# Research article

# **Open Access**

# Application of comparative genomics in the identification and analysis of novel families of membrane-associated receptors in bacteria

Vivek Anantharaman and L Aravind\*

Address: National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA Email: Vivek Anantharaman - ananthar@ncbi.nlm.nih.gov; L Aravind\* - aravind@ncbi.nlm.nih.gov

\* Corresponding author

Published: 12 August 2003

BMC Genomics 2003, 4:34

This article is available from: http://www.biomedcentral.com/1471-2164/4/34

Received: 10 April 2003 Accepted: 12 August 2003

© 2003 Anantharaman and Aravind; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** A great diversity of multi-pass membrane receptors, typically with 7 transmembrane (TM) helices, is observed in the eukaryote crown group. So far, they are relatively rare in the prokaryotes, and are restricted to the well-characterized sensory rhodopsins of various phototropic prokaryotes.

**Results:** Utilizing the currently available wealth of prokaryotic genomic sequences, we set up a computational screen to identify putative 7 (TM) and other multi-pass membrane receptors in prokaryotes. As a result of this procedure we were able to recover two widespread families of 7 TM receptors in bacteria that are distantly related to the eukaryotic 7 TM receptors and prokaryotic rhodopsins. Using sequence profile analysis, we were able to establish that the first members of these receptor families contain one of two distinct N-terminal extracellular globular domains, which are predicted to bind ligands such as carbohydrates. In their intracellular portions they contain fusions to a variety of signaling domains, which suggest that they are likely to transduce signals via cyclic AMP, cyclic diguanylate, histidine phosphorylation, dephosphorylation, and through direct interactions with DNA. The second family of bacterial 7 TM receptors possesses an  $\alpha$ -helical extracellular domain, and is predicted to transduce a signal via an intracellular HD hydrolase domain. Based on comparative analysis of gene neighborhoods, this receptor is predicted to function as a regulator of the diacylglycerol-kinase-dependent glycerolipid pathway. Additionally, our procedure also recovered other types of putative prokaryotic multi-pass membrane associated receptor domains. Of these, we characterized two widespread, evolutionarily mobile multi-TM domains that are fused to a variety of C-terminal intracellular signaling domains. One of these typified by the Gram-positive LytS protein is predicted to be a potential sensor of murein derivatives, whereas the other one typified by the Escherichia coli UhpB protein is predicted to function as sensor of conformational changes occurring in associated membrane proteins

**Conclusions:** We present evidence for considerable variety in the types of uncharacterized surface receptors in bacteria, and reconstruct the evolutionary processes that model their diversity. The identification of novel receptor families in prokaryotes is likely to aid in the experimental analysis of signal transduction and environmental responses of several bacteria, including pathogens such as *Leptospira*, *Treponema*, *Corynebacterium*, *Coxiella*, *Bacillus anthracis* and *Cytophaga*.

## Background

Cells have evolved several strategies to recognize and respond to diverse stimuli that constantly bombard their cell surfaces. The most common strategy involves receptors that are embedded in the cell membranes [1,2]. Typically, these receptors comprise of an external sensory surface, a membrane-spanning module, and an intracellular surface that transmits signals to the internal cellular machinery. Numerous receptors, which are constructed on this basic architectural principle, are known from all the three domains of life. Particularly common, in both eukaryotes and prokaryotes, are the receptors that combine an extracellular ligand-binding domain with a single transmembrane segment followed by an intracellular signaling module [1,2]. In bacteria, the most frequently occurring intracellular signaling domain is the histidine kinase domain that ultimately catalyzes phosphotransfer to a receiver domain, as part of a two-component relay system [3-5]. In the more complex crown group eukaryotes, receptors with an intracellular kinase domain that catalyzes the phosphorylation of serine, threonine or tyrosine, are the most common receptors [6,7]. In both eukaryotes and prokaryotes, receptors with intracellular catalytic domains that signal via diverse cyclic nucleotides are also fairly widespread. In contrast, certain classes of receptors are relatively limited in their distribution. For example, the classic bacterial-type chemotaxis and temperature receptors are thus far restricted to prokaryotes [8,9].

Amongst the crown group eukaryotes, such as slime molds, fungi and animals, serpentine or seven-transmembrane receptors (7TMR) are a very widely used class of receptors. Members of this class are characterized by seven membrane-spanning segments, which are arranged approximately in two-layers [10,11]. In some cases such as rhodopsin, a light receptor, they may covalently bind a prosthetic group like retinal in the cavity formed by the helices. Alternatively, they bind to a variety of soluble or surface-anchored ligands such as odorants, neurotransmitters and peptides [11]. In certain cases, such as the animal metabotrobic glutamate receptors, frizzled and latrophilin-like receptors, the 7TMRs possess additional extracellular globular domains that specifically interact with their ligands. The structural scaffold of the 7TMRs apparently possesses a great degree of flexibility that allows them to sense a remarkable diversity of ligands, such as odorants, in animals [12]. As a result, the 7TMRs form some of the largest multigene families in the genomes of vertebrates and nematodes [13]. In animals the 7TMRs predominantly function via heterotrimeric GTPases (G-proteins), which in turn relay a signal to a variety of effectors, such as adenylyl cyclases, phospholipases and ion channels. In the fungi, the 7TMRs additionally activate signaling via Ras-like small GTPases, while in

*Dictyostelium* they may also directly activate MAP kinase cascades and calcium channels though alternative pathways [11]. There is also some evidence for G-protein-independent pathways downstream of 7TMRs in animals and plants [11,14].

Though 7TMRs are currently unknown in eukaryotes other than animals, slime molds, fungi and plants, distantly related proteins, namely the prokaryotic rhodopsins, are encountered in bacteria and archaea [15,16]. The animal and the prokaryotic rhodopsins widely differ from each other in the residues that bind retinal and the actual location of the ligand in the internal pocket. However, structural comparisons between the animal rhodopsins and the prokaryotic proteins reveal that they adopt essentially the same topology and three-dimensional fold [10,17,18]. This suggests that they have most probably descended from a common ancestor despite extensive divergence of their sequence. The prokaryotic rhodopsins perform several different functions: 1) Classical bacteriorhodopsin and halorhodopsin from halophilic archaea and the proteorhodopsins from uncultured marine  $\gamma$ -proteobacteria act as photon-dependent proton or chloride transporters [19]. 2) The sensory rhodopsins from halophilic archaea function as light sensors that transmit a signal in the form of a light-induced conformational change to the transmembrane helices of receptors of the chemotaxis receptor family [16]. 3) The signaling rhodopsins from cyanobacteria, like Anabaena, function as light receptors that transduce a signal via a small intracellular conserved protein that is only found in bacteria [20].

Additionally, relatives of these prokaryotic rhodopsins are also found in several eukaryotes such as chlorophytes, dinoflagellates and fungi. While they appear to be light sensors in these organisms, their exact mode of action is poorly understood [21].

The prevalence of prokaryotic rhodopsins raises the question as to whether other, as-yet-uncharacterized 7TMRs might be deployed in prokaryotic signaling. The availability of prokaryotic genome sequences from across a wide phyletic spread allows one to address this question by using comparative genomics. Comparative genomics has extensively aided the detection of novel domains involved in signal transduction [22-27]. Furthermore, the use of contextual information that emerges from gene neighborhoods or predicted operons in prokaryotes and domain or gene fusions has provided several functional leads regarding the novel signaling domains [28]. Conserved gene neighborhoods or operons are often indicative of the products of those genes interacting physically to form complexes, or their involvement in successive steps of biochemical pathways [29,30]. Likewise, gene fusions also

suggest the close physical interactions between the products of the fused genes. Recurrent fusions of uncharacterized domains with other functionally characterized domains also help in elucidating the functions of the former through the principle of "guilt by association" [31,32]. New genomic information coming from diverse organisms often improve these analyses, because they provide newer contextual connections and allow testing of previously observed connections. The increasing flow of genomic information also helps in the identification of new domains that are absent or infrequent in the proteomes of well-studied organisms.

In this work, we apply the tools of sequence profile analysis and comparative genomics to the wealth of new information from prokaryotic genomes to identify novel membrane-associated receptors. We identify new types of bacterial 7TMRs, and show that they are far more prevalent than previously suspected. They transduce downstream signals via various intracellular pathways and are likely to play an important regulatory role in several pathogenic and free-living bacteria. These bacterial 7TMRs are also associated with novel, extracellular, ligand-binding domains, some of which appear to have undergone lineage specific radiation to recognize diverse ligands. These bacterial receptors may also provide a model for the generalized principles of 7TMR function, and even help in understanding non-G protein linked signaling mechanisms via analogous receptors in eukarvotes. We also identified two other groups of widespread membraneassociated receptors, with five and eight membrane-spanning segments respectively, in diverse bacterial lineages.

### Results and Discussion Identification of novel putative receptors in bacterial proteomes

In order to characterize potential novel domains that may play a role in bacterial signal transduction we collated all available predicted proteomes of prokaryotes from across the entire phyletic spectrum (For details see Methods section below). We laid particular emphasis on including all the recently sequenced proteomes that had not been subjected to sensitive comparative sequence analysis by others or us. Using sensitive PSI-BLAST derived profiles, we collected all the proteins in these proteomes that contained one or more of the commonly occurring domains involved in signal transduction, such as the histidine kinase, chemotaxis receptor, GGDEF, EAL, HD hydrolase, PAS and GAF domains [3,22,33]. In order to identify different kinds of novel signaling receptors, we isolated all proteins in this set which satisfied at least one of the following criteria: 1) They possessed multiple (three or more) membrane-spanning sequents that could be predicted in them using the TOPRED [34], TMPRED [35], TMHMM2.0 [36] and PHDhtm [37] programs. This allowed us to enrich potential multi-TM signaling receptors that are distinct from the common single-pass (1TM) or double-pass (2TM) receptors. 2) They showed large globular extracellular regions that could not be mapped to any other previously characterized domains. This allowed us to identify potential uncharacterized extracellular domains that may function as extracellular sensors.

The regions from signaling proteins fulfilling the abovespecified criteria were then clustered based on gapped-BLAST bit-score densities in the range of 0.8 to 0.4 per position, using the BLASTCLUST program. We specifically concentrated on those regions that formed distinct clusters with multiple representatives from the same or different organisms because they were likely to represent evolutionarily conserved domains with functional relevance in a wide range of organisms. We then used representative versions of each these regions of similarity as seeds in PSI-BLAST searches of the non-redundant protein database (NR database, National Center for Biotechnology Information). Through these searches, we were able to identify all currently available occurrences and characterize the diverse domain contexts in which they occurred. In searches involving membrane-spanning regions, we took care to avoid the inclusion of false positives arising due to their bias towards hydrophobicity. To achieve this, all searches were conducted using the correction for PSI-BLAST-statistics based on sequence composition [38] and the e-value threshold for inclusion in the profile was set at .001. We also ensured that all the detected TM domains were approximately the same size and adopted the same topology in predictions with the above-mentioned algorithms for TM prediction. Finally, we used reciprocal searches to determine whether a consistent set of proteins were recovered from different starting points with significant e-values (e < .001), and examined the sequence alignments for characteristic patterns that could distinguish them from other membrane proteins.

We describe below the novel classes of bacterial membrane receptors that were identified as a result of this analysis and the potential gleanings regarding their functions.

### Characterization of a bacterial family of seven transmembrane receptors with diverse intracellular signaling modules

The proteins PA4856 from *Pseudomonas aeruginosa* and TP0040 from *Treponema pallidum* emerged as representatives of a large cluster of proteins identified in our receptor-search procedure. These proteins shared a homologous transmembrane domain with 7 predicted membrane-spanning helices (Figure 1) fused to histidine kinase catalytic domains and receiver domains in the case of PA4856, and a chemotaxis receptor domain in the case of TP0040 (Figure 2). An examination of their predicted



### Figure I

**Multiple sequence alignment of the 7TM domains of the 7TMR-DISM family.** Multiple sequence alignment the 7TMR-DISM family was constructed using T-Coffee [73] after parsing high-scoring pairs from PSI-BLAST search results. The PHD-secondary structure [78] is shown above the alignment with | representing an  $\alpha$ -helix. The 85% consensus shown below the alignment was derived using the following amino acid classes: hydrophobic (h: ALICVMYFW, yellow shading); small (s: ACDGNPSTV, green) and polar (p: CDEHKNQRST, blue). The limits of the domains are indicated by the residue positions, on each end of the sequence. The two major groups of these receptors, typically associated with either 7TMR-DISMED2s (upper group) or 7TMR-DISMED1s (lower group), are separated by a spacer. The numbers within the alignment are non-conserved inserts that have not been shown. The sequences are denoted by their gene name followed by the species abbreviation and GenBank Identifier (gi). The species abbreviations are as provided in Table 1. This alignment is provided as an additional file in the MS-WORD format (additional file 1).

membrane-spanning topology showed that it was identical to that observed in the eukaryotic 7TM receptors and the prokaryotic rhodopsins, with the N-terminus projecting into the extracellular (or periplasmic) space and the Cterminus into the intracellular space (Figure 2). Furthermore, they were approximately the same size as the prokaryotic rhodopsins and eukaryotic 7TMRs (250–300 residues) and did not deviate in terms of the size distribution of the inter-helix loops from the latter class of proteins. In order to further investigate their affinities and phyletic spread, we initiated PSI-BLAST searches with these proteins. These searches recovered numerous homologous 7TM domains from several proteobacterial lineages, such as *Azotobacter*, Rhizobia, *Pseudomonas*, *Vibrio*, *Coxiella* and *Xylella*, spirochetes like *Treponema* and *Leptospira*, Gram positive bacteria, like Clostridia and *Bacillus halodurans*, and other bacterial lineages, such as *Cytophaga* and *Chlorobium* (Figure 1, Table 1). In particular, several proteins with these 7TM domains were seen in *Cytophaga hutchinsonii* (13 copies) and *Leptospira interrogans* (14 copies) (Table 1). All complete versions of these domains were predicted to possess a characteristic topology with an outward facing N-terminal region and cytoplasmic C-terminus. The seven predicted TM helices corresponded precisely with the seven hydrophobic segments that were strongly conserved in all these proteins. When a profile including all the above-detected bacterial 7TMR domains was used to search the NR database, the eukaryotic 7TMR receptor domains of latrophilin (gi: 4185804), ETL (gi: 4423362) and a 7 TMR domain



**Phylogenetic tree, domain architectures and gene neighborhoods of the 7TMR-DISM family.** Phylogenetic relationships of the 7TMR-DISM domain containing proteins along with the domain architectures are shown. The seed alignment used for constructing the tree was one similar to that shown in Fig. I. The RELL bootstrap values for the major branches are shown at their base. The thickness of a given branch is approximately proportional to the number of proteins contained within it. Domain architectures of the proteins in each branch of the tree are shown in boxes pointed to by the black arrows. The phyletic pattern of each family is shown, along with the number of proteins (if there are more than one). The gene neighborhood data for some of the genes encoding 7TMR-DISM encoding genes is depicted using block arrows. A red arrow indicates the domain architectures of proteins encoded by each gene. The species abbreviations are as shown in Table I. Domain abbreviations are: DISMEDI – 7TMR-DISMEDI; DISMED2 – 7TMR-DISMED2; A. cyclase-Adenylyl cyclases; GGDEF-GGDEF-motif-containing nucleotide cyclase domains; His Kin – Histidine Kinase; EAL-EAL motif containing cyclic nucleotide phosphodieste-rases; REC – Receiver domain; PAS-Ligand binding domain found in *Drosophila* <u>Period</u> clock proteins, vertebrate <u>A</u>ryl hydrocarbon receptor nuclear translocator and *Drosophila* <u>S</u>ingle minded proteins; ZR, Zinc Ribbon HTH; Helix-Turn-Helix domain (of AraC, OmpR and TetR variety); PP2C – Sigma factor PP2C-like phosphatases ; TPR – etratricopeptide repeats; CTR – Chemotaxis receptor domain; HAMP – domain present in <u>H</u>istidine kinases, <u>A</u>denylyl cyclases, <u>M</u>ethyl-accepting proteins and <u>P</u>hosphatases.

encoded by the prawn nidovirus (gi: 9082017) were recovered as the best hits (e-values =  $10^{-2}-3 \times 10^{-3}$ ) outside of the bacterial family.

A sequence alignment of the 7TM domains that were recovered in these searches showed that they shared a characteristic pattern of sequence conservation (Figure 1) including two well-conserved polar residues at the C-termini of the first and the last helix (typically basic

#### Table 1: Phyletic patterns and number of proteins \*

Domain	Firmicutes	Proteobacteria	Actinomycetes	Cyanobacteria	Spirochetes	other
7TMR	Bhal (2); Cace; Ddeh; Dhaf (5)	Alpha- Atum; Bjap; Bmel; Ccre; Mmag (2); Rpal; Rrub (2); Smel. Beta- Rmet (2); Rsol. Gamma- Avin (4); Cbru; Mdeg (2); Paer (4); Pflu (2); Pput (2); Psyr (2); Sone (2); Xcam (2); Xf; Vvul Delta- Ddes. Unclassified- Msp (2)	-	-	Lint (14); Tpal.	Chut (13); Ctep.
7TMR-HD	Bsub; Bant; Cace (2); Cper; Ctet; Cthe; Dhaf; Efae (2); Linn; Lmon; Oihe; Tten.	<b>Delta</b> – Gmet. <u>Unclassified</u> – Msp.	-	Ana; Npun (2); Pmar (2); Syn; Ssp; Tery	Lint; Tpal.	Cpne; Caur; Fnuc (2); Tmar.
5TMR-LYT	Bsub (2); Bant (2); Cace; Ctet; Linn; Lmon; Oihe; Ooen; Saga (2); Sau (2); Sepi; Smut; Tten.	<u>Alpha</u> – Rsph; Rrub. <u>Gamma</u> – Ec (2); Styp; Sone (2); Vcho; Ypes. <u>Delta</u> – Ddes; Gmet.	-	-	-	Drad (2); Fnuc (2)
8TMR-UT	Llac	Alpha- Bjap (2); Ccre; Mlot; Naro; Rsph; Smel (4). <u>Beta-</u> Bfun; Rmet (2); Rsol. <u>Gamma-</u> Ec (8); Mdeg (3); Pmul; Paer; Pput (2); Styp (4); Sone(2); Sfle (2) Vcho; Vpar; Xcam (2); Ypes <u>Delta-</u> Ddes; Mxan <u>Unclassified</u> - Msp.	Cglu; Tfus; Scoe.	Ana; Syn; Ssp (3).	Lint	Caur ESV
7TMR- DISMEDI	Bhal (3); Cace; Ddeh; Dhaf (4)	<u>Alpha</u> – Ccre <u>Beta</u> – Rmet (2); Rsol. <u>Gamma</u> – Mdeg; Xcam (2).	-	-	Lint (9)	-
7TMR- DISMED2	Dhaf	Alpha- Atum; Bjap; Bmel; Mmag (2); Rpal; Rrub (2); Smel. Gamma- Avin (4); Cbru; Mdeg (2); Paer (4); Pput (2); Psyr (2); Sone (2); Xf; Vvul. <u>Delta-</u> Ddes. <u>Unclassified</u> - Msp.	-	-	Lint (4)	Chut (13)

\*Firmicutes: Bant – Bacillus anthracis; Bsub – Bacillus subtilis; Bhal – Bacillus halodurans; Cace – Clostridium acetobutylicum; Cper – Clostridium perfringens; Ctet – Clostridium tetani; Cthe – Clostridium thermocellum: Ddeh – Desulfitobacterium dehalogenans; Dhaf – Desulfitobacterium hafniense; Efae Enterococcus faecium; Llac – Lactococcus lactis; Linn – Listeria innocua; Lmon – Listeria monocytogenes; Oihe – Oceanobacillus iheyensis; Ooen -Oenococcus oeni; Saga - Streptococcus agalactiae; Saur - Staphylococcus aureus; Sepi - Staphylococcus epidermidis; Smut - Streptococcus mutans; Tten -Thermoanaerobacter tengcongensis. Alphaproteobacteria: Atum – Agrobacterium tumefaciens; Bjap – Bradyrhizobium japonicum; Bmel – Brucella melitensis; Ccre – Caulobacter crescentus; Mmag – Magnetospirillum magnetotacticum; MIot – Mesorhizobium loti; Naro – Novosphingobium aromaticivorans; Rsph – Rhodobacter sphaeroides; Rpal – Rhodopseudomonas palustris; Rrub – Rhodospirillum rubrum; Smel – Sinorhizobium meliloti. Betaproteobacteria: Bfun – Burkholderia fungorum; Rmet – Ralstonia metallidurans; Rsol – Ralstonia solanacearum. Gammaproteobacteria: Avin – Azotobacter vinelandii; Cbru – Coxiella brunettii; Ec – Escherichia coli; Mdeg – Microbulbifer degradans; Pmul – Pasteurella multocida; Paer – Pseudomonas aeruginosa; Pflu – Pseudomonas fluorescens; Pput – Pseudomonas putida; Psyr – Pseudomonas syringae; Styp – Salmonella typhimurium; Sone – Shewanella oneidensis; Sfle – Shigella flexneri; Vcho – Vibrio cholerae; Vpar – Vibrio parahaemolyticus; Vvul – Vibrio vulnificus; Xcam – Xanthomonas campestris; Xf – Xylella fastidiosa; Ypes – Yersinia pestis. DeltaProteobacteria: Ddes – Desulfovibrio desulfuricans; Gmet – Geobacter metallireducens; Mxan – Myxococcus xanthus. Unclassified Proteobacteria: Msp – Magnetococcus sp. Cyanobacteria: Ana – Anabaena sp.; Npun – Nostoc punctiforme: Pmar - Prochlorococcus marinus; Syn - Synechococcus sp.; Ssp - Synechocystis sp.; Tery - Trichodesmium erythraeum. Spirochetes: Lint - Leptospira interrogans; Tpal – Treponema pallidum. Actinobacteria: Cglu – Corynebacterium glutamicum; Tfus – Thermobifida fusca; Scoe – Streptomyces coelicolor. Other: Cpne – Chlamydophila pneumoniae; Ctep – Chlorobium tepidum; Caur – Chloroflexus aurantiacus; Chut – Cytophaga hutchinsonii; Drad -Deinococcus radiodurans; ESV - Ectocarpus siliculosus virus; Fnuc - Fusobacterium nucleatum; Tmar - Thermotoga maritima. The number of proteins (if more than one) is given in parenthesis. Incomplete genomes are underlined.

residues). A comparison of these 7TM domains against a library of PSI-BLAST profiles and hidden Markov models (See Methods for details) for previously characterized membrane proteins gave the nematode 7TM receptor family and the prokaryotic rhodopsins as the top-scoring hits (e-values ~.01–.05), suggesting a closer relationship with the classic 7TM receptor families to the exclusion of various other membrane-associated proteins (e-values ~.9–3.5). An examination of the domain architectures of these 7TM proteins reveals considerable diversity around a shared basic architectural blue print. At their N-terminus, these 7TM domains were either directly preceded by a predicted signal peptide, or by different extracellular globular modules, analogous to the domain organization of the

animal metabotropic glutamate receptors. At their C-termini they were typically fused a range of catalytic and non-catalytic signaling domains (Figure 2). The former category includes the quintessential bacterial two-component-system-modules, namely the histidine kinase and receiver domains, cyclic diguanylate signaling enzymes such as the GGDEF-type cyclase and EAL-type phosphodiesterase domains, cNMP generating cyclases and PP2C phosphatases [3,22,39,40]. The non-catalytic domains include the PAS, chemotaxis receptor, TPR, and HAMP domains [8,26,27,41–43]. Interestingly, three DNA binding domains, the AraC-type HTH [44], the OmpR-type HTH [45,46] and the bacterial IS1-like Zn finger domains (VA & LA unpublished), are also fused C-termini of certain 7TM receptors from *Leptospira* and *Cytophaga* (Figure 2). While a small subset of the receptors lack any intracellular domains, they could non-covalently associate with soluble catalytic domains on their intracellular surface. Hereinafter, we refer to these bacterial receptors as 7TMR-DISM (for <u>7TMr</u>eceptors with <u>d</u>iverse intracellular <u>s</u>ignaling <u>m</u>odules).

The great diversity of domain architectures of the 7TMR-DISM family, particularly in terms of their intracellular modules, suggests that they activate number of different intracellular signals in response to different external ligands. With respect to transmitting different cytoplasmic signals via their intracellular regions they resemble the eukaryotic 7TMRs, rather than the prokaryotic bacteriorhodpsins, which are mainly photon-dependent ion pumps. The presence of intracellular DNA-binding domains in certain 7TMR-DISMs suggests that they may take advantage of the non-compartmentalized state of the chromosome in bacteria to directly bind DNA and regulate transcription in response to ligand-induced conformational changes. Additionally, analysis of the gene neighborhoods reveals that in Ralstonia, Pseudomonas and Leptospira the genes encoding 7TMR-DISMs form predicted operons with HTH-transcription factors with receiver domains. These are likely to represent two-component systems in which the transcription factors are modulated by the signal-activated 7TMR-DISM proteins (Figure 2). The presence of TPR repeats in the intracellular regions of certain 7TMR-DISMs is reminiscent of components of eukaryotic signaling systems [42]. These repeats may act as structural scaffolds that link the 7TMR-DISMs to intracellular protein complexes.

## The identification and functional analysis of the novel extracellular ligand-binding domains of 7TMR-DISM proteins

The N-termini of most 7TMR-DISMs are linked to large extracellular regions that are predicted to assume a globular structure. As these regions were also recovered in our procedure for identifying novel extracellular ligand-binding domains of receptors, we investigated them in greater detail. Clustering using BLASTCLUST showed that most of these extracellular domains associated with the 7TMR-DISMs fell in either of two distinct clusters. While some of the extracellular regions did not initially fall into any of the clusters, iterative PSI-BLAST searches with representative seed sequences unified all these extracellular regions with one or the other cluster. This suggested there are two distinct varieties of extracellular domains associated with the bacterial 7TMR-DISMs, which we accordingly refer to as 7TMR-DISMED1 and 7TMR-DISMED2 (for 7TMR-DISM extracellular domains 1 and 2).

Iterative PSI-BLAST searches of the NR database with 7TMR-DISMED1 additionally recovered a globular domain inserted in the middle of the sialate acetylesterase domain from various proteobacteria (e-value = 10-3-10-5 iteration 3), and in subsequent iterations carbohydrate metabolism enzymes (e-value = 10<sup>-2</sup>-10<sup>-4</sup> iteration 4-5) such as  $\beta$ -galactosidases,  $\beta$ -mannosidases and  $\beta$ -glucuronidases [47,48]. For example, a search with the 7TMR-DISMED1 from the protein LA2676 from Leptospira recovered the insert domain of the sialate acetylesterases in iteration 3 (eg. Mdeg0217, Microbulbifer degradans,  $e = 10^{-4}$ ),  $\beta$ -galactosidase in iteration 4 (LacZ, E. coli, e = 10<sup>-5</sup>) and glucuronidases in iteration 6 (eg. GUS, Homo sapiens, e = 10-2). The 7TMR-DISMED1 corresponds precisely to a distinct domain in the galactosidases and the glucuronidases, which is seen to adopt a  $\beta$ -jelly roll topology in their crystal structures [49] (Figure 3A). These domains function as accessory carbohydrate binding domains, rather than catalytic domains of the enzymes in which they occur [49]. An examination of the sequence alignment (Figure 4) shows that the 7TMR-DISMED1s and the  $\beta$ -jelly roll domain of the carbohydrate-metabolism enzymes share several conserved residues, including certain characteristic aromatic positions. The sequence alignment also suggests that the 7TMR-DISMED1s are likely to preserve the spacious cavity of these jellyrolls with a characteristic triangular outline (Figure 3A). The projection of some of the highly conserved residues into this cavity (Figures 3A and 4) suggests that the core structure of the ligand-binding pocket is also likely to be conserved across all these proteins. These observations imply that the 7TMR-DISMs with the 7TMR-DISMED1 are most likely to function as receptors for carbohydrates or related derivatives.

In contrast to the 7TMR-DISMED1s, the 7TMR-DISMED2 domains did not recover any statistically significant hits to sequences with known structures. However, secondary structure prediction based on the multiple sequence alignment predicted that the 7TMR-DISMED2s are likely to adopt an all  $\beta$ -fold with at least 8 extended regions (Figure 5). The average size of the domain and the distribution of the lengths of the extended regions matched that of several carbohydrate-binding domains, such as the discoidin domain, the cellulose binding domains of cellulases and the fucose-binding domain, which share a common jelly roll topology with the 7TMR-DISMED1s [50,51]. Hence, it is plausible that the 7TMR-DISMED2s represent yet another distinct superfamily of the carbohydrate binding jelly roll fold. This would imply that the 7TMR-DISMED2s could also potentially function as sensors for carbohydrate or related ligands.

Previous studies have shown that mapping of residues conserved in specific subgroups of a protein family on the surface view of a representative structure of that domain



Models of 7TMR-DISMED1 and the TM domain of the 7TMR-DISMs. (A) Prototype of the  $\beta$ -jellyroll seen in 7TMR-DISMED1, sialate 9-O-acetylesterases, β-glucoronidases and  $\beta$ -glucosidases. The  $\beta$ -jellyroll domain shown here is a cartoon representation of the domain from the crystal structure of  $\beta$ -galactosidase (PDB:IGHO). Conserved residues typical of the 7TMR-DISMED1 are shown in ball stick representation. "a" stands for a conserved aromatic position. (B) A homology model of the TM domain of the 7TMR-DISMs showing the distribution of conserved residues in the 7TMR-DISMs with 7TMR-DISMED1 domains. The model was constructed using bacteriorhodopsin (PDB: IC3W) and bovine retinal rhodopsin (PDB:1F88) as templates. The N terminus of the 7TMR-DISM domain, where the extracellular domain is attached, is shown in yellow. The red color shows the distribution of residues on the external surface, which are uniquely conserved in 7TMR-DISMs with 7TMR-DISMED Is. This set of proteins essentially corresponds to the lower group of sequences in Fig. 1.

may throw light on regions involved in specific interactions of those subgroups [52,53]. As such analysis could throw more light on the mechanisms of action of 7TMR-DISMs, we constructed a homology model for the 7TM domain of a representative bacterial receptor using the vertebrate visual rhodopsin and the bacteriorhodopsin as templates. We then plotted the residues conserved in the two major subgroups (see below) of 7TMR-DISMs on to the surface view of this model. Several residues that were specifically conserved in the 7TMR-DISMs, which possessed 7TMR-DISMED1s, formed distinctive patches on the rim of the tubular 7TM structure (Figure 3B). These regions could represent regions of contact between the extracellular (or periplasmic) 7TMR-DISMED1 and the outer surface of the 7TM domain. This would imply that the alterations of the contacts between the extracellular domain and the 7TM domain upon ligand-binding, are likely generate the necessary conformational change for propagating an internal signal. The domain-architectural organization of most of the 7TMR-DISMs closely resembles that of the animal glutamate receptors and vertebrate taste receptors [11,54]. Hence, it is possible that these receptors could act through a similar mechanism in which the signal is relayed via an interplay between the extracellular ligand-binding domain and the 7TM domain.

# Evolutionary diversification of the 7TMR-DISMs in bacteria

We analyzed the evolutionary history of the 7TMR-DISMs by constructing phylogenetic trees with the alignments of the conserved 7TMR domains using the neighbor-joining, least square and maximum-likelihood methods (Figure 2). These trees showed that the 7TMR-DISMs were divided into two major clusters that corresponded to forms fused to either 7TMR-DISMED1 or 7TMR-DISMED2 at their Ntermini. A small number of proteins in either group lacked a distinct extracellular globular domain, suggesting secondary loss of these domains. Phyletic pattern of the 7TMR-DISMs is patchy: two relatively closely related bacteria may differ in having or lacking a gene for such a receptor, whereas two distantly related bacteria may possess closely related receptors. The phylogenetic tree shows that forms from distantly related bacteria occasionally group together, with statistically significant support for their grouping (Rell BP>80%, for 10000 boot strap replicates). For example, one such well-supported cluster (Fig. 2) contains proteins from phylogenetically distant bacteria, such as, GC Gram positive bacteria, spirochaetes and  $\beta$ -proteobacteria. This suggests a dynamic history for the 7TMR-DISM genes, which as in the case of many other bacterial signaling proteins, is likely to have involved lateral transfer between distantly related taxa and sporadic gene loss.

However, the most striking pattern observed in the evolutionary tree of the 7TMR-DISMs was the presence of multiple well-supported clusters (Rell BP~70–100%)

