIMD7-SGDFLIRVILGIDP Mari 215 -AQQVPNRYLGNGSSTDIKVMNLPALKYKTSHKETHSVPLSFAE Haap 232 -ASNYLQQVVSTFSDMNIILAGGQVDNLSSLTSE-10-IDASGVVGLSFSGHRI 80%thtphlthslhh.tt.hhps.thp EEEEEEEEEEHHHHHHHHHHHHHHHHH	Mari	SDKLVIENDAFVDSIDI-GKKMRVTAMODNILOOIDHLP 214 87118909
<pre>80%th.h.wh.tthpsh.s.whlTtsptphlhplstp. Bth.h.wh.tthpsh.s.whlTtsptphlhplstp. BthHHHHHHHHtEEEtEEEEEEE Hmar 235 -AFEAWRDAIRED-14-GSEDLVMLLGRYELGIESE-22-TTGYNIRWPGHTTD Mtub 221 -PLHELREIVLGMAPDEQELVSRGLQIGIVUD-7-QODFLIRGLLGADP Cbei 218 -VKEFYNKVLGTNNNDQILRMCNKFPLMVKKNNRYFSTHILKFINA Syne 268 -PLAVLQDLAERLSPSDQRLARQALFIGLLMD-7-SGDFLIRVILGIDP Mmag 212 -ALDVLKEEVGELLARDLRRIGYHVGLPAEDD-SHDYQVRTLIGIDP Mmag 212 -ALDVLKEEVGELSTDIKVNLFALKYKTSHKENSVPLSFAE Hsap 232 -ASNYLQQVVSTFSDMNIILAGGQVDNLSSLTSE-10-IDASGVGLSFSGHRI 80% .sht.htphlth.slhhh.tt.hhps.thpEEEEEPEEEEHHHHHHHHHHHHHHHHH</pre>	Atha	R-8-ERIQFDTAISNGMSSVDLRYKAANVNVSLGPSC 319 26451740
 HHHHHHHHH EEEE EEEEE HHHHHHHHH EEEE EEEEE Mub 221 -PLHRLRED14GSEDLVMLLGRYELGIESE-22-TTGYNRWPGHTTD- Mub 221 -PLHRLRETVLGMAPDEQELVSRGLQIGIVVD7-QGDFLIRQLLGADP Chei 218 -VKEFYNRYLGTNNNDQILMMGNKPFLMVKRNNRYFSTHILKFINA Syne 268 -PLAVLQDLAERLSPSDQRLARQALFIGLLMD7-SGDFLIRVILGIDP Marg 212 -ALDVLKEEVGELLARDLRRIAGYIHVGLPAEGDD-SHDYQVRTLIGIDP Marg 212 -ALDVLKEEVGELSTDIKVNNLPALKYKTSHKETHSVPLSFSGHRI 80% .sht.htphth.slhhh.tt.hhpsthp EEEEEEEEEE HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	80%	th.hhh.tthpsh.shhlTtsptphlhplstp.
<pre>Hmar 235 -AFEJARDAIRED-14-GSEDLVWLLGRYELGIESE-22-TTGYNIRWGGHTTD Mtub 221 -PLHRLREIVLGMAPDEQELVSRGLQIGIVVD-7-QGDFLIRGLLGADP Chei 218 -VKEFYNKYLGTMNNPQILRMGNKFPLMVKRNNRYFSTHILKFINAD- Syne 268 -PLAVLQDLAERLSFSDQKLARQALFIGLMD-7-SGDFLIRVILGIDP Mmag 212 -ALDVLKEEVGELSFSDQKLARQALFIGLMD-7-SGDFLIRVILGIDP- Mari 215 -AQQVFNKYLGNGSSTDIKVNNLFALKYKTSHKETHSVPLSFAE Haap 232 -ASNYLQQVVSTFSDMIILAGGVVDNLSSLTSE-10-IDASGVVGLSFSGHRI 80% .sht.htphlth.slhhh.tt.hhps.thp. EEEEEEEEEEHHHHHHHHHHHHHHHHHHH</pre>	в	
<pre>Mtub 221 -PLHLREIVLGMAPDEQELVSRGLQIGIVVD7-QGDFLIRGLLGADP Cbei 218 -VKEFYNKYLGTNNNDQILRMGNKFPLMVKRNNRYFSTHILKFINA Syn2 268 -PLAVLQDLAERLSPSDQLAAGALFIGLLMD7-SGDFLIRVILGIDP Maari 212 -ALDVLKEEVGELLARDLRRIAGYLHVGLPAEGDD-SHDYQVRTLIGIDP Mari 215 -AQQVFNRYLGNGSTDIKVMNLPALKYKTSHKETHSVPLSFAE Hsap 232 -ASNYLQVVSTFSDMNIILAGGVUDNLSSLTSE-10-IDASGVVGLSFSGHRI 80% .sht.htphlth.slhhh.tt.hhps.thp. EEEEEEEEEEHHHHHHHHHHHHHHHHHHH</pre>		
Cbei 218 -VKEFYNKYLGTNNNDQILEMGNKPFLWVKENNRYFSTHILKFINA Syne 268 -PLAVLQDLAERLSPSDQRLARQALFIGLLMD7-SGDFLIRVILGIDP Mari 215 -AQQYPNRYLGNGSSTDIKVMNLFALKYKTSHKETHSVPLSFAE Hsap 232 -ASNYLQQVVSTFSDMNIILAGGYLMULFALKYKTSHKETHSVPLSFAE- sht.htphlthslhh.tt.hhps.thps.thp. EEEEEEEEEEHHHHHHHHHHHHHHHHHHH		
<pre>Syne 268 -PLAVLQDLAERLSPSDQRLARQALFIGLLMD7-SGDFLIRVILGIDP Mmag 212 -ALDVLKEEVGELLARDLRRIAGYIHVGLPAEGDD-SHDYQVRTLIGIDP Mari 215 -AQQVFNRYLGNGSSTDIKVMNLFALKYKTSHKETHSVPLSFAE Hsap 232 -ASNYLQQVVSTFSDMNIILAGGVVDNLSSLTSE-10-IDASGVVGLSFSGHRI 80% .sht.htphlth.slhh.tt.hhps.thp EEEEEEEEEEHHHHHHHHHHHHHHHHH</pre>		
Mmag 212 - ALDULKEEVGELLARDLERIAGYIHVGLPAEGDD-SHDYQVRTLIGIPD Mari 215 - AQQVFNRYLGNGSTDIKVMNLFALKYKTSHKETHSVPLSFAE- Hsap 232 - ASNYLQQVVSTFSDMNIILAGGVDNLSSLTSE-10-IDASGVVGLSFSGHRI 80% .sht.htphlthslhhh.tt.hhps.thp. EEEEEEEEEEHHHHHHHHHHHHHHHHH	Sime	269 - DLAVLODLAFPLSPSDORLAPOALFICLIMD7-SCDFLTPVILCIDD
<pre>Mari 215 -AQQVPNRYLGNGSSTDIKVMNLFALKYKTSHKETHSVPLSFAE Hsap 232 -ASNYLQQVVSTFSDMNIILAGGQVDNLSSLTSE-10-IDASGVVGLSFGHRI 80% .sht.htphlthslhh.tt.hhpsthpsthps EEEEEEEEEEHHHHHHHHHHHHHHHHH</pre>	Mmag	212 -ALDVLKEEVGELLARDLRRIAGVIHVGLPARGDD-SHDVOVRTLIGIDP
<pre>Haap 232 -ASNYLQQVVSTFSDMNIILAGGQVDNLSSLTSE-10-IDASGVVGLSFSGHRI 80% .sht.htphlth.slhhh.tt.hhps.thp. EEEEEEEEEEHHHHHHHHHHHHHHHH</pre>		
<pre>80% .sht.htphlthslhhh.tt.hhps.thp. EEEEEEEEEEHHHHHHHHHHHHHHHH</pre>	Hsap	232 -ASNYLQQVVSTFSDMNIILAGGQVDNLSSLTSE-10-IDASGVVGLSFSGHRI
Hmar TEGPLDFAVTVSEGTEVVVTHSNKSDQVYAVENAANNAVNELRGGSVAGGFVVTGA Mtub TTGAIGGEVVEVGATVQFQVRDAAAADKDLRLAVERAAABLPGPPVGGLLFTGM Mtub TGAIGGEVVEVGATVQFQVRDAAAADKDLRLAVERAAABLPGPPVGGLLFTGM Syne RVGAIAIGDRVRPQTVQFHLRDAQTSAEDLRWALSRYCAERNLQ-14-PCGALMFGL Mag GQUMIAIGEHVEGGRLIFVRDANAAADLRLANELRAKELLDGRDIRAGLTVSGT Mari NGGIIMSEPL-PVGAQIKFVYFHBWPSLQGSLEKLNLLYQHPPESIFFFNGI Hsap QSATVLLMEDVSDEKTAEAAMQRLKAANIPEHNTIGFMFAGV 80% .ptsh.hhstl.u.l.hh.pth.hht. HHHHHHHHHHEEEE GGRMRGVTDHDASTIEDLGG-IPLAGFFAGEIGP 364 81255484 SRNIFKKVNAKBIIFFKDLSVGCFYTFGEVCSAD 409 55377251 Mtub GRGRRMFGVTDHDASTIEDLGG-IPLAGFFAGEIGP 364 81255484 Spne GSKGLVGTPNFDSQFRELLGE-LDGGFFCNGEIGP 351 23015999 Mari SRQDPDILNENSELKLELDGFFCNGEIGP 351 23015999 Mari SRQDPDILNENSKLKLLSQVQEINGTYCYEFFEDTSD 344 87118909	80%	