Secondary Structure			.нннн.	EEEEE	<u>.</u> .		EEEEEEEEE.HHHHHE
Desu7352_Dhaf_23121831	50 8	SDQVVPL <mark>D</mark> GQ <mark>W</mark> EFYWHQLLEP	GGEPDS-	AANV <mark>IQ</mark> VI	PGF <mark>W</mark> F	RNDMATRIDHAR	GAAT <mark>Y</mark> RLT <mark>I</mark> KLKPSDTV
Desu1817_Dhaf_23113315	54 1	EDGIIELGGE <mark>W</mark> EFYWGQLLSP	DDFKDAT	PPALTGF	VNV <mark>L</mark> E	FNWVGKLDETK <mark>L</mark> TGM-	GSGT <mark>Y</mark> RLL <mark>V</mark> KNMPSEKV
BH1549_Bha1_10174166	55 5	FEQTITLDGE <mark>W</mark> AFYPSTWLID	KNNLNNV	3 NALS <mark>IE</mark> V	PGG <mark>W</mark> I	KHVLSPEKE <mark>V</mark> PY	GYGS <mark>YQLK</mark> IFVDPEDNMT
Desu7142_Dhaf_23121547	17 1	RDKTVKL <mark>S</mark> GE <mark>W</mark> EFYFGQLLTP	EDFKNQE	2 GKTIQN <mark>V</mark>	PAS <mark>W</mark> O	GSYQIEGEH <mark>L</mark> PPQ-	GSAT <mark>Y</mark> RLR <mark>I</mark> LLPAEGGN
hrd22-1_Ddeh_6466833	16 1	DDGIIEL <mark>N</mark> GE <mark>W</mark> EFYWGQLLSP	DDFKHSN	3 MTGY <mark>L</mark> NV	LSD <mark>W</mark> I	IGNLDQTQ <mark>L</mark> AAM-	GSGT <mark>Y</mark> RLL <mark>V</mark> KNVPSEEVL
Desu1754_Dhaf_23113246	58 1	KDRSVFL <mark>N</mark> GE <mark>W</mark> DFYPGKLLQP	EELPKE-	THYF <mark>M</mark> K <mark>V</mark>	PGG <mark>W</mark> -	DKTKGAF <mark>L</mark> QKSR	GNGT <mark>Y</mark> RLS <mark>V</mark> KLPELQEQ
LA0289_Lint_24212989	49 (	QHGPVAL <mark>Q</mark> GD <mark>W</mark> IFRWKEFIKN	PKIDSE-	KNRI <mark>M</mark> P <mark>V</mark> I	PKA <mark>W</mark> I	FRIQEPNGKN <mark>Y</mark> PGT-	GIATYFLKVILPENLS-SNN
LB241_Lint_24202368	75 1	NIESINLKGN <mark>W</mark> EFCWDQWIPP	NAEESQW	3 CNGF <mark>Y</mark> PA	PAY <mark>w</mark> i	KFYNIPGKN <mark>L</mark> SPF-	GKAT <mark>Y</mark> RLK <mark>V</mark> ILPTSFHDS
LA0815_Lint_24213515	46 1	K-NTVQL <mark>D</mark> GN <mark>W</mark> EFYWKELTHG	NFTTPK-	NTSY <mark>F</mark> PV	PGI <mark>W</mark> -	RDYDPN <mark>F</mark> TPE-	GYAT <mark>Y</mark> RLR <mark>V</mark> LCECINKN
LA3610 Lint 24216308	33 1	DFGMTKLNGE <mark>W</mark> EFSWLQGPEK	GIENH	SVEF IGV	PGS <mark>W</mark> I	INESKSNQT <mark>Y</mark> PKF-	GYGI <mark>Y</mark> RLK <mark>V</mark> FLPEIWK-KKI
CAC0818 Cace 15894105	31 1	VYKGNINLKN <mark>Y</mark> DFDKNKIVKLE-G	QLELYNN	12 SSRYLTV	PSEL	КDQ <mark>L</mark> К	GKVDEYMTLHL <mark>M</mark> VQAKDDVV
LA3614_Lint_24216313	34 1	FAFPIDLTKD <mark>W</mark> KIISGKNLNAS-I	KDVSWK-	ELKS <mark>L</mark> PI	PEDSI	ISFSEG <mark>I</mark> Y	TLTLLKT <mark>F</mark> EVS <mark>A</mark> NDFQKLA-LDG
LA2196 Lint 24214896	23 1	EENFQDL <mark>S</mark> GI <mark>W</mark> EVYVSLVSEDL-Q	TNKNLKQ	3 RWEPIPI	PSNL	KDFIK <mark>W</mark> N	SSNEEIL <mark>L</mark> RKQ <mark>F</mark> DFRNVADTS
LA2676 Lint 24215376	95 1	VSLDSTKEHG <mark>W</mark> EITLQKIDPVA-F	SDSYLKG	8 EAYEAPG	VYI <mark>L</mark> -	PEE	SIQTAFI <mark>V</mark> KKF <mark>I</mark> APKNWKGSG
LA0027 Lint 24212727	58 1	SFLKNKK <mark>E</mark> KL <mark>N</mark> IETNSREKESK	DFKNWK-	LITRFPV	NFN <mark>F</mark> I	LFSIPKQSG <mark>F</mark> H	DVTVKVE <mark>F</mark> SIH <mark>S</mark> DSIFLTFLKQP
cyaA18 Lint 24217179	24 1	WKPIDLR <mark>K</mark> GN <mark>W</mark> IAVEGFQREYL-N	GIDSTFS	3 KISHFPV	VLN <mark>E</mark> I	IFETSVETG <mark>L</mark> K	EYTLQTR <mark>F</mark> HIQ <mark>E</mark> DFQKTKV-YKP
Reut4832 Rmet 22980118	46 1	AASAWAQTPT <mark>A</mark> QCSVRILTVST-A	EAGPHGA	5 QWRSVTL	PDDW	NRRQP	GFGGSVW <mark>Y</mark> RID <mark>W</mark> QRECPGTLREP
Reut4777 Rmet 22980062	144 1	VLLAGAASARAQSEAVAFTRIE-A	AQSTWOD	6 GWVPVSL	PDIWS	STR <mark>W</mark> P	GFDGVVWYRLSWQGADPRQP
Mdeg0217 Mdeg 23026370	246	INSDKKAOAOWHNOLTONDKGT-A	EELPWYA	6 NWHSMOV	PGYW	ANHPSFKDES	TINGNVWFRKTLTLPNDAA-NEA
XCC1753 Xcam 21112846	227 1	KORDAAAOAATGKRIGRWPKVD-G	DTPOWRT	6 DWDSIPV	TOOW	ELSGYE	GMDGIAWYRTTFTLTAAEA-KAG
XCC4103 Xcam 21233525	234	ATSPTDAAPRWGKLWETWWRAH-G	DGAPWOP	2 PGAWOPA	PATLO	GAWDEWGVPOLV	GENGMVWYRSTVELTAAOA-AOD
C87348 Ccre 25398225	227	ARDPALGVARWSOSLGRWADAKIP	AAKGWEK	6 AWKPMPT	EGFW	EQAAPDLA	GFDGTVWLRLELTLTKAOA-AOG
lacz Ec 18073591	54 4	SOOLRSLNGEWRFAWFPAPEA	VPESWLE	4 EADTVVV	PSNW	OMHGYDAPIYTNVTYPITV 6	TENPTGCYSLTFNVDESWLOEGO
lacz Ldel 149566	50 /	SSLVOSLDGDWLTDYAENGO	GPUNEVA	6 NEKSVKV	PONT	ELOGEGO POVVNVOVPWDG 10	SKNPLASVVRVEDLDEAFW_DKE
lact. Lpla 28379799	38	SREIKSINGAWRENFAKTPAE	RPVDEVO	6 DEDTTOY	PGHT	ELAGYGOTOYINTLYPWEG 23	ADNTVGSVI.KTEDI.DDVFK-GOR
LAC4_Klac_173305	29 0	DIFEST.NGPWAFAI.FDAPI	DAPDAKN	7 KWSTISV	PSHW	ELOEDWKYCKPIYTNVOVPIPI 6	TUNPTGUVARTELDSKSIESEE
ebga Ec 26110092	97	SSIFI.PI.SCOWNFHEEDHPI.===0	VPEAFTS	4 DWGHTTV	PAMW	OMEGHCKLOYTDECEPET	SDNPTGAYORI FTI.SDGWO_GKO
LACZ Kone 114941	55 /	SSERBOLDOSG-SSLTPAARI.	PSMRVVT	4 DCRGTPV	PSNW	OMEGYDAPIVTNVRVPIDT	EDNPTGCVSI.HETVEDTWRENGO
TM1103 Tmar 15643040	40 1	FFT_SINGNMERTEARNDEF	VDEDEES	6 NWDETEV	DCMW	FMKGYCKBIVTNVVV	DDNDTGUVDDWTFIDEDWF_KKF
log7 Scol 19265742	40 1	WI DOGI NOVINATI UT ARAFT	WI VDEVV	6 DINETNU	CHI	EMRGIGR FIIINVVIFFEF /	ECNAVACYURUEMINDAL V DVV
1802_5581_18265745	42 1	INLEQUENCE AND THE ADVID	DIVEEVV	6 DLNF INV	CHIC	CONTRACTOR CONTRACTOR IN	TENAVASIVANTILNDALA-DAA
AAA23210_CaCe_144740	431	PLRONLINGRARFSISENSSL	RIKEF IK	7 DUMNTET	PGH LY	DEVNY OTVCOVY	EVVD DVUTVIUECU
CUCR_Mmus_114064	25 1	PREI VAL DOL NUEDADI CNNDI OG	ERVDDLR	7 PHINIEL		DERMIQIRSCIRIFLHI	BODTOWDUNU DTNCA
GUSB_Mmus_114964	35 8	SRELKALDGLWHFRADLSNNRLQG	FEQQWIR		PSSFI	NDITQEAALRDFIGWVWIE 9	TODTDMRVVLRINSA
GUSB_HSap_4504223	35 8	SRECKELDGLWSFRADFSDNRRRG	FEEQWIR	/ PTVDMPV	PSSFI	NDISQUWRLRHFVG <mark>W</mark> VWIE	TODLATRVVLRIGSA
u1dA_EC_15804986	9 :	FREIKKLDGL <mark>W</mark> AFSLDRENCG	TEORMME	4 ESRALAV	PU-5 P -	NDOPADADIRNYV(4NVWYO) P	KGWAGQKIVLRFDAV
						by bin binary a	
consensus/85%		ps.a	• • • • • • • •	h.l:	sh.	hh	shh
consensus/85%		ps.a		h.l:	sh.	h	shh
consensus/85% Secondary Structure	EEEI	ps.a	I	h.l:	sh.		shh
consensus/85% Secondary Structure Desu7352_Dhaf_23121831	EEEI YG <mark>L</mark> I	EEEEEE.EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE	PALTQI	EEEEh.l: DEYQ <mark>Y</mark> ENKPYN	sh.  V 3	.EEEEEEEEEEE. VQGDTAETLIHVANY	
consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu1817_Dhaf_23113315	EEEI YG <mark>L</mark> LPI	ps.a EEEEEE.EEEEEEEEEEE RISNIRMASA <mark>IYV</mark> NGNQ <mark>A</mark> GGGGK KKSNIRMSSV <mark>IFV</mark> NGKK <mark>I</mark> IQDGR	PALLEN	EEEE DEYQ <mark>YENKPYN</mark> NEYL <mark>F</mark> SNIPRLO	sh.  V 3 G 1	EEEEEEEEE VQGDTAETLIHVANY FEQRSPETLYVQVSNY	
consensus/85% Secondary Structure Desu7352 Dhaf_23121831 Desu1817_Dhaf_23113315 BH1549_Bhal_10174166	EEEI YGLI LPII YSII	ps.a EEEEEE.EEEEEEE.EEEEE. RISNIRMASA <mark>IYU</mark> NONQ <mark>A</mark> GGSGK KKSNIRMSSVIFVNGKKIIQDGR RVPSVRSASALYVDGRLIAGSGQ	PALTQI PALLEN LGESEN	EEE DEYQ <mark>Y</mark> ENKPYN NEYLFSNIPRL ADYV <mark>A</mark> SNLPYT	sh. v 3 G 1 A 4	EEEEEEEEEE VQGDTAEILHVANY FEQRSPEILVQVSNY NGSGVIEVVIQAANFK	
consensus/85% Secondary Structure Desu7352 Dhaf_23121831 Desu1817_Dhaf_23113315 Bh1549 Bha1_10174166 Desu7142 Dhaf_23121547	EEEI YGLI LPII YSII FGII	ps.a EEEEEE.EEEEEEE.EEEEE. RISHIRMASA <mark>IYY</mark> WAQAAGGSGK KKSNIRMSSVIPYNGKKIIQDGR RVPSVRSASAIYVDGRLIAGSGQ KIITCISASAR <mark>IPA</mark> DDQL <mark>I</mark> LECGS	PALTQI PALLEN LGESEA PGNSPH	EEEE DEYQ <mark>Y</mark> ENKPYN NEYLFSNIPRLA ADYV <mark>A</mark> SNLPYT ETTVHKYYADT	sh. V 3 G 1 A 4 G 3	LEEEEEEEEEE VQGDTAEILTHVANY FEQRESPEILVQVSNY DGGVIEVVIQAADF TDDGEVEILVQAADF	
consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu1817_Dhaf_23113315 BH1549_Bha1_10174166 Desu7142_Dhaf_23121547 hrd22-1_Ddeh_6466833	EEEI YGLI LPII YSII FGII AIKI	ps.a. EEEEEE.EEEEEEE.EEEEEE. RISHIRMASAITYNGNQÀGGSGK KKSNIRMSSVIFYNGKKIIQDGR RVPSVRSASAIYVDGRLIAGSGQ KISNIRTSSAVFVNGEKIIQQGRP	PALTQI PALLEN LGESEJ PGNSPH -ALLEN	EEE DEYQYENKPYN IEYLFSNIPRL ADYVASNLPYT STTVHKYYADT IEYLFSNIPRV	sh. V 3 G 1 A 4 G 3 G 3	VQGDTÄEILTHVANY PEQRSPELTUVQVSNY NGSGVIEVVTQAANFK TDDGEVELVQAADF LHSDIEILVQVSNY	
consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu817 Dhaf_23113315 BH1549_Bha1_10174166 Desu7142_Dhaf_23121547 hrd22-1_Ddeh_646633 Desu1754_Dhaf_23113246	EEEI YGLI LPII YSII FGII AIKI LG <mark>L</mark> I	ps.a. REFERE . EEEEEEE . EEEEEEE RISNIRMASAIYYORMQAQGSGK KWFSVRSASAIYYORALAGSGQ KITCISASARIFADDQLLECGS KSNIRTSSAVFWOREHIGDGRP KIHNIMAHR <mark>LFI</mark> NQQLYKESGL	PALTQI PALLEN LGESEJ PGNSPH -ALLEN PSDSLH	EEEE DEYQYENKPYN IEYLFSNIPRLA DIYYASNLPYT STTVHKYYADT IEYLFSNIPRVG SGYQALNTPYV	sh. V 3 G 1 A 4 G 3 G 3 V 2	EEEEEEEEEE VQQDAAFLLHVANY FEQRSPETLVQVSNY NGSGVTEVVQAAFK TDDGEVETLVQAAFF LHSSDTETLVQASNY LHSSDTELEVIQVSNY	
consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu1817_Dhaf_23113315 BH1549_Bhal_10174166 Desu7142_Dhaf_23121547 hrd22-1_Ddeh_6466833 Desu1754_Dhaf_23113246 La0289_Lint_24212989	EEEI YGLI LPII YSII FGII AIKI LGLI LAII		PALTQI PALLEN LGESEA PGNSPI -ALLEN PSDSLI PGETAA	EEEE DEYQYENKPYNY IEYLFSNIPRLA ADYVASNLPYT STTVHKYYADT IEYLFSNIPRV( SGYQALNTPYV ATSTPEWNVKI)	sh. V 3 G 1 G 3 G 3 V 2 L 3	LEEEEEEEEEE VQCUTAETLEHVANY FEGRSPELVVQVSNY NGSVJEVVIQAAPK LHSSDIELVVQAAPK EPAEEEIVIQVSNY EPAEEEIVIQVSNQ INKKE7QLRPJSNF	
consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu1817_Dhaf_23113315 BH1549_Bha1_10174166 Desu7142_Dhaf_23121547 haf22-1_Deh_6466633 Desu1754_Dhaf_23113246 LA0289_Lint_24212989 La0241_Lint_24202368	EEEI YGLI YSII FGI AIKI LGLI YGII		PALTQI PALLEI LGESEJ PGNSPI -ALLEI PSDSLI PGETAJ LGTDFI	DEEE DEEQYENKPYN HEYLFSNIPRL ADYVASNLPYT STTVHKYYADT IEYLFSNIPRV IGYQALNTPYV TSTPEWNVKII ITSTPEWNVKII	sh. V 3 G 1 G 3 G 3 G 3 V 2 L 3 T 3	LEEREFEREER VOODTATELLHVANN PEGRSPETLOVSN NOSGUTEVIQAANFK TDOGRELLQAANFK LHSDIELLQAANF LHSDIELLQOSN INKKFQIRIPLSN PARLELTVOSN INKKFQIRIPLSN PANNATUVSNF	
consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu1817_Dhaf_23113315 BH1549_Bhaf_10174166 Desu7142_Dhaf_23121547 hd22-1_Ddah_5466633 Desu1754_Dhaf_23113246 L0289_Lint_24212989 L0281_Lint_24212989 L0281_Lint_24213515	EEEI YGLI YSII FGII AIKI LGLI LAII YGII FKII	PS. a. PEEEEE. EEEEEEE. RISNIRMASAIYYNGNQAGGSGK KKSNIRMSSVIPYNGKKIIQDGR RVPSVRSASAIYYDGKLIAGSGG KITCISASAFTADDQLLAEGS KSNIRTSSAVFVNGEKIIQDGRP KININMARTFINGQLWESGL LAETSETAYEVMIDDNKIGAQGV RWTELLGAFQIFINNSVAQIGV RWTELGAFQUFIDNRKVANGF	PALLEN PALLEN LGESEJ PGNSPH -ALLEN PSDSLH PGETAN LGTDFN VGTNSV	EEE. DEYQYENKPYN HEYLFSNIPRL IDYVASNLPYT ISTYHKYADT NEYLFSNIPRV GGYQALNTPYV ATSTPEWNVKIJ YETLFLAHPLI	sh. V 3 G 1 A 4 G 3 G 3 V 2 L 3 T 3 T 3	LEEEEEEEEEE VQCUTAETLIHVANY PSGRSPELUVUSNY NGSUTEVVIQAAPK LHSSDFTLIQVSNY ENERTYVVVQAAP INKKEPQIRIPLSNF PANELEYVVVSNY NSGOPITTVNVSNF	
consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu1817_Dhaf_23113315 BH1549_Bha1_10174166 Desu7142_Dhaf_23121546 LA0289_Lint_24212989 Lint_242108 LA0285_Lint_24212989 LA0285_Lint_24213515 LA0361_Lint_24213518	EEEI YGLI LPII YSII FGII AIKI LGLI YGII FKII LS <mark>V</mark>		PALLEN PALLEN LGESEN PGNSPH -ALLEN PSDSLH PGETAN LGTDFN VGTNSV PGIDPI	h in SEEE DEYQVENKPYN HEYLFSNIPRLA DYV <sub>A</sub> SNLPYT STVHYYADT GYQALNTPYV TSTPEWNVKI YTMTPKLKPDR VTMTPKLKPDR DSIVSRIEPRD	sh. V 3 G 1 G 3 G 3 G 3 V 2 L 3 T 3 F 3	LEEREFEREER VOODTELLHVANN HOGSVIEVUQAAN NOSSVIEVUQAAN LHSDIELLVQAAN LHSDIELLVQAAN INKKFQLRIPLSN INKKFQLRIPLSN VSSORFULTVNVSFF	
consensus/85% Secondary Structure Desu7352 Dhaf 23121831 Desu817 Dhaf 23113315 BH1549 Bhal 10174166 Desu7142 Dhaf 23121547 hd22-1 Deha 5466633 Desu1754 Dhaf 23113246 L0289 Lint 24212989 L0281 Lint 24212989 L0281 Lint 24213515 L03610 Lint 24213515 L03610 Lint 24213515	EEEI YGLI YSII FGII AIKLI LGLI LAII YGII FKII LSV: YGLI	PS. a. PEEEEE. EEEEEEE. RISNIRMASAIYNONQAGGSGK KKSNIRMSVIFYNORKIIQDGR RVPSVRSSAJYNORKIIQDGR KVTSISASTAITADOLLLACSG KSNIRTSSAVFYNOEKIIQDGR KINNIMMARFJINGQLWESGL LAETSETAYEVHIDDNKKIQAGW RIPRLFGVYEVILDNQKVYSNGF SLGAIASAIRIKINGQIGCGCT KIDELLMSKWINGIIQEKAGI	PALTQI PALLEN DALLEN DGNSPH -ALLEN PSDSLH PGETAM LGTDFH VGTNSV PGIDPI 7 ERQVYI	h in DEYQYENKPYN HEYLFSNIFRL ADYVASNLPYT TTVHKYKADT HEYLFSNIFRV USTFEWNVKI TSTFEWNVKI TSTFEWNVKI TSTFELAHFLI SIVSRIEFRD DSIVSRIEFRD	sh. V 3 G 1 G 3 G 3 G 3 V 2 L 3 T 3 F 3 -	LEEEEEEEEEE VQCUTAELLHVANY PSGRSPELLVVSNY NSGVJEVVJQAAPF LHSSDTELLVOVSNY ENABLEVVJQAAPF PARLEVVVSNP INKKFQIRIPLSNP PONOMIVIVNSPF VSGOPTITVNVSFF ADEELEVVESFYSNY NGIDIVVGSF	
consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu1817_Dhaf_23113315 BH1549_Bhal_10174166 Desu7142_Dhaf_23121547 hd22-1_Dehh_6466633 Desu1754_Dhaf_23113246 La0289_Lint_24212899 LB241_Lint_24212899 LB3610_Lint_24213515 LB3610_Lint_24216308 CAC0818_Cace_15894105 LB3614_Lint_24216313	EEEI YGLI YSII FGII LGLI LAII YGLI LSV YGLI LS <mark>I</mark> I		PALLEN PALLEN LGESEN PGNSPH -ALLEN PSDSLH PGETAN LGTDFN VGTNSV PGIDPI 7 ERQVYI 7 KDGFKH	EEE DEVQENKEVEN IEVLFSNIPRI JUVASNLPYT TTUKKYXADT IEVLFSNIPRV IEVLFSNIPRV ITSTPEWNVKI ITSTPEWNVKI ITSTPEWNVKI SIVSRIEPRD JPGYYYFAK	sh. V 3 G 1 G 3 G 3 V 2 L 3 T 3 F 3 - V 1	LEEBEREBEER PORTELLHVANN- HORSPELLOVSNN- HORSPELLOVSNN- DOERVELVQANFK- TODERVELVQANFK- LHSDFELVQVSNG- INKKFQLRIPUSNF- PORUGNTUVVSNF- VSSUDFVITVNVSFF- VSSUDFVITVNVSFF- NGIDIVVQFSN	
consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu817_Dhaf_23113315 BH1549_Bhaf_10174166 Desu7142_Dhaf_23121547 haf22-1D4haf_243113246 La0289_Lint_24212989 LB241_Lint_24212989 LB241_Lint_24213515 LA3610_Lint_242145315 LA3610_Lint_242145315 LA3614_Lint_242145313 LA2164_Lint_24214531	EEEI YGLI YSII FGII AIKI LGLI LAII YGII LSV YGLI LSII FSI		PALTQI PALLQI PALLEN LGESEJ PGNSPI -ALLEN PSDSLE PGETAJ LGTDFY VGTNSY PGIDPI 7 ERQVYI 7 ERQVYI 7 ERQVYI 7 PQGYDF	EEE DEYQYENKPYN TEYLFSNIFRL DIYVSSNIFRL STTVHKYYADT TEYFSNIFRW ATSTPEWNVKIJ TEMTFKLRPDR JETLFLAHPLI DSIVSRIEPRDJ DFGYYFTAK RTVILPIPENK	sh. V 3 G 1 G 3 G 3 V 2 L 3 T 3 F 3 F 3 V 1 I	LEEEEEEEEEE VQCUTAELLHVANY	
consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu1817_Dhaf_23113315 BH1549_Bhal_10174166 Desu7142_Dhaf_23121547 haf22-1_Dehah_64666833 Desu1754_Dhaf_23113246 La0289_Limt_24212899 LB241_Limt_24212899 LB241_Limt_24216316 LA0815_Limt_24216308 CAC0818_Cace_15894105 LB3614_Limt_24216313 LB3614_Dimt_24216313 LB2676_Limt_24218376	EEEI YGLI YSLI FGLI LALI LALI YGLI LSV: YGLI LSI LSI LAVI		PALIQI PALLEP LGESEJ PGNSPI -ALLEE PSDSLI PGETAJ LGTDFI VGTNSS PGIDPI 7 ERQVII 7 ERQVII 7 KDGFKI 7 PQGVDB 5 PQAYDB	h EEEE DEVQENKPENN HEVLFSNIPRLA DEVVASNLPYT TTTVKKYADT EVLFSNIPRVC EGVQALNTPVV TSTPEWWVKII VETLFLAHPLI SSIVSRIEPRDLI SSIVSRIEPRDLI PGVYPTAK RHVILPIPENKK RTRIYSISENK:	Sh. V 3 G 1 G 3 G 3 V 2 L 3 T 3 F 3 - V 1 I 1	LEEBEREBEE POGRETELHVANN- FEGRSPELLOVSNN	
consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu817_Dhaf_23113315 BH1549_Bhaf_10174166 Desu7142_Dhaf_23121547 hd22-1_Ddah_6466633 Desu1754_Dhaf_23113246 La0289_Lint_24212989 La241_Lint_24212899 La241_Lint_24213515 La3610_Lint_24213515 La3610_Lint_242145313 La2164_Lint_242145313 La2164_Lint_242145313 La2165_Lint_2421215376 La0027_Lint_2421215376	EEEI YGLI LPII YSII FGII LAI YGLI LSII FSII LSVI TAI		PALIQI PALLEP PGNSPF -ALLEP PGSTAJ PGETAJ PGETAJ PGETAJ 7 ERQVII 7 EQGYDI 7 EQGYDI 10PSVR82	blick bl	Sh. V 3 G 1 A 4 G 3 G 3 V 2 L 3 T 3 F 3 F 3 F 3 F 3 I 1 G 3	LEEEEEEEEEE VQCUTAELLHVAW PSQRSPELLVVQVSW NSGVIEVVQAAF DDGVEVVQAAF LHSSDFELVQVSW LHSSDFELVQVSW LHSSDFELVQVSW PARELEVVVSF PARELEVVVSF NGVVTVVVSF NGVDEFVSV NSGDFVTTVWSF ADEDEVQTEFVSV NGIDVVQFSM HINDVVSF NGIDVVQFSM HINDVF SGNTLVIEVSF HKNTLVIESSM -AGEENT	
consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu817_Dhaf_23121831 Desu817_Dhaf_23113315 BH1549_Bhal_10174166 Desu7142_Dhaf_23121547 hd22-1_Ddah_6466633 Desu1754_Dhaf_23113246 La0289_Lint_24212899 L8241_Lint_24212899 L8241_Lint_24212815 LA3614_Lint_24216313 LA3614_Lint_24216313 LA266_Lint_24218376 LA007_Lint_2421376 LA007_Lint_2421727 cyala8_Lint_2421727	EEEI YGLI LPII YSII FGII AIKI LGLI YGLI SKII LSV YGLI LSII FSI LSV TAI TAI		PALTQI PALLEP PGNSPF -ALLEP PSDSLL PSDSL UGTDPT VGTNSV PGIDPI 7 ERQVII 7 ERQVII 7 ERQVII 7 ERQVDI 5 PQAYDI 10PSVRRS 9 LRRTIF		sh. V 3 G 1 G 3 G 3 G 3 C 3 G 3 C 3 T 3 F 3 F 3 F 3 L 1 G 3 L 1	LEEBEREBEER VOORDELLHVANN- NOSCUTEVIQAANFK- TDOGRVILVQAANFK- TDOGRVILVQAANFK- TDOGRVILVQAANF LHSDTELVQVSNG- FARELEVIQVSNG- VSSGDFVITVNVSFF VSSGDFVITVNVSFF NGIIDVVQPSNBINIFN NGIIDVVQPSNDINIFN FKNNELVIKSNDISELNY FKNNELVIKSDENTIF FKNNELVIKSDESDFTU FKNNELUIGAP-ASFLSKN	
consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu817_Dhaf_23113315 BH1549_Bha1_10174166 Desu7142_Dhaf_23121547 hd22-1_Ddah_6466633 Desu1754_Dhaf_23113246 La0289_Lint_24212899 LB241_Lint_24212899 LB241_Lint_24213515 LA3610_Lint_242145315 LA3610_Lint_242145315 LA3614_Lint_242145315 LA3614_Lint_242145315 LA3614_Lint_242145315 LA3614_Lint_242145376 LA0027_Lint_242125376 LA0027_Lint_24212777 cyaA18_Lint_24217179	EEEI YGLI YSLI FGII LAII YGLI FKII LSLI FSLI LAVI TAI IFLI VAL		PALIQI PALLQI PALLEI GESEZ PGNSPI -ALLEI PSDSLI PGETAJ LGTDFP VGTNSV FGIDPI 7 ERQVYI 7 KOGFKI 7 PQGVDI 5 PQAYDI 10PSVRR8 9 LRRTII 7 SWNLPI	b. 1 SEEE DEVQENKPYN HEILFSNIFEL DIVASNLFYN HEILFSNIFEL HEILFSNIFEL HEILFSNIFEL HEILFSNIFEL SIVSRIEPED DGYTYFAK- NUILPIPEN KINILPIPEN SIKSVILPIPE SIKSVILPIPE SIKSVILPIPE SIKSVILPIPE	sh. V 3 G 1 G 3 G 3 G 3 C 3 F 3 F 3 F 3 F 3 I 1 I 1 G 3 L 1	LEEEEEEEEEE VQCUTXELLHVAW PSQRSPELLWVAW NSGVIEVVIQAAPK TDJGVEVIQAAPK LHSSDIEILVQVSW LHSSDIEILVQVSW LHSSDIEILVQVSW PARELEVIVVSNF PARELEVIVVSNF SGNDFIITVWSFF NGIIDIVIQPSN HINTKIIISSN ADEDEVOLEFFVSW IGKNELLVIKKY FKNELLVIKKY GENTILVIEVKK FKNELVIEVKK ADETEVIEVE	
consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu817_Dhaf_23121831 Desu817_Dhaf_23113315 BH1549_Bhal_10171466 Desu7142_Dhaf_23121547 hd22-1_Ddah_6466633 Desu1754_Dhaf_23113246 La0289_Lint_24212899 La241_Lint_24212899 La241_Lint_24212899 La3614_Lint_24216308 CAC0818_Cace_15894105 LA3614_Lint_24216313 LA2166_Lint_24218376 LA007_Lint_2421376 LA007_Lint_24217179 Reut4832_Rmet_22980162	EEEI YGLI YSII YSII LQLI LAII YGLI FKII FKII LAVI TAI IFL VALI		FALTQI PALLQ PALLQ PGNSP -ALLEP PGDSJL FGETAJ LGTDFY VGTNSY FGIDPI 7 ERQVYI 7 ERQVYI 7 ERQVYI 7 ROGYNF 5 PQAYDI 10PSVRR 9 LRRTII 7 SWNLPI 7 SWNLPI 7 AWNLPI		Sh. 	LEEBEREBEE POGROPETLUVIA NGSUTEVIQAAPK NGSUTEVIQAAPK LHSDTELUVGAAP LHSDTELUVGAAP PARELEVIUVSF PARELEVIUVSF NGIDFVICVOPSH NGIDFVICVOPSH NGIDFVICVESH NGIDFVICSH FXNRELIXIKS PSNCATUVESH PSNCATUVESH SGELLYIKS SGELLYIKS SGELLYIKS SGELLYIKS SGELLYIGD SGENLITSILDA SGELLYIGD SGENLITSILDA SGENLI	
consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu817_Dhaf_23113315 BH1549_Bha1_10174166 Desu7142_Dhaf_23121547 hd22-1_Ddeh_6466633 Desu1754_Dhaf_23113246 La0289_Lint_24212899 LB241_Lint_24212899 LB241_Lint_24213515 LA3610_Lint_24213515 LA3614_Lint_242145313 LA2164_Lint_242145313 LA2164_Lint_242145313 LA2164_Lint_242145314 LA2067_Lint_242125376 LA0027_Lint_24212737 cyaA18_Lint_24212777 cyaA18_Lint_24212777 cyaA18_Lint_24212737	EEEI YGLI YSII YSII LQLI LAII YGLI LSII SSI LSII LSII LSII LSII LSII LAVI YGLI SQLI YAL		PALTQI PALTQI PALLEI LGESEZ PGNSPI -ALLEI PSDSLL PGETAJ LGTDFI VGTNSY PGIDPI 7 ERQVII 7 ERQVII 7 ERQVII 7 ERQVII 5 FQAYDI 10PSVRRS 9 LRRTII 7 SWNLPI 7 AWNTPI 7 QVIPI		Sh. 	LEEEEEEEEEEE VQCUTXELLHVAW PSCRYELLHVAW NSGVTEVVIQAAPK	
consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu817_Dhaf_23121831 Desu817_Dhaf_23113315 BH1549_Bhal_10171466 Desu7142_Dhaf_23121547 haf22-1_Ddah_6466633 Desu1754_Dhaf_23113246 La0289_Lint_24212899 La241_Lint_24212899 La241_Lint_24212899 La3614_Lint_2421515 La3610_Lint_2421515 La3614_Lint_24215376 La0027_Lint_24213376 La0027_Lint_24213776 Reut4832_Rmet_228900182 Mdeg0217_Mdeg_23026370 Mdeg0217_Mdeg_23026370	EEEI YGLI LPI YSII FGI LGLI LGLI YGII LSV YGLI LSV TAI TSI TSI YAL YAL YQLI YQLI YQLI YQLI YQLI		PALLQ PALLQ PALLQ PALLQ PALLQ PGDS PSDSLL GTDPF VGTNSV FGIDPI 7 ERQVII 7 ERQVII 7 ERQVII 7 ERQVII 7 ERQVII 10PSVRR 9 LRRTII 7 SWNLPP 7 SWNLPP 7 SWNLPP QVNLPI QWNLPI	h. 1: EEEE	Sh. V 3 G 1 A 4 G 3 G 3 G 3 V 2 L 3 G 3 V 2 L 3 F 3 - V 1 I 1 G 3 L 1 - - - -	LEEBEREBEEN NGSUTEVIQAAPK NGSUTEVIQAAPK IDJURVIQAAPK LHSDIELIQUSA LHSDIELIQUSA INKERQIRIPLSN PARELEVIQUSA INKERQIRIPLSN NGIDIVUQPSN NGIDIVUQPSN HDEDVOLEFYSN NGIDIVUQPSN HDEDVOLEFYSN FKNRILISSN ABEENVOLEFYSN SGENLITIGSN FKNRILISSN ABEENVOLESSN ABEENVOLESSN FKNRILISSN ABEENVOLESSN FKNRILISSN ABEENVOLESSN	
Consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu817_Dhaf_23113315 BH1549_Bha1_10174166 Desu7142_Dhaf_23121547 hd22-1_Ddah_6466633 Desu1754_Dhaf_23113246 La0289_Lint_24212899 LB241_Lint_24212899 LB241_Lint_24213515 LA3610_Lint_24213515 LA3610_Lint_242145308 CACOB18_Cacc_15894105 LA3614_Lint_242145315 LA3614_Lint_242145308 LA267_Lint_242145376 LA0027_Lint_242125376 LA0027_Lint_24212737 cyaA18_Lint_24212777 cyaA18_Lint_24212777 cyaA18_Lint_24212777 cyaA18_Lint_24212777 cyaA18_Lint_24217179 Reut4327_Rent_22980018 Reut43777_Rent_22980018 Reut4327_Rent_2298018 Reut4327_Rent_22980118 Reut4327_Rent_22980118 Reut4327_Rent_22980118 Reut4327_Rent_22980118 Reut4327_Rent_22980118 Reut4327_Rent_22980118 Reut4327_Rent_22980118 Reut4327_Rent_22980118 Reut4327_Rent_22980118 Reut4327_Rent_22980118 Reut4327_Rent_22980118 Reut4327_Rent_22980118 Reut4327_Rent_22980118 Reut4327_Rent_22980118 Reut4327_Rent_22980118 Reut4327_Rent_22980118 Reut4327_Rent_22980118 Reut4327_Rent_2298018 Reut4327_Rent_2980118 Reut4327_Rent_2980118 Reut4327_Rent_2980118 Reut4327_Rent_2980118 Reut4327_Rent_2980118 Reut4327_Rent_2980118 Reut4327_Rent_298018 Reut4327_Rent_298018 Reut4327_Rent_298018 Reut4327_Rent_298018 Reut4327_Rent_298018 Reut4327_Rent_298018 Reut4327_Rent_298018 Reut4327_Rent_298018 Reut4327_Rent_298018 Reut4327_Rent_298018 Reut4327_Rent_298018 Reut4327_Rent_298018 Reut4327_Rent_298018 Reut4328_Rent_298018 Reut4328_Rent_298018 Reut4328_Rent_298018 Rent_298018 Reut432	EEEI YGLI YSII FGII AIKI LAII YGLI LSV YGLI LSV TAI IFL: VAL GQII VAL QLI		PALLEY PALLEY PALLEY PGNSP -ALLEY PGETAJ LGTDFY VGTNSY PGIDPI 7 ERQVYI 7 ERQVYI 7 ERQVYI 7 ERQVYI 7 ERQVYI 7 ERQVYI 7 PQGYDB 9 LRRTIT 7 SWNLPI 7 SWNLPI 7 AWNTPY QUPDI QMNLPI GADQPI	LEEE. LEEEE. LEEEE. LEEV JENN PY IN HEYLFSN I PALA DY NANN PY II TY UNK YY ADY UNY IN TY UNK YA TY TY UNK YA TY SI SANK SI SANK	Sh. V 3 3 G 1 4 G 3 4 G 3 3 G 3 2 C 3 G 3 V 2 C 3 F 3 F 3 F 3 F 3 F 1 I 1 I 1 G 3 L 1 	LEEEEEEEEEEE VGCUTÄELLINVANY FEGRSPELLVQVSNY DGCVTEVVIQAAPK LHSSDIELVQVSNY LHSSDIELVQVSNY HKNEFQIR.PLSNF PARELEYVQVSNQ HKNEFQIR.PLSNF NGIDTVVQPSN	
Consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu817_Dhaf_23113315 BH1549_Bhal_10174166 Desu7142_Dhaf_23121547 hd22-1_Ddah_6466633 Desu1754_Dhaf_23113246 La0289_Lint_24212889 LB241_Lint_24212899 LB241_Lint_24212899 LB3614_Lint_24213515 LA3610_Lint_24213515 LA3614_Lint_24216313 LA2165_Lint_2421376 LA007_Lint_2421376 LA007_Lint_24213776 LA007_Lint_24217179 Reut4832_Rmet_228900162 Mdeg0217_Mdeg_23026370 Md	EEEI YGLI FGII AIKI LAII FKII SV: YGLI FKII LAVI TAII IFL VAL VAL VAL VAL VAL ATLI		PALLEY PALLEY PGNSP PGNSP PGNSP PGDSL PGETAJ UGTNS PGIDH 7 ERQVI 7 ERQVI 7 ERQVI 7 ERQVI 7 ERQVI 9 LRTTI 7 SWNLPY QWNLPY GADQPI RWDTPI	LEEE. SEQUENCE VIEW AND AN AND AND AND AND AND AND AND AND	sh. 3 G 3 G 3 G 3 G 3 G 3 G 3 G 3 C 2 G 3 F 3 F 3 F 3 F 3 F 3 F 1 I 1 I 1 G 3 L 1 	LEEBEREBEEN NGSUTEVIQAAPK NGSUTEVIQAAPK IDJURVIQAAPK LHSDIELIQUSN LHSDIELIQUSN HUVUQAAP PARELEVIQUSN HUVUSH PARELEVIQUSN HUVUSH PARELEVIUVSF SGBPTITNVSF HUVUSH HUV	
Consensus/85% Secondary Structure Desu7352 Dhaf 23121831 Desu817 Dhaf 23121831 Desu817 Dhaf 23113315 BH1549 Bhaf 10174166 Desu7142 Dhaf 23121547 hd22-1 Ddah 6466633 Desu1754 Dhaf 23113246 L0289 Lint 24212899 L8241 Lint 24212899 L8241 Lint 24213515 LA3610 Lint 242145308 CACOB18 Cace 15894105 LA3614 Lint 242145308 LA267 Lint 24214537 LA267 Lint 24214537 LA267 Lint 24214537 LA267 Lint 24214537 LA267 Lint 24214537 LA267 Lint 2421457 LA267 Lint 242	EEEI YGL YSII YSII YGL AIKI YGL YGL YGL YGL YGL YGL YGL YGL YGL YGL		FALTQI PALLER PALLER PGETAJ PG	h. 1t EEEE 	sh. V 3 1 A 4 G 3 G 3 V 2 L 3 T 3 F 3 V 1 I 1 I 1 I 1 I 1 I 1 I 1 I - - - - - - - - - - - - - - - - - - -	LEEEEEEEEEE VQCUTAELLHVANY PEGRAFELLVQVSNY NSGVIEVVIQAAPK LHSSDIELVQVSNY LHSSDIELVQVSNY EPAELEVYQVSNQ HINKORPOIRTYLSNF PAELEVYQVSNQ HINKORPOIRTYLSNF NGIDTVVQPSN	
Consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu817_Dhaf_2312313 BH1549_Bhaf_10174166 Desu7142_Dhaf_23121547 hd22-1_Ddah_6466633 Desu1754_Dhaf_23113246 La0289_Lint_24212889 La241_Lint_24212889 La241_Lint_24212889 La241_Lint_24213151 La3610_Lint_24213151 La3610_Lint_242145315 La2166_Lint_242145316 La2067_Lint_242145316 La0267_Lint_2421896 La2067_Lint_2421896 La2067_Lint_24218971 Reut4832_Rmet_22800128 Medg0217_Mdeg_23026370 Mc20135_Kcam_2112846 XCC1103_Kcam_21233525 La2_E67348_Ccre_25398225 Lacz_L62_149566	EEEI YGLI YSII FGII LGII LAI YGLI LSV YGLI SSI ISI FSI IFL VAL YQLI ATLI ATLI ATLI XSLI		PALLEY PALLEY PGNSPF PGNSPF PGNSPF PGDSL PGETAJ LGTDFI VGTNSS PGIDPI 7 ERQVII 7 ERGVII 7 ERGVII 7 ERGVII 7 ERGVII 7 ERGVII 9 LRRTII 7 SWNLFP GADQPP RWDTPI -SFTFF	h. 1 TEEEE. TEEEE. TEEVATENPENT TETVIKYADTOT TETVIKYADTOT TETVIKYADTOT TETVIKYADTOT TETVIKYADTOT TETJE T	sh.  G 1 G 3 G 3 G 3 G 3 G 3 G 3 G 3 G 3 C 1 C 1 C 1 C 1 C 1 C 1 C 1 C 1 C 1 C 1	LEEBEREBEEN VOOLTNELL HVANY	
Consensus/85% Secondary Structure Desu7352 Dhaf 23121831 Desu817 Dhaf 23121831 Desu817 Dhaf 23113315 BH1549 Bhaf 10174166 Desu7142 Dhaf 23121547 hd22-1 Ddah 6466633 Desu1754 Dhaf 23113246 L0289 Lint 2421289 L0241 Lint 2421289 L0241 Lint 2421289 L0341 Lint 2421315 L13610 Lint 242145315 L13614 Lint 242145315 L13614 Lint 242145315 L12166 Lint 242145316 L12266 Lint 24214536 L12267 Lint 24212577 cyalls Lint 2421277 cyalls Lint 2421277 cyalls Lint 24212376 Reut4327 Rent 22980018 Reut4777 Rent 22980018 Reut4327 Rent 22980118 Reut4327 Rent 2298018 Reut4327 Re	EEEI YGL YSII YSII LGI LGI LGI LGI LGI LSI SSI SSI SSI SSI SSI SSI SSI SSI SS		PALITQI PALLER FGNSPF -ALLER PSDSLL PGETAJ LGTDFI VGTNSY FGIDPI 7 ERQVII 7 ERQVII 7 ERQVII 7 ERQVII 7 ERQVII 9 LRRTII 7 SWNLPI 7 SWNLPI 9 LRRTII 7 SWNLPI 7 SWNLPI 7 SWNLPI 7 SWNLPI 9 CADQPI RADQPI RADQPI RADDPI SFTPI -SFTPI	h. 1t EEEE 	sh. 	LEEEEEEEEEEE VQCDTAETLINVANY TSCRPELLIVQVSNY DSCRPELUQVSNY DSCRPELUQVSNY LHSSDIETLIQVSNY EPAELEIVQVSNY EPAELEIVQVSNY DSCRPTTVNVSTF DAUDEVVLQAADFT NGIDTVVQPSN	
Consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu1817_Dhaf_23113315 BH1549_Bhaf_10174166 Desu7142_Dhaf_23121547 hd22-1_Ddah_6466633 Desu1754_Dhaf_23113246 La0289_Lint_24212889 La241_Lint_24212889 La241_Lint_2421315 La3610_Lint_2421315 La3610_Lint_242145315 La3616_Lint_242145315 La2166_Lint_242145316 La2166_Lint_242145316 La2067_Lint_2421277 cyal8_Lint_24215376 La0027_Lint_24215376 La0027_Lint_24215376 La0027_Lint_2421890118 Reut4777_Remt22380062 Mdeg0217_Mdeg_23026370 C07348_Ccre_25398225 lacz_L6_149556 lacz_L91a_28379799 LaC4Klac_17305	EEEI YGLL YSI YSI YGL YGL YGL YGL YGL YGL YGL YGL YGL YGL	PS a PEREER. FREEFER. EEFEFER. ISNIRMAS.TYNNONGAGGGK KKNNIKNSVIFYNORKLIAGGG KITTISASARIFADDQLILEGG SWIFWISASARIFADDQLILEGG KITHISASARIFADDQLILEGG KITHISASARIFADDQLIKEGGL LAETSETAFVEWIDDKKOAGGW KITHISAFQLFINNGKGAGGW RIPRLPGVTEWIDDKKOAGGW KITHISAFQLFINNGKGAGG ISLAIASATRIAINGILGECGT ISLAIASATRIAINGILGECGT ISLAIASATRIAINGILGECGT ISLAIASATRIAINGILGECGT ISLAIASATRIAINGILGECGT LIGHIDGKWEININGLGECGT LIGHIDGKWEININGILGECGT LIGHIDGKWEININGILGECGT ULGIUTUKKINNINGLGECGT ULGIUTUKKINNINGLGECGT ILSIGENKEITINGILGECGT ULGIUTUKKINNINGLGECGT ILSIGENKEITINGILGECGT ILSIGENKEITINGILGECGT ILSIGENKEITINGILGECGT ILSIGENKEITINGILGECGT ILSIGENKEITINGILGECGT ILSIGENKEITINGILGECGT ILSIGENKEITINGILGECGT ILSIGENKEITINGILGECGT ILSIGENKEITINGILGECGT ILGIUTUKAARINGING KUGUNDITINGKEIGSTQ- IFDGUNSAFHMCUGVGEKK	PALLEP PALLEP PGNSPP -ALLEP PGNSPP -ALLEP PGETAA LGTDFF VGTNSS FGIDPF 7 ERQVIX 7 ERQ	h. 1 THEEE. THE SECONTENT OF THE SECONT OF	sh. 	.EEEEEEEEEEE .UQGDTAELLHVANY	
Consensus/85% Secondary Structure Desu7352_Dhaf 23121831 Desu817_Dhaf 23121831 Desu817_Dhaf 23113315 BH1549_Bhaf 10174166 Desu7142_Dhaf 23121547 hd22-1_Ddeh 6466633 Desu1754_Dhaf 23113246 L0289_Lint 24212899 L0241 Lint 24212899 L0241 Lint 24212899 L0341 Lint 242145315 L35616_Lint 242145315 L35616_Lint 242145315 L3616_Lint 242145315 L3616_Lint 242145315 L3616_Lint 242145315 L3616_Lint 242145315 L3616_Lint 242145315 L3616_Lint 242145315 L3626_Lint 242145315 L3626_Lint 242145315 L3626_Lint 242145315 L3626_Lint 242145315 L3626_Lint 242145315 L3626_Lint 242145315 L3626_Lint 24214535 L3626_Lint 24214535 L3626_Lint 24214535 L3626_Lint 24214535 L3626_Lint 24214535 L3626_Lint 24214535 L3626_Lint 2421455 L3626_Lint 24214555 L3626_Lint 24214555 L3626_	EEEJ YGL YSII YSII LGL LGL LGL LGL LGL LGL LGL LGL LGL L		PALIQI PALLEP FGNSPF -ALLEP PSDSLL PGCTAA LGTDFN VGTNSS PGIDPI 7 ERQVID 7 ERQVID 7 ERQVID 7 ERQVID 7 ERQVID 7 ERQVID 7 ERQVID 7 ERQVID 7 ERQVID 9 LRRTII 7 SWNLPI 7 S	h. 1t EEEE 	sh. V 3 G 1 G 3 G 3 V 2 G 3 V 2 L 3 F 3 F 3 F 3 F 3 I 1 I 1 G 3 L 1 - - - - - - - - - - - - -	VQGDTAETLINVANY VQGDTAETLINVANY VQGDTAETLINVANY NGGUTEVUQAAPK TDGRPELLQVUQAAP LHSSDIETLQVSNY HEALEN HEALEN VQSNQ HEALEN VSSGPFITTNVSTR NGIDTVUQPSN	
Consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu1817_Dhaf_23113315 BH1549_Bhaf_10174166 Desu7142_Dhaf_23121547 hd22-1_Ddah_6466633 Desu1754_Dhaf_23113246 La0289_Lint_24212899 La241_Lint_24212899 La241_Lint_24212899 La241_Lint_2421315 La3610_Lint_242145315 La3610_Lint_242145316 La216f_Lint_242145316 La2616_Lint_242145316 La2616_Lint_242145316 La2616_Lint_242145316 La2616_Lint_242145316 La2616_Lint_242145316 La2616_Lint_24212727 cyals_Lint_24212727 cyals_Lint_24213516 La0027_Lint_24212727 cyals_Lint_24213536 Mdeg0217_Mdeg_23026370 CXC1753_Kcam_2112846 XCC1753_Kcam_251328255 lact_L67148_Ccre_25398225 lact_L67148_56 lact_Lpla_28379799 LaC4_Klac_173305 ebgA_Ec_26110922 LaC2_Kpine_L14911	EEEI YGLL YSLL YSLL YSLL YGL YGL YGL YGL YGL YGL YGL YGL YGL Y	PS a PEREER. FREEFER. EREFERE PEREER. FREEFER. EREFERE ISNI MINAS.TYNONGAGGGK KKNN IKNSSV FYNORKLIAGGG KITTISASAR FADDQLILEGG KITTISASAR FADDQLILEGG KITTISASAR FADDQLIKEGGL LAFTSETAFVENTUDIKTGAGG KITHISAFOLFUNDKVINGAGG KITHISAFOLFUNDKVINGAGG SIGAIASATRIKINGIIGECGT IIPELGYNUFUNDKVING SIGAIASATRIKINGTIGECGGI IIPELGYNUFUNDKUSG SIGAIASATRIKINGTIGECGGI FPDLGENNEITINGIS LGGIDBCHNEITINGI FIKSKF HJGTUTDKKITINGTIGECGGI ILSIGENNEITINGI KGRGG ILSIGENNEITINGI KGRGG ILSIGENNEITINGTIGECGT PFDLGENNEITINGTI KGRGG ILSIGENNEITINGKUGT LGGVUNGERFINDREGSS2 ILGFUNDATUTINGKUGYGE JUGUNSAFHMCUNGY KDGAATIYVINGGVUGFKK KFDGVERLIYVINGYUGFKG	PALIQ PALLE PGNSPH -ALLE PGDSPH -ALLE PGDETA IGTDPF VGTNS FGIDPT 7 ERQVII 7 ERQVII 7 ERQVII 7 EQVII 7 EVNIP PR MDTPJ -SETPF -SFTPF -SRLPI -SRLPI -SRLPI	h. 1 THEEE. THEEE. THEEE. THEET.	sh. .v 3 G 1 G 3 G 3 V 2 L 3 T 3 F 3 V 2 L 3 T 3 F 3 V 1 I 1 G 3 L 1 - - - - - - - - - - - - - - - - - - -	LEEBEREBEER VOGUTATEL LIVANY	
Consensus/85% Secondary Structure Desu7352 Dhaf 23121831 Desu81752 Dhaf 23121831 Desu81754 Dhaf 23121847 Hrd22-1 Dedm 5466833 Desu1754 Dhaf 23113246 L0289 Lint 2421289 L0241 Lint 2421289 L0241 Lint 2421289 L0341 Lint 2421375 L03615 Lint 2421375 L03616 Lint 242145308 L0267 Lint 242145308 L0267 Lint 242145308 L0272 Lint 2421454 L0272 Lint 2421454 L0	EEEJ YGLL YSLL YSLL YSLL YSLL YSLL YSLL YSLL		PALIQY PALIQY PALIQ PALS PGNSP PSDSLL PGDPU PSDSLL GTDPY VGTNS FGIDPY 7 ERQVI 7 ERQVI 7 ERQVI 7 ERQVI 7 ERQVI 7 ERQVI 9 LRRTI 1055VR8 9 LRTI 1055VR8 9 LRTI 7 SWNLPJ 9 LRTI 7 SWNLPJ 7	h. It EEEE IEEEE IEEEE IEEETSIN FRIARDAWN IEITSIN FRIANDAWN IEITSIN FRIANDAWN IEIT	sh. V 3 G 1 A 4 G 3 G 3 G 3 G 3 C 1 G 3 C 1 C 3 C 1 C 3 C 1 C 1 C 1 C 1 C 1 C 1 C 1 C 1	VQGDTAELLIVANY VQGDTAELLIVANY VQGDTAELLIVANY NGGVIEVVIQAAPK DDGRVELUQVSNY INSKOTEVVIQAAPK LHSSDIELLQVSNY INSKOTFOLA PARELEVVQVSNG PARELEVVQVSNG DGLDYVQSN DGLDYVQSN NGIDIVVQSN NGIDIVVQSN NGIDIVVQSN HINTKFF NGIDIVVQSN HINTKF KNNTLIVIKSN AGNATLUVKX MGNLLIVKX MGNLLIVKX MGNLLIVKX MGNLLIVKX MGNLLIVKX MGNLLIVKX MGNLLIVKX MGNNTLUXX MGNNTLUXX	
Consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu1817_Dhaf_23113315 BH1549_Bhaf_10174166 Desu7142_Dhaf_23121547 hd22-1_Ddah_6466633 Desu1754_Dhaf_23113246 La0289_Lint_24212899 La241_Lint_24212899 La241_Lint_24212899 La241_Lint_2421315 La3610_Lint_242145315 La3610_Lint_242145316 La216f_Lint_242145316 La216f_Lint_242145316 La2616_Lint_242145316 La2616_Lint_242145316 La2616_Lint_242145316 La2616_Lint_2421896 La2616_Lint_2421896 La2616_Lint_242189118 Reut4777_Reut4832_Remet_22380062 Mdeg0217_Mdeg_23026370 CG7348_Core_25398225 lacz_L62_149556 lac1_L91a_28379799 LaC4_Klac_173305 ehgA_Ec_26110922 LaC2_Kpne_1182464 Mc2193_Tmar_15643949 TM153_Tmar_15643949	EEEI YGL YSI HIY YGL YGL YGL YGL YGL YGL YGL YGL YGL YG	PS a PEREER. FREEFER. EEFEFER. ISNI MINAS TYNN NNGAGGGK KKNN IKNS V FYN NKKI LODGR KINT ISNA SYN YN ORLLAGSG KINT ISNA SYN YN ORLLAGSG KINT ISNA YN YN ORLLAGSG KINT ISNA YN YN ORLLAGGG KINT YN SAL YN ORLLAGGG KINT YN ARW DD KLOGOG KINT YN ARW DD KLOGOG KINT YN ARW DD KLOGOG SIGAIAS YR KINE DI LODKAGI HFELLWY YN DI MELODKAG SIGAIAS YR KINE DI LODKAGI HFELLWY YN GRUG SGGI HFELLWY YN GRUG SGGI HFFLIAN YN FLW YN GRUG SGGI HFFLIAN YN FLW HWL GAFG HFFLIAN YN FLW HWL GAFG HGFLYD MAG RAF IND FLW RAF HGFLYN YN GRUG YN GAFG HGFLYD HYN NAG YN GRUG YN GAFG HGFLYN HWL MAR GNYG HGFLYN HWL MAR GNYG HGFLYN HWL MAR GYNG HGFLYN HWL MAR GYNG HGFLYN HWL MAR FLW YN MAR YN YN G HGFLYN HWL MAR FLW YN MAR YN YN G FGUN YN HWL MAR FLW YN MAR YN YN FLW YN YN G FGUN YN HWL MAR FLW YN MAR YN YN G FGUN YN HWL MAR FLW YN MAR YN YN FLW YN YN G FGUN YN HWL MAR FLW YN MAR FLW YN YN HWL YN YN HWL YN YN FLW YN YN G FGUN YN HWL MAR FLW YN MAR FLW YN YN HWL YN YN HWL YN YN FLW YN YN HWL YN YN FLW YN HWL YN YN YN HWL YN Y	ALLEP PALTQU PALLED PGNSPH -ALLEP PGDSJL GTDPF VGTNS FGIDPF VGTNS FGIDPT FGPQYD FGPQYD FGPQYD FGPQYD FGPQYD FGPQYD FGPQYD FGPQYD FGPQ FGPQ FGPQ FGPQ FGPQ FGPQ FGPQ FGPQ	h. 1 THEEE. THE SET OF THE SECOND S	sh. V 3 G 1 A 4 G 3 G 3 G 3 C 3 G 3 C	. BEBEBEREBER . BEBEBEREBER . BEBEBEREBER . BODE LLOVSNY	
Consensus/85% Secondary Structure Desu7352 Dhaf 23121831 Desu81752 Dhaf 23121831 Desu81754 Dhaf 23121847 http://downer.com/ http://downer.co	EEEJ SUBJECT SUBJEC	PS a SEEEEE. FEEEEEEE. EEEEEEEE. SEEEEE. FEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE	PALIQ PALIQ PALIQ PALIQ PGNSP -ALLE PGNSP -ALLE PGDSL GDP VGTNS FGIDP 7 ERQVI 7 ERQVI 7 ERQVI 7 ERQVI 7 ERQVI 7 ERQVI 7 ERQVI 9 LRRTI 7 SWNLP 9 LRRTI 7 SWNLP 9 LRTT 9 LRTT 9 LRTT 9 SRLP -SFTP -SFTP -SRLP -SFTP -SFTP -SFTP	h. It EEEE 	sh. V 3 G 1 A 4 G 3 G 3 G 3 C 3 G 3 C 3 F 3 F 3 F 3 F 3 F 3 F 1 I 1 I 1 G 3 L 1 F - F - F - F - F - F - F - F - F - F -	LEEBEEBEBEE VQGDTAETLIYVANY TEQRSPELLIYQVSNY NGGVTEVVIQAAPF TDGEVELUQVSNY LHSSDIELUQVSNY LHSSDIELUQVSNY HNGKDYEVVIQAAPF PABELEVVVVSNF DGEVELUQVSN HINKEFORRPLSN SGDPYTTYNVSTF NGIDTVVQSN HINKEFORRPLSN SGDPYTTYNVSTF NGIDTVVQSN HINKEFORRPLSN SGDPYTTYNVSNF SGELLIYRKS GENLIYRKSL SGELLIYRG GENTLIYKSL SGENLIYRKSL AGENSTVVKTSSL AGENSTVVKTSSL AGENSTVVKTSS SGELLIYR AGENSTVVKTSS SGELLIYR SGENLIYRKS SGENLIYR	
Consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu1817_Dhaf_23113315 BH1549_Bhaf_10174166 Desu7142_Dhaf_23121547 hd22-1_Ddah_6466633 Desu1754_Dhaf_23113246 La0289_Lint_24212899 La241_Lint_24212899 La241_Lint_24212899 La241_Lint_2421315 La3610_Lint_2421315 La3610_Lint_242145316 La216f_Lint_242145316 La216f_Lint_242145316 La2616_Lint_242145316 La2616_Lint_242145316 La2616_Lint_242145316 La2616_Lint_242145316 La2616_Lint_242145316 La2616_Lint_2421896 La2616_Lint_2421896 La2616_Lint_2421897 cyahl8_Lint_24217179 Reut4832_Rmet_22800128 Mdeg0217_Mdeg_23026370 CG7348_Core_25398225 lac2_L62_149566 lac1_Lpla_28379799 LaC4_Klac_173305 ebgA_Ec_26110922 LaC2_Kpne_1182464 Mc3055 ebgA_Ec_261395 lac2_Ka13255 ebgA_Ec_2731 Max52216_Cace_144746 Max52216_Cace_1243751	EEEI JGL JFI JFI JFI JFI JFI JFI JFI JFI JFI JFI	PS a PEREER. FEREFER. EERFER. FEREFER. FEREFER. FEREFER. FEREFER. FINI INMAS. TYWNORGAGGGK KKNN IKNSSV FYWNORK I LOGOR KITHI SASAR FADOLLLEGG KITHI SASAR FADOLLLEGG KITHI SASAR FADOLLLEGG KITHI SASAR FADOLLEGG KITHI SASAR FADOLLEGG KITHI SASAR FADOLLEGG SIGAI SASAR SASAR SASAR SASAR SIGAI SIGAI SASAR SASAR SASAR SASAR SIGAI SIGAI SASAR SASA	PALTQI PALTQI PALTQI PALTQI PALLQE PGETAJ PGETAJ VGTNST PGIDDI 7 RGGFKI 7 RGGFKI 7 RGGFKI 7 RGGFKI 7 RGGFKI 7 RGGFKI 9 LRRTIT 10PSVRRS 9 LRRTIT 7 SWNLPI 7 SWNLPI 7 SWNLPI 7 SWNLPI 7 SWNLPI 7 SWNLPI 7 SWNLPI 9 LRTTI 9 LRTTI 9 SFTPI -SFTPI -SFTPI -SFTPI -SRLPI -S	h. 1 THEEE. THEEE. THEEE. THEET. THEY AND THE	S V 3 G 1 A 4 G 3 G 3 G 3 C	. BEBEBEREBER . BEBEBEREBER . BEBEBEREBER . BOD STATUS . BEBEBEREDUS . BOD STATUS . BEBELEU QUSN . DOG DY ET LY WUSH . DOG DY E	
Consensus/85% Secondary Structure Desu7352_Dhaf 23121831 Desu817_Dhaf 23121831 Desu817_Dhaf 23121831 Desu817_Dhaf 23113315 BH1549_Bhaf 2017166 Desu7142_Dhaf 23121547 Ha22-1_Ddeh 6466633 Desu1754_Dhaf 23113246 La0289_Lint 24212899 La241 Lint 24212899 La241 Lint 2421315 La3610_Lint 24214508 CACOB18_Cace_15894105 La3614_Lint 24214508 CACOB18_Cace_15894105 La3614_Lint 24214508 La266_Lint 24214508 La266_Lint 24214508 La266_Lint 24214508 La266_Lint 24214996 La267_Lint 2421277 cyal8_Lint 2421717 Reut432_Rent 22980018 Reut4777_Rmet 22980018 Reut432_Rent 22980118 Reut4777_Rmet 22980012 KCC103 Kccm 21132525 CR7348_Crce 2539225 lacz Ldel 149566 lact_Df1 28379799 LaC4 Klac 173305 ebgA Ec 26110092 LaC5 Kpme 114941 TH193_Tmm 15643949 lacz Sal 18265743 AnA52216 Cace 144746 lacA Tthe 22138751 GUSB Mum 114964	EEEI JGLI JGLI JGLI JGLI JGLI JGLI JGLI JGL		PALIQU PALIQU PALIQU PALLE LGESEZ PGNSP -ALLE PGETAJ LGTDF VGTNSY FGIDPI 7 ERQVTI 7 ERQVTI 7 ERQVTI 7 ERQVTI 7 ERQVTI 7 ERQVTI 7 ERQVTI 7 ERQVTI 7 ERQVTI 9 LRTTI 7 SWNLPF GADQPT SFTPS -SFTPS	h. It EEEE IEEEE IEEEE IEEET IEEET IEEET IEET	S S	BEEREFEREERE           VQCOTART LHVANY           FEDEREFEREER           VQCOTART LHVANY           FEDEREFERE           SGUTEVUQUSNY           FEDEREFERE           SGUTEVUQUSNY           DGGEVELUQUANFK           LISSDIELLQUSNY           DREVELUQUANFK           SGUTEVUQUSNY           EPAERLETVUQUSNY           DNSGPVILVIVVSTFK           DSGUTEVUNSTFK           DSGUTEVUNSTFK           DGENVOLEFVSNY           DKINTULVIVSTFK           DGENTULVIVSTFK           DGENTULVIVSTFK           GENTILVISKY           GENTILVISKY           GENTILVIKSVG           GENTILVIKSVG           GENTILVIKYGL           GENTILVIKYGL           GENTILVIKYGL           GENTILVIKYKSLAFT           GENTILVIKYKSLAFT           GENTULVIKYKSLAFT	
Consensus/85% Secondary Structure Desu7352_Dhaf_23121831 Desu1817_Dhaf_23113315 BH1549_Bhaf_10174166 Desu7142_Dhaf_23121547 hd22-1_Ddah_6466833 Desu1754_Dhaf_23113246 L0289_Lint_24212899 L8241_Lint_24212899 L8241_Lint_2421375 LA3610_Lint_2421375 LA3610_Lint_242145315 LA3614_Lint_242145315 LA3614_Lint_242145315 LA3614_Lint_242145315 LA3614_Lint_242145316 LA2067_Lint_242145316 LA2067_Lint_242145316 LA2067_Lint_242145316 LA2067_Lint_242145316 LA2067_Lint_242145316 LA2067_Lint_242145316 LA2067_Lint_242145316 LA2074_Lint_242145516 LA2074_Lint_242145516 LA2074_Lint_242145516 LA2074_Lint_242145516 LA2074_Lint_242145516 LA2074_Lint_242145516 LA2074_Lint_24214	EEEL LPII JSII JSII JAIKI LAIL JSI JSI SSI SSI SSI SSI SSI SSI SSI SSI		PALTQUE PALLQUE PALLQUE PALLQUE POSTAL LOTTPF POETAA POETA	h. 1 THEEE. THEEE. THEEE. THEEE. THEY AND THE THEE THEE THEE THEE THEE THEE THEE	S V 3 G 1 G 3 G 3 G 3 G 3 C	. EEEEEEEEEEE 	
Consensus/85% Secondary Structure Desu7352_Dhaf 23121831 Desu817_Dhaf 23121831 Desu817_Dhaf 23113315 BH1549_Bhaf 10174166 Desu7142_Dhaf 23121547 Ho22-1_Ddeh 5466833 Desu1754_Dhaf 23113246 L0289_Lint 2421289 L021 Lint 2421289 L021 Lint 2421289 L021 Lint 2421375 L0361 Lint 242145315 L0361 Lint 242145308 L0267_Lint 242145315 L0361 Lint 24214536 L0267_Lint 2421277 cyall8_Lint 24214996 L0267_Lint 2421277 cyall8_Lint 24214996 L0267_Lint 2421277 cyall8_Lint 24214996 L0267_Lint 2421277 cyall8_Lint 24214996 L0267_Lint 2421277 cyall8_Lint 24214996 L0267_Lint 2421277 cyall8_Lint 24214996 L0267_Lint 2421277 cyall8_Lint 2421375 Cd7348_Core 2539018 Reut4377_Rmet 22980018 Reut4377_Rmet 22980018 Reut4377_Rmet 2298018 Reut4377_Rmet 2298018 Reut4377_Rmet 2298018 Reut4377_Rmet 2298018 Reut4377 Cd7355 L025_C10352 L025_C10352 L025_K1ac 173305 ebg3_Ec 2610092 L025_Kne 114941 TH193_Tmm 15643949 L025_Sall 12265743 AAA22216 cace 144746 L028_Hsap 4504223 uda Ec 15804986	EEEI LIPI LIPI LIPI LIPI LIPI LIPI LIPI	PS 8 SEEEEE. FEEEEEEE. EEEEEEE. SEEEEE. FEEEEEEE. SEEEEE. FEEEEEEE. SEEEEE. FEEEEEEE. SEEEEE. FEEEEEEE. SEEEEEE. SEEEEEEEEEEEEEEEEEEE		h. 1 LEEE LEEE LEEE LEET	s s	VQGDTAELEFEEEEE VQGDTAELIAVANY PGCRSPELLAVONY NGSGUTEVUQAADF DDGEVELUQUSAY INSKUTEVUQAADF INSKUTEVUQAADF DDGEVELUQUSA INSKUTEVUQAADF INSKUTEVUQAADF DDGEVELUQUSA DDGEVELUQUSA DGEDEVOLEFYSNY NGIDIVUQPSN DSGDEVITWVSNF VSSGPFITWVSSF NGIDIVUQPSN HINFK FKNNSLLTIKFY FKNNSLLTIKFY FKNNSLLTIKFY FKNNSLLTIKFY GENLIFKSL GGVNTLUKVSL DGVNTLUKVSL GGNNTUKTSNA GGNVLARV MGNDGAV GGNNTUKTSNA GENNLAVAVINSD GSILED GGNNTUKTSNS VINULATYKSS MGNNUALTYKSS MGNNUA MGNNUALTYKSS MGNNUA MGNNUALTYKSS MGNNUA MGNNUALTYKSS MGNNUA	

#### Figure 4

Multiple sequence alignment of the 7TMR-DISMED1 and accessory domains of sialate 9-O-acetylesterases,  $\beta$ glucoronidases and  $\beta$ -glucosidases. Multiple sequence alignment the 7TMR-DISMED1 was constructed as detailed in the legend to Figure 1. The PHD-secondary structure [78] is shown above the alignment with E representing a  $\beta$  strand, and H an  $\alpha$ -helix. In addition to the convention described in Fig. 1 the consensus also shows the aliphatic subset of the hydrophobic class (I; ALIVMC, yellow shading) and the aromatic subset of the hydrophobic class (a; FHWY, yellow shading). The families shown to the right are A – 7TMR-DISMEDI, B – accessory domains of sialate 9-O-acetylesterases and C – accessory domains of  $\beta$ glucoronidases and  $\beta$ -glucosidases. The species abbreviations are as shown in Table I and Kpne – Klebsiella pneumoniae; Klac – Kluyveromyces lactis; Ldel – Lactobacillus delbrueckii; Lpla – Lactobacillus plantarum; Ssal – Streptococcus salivarius; Tthe – Thermoanaerobacterium thermosulfurigenes; Hsap – Homo sapiens; Mmus – Mus musculus.

Secondary Structure	EEEEE		EEE	CEEEEEEE	EEEEEE	EEEEEEE	EEEEE	EEEEEE	EEEEEEEE.	EEEEEEE	
Psyr2444_Psyr_23470811	41 LPLGRAIQVFEDP	rgeaties <mark>v</mark> saaahasa <mark>f</mark> k	PVPPGTFNAGYSRSAF	<mark>VL</mark> K <mark>VE</mark> LTYRPTDAS	IHNDWL <mark>L</mark> ELAYPP <mark>N</mark>	<mark>IDHIDFY</mark> GPDADGQ	-PTRAWHTGDMLPFS	5- <mark>R</mark> QFAQNNYL <mark>F</mark> Q <mark>L</mark> D	LP-PGQTRT <mark>LYM</mark> R <mark>I</mark> RS	SEGAVQAPLY <mark>L</mark> WST	185
PP4781 Pput 24986538	28 LPLGKAMQVYEDPI	)GNASIAQ <mark>V</mark> SAPGFAKH <mark>F</mark> Q	PHHEDVLNAGYSTSVF	<mark>VL</mark> KVE <mark>L</mark> RPTAAPNA	APRSWL <mark>L</mark> ELAYPP <mark>I</mark>	LDHLELYLPDSSGL	-YRLAQRTGDALPYDS	5- <mark>R</mark> QIRQNNYL <mark>F</mark> ELQ	LP-PGKVTT <mark>AYL</mark> R <mark>L</mark> HS	SQGSVQAPLA <mark>L</mark> WSP	172
Pflu4318 Pflu 23062251	25 AAQGSGWSVLLDD	QGNLQLSD <mark>I</mark> RSARYTNQ <mark>F</mark> S	PIELDRLTAAEPDGAL	VLRFRLAPG	-KHEQVLRIFAPDI	LSHLN <mark>LY</mark> VLDGDR-	-LIEQRDTGTAQPQA	E-RPLPSSDFM <mark>L</mark> PLP	QSDKPLD <mark>VYL</mark> RLVS	SDHQLRPYIT <mark>L</mark> QAA	161
PP4824_Pput_24986585	25 AEHDGGWTVLLDE	QANLQLSD <mark>V</mark> RSERYRNQ <mark>F</mark> S	PLPLADLDAAPAGQAL	VLHYRLAPG	DQEQLLRVFAPDI	<mark>LSGLDLY</mark> ALEGEQ-	-LLRQLHHGRQAGNAS	5- <mark>P</mark> TLRGSDHV <mark>L</mark> P	NSRQPLD <mark>IYL</mark> RLVS	SEHQLRPAIS <mark>L</mark> EPA	161
PA3974 Paer 9950164	28 LPLGQSIDVFEDVE	RGSADIND <mark>I</mark> TSRAIDSS <mark>F</mark> F	RHDKDVLNAGYSRSVF	<mark>VL</mark> RLDLDYRPVASS	DPRTWL <mark>l</mark> elaypp <mark>i</mark>	L <mark>DKLD<mark>LY</mark>LPDGQGG</mark>	-YRLAQRTGDTLPFAS	5- <mark>R</mark> PIRQNNYL <mark>F</mark> ELG	LE-PNKPQR <mark>VYL</mark> R <mark>L</mark> ES	SQGSIQAPLT <mark>L</mark> WSP	172
Magn8928_Mmag_23016342	44 QPLGGQWSALRDPS	SGTLRIED <mark>V</mark> AQSEAAHG <mark>F</mark> A	PL-PGNLAAGYDRGAW	VLSAEIRRTADA	-SADWLLEVEPAFI	LDEVTLFTGDGQG-	-GFRAVTIGDRVPFS	A- <mark>R</mark> QLAYRTLV <mark>F</mark> R <mark>L</mark> H	LP-DDLPLR <mark>VYL</mark> R <mark>I</mark> KS	STSTISVSAQ <mark>I</mark> LSP	183
Mdeg0804_Mdeg_23026970	20 LVIQQGLAFLEDKS	5GALSFED <mark>V</mark> QTRGD <mark>W</mark> Q	LQASGNFSQGYNRSAW	VLKFELLNTQQ	ANIWL <mark>L</mark> EVAYPI <mark>I</mark>	LDYVT <mark>VY</mark> VGNPQG-	-VYTKVEMGDKLPFDS	5- <mark>R</mark> PIKHRNFL <mark>V</mark> PLD	LT-KREPVT <mark>IYI</mark> R <mark>V</mark> AS	3SSSVQFPLV <mark>L</mark> WEP	156
PA3462_Paer_9949605	32 LSLGAYAEYYRDAG	GKARLGD <mark>I</mark> LALPA-QA <mark>F</mark> A	ALRGDHANFGKNAAAW <mark>W</mark>	<mark>vF</mark> R <mark>V</mark> RLDNRNGA	-DLAGF <mark>L</mark> EVNYPL <mark>I</mark>	LDDLK <mark>VY</mark> LLTADG-	-RIEQQESGDLFAFS	Q- <mark>R</mark> PVQVRNFW <mark>F</mark> P <mark>L</mark> R	LP-PGES-T <mark>LLL</mark> R <mark>V</mark> QS	STSTVYLPLY <mark>F</mark> STY	170
Mdeg0691_Mdeg_23026847	52 LELDNNFKLYEDVS	SASATIDD <mark>V</mark> AQVFEDEA <mark>T</mark> I	QFSVGDTNIGYSSSAW	WKLSFHNPQSA	PKEII <mark>L</mark> RQAYPL <mark>I</mark>	DNLQ <mark>LW</mark> TRSYES-	-HWSVKQYGDKFPFSH	I- <mark>R</mark> EIYHREFL <mark>F</mark> P <mark>L</mark> I	IP-ARSTKE <mark>YFL</mark> R <mark>Y</mark> ES	<b>SEGSLDIALS<mark>I</mark>HE</b> P	192
C83029_Paer_11350324	81 AWLNGSLDLLEDPI	DGNLAVED <mark>L</mark> EQAEQAGR <mark>F</mark> V	AA-AGRTSVGLSRSAW	<mark>√LRL</mark> DLPRREAV	-SGGWWLEVASASI	LHDLR <mark>LY</mark> LPDERG-	-GFREHRSGEAVPFA	EG <mark>R</mark> DHAYRHPL <mark>F</mark> R <mark>I</mark> P	PGDGPLR <mark>VYL</mark> RSYI	)PGGNAFPLR <mark>L</mark> WSH	220
Chut0509_Chut_23135396	35 EIQPQFVEQLQDPA	AASYTIDQ <mark>V</mark> SSPLFAAK <mark>F</mark> K	ATGETMLRNTNRQAAY <mark>W</mark>	VLKLRIINHTDK	-SQQWL <mark>I</mark> ESFNFR	INEIS <mark>CF</mark> LQTDS	-GFTEVVEGDMYAFK	Q- <mark>R</mark> SVGHKNFE <mark>F</mark> LLP	NDRDSII <mark>CYL</mark> R <mark>I</mark> KI	<i>l</i> KQPASFELF <mark>I</mark> RRF	173
Avin0379_Avin_23102206	42 LSLGPYTLYVEDRI	DADLDAAQ <mark>V</mark> LAMPE-SA <mark>F</mark> F	PVEGDQANLGQSRSAW	<mark>WF</mark> R <mark>VDL</mark> QNARDQ	SLGGF <mark>L</mark> EVNYPL <mark>I</mark>	LDDVR <mark>LT</mark> VLTPDG-	-RQLQQESGDTRPPG	4- <mark>R</mark> PVPIRNLW <mark>F</mark> P <mark>L</mark> E	LP-PGTS-T <mark>LLV</mark> R <mark>V</mark> KS	SSSAITIPLY <mark>F</mark> GTS	180
Chut1543_Chut_23136440	34 EWRKGNYAIYRDT	/RNLPLSR <mark>I</mark> LELQREGI <mark>F</mark> #	KTTSDAPRLQSSRQDM <mark>V</mark>	<mark>√A</mark> K <mark>VIV</mark> YNGSLN	-QTEWL <mark>I</mark> ELYDFH <mark>I</mark>	<mark>LDEYD<mark>IY</mark>ILQKDS-</mark>	-LYAHFTGGDLQPFH	K- <mark>R</mark> KIEHKNFI <mark>HE</mark> IV	FE-PHQLYT <mark>LFI</mark> R <mark>I</mark> HS	JRQSVAVNGV <mark>V</mark> RTF	174
PA4856_Paer_15600049	44 PSANQNWRLLRDES	5AQLRIAD <mark>V</mark> LQRKEQ <mark>F</mark> F	PLAKRSFIFPASPQAV	<mark>VL</mark> QVQLPAQK	VPSW- <mark>L</mark> WIFAPR <mark>\</mark>	<mark>/Q</mark> YLD <mark>YY</mark> LVQDGQ-	-LVRDQHTGESRPFQH	E- <mark>R</mark> PLPSRSYL <mark>F</mark> S <mark>L</mark> P	V––DGKPMT <mark>LYV</mark> R <mark>M</mark> TS	SNHPLMAWFDQIDE	178
Mmc12300_Msp_23000642	49 QSLQGHFQLLPAQ	ATPLTIEQ <mark>V</mark> IKPQITQQ <mark>F</mark> I	NINWGFAGGFQS-GER <mark>W</mark>	<mark>VLRVNL</mark> EREPDA	PSRWI <mark>I</mark> ELGSAL <mark>I</mark>	L <mark>DRLE<mark>LY</mark> I PTPTPG</mark>	-HYQRVVLGDHTPLN	Q- <mark>R</mark> PMASRLYS <mark>V</mark> P <mark>L</mark> D	LG-SAPQTT <mark>LYL</mark> R <mark>L</mark> QI	<i>ILQMMTLFGT<mark>IWQ</mark>E</i>	189
Chut1592_Chut_23136490	31 IPIGGYSEIFTDPO	GNTHTIQD <mark>V</mark> IKTGQ <mark>F</mark> V	AANSSTPNLGISPNAF	WKFTIQNQSD	-ADKLLLEYDLPF	<mark>LDEIS<mark>VY</mark>EVDKDKN</mark>	EVTGVHTSGDKYNFSI	¶− <mark>R</mark> EYDYQNYI <mark>F</mark> S <mark>I</mark> A	IK-KNEIKD <mark>IYL</mark> F <mark>I</mark> KS	3GEQIVLPIM <mark>L</mark> GST	169
Chut3598_Chut_23138550	41 MKIDKGIMYFQDE	INLLTIND <mark>I</mark> VSRKD <mark>F</mark> I	AVNQKVPNLGVTSSAQ	<mark>≪L</mark> KITIKNTTD	-VP <mark>KLL</mark> KVDFPI	DEIE <mark>FY</mark> SFEDG	-KMRIEKMGEYKDFS	E- <mark>R</mark> RYNDPNYI <mark>Y</mark> D <mark>I</mark> T	IK-NNQTQT <mark>YYL</mark> K <mark>I</mark> KS	3GEQIMLPIT <mark>I</mark> GTP	176
Chut2183_Chut_23137079	34 SLSDNYITILEDT	TAAFSITE <mark>V</mark> IAADSAQQ <mark>F</mark> K	QA-SYNFNI-HPASVY	<mark>WIKFTV</mark> KNEGDV	FTQYV <mark>I</mark> ENYYAH <mark>7</mark>	AKEFS <mark>VF</mark> YQEHG	-KLFQQKTGEYVTYK	¶− <mark>R</mark> NFSHKNLV <mark>L</mark> DLP	APQAGNERT <mark>FYL</mark> K <mark>V</mark> YS	3GLYVNFNFV <mark>INN</mark> Q	172
Magn8953_Mmag_23016367	45 RPLDGYLTFLEDR	ESAFSIED <mark>V</mark> ARHGV <mark>F</mark> E	AV-TPRRPGLISGGTL	<mark>WRFTV</mark> KRTAGA	-AEDWVLAFGEPD	<mark>EDDVR<mark>VY</mark>VTKPGG-</mark>	-SFTETLLGRRLPAR	2-ldvavrlha <mark>a</mark> n <mark>l</mark> s	LP-EEIPTT <mark>VYI</mark> R <mark>L</mark> SS	3LHKIRFEAA <mark>A</mark> LWR	182
Chut0774_Chut_23135667	32 LNIGNKASIRIDT	NQTYTIHQ <mark>I</mark> LSDTPE <mark>Y</mark> V	YP-EAMPNLGLSSYVY	<mark>∛ATFDL</mark> VNTS	-SENLI <mark>L</mark> KFNQPL <mark>I</mark>	LNEIT <mark>LY</mark> RTDGK	-SLSEITYSESEPFLS	5- <mark>r</mark> kyyypsya <mark>f</mark> d <mark>m</mark> n	AN-MGDTIT <mark>YVL</mark> R <mark>F</mark> KS	3NEPIIFDVD <mark>IAN</mark> Q	166
Chut0775_Chut_23135668	34 INIGNQAAIYADTH	NQTYTIHQ <mark>I</mark> LSGTP-KR <mark>I</mark> Y	SSAMPNLGLSNHVY	WATFNFVNIS	-PEKLI <mark>L</mark> KFNQPL <mark>I</mark>	LNEVT <mark>LY</mark> RADGK	-TLSEVTYSEKNVFF8	5- <mark>r</mark> kyyypfya <mark>f</mark> d <mark>m</mark> n	AG-MGDTIT <mark>YVL</mark> K <mark>F</mark> KS	3NEPIIFDVD <mark>IAT</mark> E	168
Chut2463_Chut_23137358	49 YVVRKYAEYYEDQQ	QRGKKITE <mark>I</mark> AGKVPFKK <mark>I</mark> S	EK-DLDFTNNNLTSAY	VLH <mark>F</mark> YVVNHTPD	-LTDFI <mark>I</mark> EMYDYD <mark>1</mark>	<mark>ENEVDLY</mark> IKDSKG-	-SFIEKKSGLDFPFD	r-rdfahknic <mark>f</mark> k <mark>l</mark> s	IP-LADTTE <mark>VYM</mark> R <mark>L</mark> YS	SSINVFEPV <mark>IKTH</mark>	188
Avin0870_Avin_23102702	57 LDLHPHLQLLHDPO	GRLQAEE <mark>V</mark> MASGQ-AF <mark>A</mark> F	AAGRHDLNFGYTRDIV	VLRLDLESRAKE	VREWQ <mark>I</mark> EFLYPS <mark>I</mark>	<mark>LD</mark> RIE <mark>LF</mark> GIGE	APLLGGDRVPAG	2– <mark>R</mark> DSPHLSPA <mark>F</mark> S <mark>V</mark> R	LA-PGERRS <mark>LFF</mark> R <mark>T</mark> QS	3SGTLTLDAQ <mark>LWE</mark> S	192
Rrub0628_Rrub_22966033	27 ADHILARAFLEDP	AGVLTIDE <mark>V</mark> ARGD <mark>F</mark> I	PF-GPNFSKGYTQSVY	WRLLVRASAA	-PEKTV <mark>L</mark> FIRPSF <mark>I</mark>	LNEVR <mark>LF</mark> YRDRASP	DGWATRVSGNRYPFSI	E-RDRKSVALG <mark>F</mark> V <mark>V</mark> D	VTHPQEE <mark>FYL</mark> R <mark>I</mark> KI	LNTQMQIDVQ <mark>A</mark> VSP	162
Chut2884_Chut_23137778	42 YVDEKSQVCIIDS	SNFQEWNP <mark>V</mark> NRMS <mark>Y</mark> F	PTSKFFRNRSID-HAA	VYT <mark>F</mark> KITNTSRT	-ALEWY <mark>L</mark> VSYNYS <mark>I</mark>	<mark>EDEID<mark>LT</mark>TVSDAG-</mark>	QTEEFQFRDTTSIYN-	<mark>R</mark> IIKHKQPV <mark>F</mark> KVS	lk-anetkt <mark>yyl</mark> rlki	MESTYNYVFA <mark>L</mark> YSH	177
Chut2495_Chut_23137390	54 PLRSGLIHVFEDTI	DNSCTSVE <mark>S</mark> VQTKT <mark>F</mark> F	PANSYFFSTPNPSSTY	<mark>NGKFIL</mark> TDNSSI		L <mark>DSLDIF</mark> AFKKSTL	QFHKKYRFGSPNETT-	<mark>KEIRHKNFT</mark> VD <mark>F</mark> P	VS-KNDTLT <mark>VYI</mark> K <mark>L</mark> KI	<u>IKNATQYDFA<mark>LVE</mark>H</u>	192
Rrub0627_Rrub_22966032	15 EDLVVSRAVIEDPO	3GVLSIDD <mark>V</mark> VSRAA <mark>V</mark> F	FDAVLSKGFSASVY	WRLRVRAPEA	GARTV <mark>L</mark> LISPSF <mark>I</mark>	LNDVR <mark>LF</mark> RRDPKAA	QGWEVRVTGNLHPFA	E- <mark>RERKAVSLS</mark> FVVG	VSPPQED <mark>FYF</mark> RIKI	lrspmtfevq <mark>als</mark> v	150
Avin3704_Avin_23105533	61 LVLQPWLRVFEDA	RADMSLEQ <mark>I</mark> LALPE-ER <mark>L</mark> S	AATAHLLAPGYSRSAF	VLRLELENRTGQ	ACQWW <mark>L</mark> FPGSAR <mark>7</mark>	ARDMI <mark>LH</mark> QQEAG	-GWSRQVAGAWHPLAN	E-WTMPTRLPA <mark>F</mark> EVG	LP-AGTSKV <mark>LWL</mark> RLAS	3DYAFVMQPM <mark>LLS</mark> H	199
XF0986_Xf_9105918	65 AQSLMLERLDDDPI	PAHEVLAG <mark>V</mark> YDAMLRF	NDTGGASIYETARHPV <mark>V</mark>	WRIRADRQISA	AGQPK <mark>L</mark> QIEFPY <mark>I</mark>	LNWVE <mark>AW</mark> VPGRSV-	-PSRHAIYGAAAD	<mark>R</mark> RYATRALV <mark>I</mark> DLP	EG-LPQGRA <mark>VWL</mark> R <mark>V</mark> HA	AQSTIPMPVS <mark>IVS</mark> N	199
Avin2657_Avin_23104497	42 GAPNGDWRILLDES	SASLTLKD <mark>V</mark> IERRDH <mark>F</mark> A	PL-GHRSLTLPANQAA	VLRVSIAGHD	TPRW- <mark>I</mark> WVFAPR <mark>\</mark>	<mark>/DRVD</mark> FFLTNRGA-	-TERRIETGAMLPDG-	LSTSGQAHL <mark>F</mark> DLP	T––DQTTRE <mark>VWL</mark> RLAI	?RQAAPAWFD <mark>YVD</mark> T	174
Pflu4190_Pflu_23062112	27 LPLGRSLQVFEDPS	5GQASIAD <mark>V</mark> RAQAAAGN <mark>F</mark> K	PHDKATLNAGYSRSAF	<b>VLKIDL</b> HYRPTNPS	SAQRSWL <mark>L</mark> ELAYPP <mark>I</mark>	LDHLD <mark>LY</mark> LPDAAGD	-YRLVRQTGDAC	RSPVAKS-ART <mark>I</mark> T	CS-TWRSSR <mark>ISS</mark> EPFI	CGCPAKARSRRR-	165
LA3996_Lint_24198039	36 SLSFFVYYFSSDSI	DENFRERI <mark>F</mark> AKDSS-LS <mark>F</mark> Ç	NIPAEVFSLGFTSNTV <mark>P</mark>	<b>FYIPLKNDTEK-</b> -	-DYRGEFEIFNSY <mark>I</mark>	LEEVDIFYRYGSKD	-SIHEILAGTS	<mark>R</mark> VYEKSFPA <mark>LNF</mark> Y	LR-PGEEIQ <mark>IVC</mark> KIKS	3GTPMRIPIV <mark>LES</mark> E	171
S04207_Sone_24350618	36 NLMPWLTVTHLNT	rseladiq <mark>a</mark> lpktk <mark>w</mark> h	QF-TSGDIQRLSQHNF	VLTFSIQSGDE	SLSRI <mark>L</mark> ALDNPL <mark>I</mark>	<mark>LDKVT<mark>LY</mark>HLVGNQ-</mark>	-LINTTYMGDTLPYQQ	Q- <mark>R</mark> PLLSNIFL <mark>Y</mark> P <b>F</b> K	IN-ANEQHT <mark>FYL</mark> HIET	LEGNAAVPIN <mark>L</mark> WSA	171
LA3986_Lint_24216685	39 ICEFDQIEFALDPI	DLSNEVPK <mark>E</mark> PKKSLV <mark>F</mark> L	PKENSFLKLGFIKESV	VIR <mark>FNI</mark> KQYP	RSRCF <mark>L</mark> RIPQVT <mark>I</mark>	LDGAALFAK	SSVQITGDRFRYS	E- <mark>RFVDDYYVF<mark>Y</mark>PLE</mark>	PLDIQKEKNQ <mark>YY</mark> L <mark>W</mark> IH	(TSSIINFPI <mark>FLE</mark> S	171
LA4269_Lint_24216967	42 NISSLIEYRYRGQ	QFAGCSPE <mark>H</mark> IDGLEDLE <mark>W</mark> H	<b>QIPTEVLRVKRTPFGN<mark>V</mark></b>	VLR <mark>F</mark> SVKNSEST	-IQSRILLLGWLN	/PDTQLCFFDKSG-	-KFVSARSGYSNDIEI	D- <mark>E</mark> KILTNLPH <mark>F</mark> RID	LE-PNENRI <mark>FYL</mark> F <mark>V</mark> LS	SNEDINYRIQ <mark>I</mark> MGL	182
S01570_Sone_24373140	28 IALDSPYLFHAEA-	-KQLPPAD <mark>F</mark> KEVSQ <mark>W</mark> M	IGQLKEASSVSLTGGDY <mark>N</mark>	MVSPVMVNSR	QTRWV <mark>V</mark> DASNSI	LESVD <mark>YW</mark> LLGSDG-	-SVQFAHSGYYAPY	<mark>E</mark> FLFDYGRK <mark>V</mark> RLN	MGTDYWL <mark>VTR</mark> LSS(	2YFSSAPEVA <mark>LES</mark> Q	160
Psyr2376_Psyr_23470745	109 AEHGAGWSTLLDER	KAHLTLDE <mark>I</mark> RSARYQNQ <mark>F</mark> S	PIELERVTAADRDSAL	<mark>VLHYRL</mark> QPT	QHEQL <mark>L</mark> RIFAPD <mark>I</mark>	LSSAD <mark>MY</mark> VMDGDR-	-QIDHVRTGNDVPIE	D-QRLPSNDFL <mark>L</mark> PVP	QSSAPLD <mark>IYL</mark> RLVS	3AQKMRPSIT <mark>L-E</mark> P	244
SMc00074_Sme1_15073849	49 LDLTATTEIYSGRO	3DAFQVST <mark>A</mark> PGTDGI <mark>V</mark> F	RIEVRSSTENHQGD	NAVFALANVSEEQ-	-LERVI <mark>V</mark> APHFRL <mark>V</mark>	NSKL <mark>FW</mark> PDLGS	-QRILSITPSEGFALI	D- <mark>R</mark> QPSEEADV <mark>F</mark> RIT	LN-PGAVIT <mark>FVA</mark> EL	-TTPELPQIY <mark>L</mark> WQP	183
AGR_C_1434_Atum_15888127	49 LDLTATTDIYANQO	3EAFQVST <mark>A</mark> PGPDGI <mark>R</mark> F	RIEVRASSEDHQGD	<mark>NAVFAL</mark> ANVSEEQ-	LERVI <mark>V</mark> APHFRL <mark>V</mark>	<mark>/N</mark> SKL <mark>FW</mark> PDLGS	-QRIIAITPSEGFALI	D- <mark>R</mark> QPSPDADV <mark>F</mark> RIT	LN-PGSVIT <mark>FVA</mark> EL	-STPQLPQIY <mark>LWE</mark> P	183
AG3433_Bme1_25526193	44 LDLSRAVELLRNKO	3ESVQVST <mark>M</mark> PGPDGI <mark>V</mark> F	RIEVQSDQNA-NASGD <mark>V</mark>	<mark>WAAFSI</mark> ANPTDEQ-	-IDRLI <mark>V</mark> APHFRL <mark>V</mark>	/GSGV <mark>IW</mark> PDLGS	-PRIASITPSEGFALI	D- <mark>R</mark> QPSADADV <mark>F</mark> R <mark>I</mark> T	LN-PGSVIT <mark>FVA</mark> EL	-SSHNLPQLY <mark>L</mark> WEP	179
Rpa12189_Rpa1_22962895	25 IDLTGVLEHLHSD	NDRVQIST <mark>A</mark> PGNDGI <mark>V</mark> F	RVEVRAREGGQN	WVFALANNSDDQ-	-LDRLI <mark>V</mark> APHYRM <mark>V</mark>	/GSGL <mark>LW</mark> PDLGM	-SRIATITPSTGDRPH	E- <mark>R</mark> QDSATADI <mark>F</mark> R <mark>V</mark> T	LD-PGSVIT <mark>FVA</mark> EL	-RTDKLPQLY <mark>L</mark> WEP	157
bll1502_Bjap_27349754	24 IDLTGVLEHQRSD	ADRIQVST <mark>A</mark> PGTDGI <mark>V</mark> F	RIEVRAREGGQN <mark>W</mark>	WVFALANNTDDQ-	-LDRLI <mark>V</mark> APHYRI <mark>V</mark>	<mark>/S</mark> SGL <mark>LW</mark> PDLGL	-SRIATITPSTGDRP	E- <mark>R</mark> QESPTADV <mark>F</mark> R <mark>V</mark> T	LD-PGAVIT <mark>FVA</mark> E <mark>L</mark>	-RTDKLPQLY <mark>L</mark> WEP	156
consensus/90%	s.	phh.	s.	Wh.h.l	p11	lpha	·····s····s	ph.h.	hhhch	h.p.	

**Multiple sequence alignment of the 7TMR-DISMED2s.** Multiple sequence alignment the 7TMR-DISMED2 domain was constructed as detailed in the legend to Figure 1. The 90% consensus follows the same convention as in Figure 1 and 2. The species abbreviations are as shown in Table 1.

comprised entirely or largely of proteins from a single organism (Figure 2). These lineage specific expansions [55] accounted for almost all of the 7TMR-DISMs in organisms such as, Desulfitobacterium, Leptospira and Cytophaga, which contained multiple copies of these 7TM domains. Furthermore, some diversity was observed in the intracellular domains associated with the 7TM domains from the proteins belonging to these lineage specific clusters (Figure 2). This suggested that the 7TMR-DISM evolved through lineage specific expansion of the 7TM domain in various bacterial lineages, accompanied by some domain shuffling in their C-terminal intracellular modules. The lineage specific expansion of these 7TMRs is reminiscent of the evolution of the eukaryotic 7TMRs such as the odorant receptors of vertebrates and chemoreceptors of nematodes [55]. The predicted role, for at least a subset of these receptors, in carbohydrate sensing, is consistent with their expansion in Cytophaga hutchthat is known to actively metabolize insonii polysaccharides [56]. The expansion in Leptospira implies that this spirochete too may respond to diverse carbohydrates. Alternatively, it could also utilize its numerous

7TMR-DISMs to recognize different carbohydrates on host cell surfaces to regulate its motility.

# 7TMR-HD: a novel family of bacterial receptors with a HD hydrolase domain

Another family of potential bacterial 7TMRs that was recovered by our receptor search procedure was typified by slr0104 from Synechocystis. This family of receptors is present in cyanobacteria, spirochetes like Leptospira and Treponema, most low GC Gram positive bacteria, some proteobacteria, like Magnetospirillum and Geobacter, Chloroflexus, Fusobacterium and Chlamydia pneumoniae (Figure 6). All members of this family are characterized by the presence of an intracellular HD hydrolase domain [57] Cterminal to the 7TM domain. Accordingly, we named these receptors as 7TMR-HD (for 7TM receptors with intracellular HD domains). The majority of 7TMR-HD proteins also contain a distinct domain N-terminal to the 7TM domain that is predicted to localize to the extracellular or periplasmic compartment, based on the presence of a signal peptide. Analysis of the predicted topology and searches against a position specific score matrix library of