Mtub TTGAIGIGEVVEVGATVQFQVRDAAAADKDLRLAVERAAAELPGPPVGGLLFTGM -EEALISTKL-DIGEKIRIGFGDLRBILQSAKEMYSEITRCPCESLLIYSD Syne RVGAIAIGDRVFQGTVQFHLRDAQTSAEDLRWALSTVACBRNLQ-14-PCGALMFSGL Mmag GQGWIAIGEHVEEGGRLIFVRRDAQTSAEDLRWALIGLKERLDGRPIRAGLYVST NGGIIMSEPL-PVGAQIKFYYFHWFSLQGSLEKLNLLYQHPPESIFTRGI Hmap QSATVLLNEDVSEKTAEAAMQRLKAANIFEHNTIGFMFAGV 80% .ptsh.hhstl.u.l.hh.pth.hht. 	Hmar	
Cbei -EEALISTKL-DIGEKIRIGFGDLREILQSAKEMYSEITRCPCESLIYSD Syne RVGAIAIGDRVRPGQTVQFHLRDAQTSAEDLRWALSRYCAERNLQ-14-PCGALMFSL Mag GQWIAIGEHVEEGGRLIFVRDANARADLRRMLIGLKERLDGRPIRAGLYVST Mari NGGIIMSEPL-PVGAQIKFVYFHPWPSLQGSLEKLNLLYQHPPESIFTFFVI UGGIIMSEPL-PVGAQIKFVYFHPWPSLQGSLEKLNLLYQHPPETIGFMFAV 80% .ptsh.hhstl.u.i.hhpth.hh.t	Mtub	TTGAIGIGEVVEVGATVQFQVRDAAAADKDLRLAVERAAAELPGPPVGGLLFTCN
Syne RVGAIAIGDRVRPGGTVQFHLRDAQTSAEDLRWALSRYCAERNLQ-14-PCGLNPSL Mmag GQGWIAIGEHVEEGGRLIFVRRDANAARADLRRMLIGLKERLDGRPIRAGLYVSOT Mari NGGIIMSEPL-PVGAQIKFVYFHPWSLGSLEKLNLLYQHPPESIFTFNG Hsap QSATVLLNEDVOSQIKFVYFHPWSLGSLEKLNLLYQHPPESIFTFNG 80% .ptsh.hhstl.u.l.hh.pth.hht. Hmar CRAMILGEDFDDAVDTIHSAIDA-PFAGFETYGEVCSAD 409 55377251 Mtub GCRRMFGVFDDAVDTIHSAIDA-PFAGFETYGEVCSAD 409 55377251 Mtub GCRRMFGVFDDAVDTIHSAIDA-PFAGFETYGEVCSAD 409 55377251 Syne GRCKRLYGTNAKEIIFPKDLSVGCFYTFGEVCSAD 409 55377251 Syne GRCKGLYGTNAKEIIFPKDLSVGCFYTFGEVCSAD 409 55377251 Syne GRCKGLYGTNAKEIIFPKDLSVGCFYTFGEVCSAD 409 55377251 Syne GRCKGLYGTNAKEIIFPKDLSS151 2015999 Mmag GRCKGLYGTNAKEILSPLUG-FPLIGFFANGEIS351 2015999 Mari SRDQDPILNENSKLKLLSQVQEINGTYCYGEFFSD 344 87118909 Magi SRDQDPILNENSKLKLLSQVQEVHEGFFGNEEIGCDR 367 62898732	Chei	-EEALISTKL-DIGEKIRIGFGDLREILOSAKEMYSEITRCPCESLLIYSC