Secondary structure	
YQFF Bsub 16079587	276 FKP <mark>I</mark> SGLLIM <mark>I</mark> GLFIATLVYYFEKQKQNLKFK-NQS <mark>ILLF</mark> SIITTLLLVIMEVVSLFQK 6 GYLVPIA-AGA <mark>IL</mark> IKLL <mark>M</mark> NERIAILGSI <mark>I</mark> LAICGSMMFN
lin1503 Linn 16800571	290 VKQ <mark>Y</mark> AGFAVF <mark>I</mark> IALAALLFLYTKKQTISKAKK-MQT <mark>MLIF</mark> SSVYVVSLFMLMIILFLEN 6 AFLFPAA-FAP <mark>MI</mark> LKIL <mark>L</mark> NEKYAFLSVL <mark>Y</mark> IAVTSLFAFQ
1mo1466 Lmon 16803506	290 VKQ <mark>Y</mark> AGFAIF <mark>I</mark> IALAAILFLYTKKQTQPKAKK-MQT <mark>MLIF</mark> SSVYLVSLFMLFIILFLET 6 AFLFPAA-FAP <mark>MI</mark> LKIL <mark>L</mark> NEKYAFLSVIFIAVTSLLTFQ
Efae2969 Efae 22993455	288 IFPLIAMVLAVLLOIGVLIYNSL-OFONAGER-TKYVLFYVTAMSVSIVLMKFFOLFOT 6 PLFYPAA-FAPLVLSFFLNRRAGIMAGIFOAVSALFIFY
Desu2389 Dhaf 23113918	289 WRSVAGIALIVLASIVTMTFYAHOYRKNYSDI-SKKLLLTALMMFLVLAMGKAVISLNL 9 GILIPVA-WATMTIAILV
CAC1292 Cace 15894574	260 WYTYVALAVAVI.IVI.FLOTYYTYRYYKETYRD_NKKTINC.CULNTISTILARSVGTIS_ PFLIPLA_CAPMIMILLMDSRISLFESTINCTFISLICK
CTC02023 Ctet 28204087	
CRE2010 Cror 19211001	
CFE2019_CPE1_18311001	200 BITTINGSULVITVILGUIGIIRATIFAIRE-FORTVILGUIGVU ULAREGUMUS- NITTEA-MEMITILLARATISDUFOMUNUUUUUUUU
Chte1004 Cthe 23020923	290 FAFAGGILAIILFISELLLIMMNFCKKVIIN-KIDILLESVILMILFIARWVMEIS- PLIIPIF-IATMLISELLDLKLAIMVNVLTVAISTAI
TTE09/1_Tten_2080/451	261 YALTVGFLILSSSLFLSVYYIIRLDKKIATKMYIELCCTGILYLVLAVAFKGIE- PLLIPAA-MLPHLVSILIDPYIAIMVDIIYSLLVGLMVS
Gmet0579_Gmet_23053661	320 LFTGIGLFGLVLVILYFPYRFARKNIRKFSPV-TKDLLLISLVTAGMFFLLKIALTISA 14 FYLFFFA-VGPMLIRIILNSEVAMVYTAITAPLLGIMFN
LA1683_Lint_24214383	327 FAS <mark>I</mark> VSILLI <mark>Q</mark> TVFVVIIYIFLKKYNPKRLNDVSSN <mark>VIVF</mark> SLIWVLVLSCTIASKIFFN 11 ALFVPVG-MVC <mark>LI</mark> ISFI <mark>Y</mark> DEQLSIAIGF <mark>Y</mark> LSFFIFMASH
Ddes1875_Ddes_23474958	301 APR <mark>V</mark> AGVFLT <mark>A</mark> LILAAGLL-FSHRTGSFSFKS-SKD <mark>IIFL</mark> SCVVLAFAGMAKGLDVLGA 14 PFLLPVA-SAGG <mark>L</mark> IGLI <mark>F</mark> SFRRCVTTSL <mark>L</mark> LAFICMAMMD
TP0651_Tpa1_15639638	364 SLL <mark>V</mark> SSLLLL <mark>A</mark> FLYLLAFFLFSKRMAHPPLKL-RVE <mark>LLIL</mark> YTSVAGYLCTLFLSKIAAL 6 IPFQPTA-LCI <mark>ML</mark> VTAL <mark>V</mark> SHRSAVTSSF <mark>L</mark> IAFAVLIASQ
FNV0454 Fnuc 27886299	282 FII <mark>A</mark> LNLVFL <mark>L</mark> VISSIFNVVTIKFYSKEILEK-NKYRS <mark>IM</mark> LLTIATLLVFRIVPSSM IYLLPLD-TML <b>LL</b> LLFI <mark>V</mark> RPRFSVFLTMMVISYMLPITD
Chlo0060 Caur 22969909	285 WSELLGYGLLAAILTAGISGYLHVFQPKLMTH-PRALTTLMFSIVVTLLLARLTLLVYN 2 PIFFPLA-VLA <mark>VV</mark> TTIVFTGQIGFAVSLLAAVVISVME-
Tery2872 Tery 23042199	333 WRS <mark>L</mark> VVFVCIVSASVAIVVIVGKKFQSCLRRRDH <mark>ILLL</mark> LLSLSAPLLVALRI PWTSSLP-AIG <mark>LL</mark> VGNF <mark>Y</mark> GSTMGVTLVG <mark>L</mark> ISVVLPIGMD
slr0104 Ssp 16331836	376 WOGLLRTAGLVGGALIIFCGVSRRIHRPLRRRDHILLCLLSVSTPVLFLLD PVYNNLP-AISLLTSSFYGPTLAITOVVLVGGLSAFAME
Npun2210 Npun 23125890	326 WIHLIRFSGMVGLAVAIFKLVERRIYAKLRPSDYLLILLTLSAPLLIFLT KSFTGLP-AIGLLVGSFYGTAIGGTYMVLLTILLSMGTE
alr2784 Ana 17230276	415 WIALLKIGGI VTGGTCTFALVETRSKCPLROR-DRILUTLITISTPGVLAMGV PYTTWS-AUGULIGSPYGRELSMTVIGLILFTLPMSME
Npun0472 Npun 23124089	
Bmi+0951 Bmar 23131657	
Funite0551_Fmat_25151057	$2/3$ We below standowning three clark - induce volume $\sqrt{2}$ show in $\sqrt{3}$ - induce $\sqrt{3}$ and $\sqrt{3}$ and $\sqrt{3}$
Synwhooso_Syn_23133329	202 FLINDERF LEADVACEVENDURAREKFOLDVACHARINELTVAL
Mmc12334_Msp_23000875	392 WERVLEDSTTVALF LITIGELFLEKTAAAF FRD-RETVELIGTELVISASESAVTLALGO IS TILPPVA-MUSAMASLITGARVSDPGGTWVVVTALSFLT
ydir_Cpne_1561/9/2	200 CRSLWGAIFVVLILLWGYGALKALCPEMLKS-PORFMLYLAILTLSLLWCRGTEIFCA 8 PPILPFTAVLGFFLGFPIAGFSCTFLALLYTLGSD
TM1508_Tma_15644256	32 TFEALITLIVWFALVEMSTRYYRKYWLSEVFT-YTHLALILLGSSFIGFSFPEIGPFVTPVYIPVALIELVFFSPEIAITSGFLMSLFALYRWS
BA_4976_Bant_21402351	276 FQPYVGLAVL <mark>I</mark> GVLLYFMHKQFEVFLQRKRED-RPY <mark>ILAY</mark> ITILSITIVLMKIISLFQK 10 PVAMGTILVKLMIGDRFVFLTSMIFSVCGSIMFNEGVTS
consensus/95%	hh
Secondary structure	
Secondary structure	
Secondary structure YQFF_Bsub_16079587 lin1503 Linn 16800571	
Secondary structure YQFF_Bsub_16079587 lin1503_Linn_16800571 lmo1466_Imco.16803506	
Secondary structure YQFF_Bsub_16079587 lin1503_Linn_16800571 lmo1466_Lmon_16803506 Ffam2060_Ffam200_Ffam2000_Ffam2000_Ffam200_Ffam20000_Ffam20000000_Ffam20000_Ffam2000000000000000000000000000000000000	
Secondary structure YQFF_Bsub_16079587 lin1503_Linn_16800571 lm01466_Lmon_16803506 Efae2969_Efae_22993455 perm0380_hbcf_22113010	PNVUGIYYLISGISOVLFL GKHNARSKILGTGLFVAFINWVVLSLLLIQNTALSG 2 IGTLMLMGVYSGRASSVLIIGLMPFFTGFGLS 476 STSGIAIFILLSGITSVVAL RDYSRRSAILSGTGUVGLINNVYVLSLLLIQNTALSG 2 IGTLMLMGVYSGRASSVLIIGLMPFFTGFGLS 476 ATSGITFILLSGITSVVAL RDYSRRSAILSGTGWVGLINNVYULLLLINNSTLLQ 2 TUAAIGYAFLGGGAFLGGVOIPLFTTFGLT 486 IGTNFLTIILMSVPFGLMA 3 RKRISEQGSSAAWWYIFPVMDVLLVYVGKSFG 2 ZTUAAIGYAFLGGGAFLGVVIPLFTGLT486 GONTULIDEGGGYDUNGU SI JOOGEN DDVLTAVYVGANSFG 2 ZTUAAUGYFGAFLGGGAFLGGVOIPFFTGLT 486 GONTULIDEGGGYDUNGU SI JOOGEN DDVLTAVYVGANSFG 2 ZTUAAUGYFGAFSFLTGGLRFYILLAWNDDS 486
Secondary structure YQFF Bsub 16079587 lin1503 Linn_16800571 lm01466 Lmon_16803506 Efae2965 Efae2967 Bfae2993455 Desu2389 Dhaf_23113918	FNYUIGIYYLISGISGVLFL       GKHNARSKILGTGLFVAFINWVVULSLLLIONTALSG       2 IGTLMLMGVVSGRASSVLIIGLMPFFTGFGLS       476         STSGIAIFILLSGISSVVAL       GKHNARSKILGTGLFVAFINWVVULSLLLIONTALSG       2 IGTLMLMGVVSGRASSVLIIGLMPFFTGFGLS       476         ATSGITFFILLSGATSVVAL       RDYSRRSAIMSGPMVGLINNVVULSLLINNSTLQ       2 TLMAIGYAFLGGFGAFLGVGVIPFFTIFGLIT       486         ATSGITFFILLSGATSVVML       RDYSRRSAIMLSGPMVGLINNIVVLILLLINNSTLQ       2 TLMAIGYAFLGGFGAFLGVGVIPFFTIFGLIT       486         GGNULALIALGGFGVGVSUS       SLLSQRSDLARAGUYLAFNVLTASCIALISGNGISA       YMLANUCGFTAGFFFLITGLHFTTGGFGTS       490         SUMUNDUNUG       VIDAUBADDADING       VIDAUBADDADING       VIDAUBADDADING       VIDAUBADDADING       VIDAUBADDADING
Secondary structure YQFF Bsub_16079587 lin1503_Linn_16800571 lm01466_Lmon_16803506 Efae2969_Efae_22993455 Desu2389_Dhaf_23113918 CAC1292_Cace_15894574	PNVUGIYYLISGISOVLFL GKHNARSKILGTCLFVAFINWVVUSLLLIQNTALSG 2 IGTIMLMGVYSTPASSVLIIGLMPFFTGFGLS 476 STSGIAIFILLSGITSVVAL RYSRRSAILSGTCMVGLINWVVUSLLLIQNTALSG 2 IGTIMLMGVYSTPASSVLIIGLMPFFTGFGLS 476 ATSGITFILLSGITSVVAL RYSRRSAILSGTWVGLINWVVULLLILNNSTLLQ 2 TUAAIGYAFLGGGAFLGGVOIPFFTIFGLT 486 IGTNFLTIILMSVPFGLMA 3 RKRISEQGSSAAWWIIFPVMHOVILVIVGGNSFG 2 TUAAIGYAFLGGGAFLGGVOIPFFTFGLT 486 SGWLIALLAFGGFUVSV SLLSGRSDLARGGYLAAFNVLASCIALISGGGISM HISIGLGIVGFTSGLTAGGTSF540 SWULALAFGGFUVSV SLLSGRSDLARGGYLAAFNVLASCIALISGGGISM VISGFUSSNLVD NVQKASFAFLAGTLAFFGGFIS 490 FNIEVTILAMSSVIAFMSF RKMCQRNDTIYAAFT
Secondary structure YOFF Bsub 16079587 lin1503 Linn 16800571 lm01466 Lm00-16803506 Efae2965 Efae29293455 Desu2389 Dhaf 23113918 CAC1292 Cace 15894574 CTC02022_Ctet 28204087	PNTVIGIYYLISGISGVLFL         GKHNARSKILGTGLFVAFINWVVLSLLLDNTALSG         2 IGTIMLMGVVSGFASSVLIIGLMFFFTGFGILS         476           STSGIAIFILLSGITSVVAL         GKHNARSKILGTGLFVAFINWVVLSLLLDNTALSG         2 IGTIMLMGVVSGFASSVLIIGLMFFTGFGILS         476           STSGIAIFILLSGITSVVAL         RDYSRKSAIMLSGFMVGLINNVVLILLLUNNSTLLQ         2 ILMALGYAFLGGFGAFILGVGVIPLFFTIFGLLT         486           ATSGITFILLSGATSVVML         RDYSRKSAIMLSGFMVGLINNVVLILLLUNNSTLLQ         2 ILMALGYAFLGGFGAFILGVGVIPLFFTIFGLLT         486           SGMLIALIALFGGFIGVHSV         SLLSGRSDLARAGLTLARAGUTLARVUTFSKUSSGO         2 WILMLVGCFTGLFFFLIGGLFF         486           SGMLIALIALFGGFIGVHSV         SLLSGRSDLARAGUTLARAGUTLARSFIGFLVSNLV         NVQKASFFLGLISAILTGELFFTSTIPUVT         453           FINEVEITLALMSISIIL         KKMQERNDILSVLTZSIVVILTFSAGVLISSNUT         VVKKAGFSLIGGLFTGFLFFFSSTPUVT         453
Secondary structure YQFF Bsub_16079587 lin1503_Linn_16800571 lm01466_Lmon_16803506 Efae2969_Efae_22993455 Desu2389_Dhaf_23113918 CAC1292_Cace_15894574 CTC02023_CCtet_28204087 CTP2019_Cper_18311001	FNYUIGIYYLISGISOULFL       GKHNARSKILGTCLFVAFINWVVLSLLLIQNTALSG       2 IGTHMLMGVYSTRASSVLIGLMPFFTGFGLS       476         STSGIAIFILLSGITSVVAL       RDYSRRSATMSGFMVGLINNVYULSLLIQNTALSG       2 IGTHMLMGVYSTRASSVLIGLMPFFTGFGLS       476         ATSGIAIFILLSGITSVVAL       RDYSRRSATMSGFMVGLINNVYULSLLINNSTLQ       2 ITHALGYAFLGGGAFLGGVGFLFTTFGLT       486         IGTMFLTIILMSVFSGLMA       RKRISEQGSSAAWWIIFPVMHDVILVYQGMSFG2       2 ITHALGYAFLGGGAFLGGVGFFFLTTGLT       486         SGMLIAILAFCGFURVSV       LSOSGSLAARGUSTAAWVITSVILVQGMSFG2       2 ITHALGYAFLGGGAFLGVGFFFLTGLTGLT       486         SGMLIAILAFCGFURVSV       LSOSGSLAARGUSTAAWVITFPVMHDVILVYQGMSFG2       2 ITHALGYAFLGGGAFLGVGFFFLTGLTPVITAVNDUS       486         SMILIAFCGFURVSV       LSOSGSLAARGUSTAAWVITFPVMHDVILVYQGMSFG2       2 ITHALGYAFLGGGAFLGVGFFFSVGFTTGFURVF       455         SMILVGRMDIISSVIAFMSF       RKMCQRNDIISALFTAANNITSV       NVQKASFAFIAGLSGLIGLGGFTGFLFFFSTFDVT       445         SNVERIALAILMULGFLL       KMGQRNDIISTTAUVISTINVILTFSAGVLISNNIT       ILADSTFAACALSGLIGGLIGFTIGFLFFFSTFDVT       453         SNVERIALLALMOGGLL       KMGQRNDIISTTAUVISTINVILTFSAGVLISNNIT       ILADSTFAACALSGLIGGLIGFTIGFLFFFSTFDVT       453         SNVERAGUGTL       KMGQRNDIISTTAUVISTINVILTFSAGVLISNNIT       ILADSTFAACALSGLIGGLIGFTIGFLFFF       57571174       453
Secondary structure YQFF Bsub 16079587 lin1503 Linn 16800571 lm01466 Lmon 16803506 Efae2969 Efae 22993455 Desu2389 Dhaf 23113918 CAC1292 Cace 15894574 CTC02023 Ctet 28204087 CPE2019 Cper 18311001 Chte1004 Cthe 23020923	FNYUIGIYYLISGISGVLFL         GKHNARSKILGTGLFVAFINWUVULSLLIQNTALSG         2 IGTLMLMGVYSGFASSVLIIGLMFFFTGFGILS           FNYUIGIYYLISGISGVLFL         GKHNARSKILGTGLFVAFINWUVULSLLIQNTALSG         2 IGTLMLMGVYSGFASSVLIIGLMFFFTGFGILS         476           STSGIAIFILLSGITSVVAL         RDYSRRSAINFSGFMVGLINNUYULLLINNSTLLQ         2 ITMALGYAFLGGFGAFILGYGVIPLFFTIFGLIT         486           ATSGITIFILSGATSVVML         RDYSRRSAINESGFMVGLINNUYULLLINNSTLQ         2 ITMALGYAFLGGFGAFILGYGVIPLFFTIFGLIT         486           IGTNFLTILMSVYFSGLMA         3 RKRISEGQSSAAMWUIFPYMMOVILIVYQGNSFGD         2 TMALGYAFLGGFGAFILGYGVIPLFFTIFGLIT         486           SGMLIALIALFGGFIGVHSV         SLLSQRSDLARAGUTLAFNVLTASCIALISGNGISA         YMLAUYCGFTALFFFILIGHINKYERTGFGITS         490           PNIEVIILALMSVIILALMSVINKS         KKMQERNDILYSVLTAFINVITFSIGVLISNNTID         VVQKASFFIIGULSGIFTIGFLPFFSIFDVY         453           PNPNIILALINVULGGTLL         RKMQQNDLISSTAVISSITVAVISSITTAVIANIHKYKGKEI         LLKSCALVGFRIGULSGIFTIGFLPFFSIFDVIT         453           NDFFFITMALVTGEFIT         SKMKRRUSLAGUTGSISTVALNANIHKYKGKEI         LLKSCALVGFRIGULSGIFTIGLPFF         STFDIYT
Secondary structure YQFF Bsub_16079587 1in1503_Linn_16800571 1m01466_Lmon_16803506 Efac2969_Efac_22993455 Desu2389_Dhaf_23113918 CAC1292_Cace_15894574 CTC02023_Ctet_28204087 CTC02023_Ctet_28204087 CTE2019_Cper_18311001 Chte1004_Cthe_23020923 TTE0971_Tten_72807451	
Secondary structure YQFF Bsub_16079587 lin1503 Linn_16800571 lm01466 Lmon_16803506 Efae2969_Efae_22993455 Desu2389 Dhaf_23113918 CAC1292_Cace_15894574 CTC02023 Ctet_28204087 CFE2019 Cper_18311001 Chte1004_Cthe_23020923 TTE0971_Tten_20807451 Gmet0579_Gmet_23053661	
Secondary structure YQFF Bsub_16079587 1in1503_Linn_16800571 1m01466_Lmon_16800571 Desu2389_Dhaf_23113918 CAC1292_Cace_15894574 CTC02023_Ctet_28204087 CPE2019_Cper_18311001 Chte1004_Cthe_23020923 JTE0971_Tten_7807451 Gmet0579_Gmet_23053661 LA1683_Lint_24214383	Signification           PNTVIGITYLISGISOVLFL         GKHNARSKILGTCLFØAFINWVVLSLLLQNTALSG         2 IGTIMLINGVVSLØASSVLIGLMPFFTGFGLS         476           STSGIAFFILLSGITSVVAL         RNJSRRSATLSGFØNGLINNVVVLSLLLUNNSTLQ         2 ITMALGYAFLOGGGAFLGGGAFLGGGAFTLGGUFFFTFGLT         486           ATSGITFILLSGITSVVAL         RNJSRRSATLSGFØNGLINNVVLLLLUNNSTLQ         2 ITMALGYAFLOGGAFTLGGGAFTLGGGAFTLGGGAFTLGGGAFTLGGGAFTLGGGAFTLGGUFFF         486           IGTNFLTTILNSGTSVVAL         RNJSRRSATLSGFØNGLINNVVULLLUNNSTLQ         2 ITMALGYAFLOGGAFTLGGVGAFTLGGVGAFTLGGAFTLGGAFT
Secondary structure YQFF Bsub_16079587 lin1503 Linn_16800571 lm01466 Lmon_16803506 Efae2969 Efae_22993455 Desu238° Dhaf_23113918 CAC1292 Cace_15894574 CTC02023 Ctet_28204087 CPE2019 Cper_18311001 Chte1004 Cthe_23020923 TTE0971 Tten_20807451 Gmet0379 Gmet_23053661 LA1663 Lint_24214383 Ddes1875 Ddes_23474958	
Secondary structure YQFF Bsub_16079587 lin1503_Linn_16800571 lm01466_Lmon_16800571 Desu2389_Dhaf_23113918 CAC1292_Cace_15894574 CTC02023_Ctet_28204087 CPE2019_Cper_18311001 Chte1004_Cthe_23020923 TTE0971_Tten_27807451 Gmet0579_Gmet_23053661 LA1683_Lint_24214383 Ddes1875_Ddes_23474958	
Secondary structure YQFF Bsub_16079587 lin1503 Linn_16800571 lm01666 Lmon_16803506 Efac2969 Efac_22993455 Desu2389 Dhaf_23113918 CAC1292 Cace_15894574 CTC02023 Ctet_28204087 CTE02019 Cper_18311001 Chte1004 Cthe_23020923 TTE0971 Tten_20807451 Gmet0579 Gmet_23053661 LA1683 Lint_24214383 Ddes1875 Ddes_23474958 TP0651_Tpa1_15539638 TP00514 Func 27886299	PNVUIGIYYILSGISOULFL         GKHNARSKILGTGLFVAFINWVVUSLLLIQNTALSG         2 IGTLMLMGVYSGPASSVLIIGLMPFFTGFGIS         476           STSGIAIFILLSGITSVVAL         RDYSRRSATMSGGPWGLINNVYULLLLINNSTLQ         2 ITAALGYAFLGGGAFLGGVAFILGGVIPFFTGFGIS         476           ATSGITFILLSGITSVVAL         RDYSRRSATMSGGPWGLINNVYULLLINNSTLQ         2 ITAALGYAFLGGGAFLGGVAFILGGVIPFFTGGLT         486           IGTNFLTIILMSYDFGGMA         RKRISEQGSSAAWNITPVMOULNIYUGGKSGO         2 ITAALGYAFLGGGAFLGGVAFILGGVIPFFTGGLT         486           SGNLIALLAFGGFIGVHSV         SLSQRSDLARAGYTAAFVLTASGIALSSGMISA         WLISIGLGIVMTFSSVLAVGTLHMFTGFGIS         490           PNVEITLAAINVLSISIIL         KKMQERNDIIXSVITSIVVLTSIVNULTFSAGVLISNNID         NVQKASFAFLGGLASALLTIGUPFFSITDVT         453           PNVEITLAAINVLSTSIIL         KKMQERNDIIXSVUTSTVAULSSITTVAUSSITTVAUSSITTVAUSULSNNID         NVQKASFAFLGGLASALLTIGUPFFSITDVT         453           NDFKFIYMALVTGFFSAFIV         SKAINKRNKISLGGIVYSANUVSVUVIVANINVKYGKEI         LLASSTFAAGATLSGULSGIFTGFLPFF         STFDIVT         453           NDFKFIYMALVTGFSAFIV         SKAINKRNKISLGGIVYSANUVSVUVIVANINVKYGKEI         LLASSTFAAGATLSGULSGIFTGFLPFF         STFDIVT         453           NDFKFIYMALVTGFSAFIV         SKAINKRNKISLGGUVSVINFAMALAFQTMSDSFLSL         ITMCALFALAGGLCSALVTGTIPLISTFNVIT         755           NDFKFIYMALVTGFSAFW         SKAINKRNKISLGGUVSVINFAMALAFQTMSDSFLSL<
Secondary structure YQFF Bsub_16079587 lin1503_Linn_16800571 lm01466_Lmon_16800571 Desu2389_Dhaf_23113918 CAC1292_Cace_15894574 CTC02023_Ctet_28204087 CPE2019_Cper_18311001 Chte1004_CCte_23020923 TTE0971_Tten_20807451 Gmet0579_Gmet_23053661 LA1683_Lint_24214383 Ddes1875_Ddes_23474958 FP0651_Tpal_15639638 FNV0454_Fnuc_27886299	FNYVIGIYYLISGISOVLFL       GKHNARSKILGTCLFVAFINUVVULSLLLQNTALSG       2 IGTIMLINGVYSGPASSVLIIGLMPFTTGFGLS       476         STSGIAIFILLSGITSVVAL       RNYSRRSAILSGTCLFVAFINUVVULSLLLQNTALSG       2 IGTIMLINGVYSGPASSVLIIGLMPFTTGFGLS       476         ATSGIAIFILLSGITSVVAL       RNYSRRSAILSGTCLFVAFINUVVULSLLLUNNSTLQ       2 ITMALGYAFLGGGAFLGGVGAFLGGVGAFTFGGLAF       476         GKNARSKILGTCLFVAF       RNYSRRSAILSGTCLFVAFINUVVULSLLLUNNSTLQ       2 ITMALGYAFLGGGAFLGGVGAFLGGVGAFTFGGUAFLGGVGAFTFGGVGAFLGGVGAFLGGVGAFTFGGUAFLGGVGAFTGGVGAFLGGVGAFTGGVGAFLGGVGAFTGGVGTFFTFGLLT       486         IGTNFLTIILSGITSVVAL       RNKISSRGSSAAWVITPVMOLLNIVVIGGUSGA       HILSIGLGUIGVGFGAFLGGVGAFTFGGUAF       486         SGMLIALLFCGFUGVSV       SLGSRSLARAGUITAAFNUTARSCILLSGILSUNGVGAS       HILSIGLGUIGVGFFSVLUTGUAFNTFFGFGTS       490         PNNEVILLAITINUSUTGFSAFI       RKMQORDITAAFFAATNALMSFSIGFLVSSNLVD       VVKRASFAFLGGLGSLGGLIGULSGTTGGFIS       11555777774         PNNFIILLAITINUSUGGTL       RKMQORDITAAFTAGTVSINVLVALINANNUTGFSAFLGGLGSLGGLIGULGUFFF       STFDIVT       455         NDFKFINALVTGFSAFIV       SKAINNRNELSLGCIVSINVLVALISIUNUVVALINANNUTGUSUNSUS       LKRCALVFANGINASNUTIGLGPLSTFDIVT       455         NDFKFINALVTGFSAFIV       SKAINNRNELSLGCIVSINVLVAALHAGVKSVVNPANALAFGTNSDFSLSL       ITCALFALGGLGALSATVGTIFFLTFFLSTFNUTGUS       11600000000000000000000000000000000000
Secondary structure YQFF Bsub_16079587 lin1503_Linn_16800571 lm01466_Lmon_16803506 Efac2969_Efac_22993455 Desu2389_Dhaf_3113918 CAC1292_Cace_15894574 CTC02023_Ctet_28204087 CTE2019_Cper_18311001 Chte1004_Cthe_23020923 TTE0971_Tten_20807451 Gmet0579_Gmet_23053661 LA1683_Lint_2214383 Ddes1875_Ddes_23474958 TP0651_Tpal_15539638 FPW0454_Fnuc_27886299 Chl00060_Caur_22969909 Parv2872_Terv_23042199	FNVUIGIYYILSGISOULFL         GKHNARSKILQTGEFVAFINWVVUSLLLIQNTALSG         2 IGTHMLMGVYSERASSULIGLMPFFTGFGLS         476           STSGIAIFILLSGITSVVAL         RDYSRRSAILSGOFWOLINNVYULLLLINNSTLQ         2 ITAALGYAFLGGGAFLGGVAFLIGVUFFFTGELT         486           ATSGITFILLSGITSVVAL         RDYSRRSAILSGOFWOLINNVYULLLINNSTLQ         2 ITAALGYAFLGGGAFLGGVAFLIGVUFFFTGELT         486           IGTMFLTIILMSYDFGIMA         RKRISEGOGSSAAWWITFPVMODVILVYUGGNSFO2         2 ITAALGYAFLGGGAFLGGVAFLITGULPFTFGLIT         486           SGWLIALLAFGGFIGVNS         SLLSGRSDLARGEYIAAFNUTASCIALISGMGIA         WLISIGLGVNGTFSSULJUGULPFTSVJAVGTLMMFTGFGIS         490           PNVEITLAAINSVIFFNS         RKKGRRDDIIASVITSTVVLTSIVNVIFFSAGFLUSNNID         NVQKASFAFLGGLSGITGFLFFFSSTDVT         453           PNDFFIYMALVTGFFSAFIV         SKANKRNKISLGGIVSAINVUVVSULSNNID         NVQKASFAFLGGLGSLIGGIFTGFLFFFSSTDVT         453           NDFKFIYMALVTGFFSAFIV         SKANKRNKISLGGIVSAINVUVVSUVSULSNNID         NVQKASFAFLGGLGSLIGGUTGFLFFF         5770171453           NDFKFIYMALVTGFFSAFIV         SKANKRNKISLGGIVSANVUVSUVSUVSUVSUVSUVSUVSUSUSUSUSUSUSUSUSU
Secondary structure YQFF Bsub_16079587 lin1503_Linn_16800571 lm01466_Lmon_16800571 Dsu2389_Dhaf_23113918 CAC1292_Cace_15894574 CTC20203_Ctet_28204087 CPE2019_Cper_18311001 Chte1004_Cthe_23020923 TTE0971_Tten_20807451 Gmet0579_Gmet_23053661 LA1683_Lint_24214383 Ddss1875_Ddss_23474958 PT0651_Tpal_15639638 FNV0454_Fnuc_27886299 Chlo0060_Caur_22669909 Tery2872_Tery_23042199 S1=0104_5ss_16331836	FNYVIGIYYLISGISOVLFL       GKHNARSKILGTCLFØAFINWVVLSLLLIQNTALSG       2 IGTIMLINGVYSLØASSVLIGLMPFTGFGLS       476         STSGIAFFILLSGITSVVAL       ROYSRRSATLSGEMVGLINNVYVLSLLLIQNTALSG       2 IGTIMLINGVYSLØASSVLIGLMPFTGFGLS       476         ATSGIAFFILLSGITSVVAL       ROYSRRSATLSGEMVGLINNVYVLSLLLIQNTALSG       2 IGTIMLINGVYSLØASSVLIGGUPFTTFGLT       486         IGTNFLTTILLSGITSVVAL       ROYSRRSATLSGEMVGLINNVYULSLLINNTSLQ       2 THAALGYRFLOGGAFILGGVIFFTFTGLT       486         IGTNFLTTILLSGITSVVAL       ROYSRRSATLSGEMVGLINNVYULSLLINNTSLQ       2 THAALGYRFLOGGAFILGGVIFFTFTGLLT       486         SGMLIALTAFCGFIONSV       SLØSRDLARAGGVITAAPNUTASCILLSGILSVALVUGGNSG       2 THAALGYRFLOGGAFILGGVISCILVAUVGTHFFTGFGIST       490         PNNEVTILALINNVGGFLL       REMCORNDITAAFFAINNALKSFSIGFLVSSNLVD       NVKRASFAFIAGLIGSLGGTIGULSGILFTGFLPFTFSTDIVT       455         PNVFIILLAITINUSGTLI       REMCORNDITAAFFAINNALKSSILFSVGULSSNLVD       NVKRASFAFIAGLIGSLGGLIGLIGULGGTIFTGFLPTFTSFDUVT       455         NDFKFINALVTGTSAFI       SKAIKRNRISLAGTIVSINVLVAININKTGKEI       LLABCSTAAKGALLSGILTIGULPFFSTDIVT       455         NDFKFINALGVSCUVAGGVAGKK       RRACRDSFKGUVSVVNPANALAFGTNSDFLSL       ITCALFALGGLGSLVTGTIFLITVFHVT       57         SGTUNHFFGLSSVNGAWAVI       RRACRDSFKGUVSVVNPANALAFGTNSDFLSL       ITCALFALGGLGSLVTGTIFLITVFHVT       51         FNTEFTIGLLSSVNGAWAVI       RRACRDSFKGUVSVVNPANALAFGTNSDFLSL       IT
Secondary structure YQFF Bsub_16079587 lin1503_Linn_16800571 lm01466_Lmon_16803506 Efae2969_Efae_22993455 Desu2389_Dhaf_23113918 CAC1292_Cace_15894574 CTC02023_Ctet_28204087 CTE2019_Cper_18311001 Chte1004_Cthe_23020923 TTE0971_Tten_20807451 Gmet0579_Gmet_23053661 LA1683_Lint_2241383 Ddes1875_Ddes_23474958 TP0651_Tpal_1563963 Dd0512_Tpal_1563963 PN00454_Fnuc_27886299 Chl00066_Caur_22969909 Dery2872_Tery_23042199 slr0104_Ssp_16331836 Nun2210 Nunn_23125890	GRHNARSKILGTGEFVAFINNUVULSLLIQNTALSG 2       IGTMLINGUVSBPASSVLIIGLMPFTGFGLS 476         STSGIAIFILLSGITSVVAL       RUSRRSAILSGOFVALINNUVULSLLIQNTALSG 2       IGTMLINGUVSBPASSVLIIGLMPFTGFGLS 476         ATSGIIFILLSGITSVVAL       RUSRRSAILSGOFVALINNUVULSLLINNNSTLQ 2       ITMAIGYAFLGGGAFLGGVAFLIGUVSPTFTGELT 486         IGTMFLTIILMSUPSGLMA 3       RRKISEGGSSAAWWITFPVMDVILVIULLINNNSTLQ 2       ITMAIGYAFLGGVAFLIGUVSPTFTTGELT 486         SGWLIALIALFGGFUVSV       SLSGRSDLARGEYLAAFNVLASCILSISGIGISSA       VILSIGLGUVGFTGALFSFLITGLAPFTIFGLT 486         SGWLIALIALFGGFUVSV       SLSGRSDLARGEYLAAFNVLASCILSISGIGISSA       VILSIGLGUVGFTGALFSFLITGLAPFTIFGLT 486         SWLINGUNDTITALAFTANVLASCILSISTI       KKMQGRNDTITALAFTANVLASCILSISNITD       NVQKASFAFLAGTLGULGUVGTFSFLUTGULPFF         FNNEITILALINVLSISIIL       KKMQGRNDTITALGTVALFSVGLTANNULVAAININKTGMEI       ILADSTFAAAGALISGLIJGUPFTISTDIVT 453         NDFKFIYMALVTGFFSAFIV       SKAIKRNRLSLAGTUVSAUVVEVVLUSULSNITD       NVQKASFAFLAGLIGGUCLSAIVTGULPFF         SGDVVVGGLGGUVAAGV       RRCKDRPKRGUVGUUSSUNVENAALAFGMSDEFLSL       ILADSTFAAGALISGLUCLSAIVTGUPFTISTTVTVFVT 521         YNPESFMGGVGCUVSSGG       RRAGTRQEVVVSLLPLAGGULAATLAFFFHATD       FGGLMLASANAVLSLLIFALSGVAASAC         GGTDVMLFFFLSSUNGANV       RRAGTRQEVVVSLLPLAGGULATTIFEFFKLSQQAIFTL       FGLGLMLASANAVLSLLIFALATLAFFFHATD       FGLGLAMSANAULSLLIFALAGPLIFT 475         GGTDVMLFFFLSAUGGVGUVV       RNMSSRLDIVKSSVLAVAQPFLAATLHFFFHATD <t< td=""></t<>
Secondary structure YQFF Bsub_16079587 lin1503_Linn_16800571 lm01466_Lmon_16803506 Efae2969_Efae_22993455 Desu2389_Dhaf_23113918 CAC1292_Cacc_15894574 CTC20203_Ctet_28204087 CPE2019_Cper_18311001 Chte1004_Cthe_23020923 TTE0971_Tten_20807451 Gmet0579_Gmet_23053661 LA1683_Lint_24214383 Ddes1875_Ddes_23474958 PM0454_Fnuc_27886299 Chl0006_Caur_22969909 Tery2872_Tery_23042199 S1c0104_Ssp_16331836 Npun2210_Npun_23125890	FNYVIGIYYLSGISOVLFL       GKHNARSKILGTCLFVAFINWVVLSLLLQNTALSG       2 IGTLMLMGVYSGPASSVLIGGLMPFTGFGLS       476         STSGIAIFILLSGITSVVAL       RNYSRRSAILSGCFVVULTULLUNNSTLQ       2 ITMAIGYAFLOGGGAFLGGVGAFLGGVGAFLGGVGAFTGGVGFF       476         ATSGITFILLSGITSVVAL       RNYSRRSAILSGCFVVULTULLUNNSTLQ       2 ITMAIGYAFLGGGAFLGGVGAFLGGVGAFTGGVGFF       476         GINTFITILLSGITSVVAL       RNYSRRSAILSGCFVVULTULLUNNSTLQ       2 ITMAIGYAFLGGGAFLGGVGAFLGGVGAFTGGVGFF       476         GSGMLIALTGFGFFVVUL       RNSRRSAILSGGVGVAL       RNKTSEQGSSAAWVIFPVMOULDIVYUGGNSG2       2 ITMAIGYAFLGGGAFLGGVGVFFFTFFGGLS       486         SGMLIALTGFGFFVUSS       SLSGRSDLARGGTVAFNFFF       RKMCONDTIVAAFTAINALMSFFIGFLVSSNLVD       VVKRASFAFLGGLGGLGGLGALUNGVGTHFFTFGFGTS       490         PNVETILLAINVUGGTL       RKMCONDTIVAAFTAINALMSFFIGFLVSSNLVD       VVKRASFAFLGGLGGLGGLGALGGLGSTTGGVFFFTSFDTVT       455         NDFKFINALVGGTLGALGGLGALK       SSRQNDPVKAGCIVSAINNINVKISUVUSLLNSNDTV       VVKRASFAFLGGLGGLGALGSTTGGVFFFTSFDTVT       455         NDFKFINALGVGUVSASG       RKKCRNELSGGTVGSAINTAGVGUVSAGCIVSASG       RKKCRNELSGGVGVAKGCIVSASG       VKKAGFTFAGLGGLGALMSVT       500         SGTDVMLFFELSSVMGAVI       RRAGTRGEVVNSLVAGGVGVAKGCIVSASGTALALAFGVFFUNT       15 FKLVULLINGAAGALGGLGALA       15 FKLVULLINGAGVGUVAGGVVXISTUSGVGUVAGV       15 FKLVULLINGAGVGUVAGUSGVVX       15 FKLVULLINGAGGUGLGSALVGGVFFUNT       15 FKLVULLINGAGGULGSALVGGVFFUNT       15 FKLVULLING
Secondary structure YQFF Bsub_16079587 lin1503_Linn_16800571 lm01466_Lmon_16803506 Efae2969_Efae_22993455 Desu2389_Dhaf_23113918 CAC1292_Cace_15894574 CTC02023_Ctet_28204087 CTE2019_Cper_18311001 Chte1004_cthe_23020923 JTE0971_Tten_20807451 Gmet0579_Gmet_23053661 LA1683_Lint_2241383 pdes1875_Ddes_234749588 TP0651_Tpal_156398 TP0051_Tpal_156398 TP0051_Tpal_156398 TP0051_Tery_23042199 S1C0104_Ssp_16331836 Npun2210_Npun_23125890 alr2784_Ana_17230276	
Secondary structure YQFF Bsub_16079587 lin1503_Lin_16800571 lm01466_Lmon_16803506 Efae2969_Efae_22993455 Dsau2389_Dhaf_23113918 CAC1292_Cace_15894574 CTC20203_Ctet_28204087 CFE2019_Cper_18311001 Chte1004_Cthe_23020923 TTE0971_Tten_20807451 Gmet0579_Gmet_230536638 PM0451_Dales_23474958 Dd681875_Dd6s_23474958 Df051_Tpal_15539638 FNV0454_Fnuc_27886299 Chi00060_Caur_22869909 Tery2872_Tery_23042199 Slc1014_Spg_16331836 Npun2210_Npun_23125890 alc2784_Ana_17230276 Npun0472_Npun_23124887	FNYVIGIYYLSGISOVLFL       GKHNARSKILGTCLFVAFINWVVLSLLLQNTALSG       2 IGTANLMGVYSGPASSVLIGGLMPFTGFGLS       476         STSGIAIFILLSGITSVVAL       RYSRRSATLSGCMWGLINNVYVLSLLLUNNSTLQ       2 THAAIGYAFLGGGAFLGGVGAFLGGVGAFTGFGLS       476         ATSGIAIFILLSGITSVVAL       RYSRRSATLSGCMWGLINNVYLLLLUNNSTLQ       2 THAAIGYAFLGGGAFLGGVGAFLGGVGAFTGGVGAFLGGVGAFTGGAFTG
Secondary structure YQFF Bsub_16079587 lin1503_Linn_16800571 lm01466_Lmon_16803506 Efae2969_Efae_22993455 Desu2389_Dhaf_23113918 CAC1292_Cace_15894574 CTC02023_Ctet_28204087 CTE2019_Cper_18311001 Chte1004_cthe_23020923 JTE0971_Tten_20807451 Gmet0579_Gmet_23053661 LA1683_Lint_24214383 Ddes1875_Ddes_234749588 TP0651_Tpal_156398 TP0051_Tpal_156398 TP0051_Tpal_156398 TP0051_Tery_23042199 Slr0104_Ssp_16331836 Nyun271_Cmyn_23125890 alr2784_Ana_17230276 Nyun271_Nyun_23124089 Pmit0951_Pmar_23124089	GRHNARSKILGTGLFVAFINNUVULSLLLQNTALSG       2 IGTMLINGVYSGPASSVLIIGLMPFTGFGILS       476         STSGIAIFILLSGITSVVAL       RNJSRRSATHSGPMVGLINNUVULSLLLQNTALSG       2 IGTMLINGVYSGPASSVLIIGLMPFTGFGILS       476         ATSGIAIFILLSGITSVVAL       RNJSRRSATHSGPMVGLINNUVULSLLLUNNSTLQ       2 ITMAIGYAFLGGGAFLGGVAFLIGVGVIPFTTIFGLI       486         IGTMFLTIILNSGITSVVAL       RNJSRRSATHSGPMVGLINNUVULSLLUNNSTLQ       2 ITMAIGYAFLGGGAFLGGVAFLIGVGVIPFTTIFGLI       486         IGTMFLTIILSGITSVVAL       RNJSRRSATHSGPMVGLINNUVULSLULUNNSTLQ       2 ITMAIGYAFLGGVAFLGVGVIPFTTIFGLI       486         SGMLIALAFCGFURVS       SLSGPSDLARGCYLAPAVUTAFNJARACILSSICIAUGYAFUNTAFNGT       NVGKASFAFIAGILSGLTIGVFFTSTDVIV       445         PNVEITILALNVLSISIIL       KKMQERNDILSVLYISINNILTSSVITAFNNTAFNOTTINNUVKAGFSLGSLGGLTIGVFFTSTDVIV       453       NVGKASFAFIAGILSGLTIGVFFTSTDVIV       453         NDFKFIYMALVTGFFSAFIV       SKANKRNRLSLAGTUSANVLVVSULLISNNTD       NVGKASFAFIAGALSGLIGGUTGVFTSTDVIV       453         NDFKFIYMALVTGFFSAFIV       SKANKRNRLSLAGTUSANVLVVSUVSULSNNTD       NVGKASFAFIAGALSGLIGGUTGVFTTUSTTFDIT       453         NDFKFIYMALVTGFFSAFIV       SKANKRNRLSLAGTUSANVLVVSUVSUVSUVSUVSUVSUS       SKANKRNRLSLAGTUSANVLVVSUVSUVSUVSUSUNSUSSUT       NVGKASFAFIAGALSGLIGGUTGUTGVFTTUSTTFTTVTTVTT       453         NDFKFIYMALVTGFFSAFIV       SKANKRNRLSLAGTUSANVUVSUVSUVSUVSUVSUSUS       SKANKRNRLSLAGTUSANVUVSUSUNSUN       154       1
Secondary structure YQFF Bsub_16079587 lin1503_Linn_16800571 lm01466_Lmon_16803506 Efae2969_Efae_22993455 Desu2389_Dhaf_23113918 CAC1292_Cace_15894574 CTC20203_Ctet_28204087 CPE2019_Cper_18311001 Chte1004_Cthe_23020923 TTE0971_Tten_70807451 Gmet0579_Gmet_230536638 PM0451_Dales_23474958 PM0454_Fnuc_27886299 Chi0066_Caur_22969909 Tery2872_Tery_23042199 Slc1044_Spg_16331836 Npun2210_Npun_23125890 alt2784_Ana_17230276 Npun0472_Npun_23124089 Pmit0951_Pmar_23131657 Shum06565grg_2131332	FNYVIGIYYLSGISOVLFL       GKHNARSKILGTCLFVAFINWVVLSLLLQNTALSG       2 IGTHNLMGVYSGPASSVLIGGLMPFTGFGLS       476         STSGIAIFILLSGITSVVAL       RNYSRRSAILSGCHFVGLINNVYVLSLLLQNTALSG       2 IGTHNLMGVYSGPASSVLIGGLMPFTGFGLS       476         ATSGIAIFILLSGITSVVAL       RNYSRRSAILSGCHFVGLINNVYVLSLLLUNNSTLQ       2 ITMALGYAFLGGGAFLGGVGIPTFTGLT       486         IGTNFLTIILNSGITSVVAL       RNYSRRSAILSGCHVGLINNVYULSLLUNNSTLQ       2 ITMALGYAFLGGGAFLGGVGIPTFTGLT       486         IGTNFLTIILSGITSVVAL       RNKSRRSAILSGCHVGLINNVYULSLLUNNSTLQ       2 ITMALGYAFLGGGAFLGGVGIPTFTGLT       486         SGMLIALFGGFGIVSVS       SLSGRSDLARGGIVGASNAWUIFPVMBULLVYQGMSFG2       2 ITMALGYAFLGGGAFLGVGVIPTFTGLT       486         FNVEITILAINVGGGLL       RKMCORDDIVAAFTALNALMSFSIGFLVSSNLVD       NVKRASFAFLGGLGGLGGLISALVGVTHFTGFGST       490         FNVEITILAINVUGGLL       RKMCORDDIVASCLIVSSNLVD       VVKRASFAFLGGLGGLGGLGALLSGTUFFFSTDIVT       455         NDFFTIVALVGGFLGGUGAAHK       RKMCORDDIVASCLIVSSNLVD       VVKRASFAFLGGLGGLGGLGALSTIGFFFSTDIVT       455         NDFFTIVALVGGRSSAGN       RRCKDRSRLTTAGVKNVNPAMALAPGYNDDIVS       VVKRASFAFLGGLGGLGCALSTIGFFFSTDIVT       456         NDFFTIFFALSGVSNGAWY       RRCKDRSRLTTAGVKVSVNPAMALAPGYNDDIFS       VLKSGLMGINNTYSILLVGLLPAFTTSTDIVT       517         NDFFTIFFALSGVSNGAWA       RRAGTRGEVKVSVLVAGVCLANGFTLATTHFFFAGTLFFFNSGESSGGAR       1157       110100000000000000000000
Secondary structure YQFF Bsub_16079587 lin1503_Linn_16800571 lm01466_Lmon_16803506 Efac2969_Efac_22993455 Desu2389_Dhaf_23113918 CAC1292_Cace_15894574 CTC02023_Ctet_28204087 CFB2019_Cper_18311001 Chte1004_Cthe_23020923 JTE0971_Tten_70807451 Gmet0579_Gmet_23053661 LA1683_Lit_22414383 Ddes1875_Ddes_23474958 TP0551_Tpal_15639638 TP0551_Tpal_15639638 TP0054_Fnuc_27886299 Chl00066_Caur_22060999 Ch272872_Tery_23042199 slr0104_Ssp_16331836 Npun210_Npun_23125890 alr2784_Ana_17230276 Npun0470_Npun_23124089 Pmit0951_Pmar_23131657	
Secondary structure YQFF Bsub_16079587 lin1503_Lin_16800571 lm01466_Lmon_16803506 Efae2969_Efae_22993455 Desu2389_Dhaf_23113918 CAC1292_Cace_15894574 CTC20203_Ctet_28204087 CPE2019_Cper_18311001 Chte1004_Cthe_23020923 TTE0971_Tten_70807451 Gmet0579_Gmet_230536638 PM0451_Dpal_15639638 PM0454_Fnuc_27886299 Chi0066_Caur_22969909 Tery2872_Tery_23042199 Slc1044_Spg_1631836 Npun2210_Npun_231258900 alt2784_Ana_17230276 Npun0472_Npun_23124089 Pmit0951_Pman_23131657 SymWh056_Sym_2313329 Mmc12334_Msp_23000675	MAINTERSTORMENT       GRHNARSKILDTGLFVAFINNVVLSLLLQNTALSG 2       IGTHNLMGVYSTRASSVLIGLMPFTGFGLS 476         STSGIAIFILLSGITSVVAL       RNSRSAILSGIGVFU       GRHNARSKILDTGLFVAFINNVVLSLLLQNTALSG 2       IGTHNLMGVYSTRASSVLIGUGVFFTGFGLS 476         ATSGITFILLSGITSVVAL       RNSRSAILSGIGVFU       RNSRSAILSGIGVFU       RNSRSAILSGIGVFU       476         ATSGITFILLSGITSVVAL       RNSRSAILSGIGVGULINNVVLILLLINNSTLQ 2       THAAIGYAFLGGGAFLGGVGIFFTFTGLIF 486         IGTNFLTIILMSVFSGLMA 3       RKKISEQGSSAAWWITFPVMDVLLVYQGMSFG 2       THALGYAFLGGVGAFLGGVGVFFTTGGLFFT       486         SGMLIALFGGFGIGVSV       SLSGRSDLARGGIGVAFLAFNVLTARSCILLSGILSGIGVGFFGGTS 400       HISSIGLGUNGTSSVLAVGVTHFTGFGST 490       486         FNVETILLAINVLGGLL       RKMQORDTIXAAFVLTASSULFSKOVLVSNUD       VVKRASFAFLGGLGGLGGLGALLSGILVGFFFSTDIVT 455       FNVETITGLFFFSTDIVT 455         NOFFTIMAUTGGFFAFVS       RKMQORDDILYSSTVAVLSSULFSVOULANNINKNUD       VVKRASFAFLGGLGGLGGLGALLSGIVSTVUTGLLPFF       STFDIVT 455         NOFFTIMAUTGGFSAGA       RRCDRSRITTAGVKSVVNPAMALAPGMSDFLSL       ILKEGLVFRGULSKATUGFFFITTGFLFFF       517         SGTDVMLFFLALSGVSGUNAGY       RRAGTRGEVVNSLVAGGVGUNAGY       VLKSGLMGTINFFSILVIGLLPFF       STFDIVT 455         YNDKTSFRGAVSCUNAGY       RRAGTRGEVVNSLVGVGUNAGY       SARQRADFVKAGLTVSANNLVSIVGULLSNDIVS       VLKSGLMGINTFFSILLSVMALAPG       AIFLIGVGUNAGGUNAGGV       YLKSGLMAGILSGUNAGGNAG       YLKSGLMAGILSGUNAGGNAG       KKAGLDS
Secondary structure YQFF Bsub_16079587 1in1503_Linn_16800571 1m01466_Lmon_16803506 Efac2969_Efac_22993455 Desu2389_Dhaf_23113918 CAC1292_Cace_15894574 CTC02023_Ctet_28204087 CFB2019_Cper_18311001 Chte1004_Cthe_23020923 TE0971_Tten_20807451 Gmet0579_Gmet_23053661 LA1683_Lint_2214383 Dd651875_Dd6s_23474958 FN00454_Fnuc_27886299 Chlo0060_Caur_2296909 Tery2872_Tery_23042199 slr0104_Ssp_16331836 Npun210_Npun_23125890 alr2784_Ana_17230276 Npun0471_Npun_23125890 alr2784_Ana_17230276 Npun0471_Npun_23124089 Pmit0951_Pmar_23131529 SymMo655_Syn_2313229 Mmc12334_Msp_23000675 yq4F_Cpne_15617972 TM1506_Tma_15644256	FNYVIGIYYLISGISOULFL       GKHNARSKILQTCLFVAFINUVVULSLLLQNTALSG       2 IGTMULMGVYSGRASSVLIIGLMFFTGFGLS       476         STSGIAIFILLSGITSVVAL       RNSRRSATLSGOTMUCLINNVYULSLLLUNNSTLQ       2 ITMAIGYAFLOGGGAFLGUGVIFFTTGLI 486         ATSGITFILLSGITSVVAL       RNSRRSATLSGOTMUCLINNVYULSLLLUNNSTLQ       2 ITMAIGYAFLOGGAFLGUGVIFFTTGLI 486         IGTMFLTIILMSGITSVVAL       RNSRRSATLSGOTMUCLINNVYULSLLLUNNSTLQ       2 ITMAIGYAFLOGGAFLGUGVIFFTTGLI 486         IGTMFLTIILMSVFSGLMA       RKKISEQOSSAAWVIFPVMOULINVYUGULSUNTIQ       2 ITMAIGYAFLOGGAFLGUGVIFFTGLUFF         SOMIALIAPCGFIGUSSV       SLSORSDLARAGUTAPHVITAPVUTALSULLINNTIG       41151GLIGHTGFSFLUTGLFFF         FNVEITLAAINVISISIL       KKMQRNDTINALFTANUNISFSIGFUSSNUD       VVKKAGFSLIGELSGIFTIGLPFF       51FDVT 445         NDFKFIYMALVTGTFSAFIV       SKANKRNRLSLAGITSINVLVAAININKYKGKEI       LLADSTFAAGALSGILGULSGIFTIGUPFF       51FDVT 455         NDFKFIYMALVTGTFSAFIV       SKANKRNRLSLAGITSINVLVVAAININKKGKEI       LLADSTFAAGALSGILGULSGITTGUPFF       51FDVT 455         NDFKFIYMALVTGTFSAFIV       SKANKRNRLSLAGITSSGVIPSIKNUV       STICLALFAAGALSGILCSA TUTGUPFF       51FDVT 455         NDFKFIYMALVTGTFSAFIV       SKANKRNRLSLAGITSSAVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV
Secondary structure YQFF Bsub_16079587 lin1503_Lin_16800571 lm01466_Lmon_16803506 Efae2969_Efae_22993455 Dsau2389_Dhaf_23113918 CAC1292_Cace_15894574 CTC20203_Ctet_28204087 CPE2019_Cper_18311001 Chte1004_Cthe_23020923 TTE0971_Tten_70807451 LA1683_Lint_24214383 Dd0s1875_Dd0se_23474958 PM0515_Tpal_15539638 FNV0454_Fnuc_27886299 Chc00066_Caur_22969909 Tery2872_Tery_23042199 S1c0104_Ssp_16331836 Npun2210_Npun_23125890 01c12784_Ana_17230276 Npun0472_Npun_23124089 Pmi00951_Fnar2_3131657 Synwh0656_Syn_2313329 Mmc1233_Msp_23000675 YqfF_Cne_15617972 TM1508_Tma_15644256 Ba_4976_Bant_21402351	FNYVIGIYYLISGISOVLFL       GKHNARSKILQTCFYAFINWVVLSLLLQNTALSG       2 IGTHALMOVYSTRASSVLIGLMPFTTFGELS       476         STSGIAIFILLSGITSVVAL       RNYSRSATHSGPMVGLINNVYVLSLLLQNTALSG       2 IGTHALMOVYSTRASSVLIGUVFFTTFGLT       486         ATSGITFILLSGITSVVAL       RNYSRSATHSGPMVGLINNVYULLLLNNNSTLQ       2 ITMALGYAFLGGGAFLGGVGIPTFTTFGLT       486         IGTNFLTIIMSVFSGLMA       RKKISEQGSSAAWWITFYMMULLLINNNSTLQ       2 ITMALGYAFLGGGAFLGGVGIPTFTFGLT       486         SGMLIALTGFGFTOVINS       SLSGRSDLARGGTVARNVILTPVMMULLLINNSTLQ       2 ITMALGYAFLGGGAFLGGVGIPTFTFGLT       486         SGMLIALTGFGFTOVINS       SLSGRSDLARGGTVARNVITSINVULLLINNSTLQ       2 ITMALGYAFLGGGAFLGGVGVIPTFTFGLT       486         SGMLIALTGFGFTOVINS       SLSGRSDLARGGTVARNVITSINVULLLINNSTLGL       2 ITMALGYAFLGGAFLGGLISGITGGLISGTTGFFTFTFGGTS       490         FNNFIILLAITINGSVIAGGYL       RKMQORNDILYSSTVAVLSSILFFSKOTLTNNFME       HISSIGLGGLISGTUGFFTFTGFFTFTSTSTVIT       455         NOFFTIFNALOGGIGAIK       SSAKQRNDPVKAGLYSAVNINSINGVGULSNDIVS       VVKKAGLYFARUMKYTIGLFFT       55771445         NOFFTIFNELOGGIGAIK       RRAGTRGEVVKSUVSAGG       SSAKQRNDPVKAGLYSAVNINSINGVGULSNDIVS       VKKSGLARGFLFTFTFFTFTFTFTFTFTFTFTFTFTFTFTFTFTFTFT

**Multiple sequence alignment of the 7TM domains of the 7TMR-HD family.** Multiple sequence alignment the 7TM domains of the 7TMR-HD family was constructed was constructed as detailed in the legend to Figure 1. The 95% consensus follows the same convention as in Figure 1 and 2 and also shows the following classes alcohol (o: ST, Blue), the tiny subclass of small (u; GAS, Green shading) and an 'E' shows the completely conserved amino acid in that group. The species abbreviations are as shown in Table 1.

membrane-spanning domains with RPS-BLAST, suggests that the 7TM domains of the 7TMR-HDs are distant relatives of other eukaryotic and bacterial 7M receptors (e-values ~.01–.07). However, the 7TMR-HDs possess several distinctive features in their 7TM domain that clearly demarcate them from all other membrane proteins. These include, positively charged patches in the intracellular loops between helix 1–2 and helix 5–6, a glycine in helix 7, and conserved glutamate and alcoholic residues in the C-terminal cytoplasmic tail (Figure 6).

In terms of domain architecture, the 7TMR-HDs are most similar to the 7TMR-DISMs in possessing a large N-terminal extracellular domain and an intracellular signaling domain. This architectural pattern suggests that the extracellular domain (7TMR-HDED, for 7TMR-HD Extracellular Domain) is most likely to function as a sensor domain that transmits the signal via the 7TM domain to the intracellular catalytic domain. Sequence profile searches with the extracellular domain did not establish any relationship with previously known globular domains. A multiple

Secondary structure		ЕЕНИНИНИНИНИНИНИЕЕНИНИНИНИ	
YQFF Bsub 16079587	51	LDLFSVSDKTIYAPATVEDQKATEEKKQAAEDA-VEDQYTLKKEYTDNRIDL-VSSIFDSISEVKKS-SEEGSK 14 LTSDVNDSISEDS-IKTLLKADSEDFSFVRDTVITA	
lin1503 Linn 16800571	46	VKLFQVAEKT <mark>IRS</mark> PQTVEDTEKTKEERTK <mark>A</mark> SNA-VEDVYVYNRETGQNRVAL- <mark>I</mark> QSLFAYVIEVNSE-AEENDK 33 VSEKITSNISDEV-FTTLIEADSKDFNVMEDVITTF	
lmo1466 Lmon 16803506	46	VKLFQVAEKT <mark>IRS</mark> PQTVEDTEKTKEERTK <mark>A</mark> SDA-VEDVYVYNRETGQNRVAL- <mark>I</mark> QSLFAYVNEVNAE-AQEKDT 33 VSEKITSNISDEV-FTTLIEAKSKDFNVMEDVVTTF	
Efae2969 Efae 22993455	44	FREGQVAEESIRANKTIENTEETEQKRKLAAEA-VTPEYTYQKDLADDQNNR-IKQLFELIDKTNDAINKSYDE 32 ENVAFYQKLPNNF-YSTIFEMTETEIHTVRDESLKI	
Desu2389 Dhaf 23113918	64	LKLGDPSPQLVTAPYEKNIEDLKKYYQDQEAAAAEA-VKPVYTQDEEYLTSISRD-LSTAFTALEEAIAS-DEDKAD 10 FNVLPQEALEGLL-DSSLDLLEEKEQKATEVILVNA	
CAC1292 Cace 15894574	40	LKEGDIAKNDIKATRDVNDEAATEERRKQAVNS-VGIQYDKNTEIINNIIDN-INNDFTIMNKVKDE-NSDDKT 4 LKSSLKTDLDDSN-LSVILSMNKEDLKDLQOFIIKT	
CTC02023 Ctet 28204087	44	LMEGEIANVDIKATREVVDDIATKARVKQAIES-VPLQYNKKPEVKNNVLED-LDKLFNKIPSVVDV-NISEKD 4 LREGNNIKLSDED-YMEFLKLDKDNLDKLKVVLVKT	
CPE2019 Cper 18311001	43	LKVGDIPKSDIKAHREIIDESATEARKKEAEEK-VDKOVSLRTDVOKOSEEK-IKGFFSSVEKVLAQ-DKPEEE 3 LIPRAPFKITDAQ-ANKIASMNEOSTKKLESVCIDG	
Chte1004 Cthe 23020923	50	VTLGAISEYDIISPRDIVNTVKTEENAKKAASO-VSPVMRDIPNAPIEVINL-VOKLFFLINDAONT-YKSKIS 25 SIGELGIKLEDAOIDYLISNAGEEDINSLEVVIRSK	
TTE0971 Tten 20807451	44	LKAGDVAPODIKAPKDVIDVIATOKKIOEAVNA-VNPKYDYNENIAKESYLK-LADFFNKLREVRKS-SESEEO 4 FKNITPINLEDKD-IAFLLKIDDSTLIKMETVILS2	
Gmet0579 Gmet 23053661	67	YKAGDIATTDIRATODYLLEDRILLTEKKRNEAEAA-VPYVYAFSLDGAGELVGR-VEOALALLOOAATA-EPSOGV 46 IVVDKRSFAGDTAHGIVLVDARTGESLGKMEYTV	
LA1683 Lint 24214383	65	VSEGKTAPEKIISAKEIVYEDEDKTKAKKLTAYOS-APFVFDRDFGVLOEOINNFTOEDMENFRSFKPS-AEGRIY 37 SNLVFSSFCILRDLPPDYAALKTSGARVRNOGIKEC	
Ddes1875 Ddes 23474958	53	VLAGEVAERDIOADOELRVEDAAATEARRNEVART-OPPLFDLSTATIDVLRAN-VNATLESLRNLSADRHETHRO 36 TRLASGVVADORILLOERNGIIVRDLSSGSEVLRET	
TP0651 Tpal 15639638	47	FAVGRVAERNVVSPVFISLTDELATOROXARRKKE-IPAVFERRIDLEOAEVRA-YOAFCHALOKARVGMALDTSS 107 OASAGDLPSASVP-ISPDASADEAENAAGDVFSRTC	
FNV0454 Fnuc 27886299	68	WWGDVVKSDIVSPKTTVFRDKIGKDKLIODMIDR-LDKDYIYSSDAADIYKEE PDNFHKEIIAIKKGNLKSFDY 1 GFERKTGKAMPOSIIDKLLEEDEEKIDATFSKLETC	
Chlo0060 Caur 22969909	58	LOVGOPSPLSTAPEEVTETSDVLTAERETOAANNPDNLVYFNDPKIPIOORNULFTLLETIGRLENDPTLDEAG 6 NLPSADVOLNTDO-ARLILALDDEEWSLLETTILNI	
s1r0104 Ssp 16331836	78	LSUGTESPUTINAPEDCSFVDERTTEEKRKEVRAG-TIPRLORDNELTAOLKOE-RSOYLDAINOLRYL-AGTFPY 82 AEVKADROITNNOIGALKLDGPTWOTTOOTIOV	
Npun2210 Npun 23125890	23	LOVDKIAPOTUTAPSSANVEDVKTTEERRREARSG-ASAVWVDDPVINEOIHON-LOOLFTKGNOIREN-LDNFPF 89 TOSVYAOPETLYDATFLNLSDTDWOKTOIOVROT	
alr2784 Ana 17230276	49	KIGTVAPOTTKAPHTASTENKKRTEVERKAASKS-STPULMUDAKITAOTDON, LEKMI.EOGNETETS-AGSEPF 151 LHUDTITTOTIVH-ETVILELSDVEWTOTOKGROG	
Npup0472 Npup 23124089	122	LOUGHS POTTAPYTAKTEDOKKTEAERKAVSSS-SLOVIMI.DARINEOINEN, LOUI.DDGNEIRAV-ACAPPF 119 LOUETVTPEAVYE-ESFLUDI.SDVEKTOMGTHOS	
Pmi+0951 Pmar 23131657	60		
Synwh0656 Syn 23133329	50		
Mmc12334 Msp 23000675	76	TEVICEUTRNIKAERDULVEDRAATELREEKAVAA_VPEVENMPHMVASI.SSG_TSDRIVELHEAMHTLAKSGOW 112 SATOTRIORPIN_INPEKERDI.TEDRUVTI.	
waff Cone 15617972	51		
BA 4976 Bant 21402351	47		
Sonsonsus / 95%			
consensus/95%			
		•	
Secondary structure			
Secondary structure			75
Secondary structure YQFF_Bsub_16079587 lin1503_Linn_16800571		HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	75
Secondary structure YQFF_Bsub_16079587 lin1503_Linn_16800571 ump1466_Lumon_16803506		HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	75 89
Secondary structure YQFF Bsub_16079587 lin1503_Linn_16800571 lm01466_Lmon_16803506 Ffac2060 Ffac_22003455		HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	75 89 89
Secondary structure YQFF Bsub_16079587 lin1503_Linn_16800571 lmo1466_Lmon_16803506 Efae2969_Efae_22993455 Desu3280_Thef=21113018		HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	75 89 89 87
Secondary structure YQFF_Bsub_16079587 lin1503_Linn_16800571 lm01466_Lmon_16803506 Efae2969_Efae_22993455 Desu2389_Dhaf_23113918 CAC1292_Cace_1584574		Inhuman       Inhuman	75 89 89 87 88
Secondary structure YQFF Bsub 16079587 lin1503_Linn 16800571 lm01466_Lmon_16803506 Efae2969_Efae_22993455 Desu2389_Dhaf_23113918 CAC1292_Cace_15894574 CTC02023_ct+2_28904087		HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	75 89 89 87 88 57
Secondary structure YQFF_Bsub_16079587 1in1503_Linn_16800571 1m01466_tmon_168003506 Efae2969_Efae_22993455 Desu2389_Dhaf_23113918 CAC1292_Cace_15894574 CTC02023_Ctet_28204087 CTC02023_Ctet_28204087		INNERSITESD       INNERSITESD         INNERSITESD       KLSDAKDKVEKELKSNSIPSKYLGÅRTEIGRATIPNYVFDPKATEAKROEASDNVQQV-QLVGGQUVDEENDLDREVYRKLELGGLINNSKL         VNTVNSSEIPSD       KLSDAKDKVEKELKSNSIPSKYLGÅRTEIGRATIPNYVFDPKATEAKROEASDNVQQV-QLVGGQUVDEENDLDREVYRKLELGGLINNSKL         VAKTMENKIRDE       KLNAAKIRARDDIELSAISSYKNVSKUUSYAIVP-NEIYNEEQTDERKREAAGSVVPV-KLQGQVIVQEGUVDEETYRGLMUHLLDQKMP         VEKSMENKIRE       KLNAKIRARDDIELSAISSYKNVSKUUSYAIVP-NEIYNEEQTDERKREAAGSVVPV-KLQGQVIVQEGUVDEETYRGLMUHLLDQKMP         VEKSMENKIRE       KLNAKIRARDDIELSAIFSYKNVSKUUSYAIVP-NEIYNEEQTDERKREAAGSVVPV-KLQGQUIVDEETYRGLMUHLLDQKMP         VEKSMENKIRE       KLNAGUITARADDIELSAIFSYKNVSKUUSYAIVP-NEIYNEEQTDERKREAAGSVVPV-KLQGQUIVQEGUVDEETYRGLMUHLLDQKMP         KESMENKIRARDIELSGIFSYKNVSKUVSVAVVV-NUP-NNFUDEGTDERKEAAGSVVPV-KLQGQUIVQEGUVDEETYRGLMUHLLDQKMP         LDEQMSKQVRES       ELEAFKQEAEQIQINNTPAQQOMIRYLVDGGUV-NDVNVVV-NLLDCKVPV-KLQGUIVQEGUTVDETYRGLMUHLDQKMP         LKDAINGEIRG       DVPALRQKIKQDINEASFPDEFKAFLARFVDAKTQP-TILEDEETTELKQARARSVLMESGRXANGKIVGFOETVDEKTMQVUGUGUSUGUSPG         LKDAINGEIRG       DVPALRQKIKQSIAAEFSKEFSKDAILGIAINSVVP-NNFFVDEXTELKKDISKNERVVV-VVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV	75 89 87 87 88 57 66
Secondary structure YQFF Bsub 16079587 lin1503 Linn 16800571 lm01466 Lmon_16803506 Efae2969 Efae 22993455 Desu2389 Dhaf 23113918 CAC1292 Cace 15894574 CTC20202 Ctet 28204087 CPE2019 Cper 18311001 CPE2019 Cper 18311001		HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	75 89 87 87 88 57 66 58
Secondary structure YQFF Bsub 16079587 1in1503_Linn_16800571 1m01466_Lmon_16803506 Efac2969_Efac_22993455 Desu2389_Dhaf_23113918 CAC1292_Cace_15894574 CTC02023_Ctet_28204087 CTE2019_Cper_18311001 Chte1004_Cthe_23020923 CHEC1004_Cthe_23020923		HHHHHH       HHHHHHHHHHHH         WITVMSSEIPSD       KLSDAKDKVEKELKSNSIPSKYLGÅRTEIGRAIIPNYVFDPKATEAKRORASDNVQQVQIXQQQUVEENDLDDREVYRKIELGGLINNSKI         VNTVNSSEIPSD       KLSDAKDKVEKELKSNSIPSKYLGÅRTEIGRAIIPNYVFDPKATEAKRORASDNVQQV-QIXQQQUVEENDLDDREVYRKIELGGLINNSKI         VAKTHENIRDE       KLNAAKIRARDDIELSAISSYIKN'SKUUSYAIVP-NETINEEQTDERKREAQSVVPV-KLQQQUTVQEQIVDRETYRQLMHLLDQKMP         VEKSMENINE       KLNAKIRARDDIELSAISSYIKN'SKUUSYAIVP-NETINEEQTDERKREAQSVVPV-KLQQUTVQEQIVDRETYRQLMHLLDQKMP         JUESNSKQVES       ELEAFKQEAEBQIQILNVTPAQQQMIRIUDQCIV-NDVLNEKKTEELKGSAREAVQPV-MIFQCEIIVEENDASAMKKLELIGINQTS         ZIKAQSSIAAEFSKEKFSDEFKAT       LNATAGINSVKPNMFIDSKKTEELKROISKKVENV-VIKKOGEIVDEKINQUVGVGUVUVGYGLUENGSP         LIKAVAENED       JIKKAQSSIAAEFSKEKFSKDARTEGIAISNYKPNMFIDSKKTEELKROISKKVENV-VIKGQIUVGCIVKEQEPVTARISUEDIGLISNSKO         LINKAYEAPIEG       JIKKAQSSIAAEFSKEKFSKDARTALGIAISNYKPNMFIDEKKTEELKROISKKVENV-VIKKOQIUVKGIUVKGGEIVDEKINQULGUSGIUVUVGUUVGUUVGUUVUVGUUVUVGUUVUVGUUVUVUVUVUVUVUVUVUVUVUVUVUVUVUVUVUVUVUVU	75 89 87 88 57 66 58 87
Secondary structure YOFF Bsub 16079587 lin1503 Linn 16800571 lm01466_Lmon_16803506 Efa2969 Efae 22993455 Desu2389 Dhaf 23113918 CAC1292 Cace_15894574 CTC20202 Ctet 28204087 CPE2019 Cper_18311001 Chte1004 Cthe 23020923 TTE0971 Tten 20807451 Cmet0576 Comt 2080661		HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	75 89 87 88 57 66 58 87 58
Secondary structure YQFF Bsub 16079587 1in1503_Linn_16800571 1m01466_tmon_16803506 Efae2969_Efae_22993455 Desu2389_Dhaf_23113918 CAC1292_Gace_15894574 CTC02023_Ctet_28204087 CTP2019_Cper_18311001 Chte1004_Cthe_23020923 TE0971_Tten_20807451 Gmet0579_Gmet_23053661		HHHHHH       HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	75 89 87 88 57 66 58 87 58 18
Secondary structure YOFF Bsub 16079587 lin1503 Linn 16800571 lm01466_Lmon_16800570 Efa2936_Efac_2293455 Desu2389 Dhaf 23113918 CAC1292_Cace_15894574 CTC20203_Ctet_28204087 CPE2019_Gper_18311001 Chte1004_Cthe_23020923 TTE0971_Tten_20807451 Gmet0579_Gmet_23053661 LA1683 Lint_24214883 Dden1075 Eda-23240858		HHHHHH	75 89 87 88 57 66 58 87 58 18 26
Secondary structure YQFF Bsub 16079587 lin1503_Linn 16800571 lm01466 Lmon_16803506 Efae2969_Efae_22993455 Desu2389_Dhaf 23113918 CAC1292_Cace_15894574 CTC02023_Ctet_28204087 CTC02023_Ctet_28204087 CTE2019_Cper_18311001 Chte1004_Cthe_23020923 TTE0971_Tten_20807451 Gmet0579_Gmet_23053661 LA1683_Lint_24214383 Pdes1875_Pdes_23474958		HHHHHH       HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	75 89 87 88 57 66 58 87 58 18 26 99
Secondary structure YOFF Bsub 16079587 1in1503 Linn 16800571 1m01466_Lmon_16803506 Efae2969_Efae 22993455 Desu2389 Dhaf 23113918 CAC1292 Cace 15894574 CTC20202 Ctet 28204087 CPE2019_Cper_E311001 Chte1004 Cthe 23020923 TTE0971_Tten 20807451 Gmet0575 Gmet 23053661 LA1683 Lint 24214383 Ddes1875 Ddes 23747958 TP00551 Tpal 15639638		HHHHHH       HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	75 89 87 88 57 66 58 58 58 26 99 62
Secondary structure YQFF Bsub 16079587 lin1503 Linn 16800571 lm01466_Lmon_16803506 Efac2969 Efac 22993455 Desu2389 Dhaf 23113918 CAC1292 Cace 15894574 CTC20202 Ctet 28204087 CPE2019 Cper 18311001 Chte1004 Cthe 23020923 TE0971 Tten_20807451 Gmet0579 Gmet 23053661 LA1683 Lint_23214383 Ddss1875 Ddes 23474958 FNW0454 Fnuc 27886299		HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	75 89 87 88 57 66 58 75 87 58 26 99 62 80
Secondary structure YOFF Bsub 16079587 lin1503 Linn 16800571 lm01466_Lmon_16803506 Efa22669 Efa2 22993455 Desu2389_Dhaf_23113918 CAC1292_Cace_15894574 CTC20202 Ctet 28204087 CPE2019_Cper_18311001 Che1004 Cthe 23020923 TTE0971_Tten_20807451 Gmet0575 Gmet 23053661 LA1683 Lint 24214383 Ddes1875 Ddes 23474958 TP0651 Tpal 15639638 FNV0454 Truc 2788299 Che0060_Cau_22969099		HHHHHH       HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	75 89 87 88 57 66 58 58 58 18 26 99 62 80 84
Secondary structure YQFF Bsub 16079587 lin1503 Linn 16800571 lm01466_Lmon_16803506 Efac2969 Efac 22993455 Desu2389 Dhaf 23113918 CAC1292 Cace 15894574 CTC20202 Ctet 28204087 CPE2019 Cper 18311001 Che1004 Cte 23020923 TTE0971 Tten 20807451 Gmet0579 Gmet 23053661 LA1683 Lime 23043638 PM0545 Fnuc 27886299 FNU0454 Fnuc 27886299 slr0104 Sep 1631836		HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	75 89 87 88 57 66 87 58 87 58 87 58 80 80 84 75
Secondary structure YQFF Bsub 16079587 lin1503_Linn 16800571 lm01466_Lmon_16803506 Efae2969_Efae_22993455 Desu2389_Dhaf_23113918 CAC1292_Gace_15894574 CTC02023_Ctet_28204087 CTC02023_Ctet_28204087 CTC02019_Cper_18311001 Chte1004_Cthe_23020923 TTE0971_Tten_20807451 Gmet0579_Gmet_23053661 LA1683_Lint_24214833 Ddes1875_Ddes_23474958 TP0651_Tpal_15639638 FTV0454_Tnuc_27886299 Chl00060_Caur_22969909 Slc0104_Sp_16331836 Npun2210_Npun_23125890		HHHHHH       HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	75 89 87 65 87 58 87 80 80 84 75 80 84 75
Secondary structure YOFF Bsub 16079587 lin1503 Linn 16800571 lm01466_Lmon_16803506 Efa2969 Efae 22993455 Desu2389 Dhaf 23113918 CAC1292 Cace_15894574 CTC20202 Ctet 22804087 CPE2019 Cper_18311001 Chte1004 cthe 23020923 TTE0971 Tten 20807451 Gmet0579 Gmet 23053661 LA1683 LInt_24214383 Ddes1875 Ddes 23474958 TP0651 Tpal 15539638 FNV0454 Fnuc_27886299 Slr0104 Ssp_16331836 Npun2210 Npun 23125890 slr0104 Ssp_16331836 Npun2210 Npun 23125890		HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	75 89 87 65 87 58 87 58 87 58 82 99 62 84 75 24 51
Secondary structure YQFF Bsub 16079587 lin1503_Linn 16800571 lm01466 Lmon_16803506 Efae2969_Efae_22993455 Desu2389 Dhaf 23113918 CAC1292_Cace_15894574 CTC20202 Ctet 28204087 CPE2019_Cper_18311001 Chte1004_Cthe_23020923 TRE0971_Tten_20807451 Gmet0579_Gmet_23053661 LA1683 Lint_24214383 Ddes1875_Ddes_23474958 TP0651_Tpal_15639638 FP00454_Fnuc_27886299 Chl00060_Caur_22969909 S1c0104_Ssp_16331836 Npun2210_Npun_23125890 alt2784_Ana_17230276 Npun0472_Npun_23124089		HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	7598978876587889877598877598877598775987
Secondary structure YOFF Bsub 16079587 lin1503 Linn 16800571 lm01466_Lmon_16800570 Bfa2936_Bfa22939Dhaf 23113918 CAC1292_Cace_15894574 CTC20203_Ctet_28204087 CPE2019_Cper_18311001 Chte1004 Cthe 23020923 TTE0971_Tten 20807451 Gmet0575 Gmet 23053661 LA1683 Lint 24214383 Ddss1875_Dds=23474958 TP0651_Tpal_15533638 FNV0454 Fnuc_27886299 Slr0104_Ssp_16331836 Npun2210 Npun 23125890 alt2784 Jana 17230276 Npun0475_Pmar_2131687		HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	759 897 857 658 758 857 658 758 857 658 752 844 752 14572
Secondary structure YOFF Bsub 16079587 lin1503 Linn 16800571 lm01466 Lmon_16803506 Efae2969 Efae 22993455 Desu2389 Dhaf 23113918 CAC1292 Cace 15894574 CTC20202 Ctet 28204087 CPE2019 Cper 18311001 Chte1004 Cthe 23020923 TE0971 Tten_20807451 Gmet0579 Gmet 23053661 LA1683 Lint 24214383 Ddss1875 Ddes 23474958 FNV0454 Pnuc 27886299 Chl00060 Caur 22969909 ShC0065 Caur 23125890 Sh20164 Sp_ 16331836 Npun2210 Npun 23125890 Da12784 Ana 17230276 Npun0472 Npun 23124089 Pmit0951 Pmar 23131657 Symbubes Sym 23131657		NHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	75987 89987 8576587 5182992 84475214 5723 630
Secondary structure YOFF Bsub 16079587 lin1503 Linn 16800571 lm01466_Lmon_16800570 Efae2069_Efae_2293455 Desu2389 Dhaf 23113918 CAC1292_Cace_18294574 CTC20203_Ctet_28204087 CPE2019_Cper_18311001 Chte1004_Cthe_23020923 TTE0971_Tten_20807451 Gmet0575_Met_23053661 LA1683 Lint_24214383 Ddes1875_Ddes_23474958 TP0651_Tpal_15639638 FNV0454_Fnuc_27886299 slr0104_Ssp_16331836 NPun2210_Npun_23125890 alc2784_Ana_17230276 Npun0475_Mpun_2312689 Pmic0951_Pmar_23131657 Symwh0556_Syn_2313329 Mmc12334_Msp_23000675		HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	75 89 87 88 57 65 87 85 85 85 85 85 85 85 85 85 85 85 85 85
Secondary structure YQFF Bsub 16079587 1in1503_Linn 16800571 1m01466_Lmon_16800571 Desu2389_Dhaf_23113918 CAC1292_Gace_15894574 CTC20202_Ctet2_28204087 CPE2019_Gper_18311001 Chte1004_Cthe_23020923 TE0971_Tten_20807451 Gmet0579_Gmet_23020923 DtA1683_Lint_24214383 Ddss1875_Ddcs_23474958 TP0651_Tpa1_15639638 FNV0454_Fnuc_27886299 Chl00060_Caur_22969909 S1C0104_Ssp_16331836 Npun2210_Npun_23125890 a1z2784_Ama_17230276 Npun0472_Npun_23124089 Pmit0951_Pmar_23131657 Symwh0656_Syn_2313329 Mmc12334_Msp_23000675		HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	7599878876587899608475763055 887658785865872545763055
Secondary structure YOFF Bsub 16079587 1in1503 Linn 16800571 1m01466_Lmon_16800570 Efae2969_Efae 22993455 Desu2389 Dhaf 23113918 CAC1292 Cace_15894574 CTC20203 Ctet 28204087 CPE2019_Cper_18311001 Chte1004 Cthe 23020923 TTE0971_Tten 20807451 Gmet0575 Gmet 23053661 LA1683 Lint 24214383 Ddes1875 Ddes 23474958 TP0651_Tpa1_15639638 FNV0454 Fnuc_27886299 slr0104_Ssp_16331836 NPun2310 Npun 23125890 slr2104_Ssp_16331836 Npun2470 Npun 23125890 slr21784 Āna_17230276 Npun0475 Mpun 23125890 slr21344 Msp_23000675 Yq#F Cpne_15611972 BA_4976_Bant_21402351		HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	759897887668788967254576887887688758875887518699208455245576305575

**Multiple sequence alignment of the 7TMR-HDEDs.** Multiple sequence alignment was constructed as detailed in the legend to Figure 1. The species abbreviations are as shown in Table 1.

alignment of the 7TMR-HDEDs shows that it is highly enriched in polar residues and is predicted to assume a predominantly  $\alpha$ -helical fold with several amphipathic helical hairpins (Figure 7).

To gain further insight into the functions of the 7TMR-HDs we used the contextual information available in the form of their gene neighborhoods (Figure 8A). In Grampositive bacteria, *Fusobacterium*, *Treponema*, *Leptospira*, *Geobacter* and *Thermotoga*, the 7TMR-HD gene is always associated with the conserved PhoH-YbeY-diacylglycerol kinase gene neighborhood, and is specifically located adjacent to the YbeY and PhoH genes. The PhoH-YbeYdiacylglycerol kinase gene neighborhood [58] is defined by several conserved genes (Figure 8A), which often occur adjacent to each other or the gene for the well-studied bacterial membrane-associated enzyme diacylglycerol kinase (DgkA) [59,60]. The PhoH gene encodes a member of the helicase-like superclass of the P-loop NTPase fold, and has been linked to the response to phosphate starvation in *E. coli* [58]. The third major member of this neighborhood encodes a highly conserved, uncharacterized protein typified by *E. coli* YbeY. An alignment of the YbeY orthologs (unpublished data VA & LA) indicates that it possesses 3 conserved histidines that are arranged in a manner very similar to the Zincin-like metallopeptidases [61,62]. The occurrence of YbeY next to the DgkA gene, and the fusions of the YbeY protein with the DgkA protein, suggest a function for it in the same pathway as DgkA, namely lipidhead-group metabolism. This function is also supported by the fusion of the YbeY protein to the acyl carrier protein synthase domain in *Plasmodium*. These observations, taken together with the predicted metalloprotease-like active site of the YbeY proteins, suggest that YbeY is most likely to function as an endogenous lecithinase (phospholipase C) in lipid metabolism to generate diacylglycerol, the substrate for DgkA, from phosphatidylcholine. This would imply that the 7TMR-HDs, which show a strong association with the PhoH-YbeY-diacylglycerol kinase gene neighborhood, are likely to function as a regulator of this pathway of lipid metabolism.

It is possible that the highly polar 7TMR-HDED may sense particular changes to ion concentrations, and regulate the YbeY-DgkA-dependent lipid metabolism pathway in order to regulate membrane properties. The nature of the intracellular signal transmitted by the HD hydrolase domain of the 7TMR-HDs is unclear. This HD domain is very distinct from the HD-GYP variety, which acts as a cyclic diaguanylate phosphodiesterase. Contextual analysis also provides no evidence for any association with GGDEF proteins, thus ruling out a role in cDGMP signaling [39]. Likewise, contextual information does not provide any evidence for association with cyclic NMP signaling even though HD domains are known to act as phosphodiesterases in this signaling pathway. These observations imply that the HD hydrolase domain of the 7TMR-HDs may have a distinct function of its own. One possibility is suggested by the contextual association of the 7TMR-HD genes with the genes for the P-loop protein PhoH (Figure 8A). These two proteins could potentially constitute a kinase-phosphoesterase couple that regulates YbeY-DgkA-dependent lipid metabolism. In Chlamydia pneumoniae 7TMR-HD gene occurs in the neighborhood of genes for several uncharacterized membrane proteins with no detectable homologs in other bacterial lineages. Hence, it is likely that in C. pneumoniae the 7TMR-HD has acquired a distinct function, which may be related to the expression of these pathogen-specific surface proteins.

# Other bacterial receptors with evolutionarily mobile membrane-associated sensory domains

Most of the above-identified bacterial 7TMRs contain additional N-terminal domains that are likely to play a crucial role in recognition of an extracellular signal. We were also interested in identifying novel families of membrane-associated receptors that do not contain any extracellular N-terminal domains, but primarily utilize their multi-TM domains for sensory purposes. While there are numerous prokaryotic signaling proteins with TM domains, a widely utilized sensory TM domain is likely to exhibit the following characteristics: 1) distinctive sequence or structural features that clearly distinguish them from generic multi-TM proteins and previously characterized transporters. 2) Evolutionary mobility, which means that the same conserved multi-TM domain could be associated with different types of intracellular signaling domains. These criteria are supported by the precedence offered by the domain architectures of previously identified membrane-associated bacterial receptors, such as those of the MHYT family [63]. Analysis of the clusters of conserved Multi-TM domains recovered in our receptor search procedure identified two widespread groups of conserved membrane-associated domains that were combined in different proteins with different types of intracellular signaling domains, but lacked any other extracellular domains.

The first group of these domains is typified by the Bacillus proteins, LytS and YhcK, which share a conserved membrane-spanning domain with 5 TM helices (Figure 8B and 9). In LytS-type proteins the 5TM domains are combined with C-terminal intracellular GAF and histidine kinase domains, while in the YhcK-type proteins the 5TM domain is combined with intracellular GGDEF (diguanvlate cyclase catalytic) domains. Occasionally, some members of the latter group, such as SO1500 from Shewanella, are also combined with additional intracellular EAL (cyclic diguanylate phosphodiesterase) and PAS domains (Figure 8B). We named this family of conserved 5TM domains the 5TMR-LYT family (for 5 transmembrane receptors of the LytS-YhcK type). 5TMR-LYTs are widely distributed in bacteria, with multiple members in Gram-positive bacteria, various proteobacteria, Fusobacteria, and Deinococcus (Table 1). The presence of a strongly predicted signal peptide in all members of this family suggests that it adopts a topology analogous to the classic 7TMRs: the N-terminus of the first helix is extracellular (or periplasmic), while the C-terminal tail with the fused signaling domain is intracellular (Figure 9). The membranespanning domain of the 5TMR-LYT family is distinguished from other membrane spanning domains by the presence of certain distinctive sequence features. These include the presence of a characteristic NXR motif in the loop between helix-1 and 2, multiple small residues, like glycine and proline, in the middle of helix-2, and a small residue (typically glycine) in the midst of the 5<sup>th</sup> helix. These small residues in the middle of the TM helices are likely to distort them, and this conformation may be critical to accommodate a ligand, or provide flexibility for transmission of a signal.

In several bacterial lineages, such as the Gram-positive bacteria and Vibrionaceae, the 5TMR-LYTs of the LytS variety is encoded by a gene that occurs in the same operon as a gene for a LytR type transcription factor (Figure 8B). This suggests that they transmit a signal via a LytR protein to regulate transcription. In Gram-positive bacte-



**Gene neighborhoods of 7TMR-HD, 5TMR-LYTs and the Phylogenetic tree, domain architectures and gene neighborhoods of the 8TMR-UT.**A) The domain architectures found in 7TMR-HD proteins and the PhoH-YbeY gene neighborhoods are shown. The upper panel shows the PhoH-YbeY neighborhood in proteobacteria (*E. coli* DgkA-YqfG-YqfF-YqfE operon), while the lower one shows the neighborhood typical of bacteria with the 7TMR-HD proteins. The organisms, possessing a particular domain architecture, are indicated by abbreviations in brackets). YbeY or its ortholog YqfG is the predicted lecithinase with a metal binding active site. YbeX is a Cystathionine beta-synthase domain (CBS) containing protein.B) The domain architectures found in 5TMR-LYT containing proteins and the conserved operon LytT-LytS found in *Bacillus* and gram-positive bacterial genomes are shown. C) Phylogenetic relationships of the 8TMR-UT domain containing proteins along with their domain architectures are shown. The RELL bootstrap values for the major branches are shown at their base. The thickness of a given branch is approximately proportional to the number of proteins contained within it. Domain architectures of the genes encoding 7TMR-DISM containing proteins are shown. The red arrow points to the architecture of protein encoded by a particular gene in a depicted neighborhood. Domain abbreviations are as shown in Figure 2 and HD, hydrolase of the HD superfamily; HD-GYP – cyclic diaguanylate phosphodiesterases of the HD-GYP variety; GAF – domain found in c<u>G</u>MP-specific phosphodiesterases, <u>A</u>denylyl cyclases and *Escherichia coli* <u>E</u>hlA. Species abbreviations are as shown in Table I

ria, LytS and LytR affect the composition of the cell wall by regulating murein hydrolases, and disruption of the LytS gene results in increased autolysis [64,65]. This suggests that the membrane-spanning domain of the 5TMR-LYT family may act as a receptor for derivatives of murein. Some of the 5TMR-LYTs of the YchK variety occur in gene