Mmag GQ@WIAIGEHVEEGGRLIFVRRDANAARADLRRMLIGLKERLDGRPIRACLYGT Mari MGGIIMSEPL-PVGAQIKFVYFHWPSLQGSLEKLNLLVYHPPESIFTFNG Hsap QSATVLLNEDVSDEKTAEAAMQRLKAANIPEHNTIGFMFAGV 80% .ptsh.hhstl.u.l.hh.pth.hht Hmar CMAMILGEDFDDAVDTIHSAIDAPFAGFETYEEVCSAD 409 55377251 Mtub GZRRMFGVFDDDAVDTIHSAIDAPFAGFETYEEVCSAD 409 55377251 GZRRMFGVFDDDAVDTIHSAIDAPFAGFETYEEVCSAD 409 55377251 Mtub GZRRMFGVFDDASTIEDLLGG-IPLAGFFAAGSIGP 364 8125484 Cbei SRNIFKKVDNASTIEDLLGG-IPLAGFFAAGSIGP 364 8125484 Cbei SRNIFKKVPNFDSQFRELLGE-LPLAGFFAAGSIGP 36606541 Mmag GZGYMFGHFDAFELLEE-LPLGFFCMEEIS 351 2015999 Mari SRQQDPILNFDASTELLEVLOF-FPLGFFANGEIS 351 2015999 Mari SRQDPILNFDASPELLEVLOF-FPLGFFANGEIS 351 2015999 Mari SRQDPILNFDASPELLEVLOF-FPLGFFANGEIS 351 2015999 Mari SRQDPILNFDASPELLEVLOF-FPLGFFANGEIS 351 2015999 Mari SRQDPILNFDASPELLEVLOF-FPLGFFANGEIS 354 3718909 <td>Syne</td> <td>${\tt RVGAIAIGDRVRPGQTVQFHLRDAQTSAEDLRWALSRYCAERNLQ-14-PCGALMFSCL}$</td>	Syne	${\tt RVGAIAIGDRVRPGQTVQFHLRDAQTSAEDLRWALSRYCAERNLQ-14-PCGALMFSCL}$
Heap QSATVLLMEDVSDEKTAEAAMQRLKAANIPEHNTIGFMFAGV 80% .ptsh.hhstl.u.l.hh.pth.hht Hmar CRAMILGEDFDDAVDTIHSAIDAPFACFETYGEVCSAD 409 55377251 Mtub GRCRMFGVFDDAVDTIHSAIDAPFACFETYGEVCSAD 409 55377251 Htub GRCRMFGVDDASTIEDLLGG-IPIAGFFAAGEIGP 364 81255484 Cbsi SRINIFKVNNKEIIFPKDLSVGCFTTFGEFNNU 347 82748819 Syne GRCKGLYGTPNFDSQRFRELLGE-LPLGGFFCNGEIGP 425 86606541 Mmag GRCEYMFGHKDAEPELLHEVLGD-FPLIGFFANGEIS 351 23015999 Mari SRDQDPILNENSKLKLLSQVQEINGTYCTGEFFSD 344 87118909 MagG GRCFYYRAKONVEADAFRFFFSVPLFGFFSMEEIGCDR 367 62898732	Mmag	GOGWIAIGEHVEEGGRLIFVRRDANAARADLRRMLIGLKERLDGRPIRAGLYVSOT
Heap QSATVLLMEDVSDEKTAEAAMQRLKAANIPEHNTIGFMFAGV 80% .ptsh.hhstl.u.l.hh.pth.hht Hmar CRAMILGEDFDDAVDTIHSAIDAPFACFETYGEVCSAD 409 55377251 Mtub GRCRMFGVFDDAVDTIHSAIDAPFACFETYGEVCSAD 409 55377251 Htub GRCRMFGVDDASTIEDLLGG-IPIAGFFAAGEIGP 364 81255484 Cbsi SRINIFKVNNKEIIFPKDLSVGCFTTFGEFNNU 347 82748819 Syne GRCKGLYGTPNFDSQRFRELLGE-LPLGGFFCNGEIGP 425 86606541 Mmag GRCEYMFGHKDAEPELLHEVLGD-FPLIGFFANGEIS 351 23015999 Mari SRDQDPILNENSKLKLLSQVQEINGTYCTGEFFSD 344 87118909 MagG GRCFYYRAKONVEADAFRFFFSVPLFGFFSMEEIGCDR 367 62898732	Mari	NGGIIMSEPL-PVGAQIKFVYFHPWPSLQGSLEKLNLLYQHPPESIFTFNCI
Hmar CRAMILGEDPDDAVDTIHSAIDAPERGERTYEEVCSAD 409 55377251 Mtub GRCRMFGVTDHDASTIEDLLGG-IPLAGFFAAGEIGP 364 8125548 Cbei SRINIFKVNAKEIIPFKDLSVCGFYTFEGENNVD 347 82748819 Syne GRCKGLYGTPNFDSQRFRELLGE-LPLGGFFCNGEIGP 425 86606541 Mmag GRCEYMFGHKDAEPELLHEVLGD-PFLIGFFANGEIS 351 23015999 Mari SRDQDPILNENSKLKLLSQVQEINGTYCYGEFFSD 344 87118909 Hasq GGCFYYFAKONVEADAFRFFFSVFLGFFGNGEIGCDR 367 62898732	Hsap	QSATVLLNEDTIGFMF#CV