Secondary structure		EEEEEEE			
lytS_Bsub_1770001	37	GYKGKAILISI <mark>F</mark> SLFSIISNYTGIE <mark>I</mark>	16	SGS <mark>IANTRILGV</mark> EIGG <mark>L</mark> LG <mark>G</mark> PFVGAGIGILAGLHR <mark>F</mark> SLGGSTALSCAVSSILAG <mark>V</mark> LAG	LIC.
CTC01978_Ctet_28204044	37	KEFNKLALI <mark>V</mark> F <mark>F</mark> SLLSILGTYTGVN <mark>I</mark>	17	PHA <mark>IANTRP<mark>I</mark>GA</mark> ITAG <mark>Y</mark> LG <mark>G</mark> PLVGICVGLIAGLQR <mark>Y</mark> YVGGFTALACAVATIMGG <mark>I</mark> CGG	LAF
FNV0759_Fnuc_27886700	39	TKKDIIALS <mark>F</mark> F <mark>F</mark> SLLSISGTYIGLN <mark>F</mark>		NGA <mark>IL</mark> NTRN <mark>VGV</mark> IAGG <mark>I</mark> LG <mark>G</mark> PYVAIITGLVAGIHR <mark>A</mark> FVNLGRETAIPCAIATITGG <mark>F</mark> LTA	YV I
BA_0550_Bant_21397925	37	GYVDKLKLICIFSVFTIVSNYTGIEI	16	SST <mark>IANTRIMGV</mark> GISGLLG <mark>G</mark> PIVGIGVGSIAGIHR <mark>Y</mark> MLGGTTALSCAISSILAG <mark>V</mark> ITG	YIC
Ddes2593_Ddes_23475675	39	SPLQKSFCILFFGLFGILGTYSGNAI		FESVANLRAMAVITAGLFGGPAVGLGAGLMAGLHRNIIDPGGFSALPCGLATVLEGLFAG	
yehU_EC_26248504	44	RLPHKFLCYIVFSIFCIMGTWFGLHI		DDSIANTRAIGAVMGGLLGGPVVGGLVGLTGGLHRYSMGGMTALSCMISTIVEGLLGG	
YPdA_EC_26109167	41	SPRELLAVIAIFSLFALFSTWSGVPV		EGSLVNVRITAVMSGGILFGPWVGIITGVTAGIHRYLIDIGGVTAIPCFITSILAGCISG	W 1 1
FN0220_Fnuc_19/03565	22	TKKDIVALSFFFSLLSISGTIIGLNF		NGAILNTRNYGYIAGGILGOPIVSIIIGLVAGIHRAFVNLGRETAIPCAISTIIGGFLTA	
OB1642 Oibe 22777325	37	TI KTVI VHST. TPGLI STIGAHVGVVM	16	VENTANSK V V V V V V V V V V V V V V V V V V V	WAG
Open1099 Open 23037886	37	OISTKEVI.I. FAGEALESNLTGVET	17	SDSTANTETLUTTVAGLUGGEVUGTITGLIAGUHEVE-OGNGADSEVIESSIIIG	мто
Brub1268 Brub 22966664	41	RLPHKAACYAIFSLFCIMGTYFGLRI	1,	DDSIANTRAIGAVLGGMLGGPSVGLAVGLTGGLHRYSMGGFTALACAISTTAEGLIGG	LVI
vehU Styp 16420695	39	RLPHKLLCYVTFSIFCIMGTYFGLHI		EDSIANTRAIGAVMGGLLGGPVVGGLVGLTGGLHRYSMGGMTALSCMISTIVEGLLGG	LVE
S02822 Sone 24374352	36	RLPHKIFIYLVFSSFCIMATYFGEOT		SGAIANTRAMGAVLGGLLGGPVTGFLVGLTGGLHRYSMGGFTDLACAISTTLEGLSAG	MIS
lvtS Saur 21203400	37	TWKARWOLCIIFSLFALMSNLTGIVI	16	DVSLANTRVLTIGVAGLVGGPFVGLFVGVISGIFRVYMGGADAOVYLISSIFIGIIAG	YFO
SE2011 Sepi 27316477	37	KWRSKFOLI <mark>IIF</mark> GIFSMISNFTGIE <mark>I</mark>	16	DASMANTRVLTIGVSGLIGGPWVAIIVGIISGLCRLYIGGADAYTYLISSIVIATISG	YFO
qbs1052 Saga 23095505	36	LKKHQYILL <mark>I</mark> LFSLFAIISNFTGVE <mark>I</mark>	16	QSS <mark>VANTRVLTI</mark> GVSGLIG <mark>G</mark> PIVGIIVGLLSVFVR <mark>Y</mark> LQGGLAPHIYVISSLLIG <mark>L</mark> CSG	LSC
gbs0180 Saga 25010255	34	SKRETVVLV <mark>I</mark> I <mark>F</mark> GLFVIISNITGIE <mark>I</mark>	17	SDS <mark>LANTRTLVI</mark> TTASLVG <mark>G</mark> PLVGSIVGFIGGVHR <mark>F</mark> FQGSFSGSFYIVSSVLVG <mark>I</mark> VSG	ĸIC
TTE0871 Tten 20515869	38	TKKDLLITM <mark>L</mark> I <mark>F</mark> GLIGIAGTYLGVP <mark>I</mark>		NDA <mark>IANSRV<mark>VGP</mark>MVAGLIG<mark>G</mark>PLVGLGAGLIAGVHR<mark>F</mark>FLGGFTAISCGISTIIEG<mark>F</mark>AGG</mark>	LIE
C82292 Vcho 11278601	36	RFDHKISVY <mark>V</mark> L <mark>F</mark> SLFCIMGTYFGLQ <mark>I</mark>		NDA <mark>IA</mark> NTRA <mark>IGA</mark> VMGG <mark>L</mark> FG <mark>G</mark> PVIGFAVGLTGGLHR <mark>Y</mark> TLGGFTDLACAISTTAEG <mark>L</mark> IGG	LLE
y3885_¥pes_22127756	41	RLPHKLVCY <mark>L</mark> T <mark>F</mark> SMFCIMGTYFGLN <mark>I</mark>		DDS <mark>IA</mark> NTRA <mark>I</mark> GAVLGG <mark>M</mark> LG <mark>G</mark> PSVGFLVGLTGGLHR <mark>Y</mark> SMGGMTATACMLSTIAEG <mark>L</mark> LGG	LLE
yhcK_Bsub_7474584	26	RLVHKKDS-ISFQAVKGLACGLLGVI	8	©QHS <mark>II</mark> DLRN <mark>I</mark> P <mark>I</mark> MIAA <mark>L</mark> YG <mark>G</mark> WVSTATALAMITAGR <mark>L</mark> LITMNTSALYSVIIICIA <mark>A</mark> IPS	LIV
S01500_Sone_24347254	18	PKKHQRQYF <mark>L</mark> LWRLGIGVLIGGIGVA	9	PGI <mark>IF</mark> DTRS <mark>VLL</mark> CVSG <mark>L</mark> FF <mark>G</mark> GLPTAIAVVITALYR <mark>L</mark> SIGGV-AMWVGVGVIITSGLIG	۷IV
Rsph2038_Rsph_22958458	31	RRLRHVLLG <mark>L</mark> AFGFGASVQMMQSIAV	1	. PGV <mark>IV</mark> DSRA <mark>L</mark> FIGFAG <mark>A</mark> FL <mark>G</mark> PLGATVALMVSLLLR <mark>I</mark> SL-VSTAMPALAAGSMAASA <mark>L</mark> MGS	LWB
SAV0746_Saur_15923736	30	MVFSKAYVT <mark>VLM</mark> TIVSLLLSVYPIP <mark>Y</mark>	1	. EDY <mark>LI</mark> HLTF <mark>VPL</mark> LFLGRFT <mark>N</mark> MVYTLSATVIVAIVE <mark>I</mark> VVFNNSIMYGVTLIVIAA <mark>V</mark> TSA.	IGI
CAC3544_Cace_15026638	29	DKVRTIWGK <mark>V</mark> IVGVFCGFAGILLLM <mark>Y</mark>	6	TDT <mark>LI</mark> DLRA <mark>YAV</mark> MIAS <mark>Y</mark> TA <mark>G</mark> TLPTIISGIMIMLYR <mark>V</mark> FHFGVSVSSIVAAFRILS <mark>Y</mark> IIFI	FY ]
BA_0523_Bant_21397898	33	SFWQKCVVG <mark>ILT</mark> GILSVLLMYFGVH <mark>I</mark>		ENIMLDLRYLAVILAIIIGGPIASSITVTIILITRLFLTEYSLASELACYTIVLIGIG	VTI
lin2277_Linn_16801341	32	SINRKMVIG <mark>VWA</mark> GLLGFALMIFGIP <mark>I</mark>	1	NNV <mark>IVDLRHI</mark> PIIMVGFYG <mark>G</mark> PIPAIVSAVIITTSR <mark>F</mark> LISVNSA-AVMAGVIMM <mark>L</mark> IGI	IT?
DRA0181_Drad_15807849	36	DSRLQVPRFLLVSLMCVPLMFFPAQV	1	PGVFVDLRAVPIAFLTLRLGWGWGLVGAVPLLVYR <mark>Y</mark> L-LGGVGWPPAMVSAIGVVLVAG	LFI
DR1090_Drad_6458824	32	TAIHVPR-LLLAVMAVVLMLYPAEV	1	PGVIIDMRAVPIAYLALRKGVWAGLLALIPLLLYR <mark>F</mark> HLGGVGTWSAVFSAVGVVLLGG	LLI
consensus/95%		·····n.	• • • •	nns.p.n.nsnGss	<b>n</b> .
Secondary structure		1		1	
lvtS Bsub 1770001		RYFTKRY 3 TPRIAALVGIGMESLON	MII	TLLMAKPFSDAWELVSMIGIPMILINGTGSFIFLSIIOAIIRKEEOA 215\LvtS	
lytS_Bsub_1770001 CTC01978 Ctet 28204044		RYFTKRY 3 TPRIAALVGIGMESLQM KHSKSKT 1 NPKLAFITTIIAEILEM	MII	ILLMAKPFSDAWELVSMIGIPMILINGTGSPIFLSIQAIIRKEEQA 215\Lyts ILLISKPFNHALOLEKIFALPMIINSVGVFVFVYITKDALEOYN 212	
lytS_Bsub_1770001 CTC01978_Ctet_28204044 FNV0759 Fnuc 27886700		RYFTKRY 3 TPRIAALVGIGMESLQM KHSKSKT 1 NPKLAFITTIIAEILEM RFIKSKD KIFFGFFLACVVENLSM	MII	LLIMAKPFSDAWELVSMTGIPWILLNG7GSPIFLSIQAIIRKEEQA 215\Lyts TLILSKPPNHALQLEKIFALPWIIINSVGVFVYVITKDALEQYN 212 LILIKKRILQNIVANFVVPWVFNSIGSSVILLVEDIIGKSEL 200	
lytS_Bsub_1770001 CTC01978_Ctet_28204044 FNV0759_Fnuc_27886700 BA 0550 Bant 21397925		RYFTKRY 3 TPRIAALVGIGMESLQM KHSKSKT 1 NPKLAFITTIIAEILEM RFIKSKD KIFFGFFLACVVENLSM YIFKKYN 3 TPKFSAVLSVFIVSLEM	MII MII MGL MIM	ILLMARPFSDAWELVSWIGIPMILINGTGSFIFLSIGAIIRKEEQA 215\Lyts ILLISKPPNHALQLEKIFALPMIIINSVGVFVFVYITKDALEQIN 212 ILIILKKILAQNIVANFYVPMVFMNSIGSSVLILLVEDIIQKSELI 200 ILLIVEDGISIVKTIAPMILVNSFGSFILLSMICAILROEENA 212	
lyts Bsub 1770001 CTC01978 Ctet 28204044 FNV0759 Fnuc 27886700 BA 0550 Bant 21397925 Ddes2593 Ddes 23475675		RYFTKRY 3 TPRIAALVGIGMESLQM KHSKSKT 1 NPKLAFITTITAEILEN RFIKSKD KIFFGFFLACVVENLSM YIFKKYN 3 TPKFSAVLSVFIVSLEM WRVKESM DWRLAAMLGLVGETVHM	MIII MIII MGLI MIMI MIMI	LLMAKPFSDAWELVSMTGIPHILINGTGSPTFLSTQAIIRKEEQA 215\LytS LILSKFPNHALQLEKIPALPHIINSVGPVFVYTRTNALQIN 212 LILSKFPNHALQLEKIPALPHIINSVGFVFVYTRTNALQIN 212 LILIVEGISIVKTAIPHILVNSFGSFILLSMIQAIRQEENA 212 VLALSRFPPQAVEVKTQTQHMIVVNSGALFVQINLVSKGEENA 200	
lytS Bsub 1770001 CTC01978 Ctet 28204044 FNV0759 Fnuc 27886700 BA 0550 Bant 21397925 Ddes2593 Ddes 23475675 yehU Ec 26248504		RYFYKRY 3 TPRIAALVGIGMESLQ) KHSKSKT 1 NPKLAFITTIIAEILEN FFIKSKD KIFFGFFLACVVENLSN YIFKKYN 3 TPKFSAVLSVFIVSLEN WRVKESM DWRLAAMLGLVGETVHN SLILTRG 6 NPITAGAVTFVAEMVQ)	MIII MIII MGLI MIMI MILI	LLIAAKPFSDAWELVSNIGIPHILINGTGSPIFLSIGAIIRKEEQA 215\Lyts LLISKFPNHALQLEKIFALPHIINSVGYFVFVYITKDALEQYN 212 LIILKKFLLQNIVARYVFWYFNSIGSSVLLLVENIIQKSELI 200 ILLIVEDGISIVKITAIPHILVNSFGSFILLSMIQAILRQEENA 212 VLALSRFFPQAVELVNIGHPMIVVNSLGAALFVQIINLVSKGRER 200 LILAIRFYEDAVKLSWARPMVTNVGAALFVRILLOKRAMFEKY 209	
lytS Bsub 1770001 CTC01978 Ctet 28204044 FNV0759 Fnuc 27886700 BA_0550 Bant_21397925 Ddes2593 Ddes_23475675 yehU_EC_26248504 ypdA Ec 26109167		RYFTKRY 3 TPRIAALVGIGMESLQ) KHSKSKT 1 NPKLAFITTIIAEILEN RFIKSKD XIFFGFLACVUENLSN YIFKKYN 3 TPKFSAVLSVFIVSLEN WRVKESM DWRLAAMLGLVGETVHN SILIRRG 6 NPITAGAVTFVAEWVQN LKIPKAQ RWRVGILGGMLCETLTN	MIII MGLI MGLI MINI MILI MILI MILI	ILLMARPFSDAWELVSNIGIPHILINGTGSFIFLSIGAIIRKEEQA 215\Lyts ILLLSKPFNHALQLEKIFALPHIINSVGYFVFVYITKDALEQIN 212 ILIILKKILAQNIVANFYVPMVFMNSIGSSVLILLVEDIIQKSELI 200 ILLIVEDG	
lyts Bsub 1770001 CTC01378 Ctet 28204044 FNV0759 Fnuc 27886700 BA 0550 Bant 21397925 Ddes 2539 Ddes 23475675 yehu Ec 26248504 ypda Ec 26109167 FN0220 Fnuc 19703565		RYFTKRY 3 TPRIAALVGIGMESLQ) KHSKSKT 1 NPKLAFITIAEILEN FIKSKD KIFFGFLACVVENLSN YIFKKYN 3 TPKFSAVLSVFIVSLEN MRVKESM DWRLAAMLGLUGETVHH SILIRRG 6 NPITAGAVTFVAEMVQN LKIPKAQ RNFVGILGGMLCETLTN RFFKNKD	MIII MGLI MGLI MINI MILI MILI MILI	LLIAAKPFSDAWELVSKTGIPHILINGTGSPIFLSIQAIIRKEEQA 215\LytS LILSKFFNHALQLEKIFALPHIINSVGYFVFVYITKDALEQIN 212 LILSKFFNHALQLEKIFALPHIINSVGYFVFVYITKDALEQIN 212 LLIVEGCISIVKTAIPHILVNSFGSFULLSKIQAILRGEENA 212 VLLSRFFPGAVELVKYTGHPHILVNSFGALFVQIHLVSKGRERF 200 ILAIARPYEDAVELVKYTGHPHILVNSFGALFVGIHLVSKGRERF 200 ILAIARPYEDAVELVSTGAPHIVTNTGAALFWRILLDKRAMFEKY 209 VIVMAPTTALGIDIVSKIGIPHIL-GSVCIG-FIVLLVQSVEGEKEA 200 LILLIDKILGNUTSFYLPHVLMNSSGSVILLZVEDIGKSEIV 183	
lyts Bsub 1770001 CTCO1978 Ctet 28204044 FNV0759 Fnuc 27886700 BA 0550 Bant 21397925 Ddcs2593 Ddcs 23475675 yehU Ec 26248504 ypdA Ec 26109167 FN0220 Fnuc 19703565 Gmet2026_cmet_23055156		RYFTRY 3 TPRIAALVGIGHESLØV KHSKSKT 1 NPKLAFITTIIAEILED RFIKSKD KLFFGFIACVVENLSN YIFKKN 3 TPKFSAVLSVFIVSLEM WRVKESM DWRLAANLGLVGETVH SILIRRG 6 NFITAGAVTFVAEWVØN LKIFKAQ RNRVGILGGMLCETITM RFIKNKD RIFFGFIACTIENLSN	MIII MGLI MIMI MILI MILI MILI MGLI MVII	LLIAAKPFSDAWELVSNIGIPHILINGTGSPIFLSIGAIIRKEEQA 215\Lyts LLISKFPNHALQLEKIFALPHIINSVGYFVFVYITKDALEQYN 212 LIILKOKILAQNIVARYVPHVPNYSISSSVLLLVENIIQKSELI 200 ILLIVEDGISIVKITAIPHILVNSFGSPILLSMIQAILRQEENA 212 VLALSRFFPQAVELVKVIGHPMIVVNSLGAALFVQIINLVSKGRER 200 ILLAIRPYEDAVKUSVGAPHIVVNSLGAALFVQIINLVSKGRER 200 ILLILDKILAQNIVTSFIIPHILMSIGASVLILIVEDIIQKSEIV 183 LLVMAKPTAAVSLVSVIGPMILVNSIGLALFVELVSVFREKER 199	
lyts Bsub 1770001 crc0378 [cte 28204044 FW0759 Fnuc 27886700 BA_0550 Bant_21397925 Ddes2593 Ddes 23475675 yehU_E_226248504 ypdA Ec 26109167 FM0220 Fnuc 19703565 Gmet2026 Gmet 23055156 Gml642_0ihe_22777325		RYFTKRY 3     TPRIALVGIGHESLOW       KHSKSKI 1     NPKLAFITTILAEILEN       RFIKSKD     NPKLAFNTITIAEILEN       RYFKSM     TPKFSAVLSVFIVSLEN       SULIKPKQ     NPKLAANLGUVEVTHN       SLIKPKQ     NPKROKIGGAUTEVALEWUG       KIFKAQ     RHVRVGLIGGAUTEVALEWUG       RFIKNKD     RIFPGPFLACTIENLSN       HALKRQ     1     DPAVAFITGVAVETLQ       QFFSDDR 2     SPKNALFIGIFPVLIN	MIII MGLI MGLI MILI MILI MILI MILI MULI MULI	LLMAKPFSDAWELVSKIGIPHILINGTGSPIFLSIGAIIRKEEQA 215\LytS LLISKFPNHALQLEKIPALPHIINSVGPFVFVYITKDALEQIN 212 LIISKFPNHALQLEKIPALPHIINSVGFFVFLYITKDALEQIN 212 ILIVEGGISIVKTIAIPHILVNSFGSFILLSWIGAILRQEENA 212 VLLSRFPFQAVELVKYIGHMILVNSFGALAFVQINLVSKGRERR 200 ILAIRFYEDAVELVSNIGHMILVNSFGGIFVILLSWIGAVEREX 209 VLUMAPTALGIDIVSKIGIPHIL-GVCIG-FVLLVSVGVEGEKEA 200 ILIILDKILAQNIVTSFIPHVLMNSIGASVLILIVEDIQKSEIV 183 LLVMAKFFTAAVSLVSVIGFPHILVNSIGLAFVELVSVFREKERF 199 LLIMLPQCIQVUNTIGFVILVNSIGLAFVELVSVFREKERF 199 LLIMLPDQCIQVUNTIGFVILVNSIGLAFVELVSVFREKERF 199	
lyts Bsub 1770001 CTC0378 Ctet 28204044 FNV0759 Fnuc 27886700 BA 0550 Bant 21337925 Ddes2539 Ddes 23475675 yehu Ec 26248504 ypdh Ec 26109167 FN0220 Fnuc 19703555 Gmet2026 Gmet 23055156 081642 Oihe 22777325 0cen1039_coen 23037886		RYFTRY 3 TPRIALVGIGHESLOW KHSKSKT 1 NPKLAFITTIIAEILER RYKSKD KIFGFFLGCVVENLSN YIFKKNN 3 TPKFSAVLSVFIVSLEM SILIRRG 6 NPITAGAVTFVAEMVOM LKIFRAQ RVRLAGKLGVETTH SILIRRG 1 DPAVAFITGVAVETLOM QFFSDDR 2 SPNKAFIGIFPVLHN HSYSKRR 3 TAGVAATIGLGMEFID	MIII MGLI MGLI MILV MILV MILV MGLI MGLI MLFI	LLMAKPFSDAWELVSKTGIPHILINGTGSPIFLSIQAIIRKEEQA 215\LytS LLISKFFNHALQLEKIFALPHIINSVGFFVEVYITKDALEQIN 212 LLISKFFNHALQLEKIFALPHIINSVGFVEVYITKDALEQIN 212 ILIILKOKILAQNIVARYVPMYFNHSIGSSVILLVENIIQKSELI 200 ILLIVEDCISIVKTIAIPHILVNSFGSFILLSHIQAILRQEENA 212 VLALSRFPFQAVELVKVGMPHIVVNSIGAALFVQIINLVSKGRERR 200 ILAIARPYEDAVRLVSNIMAPHMVINTVGAALFVRILLDKRAMFEKY 209 VIVMAPTTALGIDIVSKIGIPHIL-GSVCIG-FIVLUVGSVEGKERA 200 ILLILDKLIQNIVTSFIPHVLMNSIGAALVELVSVFREKER 199 LLILLDKLIQNIVTSFIPHVLMNSIGAALVELVSVFREKERF 199 LLIMLPDQQIGVIVNTGFPHILVNSIALTFTMIRLALNETEQE 214 LVVFSVSGLQLVKLIFLPHMILNSFGTVVFNSIALTISALHOEEQL 213	
lyts Bsub 1770001 CTC01978 Ctet 28204044 FNV0759 Fnuc 27886700 BA 0550 Bant 21397925 Ddes2593 Ddes 23475675 yehU Ec 26248504 ypdA Ec 26109167 FN0220 Fnuc 19703565 Gmet2026 Gmet 23055156 OB1642 OIhe 2777325 Ocen1099_Ocen 23037886 Rrub1266_Rrub 22966664		RYFTKRY 3 TPRIALVGIGHESLOW KISSKKI 1 NPKLAFITTITAEILEN KRISKKI NFKLAFITITAEILEN KRISKKI KIFFGFLACUVENLSI WRVKESM DURLAANLGLVGETVHN SILIRG 6 NPITAGAVTFVAEWVOM KINFKAQ RWRVGILGGMLCETIT RFIKNKD RIFFGFLACTIENLSN QFFSDDR 2 SPNKALFIGIFPPULM HSYSKR 3 TAGVAAITGLGMEFIQM RIVSRG 6 QPLVVGLVTLFAEVCOM	MIII MGLI MGLI MILI MILI MILI MGLI MGLI	LLIMAKPFSDAWELVSNTGIPHILINGTGSPIFLS IQAIIRKEEQA 215\LytS LLISKFPNHALQLEKIFALPHILINGTGSPIFLS IQAIIRKEEQA 215\LytS LLISKFPNHALQLEKIFALPHILINSVGFVEVVIITKDALEQIN 212 ILLIKONTANTVHVNYNSSSVILLVENIIQKSELI 200 ILLINGTSPFQAVELVKIGMPHIVNSIGAALFWRILLOKRAMFEKY 200 JULAIRFYEDAVKLSWIAPHIVNSIGAALFWRILLOKRAMFEKY 200 JULAIRFYEDAVKLSWIAPHIVNSIGAALFWRILLOKRAMFEKY 200 ILLILIDKILAGNIVTSFIIPHILNSIGASVILLIVEDIIQKSEIV 183 LLIMLPAQTAUSVISTGIPHILNSIGLALFVELVSSVFREKER 199 LLIMLPAQGIGVQIVNTGIPHIVNSIGLALFVELVSSVFREKER 199 LLIMLPQQIGVQIVNTGIPHIVNSIGLAFVVELVSSVFREKER 199 LLIMLPQQIGVQIVNTGIPHVMINSIGLAFVVELVSSVFREKER 199 LLIMLPQQIGVQIVNTGIPHVMINSIGLAFVVENSISSIHLEGEL 213 ILVFSNGGLQUKLFLPHMILNSFGTWVFNSILAINTEEQE 206	
lyts Bsub 1770001 crc0378 [cte 28204044 FW0759 Fnuc 27886700 BA_0550 Bant 21397925 Ddes2593 Ddes 23475675 yehU_Ec_25248504 ypdAEc_26109167 FN0220 Fnuc 19703565 Gmet 2026 Gmet 23055156 Obl642_016 22777325 Ocen1039 Ocen 23037886 Arub1268 Fnub 22966664 yehU_Styp_16420695 -		RYFTKRY 3     TPRIALVGIGHESLOW       KHSKSKT 1     NPKLAFITTITLELLEN       RFIKSKD 4     NFGFFLACUVENLSN       YIFKKNN 3     TPKFSAVLSVFIVSLEN       SILIRG 6     NPITAGAVTFVALEWOW       RFIKNKD 7     RHRVGILGGMLCETITN       RFIKNKD 7     DPAVAFITGVAVETLOW       OFFSDDR 2     SNKLAFIGGHPFVLIN       HSYSKR 3     TGVAATTGLGHEFIQ       RVJLVSRG 6     SPLTAGAITCLAEUCM	MIII MGLI MILI MILI MILI MILI MGLI MUII MLII MLII	LLIAAKPPSDAWELVSKIGIPHILINGTGSPIPLSIQAIIRKEEQA 215\LytS LLIAKFPNHALQLEKIPALPHIINSVGYPVVYITKDALEQYN 212 LIIKKFPNHALQLEKIPALPHIINSVGYPVVYITKDALEQYN 212 ILIIKORIUSQNIVARYVMYPMYSHSSSSVILLVENIQKSELI 200 ILLINSFCFPQAVELVKY GMPHIVVNSCGALFVQINLVSKGERER 200 ILAIARPYEDAVELVSNIGAPMYUTYUGALFWRILLDKRAMFEKY 209 VIVWAPTTALGIDIVSKIGIPHIL-GSVCIG-FIVLLVSVEGEKEA 200 ILILINGANUTSYITHVLANSIGASVILLVEDIIQKSEIV 183 LLVMAKPTTALGUIVSVIGPHILVNSIGALIINTTMIKLAINETEGE 214 ILVIKSGCLQUKLFIPHVLNNSIGALFVTLJORKTMFEQ 213 ILVARPANNAVHLVETALPHVVTNVGAAFMTLLDKRAMFEXY 204	
lyts Bsub 1770001 crc0378 ctet 28204044 FW00759 Fnuc 27886700 BA 0550 Bant 21397925 Ddes2533 Ddes 23475675 yehu Ec 26109167 FW0220 Fnuc 19703565 Gmet2026 Gmet 23055155 OB1642 Oihe 2777325 Ocen1039 Ocen 2337886 Rrub1268 Rrub 22966664 yehu styp 16420695 S02822 Some_24374352		RYFTKRY 3     TPRIALVGIGHESLOW       KHSKSKT 1     NPKLAFITTILAELLEN       RFIKSKD 1     NFKLAFITTILAELLEN       WRVKESM 0     NFKTAGAVFTVALENUD       NRUKESM 0     DNRLAAMLGLVGETVHN       SILIRG 6     NFITAGAVFTVALENUD       RFKRM 1     DNRLAAMLGLVGETVHN       RIKIRG 7     DPAVAFTGVALEVUD       QFFSDDR 2     SPNKALFIGIFPPLACTIENLSN       SVLIRG 6     OPLVAGUGVTLAGKVETLD       SVLIRG 6     SPLTAGATCIENLEND       SVLIRG 6     SPLTAGATCIENLEND	MIII MGLI MGLI MILV MILV MGLI MUII MGLI MLII MLII MLII	LLMAKPFSDAWELVSNTGIPHILINGTGSPIFLSIQAIIRKEEQA 215\LytS LLISKFFNHALQLEKIFALPHILINGTGSPIFLSIQAIIRKEEQA 215\LytS LLISKFFNHALQLEKIFALPHILINSVGFVEVVITKDALEQIN 212 ILIILKONIUSKIGAPHIVNSFGSFILLSHIQAILKOEENA 212 VLASRPFPQAVELVKVIGMPHIVNSIGAALFVQIINLVSKGRERR 200 LLAIARYFEDAVRLVSNIGAPHIVNSIGAALFVQIINLVSKGRERR 200 ILLINGNITSFIPHILNSSGGIPHILGSVCIG-FIVLLVQSVEGKEA 200 ILLILDKRIJQNITSFIPHVLNSIGAALFVRILLDKRAMFEKY 209 LILLIDKIJQNITSFIPHVLNSIGAAVILIVEUVSSVEREKEA 200 LLILLDKIPHONUSSIG FPHILVNSIGALFVRILLDKRAMFEKY 209 LILLIDKIPHONUSSIG FPHILVNSIGALFVRILLDKRAMFEKY 183 LLVMAKPFTANSLVSVIGFPHILVNSIGALFVRILLDKRAMFEKZ 214 ILVFSVGQLVKLIFLPHNILNSFGTVVFNSISISAILHOEEQL 213 ILLVARPAVNAVELVETIALPHUVANSVGAALFWRILLDRRAMFETQ 206 ILLIARFDDALHUSTIAPHWVINYGGALFWRILLDKRAMFETQ 204 ILLIARFDALHUSTIAPHWVINSIGALFWRILLDKRAFFETQ 204 ILLIARFDALHUSTIAPHWVINSIGALFWRILLOKRAFFETQ 204 ILLIARFDALHUSTIAPHWVINSIGALFWRILLOKRAFFETQ 204	
lyts Bsub 1770001 CrC01978 Ctet 28204044 FNV0759 Fnuc 27886700 BA 0550 Bant 21397925 Ddes2533 Ddes 23475675 yehU Ec 26248504 ypdA Ec 26109167 FN0220 Fnuc 19703565 Gmet2026 Gmet 23055156 G081642 Oihe 22777325 Goen1039_Ocen 23037886 Gnetb2026 Gmet 23055156 G081642 Oihe 22777325 Goen1039_Ocen 23037886 yehU Styp 16420695 S02822 Sone 24374352 lyts Saur 21203400		RYFTRRY 3 TPRIALVGIGHESLOW KHSSKKI 1 NPKLAFITTITLELLEN RFIKSKD KIFFGFPLACVVENLSN YIFKKNN 3 TPKTSAVLSVFIVSLEN SILIRKG 6 NPITAGAVTFVAENVOM KHVKESK 0 NPKTAANLGUCETITH RFIKNKD RIFFGFLACTIENLSN UGFSDDR 2 SNKALFIGIFPPVLIN HSYSKRR 3 TAGVAATTGLAVETLOM GPLVVGLVTLFAEVCOM SVLIRG 6 SPLTAGAITCIAELVOM SVLIRG 6 SPLTAGAITCIAELVOM SVLIRG 6 SPLTAGAITCIAELVOM LQAQRKK 3 SIKKAMIGIVHENUS	MIII MGLI MIMI MILV MILV MGLI MILV MGLI MLFI MLII MLII MLII MLII	LLMAKPFSDAWELVSKIGIPHILINGTGFIFLSIGAIIRKEEQA 215\LytS LLIAKKPFNHALQLEKIFALPHIINSTGFIFLSIGAIIRKEEQA 215\LytS LLISKFFNHALQLEKIFALPHIINSTGFFILSKIGAIIRKEEQA 212 ULIVEGGISIVKTAIPHILVNSFGSFILSKIGAIRVQINUSKGEENA 212 VLLSRFFPGAVELVKYGIPHILVNSFGARAFVQINUSKGEERA 200 ILAIARPFEDAVELVSTGFIFHIL-GSVCIG-FVLLVSKGERER 200 ILILIDKILAQNIVTSFIFHVLMSSIGAIVVQINUSKGERER 200 ILILIDKILAQNIVTSFIFHVLMSSIGASVLILIVEDIQKSEIV 183 LUVAKFFTAAVSLVSVIGFPHILVNSIGALFVVELVSSVFREKEF 199 LLILLAKFFTAAVSLVSVIGFPHILVNSIGALFVELVSSVFREKEF 199 ILILIARFFEDAVELVSVIGFPHILVNSIGAANFYTLIDKRAMFESI 214 ILVARPAVAVULVETAIPHUVANSIAITIFTMIKIALNETEGE 214 ILVARPAVAVULVETAIPHUVANSAANFYTLIDKRAMFESI 206 ILILARFFDDALHLVSTAPHUVTTVGAALFYRILDKRAMFESI 206 ILLIARFPDDALHUSTAAPHUVTTVGAALFYRILDKRAMFESI 201 ILLIACFPDAAMEUVQF2PHILVSIGAALFYSIINGKMTPESI 201	
lyts Bsub 1770001 crc0378 Ctet 28204044 FW0759 Fnuc 27886700 BA_0550 Bant_21397925 Ddes2539 Ddes_23475675 yehU Ec_26248504 ypdA Ec_26109167 FM0220 Fnuc 19703565 Gmet2026 Gmet 23055156 O081642 016 22777325 Ocen1039 Ocen 23037886 Rrub1268 Rrub_22966664 yehU Styp 16420695 S02822 Sone 24374352 Jyts Saur 21203400 SE2011 Sepi 27316477		RYFTRRY 3     TPRIALVGIGHESLOW       KHSKSKI 1     NPKLAFITTITLELLEN       RFIKSKD 1     NFKLAFITTITLELLEN       RFIKSKD 2     NFKLAFMELGUKVERTUN       SILIRG 6     NPITAGAVTFVALEWOW       LKIPKAQ 2     NRKVGILGGMLCETLTN       RFIKNKD 3     TAGVAATTGGAVETLOW       QFFSDDR 2     SPNKLAFIGICFPVLIN       HSYKRR 3     TAGVAATTGGAVETLOW       RYLVSRG 6     QPLVVGLVTLFAEVCQD       VILRRG 6     SPLTAGATTGLEVUP       FYLRRAG 6     SSLKSAMIGTUKENIQ       LQOQREK 3     SIKKABAIGAITELIQUE	MIII MIII MIMI MILV MILV MILV MILV MILV	LLIAAKPFSDAWELVSNTGIPHILINGGGPIFLSIQAIIRKEEQA 215\LytS LLISKFFNHALQLEKIFALPHIIINSVGYFVFVYITKDALEQIN 212 LIISKFFNHALQLEKIFALPHIINSVGYFVFVYITKDALEQIN 212 LIISKFFNHALQLEKIFALPHILVNSFGSFULLSMIQAILQSEENA 212 VLUASRFFPAQVELVKYTGHPHIVVNSFGALFVQIHLVSKGRERR 200 ILAIARPYEDAVRLVSNTAAPMVTNTVGAALFWRILLDKRAMFEKY 209 VIVMAPTTALGIDIVSKIGIPHIL-GSVCIG-FIVLLVQSVEGKERA 200 LILLINGNIVSSYTFPHVLMNSIGASVLILVEGVIQKSEIN 813 LLVMAKFFTAAVSLVSVIGFPHILVNSIGALFWRILLDKRAMFEKY 199 LLUARFPTAAVSLVSVIGFPHILVNSIGALFWRILLAINETEGE 214 ILVMSVGQLVKLIFIPHMILNSFGTWVFNSISAYLAINETEGE 214 ILVAFAVANVHLVETIAPHVVNNSUGALFWRILLDRTMFETQ 206 ILLIARFDDALHUSVIGFPHILVNSIGALFWRILLDRTMFETQ 206 ILLIARFDDALHUSVIGFPHILVNSIGALFWRILLDRTMFETQ 206 ILLIARFDALHUSVIGFPHILVNSIGALFWRILLOKRAMFEKY 204 ILLIARFAVANVHUSTAPHVINVSUGALFWRILLDRTMFETQ 206 ILLISFBHAXVAVHLSVGIAPHVLINSIGALFWRILLDRTMFETQ 201 ILLIAFSHDALHUSVGIAPHVLINSIGALFWRILLDRTMFETQ 201 ILLISFBHAXVAVHDISLALFMIUNSIGALFWRILLDRTMFETQ 201 ILLISFBHAXVAVHDISLALFMIUNSIGALFWRILLDRIKAVESKI 201 ILLISFBHAXVAVHSISALFMILINSIGALFWRILLDRIKKAVESKI 201 ILLFSHBAXVAVHLSVGIAPHILINSIGALFWRILLSKARFEKS 214	
lyts Bsub 1770001 CTC01978 Ctet 28204044 FW00759 Fnuc 27886700 BA 0550 Bant 21397925 Ddes2593 Ddes 23475675 yehU Ec 26248504 ypdA Ec 26109167 FW0220 Fnuc 19703565 Gmet2026 Gmet 23055155 OB1642 Oihe 2777325 Ocen1099_Ocen 23037886 Rrub1268 Rrub 22866664 yehU Styp 16420695 S02822_Some 24374352 lyts Saur 21203400 SE2011 Sepi 27316477 gbs1052_Saga 23095505		RYFTKRY 3 TPRIALVGIGHESLOW KISSKKI 1 NPKLAFITTITAEILEN KRISKKI NFKLAFITITIAEILEN KRISKKI KIFFGFLACUVENLSJ VIFKKYN 3 DYKLAANLGLVGETVHN SILIRG 6 NPITAGAVTFVAEWUGD KINFKAQ KNRVGILGGMCLEFITT RFIKNKD RIFFGFLACTIENLSD QFFSDDR 2 SPNKALFIGIFPPULH HSYSKR 3 TAGVAAITGLGMEFIQ RILVSKG 6 QPLVVGLVTLFAEVUGD SVLIRG 6 SPLTAGAITCIAELVGD SVLIKG 6 SPLTAGAITCIAELVGD VJLKRKG 8 SIAKGAITCIAEUUD LQAQRK 3 SIAKSANIGIVMENIG NFLVRKN 3 RVLDAMVVGFGHELLQD	MIII MGLI MINI MILV MILV MGLI MMILV MGLI MMILI MMLII MMLII MMSLI MMSLI MMSLI	LLMARPPSDAWELVSKIGIPHILINGTGPFILSIGAIIRKEEQA 215\LytS LLISKPPNRALQLEKIPALPHIIINSVGPFVEVYITKDALEQIN 212 LIISKPPNRALQLEKIPALPHIINSVGPFVEVYITKDALEQIN 212 LIISKPPNRALQLEKIPALPHIINSVGFVEVITKUSALER LIISKPICAUEVKVTGPWRVPNNSGSVLILVEDIIQKSELI 200 LLAIRPPEDAVELVSVTGPMRVVNSGRALFVQINLVSKGRERF 200 LIAIRPPEDAVELVSVTGPMLVNSTGGALFVQINLVSKGRERF 200 LILIDKILAQNIVSFYIPWLNNSIGASVLILVEDIIQKSELI 200 LILIDKILAQNIVSFYIPWLNNSIGASVLILVEDIIQKSELI 200 LILIDKILAQNIVSFYIPWLNNSIGASVLILVEDIIQKSELI 200 LILIDKILAQNIVSFYIPWLNNSIGASVLILVEDIIQKSELI 200 LILIDKILAQNIVSFYIPWLNNSIGASVLILVEDIIQKSELI 200 LILIDKILAQNIVSFYIPWLNNSIGASVLILVEDIIQKSELV 200 LILIDKILAQNIVSFYIPWLNNSIGASVLILVEDIIQKSELV 200 LILIDKILAQNIVSFYIPWLNNSIGASVLILVEDIIQKSELV 200 LILIALPFTANAVULVETIAPMULVNSIGAALFVSINSIGANETTELLVEDIQUS 214 LVFSVSGQULVKLIFIPMILNSFGTVFWSISISAILHVEDIQUS 200 LILIARPFDAAKUSNISAILAPMULVNSIGAALFVSILDKRAFFEX 200 LILIARPFDAAKUSNISAILAPMIINSGALFVSILDKRAFFEX 201 LILIACPFDAAKUSNISAILAPMILVNSVGAALFVSIISSIKSEGM 215 LILISPNDAAKUSVSTAPHULVSISIAAILFNSIISSIKEGM 215 LILFSVDAAKUSVSTAPHULVSISIAAILFNSIISSIKEGM 215 LILFSVDAAKUSVSTAPHULVSISIAAILFNSIISSIKEGM 215 LILFSVDAAKUSVSTAPHULVSISIAAILFNSIISSIKSEGM 215 LILFSVDAAKUSVSTAPHULVSISIAAILFNSIISSIKEGM 215 LILFSVDAAKUSVSTAPHULVSISIAAILFNSIISSIKKEGM 215	
lyts Bsub 1770001 CrC01978 Ctet 28204044 FW0759 Fnuc 27886700 BA 0550 Bant 21397925 Ddes2533 Ddes 23475675 yehU Ec 25248504 ypdA Ec 26109167 FW0220 Fnuc 19703565 Gmet2026 Gmet 23055156 OB1642 Oihe 22777325 Ocen1059 Ocen 23037886 Krub1268 Frub 22966664 yehU Styp 16420695 S02872 Sone 24374352 1yts Saur 21203400 SE2011 Sepi 27316477 gbs1052 Saga 2305505 gbs1080 Saga 25010255		RYFTRRY 3     TPRIALVGIGHESLOW       RKISKSKI 1     NPKLAFITTITLELLEN       RFIKSKD     NFLAFITTITLELLEN       RFIKSKD     NFLAFARTITLELLEN       RFIKSKD     NFLAFARTITLELLEN       SLIFRAG     NPITAGAVTFVALENOW       GYFSDDR 2     SNKALFIGLETITN       RYLVSRG 6     SPITAGATTGVAFELOW       RYLVSRG 6     SPITAGATTGLAFETQU       SVLIRRG 6     SPLTAGATTGLAFETQU       SVLIRRG 6     SPLTAGATTGLAFETQU       SVLIRRG 6     SPLTAGATTGLAFETQU       SVLIRRG 6     SLITAGATTGLAFETQU       MUTIRNG 7     SIXKSANGTONEMUM       MUTIRNG 8     SLASAMGTONEMUM       MUTIRNG 8     SULABAWGRELLD	MIII MGLI MGLI MINI MILV MGLI MMILV MGLI MMGLI MMLI MMLSI MMSCI MMCCI MICI	LLIAAKPPSDAWELVSKIGIPHILINGTGSPIPLSIQAIIRKEEQA 215\LytS LLIASKPPNALQLEKIPALPHIINSVGYPVVYITKDALEQIN 212 LIILSKFPNALQLEKIPALPHIINSVGYPVVYITKDALEQIN 212 LIILKOKILAGNIVASYVYPWYPMYSISSSVILLVSVIQKSEL 200 ILLINGTON VALLSKIPPGAVELVSKIGARVOINVSKGERE 200 ILAIAPPTEDAVELVSKIGIPHILOSSCALAFVOINVSKGERE 200 LILLIDKIDQUISSKIGIPHILOSSCALAFVOINVSKGERE 200 LILLIDKIDQUISSKIGIPHILOSSCALAFVOINVSKGERE 200 LILLIDKIDQUISSKIGIPHILOSSCALAFVOINVSKGERE 200 LILLIDKIDQUISSKIGIPHILOSSCALAFVOINVSKGERE 200 LILLIDKIDQUISSKIGIPHILOSSCALAFVOINVSKGERE 200 LILLIDKIPQUIGVUSVIGPHILVNSIGASVILLVSUFGERE 200 LILLIDKIPQUIGVUSVIGPHILVNSIGALFVELVSUFFEREF 199 LIVKAFVTANSLUSVIGPHILVNSIGALFVELVSUFFEREF 199 LILLIALPODALENSVIGPHILVNSIGALFVELVSUFFEREF 204 ILLIASPDANKUSVIGPHILVNSIGANFWILLDKRAMFET 206 ILLIASPDAAKUSVIGPHULVNSIGAAFWKILDKRAMFET 206 LILLIASPIDAAKUSVIGPHULVNSIGAAFWKILDKRAMFET 204 LLIASPDAAKUSVIGPHULVNSIGAAFVKILDKRAMFET 204 LLIASPDAAKUSVIGIPHULVNSIGAAFVKILDKRAMFET 204 LLITFSBUKAVAVDLSILAIPHVINSIGAAFVKILDKKAMFET 204 LLIFSBUKAVAVDLSILAIPHVINSIGAAFKIKSIKSVELEGU 215 ILLFSVNIHAATUVSFIAPHILNSIGCATFVITUSTKEEGU 215 ILLFSVNIHAATUVSFIAPHILSGGSTAFLALKKYLSNESQL 208 VIFFSBUKAVUSUVIGPHULNSIGATIFVSIISSVELEEN 214 VGIFFGWELVKKIVIPHULNSIGATFVITUSTERSUF1200 208	
lyts Bsub 1770001 crc0378 Ctet 28204044 FW00759 Fnuc 27886700 BA 0550 Bant 21397925 Ddes2593 Ddes 23475675 yehu Ec 26109167 FW0220 Fnuc 19703565 Gmet2026 Gmet 23055155 OB1642 Oihe 2777325 Ocen1039 Ocen 23037886 Rrub1268 Rrub 22966664 yehu Styp 16420695 S02822 Sone 24374352 lyts Saur 21203400 S82011 Sepi 27316477 gbs1052 Saga 25010255 TTE0871 Tten 20515869 G03202 Dten 117ten 20515869		RYFTRRY 3 TPRIALVGIGHESLOW KISKSKT 1 NPKLAFITTITLELLE RFIKSKD KIFFGFLACVVENLS YIFKKYN 3 TPKFSAVLSVFYVSLE WRVKESM DWRLAANLGLVGETVH SILIRG 6 NFITAGAVTFVAEWVOM LKIPKAQ RRVGVIGGMCLECITT RFIKNKD RIFFGFLACTIENLS NFNFRG 6 QPLVVGLVTLGAEVVOM RYLVSRG 6 QPLVVGLVTLGAEVCQ SVLIRG 6 SPLTAGAITCIAELVOM RYLVSRG 6 SPLTAGAITCIAELVOM SVLIRG 6 SPLTAGAITCIAELVOM NULRKSN 3 SIKKSANIGITVAEMIM NTIRKNY 3 RVLDAWTVGFMELLOM NYLRKY 2 WEXTFANGMELTAENIM NYLRKY 2 WEXTFANGMELTAENIM NYLRKY 2 DESTUMING	MIII MGLI MGLI MILI MILI MILI MGLI MGLI	LLMAKPFSDAWELVSNTGIPHILINGGGPTFLSIQAIIRKEEQA 215\LytS LLISKFFNHALQLEKIFALPHIINSVGFFVEVYITKDALEQIN 212 LLISKFFNHALQLEKIFALPHIINSVGFVEVYITKDALEQIN 212 LLILKOKILQNIVARYVPMYFNHSIGSSVILLYVENIQKSELT 200 LLILKOKILQNIVARYVPMYFNHSIGSVILLYVENIQKSERA 212 VLALSRFFPQAVELVKVGMPMIVVNSIGAALFVQIINLVSKGERR 200 LLIAIARYVEDAVRLVSNTAPHMVINTGGALFVRILLDKRAMFEKY 209 VIVMAPTTALGIDIVSKIGIPHI-GSVCIG-FIVLUVGSVEGKER 200 LLILLOKILQNIVTSYIPHVLMNSIGAALFVRILLDKRAMFEKY 209 LLVNSVGGLQLVKLFIPHVLMNSIGAALVRILVELVSSVFREKER 199 LLVNSVGGQLVKLFIPHVLMNSIGAALVRILVELVSSVFREKER 199 LLVNSVGGQLVKLFIPHVLMNSIGAALFVRILLDKRAMFETQ 206 ILLIARPDALHUSTIGIPHUVNSIGAALFVRILLDKRAMFETQ 206 LLILARPDALHUSTIAPHWIINVSGAALFVRILLDKRAMFETQ 206 LLILARPAVAAVHUETIAPHUVNSVGAALFVRILLDKRAMFETQ 206 LLILARPAVAAVHUETTAPHVINSIGAALFVRILLDKRAMFETQ 206 LLISFSHGAVAVLSVIGIPHILNSIGAALFVRILLDKRAMFETQ 206 LLISFSHGAVAVLSVIGIPHILNSIGAALFVRILLDKRAMFETQ 206 LLIFISHHATILVSFIALPHIINSIGAALFVRILLDKRAMFETQ 206 LLIFISHJANDLISJIALPHIINSIGAALFVRILLDKRAMFETQ 206 LLIFISHJANDLISJIALPHIINSIGAALFVRILLDKRAMFETQ 206 LLIFISHJANDLISJIALPHIINSIGAALFVRILLSKIEQEQM 215 LLIFSNHAATUVDISFIALPHIINSIGAALFVRILLSKIEQE200 215 LLIFSNHAATUVDISFIALPHIINSIGATFLIISIISSTGLEBH 214 LLFISHJANDLISJIALPHIINSIGAALFVRILLSCU 208 LLATAKFPHQAVELVKIIGIPHWVNATGIGIFHIMIKSIFDEKEQI 198 LLVANAFFORVANJELDSTIJALPHIINSIGALFVRILDSTUSSEN 201 LLISTSOFVANJEN ALPHIINSIGAALFVRILDSTUSSEN 208 LLATAKFPHQAVELVKIIGIPHMINISTGAALFVRILDSTUSSEN 208 LLATAKFPHQAVELVKIIGIPHMINISTGAFFUNISSIFUSIESEN 201 LLISTSOFVANJEN 2000000000000000000000000000000000000	
1yts Bsub 1770001 CrC01978 Ctet 28204044 FNV0759 Fnuc 27886700 BA 0550 Bant 21397925 Ddes2533 Ddes 23475675 yehU Ec 26248504 ypdA Ec 26109167 FN0220 Fnuc 19703565 Gmet2026 Cmet 23055156 OB1642 Oihe 22777325 Ocen1039 Ocen 23037886 Rrub1268 Cmet 2305515 O081642 Oihe 22777325 Ocen2037 Comet 2305156 081642 Oihe 2717325 Ust Sur 21203400 S2021 Sepi 27316477 gbs1052 Saga 23010255 gbs1080 Saga 2501025505 gbs1080 Saga 2301025505 gbs1080 Saga 2301025505 gbs1080 Saga 2301025505		RYFTRRY 3 TPRIALVGIGHESLOW KHSSKKI 1 NPKLAFITTITLELLEN RFIKSKD KIFFGFFLACUVENLSN YIFKKNN 3 TPKTSAVLSVFIVSLEN SILIRKG 6 NPITAGAVTFVAEWUGD KIFFAG NRVGILGGMUCETITN RFIKNKD RIFFGFLACTIENLSN UGFSDDR 2 SPNKALFIGIFPFVLM HSYSKRR 3 TAGVAATTGLAWETLOW GYFSDDR 2 SPNKALFIGIFPFVLM HSYSKRR 3 TAGVAATTGLAWETLOW SVLIRG 6 SPLTAGAITCIAELUGD LQAQRKK 3 SIAKSANIGIVMENIOM HGTIKQN 3 SIKKSANIGIVMENIOM HGTIKQN 3 SIKKSANIGIVMENIOM HGTIKQN 3 SIKKSANIGIVMENIOM HGTIKGN 5 SSUTURATIALEUTD NILRNN 3 STSQUILISIAESLOW NIXRNN 3 STSQUILISIAESLOW NIXRNN 5 OPLUNCUTTAELUGD NIXRNN 5 OPLUNCUTAELUGD NIXRNN	MIII MGL MGL MIL MGL MIL MGL MGL MGL MGL MGL MGL MGC MGC MGC ML ML ML ML ML ML	LLMAKPFSDAWELVSKIGIPHILINGTGFTFLSIQAIIRKEEQA 215\LytS LLIAKFPNHALQLEKIFALPHIINGTGFTFLSIQAIIRKEEQA 215\LytS LLIKKFPNHALQLEKIFALPHIINSVGFFVEVYITKDALEQIN 212 LLIKKFPNHALQLEKIFALPHIINSVGFVEVYITKDALEQIN 212 VLLSRFFPGAVELVKYIGHTVUNSGSTLLLSKIQAILQKEENA 212 VLLSRFFPGAVELVKYIGHTVUNSGCALFVQINUSSGGEKE 200 LLILLIDKALAGINISKSIFHTLIGSVGIG-FVLLVSVGVEGEKEA 200 LLILLIDKALAGINISKSIFHTLIGSVGIG-FVLLVSVGVEGEKEA 200 LLILLIDKALAGINISKSIFHTLISSGFTVEVSISSTLEFT LLIMLPOQUGVINTGFTIPHULGNSIGASVLLILVEDIIQKSEIV 183 LLVMAKFFTANSLISVIGFPHILVNSIGLAEVUEVSSVREKEKF 199 LLILLARPFDALKISVIGFPHILVNSIGAALFNENSISSTREKEF 199 LLIMLPOQUGVINTGFTPHVINSIGAALFNENSISSTREEDI 213 ILVARAVAVELVETALPHUNANSVGAAMFHTLLDKRAMFEKZ 204 LLIARFPDDALHUSNITAPHUVINTVGAALFMRLLDKRAMFEKZ 204 LLIARFPDDALHUSNITAPHULNSIGAALFNSIKDKAMFFKL 201 LLISSUFPAAKEUNGIAPHULNSIGAAFFTIINSIGTAFFKL 201 LLISSUFPAAKEUNGIAPHULNSIGAAFFTSIISISKQEEQU 215 LLFFSHDAALKISFIALPHIINSIGGIFISIISISKQEEBU 214 VGFFGGWEUKMIVIPHUINSIGSTLFIALKKTISDKEESQL 208 LLIARAKFPAUELKUSTIAPHIINSIGGIFISIISISKEEPEK 201 LLISAKFPHAVELKUSTIAPHINSIGSTLFIALKKTISDKEESQL 208 LLISAKFFYBAUELVSTIAPHINSIGSTLFIALKKTISDKEEQU 215 LLIFSVBGALKUSFIAPHIINSIGGIFISIISISKGELEFIA 214 VGFFGGWEUKMIVIPHUINSIGSTLFIALKKTISDKEEQU 208 LLVAKFFEGSIALVSTIAPHUINSIGSTLFIALKTISDKEESQL 208 LLVAKFFEGSIALVSTIAPHULNSIGATFIIINSIGTIFFHSSILDREFTEE 201 LVINDENEUNGUVIPHUNSIGIFIFALAKTISDKITPEKEZ 201 LVINDENEUNGUVIPHUNSIGSTLFIALKTISDKITPEKEZ 201 LVINDENEUNGUVIPHUNSIGSTLFIALKTISDKITPEKEZ 201 LVINDENEUNGUVIPHUNSIGSTLFIALKTISDKITPEKEZ 201 LVINDENEUNGUVIPHUNSIGSTLFIALKTISDKITPEKEZ 201 LVINDENEUNGUVIPHUNSIGSTLFIALKTISDKITPEKEZ 201 LVINDENEUNGUVIPHUNSIGSTLFIALKTISDKEENGUVIPHUNGUVIPHUNSIGSTLFIALKTISDKEENGUVIPHUNGUVIPHUNSIGSTLFIALKTISDKEENGUVIPHUNGUVI	
lyts Bsub 1770001 crC0378 Ctet 28204044 FW00759 Fnuc 27886700 BA 0550 Bant 21397925 Ddes2533 Ddes 23475675 yehu Ec 26109167 FW0220 Fnuc 19703565 Gmet2026 Gmet 23055155 O81642 oihe 2777325 Oceni039 ocen 23037886 Rrub1268 Rrub 2296664 Scoll Sepi 27316477 gbs1052 Saga 23095505 gbs180 Saga 25010255 TTE0871 Tten 20515869 G2292 Vcho 11278601 y3885 Tpes 22127756 ther Rsth 7474584		RYFTRRY 3     TPRIALVGIGHESLOW       RKHSKKT 1     NPKLAFITTITLELLEN       RFIKSKD     NFKLAFITTITLELLEN       RFIKSKD     NFKLAFLTITLELLEN       RFIKSKD     NFKLAFLGUVENLSN       SILIRKA     OMRLAANLGUVESTUH       SLIRKG 6     NFITAGAVTFVALEWO       RFIKNKD     RHRVGLIGGMUCESTUH       RFIKNKG 1     DPAVAFITGVANETLOW       QFSDDR 2     SPNKALFIGITCIAELVOM       RILSKR 3     TAGVAAITGLGMEFIQ       SVLIRRG 6     SPLTAGAITCIAELVOM       FILRRAG 6     NPLAGUTUTFAELMOM       NULRRN 3     SIKKABIGATUTAELVEND       NULRKN 4     STSQVILISIIAESIQ       KYLLKKG 6     SPSUFVAVILISIAESIQ       KYLLKKG 6     SPSUFVAVILISIAESIQ       KYLKRN 1     STSQVILISIAESIQ       KYKLKK 2     WEASFVAGTICAEVLOM       SDRKKV0     SPSUFVAVILIAELVANGULUAEULOM       KYKLKK     QLUVAGITUAEULOM       KYKLKK     QLUVAGITUAEULOM	MIII MGL MGL MIL MIL MGL MIL MGL MGL MGL MGL MGL ML MGL MGC ML MIL MIL MML MMI MMI	LLIAAKPFSDAWELVSNIGIPHILINGGGPIFLSIQAIIRKEEQA 215\LytS LLILSKFPNHALQLEKIFALPHIINSVGYFVFVYITKDALEQIN 212 LILSKFPNHALQLEKIFALPHIINSVGYFVFVYITKDALEQIN 212 LLILKOKILAQNIVASYTVYMYPMYSNSGSVLILVSUIQKSELI 200 LLLIVEDCISIVKTIAIPHILVNSFGSFILLSKIQAILQKEENA 212 VLLSKFPFQAVELVKYIGHMILVNSFGALFVQINLVSKGERER 200 ILAIARPYEDAVRLVSNIAAPHMVINTVGAALFWRILLDKRAMFEKY 209 VIVMAPTTALGIDIVSKIGIPHIL-GSVCIG-FIVLVQSVEGKER 200 LLILLOKILQQUIVSTGAPHMVINSIGLAFVCIULVSKGERER 200 LLILLOKILQQUIVSTGIPHILVNSIGLAFVCIVLVSKFRERF 199 LLILLOKIQQUIVNTGIPHILVNSIGLAFVCIVLVSKFRERF 214 LLVARAPTAAVSLVSVIGFPHILVNSIGLAFVELVSVFREKEFF 199 LLIALRPDQUGVQIVNTGIPHVINSIGLAFVELVSKFREKEFF 199 LLIALRPDQUGVQIVNTGIPHVINSIGLAFVKILDKRAMFEKY 204 LLIARPADAHLVSTIAPHVINSVGAAFMKILDKRAMFEKY 204 LLIARPADAHLVSTIAPHVINSVGAAFMKILDKRAMFEKY 204 LLIARFDDAHLUSSFIAPHVINSIGAAFMKILDKRAMFEKY 204 LLIAFSHDAHLVSFIAPHVINSIGAAFMKILDKRAMFEKY 204 LLIAFSHDAHLVSFIAPHVINSIGAAFKKILDKRAMFEKY 204 LLIAFSHDAHLVSFIAPHVINSIGAAFKKILDKRAMFEKY 204 LLIFSSHDAVELVSFIAPHILSSGSTEFIALKYLSNESQL 205 LLIFSSHDAVELVKIGIPHVINSIGAAFKKIRSFENGEN 215 LLIFSSHDAKUVVINTYFIALSGSTEFIALKKYLSNESQL 208 LATARFFHQAVELVKIGIPHMILMSGAAFKALKYLSNESQL 208 LLATARFPHQAVELVKIGIPHVINAGGIFFIKIKSFPDENS 201 LLVVAKFFESYSIAUSTIAPHVINTYFIAALFKSLINKSFPDENS 201 LLVVAKFFESYSIAUSTIAPHVINTYFIAALFKSLINKSFPDENS 201 LLVVAKFFESYSIAUSTIAPHVINTYFIAALFKSLINKSFPDENS 206 LATARFFHQAVELVKIGIPHVVNAGGIFFIKIKSFPDENS 201 LLVVAKFFESYSIAUSTIAPHVINTYFIKISGAAFKSLINKSFPDENS 201 LLVVAKFFESYSIAUSTIAPHVINTYFIKISGAFFILLSKSFPDAFFIKS 201 LLVVAKFFESYSIAUSTIAPHVINTYFIKISGAFFIKISFFIKISFFIKISFFIKS 206 LLVVAKFFESYSIAUSTIAPHVINTYFIKISGAFFIKISFFIKISFFIKISFFIKISFFIKISFFIKS 206	
lyts Bsub 1770001 CrC0378 Ctet 28204044 FW00759 Fnuc 27886700 BA 0550 Bant 21397925 Ddes2593 Ddes 23475675 yehl Ec 26248504 ypdA Ec 26109167 FN0220 Fnuc 19703565 Gmet2026 Gmet 23055156 OB1642 Oihe 2777325 Ocen1059 Ocen 23037886 Rrub1268 Rrub 22966664 yehl Styp 16420695 S02822 Sone 24374352 lyts Saur 21203400 SE2011 Sepi 27316477 gbs0182 Saga 25010255 TTE0871 Then 20515869 C82292 Vcho 11278601 y3885 Tpes 22127756 yhck Bsub 747584		RYFTKRY 3 TPRIALVGIGHESLQ KISKSKI 1 NPKLAFITITIAEILEN KRISKKI NFKLAFITITIAEILEN KISKSKI NFKLAFITITAEILEN KIFKGD NFLAANLGLVGETVH SILIRKG 0 NPITAGAVTFVAEWUQ UKIVKSM ODWLAANLGLVGETVH KIKKGQ 1 DPAVAFITQGVAETLQ QFFSDDR 2 SPNKALFIGIFPPLL HSYSKR 3 TAGVAAITGLGMEFIQ RILVSKG 6 QPLVVGLVTLFAEVQO SVLIRG 6 SPLTAGAITCIAELUQ DLAKEN 3 SIKKGAITGLATETIQ NYLKRNY 3 RVLDAWVGFGHELLQ DKLKENH 3 SIKKGAITGLEGGUQ NYLKENY 3 SVLNATTEIQ NYLKENK 6 SPSITAVITAEILQ DKLKENH 3 SIKKGAITGLGGEGUQ KIYKGEY 2 WEIXFAKGIFGEGUQ SRIKKVQ 1 KHAFYLLITINSLISF5	MIII MIII MMIII MMII MII MMII MMII MMI	LLMARPPSDAVELVSKIGIPHILINGTGPTPLSIOAIIRKEEQA 215\LytS LLISKPPNRALQLEKIPALPHIIINGTGPTPLSIOAIIRKEEQA 215\LytS LLISKPPNRALQLEKIPALPHIINSVGPTVEVYITKDALEQIN 212 LLISKPPNRALQLEKIPALPHIINSVGPTVEVYITKDALEQIN 212 LLILKOPILAQNIVARYVPNYIGNALFVQIKUVSKGERER 200 LLAIRPYEDAVELVSVIGPHILVNSFGSFILLSKIQAIIRQSEENA 212 VILASRPPPQAVELVSVIGPHILVNSFGGERIALVQIKUVSKGERER 200 LLILKAPYEDAVELVSVIGPHILVNSIGALFVQIKUVSKGERER 200 LLILKAPIEDAVELVSVIGPHILVNSIGAEVLILVEDIIGKSEIV 183 LLVMAKPFTANSLSVIGPHILVNSIGAEVLILVEDIIGKSEIV 183 LLVMAKPFTANSLSVIGPHILVNSIGLAFVIKUSKGERER 199 LIILIACHIVSSVIGPHILVNSIGLAFVEN LIVTSVIGOUVNIG 100 LILIACPHDALEUSNIGPHILVNSIGLAFVEN LIVTSVSCLQLVKLFIPHVINSIGLAFVEN LILKAPPDAAKUSNISSIAPHUVNSIGAAMPHTLLAKRETEGU 216 LILIACPFDAAKUSNISAIAPHVINTVGAALFVKILDKRAFFEX 204 LLILACPFDAAKUSNISAIAPHVINTVGAALFVKILDKRAFFEX 204 LLISSVGFAMEUNQI 201 LLISAPHVLNSISIAANFMILSSIISIKGEMI 201 LIJFSNIHATIVSFISHPHILNSIGSTEFIALKTVISHEEM 215 LLISSVGFAMEUNQI 201 LLISSVGFAMEUNQI 201 LVVAKFFEQSVALUSSI 201 LLINGKFFAMEUNANI 201 LVVAKFFEQSVALUSSI 201 LLVAKFFEQSVALUSSTAPHILNSIGSTEFIALKTVLSNESQI 208 LLVAKFFEQSVALUSSTAPHVINTISIAGGTEFIALKTVLSNESQI 208 LLVAKFFELVELMEN 201 LVVAKFFERVELVEN 201 LVVAKFFE	
lyts Bsub 1770001 CrC01978 Ctet 28204044 FW0759 Fnuc 27886700 BA 0550 Bant 21397925 Ddes2533 Ddes 23475675 yehU Ec 26248504 ypdA Ec 26109167 FN0220 Fnuc 19703565 Gmet2026 Gmet 23055156 0061029 Coine 22777325 Ocen1039 Coine 23037886 Rrub1228 Frub 22966664 yehU Styp 16420695 S02822 Sone 24374352 lyts Saur 21203400 SE2011 Sepi 27316477 gbs1052 Saga 23095505 gbs1080 Saga 23015859 gbs1080 Saga 23015859 gbs1080 Saga 23015859 gbs1080 Saga 23015859 gbs1080 Saga 23015859 gbs1080 Saga 23015859 gbs1080 Saga 230125505 gbs1080 Saga 230125505 gbs1080 Saga 230125505 gbs1080 Saga 230125505 gbs1080 Saga 231756 yhck Bsub 7474584 S01500 Sone 24347254		RYFTRRY 3 TPRIALVGIGHESLOW RISKSKI 1 NPKLAFITTITLELLEN RFIKSKD NFLAFTTITLELLEN RFIKSKD NFLAFNETITLELLEN RFIKSKD NFLAFNLELVCETUH SILIRRG 6 NFITAGAVTFVALEWOY HRLKRG 1 DPAVAFITGVAVETLOM OFFSDDR 2 SPNKALFIGITENLES RIJKRK 6 SPLTGATTGUAVETLOM OFFSDR 2 SINKALFIGITENLES SVLIRG 6 SPLTGATTGLAEUOP FYLRRG 6 SPLTGATTGLAEUOP SVLIRG 6 SPLTAGATTGLAEUOP NTLRKN 3 SIKKALIGIVMENION HQTIKON 3 SIKKALIGIVMENION NTLRKN 3 SISVALIGIVMENION NTLRKN 6 SPSIVFAVILFAEUOP TYLLRKG 6 SPSIVFAVILFAEUOP	MIII MIII MGLI MIII MIII MIII MIII MGLI MMIII MIII	LLIAAKPPSDAWELVSKIGIPHILINGTGSPIPLSIQAIIRKEEQA 215\LytS LLILSKPPNHALQLEKIPALPHIINSUGYPVVYITKDALEQIN 212 LILSKPPNHALQLEKIPALPHIINSUGYPVVYITKDALEQIN 212 LILSKPPNHALQLEKIPALPHIINSUGYPVVYITKDALEQIN 212 ULLINGNUTSPUTAPHIVNSTGSFILLSKIQAILQKEENA 212 VILLSKPPPGAVELVKYIGHPHIVNSIGALFVQIHUNSKGERER 200 LLAIARPYEDAVELVSIGAPHIVNSIGASVLILVVSKGERER 200 LILLIDAQNUTSPITPWLANSIGASVLILVVEDIQKSEIV 183 LUVMAPFTALGIDIVSKIGIPHIL-GSVCIG-FIVLLVSVEGEKEA 200 LILLIAQNITSPITPWLANSIGASVLILVVEDIQKSEIV 183 LUVMAPFTALGIDIVSKIGIPHIL-GSVCIG-FIVLLVSVEGEKEA 200 LILLIAQNITSPITPWLANSIGASVLILVEDIQKSEIV 183 LUVMAPFTANSLVSVIGPPHILVNSIGALFVFLVSSVEGEKER 199 LILLARPTQUGVINTIGHPWLINSIGASVLILVEDIQKSEIV 213 LILARPPDAKUSVIGPPHILVNSIGAAFMTILLDKRAMFEK 204 LLILARPPDDALHUSNIAAPMVTNTVGAALFMRILLDKRAMFEK 204 LLIARPPDDALHUSNIAAPMVTNTVGAALFMRILLDKRAMFEK 204 LLIAFSPDAAKUNQIAPPHILVNSIGAAFMTILLDKRAMFEK 204 LLIFSSBOAAKUNGIAPPHILVNSIGAAFMTILDKRAMFEK 204 LLIFSSBOAAKUNGIAPPHILVNSIGAAFMTILDKRAMFEK 204 LLIFSSBOAAKUNGIAPHILNSIGGIFISIISSQKEEMA 214 VGIFFGWEUVKRIVIPMHINSIGGIFISIISSQKEEMA 214 VGIFFGWEUVKRIVIPMIINSUGAAFMKILDRATIFEK 201 LLATAKPFHQAVEUVKIGIFIPMIINSUGAIFMSILDRATIFEK 201 LLATAKPFHQAVEUVKIGIFIPMIINSUGAAFMSILDRATIFEK 201 LLAVARFPHQAVEUVKIGIFIPMIINSUGAAFMSILDRATIFEK 201 LAVARFPHQAVEUVKIGIFIPMIINTIGAAMMKILDRATIFEK 201 LAVARFPHQAVEUVKIGIFIPMIINTIGAAFMKILDRATIFEK 201 LAVARFPHQAVEUVKIGIFIPMIINTIGAAFMSILDRATIFEK 201 LAVARFPHQAVEUVKIGIFIPMIINTIGAAFMSILDRATIFEK 201 LAVARFPHQAVEUVKIGIFIPMIINTIGAAFMSILDRATIFEK 201 HAVARFPHQAVEUVKIGIFIPMIINTIGAAFMSILDRATIFEK 201 HAVARFPHQAVEUVKIGIFIPMIINTIGAAFMSILDRATIFEK 201 HAVARFPHQAVEUVKIGIFIPMIINTIGAAFMSILDRATIFEK 201 HAVARFPHQAVEUVIKIGIFIPMIINTIGAAFMSILDRATIFEK 201 HAVARFPHQAVEUVIKIGIFIPMIINTIGAAFMSILDRATIFEK 201 HAVARFPHQAVEUVIKIGIFIPMIINTGAAFMSILDRATIFEK 201 HAVARFPHQAVEUVIKITIFIMITIGAAFMSILJVKIGHIFIFEN 201 HAVARFPHQAVEUVAITANITATIFIFMIFIFIFIFIFIFIFIFIFIFIFIFIFIFIFIFIF	
lyts Bsub 1770001 CrC0378 Ctet 28204044 FW00759 Fnuc 27886700 BA 0550 Bant 21397925 Ddes2593 Ddes 23475675 yehl Ec 26109167 FW0220 Fnuc 19703565 Gmet2026 Gmet 23055155 OB1642 Oihe 2777325 Ocen1039 Ocen 23037886 Rrub1268 Rrub 22966664 yehl Styp 1642065 S02822 Sone 24374352 lyts Saur 21203400 S82011 Sepi 27316477 gbs1052 Saga 25010255 TTE0871 Ttem 20515869 C82292 Vcho 11278601 yhck Bsub 7474584 S01500 Sone 24347254 Rsph2038 Rsph 22958458 S01500 Sone 24347254		RYFTRRY 3 TPRIALVGIGHESLOW RISKSKT 1 NPKLAFTTITLELLEN RFIKSKD KIFFGFLACVVENLSJ VIFKKYN 3 TPKFSAVLSVFIVSLEB WRVKESM DWRLAAMLGLVGETVH SILIRG 6 NFITAGAVTFVAEWUG LKIPKAQ RRVGIIGGMCLEFITF RFIKNKD RIFFGFLACTIENLSD RIFFGFLACTIENLSD VFSDRR 2 SPNKALFIGIFFPVLH HSYSKR 3 TAGVAATGGAGEFIQ SVLIRG 6 SPLTAGATCIAELVGD VSULRG 6 SPLTAGATCIAELVGD VSULRG 6 SPLTAGATCIAELVGD NTLRKNY 3 RVLDAWVGFGHEILGD NTLRKNY 3 RVLDAWVGFGHEILGD NYLKRNY 3 RVLDAWVGFGHEIGGUQD TILLRKG 6 SPSIVPAVTIAEND LYKGEY 2 WEYAFNAGFIGEGUQD TILLTKNK 6 SPSIVPAVTIAEND SRKKVQ 1 KHAFYLLITINSLISF5 RIYKRAH 3 ISGTEVISFALVILLI VLRRHG 1 GFTCLULLAVVISLU	MIII MMIII MMIII MMIII MMIII MMIII MMII MMII MMII MMIII MMIII MMIII MMIII MMIII MMIII MMIII MMIII MMIII MMIII MMIII	LLIAAKPFSDAWELVSNIGIPHILINGGGPIFLSIQAIIRKEEQA 215\LytS LLISKFFNHALQLEKIFALPHIINSVGYFVEVYITKDALEQIN 212 LIISKFFNHALQLEKIFALPHIINSVGYFVEVYITKDALEQIN 212 LIISKFFNHALQLEKIFALPHIINSVGYFVEVYITKDALEQIN 212 ULISKFFPNALQLEKIFALPHILVNSFGSFILLSMIQAILQKEENA 212 VLUASRFFPAQVEVLKVIGHMHIVVNSGAALFVGILUVSKGRERF 200 ILAIARPYEDAVRLVSNIAAPMVTNTVGAALFVGILUVSKGRERF 200 ILLILDKILQNITSSYTFPMULMNSGASVLILVEGVIQKSEVIG 200 ULVANPTTALGIDIVSKIGIPHIL-GSVCIG-FIVLVQSVEGKER 200 LLULARPTAAVSLVSVIGFPHILVNSIGLALFVGILUKSKGRERF 214 LLVAKFFTAAVSLVSVIGFPHILVNSIGLALFVELVSVFREKEFF 199 LLVAKFFTAAVSLVSVIGFPHILVNSIGLALFVELVSVFREKEFF 199 LLVARPAVNAVHLVETIAPHUVNSIGALFFRILLAKINETEGE 214 ILVFSVSGQLVKLIFIPHMILNSFGTWVFNSISAYLHQEQL 213 ILVARPAVNAVHUETTAPHUVNSUGALFVRILLOKRAFFEKY 204 ILLIARFDDALHUSVIAPHVINVSUGALFVRILLOKRAFFEKY 204 ILLISFSHDAXVDUSJGAPHUINSIGAIFVKILLOKRAFFEKY 204 ILLISFSHDAXVDUSJGSFISHINTIGGGFISISISKQEEQM 215 ILLFSSHDAXVDUSJGSFISHINTIGGGFISISISKQEEQM 215 ILLFSSHDAAVELVKIGIPHMIINSIGALFVRILLOKRAFFEKY 204 ILATARPAVANUSFISAPHIINSUGALFVRILLOKRAFFEKY 204 ILATARPHQAVELVKIGTPHIINSIGAIFISISTEVERGEN 215 ILLFSSHDAMELVNGIAPHIINSIGAIFISISSISKIELALKYEEQM 215 ILLFSSHDAAVELVEFIAPHIINSIGAIFISISSEVENIKSIENESEN 208 ILATAKFFHQAVELVKIGFPHIINSIGAIFISISSEVENIKSIENESEN 208 ILATAKFFHQAVELVKIGFPHILINSIGAIFISISSEVENIKSIENESEN 208 ILAVAKFFGSXLVSTIAPHILANVINTIGAAFMKILDRAAFFEKY 206 FLDHEVKKYVPHILISIGAGFISIISISTEVENESEN 208 ILAVAKFFGSXLVSTIAPHILANVINTIGAAFMKILDRAAFFEEDEN 198 ILVVAKFFESSILVSTIAPHILANVINTIGAAFMKILDRAAFFESSIENES FLOGUNVFFUNDINTISIAGENESIISIDENTIFFESSIENES ILVVAKFFESSILVSTIAFFILLIFUNTISIAGENESIISIDENAALFFESSIENES ILVVAKFFESSILVSTIAFFILLIFUNTISIAGENESIISIDENAALFFESSIENES ILVVAKFFESSILVSTIAFFILLIFUNTISIAGENESIISIDENAALFFESSIENESENESENES ILVVAKFFESSILVSTIASSIAPHILANVINTISIAGENESIISIDENAALFFESSIENESENESENESENESEN ILVVAKFFESSILVSTIASSIAPHILANVINTISIAGENESIISIDENAALFFESSIENESENESENESENESENESENESENESENESENES	
lyts Bsub 1770001 CrC01978 Ctet 28204044 FNV0759 Fnuc 27886700 BA 0550 Bant 21397925 Ddes2593 Ddes 23475675 yehU Ec 26248504 ypdA Ec 26109167 FN0220 Fnuc 19703565 Gmet2026 Gmet 23055156 OB1642 Oihe 22777325 Ocen109_Ocen 23037886 Rrub1266 Rrub 22966664 yehU Styp 16420695 S02822 Sone 24374352 lyts Saur 21203400 SE2011 sepi 27316477 gbs1052 Saga 23095505 gbs1080 Saga 25010255 gbs1080 Saga 23095505 gbs1080 Saga 23095555 gbs1080 Saga 23095555 gbs1080 Saga 23095555 gbs1080 Saga 23012555 gbs1080 Saga 23012555 gbs1080 Saga 24374354 S01500 Sone 24374254 Rsph2038 Rsph 22558458 SAV0744 Saur 15923736 C42394 Cace 15926638		RYFTRRY 3 TPRIALVGIGHESLOW RYFTRRY 3 TPRIAFTTITAEILER RYIKSKD NFKLAFTTITAEILER RYIKSKD NFKLAFTTITAEILER RYIKSKD NFKLAFNELGVGETVHS SILIRRG 6 NFITAGAVTFVAEWUG QFFSDR 2 NFNVGIIGGHCETITF RYIKNKD RIFFGFLACTIENLS RYFTRR 6 SPLTAGATTIGHEFYU QFFSDR 2 SPNKALFIGIFPPVLH HSYSKR 3 TAGVAATTGLGHEFIG RYLVSKG 6 SPLTAGATTCIAELUG NULVCUTFAEVG QFFIRAG 6 NFLUCCUTFAETUG NULRKN 3 SILKGAIIGAITEIQ NULRKN 3 SUKGAIIGAITEIQ NULRKN 3 SUKGAIIGAITEIQ NULRKN 3 SUKGAIIGAITEIQ NULRKK 6 SPSITFAVTLAELUG DKLKENH 3 SUSULISIIAESIG DKLKKN 3 IGTIVLIAELUG SRRKKUQ 1 KHAFYLLIITNSLISFS RHYKAH 3 IGTTUVLIAEULG SRRKKUQ 1 GHTGULLIAWISLEP FL	MIII MIII MIM MGL MII MII MII MGL MM MGL MM MII MM MII MM MII MM MIIII MM MIIII MIIIII MIIII MIIII MIIII MIIIII MIIII MIIII MIIII MIIII MIIII MIIII M MIIIII MIIII MIIII MIIII MIIIII MIIII MIIII MIIII MIIIII MIIII MIIII MIIII MIIII MIIIII MIIIII MIIII MIIII MIIIII MIIIII MIIIIII	LLMARPPSDAWELVSKIGIPHILINGTGSPIPLSIGAIIRKEEQA 215\LytS LLIAKRPPNALQLEKIPALPHIINSUGYPVVYITKDALEQIN 212 LIIKKEPNNALQLEKIPALPHIINSUGYPVVYITKDALEQIN 212 LIIKKEPNALQUKKIPALPHIINSUGSVLILVERUIQKSELI 200 ILLIAGNUARPYEDAVELVYIGHPMIVNSUGALFVQINUSSGREFA 200 ILLIARPPEDAVELVYIGIPHILOSUGIAFVQINUSSGREFA 200 ILLILDRILAQNIVTS'YIPMVLMNSIGASVLILIVEDIIQKSELV 183 LUVAKPFTAAVSLVSVIGPPHILVNSIGALFVQINUSSGREFA 200 ILLILLIDRILAQNIVTS'YIPMVLMNSIGASVLILIVEDIIQKSELV 183 LUVAKPFTAAVSLVSVIGPPHILVNSIGALAFVELVSVYREKERF 199 ILLIARPPEDAVELVSVIGPPHILVNSIGALAFVELVSVYREKERF 199 ILLIARPPEDAVELVSVIGPPHILVNSIGALAFVELVSVYREKERF 199 ILLIARPPEDAVELVSVIGPPHILVNSIGALFVSINDESELV 123 ILLIARPPEDAVELVSVIGPPHILVNSIGALFVSINDESELV 206 ILLIARPPEDAVELVSVIGPPHILVNSIGALFVSINDESELV 206 ILLIARPPEDAVELVSVIGPPHILVNSIGALFVSINDESELV 201 ILLIARPPEDAVELVSVIGPPHILVNSIGALFVSINDESELV 201 ILLIARPPEDAVELVSVIGPPHILVNSIGALFVSINDESEV 206 ILLIARPPEDAVELVSVIGPPHILVSIGAFVSISISISISSUKEEN 215 ILLFSVENGALEVSFIZHPHILNSIGGIFSISISSUKEELA 214 VGIFFGWELVKKIVIPMIINSIGSTEFIALKTVISNERSGI 208 ILLVAKPPEQSVALVSTAAPHINSIGGIFSIINSIGHVERSILDRAFFES 206 ILLVAKPPEQSVALVSTAPHILNSIGSTEFIALKTVISNERSGI 206 ILVAKPPEQSVALVSTAPHILNSIGSTEFIALKTVISNERSGI 208 ILVAKPPEGSVALVSTAPHILNSIGSTEFIALKTVISNERSGI 206 ILVAKPPEGSVALVSTAPHILNSIGSTEFIALKTVISNERSGI 206 ILVAKPPEGSVALVSTAPHILNSIGSTEFIALKTVISNERSGI 206 ILVAKPPEGSVALVSTAPHILNSIGGUSITIPHES 206 ILVAKPPEGSVALVSTAPHILNSIGGUSITIPHES 206 ILVAKPPEGSVALVSTAPHILNSIGGUSITIPHES 206 ILVAKPPEGSVALVSTAPHILNSIGGUSITIPHES 206 ILVAKPPEGSVALVSTAPHILISIGGUSITIPHES 206 ILVAKPPEGSVALVSTAPHILVISIGGUSITIPHES 206 ILVAKPPEGSVALVSTAPHILVISIGGUSITIPHES 206 ILVAKPFEGSVALVSTAPHILVISIGUS 200 ILVAKPFEGSVALVSTAPHILVISIGUS 200 ILVAKPFEGSVALVSTAPHILVISIGUS 200 ISOVAVSPONSKITPILIJEPHILVSVISTAPHILVSVISTAPHILVSVISTAPHILVSVISTAPHILVISIGUS 206 ISOVAVSPONSKITPILIJESLITASSUSTAPHILVISIGUS 200 ISOVAVSPONSKITPILIJESLITASSUSTAPHILVSVISTANUSVISTAPHILVSVISTANUSVISTANUSVISTAPHILVISIS	
lyts Bsub 1770001 CrC0378 Ctet 28204044 FW00759 Fnuc 27886700 BA 0550 Bant 21397925 Ddes2533 Ddes 23475675 yehU Ec 26109167 FW0220 Fnuc 19703565 Gmet2026 Gmet 23055155 O81642 oihe 2777325 Ocen1039 ocen 2337886 Rrub1268 Rrub 2296664 yehU Styp 16420695 S02872 Sone 24374352 lyts Saur 21203400 S82011 Segi 27316477 gbs1052 Saga 2500505 gbs180 Saga 25010255 TTE0871 Tten 20515869 ghc82 Types 22127756 yhcK Bsub 71474584 S01500 Sone 24347254 Rsph2038 Rsph 22598458 S0102644 Cace 1502638		RYFTRRY 3 TPRIALVGIGHESLOW RKSKSKT 1 NPKLAFTTITLELLEN RFIKSKD KIFFGFLACVVENLSS WRVKESM DWRLAAMLGLVGETVH SILIRG 6 NPITAGAVTFVALEWUD UNRVGILGGMCLCETIT RFIKNKD RIFFGFLACTIENLSS RIFFGFLACTIENLSS SVLIRG 6 SPLTAGATCIENLSS SVLIRG 6 SPLTAGATCIENLSS SVLIRG 6 SPLTAGATCIENLSS SVLIRG 6 SPLTAGATCIENLSS SVLIRG 6 SPLTAGATCIENLSS SILKGAN 16 SFLTAGATCIENLSS SILKGAN 16 SFSUFAVTIFAEVUD LQAQRK 3 SIAKSANIGIVMENIG NULKENN 3 SIAKSANIGITEID NILKENN 3 SIAKSANIGITEID NILKENN 4 SPSUFAGTIENESUS SRKKVQ 1 KHAFYLLIITNSLISSS RHYRAM 1 GGTFUSFAFLVHLVN VLRRLG 1 GFFCLULLAVMISLEF STMUTAS 2 KRWFYKUTTLLNINV VIRKKX 2 KRWFYKUTTLLNINV	MIII MIII MIM MGL MIN MIN MIN MIN MGL MM MIN MIN MIN MIN MIN MIN MIN MIN MIN	LLIAAKPFSDAWELVSKIGIPHILINGTGSPIPLSIQAIRKEEQA 215\LytS LLILSKFPNHALQLEKIPALPHIINSVGYPVVYITKDALEQIN 212 LILSKFPNHALQLEKIPALPHIINSVGYPVVYITKDALEQIN 212 LILIKUPALQNIVASTVYVWYMYNSGSSVLILLVERIUGKSELI 200 ILLILVEGCISIVKTIAIPHILVNSFGSFILLSKIQAILQKEENA 212 VLLSKFPFQAVELVKYIGHMILVNSFGSFILLSKIQAILQKEENA 200 ILLIARPYEDAVELVSNIGAPHMYUNVGAALFMRILLDKRAMPEKY 209 UIVNAPTTALGIDIVSKIGIPHIL-GSVCIG-FULLVGSVEGEKA 200 LILLIARPYEDAVELVSNIGASVULIVEUVSIGKEEN 183 LLVMAKPFTAAVSLVSVIGFPHILVNSIGASVLILVEUVIGKSEIN 183 LLVMAKPFTAAVSUSVIGFPHILVNSIGALFVRILLSKREFF 199 LLILINCPQQUGVUNTGIPHULVNSIGALFVRILLSKREFF 214 ILVFSVGGQUKKLFIPHVLNNSIGALFVRILLDKRAMPEKY 204 ILLIARPPDAHEUSVIGPPHILVNSIGALFVRILLDKRAMPEKY 204 ILLIARPPDAHEUSVIGPPHILVNSIGALFVRILLDKRAMPEKY 204 ILLIARPPDAHEUSVIGPPHULVNSIGALFVRILLDKRAMPEKY 204 ILLIARPPDAHEUSVIGPPHULVNSIGALFVRILLDKRAMPEKY 204 ILLIARPPDAHEUSVIGPPHULVNSIGAALFVRILDKRAMPEKY 204 ILLIFSHDKAIVDISLAIPHVINNSGAALFVRILDKKIKEQEQ 215 ILLFFSHDKAIVDISJAPHVVINNGGALFVRILDKRAMPEKY 204 UGIPTGWEUVKMIVIPHIINSGSTIFLAIKKIKSIGKLEEQU 215 ILLFSHDKALKUSFIAPHILSKIGGIFFLAIKSIKRQELPUS 214 VGIPTGWEUVKMIVIPHMILNSIGGTFLAIKKIKSIFDEKEQI 198 ILVVAKFPEGSVALVSIGIPHVINNIGGALFVRILDKRAFEKY 201 ILAVARPPKGSVALVSIGIPHVINISIGGALFVSILLDKRAFEKY 201 ILAVARPPKGSVALVSIGIPHVINISIGGALFVSILLDKRAFEKY 201 ILAVARPPKGSVALVSIGIPHVINISIGGALFVSILLDKRAFEKY 201 ILAVARPPKGSVALVSIGIPHVINISIGGALFVSILLDKRAFEKY 201 ILAVARPPKGVELVKIGIPHVINAGGIGFVINKISLSFREFLEEDE 187 ILVVAKFVEJSVALVSIJIPHILISIGGALFVSILDKRAFEKY 206/ WFPLLPGQHVV-VLSKITLPILLIYLISIISGELSVILVSGFLEENAALFY 185/yhck WFPLLPGQHVV-VLSKITLPILLIYLISIISGELSVIJVSILVSSILSSKEFLEEDE 187 ILMFVOGTSUSUAJIVILPISIISUSSILVSSILVSSILSSTILVSSILSSTILVSESUS 190 ILAVARPVKSTSUSUAJIVILPILINISGILVYKINSKITI 190 IMFKV	
lyts Bsub 1770001 CrC0378 Ctet 28204044 FW00759 Fnuc 27886700 BA 0550 Bant 21397925 Ddes2593 Ddes 23475675 yehU Ec 26248504 ypdA Ec 26109167 FN0220 Fnuc 19703565 Gmet2026 Gmet 23055156 OB1642 Oihe 2777325 Ocen1059 Ocen 23037886 Rrub1268 Rrub 22966664 yehU Styp 16420695 S02822 Sone 24374352 lyts Saur 21203400 S82011 Sepi 27316477 gbs0180 Saga 25010255 TTE0871 Then 20515869 C82292 Vcho 11278601 yhcK Bsub 7474584 S01500 Sone 24347254 Rsph2038 Rsph 22958458 SAV0746 Saur 15923736 CAC3544 Cace 15026638 BA 0523 Bant 21397888		RYFTRRY 3 TPRIALVGIGHESLOW RYFTRRY 3 TPRIARTTITAEILER RFIKSKD KIFFGPLACVVENLSJ YIFKKYN 3 TPKPSAVLSVFIVSLEJ WRVKESM DWRLAAHLGLVGETVH SILIRG 6 NFITAGAVTFVAEWUG URVSLIGMCLETIT RFIKNKD RIFFGPFLACTIENLSD UPFSDR 2 SPNKALFIGIFPVLH HSYSKR 3 TAGVAAITGLGMEFIQ RYLVSRG 6 SPLIVGLVTIFAEVQD SVLIRG 6 SPLIVGLVTIFAEVQD SVLIRG 6 SPLIVGLVTIFAEVQD VLRANG 6 SPLIVGLVTIFAEVQD NYLRRNG 3 SIKSAAITGIATEEIIQ NYLRRNG 3 SIKKGAITGIATEEIIQ NYLRKN 3 RVLDAMVVGFGHEILQD DKLKENU 3 SIKKGAITGIEGUQD ILITRKN 6 SPSITPAVTIFAEIVQD SRRKVQ 1 KHAFYLLITNSLISFS RHYRKAH 3 GGTUVLSFALVILIV VLRRLG 1 GHFGLVLAVHSLPI FL	MIII MMIII MMLI MMLI MMLI MMLI MMLI MML	LLMARPFSDAWELVSKIGIPHLLINGTGSFIFLSIGAIIRKEEQA 215\LytS LLISKFPNRALQLEKIFALPHIINISGSFIFLSIGAIIRKEEQA 215\LytS LLISKFPNRALQLEKIFALPHIINISVGYFVEVYITKDALEQIN 212 LLISKFPNRALQLEKIFALPHIINISVGYFVEVYITKDALEQIN 212 LLILKORILQNIVATYPWYFNNSIGSVLILVEDIIQKSELI 200 LLILKVFQAVELVKYGFMHIVVNSGALFVGINLVSKGRERF 200 LLILKAFFTBAVELVSNIGAPHIVNNSGALFVGINLVSKGRERF 200 LLILKAFFTBAVELVSNIGFMILVSSIGEFILIVEDIIQKSELI 183 LLVMAKFFTAAVSLVSVIGFPHILVNSIGALFVEILVSVEREKERF 199 LLILLACHIVSSVIGFPHILVNSIGALFVEILVEDIIQKSELI 213 LLILKAFFTBAVELVSVIGFPHILVNSIGALFVEILVSUFREKERF 199 LLILLACHIVSSVIGFPHILVNSIGALFVEILVSUFREKERF 206 LLILLACHAWVHLVETIALPHUINSGALFVEILVSUFREKERF 206 LLILLACHAWVHLVETIALPHUNSSIGAIFYSILSKIELSEEQU 213 LLILARFFDDALHLVSNIGAPHULVNSIGALFVEILVSUFREKERF 206 LLILLACFPHAWELVGVISTALATIFTMIKLAINETEGE 206 LLILLACFFDAAKUVSVIGFPHILVNSIGALFVSITOSTISTKQEEQU 215 LLILTSVFNGAQUKKIFLPHMINSIGGIFISIISSIKEERA 214 VGFTGGWELVKNISTGALFVSITIGGIFISISSIKEEQU 215 LLIFSVFNLHHAITUSST ALPHILNSIGGIFISISSIKEEQU 215 LLIFSVFNLHHAITUSST AAPHILNSIGGIFISILSSTOKLEERA 214 VGFTGGWELVKNISTAFMILSNIGGIFISILSSTOKLEERA 214 VGFTGGKELVKNIVIPHMILNSIGGIFISILSTKREEQU 208 LLVAKFFEQSVALVST AAPHILNSIGGIFISILSSTOKLEERA 214 VGFTGGKELVKNIVIPHULSSIGGIFISILSSTOKLEERA 214 VGFTGG	
lyts Bsub 1770001 CrC01978 Ctet 28204044 FNV0759 Fnuc 27886700 BA 0550 Bant 21397925 Ddes2533 Ddes 23475675 yehU Ec 26248504 ypdA Ec 26109167 FN0220 Fnuc 19703565 Gmet2026 Gmet 23055156 OB1642 Oihe 22777325 Ocen1039 Ocen 23037886 Cmet2026 Gmet 2305515 OB1642 Oihe 22777325 Ocen1039 Ocen 23037886 S02822 Sone 24374352 lyts Saur 21203400 SE2011 Sepi 21316477 gbs1052 Saga 23005505 gbs1080 Saga 25010255 Gmet011 Sepi 21316477 gbs1052 Saga 22005505 gbs0180 Saga 25010255 Gmet011 Sepi 21316477 gbs1052 Saga 220125565 gbs0180 Saga 25010255 Gmet011 Sepi 21316477 gbs1052 Saga 220125505 gbs0180 Saga 25010255 Gmet011 Sepi 2137647 gbs055 Saga 22127756 yhck Bsub 7474584 S01500 Sone 24347254 SAV0746 Saur 15923736 CAC3544 Cace 15026638 BA 0523 Bant 21337898 lin2277 Lina 15801341		RYFTRRY 3 TPRIALVGIGHESLOW RISKSKT 1 NPKLAFITTITLELLEN RFIKSKD NFLAFITTITLELLEN RFIKSKD NFLAFNTITLELLEN RFIKSKD NFLAFNLGUVENTHN SILIRRG 6 NFITAGAVTFVALENGO UKLIRKAQ NRVGILGGULCETIT RFIKNKD RIFFGFLACTIENLSN RIFKGFLACTIENLSN RIFKGFLACTIENLSN RIFKGFLACTIENLSN SVLIRRG 6 SPLTAGAITGUAVETLOM QFSDDR 2 SUNKALFIGIFPVLHN HSYSKR 3 TAGVAATTGLACHEFIG VLVSKG 6 SPLTAGAITGUAVETLOM QLAQRKK 3 SIAKSANIGIVMENION HQTIKQN 3 SIAKSANIGIVMENION HQTIKQN 3 SIAKSANIGIVMENION HQTIKQN 3 SIAKSANIGIVMENION HQTIKQN 3 SISVAIGAITEIION NILRKN 3 STSQUILISIAESION NILRKN 4 STSQUILISIAESION SKRKVQ 1 GPTGULLAVNISLIP FL	MIII MMIII MMII MMLI MMLI MMII MMLI MML	LLIAAKPPSDAWELVSKIGIPHILINGTGSPIPLSIQAIIRKEEQA 215\LytS LLILSKPPNHALQLEKIPALPHIINGTGSPIPLSIQAIIRKEEQA 215\LytS LLILSKPPNHALQLEKIPALPHIINSUGYPVVYITKDALEQIN 212 LLILKOKILAGNIVARYVPWYPNYSGSSVLILLVGUIQKSELI 200 ILLIUPGCISIVKTIAIPHILVNSFGSFILLSKIQAIIRQEENA 212 VVLALSRPPPQAVELVKYIGHPHILVNSFGARLAFVQINLVSKGRERR 200 ILAIARPYEDAVELVSIGAPHIVUNSIGALFVQINLVSKGRERR 200 LLILLOKLAQNIVSSITPWLLNSIGASVLILVEDIQKSELV 183 LLVMAKPFTAAVSLVSVIGPPHILVNSIGALFVQINLVSKGRERR 199 LLILLOKLAQNIVSSITPWLNNSIGASVLILVEDIQKSELV 183 LLVMAKPFTAAVSLVSVIGPPHILVNSIGALFVELVSVFREKERF 199 LLILLOKLQUGVQIVNIGIPHILVNSIGALFVELVSVFREKERF 199 LLILLOKLQUGVQIVNIGIPHULVNSIGALFVELVSSVFREKERF 199 LLILARPPDALHUSVIGPPHILVNSIGALFVELVSSVFREKERF 199 LLILARPPDALHUSVIGPPHILVNSIGASVLILVEDIQKSELV 204 LLILARPPDAKUVSVIGPPHILVNSIGASVIIIFTMHIKAINETEGE 214 ILVFSFDKAAVAVLVETAIPHUANSVGAAFMFILLDKRAMFEKY 204 LLILARPPDAAKELVQIAPPHILVNSIGAAFHKILDKRAMFEKY 204 LLIFSFDKAAVDUSLSILAPHVINSIGAIFHSIISKGEEDQ 215 ILLFFSHDKAKIVDISTAPHILNSIGGIFISISTSKQEEDQM 215 ILLFSHDKAKIVJEPHILNSIGGIFISISTSVELERA 214 VGIFFGWELVKKIVIPHMIINSIGGIFISIISTKQEEDQM 215 ILLFSHDKAKIVIGIPHVINSIGAAFHKILDRATFFEX 206/ ILVVAKPFRQSIAUSTIAPHIINSUGAAFHKILDRATFFEX 206/ ILVVAKPFRQSIAUSTAPHIINSUGALFVSILDRATFFEX 206/ LILVVAKPFRQSIAUSKITAPHIINSUGALFVSILDRATFFEX 206/ ILAVAKPFRQSIAUSKITAPHIINSUGALFVSILDRATFFEX 206/ ILAVAKPFRQSIAUSKITAPHIINTIGAAMFMILDRAAFFEX 206/ LICVAKPFRQSIAUSKITAPHIINTIGAAMIKILDRAAFFEX 206/ LFQUAUSPDSEVLAATAIVUNISIGALFVSILDRATFFEX 206/ LFQUAUSPDSEVLAIAIUVINISIGALFVSILDRATFFEX 206/ LFQUAUSPDSEVLAIAIUVINISIGALFVSILDRATFFEX 206/ LFQUAUSPDSEVLAIAIUVINISIGALFVSILSAKITVDINHFSIV 185 LFFN	
lyts Bsub 1770001 CrC0378 Ctet 28204044 FW00759 Fnuc 27886700 BA 0550 Bant 21397925 Ddes2593 Ddes 23475675 yehl Ec 26109167 FN0220 Fnuc 19703565 Gmet2026 Gmet 23055155 OB1642 Oihe 2777325 Ocen1039 Ocen 23037886 Rrub1268 Rrub 22966664 yehl Styp 16420695 S02822 Sone 24374352 lyts Saur 21203400 S82011 Sepi 27316477 gbs1052 Saga 25010255 TTE0871 Ttem 20515869 C82292 Vcho 11278601 yhck Bsub 7474584 S01500 Sone 24347254 Rsph2038 Rsph 22958458 BA 0523 Bant 21397898 BA 0523 Bant 2139788 BA		RYFTRRY 3 TPRIALVGIGHESLOW RIFIKSKD NFKLAFITTITLELIER RFIKSKD NFKLAFITTITLELIER RFIKSKD NFKLAFITITLELIER RFIKSKD NFKLAFNTITLELIER RFIKRKD NFKLAFNTILELUGVETVH SILIRG 6 NFITAGAVTFVAEWUG QFFSDR 2 SPNKALFIGIFPFULN HSVSKR 6 QPLVVGLVTLFAEVUG QFFSDR 2 SPNKALFIGIFFPULN HSVSKR 6 QPLVVGLVTLFAEVUG SVLIRG 6 SPLTGGATTIELIG NULRKN 3 RVLDAWVUGFGELLO UQARK 3 SIKKGAIGGATETIG NULRKN 3 RVLDAWVUGFGELGUG NYLKRK 2 WENAPMAGFIGEGUG YILLRG 6 SPSITPAVILFAEVUG SKRKR 3 SIKKGAIGTTEILG NULKENN 3 RVLDAWVUGFGELLO KYKGEY 2 WENAPMAGFIGEGUG YILLRKG 6 SPSITPAVILAEVUG SRRKKQ 1 KHAFYLLIITNSLISF HYRKAH 3 IGGTEVISFALUNUT ULRKKG 2 KHAFYKWFTLLHINVI SRRKKV 2 KHAFYKWFTLLHINVI SRRKKV 2 NINGWFILLINVVISFILM VURKIKS 2 KRNFYKWFTLLMINV HYRKLS 2 KRNFYKWFTLLMINV SKRLKU 1000000000000000000000000000000000000	MIII MGLI MGLI MGLI MIII MIII MIII MMII MM	LLMARPPSDAWELVSKIGIPHILINGTGPIPLSIOAIIRKEEQA 215\LytS LLISKPPNRALQLEKIPALPHIINSTGPIPLSIOAIIRKEEQA 215\LytS LLISKPPNRALQLEKIPALPHIINSTGPIPLSIOAIIRKEEQA 215 LLISKPPNRALQLEKIPALPHIINSTGPILSKIQAIIRQEENA 212 ULJILKOKILQNIVATIPHIVNSTGSPILLSKIQAIIRQEENA 212 ULJILKOKILQNIVATIPHIVNSTGALAFVQINUVSKGRERK 200 LLILKOKILQNIVSTIPHILVNSTGALAFVQINUVSKGRERK 200 LLILLOKILQNIVSTIPHILSKIGTOFIVIKUSKGRERK 200 LLILLOKILQNIVSTYIPHUNSIGASULLVEDIIQKSEIV 183 LUVMAKPTANSUSVIGPHILVSSIGALFVQINUSKGRERK 200 LLILLOKILQNIVSTYIPHUNSIGASULLVEDIIQKSEIV 201 LILIARPTDAVELVSTGPHILVSSIGLAFVEINSKGRERKE 200 LLILLOKILQNIVSTYIPHUNSIGASULLVEDIIQKSEIV 201 LUVAKPTANSUSVIGPHILVSSIALTIPTMIRLAINETEQE 214 LVFSVGGQULVKLTPHNILNSFGTVYNSISSIATETMILAINETEQE 214 LUVSKAVAVLUSTIAPHUVNNSGAANMIKILAINETEQE 206 LILLARPFDAAKUSNISAILAPHIVNSVGAANMIKILAINETEQE 206 LILLIARPFDAAKUSNISAILAPHIVNSVGAANMIKILAINETERE 206 LILLSGPHOAKUSVIGPHUNNSGAANMIKILISKEGUN 215 LILFSVGHAMEUNQIANGAILMSITTAKISITSIKQEEQM 215 LILFSVDFNQALKUSTSMPHILSSTGATIFITIISIKGEEM 214 VGIPTGWEUVKMIVIPHILSSTGATIFITIISIKGEEM 216 LILARAKPFQAVELVKIGIPHUNNSGGIFHINKIKSITHGEEMQ 208 LLVAKFFEQSVALUST APHIINSIGATIFITIISIKGEDI 198 LVVAKFFEQSVALUSTAAPHIINSIGATIFITIISIKGEDI 206 FLDLBISTEHVINNISIGAANMIKILLARAIFEST 206/ FLDLBISTEHVINNISIGALFINSITIKGEEM 210 LAVAFFEQSVALUSTAAPHIINSIGAMISTIDENKALLF 185 VAKFFEQSVALUSTIAPHIINSIGANSILVIDENKALFFEET 201 LAVAFFEQSVALUSTIAPHIINSIGATIFITIISASUTUNENGU 208 LLVAKFFEQSVALUSTAAPHIINSIGAMISTIDENKILFEET 201 LAVAFFEQSVALUSTAAPHIINSIGAMISTIDENKILFEET 201 LAVAFFEQSVALUSTIAPHIINSIGANSTIDENKILFEET 201 LAVAFFEGSVALUSTIAPHIINTISIAGCHSIIDIKTIESTIAFEET LAVAFFEGSVALUSTAAPHIITIISIAGCHSIIDIKTIESTIAFEET LAVAFFEGSVALUSTAAPHIITIISIAGCHSIIDIKSIFIARAIFE LAVAFFEGSVALUSTAAIUVILVESSUANITTIESANTINNESGEN 190 HFFLLOQONV-VUSKIILPILVIVYGEGAAPHILSASUSSEN 181 LAVAFFEGSVALUSTIAPHIINSIASIIDIKSIFIBASU LAVAFFEGSVALUSTIAPHIINSIASIIDIKSIFIBASU LAVAFFEGSVALUSTIAPHIINSIASIIDIKSIFIBASU LAVAFFEGSVALUSTIAPHIINSIASIIDIKSIFIBASU LAVAFFEGSVALUSTIAPHIINSIASIIJIDENKILFELSEVIB	

**Multiple sequence alignment of the 5TM domains of the 5TMR-LYT family.** Multiple sequence alignment the 5TM domains of the 5TMR-LYT family was constructed as detailed in the legend to Figure 1. The two subgroups 1) LytS and 2) YhcK are shown on the right. The species abbreviations are as shown in Table 1.

neighborhoods containing genes encoding various signaling proteins with GAF [25], histidine kinase and receiver modules. Hence, the YchK-like proteins could be part of a distinct signaling complex that could relay signals, which may be dependent on cell wall composition. A second family of evolutionarily mobile TM domains that was recovered in our search procedure is typified by the conserved N-terminal TM domain, which is found in the *E. coli* sugar response histidine kinase UhpB (Figure 8C and 10). In addition to orthologs of the UhpB protein,

GTM Socondary structure hpp Ec. 15004264 20462 Ec. 15800409 204683 Von. 15601440 VPA0965 Vpar.28900820 RSc0982 Rpc. 17545701 yfah Styp 16765230 yfgr Dc. 26248663 part 12295 part 1229 part 111 Parce 12596778 yrge Ec. 1613007 CG16038 Cglu 19552261 yrger Ec. 26247800 Bcog T2008 Bfun 2299260 all3275 Ana 17230767 CC662 Ccrc 16126867 007116 LSW 13248664 bant 1300 Band 12231742811 Hac11833 Mand 233174281 Horits 1307 Band 233174281 Horits 1307 Band 233174 Horits 1407 Band 233174 Horits 1407 Band 1407 Band 233174 Horits 14

Secondary structure uhpm Ec\_15904264 WEMD625 Vpa-1501440 VFA0965 Vpa-28900820 RSC0962 From 28907 yfed Strp 16755736 STRV503 Eyrp 16755736 STRV503 Eyrp 16755736 STRV503 Eyrp 16755736 JfgF Ec\_26248663 VCC3131 Kcan 21232561 Ddes1160 Ddes 23474249 PA1181 Pace 15397 yg11038 Cglu 19552261 Graft Ec\_26247800 Pcep7208 Pfun 23999260 all2125 Ana 17230767 CC2632 Cccc 16126867 OKF186 ESV 13242656 Reph4145 Reph 22960563 Reu15140 Exet 2347610 HCC176 Dccm 21231210 Statistic 134415 HCC175 Sep 1633025 HCC1760 Tccm 21231210 Statistic 134415 HCC175 Sep 1633025 HCC176 Tccm 27382271 all1672 Sep 1633025 Hdeg1214 Hdeg 23029156 Hdeg1214 Hdeg 23029312 Hdeg1214 Hdeg 23029125 Hdeg1214 Hdeg 23029215 Hdeg1214 Hdeg 2302915 Hdeg1147 Lter 2991