Hmar CRAMILGEDPDDAVDTIHSAIDAPERGERTYEVCSAD 409 55377251 Mtub GRCRMFGVTDHDASTIEDLLGG-IPLAGFFAREIGN- 364 81255484 Cbei SRINTFKVNNKEIIFPKDLSVCGFYTFGEFNNVD 347 82748819 Syne GRCKGLYGTPNFDSQRFRELLGE-LPLGOFFCNOEIGP 425 86606541 Mmag GRCEYMFGHKDAEPELLHEVLGD-FPLIGFFANGEIS 351 23015999 Mari SRDQDPILNENSKLKLLSQVQELINGTYCYGEFFSD 344 87118909 Hasp GGCFYYFAKONVEADAFRFFFSVPLEGFFGNGEIGCDR 367 62898732	80%	.ptsh.hhstlul.hhpthhh.thshhhsCh
Hmar CRAMILGEDPDDAVDTIHSAIDAPERGERTYEVCSAD 409 55377251 Mtub GRCRMFGVTDHDASTIEDLLGG-IPLAGFFAREIGN- 364 81255484 Cbei SRINTFKVNNKEIIFPKDLSVCGFYTFGEFNNVD 347 82748819 Syne GRCKGLYGTPNFDSQRFRELLGE-LPLGOFFCNOEIGP 425 86606541 Mmag GRCEYMFGHKDAEPELLHEVLGD-FPLIGFFANGEIS 351 23015999 Mari SRDQDPILNENSKLKLLSQVQELINGTYCYGEFFSD 344 87118909 Hasp GGCFYYFAKONVEADAFRFFFSVPLEGFFGNGEIGCDR 367 62898732		
Mtub GRCRRMFGVTDHDASTIEDLIGG-IPLACFRAGEIGP 364 81255484 Cbei SRKNIFKKVNAKEIIPFKDLSVGGFYTFGEFNNVD 347 82748819 Syne GCKGLYGTPNFDSQFRELLGE-LPLGGFFCNGEIGP 425 86605541 Mmag GCEYMFGHFNPESQFRELLGE-LPLGGFFCNGEIGP 351 23015999 Mari SRDQDPILNENSKLKLLSQVQEINGTYCYGEFFSD 344 87118909 Hasp GCFQYYFAKCNVEADAFRFFFSVPLFGFFGNGEIGCDR 347 62898732	Hmar	CRAMILGEDFDDAVDTIHSAIDAPFACFETYCEVCSAD 409 55377251
Cbei SRKNIFKKVNAKEIIPFKDLSVGGFYTFGEFNNVD 347 82748819 Syne GRCKGLYGTPNFDSQRFRELLGE-LPLGGFFCNGEIGP 425 8660541 Mmag GRCYMFGHKDAEPELLHEVLGD-FPLIGFFANGEIS 351 23015999 Mari SROUPPLINENSKLKLSQVQELNGTYCYGEFFTSD 344 87118909 Heap GRGYYRAKGNVEADAFRKFFSSVPLFGFFGNGEIGCDR 367 62898732	Mtub	
Syne GRGKGLYGTPNFDSQRFRELLGE-LPLGGFFCNGEIGP 425 86606541 Mmag GRGEYMFGHKDAEPELLHEVLGD-FPLIGFFANGEIS 351 250599 Mari SROOPDILNENSKLKLLSQVQELNGTYCYGEFFISD 344 87118909 Maap GRGFQYYRAKGNVEADAFRKFFFSVPLFGFFANGEIGCDR 367 62898732		SRKNIFKKVNAKEIIPFKDLSVGCFYTFGEFNNVD 347 82748819
Mmag GCEYMFGHKDARPELLHEVLGD-FPLIGFFANGEIS 351 23015999 Mari SRDQDPILNENSKLKLLSQVQEINGTYCYGEFFTSD 344 87118909 Hsap GCFQYYRAKGNVEADAFRKFFFSVPLFGFFGNGEIGCDR 367 62898732		GRGKGLYGTPNFDSORFRELLGE-LPLGCFFCNGBIGP 425 86606541