	·····		····	·····
8 1	RLITVIACFFIFSAA <mark>W</mark> F <mark>C</mark> LWSISLHLVERPD-MA	AV <mark>LL</mark> FPF <mark>GLRL</mark> GL <mark>MLQC-PRGYWPVLL</mark> GXEW <mark>L</mark> LIYWLTQAV	- GLTHFSLLMIG-SLLTLLPVA <mark>LIS</mark> RYRHQ	RDWRTLLLQGA <mark>A</mark> LTAAALLQ <mark>S</mark> LPWLWH
71	HWVILLFIVLAWGTG <mark>W</mark> LMTLSFYLTNNGQ-QA	AV <mark>LF</mark> LPQ <mark>GVYL</mark> ALVILL-SRRYWPA <mark>LV</mark> LPPL <mark>L</mark> MMFWLHSE-	<ul> <li>QLLNGYLMLAT-PVISLFPALLAQNFWHRF</li> </ul>	PLYWQRLTLLLAT <mark>V</mark> TAASLLNT <mark>A</mark> LLSPFMSG
3 1	KLFPLIFAFVLYSISOFSLWNVSOYLSHNPL-OA	AYLLFPTGLRLAVYLLARPTYVWVWLLSDMLLAGVILVLLP	DOSSPLMLLM-PWVVWLAAVIAROHWPKL	HLYWOKLLLIVALVLLHTILVGITFTLLAKPLHMG
4 1	CUTTICCI FUMACANECI MUTAVVEUNDE I A	ATTIEDENT DI CIALUM, DEAVMONTVUCEMATETALAMIT	A 1 BOWL THE TARVAST DUTLI TARKY VCDON	PHI NUMCTULT TEAFTNUMAUCE
6 1	HRTSPSLAAFLWGAL <mark>Y</mark> L <mark>L</mark> AAVVSHRLNGPIDMTG	GY <mark>IWLSA<mark>G</mark>ITMAAFMLR-PYREWPG<mark>L</mark>GAAFT<mark>V</mark>GQLVLCAFE</mark>	K 1 NPAHALLFVLDEAGSAALAVA <mark>LV</mark> RLTRVPL	-DGLDFVRAMLAAGT <mark>L</mark> SALLGALP <mark>G</mark> AAWFAWSQDAPF
10 1	NIKIFLLAFCLTVPA <mark>I</mark> L <mark>L</mark> SRLISPRATIDSSYIF	FLAWLPLCVMFSVLFLF-GRRGVAPMVGGMMLTNEWNFHL-	<ul> <li>PLPQAMVLLFCTQFPVLLVCAIVRWQLGAR 4 IPN</li> </ul>	QGIWLRVFWLGLMTP <mark>F</mark> GIKISMHL <mark>A</mark> GHYLAFPVT-ISTFF
11 1	RDKWWGLPLILPSIL <mark>L</mark> P <mark>V</mark> LSSANTYALTSTGNVV	VL <mark>FYLPLAFML</mark> SLMLFF-GWAALPG <mark>IVLAIFW</mark> RRYPOTGLY	E TLSVTMHFIIT-IVLSWGGYRVFSPRRNNV 10 RIF	WOVFCSATLFLVIYO <mark>F</mark> AAFVGMYE <mark>S</mark> KASLMGVMPF
	DEFINIT DI DI DAI T <mark>I DI</mark> DULTEMENTICAMU	THE THE ALL AND A DATE OF		
11 1	KDKWWGLPLFLPSLILP <mark>I</mark> FARINIFARISSGEVF	FLFILFLALMISMMAFF-SWAALPGIALGIFVRKIAELGFI	TLSLIANFIII-IILCWGGIRVFIPRRNNV 6 LIS	QKLFWQIVFFAILFLILFQFAAFVGLLASKENLV-GVMPF
11	MEGLKSVTQGLLTAT <mark>A</mark> Y <mark>C</mark> LAFQLSWHCSLD	DQ <mark>WY</mark> LPA <mark>GL</mark> RIAALLLA-PSRLLPW <mark>IL</mark> MGDV <mark>A</mark> ALLMIRVPA	I 5 NPTWAYASPWILVPAIALVPI <mark>AI</mark> RARYGAV	-HRAGASLIPMLLVT <mark>A</mark> VWGALCTL <mark>G</mark> TNIMLDGPSSAIS
40 2	APVRQRLQLPLAGAI <mark>Y</mark> F <mark>V</mark> AVLCGLQSGIACGLPP	PS <mark>IH</mark> PAA <mark>GF</mark> GFAMLTLR-GGTMLPVLMLADFCAGMTAL	<ul> <li>PPLPALALAAGNTAGIALAWQLMQAKAPQV RIF</li> </ul>	DLPRHAVLFICAAPL <mark>V</mark> QSVTAATA <mark>G</mark> TFSLTLAGIVHHSQA
6 5	PDDRGI WAGTOAGVI. FFLAALACTVI.SPOPATTA	AATWI, DNATI, TAATI, PA-PI, POWPATI, I. CSAVANI, AANPI, V	2 1 SLWMSAAFLPANLSPATFAAML,PINPUPP DFF	ENTHEAVENT WAGAT TPPLICASUCAVI VEVECMCRE
	IDDROLWAGIQAGVL <mark>F</mark> FHARDACIILSKQFAIIA	ARTHDERATDIARDDAA-FDAQWEATDDCORVANDARMADI	J SLAMSARF LFRALSERIF ARADDRINKVKE DFE	r NEHSAV KV ENAGAELT FEEDGASV GAVEVSVFGMGKF
11 /	ALPHPLLHLVSLGLVSF <mark>I</mark> FTLFSLELSQFGTQLA	APLWFPTSIMMVAFYRH-AGRMWPGIALSCSLGNIAASILL	FITSSLNMTWTTINIVEAVVGAV <mark>LLR</mark> KLLPWY	NPLQNLADWLRLALGSAIVPPLLGGVLVVLLTPGDDP
26 5	SLLVALLRQIVFAFA <mark>Y</mark> L <mark>I</mark> AVIGSYSLVYGDSLVA	AI <mark>VW</mark> PSV <mark>GIAV</mark> WWAVTC-RSWKNFA <mark>LI</mark> CGYV <mark>F</mark> LVPAIYLHF	F 6 GVILAGIAHAIAGPGVALVMA <mark>FME</mark> NAQLPE 7 FAP	FSHIRLPGDVFRLLV <mark>A</mark> GIVMVAIS <mark>K</mark> LIVILAYALADLPYS
12 1	FOEGHLI, PNSTATEVI, TTI EVELGAEL PLUHELS	SI. FWPI.NCVMACVFAPY_VWI.NPI.HVVATSVVAMI.VVDATT	P 3 LUSLATNESNM_METUTUALLUAPDKRLCKN	KVEPVSALRIENVCL.
20 1				
201	PLMKKF SQLLASTLAFT VFACFGIVLTKATWGVS	SSIMPANGUIAGLADDG-IQIRERDIAIAVAAGGLAANLWL	L I DESAGLIFSLONVEVSAMSMS <mark>LE</mark> FFEAPL I AVP	NPQGLIRFLIRAVLLPILISALVVQVATHIALGWPA
9 1	IMRRWLMISLLGLFA <mark>L</mark> V <mark>I</mark> AHGMALIYRIQPG-VS	SL <mark>WF</mark> PPS <mark>GVAI</mark> ALT <mark>F</mark> WF-GPCG <mark>II</mark> LTGL <mark>A</mark> SFLMSPFWG	<ul> <li>LHGWERLIALVDIIEPLSAWLLYRRLWKGS RTI</li> </ul>	NTLKDATLFTVSAPL <mark>V</mark> ACATLAII <mark>G</mark> SVCLVAIGKMPLASL
10 1	ERRQTALLLALSFAV <mark>A</mark> A <mark>A</mark> IVFSENLTRTTLG-VA	AA <mark>FW</mark> PAN <mark>AFLA</mark> AG <mark>LLTL-TSVRRLV<mark>LI</mark>AVFV<mark>F</mark>FHVAVCLQA</mark>	3 1 GPLKALVFATVDGLEAGAVWAAFRFWNGA	-PRVRTLRALAQLPAITTPIAAGSALICATVLHFGYGAPF
1.	MI POVORTI COFTAOVTOVI VSMEVEVSPOOTT	TEWWTOTATATAAMI, HT_SOTHWPERU_CHTASSMI, TV TPE	A PUPILIATCASNALCOFFCVUSMEHEVPUL 2 KEU	CTI.RELATEVAEPUTTASTVASTPOSLAEVELCT-DVEL
-				
88 /	AGAPLWLVWPLAVAG <mark>Y</mark> V <mark>L</mark> MGRLGLVTAMPPTGAV	VV <mark>LWPPNAVLLTALLAS-APKHWPVVL</mark> LGGVAAEVAVNWHT.	L PIAWALAFGTVNAAESTLAAS <mark>LLR</mark> RFSDGP VRL	DGLAAVVRFVLLAPL <mark>A</mark> ASATAALA <mark>G</mark> AALIALRFPEVDY
15 1	PVPAARWGLLLAAVG <mark>Y</mark> F <mark>A</mark> LATAGIWLGRQPGTGA	AMLWMGNAFGLGLLLHR-PRQEAPALLASMLAADVLANLLA	D 1 TAGVALLLAGANLLEIGGMAMLLRLMSGSG -TL	ARTGPLATFALTIGV <mark>A</mark> ATLGPAIS <mark>G</mark> VPGAAVLATAYGTQF
8 1	MAVAKELVRGMLISV <mark>C</mark> YCLAFLMLWRLSID	DOWYLPVGLRVVTLLFR-PYREWPFLLAGDAAAMLFLRIPL	3 5 NPLWSYLSPFVHAPLIAGAIGLLRYOVPHI	VKSOOWFIPAALVIALWNAICSLVLNSTLGGPOAYPV
6 1	EVERYDRUT AT FTURYUT CCCFAUAT A TURCTCT	TERMADERT PTATINT ALV. OROSWOWNT LACCRAFUL SNELW	2 DEDAAMI TYUCNAT PAUT CACT UNWAT PDC TOT	FEI PEUT AFUUT CACUADUUCAEUCCAET AMECT I COTE
	FIRFHFRHLALFIVATVLGCGFRHALAIVFGIGI	13FHAF30DFTATDADV-QRQ3HFHHDDAGCFAEVD3AFDH	E E FFFRRILLI VORALERVI GRALVINALKKA IKL	EILKEVLARVVLOROVARVVSAIVOSAILAWFGI-LSQIF
63 3	SSRPRILHLGLFVTA <mark>Y</mark> V <mark>L</mark> AAGFAQALAVVPGTGI	IS <mark>IWPPSGLFIATLVIA-PTKSWPWWV</mark> IAGL <mark>L</mark> GELFSSALW	F Z SLPVALLIYSGNALEAITGAW <mark>LVNL</mark> ACKKP VKL	ETLQEVLAFVLLAAG <mark>L</mark> APFASATV <mark>G</mark> SATLAWFGVQSF
10 1	EIVKICGIIFISGMI <mark>Y</mark> Y <mark>V</mark> LALIGRKTAIYPSYAA	AA <mark>IW</mark> PAS <mark>GA</mark> ALGFTLLF-GNYAVLG <mark>VF</mark> LASVLFNFGTGILS	3 1 NLYLNFLIGFFSAFQSYVGKTVLTRKIPGY -KI	SDRTQFVFLFIFLEA <mark>I</mark> VCVINATG <mark>S</mark> VASMYFLGEIDLSSI
17 1	YSMSKLNEILMVAVV <mark>Y</mark> F <mark>T</mark> TARLGOLLAISPGNVT	TPVWLPSGTCLAWVLLR-GRHLWPGTFLGAFAGNIWAYFDT	7 SILAGSANGLGDVIAIVGAAWMINTISPAT ALF	NNSKNTTIFLIYGGI <mark>I</mark> GPFISASI <mark>G</mark> IFGLLLTGNILPNDA
10 1	URCENT I ENT CLUUUNT TACAL CLOVATU CNTS	CRAMPDA OVET AAT T VI OVERWROUPVCAEVICUT ON	BUUASI CUCUCSTI AAUI CUCI I PMI CED. KCI	ERTROUVAL CACACAL CLOVEAWACUAST VI CORL RETEL
10	WORNDEENDOEVVVIDDAGADGEQVAIV-GAIS	37 KWFFRGV3DAADDVD-GV3KWFGVFVGR3V13VD6DV	- FVVRSDOVGVGSILARVLGVGDLKNLGFD- KGL	BRIRD V VALCAGAGALCLG V SAWAG VASLVLGGALF SISL
52 9	QGTRYAGVLVA IGV I <mark>Y</mark> F <mark>V</mark> LAKGGLALASVHPSA T	TP <mark>IW</mark> PPT <mark>GL</mark> ALAAVLLW-GYRTWPA <mark>IF</mark> IAAV <mark>I</mark> ANATTAG	<ul> <li>SVATAIAIASGNTLEAVVGAYLVNRWSRGR HTF</li> </ul>	SRPNSVAKFALICIV <mark>I</mark> ATPISASI <mark>G</mark> LTSLATAGYIDRTNF
15 1	RSVSWOROIIVAIAI <mark>Y</mark> G <mark>T</mark> IYISOLFITYAGTASS	SPIWIPTGMAVGFLSIW-GYOVWPGMVGGLLLGEIMALHGL	E 4 FILTVAITGVV-SLVNLFSVY <mark>LS</mark> DILTANN YLF	SKIKNIIRFIVFVCFGSRLPAAIICPFLLYLFEKISFGLY
18 1	KSI. PPPI. VTAL PSI. ATAL GI CTAL AVCMI. GPG	COONT ST. CLOCHALAL WOOGTNT TALALCUTTA PMVVOC	2 LALALONTSTAVVLMACRCLEMIERSPOGO	-RLODGWVFLLAGVGLGATLTATATATILLSLWOGDCRT
25 1	LVKLGFKHSVPFILA <mark>YLA</mark> LAGASMLLAVAPSNVI	TP <mark>LYPAAGLAW</mark> GV <mark>AIL</mark> W-GRKTYPS <mark>IF</mark> VAATLANTLPLLFS	2 4 AVLAGIAIGVGEVIAVLISCK <mark>LLS</mark> AVRGRER	VFKQVQDVIAFSVIALFWVVSPTI <mark>G</mark> VSALALAGLVESGAF
61 1	LHNILLNLLVALFYV <mark>V</mark> G <mark>I</mark> RLSDQLVSTLLPGRIS	SP <mark>VWFPSALTF</mark> GVFFHF-GHWVTPG <mark>II</mark> LGSVLGLVSVLSTF	D 7 FLVLETAFAFANTLQPFIGDL <mark>WLK</mark> NKLISV 21 LNF	FQRVSTTLAFVQGAI <mark>I</mark> GPFLSAVI <mark>G</mark> VTTLLLMGIIARENF
12 1	NAFAYIOVNLFIALA <mark>Y</mark> F <mark>I</mark> LGOMGLMIALPPGYSA	AATWPAAGVATASCLUWKGRAPWGGVFIGSLUTNMNIGOEL	I LGWLPLIIALGTLLOTIISAR <mark>IIO</mark> YIDPOLT	LDKPETVIKSMVSLSASCCIASLFCNAALVLGDTITSADV
1.1	TARENCYMII ACI CYNTI CAAACEETI DRCYAS	CRTWRAA WANTCOTRE CHRATICUPEACENT NTOFENAC		A TEUNDURI AT UTODI ACT MEDEURI CETTUDUOCA
	AIRKSVGIMLERGEGIMLEGARAGFFIIFFGIRS	SFIRFARDARY ICCIRF-GREADDOVEFASEIDAIQESNAS.	S S ANVVRGETREGRAVCORGENSICAR REASA HOL	ATT VAL VAL AND A DE LAS LAS FIVOTST LAS DIVE VOSA
11 (	QMRFDAAQIAVVAVA <mark>Y</mark> H <mark>A</mark> LAQLGLYFAIPPGFAT	TV <mark>FW</mark> PAS <mark>GVAL</mark> AA <mark>VM</mark> LR-GPWMMLG <mark>VL</mark> LGSFSVNLYISLQL	N 7 NVLVALGIACSSALQAGLGAW <mark>LI</mark> SYWLKRQ 1 GRL	TQLRHLLVLVGLGGP <mark>I</mark> GCMLAATG <mark>G</mark> TTILVATGVMLPEQW
21	AQWSRRALISFGVVY <mark>A</mark> V <mark>L</mark> IWVGAQPLLANRG-IP	PLLWVAGGFAAGGVLLA-GKRLLPIIFIGAFIAYTVVFRQL	3 5 LLAAALVFGLSAIIQAWLVNWLVRFGR TLP	PSSIRYTLVVCGLLALAVLPAPFLAVMVLVVVQQPLVSSP
			· · · · · · · · · · · · · · · · · · ·	
c				DHPVDLLSLLVQSLTG <mark>LL</mark> GAGIQRLRELNQS 28
0			LALP-IIALAWHYGWQGALIATLMNAIALIASQTWR FT	DH
-	SKESWN <mark>A</mark> LLLTLT <mark>SCL</mark> TLAPICLVFWHYL 16 I PIMLPGLTSFTGGVLLTPFVYLIFEFL 16 P		LALP-IIALAWHYGWQGALIATLMNAIALIASQTWR FLP-NVVMANKFGWQGWLSGLLGSMMITIARQIGUG FFLP-NVVMANKFGWQGWLSGLLGSMMITIARQIGUG	DHPVDLLLSLLVQSLTGLLGAGIQRLRELNQS 28: FSN
-	SKESNN <mark>ALLLTLTGGLTLAPICLVFWH</mark> YL 16 I PIMLPGLTSFTGGVLLTPFVYLIFEFL 16 P RETLIA <mark>V</mark> SIASLTGGIVLTPFLYLVADYL 17 K		IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	DH
	GRESWN <mark>ALLILITGGLTLAPICLVFWHYL 16 I PINLPGLTSFTGGVLLPFVYLJFEFL 16 P RETLAVSISSLTGGVLPFVYLJFEFL 16 P VPSVYM<mark>VNLASISGGLM</mark>LVPMCTLLWNYL 17 K</mark>		LILIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	DH
	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	INMEGRALUTILLEVISION         INMEGRALUTILLEVISION           PLRTSLITURGEPTIGIGONULAPET         EBILLIV           KLKPKMLMWSLLFPSIGLABLTLEQH         FRILLIV           VUFSIANIVYIAULISISIONSIAPETGIGAPLATURE         KRAPAPCI	LAD- I TILANYICHQAALIATIMNI ALIASOYNR	DH
	SKESWARLLINGGUTAPICUUHHII 161 	INNEGRIUWYLLE VISIAUQGLPDELSRFPFC PLTSLIIVSEPTIGIOTYNUSFEISRFPFC KLEFKHAMMSLFFTGLGLAELTLLEGMKPFALLIV VYFSIRHIVYALLIGSILVGSIPELKRFAPC ETLLGLASVALVVSARLIFOGSISARFGGI-GFALTYVP AVLAREKRFTTLMFFATLATLAVICTFPETKYLAG	ABC-ITALAMRIGMOGALIATIMALALIASQTWR- 	DH
	ARSINALLITIGOTILAPICAYMAYI 16 I DIMLOGISTI GOULAPPU AULTEIL 16 I ABILIAVSIASITGQU/IPPLUVADIL 17 K VPSVINVILASISGCALUPPCULAVII 16 V QOVLINVAASILVUVICULAVII 16 V QOVLINVASICLUVUVPUCULAVII 16 V GOVLINVASICUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU	INNERGALVATILLEVISUANQUGIPDEL SAFFFFC INNERGALVATILLEVISUANQUGIPDEL SAFFFC ILIZISIINGEN ILIZISIPICE IL	ADP - 12ALAMPENGOALIATIMA TALIASQUWR - 17ALAMPENGOALIATIMA TALIASQUWR - 17ALAMPENGOALIASIACIASSACIASIACIAS - 17ALAMPENGOALIASIACIASIACIASOW - 17ALAMPENGOALIASUACIASIACIASUACI - 17ALAMPENGOALIASUACIASIACIASUACI - 17ALAMPENGOALIASUACIASIACIASUACI - 17ALAMPENGOALIASUACIASIACIASUACIASUA - 17ALAMPENGOALIASUACIASUACIASUACIASUA - 17ALAMPENGOALIASUACIASUACIASUACIASUA - 17ALAMPENGOALIASUACIASUACIASUACIASUA - 17ALAMPENGOALIASUACIASUACIASUACIASUACIASUA - 17ALAMPENGOALIASUACIASUACIASUACIASUACIASUA - 17ALAMPENGOALIASUACIASUACIASUACIASUACIASUACIASUA - 17ALAMPENGOALIASUACIASUACIASUACIASUACIASUACIASUA - 17ALAMPENGOALIASUACIASUACIASUACIASUACIASUACIASUACIASUACIASUACIASUA - 17ALAMPENGOALIASUACIAS	DH
	SKSSNARALLITIGGTILAPICLYVNYL 16 I PIHLPGLTSTFGVLTPPVLLFPF 16 P RETLIAVSIASITGVLTPPVLLFPVL 17 K VPSVIVNUASISSGCMLVPVPCLINNYL 16 V DQVLRTWAASDPLJVLIVPVLATYSSRF 10 E GTGTAIFSIVDLLGLSAALIFTLPFTYP 18 A NIHTLINVQALVWRLVGVVPCCPTIRT 18 A	BURGHELVYLLIVYLLIVYLLIVYLLIV INMEGHELVYLLIVYLLIVYLLIV PLATSLI INGLEF IGIOTQVLLEFEIBRTFBYC KLARPKLAMSLEFF IGLOLARLILLGMFRPALLI VVFSTRHIVFTAVL_IGSLIVYSLIPEELFRPALG ETLIGLASVILVYSLILTYGSTRAPTGTI-GFALTYV PYLLAKSKLIFTLIFTATLLAVLCTFFETKXIG BATKKEIVTHAVTTAFTLIAKTLONDSSIF-STNTILS.	LAND - I TALANING MODAL LATLAND ALL LASOFT FLORING MODAL CALLADD ALL LASOFT FLORING MODAL CALLADD ALL LASOFT FLORING MODAL AND ALL LASOFT FLORING MODAL AND ALL LASOFT ALL TALANING MODAL LATLAND ALL AND ALL AND ALL FLORING ALL AND ALL AND ALL AND ALL AND ALL FLORING ALL AND ALL AND ALL AND ALL AND ALL AND ALL FLORING ALL AND ALL AND ALL AND ALL AND ALL AND ALL ALL AND ALL AND ALL AND ALL AND ALL AND ALL AND ALL AND ALL ALL AND ALL AN	DRPVDLLJELUVQSLYCELLGAGIQRLREELAQG 28. PSN
	SKRSNARLLITT GOTTLAFIC VENUL 16 I - PILLPGLTST GVLTPFVLVAVULFFL 16 I RETIA VSIASIT GVLVPFVLVAVULFFL 17 K VSUVAVULASIS GGLVVPVLVLVAVULFFL 17 K VSUVAVULASIS GGLVVPVLVLVAVSIFR 10 E STGRAT SFUVALLISAALIFTLFFVT 18 A NIMTLINQALLVMLVGVLCFTIRTI 18 A NIMTLINQALLVMLVGVLCFTIRTI 18 A	INNRGRIVWILLEVISWUNGLGLPDEL	AD-IIIAANHYGNQALLIAIIMAIALIAGYTR PD-NYVANKYGNQGGUJGLIGUSMITIAGYTR LD-IIIAANYGNQGUJASLIGUSMITIAGYGG D-IIIAANYGNQGALLATUMISIALIAANGYSG PD-IIIAANYGNQALLATUMISIALIAANGYSG PPA-VYALLINGIGGGAANAVULAITULIFANGOP- LD-VULGYTGVGKISTELLANYSNALTALFISYNN LD-VULGAANYGYKISISIANYALTAISTINSYNYTPI-	DH
	INTERNAL INTERNAL AND	INNERRIVETLEVELUEVELUEVELUEVELUEVELUEVELUEVELUE	$\label{eq:result} \begin{array}{l} AB = 1 \\ AB = $	DB
	SKESNNALLITLRÖGTLAFICLYPHILL IG DEN GELTSTOLLITLRÖGTLAFFTL AFLLAFICTUREN I STALLITLLITLRÖGTLAFFTL STALLITLRÖ	INNEGRILVWILLEVISIUMQIGLEPBELSAFFPEC LIFTSLIINGSPT IGOTQNVISPEISAFFPEC KLEFKHAMNSLJPFIGLAELTLEGM	AD-1 TZLAMHYGMQGALLATIMBA TALIAGYMR D-1 VNJAMKYGMQGGUJGJUSSIMITTARJGYGG TLD-7 TLAXTYGMQGGUJASISISIMITTARJGYGG P-1 VJALIYGMQGALATSLBYTTARJGYGS PFA-VVJALIMGGRG-SAAVIVLATSTALTARSGYSN- D-7 VJALIMGGRG-SAAVIVLATSTALTARSGYSN- LDPVHCHGANTGYRTSIN NAVLITSINIYMYPI LDPUHCGANTGYRTSIN NAVLITSINIYMYPI TUP-7 VJLTHUGHRG-AALGVVLATSINIYMYPI TUP-7 VJLTHUGHRG-AALGVVLATSINISYMYPI TUP-7 VJLTHUGHRG-AALGVVLATSINISYMYPI	DH
	ARSINGLITITE GITLEPICLYPHEL 6 - DRESHNALLITITE GITLEPICLYPHEL 6 - RETLAR GITLEPICLYPHILMADL 17 RETLAR GITLEPICLYPHILMADL 17 NOTHER GITLEPICLYPHILMADL 17 NOTHER GITLEPICLYPHILMADL 17 NOTHER GITLEPICLYPHILMADL 17 NITELINGLAU MEL GAALFTER 17 NITELINGLAU MEL GYALCHTER 17 NITELINGLAU MEL GYALCHTER 17 NITELINGLAU MEL 17 JOHN 18 NITELINGLAU MEL	INNERGELVWILLEVISUNGLGLEPDEL	ADP - I TULANHEYMOGALLATING TALLAGYNR - TULANHEYMOGAULATING TALLAGYNR - TULANEYMOGAULASIACIAASIACIAC - TULANEYMOGAULASIACIAASIACIAC - TULANEYMOGAULATULASIACIAL - TULANEYMOGAULATULASIACIAL - VILANEYMOGAULATULASIACIAL - VILANEYMOGAULATULASIA - VILANEYMOGAULATULASIA - VILANEYMOGAULATULASIA - VILANEYMOGAULATULASIA - VILANEYMOGAULATULASIA - VILANEYMOGAULATULASIA - VILANEYMOGAULATULASIA - VILANEYMOGAULATULASIA - VILANEYMOGAULATULASIA - VILANEYMOGAULASIA - VILANEYMOGAULATULASIA - VILANEYMOGAULASIA - VILANEX - VILANEX - VILANEX - VILANEX - VILANEX - VILANEX - VILANEX - VIL	DH
	LINE STUDIES IN THE STATE OF TH	INNIGHI.VII.LIVII.SIV.OJOPEBRTPPC PLATSII.INGLEY IOIOTOMUL PEIBRTPDC PLATSII.INGLEY IOIOTOMUL PEIBRTPC FILICIA.VIV.SIV.OJOPEBRTPLC VVFSTRIJVETAVLI IOSILVOTSLPEERRPACT PLILOLAAVSLIVVSAILI POOSTARPCTOI - GALITYP APULAKERLETILÄPÄTLATLANLAVCTPFETKI IOS SATKKEI VIVLAVY TILHFILAMELINETDISSIF-SINTILS. SATKKEI VILLÄ YLÄLLLÄVELULENENESTIF-SINTILS. SATKKEI VILLÄ YLÄLLLÄVELULENENESTIF-SINTILS. BATKKEI VILLÄ YLÄLLILLENENESTIF-SINTILS. BATKKEI VILLÄ YLÄLLILLENENESTIF-SINTILS. BATKKEI VILLÄ YLÄLLI YLÄLLILLENENESTIF-SINTILS. BATKKEI VILLÄ YLÄLLI YLÄLLILLENENESTIF-SINTILS.	LLD - THANKS MODAL LETTING THAT AND	DE
	ARSINALLILITÄRITLAISILLILIA KRSINALLILITÄRITLAISI ARSINALLILITÄRITLAISI ARSINALLINITÄRITLAISI ARSINALLINITTAINALLINITTAINA ARSINALLINITTAINALLINILLINILLINILLINILLINIL	INNERGELVWILLEVISIUMQIGLEPDEL	AD-IIILANHYGYQGALLAILMAINA TALLAGYTM TD-IVLANHYGYQGGULGALGUSMITTARQTYG TLD-TTLAINYGYQGULASLAUSIMITTARQTYG TLD-TTLAINYGYQGULASLAUSINITTARQTYG TLD-TULAYYGYQGALLATHAUSINITALAIASUYS TALTARQTYGQALLATHAUSINI TALTARQTYGAUSINI TA	DH
	LILIUULELLILUKALTLAPICUVMINIL ISE SESNALLILUKALTLAPICUVMINIL ISE PILAISISTEN DALLIPUVLILISE VIEVINILISE DALLIPUVLILISE VIEVINILISE DALLIPUVLILISE VIEVINILISE DALLIPUVLILISE VIEVINILISE VIEVINILISE VIEVINILISE VIEVINILISE VIEVINILISE VIEVINILISE VIEVINILISE VIEVARARVONI VIEVARARVONI V	UHIHIHIHIHIHIHIHIHIHIHIHIHIHIHIHIH	$\label{eq:response} \begin{split} & AB = 1  Adamse to the second state of the second state$	DH
	INTERNITATION CONTRACTOR CONTRACT	UINKONELVIKUULLINUULULUULUULULUULULUULULUULUULUULUULUULU	$\label{eq:result} \begin{array}{l} \textbf{L} \textbf{L} \textbf{L} \textbf{L} \textbf{L} \textbf{L} \textbf{L} L$	DB
	ALLINING LITLATION AND AND AND AND AND AND AND AND AND AN	INNIGRELIVITLLIVISUUGUGUEUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU	ADP - I TALAMMEY MOQGALLATIMA TALLASQUMR DE - NYADANE MOQGAULSILATIMA TALLASQUMR DE - NYADANE MOQGAULSILATIMA TALLASQUMR DE - I TALARE MOQGAULSILATIMA DE - I TALARE MOQGAULSILATIMA DE - I TALARE MOQGAULSILATIMA DE - I TALARE MOQGAULSILATIMA DE - NYADANE MOQGAULSILATIMA DE - NYADANE MOQALISTATIANA DE - NYADANE MOQALISTATIANA DE - NYADANE MOQALISTATIANA DE - NYADANE MOQALISTATIANA NYADANE MOQALISTATIANA NYADANE MOQALISTATIANA DE - NYADANE MOQALISTATIANA NYADANE N	DH
	INTERNATION OF A CONTRACT OF A		LAD - 1 ILLANNIY GOQALI ATJAKI ALI IA SQTWR. - MY MANNY GOQALI ATJAKI ALI IA SQTWR. - MY MANNY GOQALI ATJAKI ALI IA SQTWR. - MY ALIANY GOQALI ATJAKI ALI IA SQTWR. - MY ALIANY GOQALI ATTALIS TALI AA SGTWR. - WY ALIANY GORG-SAAVIVIL AUTU ITS HAY GARAGORP- - LUY-DY ALIANY GYRT SI IA WY ALI ATJAKI SQTWR H- LI DYNGRAMRY GYRT SI IA WY ALI TS HIY GYRTWR H- LI DYNGRAMRY GYRT SI IA WY ALIANTIA IS SUMY FI - YY ALIANY GYRT GYRT SI YA WY ALIANY ALIA SUMY - YY ALIANY SYNTAI GORG HANGY ALIANY ALIANY ALIANY - WY ALIANY SYNTAI GORG HANGY ALIANY ALIANY ALIANY - YY ALIANY SYNTAI GORG HANGY ALIANY ALIANY ALIANY - YY ALIANY SYNTAI GORG HANGY ALIANY ALIANY ALIANY - LY ASARTY DUY ALIANY ANY TYY ALIANY ALIANY ALIANY - LY ASARTY DUY ALIANY ANY TYY ALIY ALIANY ALIANY - LY ASARTY DUY ALIANY ANY TYY ANY ALIANY ANY ANY ANY ANY ANY ANY ANY ANY ANY	DB
	INTERNATIONAL STATES ST	UIMINI IN SULVIDUE SAFTFEC INMEGRI VALVILI VI SULVIDUE PEI	LAND - I TALANDISTOPOLI GILLOS MITTADO INC. TEDE - NY DANNY GIQOGALI GILLOS MITTADO INC. TEDE - TEL LANY GIQOGALI GILLOS MITTADO INC. DE - TEL LANY GIQOGALI GILLOS MITTADO INC. DE - TEL LANY GIQOGALI GILLOS MITTADO INC. DE - VIVALINO SIGO - SAAVUVALI LANY LANS GINO PA- VVVALING GIO- ANA UVALI LATVILIES TANI TAABSOYS LADVUL GOTTI GIVEN SI SI INAVVALI SI HIS YON TE I- LADVUL GIART KONKI SI INAVVALI SI HIS YON TE I- LADVUL GIART KONKI SI INAVVALI SI HIS YON TE I- LADVUL GIART KONKI SI INAVVALI SI HIS YON TE I- LADVUL GIART KONKI SI INAVVALI SI HIS YON TE I- LADVUL GIART KONKI SI INAVVALI SI HIS YON TE I- LADVALI SI HIS YON TE I SI NAVVALI SI HIS YON TE I- LI JANANS SI YOTALI GILLOS YON TAMIL MIS HIM YON HIG - LANGARRUPULAN LIANTY SAIVUL INAVISI D- PALI KONKI PONGANI YOTA JI YANANSI VANIS VID- PALI KONKI YON TE YOTALI YANYI YANYI KANYI D- YU PALI KANYI YANYI YANYI YANYI YANYI KANYI D- YU PALI KANYI YANYI YANYI YANYI YANYI KANYI KANYI D- YU PALI KANYI YANYI YANYI YANYI YANYI KANYI D- YU PALI KANYI YANYI YANYI YANYI YANYI YANYI KANYI HAYI YANYI Y	DH
	HILINING BUT AND A STATE AND A		ABP - I TALAMIN GNQGALLATIMA TALL AQCYMR - NYAMAN GNQGAULATIMA TALL AQCYMR - NYAMAN GNQGAULATIMA SAN TILAYANG YAG - NYAMAN GNQGAULATIMA SAN TILAYANG YAG - I TALAIN GNQGAULATIMA SAN TILAYANG YAG - I TALAIN GNQGAULATIMA SAN TILAYANG YAG - NYAMAN GNGAULATIMA SAN TILAYANG YAG - NYAMAN GNGAULATIMA SAN TILAYANG YAG - NYAMAG YAG - NYAMAG YAG - NYAMAG YAG - LIYAANAG YAG - LIYAAAG - LIYAA -	DH
	INTERNALIZED IN TOTAL TOTAL IN THE INFORMATION INTO INTERNAL INTERNAL INFORMATION INTO INTERNAL INFORMATION INTERNAL INTERNAL INFORMATION INTERNAL INFORMATION INTERNAL INTERNAL INFORMATION INTERNAL INFORMATION INTERNAL INTERNAL INFORMATION INTERNAL INFORMATION INTERNAL INFORMATION INTERNAL INTERNAL INFORMATION INTERNAL INTERNAL INTERNAL INTERNAL INTERNAL INTERNAL INTERNAL INTERNAL INTERNAL		LAD - I TALAMITY ORGANIZATIAN ALI IS GYM	DB
	INTERNAL INTERNAL STATESTATESTATESTATESTATESTATESTATESTAT		LAND - I TALANNIS YOO AL LATANA BALLANDA - I TALANDIS YOO AL LATANA BALLANDA - I TALANDA YA	DH
	LILIULILLILLE CLAPHOLOGYMHIL 16 LIESSNAALLILLE CLAPHOLOGYMHIL 16 PETLAKSTAAC DALLEPPTLAKEN VESVTOMLASIS CLAUPPTLANDL 15 PETLAKSTAAC VALUE PETLAKSTAAC VALUE PETLAKSTAAC VALUE PETLAKSTAAC VALUE PETLAKSTAAC VALUE VESVTOMLASIS CLAUPPTLAND VESVTOMLASIS VESTAAC VESTA	UHINI HANDOLE POLICIA SALA SALA SALA SALA SALA SALA SALA SA	ABP - I TALAMIT STORY AT A THE AT A THE ASSESSMENT AND STORY AND A THE ASSESSMENT AND A STORY AND A THE ASSESSMENT AND A STORY AN	DH
	INTERNICIALISTICATION CONTROLLADION CONTROLLADION CONTROLLADION CONTROLLADION CONTROLLADION CONTROLLADIONE CONTROLADIONE CONTROLOGICO CONTROL	INHINI IN	LILD - I THANKING NGGALI JIYA NA LI LISOYA. YE D- YW MANYO NGGAU AGULOSANI TI AG IGG- YE D- YW MANYO NGGAU AGULOSANI TI AG IGG- YE D- YW ANYO NGGAU AGULOSANI TI AG IGG- PA-VWALING NGGAU ALYU ALYU AT SARABOD - LIJA NGGAU AGULAYU ALYU ALYU ALYU ALYU ALYU YW AGULAYU AGULAYU ALYU ALYU ALYU ALYU LIJYAH AGANKI YKI SI IAVALIYU ALTSI HIYOYAH LIJYAH AGANKI YKI SI ALYU ALYU ALYU ALYU HIYO ALYU AGULAYU ALYU ALYU ALYU ALYU HIYO ALYU ALYU AGULAYU ALYU ALYU ALYU ALYU HIYO ALYU ALYU AGULAYU ALYU ALYU ALYU ALYU HIYO ALYU ALYU ALYU ALYU ALYU ALYU ALYU ALYU	DB
	ALLINING ALLINING ATLANDICATION AND A STATEMENT A STATEMENT A STATEMENT AND A	INNIGRALUWILLE VISIUMQUGLEDEL SAFTPEC INNIGRALUWILLE VISIUMQUGLEDEL SAFTPEC HITSLI INSLIP I GIGIQUVIC PEFI ERLLIV KLAPKALMISLIP I GIGIQUVIC PEFI ERLLIV VISI RAI 197 ALLIGSI LUQISLEDEL SAFTAFEC DEN ANDER VISI GIGIQUVIC PEFI SAFTFEC BATKEL VINTUL INTAKI LANDETPF OFTITIS SATKEL VINTUL INTAKI LANDETPF OFTITIS SATKEL VINTUL INTAKI LANDETPF GIMSL PORBALLASGAAL ICLIV.GPBELDOS LATLIG PORBALLASGAAL ILLU, VISI SAVAR GIMSL DIRBASHREFSAILLIK, VISI VISI SAVAR VISI TIS KILLESSAILE SALLIL CALVESTANDALLIVI USU SAVAL VIAFITI LAINESSANGENOV I TAVGULAUSU SAVAL VIAFITI LAINESSANGENOV I TAVGULAUSU SAVAL VIAFITI LAINESSANGENOV I TAVGULAUSU SAVAL	ABP - I TALAMINY CHOQGALLATIMEN TALL ASQ'MR. 	DH
	INTERNATION OF A STATE		$\label{eq:response} \begin{array}{l} \\ AB_{P} = 1 \\ AB$	DB
	INTERNICT INTERNATIONAL DE LA CONTRETA LA CONTRETA LA CONTRETA LA CONTRETA DE LA CONTRETA LA CONTRETA LA CONTRETA DE LA CONTRETA DEL CONTRETA DE LA CONTRETA	INHINI IN	LLD I ILLAND GYAL I ILLAND ALL IS GYR. - THE ILLAND GYAL I ILLAND ALL IS GYR. - THE ILLAND GYAL I ILLAND ALL IS GYR. - THE ILLAND GYAL I ILLAND ALL I ILLAND ALL I ILLAND ALL - THE ILLAND GYAL I ILLAND ALL I ILLAND ALL I ILLAND ALL - THE ILLAND GYAL I ILLAND ALL I ILLAND ALL I ILLAND ALL - THE ILLAND GYAL I ILLAND ALL I ILLAND ALL I ILLAND ALL - THE ILLAND GYAL I ILLAND ALL I ILLAND ALL I ILLAND ALL - ILLAND GARAGONAL I ILLAND ALL I ILLAND ALL I ILLAND ALL - ILLAND GARANGONAL I ILLAND ALL I ILLAND ALL I ILLAND ALL - ILLAND GARANGONAL I ILLAND ALL I ILLAND ALL I ILLAND ALL - ILLAND ALL I ILLAND ALL I ILLAND ALL I ILLAND ALL I ILLAND ALL - ILLAND ALL I ILLAND ALLAND ALLAND ALLAND ALL I ILLAND ALLAND ALLAN	DE
	HIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	UHINI UNILLA VISUA MUQUE PEL INNIGRAL UNILLA VISUA MUQUE PEL LIFISI IN USUA ILO TANU SPEL LIFISI IN USUA ILO TANU SPEL LIFISI IN USUA ILO TANU SPEL PELLO LANSUL VISUA ILO TANU SPEL PELLO LANSUL VISUA ILO TANU SPEL SATKKEI VINU VISUA ILO TANU SPEL SATKKEI VINU VISUA ILO TANU SPEL SATKKEI VINU VISUA ILO TANU SPEL PELLO LANSULA VISUA ILI LUNU ALE SSITI S-SINI TLS SATKKEI VINU VISUA ILI PAUS VISUA PERLO LANSUA VISUA ILI PAUS INTER NEMALSGRAGO VINU TANU TILUFI LUNU SPEL PERCALA VISUA YILA YILA VILUS PELPOS INTEGNA VINU TILUFI LUNU SPELPOS INTERLAS VINU TILUTU VILUS VILU SPELPOS INTERLAS VILUS VILUS VILUS VILU PERCALA VISUA YILAY TILU SPELPOS INTERLAS VILUS VILUS VILUS VILUS VILUS VILUS PERCENTI VILUS VILUS VILUS VILUS VILUS VILUS VILUS SULLES SALLUS VILUS VILUS VILUS VILUS VILUS VILUS SULLES SALLUS VILUS VI	ABP - I TALAMINE MOQGALLATIMA TALL ASQ'MR. INVAMANE MOQGALLATIMA TALL ASQ'MR. INVAL AN AND SAN AND S	DH
	INTERNAL ALL AND A DEVINE AND A	UHINI HIM IN SUM OLO TOPUL SUPER- PLATSI IN SULF I DIOTONUL PEI	LARD-1 IALANNI SQGALLATJABA JALJA SQTWR. 	DB
	INTERNICIALITATION INTERNATIONAL DE LA CONSTRUIT DE LA CONSTRUIT. LA CONSTRUIT DE LA CONSTRUIT. LA CONSTRUIT DE LA CONSTRUIT. LA CONSTRUIT DE LA CONSTRUIT. LA CONSTRUIT DE LA CONSTRUIT. LA C	INHIHIIIIIIIIIIIIIIIIIIIIIIIIIIIII	LINE - I TALANDER OF ALL TALANDER ALL TALAND	DE
	HIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	UHINI HALLAVI SLALQALPOL	ABP - I TALAMIT STORY AT A THE AT A THE ADDARSE AND A THE ADDARSE ADDA	DH
	LINE SENSING LIFE DE LA DE CUMUNEL SE 2 - PILL QUITET D'YLLPPY LIFEL IS PETLAÑSISALD SOULPPY LIFEL JANNI IS PENUNAKASIS OLANYPY CLIANTLI IS VERVINAKASIS OLANYPY CLIANTLI IS UTATAN SENSING VILTPY LIFEL SE SUGULT VASDEL ULISAALITTEFYTY IS ULGTLIYYGALLY MILGYPICYTERU IS SUGULT VASDEL MAUVEFYLIANTRA I STVIGARTLAGGAF LISGGAALAMTRA I LIVANSI STITUTTARINING LIFUCUTUM STVIGARTLAGGAF LISGGAALAMTRA I LIVANSI STITUTARIANTI I LIVANSI STUTUTUM I LIVANSI TILTYLIYL VIGULTUM LIVANSI STITUTUKARI I LIVANSI STITUTUKARI I LIVANSI STITUTUKARI I LIVANSI STITUTUKARI I LIVANSI STITUTUKARI I LIVANSI STITUTUKARI I LIVANSI VIGULLIYAGTI I LIVANSI STITUTUKARI I LIVANSI VIGULLIYAGTI I LIVANSI VIGULLIYAGTI I LIVANSI VIGULLIYAGTI VIGUN I LIVANSI VIGULLIYAGTI I LIVANSI VIGUNLIYAGART I STVINAKANAN VIGUNLIYAN I STVINAKANAN VIGUNLIYAN I STANTIMUKANAN VIGUNLIYAN VIGUNLIYAN I STANTIMU		LAB - I ILAMIT GOOL I LIMIN ALL IS GYNG LAB - I ILAMIT GOOL I LIMIN ALL IS GYNG LAB - ILMIN GOOL I LIMIN ALL IS GYNG LAB - ILMIN GOOL I LIMIN ALL IS ALL AND ALL IS GYNG LAB - ILMIN GOOL I LIMIN ALL I LIMIN GOOL FAR - VIAL I LIMIN GOOL I LIMIN ALL I LIMIN ALL I LIMIN ALL I LIMIN ALL I GIT I GURLI STYLL I STALL I LIMIN ALL I LIMIN ALL I GIT I GURLI STYLL I STALL I LIMIN ALL I LIMIN ALL INGE ALL I LIMIN ALL I LIMIN ALL I LIMIN ALL INGE ALL I LIMIN ALL I LIMIN ALL I LIMIN ALL I GIT I GURLI STYLL I STALL I LIMIN ALL I LIMIN ALL I GIT I GURLI STYLL I STALL I LIMIN ALL I LIMIN ALL I GIT I GURLI STYLL I STALL I LIMIN ALL I LIMIN ALL I GURLI STYLL I STALL I LIMIN ALL I LIMIN ALL I GURLI STYLL I STALL I STALL I STALL I LIMIN ALL I GURLI STYLL I STALL I STALL I STALL I LIMIN ALL I GURLI STYLL I STALL I STALL I STALL I STALL I GURLI STYLL I STALL I STALL I STALL I STALL I STALL I GURLI STALL I STALL I STALL I STALL I STALL I STALL I GURLI STALL I STALL I STALL I STALL I STALL I STALL I STALL I STALL I GURLI STALL I STALL I STALL I STALL I STALL I STALL I STALL I STALL I GURLI STALL I STALL I STALL I STALL I STALL I STALL I STALL I STALL I GURLI STALL I STALL I STALL I STALL I STALL I STALL I STALL I STALL I GURLI STALL I ST	DB
	INTERNATIONAL STATES AND		LINE - I TAL AND YOUNG ALL AT AND ALL AS OFT. THE - NY ANNY GROUND ALL ALL AND ALL AS OFT. THE - THE LANK (GOOGAL ALL AND ALL AS OFT. GARAGE - THE LANK (GOOGAL ALL AND ALL AS OFT. GARAGE - THE AND ALL AND ALL AND ALL AND ALL AND ALL AND ALL THE AND ALL AND	DE
	INTERNAL INTERPRETATION INTERPRETATI	UHIHIHIHII UHIHIHII SHATAYA SAFAYA INMIRARI UWILLAWI ISLMA QIAL PDESRFTPE C UNITS AND	$\label{eq:response} \begin{array}{l} \label{eq:response} \\ eq:respons$	DHPVDLIJLSLIVQSIJTCILDAGIQRAFELIQG 28: SenDVLIFIFIAJALATGIGAIASQUILAD 29 SESDOGLIJTHSGAPVJ.GGIALA 20 STANDARD 20 AVLDRINGGLIBAQIJLGITALI, UNAALKTIRQUARD 20 SGUDDGLAITSSSIVJSSIVSTITSINGLAVKROGINGR 20 GUDDGLAITSSSIVJSSIVSTITSINGLAVKROGINGR 20 GUDDGL
	INTERNICIALISTICATION INTERNAL DE LA CONTRATA DE LA CONTRATA LA CONTRATA DE LA CO		LINE - INLANTING OLD LITING AND	DB
	INTERNAL INTERNAL STATES INTERNAL INTERNA INTERNA INTERNAL INTERNAL INTERNA		LAND - I TALANNIS YAQALI ATA YALAN Y	DE
	INTERNAL INTERPRETATION INTERPRETATI		ABP - I TALAMIT SUGGET, AT LIMB, AT LIM	DB. VULLISLIVQSIJCULDAGIQRIELEAG Sen V. UVLIFIFIAJCALICTICIALGAGIQRIELEAG Ses V. UVLIFIFIAJCALICTICIALGAGIQRIELAG Ses V. UVLIFIFIAJCALICTICIALGAGIQRIELAG Ses V. UVLIFIFIAJCAGICAGICAGICAGICAGICAGICAGICAGICAGICAGI
	INTERNICAL INTERPEDIATION CONTRACTOR INTERNAL INC. - PILLO GUITATI SPILLOPPYLLIPPYLLIPPYLLOPPYLLOPPYLLOPPYLLOPPYLLOPPYLLOPPYLLOPPYLLOPPYLLOPPYLLOPPYLLOPPILLOPPYLLOPPILLOPPYLLOPPIL	INHINI IN IN INTERNET AND A STATE AND	LID - THILMING TO GOLD ACTION THILMING THE THICKNER THE - THILMING TO GOLD ACTION THILMING THE CON- THE OF THILMING TO GOLD ACTION THILMING THE CON- THE OF THILMING TO GOLD ACTION THILMING THE CON- THE OF THILMING TO GOLD AT THILMING THAN THAN THILMING THILMING THILMING THILMING THILMIN	DB
	INTERNAL INTERNAL STATE		LINE AT THE ADMINISTRATION OF A LINE AND A L	DH
	INTERNAL INTERPILATION INTERNAL INTERNA		LARD-1 IN ANNIES (QGALIATIAN), ALLIA SQYM,	DB
	INTERNET INT	INHINI IN IN IN IN INTERNET	LINE - I THAN HIGH ON A LIGHT OF	Description         PVDLIJSLIVQSIJCUL GAGIGRERELIGG 28:           Ben
	INTERNAL INTERNAL DE LA POLYVILTEFIL IS INTERNAL INTERNAL DE LA POLYVILTEFIL IS PERILAÑSISAL SOULTEFIL IS VERILAÑSISAL SOULTEFIL IS UNITERNAL INTERNAL IS UNITERNAL IS UNITERNAL INTERNAL IS UNITERNAL IS UNITERNAL INTERNAL IS UNITERNAL IS UNITERNAL INTERNAL IS UNITERNAL IS UNITERNAL INTERNAL INTERNA INTERNAL INTERNAL INTERNA IN		LINE - I TALANNES YOOTAL GITLAG ALA ALA ALA ALA ALA ALA ALA ALA ALA	DB. VULLISLIVQSIJCELLGAGIQREELBAG Sen VULLISTIAJCALTOLOIAISEQUILLI 29 Sen DOBLITHESGATVIJCGIAISEQUILLI 29 Ses DOBLITHESGATVIJCGIAISEQUILLI 29 Ses DOBLITHESGATVIJCGIAISEQUILLI 29 Ses DOBLITHESGATVIJCGIAISEQUILLI 29 Ses DOBLITHESGATVIJCGIAISEQUILLI 29 Ses DOBLITHESGATVIJCGIAISEQUILLI 29 Ses DOBLITESSIJVSIJ Ses DOBLITESSIJVSI SES DOST DOLLESSIJVSI SES DOBLITESSIJVSI SES DOBLITESSIJVSI SES DOST DOLLESSI SES DOBLITESSIJVSI SES DOST DOLLESSI SES DOST DOST DOST DOST DOST DOST DOST SES DOST DOST DOST DOST DOST DOST DOST SES DOST DOST DOST DOST DOST DOST DOST SES DOST DOST DOST DOST DOST DOST SES DOST DOST DOST DOST DOST DOST DOST DOS
	INTERNAL ALL STATES AND		LAD - I TALAMIT SQUALLA TALMA LALLA SQUAR LAD - I TALAMIT SQUALLA TALMA LATALA SQUAR LAD - I TALAMIT SQUALLA TALMA LATALA SQUAR LAD - I TALAMIT SQUALLA TALMA LATALA SQUAR LAD - I TALAMIT SQUAR SQUAR SQUAR LAD - I TALAMIT SQUAR SQUAR SQUAR SQUAR LAD - I TALAMIT SQUAR SQUAR SQUAR SQUAR SQUAR LAD - I TALAMIT SQUAR SQUAR SQUAR SQUAR SQUAR SQUAR LAD - I TALAMIT SQUAR	DB

#### Figure 10

**Multiple sequence alignment of the 8TMR-UT family.** Multiple sequence alignment the 8TMR was constructed as detailed in the legend to Figure 1. The species abbreviations are as shown in Table 1.

this conserved TM domain was recovered in diverse signaling proteins that are particularly widespread in proteobacteria. Sporadic occurrences of this domain were also encountered in signaling proteins from actinomycetes like *Corynebacterium glutamicum*, cyanobacteria, spirochetes like *Leptospira interrogans*, flexibacteria like *Chloroflexus*, and the *Ectocarpus siliculosus* virus (Table 1). A search for signal peptides using the SignalP program [66,67] with bacterial signal peptide models did not yield strong signal predictions for these proteins. TM helix prediction with an alignment of this conserved TM domain using the PHDhtm program [37] suggested the presence of 8 membranespanning helices. Further, helix prediction, individually for all members of this family, with the TOPRED, TMHMM2.0 and TMPRED programs [34–36] also suggested the presence of 8 membrane-spanning helices on an average. These algorithms also predicted a topology with an intracellular N and C terminus for this TM domain, which is compatible with the C-terminal signaling domains occurring immediately after the last predicted membrane spanning helix in most instances. Accordingly, we refer to this domain as the 8TMR-UT domain (for 8 trans membrane UhpB type domain).

8TMR-UTs are approximately 290–300 residues in length and are characterized by several distinctive features that differentiate them from all other TM regions (Figure 10). These include, an aromatic position followed by a proline in the second helix, a pair of small residues typically glycine in the 5<sup>th</sup> helix, and a charged patch just C-terminal to the last helix. The conserved prolines and glycines within the predicted helices suggest that they may possess conformational distortions that could be critical for signal sensing and transduction. The clearest functional clues for the 8TMR-UT domain comes from the E. coli UhpB protein, whose C-terminal intracellular histidine kinase transfers a phosphate to the receiver domain of the transcription factor UhpA. Currently available experimental evidence suggests that the 8TMR-UT domain of UhpB interacts with the transporter UhpC, which binds glucose 6-phosphate [68,69]. When UhpC binds glucose 6-phosphate it appears to transmit a signal via the 8TMR-UT domain of the UhpB protein to activate its kinase domain. Operons related to the Uhp operon are seen in number of bacteria, suggesting that a similar signal relay system is widely employed in sugar sensing by bacteria.

Phylogenetic analysis of this family suggests that the 8TMR-UT proteins are divided into 3 major groups (Fig. 8C). Proteins belonging to each of these sub-divisions often show a sporadic phyletic pattern, and often 8TMR-UT domains from distantly related organisms group closely together in the tree (Fig. 8C). These observations suggest a dynamic evolutionary history with gene loss and lateral transfer as in the case of the 7TM-DISMs. Likewise, the 8TMR-UT domains also appear to have extensively combined with a range of intracellular domains in various bacteria (Fig. 8C). These combinations often include linkages with GGDEF and HD-GYP domains, which are cyclic diguanylate generating and degrading enzymes respectively, histidine kinase and receiver domains, and PAS domains. In some cases the 8TMR-UT is combined with extracellular CHASE domains [23,70] (Fig. 8C), which suggests that it may transmit the conformational changes arising from the interactions of ligands with the CHASE domain, to intracellular signaling domains. These observations suggest that the 8TMR-UT domain might, in general interact with other membrane-associated proteins, and act a switch that senses conformational changes in them to transmit signals. Examination of the gene neighborhoods of the 8TMR-UTs showed that they often co-occurred with genes predicted to encode molecules with HTH, HAMP, PAS, GGDEF, HD-GYP, histidine kinase and receiver domains (Fig. 8C). These predicted operon organizations suggest that the 8TMR-UTs might functionally interact with other signaling molecules to form complex signaling networks that might help in sensing a wide diversity of stimuli [9].

## Conclusions

The presence of a relatively small set of well-understood signaling domains that are combined with a variety of other accessory domains allowed us to set up a sieve for new receptors in prokaryotes. As a result, we were able to identify two distinct families of 7TM receptors, namely 7TMR-DISM and 7TMR-HD that are unique to bacteria. The discovery of these new 7TMRs in diverse bacteria suggests that they may be more widely utilized in prokaryotes than has been previously suspected. Importantly, the domain architectures of the 7TMR-DISMs suggest that they are likely to activate a variety of intracellular signaling cascades including adenylyl cyclases and kinases. This suggests that these bacterial 7TMRs are functional analogues of the eukaryotic receptors, and could serve as models for the non-G protein linked pathways downstream of the eukaryotic receptors. Most members of the 7TMR-DISM family are fused to one of two extracellular domains at their N-termini. Both these domains are predicted to adopt all- $\beta$  fold with a jellyroll topology similar to the discoidin-type sugar binding domains. One of them, 7TMR-DISMED1 can be unified with the carbohydrate binding domains of β-galactosidases and β-glucoronidases. Accordingly, the 7TMR-DISM family is predicted to function as receptors for carbohydrates and related molecules.

Based on the contextual information from gene neighborhoods, the 7TMR-HD proteins are predicted to act as receptors that regulate the highly conserved glycerolipid metabolism pathway in response to stimuli sensed by their extracellular domains. The architectures of most of the 7TMR-HD and 7TMR-DISM proteins are reminiscent of the animal metabotropic glutamate and taste receptors. These animal receptors contain an extracellular periplasmic solute-binding domain that is typical of several bacterial signaling proteins [54]. This architecture, along with the limited phyletic pattern of these 7TMR (only found in animals), could imply that they were acquired from an as yet unknown prokaryotic source. In more general terms, the eukaryotic 7TMRs are thus far restricted in their phyletic spread to a few crown group lineages (animals, slime molds fungi and plants) and appear to have proliferated principally through lineage specific expansions from a few founders. The herein-reported discovery of new representatives of 7TMRs in bacteria suggests that they are ancient and widespread in the bacterial lineage. Hence, it is possible that a subset of the crown group eukaryotes may have ultimately acquired the founding members of their 7TMR families from a prokaryotic source through lateral transfer.

We also detected two evolutionarily mobile membranespanning domains, namely 5TM-LYT and 8TMR-UT that are associated with several different types of intracellular domains. These conserved TM domains, unlike members of the 7TMR-DISM and 7TMR-HD families, are not associated with large globular extracellular domains. We propose that one of these families, the 5TMR family, is likely to function as receptors for murein or its components. 8TMR-UT in contrast may sense conformational changes in other membrane-associated proteins and relay these signals to the intracellular signaling domains.

Identification of these receptors suggests new paradigms in bacterial signal transduction, and could also provide models for the functions of an important class of eukaryotic proteins. Experimental investigation of these proteins, particularly those from pathogenic bacteria, such as *Bacillus anthracis, Leptospira* and *Cytophaga*, are likely to be of interest in understanding microbial physiology and pathogenesis.

## Methods

The non-redundant (NR) database of protein sequences (National Center for Biotechnology Information, NIH, Bethesda, Date: April 1, 2003) was searched using the BLASTP program [71]. The searches were supplemented with a new round of searches at the time of revision of the manuscript (June 20, 2003). All completely sequenced and assembled microbial genomes that were submitted to the NCBI GenBank database as of April 2003, including 16 species of archaea and 96 species of bacteria were included used in this analysis. A complete list of these genomes and the predicted proteomes in fasta format can be downloaded from: <u>http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/micr.html</u>

Additional sequences, from microbial genomes that have been sequenced but not completely assembled and submitted to the GenBank database were also used in this analysis. A list of these prokaryotic genomes, from which sequences have been deposited in GenBank can be accessed from the following URL: <u>http://</u> www.ncbi.nlm.nih.gov/PMGifs/Genomes/eub\_u.html

Profile searches were conducted using the PSI-BLAST program with either a single sequence or an alignment used as the query, with a default profile inclusion expectation (E) value threshold of 0.01 (unless specified otherwise), and was iterated until convergence [71,72]. For all searches involving membrane-spanning domains we used a statistical correction for compositional bias to reduce false positives due the general hydrophobicity of these proteins. Multiple alignments were constructed using the T\_Coffee program [73], followed by manual correction based on the PSI-BLAST results. Signal peptides were predicted using the SIGNALP program [66,67]http:// www.cbs.dtu.dk/services/SignalP-2.0/. Multiple alignments of the N-terminal regions of proteins were used additionally to verify the presence of a conserved signal peptide, and only those signal peptides that were conserved across orthologous groups of proteins were considered as true positives. Transmembrane regions were

predicted in individual proteins using the TMPRED, TMHMM2.0 and TOPRED1.0 program with default parameters [34-36]. For TOPRED1.0, the organism parameter was set to "prokaryote" [34]http://bioweb.pas teur.fr/seganal/interfaces/toppred.html. Additionally, the multiple alignments were used to predict TM regions with the PHDhtm program [37]. The library of profiles for membrane proteins was prepared by extracting all membrane protein alignments from the PFAM database http:// www.sanger.ac.uk/Software/Pfam/index.shtml) and updating them by adding new members from the NR database. These updated alignments were then used to make HMMs with the HMMER package [74] or PSSM with PSI-BLAST. All large-scale sequence analysis procedures were carried out using the SEALS package http:// www.ncbi.nlm.nih.gov/CBBresearch/Walker/SEALS/ index.html.

Structural manipulations were carried out using the Swiss-PDB viewer program [75] and the ribbon diagrams were constructed with MOLSCRIPT [76]. Searches of the PDB database with query structures was conducted using the DALI program [77]. Protein secondary structure was predicted using a multiple alignment as the input for the PHD program [78]. Homology modeling was carried out using the Swiss-PDB viewer, version 3.7 to align a target sequence with template structures. This alignment was then provided as input to the SWISS-MODEL server [75] to generate a homology model using the PROMODII program. This model was then energy-minimized using the GROMOS96 routine of the SPDBV.

Similarity based clustering of proteins was carried out using the BLASTCLUST program ftp://ftp.ncbi.nih.gov/ blast/documents/blastclust.txt). Phylogenetic analysis was carried out using the maximum-likelihood, neighborjoining and least squares methods [79,80]. Briefly, this process involved the construction of a least squares tree using the FITCH program or a neighbor joining tree using the NEIGHBOR program (both from the Phylip package) [81], followed by local rearrangement using the Protml program of the Molphy package [80] to arrive at the maximum likelihood (ML) tree. The statistical significance of various nodes of this ML tree was assessed using the relative estimate of logarithmic likelihood bootstrap (Protml RELL-BP), with 10,000 replicates. Text versions of all alignments reported in this study can be downloaded from: ftp://ftp.ncbi.nih.gov/pub/aravind

## **Authors' contributions**

Author 1 (VA) contributed to the discovery process, preparation of most of the figures. Author 2 (LA) conceived the study and contributed to the discovery process, preparation of one of the figures and the manuscript. All authors read and approved the final manuscript.

#### References

- 1. Alberts B, Bray D, Lewis J, Raff M, Roberts K and Watson JD: In Molecular Biology of the Cell Garland Publishing, Inc; 1999.
- Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J and Zipursky L: In *Molecular Cell Biology WH Freeman & Co*; 1999.
   Koretke KK, Lupas AN, Warren PV, Rosenberg M and Brown JR:
- Koretke KK, Lupas AN, Warren PV, Rosenberg M and Brown JR: Evolution of two-component signal transduction. Mol Biol Evol 2000, 17:1956-1970.
- Robinson VL, Buckler DR and Stock AM: A tale of two components: a novel kinase and a regulatory switch. Nat Struct Biol 2000, 7:626-633.
- Stock AM, Robinson VL and Goudreau PN: Two-component signal transduction. Annu Rev Biochem 2000, 69:183-215.
- 6. Blume-Jensen P and Hunter T: Oncogenic kinase signalling. Nature 2001, 411:355-365.
- Manning G, Plowman GD, Hunter T and Sudarsanam S: Evolution of protein kinase signaling from yeast to man. Trends Biochem Sci 2002, 27:514-520.
- 8. Bourret RB and Stock AM: Molecular information processing: lessons from bacterial chemotaxis. J Biol Chem 2002, 277:9625-9628.
- 9. Bray D: Genomics. Molecular prodigality. Science 2003, 299:1189-1190.
- 10. Lu ZL, Saldanha JW and Hulme EC: Seven-transmembrane receptors: crystals clarify. Trends Pharmacol Sci 2002, 23:140-146.
- 11. Pierce KL, Premont RT and Lefkowitz RJ: Seven-transmembrane receptors. Nat Rev Mol Cell Biol 2002, 3:639-650.
- 12. Mombaerts P: Seven-transmembrane proteins as odorant and chemosensory receptors. *Science* 1999, 286:707-711.
- 13. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, Szustakowki J, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S and Chen YJ: Initial sequencing and analysis of the human genome. Nature 2001, 409:860-921
- 14. Jones AM: G-protein-coupled signaling in Arabidopsis. Curr Opin Plant Biol 2002, 5:402-407.

- Lanyi JK and Luecke H: Bacteriorhodopsin. Curr Opin Struct Biol 2001, 11:415-419.
- Spudich JL and Luecke H: Sensory rhodopsin II: functional insights from structure. Curr Opin Struct Biol 2002, 12:540-546.
- Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M and Miyano M et al.: Crystal structure of rhodopsin: A G proteincoupled receptor. Science 2000, 289:739-745.
- Teller DC, Okada T, Behnke CA, Palczewski K and Stenkamp RE: Advances in determination of a high-resolution threedimensional structure of rhodopsin, a model of G-proteincoupled receptors (GPCRs). Biochemistry 2001, 40:7761-7772.
- Beja O, Aravind L, Koonin ÉV, Suzuki MT, Hadd A, Nguyen LP, Jovanovich SB, Gates CM, Feldman RA, Spudich JL, Spudich EN and DeLong EF et al.: Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science 2000, 289:1902-1906.
- 20. Jung KH, Trivedi VD and Spudich JL: **Demonstration of a sensory** rhodopsin in eubacteria. *Mol Microbiol* 2003, **47**:1513-1522.
- Bieszke JA, Spudich EN, Scott KL, Borkovich KA and Spudich JL: A eukaryotic protein, NOP-1, binds retinal to form an archaeal rhodopsin-like photochemically reactive pigment. *Biochemistry* 1999, 38:14138-14145.
- Anantharaman V, Koonin EV and Aravind L: Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. J Mol Biol 2001, 307:1271-1292.
- 23. Anantharaman V and 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.
- Anantharaman V and Aravind L: Cache a signaling domain common to animal Ca <sup>2+</sup> channel subunits and a class of prokaryotic chemotaxis receptors. Trends Biochem Sci 2000, 25:535-537.
- Aravind L and Ponting CP: The GAF domain: an evolutionary link between diverse phototransducing proteins. Trends Biochem Sci 1997, 22:458-459.
- 26. Ponting CP and Aravind L: **PAS: a multifunctional domain family** comes to light. *Curr Biol* 1997, **7:**R674-R677.
- Taylor BL and Zhulin IB: PAS domains: internal sensors of oxygen, redox potential, and light. Microbiol Mol Biol Rev 1999, 63:479-506.
- Iyer LM, Anantharaman V and Aravind L: Ancient conserved domains shared by animal soluble guanylyl cyclases and bacterial signaling proteins. BMC Genomics 2003, 4:5.
- Makarova KS, Aravind L, Grishin NV, Rogozin IB and Koonin EV: A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. Nucleic Acids Res 2002, 30:482-496.
- 30. Wolf YI, Rogozin IB, Kondrashov AS and Koonin EV: Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. Genome Res 2001, 11:356-372.
- 31. Aravind L: Guilt by association: contextual information in genome analysis. Genome Res 2000, 10:1074-1077.
- Huynen M, Snel B, Lathe W 3rd and Bork P: Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. *Genome Res* 2000, 10:1204-1210.
   Schaffer AA, Wolf YI, Ponting CP, Koonin EV, Aravind L and Altschul
- Schaffer AA, Wolf YI, Ponting CP, Koonin EV, Aravind L and Altschul SF: IMPALA: matching a protein sequence against a collection of PSI-BLAST-constructed position-specific score matrices. Bioinformatics 1999, 15:1000-1011.
- 34. Claros MG and von Heijne G: **TopPred II: an improved software** for membrane protein structure predictions. *Comput Appl Biosci* 1994, **10:**685-686.
- Hofmann K and Stoffel W: TMbase A database of membrane spanning proteins segments. Biol Chem Hoppe-Seyler 1993, 374:166.
- Krogh A, Larsson B, von Heijne G and Sonnhammer EL: Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 2001, 305:567-580.
- Rost B, Fariselli P and Casadio R: Topology prediction for helical transmembrane proteins at 86% accuracy. Protein Sci 1996, 5:1704-1718.
- Schaffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Koonin EV and Altschul SF: Improving the accuracy of PSI-

**BLAST** protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res* 2001, 29:2994-3005.

- Galperin MY, Natale DA, Aravind L and Koonin EV: A specialized version of the HD hydrolase domain implicated in signal transduction. J Mol Microbiol Biotechnol 1999, 1:303-305.
- Bork P, Brown NP, Hegyi H and Schultz J: The protein phosphatase 2C (PP2C) superfamily: detection of bacterial homologues. Protein Sci 1996, 5:1421-1425.
- Aravind L and Ponting CP: The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. FEMS Microbiol Lett 1999, 176:111-116.
- 42. Kobe B and Kajava AV: When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. Trends Biochem Sci 2000, 25:509-515.
- 43. Pei J and Grishin NV: **GGDEF** domain is homologous to adenylyl cyclase. *Proteins* 2001, **42:**210-216.
- Égan SM: Growing repertoire of AraC/XylS activators. J Bacteriol 2002, 184:5529-5532.
- Itou H and Tanaka I: The OmpR-family of proteins: insight into the tertiary structure and functions of two-component regulator proteins. *J Biochem (Tokyo)* 2001, 129:343-350.
- Kenney LJ: Structure/function relationships in OmpR and other winged-helix transcription factors. Curr Opin Microbiol 2002, 5:135-141.
- 47. el Hassouni M, Henrissat B, Chippaux M and Barras F: Nucleotide sequences of the arb genes, which control beta-glucoside utilization in Erwinia chrysanthemi: comparison with the Escherichia coli bgl operon and evidence for a new beta-glycohydrolase family including enzymes from eubacteria, archeabacteria, and humans. J Bacteriol 1992, 174:765-777.
- Davies G and Henrissat B: Structures and mechanisms of glycosyl hydrolases. Structure 1995, 3:853-859.
- Jacobson RH, Zhang XJ, DuBose RF and Matthews BW: Threedimensional structure of beta-galactosidase from E. coli. Nature 1994, 369:761-766.
- Baumgartner S, Hofmann K, Chiquet-Ehrismann R and Bucher P: The discoidin domain family revisited: new members from prokaryotes and a homology-based fold prediction. Protein Sci 1998, 7:1626-1631.
- 51. Bork P and Doolittle RF: Drosophila kelch motif is derived from a common enzyme fold. J Mol Biol 1994, 236:1277-1282.
- Lichtarge O, Sowa ME and Philippi A: Evolutionary traces of functional surfaces along G protein signaling pathway. Methods Enzymol 2002, 344:536-556.
- Madabushi S, Yao H, Marsh M, Kristensen DM, Philippi A, Sowa ME and Lichtarge O: Structural clusters of evolutionary trace residues are statistically significant and common in proteins. J Mol Biol 2002, 316:139-154.
- O'Hara PJ, Sheppard PO, Thogersen H, Venezia D, Haldeman BA, McGrane V, Houamed KM, Thomsen C, Gilbert TL and Mulvihill ER: The ligand-binding domain in metabotropic glutamate receptors is related to bacterial periplasmic binding proteins. Neuron 1993, 11:41-52.
- 55. Lespinet O, Wolf YI, Koonin EV and Aravind L: The role of lineagespecific gene family expansion in the evolution of eukaryotes. *Genome Res* 2002, 12:1048-1059.
- Holt JG: In Bergey's Manual of Systematic Bacteriology Baltimore; Williams & Wilkins; 1989.
- 57. Aravind L and Koonin EV: The HD domain defines a new superfamily of metal-dependent phosphohydrolases. Trends Biochem Sci 1998, 23:469-472.
- Kazakov AE, Vassieva O, Gelfand MS, Osterman A and Overbeek R: Bioinformatics classification and functional analysis of PhoH homologs. In Silico Biol 2002, 3:1.
- Smith RL, O'Toole JF, Maguire ME and Sanders CR 2nd: Membrane topology of Escherichia coli diacylglycerol kinase. J Bacteriol 1994, 176:5459-5465.
- 60. Walsh JP and Bell RM: Diacylglycerol kinase from Escherichia coli. Methods Enzymol 1992, 209:153-162.
- 61. Hooper NM: Families of zinc metalloproteases. FEBS Lett 1994, 354:1-6.
- 62. Bond JS and Beynon RJ: The astacin family of metalloendopeptidases. Protein Sci 1995, 4:1247-1261.

- Galperin MY, Gaidenko TA, Mulkidjanian AY, Nakano M and Price CW: MHYT, a new integral membrane sensor domain. FEMS Microbiol Lett 2001, 205:17-23.
- Brunskill EW and Bayles KW: Identification and molecular characterization of a putative regulatory locus that affects autolysis in Staphylococcus aureus. J Bacteriol 1996, 178:611-618.
- Brunskill EW and Bayles KW: Identification of LytSR-regulated genes from Staphylococcus aureus. J Bacteriol 1996, 178:5810-5812.
- Nielsen H, Engelbrecht J, Brunak S and von Heijne G: A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Int J Neural Syst 1997, 8:581-599.
- 67. Nielsen H, Engelbrecht J, Brunak S and von Heijne G: Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* 1997, 10:1-6.
- Island MD and Kadner RJ: Interplay between the membraneassociated UhpB and UhpC regulatory proteins. J Bacteriol 1993, 175:5028-5034.
- Verhamme DT, Postma PW, Crielaard W and Hellingwerf KJ: Cooperativity in signal transfer through the Uhp system of Escherichia coli. J Bacteriol 2002, 184:4205-4210.
- Mougel C and Zhulin IB: CHASE: an extracellular sensing domain common to transmembrane receptors from prokaryotes, lower eukaryotes and plants. Trends Biochem Sci 2001, 26:582-584.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997, 25:3389-3402.
- Aravind L and Koonin EV: Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. *J Mol Biol* 1999, 287:1023-1040.
- Notredame C, Higgins DG and Heringa J: T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 2000, 302:205-217.
- Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M and Sonnhammer EL: The Pfam protein families database. *Nucleic Acids Res* 2002, 30:276-280.
   Guex N and Peitsch MC: SWISS-MODEL and the Swiss-Pdb-
- Guex N and Peitsch MC: SWISS-MODEL and the Swiss-Pdb-Viewer: an environment for comparative protein modeling. Electrophoresis 1997, 18:2714-2723.
- 76. Kraulis PJ: Molscript. J Appl Cryst 1991, 24:946-950.
- 77. Holm L and Sander C: Protein structure comparison by alignment of distance matrices. J Mol Biol 1993, 233:123-138.
- Rost B and Sander C: Prediction of protein secondary structure at better than 70% accuracy. J Mol Biol 1993, 232:584-599.
- Felsenstein J: Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol* 1996, 266:418-427.
- Hasegawa M, Kishino H and Saitou N: On the maximum likelihood method in molecular phylogenetics. J Mol Evol 1991, 32:443-445.
- Felsenstein J: PHYLIP Phylogeny Inference Package (Version 3.2). Cladistics 1989, 5:164-166.