Mari SRDQDPILNENSKLKLLSQVQEINCTYCYGEFFTSD 344 87118909 Hsap GRGFQYYRAKGNVEADAFRKFFPSVPLF <mark>C</mark> FFGN <mark>GB</mark> IGCDR 367 62898732		
Hsap GRGFQYYRAKGNVEADAFRKFFPSVPLFCFGNCEIGCDR 367 62898732	Mari	
	80%	hRhht.p.t.ht.hhhhGh.s.GE.h

Fig. 1.

Representative multiple alignments of the FIST_N (**A**) and FIST_C (**B**) subdomains. Representatives of Archaea, Bacteria and Eucarya are included. Predicted secondary structure is shown above the alignment and consensus (80%) calculated for all members of the subdomain family is shown below. GI identification numbers are shown at the end of each sequence. Complete alignments are provided as Supplementary Material. Highly conserved residues are highlighted. Species abbreviations: Atha, *Arabidopsis thaliana*; Cbei, *Clostridium beijerincki*; Hmar, *Haloarcula marismortui*; Hsap, *Homo sapiens*; Mari,

Marinomonas species; Mmag, *Magnetospirillum magnetotacticum*; Mtub, *Mycobacterium tuberculosis* and Syne, *Synechococcus* species.

Author Manuscript

FIST_N FIST_C Methanopyrus kandleri 19887261
FIST_N FIST_C MCP signal Bradyrhizobium japonicum 27352453
FIST_N FIST_C HisKA HATPase_c Wolinela succinogenes 34483007
FIST_N FIST_C GGDEF Vibrio cholerae 9654530
FIST_N FIST_C TrmB Methanocaldococcus jannaschii 1500521
F-box FIST_N FIST_C Arabidopsis thaliana 79546803
F-box FIST_C Homo sapiens 62898732
FIST_N <i>Campylobacter jejuni</i> 86149651
FIST_C MCP signal Campylobacter jejuni 86149693

EAL

Fig. 2.

omain architecture of representative proteins containing the FIST domain. Species names and GenBank identification numbers are shown. Domain designations (Pfam accession numbers in parentheses): MCPsignal (PF00015), MCP C-terminal signaling domain; GGDEF (PF00990), di-guanylate cyclase; EAL (PF00563), di-guanylate esterase; HisKA (PF00512), histidine kinase dimerization domain; HATPase_c (PF02518), histidine kinase domain ATP-binding domain; TrmB (PF01978), helix-turn-helix DNA-binding domain; F-box (PF00646), protein-protein interaction domain and PAS (PF00989), the PAS sensory domain.