5260–5279 Nucleic Acids Research, 2004, Vol. 32, No. 17 doi:10.1093/nar/gkh828

Comparative genomics of the FtsK–HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging

Lakshminarayan M. Iyer, Kira S. Makarova, Eugene V. Koonin and L. Aravind*

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

Received May 25, 2004; Revised and Accepted August 27, 2004

ABSTRACT

Recently, it has been shown that a predicted P-loop ATPase (the HerA or MIaA protein), which is highly conserved in archaea and also present in many bacteria but absent in eukaryotes, has a bidirectional helicase activity and forms hexameric rings similar to those described for the TrwB ATPase. In this study, the FtsK-HerA superfamily of P-loop ATPases, in which the HerA clade comprises one of the major branches, is analyzed in detail. We show that, in addition to the FtsK and HerA clades, this superfamily includes several families of characterized or predicted ATPases which are predominantly involved in extrusion of DNA and peptides through membrane pores. The DNA-packaging ATPases of various bacteriophages and eukaryotic double-stranded DNA viruses also belong to the FtsK-HerA superfamily. The FtsK protein is the essential bacterial ATPase that is responsible for the correct segregation of daughter chromosomes during cell division. The structural and evolutionary relationship between HerA and FtsK and the nearly perfect complementarity of their phyletic distributions suggest that HerA similarly mediates DNA pumping into the progeny cells during archaeal cell division. It appears likely that the HerA and FtsK families diverged concomitantly with the archaeal-bacterial division and that the last universal common ancestor of modern life forms had an ancestral DNA-pumping ATPase that gave rise to these families. Furthermore, the relationship of these cellular proteins with the packaging ATPases of diverse DNA viruses suggests that a common DNA pumping mechanism might be operational in both cellular and viral genome segregation. The herA gene forms a highly conserved operon with the gene for the NurA nuclease and, in many archaea, also with the orthologs of eukaryotic double-strand break repair proteins MRE11 and Rad50. HerA is predicted to function in a complex with these proteins in DNA pumping and repair of double-stranded breaks introduced during this process and, possibly, also during DNA replication. Extensive comparative analysis of the 'genomic context' combined with in-depth sequence analysis led to the prediction of numerous previously unnoticed nucleases of the NurA superfamily, including a specific version that is likely to be the endonuclease component of a novel restriction-modification system. This analysis also led to the identification of previously uncharacterized nucleases, such as a novel predicted nuclease of the Sir2-type Rossmann fold, and phosphatases of the HAD superfamily that are likely to function as partners of the FtsK–HerA superfamily ATPases.

INTRODUCTION

Cell division in bacteria is mediated by several distinct protein complexes which are involved in chromosome segregation, choice of the division site and partitioning of the chromosomes between the daughter cells (1,2). In bacteria, unlike in eukaryotes, DNA replication, chromosome segregation and cell division are not temporally ordered by the phases of the cell cycle during which checkpoints ensure the proper progression of these events. The key event in bacterial cell division is the assembly of the oligomeric Z-ring formed by the tubulinrelated GTPase, FtsZ (3,4). This ring typically forms near the center of the bacterial cell, where the DNA concentration is low. Aberrant formation of the Z-ring in regions closer to the poles of the cell is prevented by the action of the MinD ATPase and an associated protein complex (5). The FtsZ ring recruits another key cell division protein, FtsK, via interactions with its N-terminal region, and FtsK, in turn, recruits several additional cell division proteins to the ring complex (6,7). FtsK is a large protein that consists of an N-terminal transmembrane domain with four membrane-spanning helices, a central coiled-coil region and a C-terminal P-loop ATPase domain. Although disruption of the ATPase domain of FtsK is not lethal in Escherichia coli, the mutant cells are defective in

*To whom correspondence should be addressed. Tel: +1 301 594 2445; Fax: +1 301 435 7794; Email: aravind@ncbi.nlm.nih.gov

chromosome segregation as well as septation and exhibit asymmetrically positioned nucleoids and large anucleate regions (8). These observations suggest that the ATPase activity of FtsK is required for proper chromosome segregation (9,10). After the replication of bacterial circular chromosomes, homologous recombination can lead to the formation of dimeric circles (9,11-13). Recombinases XerC and XerD act in concert to resolve these dimers (14). The ATPase domain of FtsK tightly regulates the Xer recombinases and mediates a switch in the catalytic state of XerCD such that XerD initiates duplex recombination (9,10). Experiments in the Bacillus subtilis and the E.coli systems indicate that the FtsK protein translocates along DNA and mediates pumping of the chromosome across the closing septum (12,15). Furthermore, FtsK also interacts with the ParC subunit of topoisomerase IV and recruits it to regions close to the septum. Additionally, FtsK activates chromosome decatenation by topoisomerase IV (16). These observations indicate that FtsK plays a central role in chromosome segregation both by activating recombination and decatenation and by pumping the chromosomal DNA across the septum.

ATPases related to FtsK from Gram-positive bacteria and actinomycetes, with multiple ATPase domains, have been proposed to function as pumps for the extrusion of small polypeptides of the ESAT-6 superfamily (17). Sequence analysis has shown that FtsK belongs to a family of P-loop ATPases which also includes two proteins of the type IV secretion systems (T4SS), VirB4 and VirD4, and the TrwBlike proteins involved in the conjugal transfer of plasmids (18,19). The VirD4-like ATPases of agrobacteria and other conjugative plasmids are required for the coupling of plasmid DNA processing by the relaxosome to the mating bridges (20). VirB4 is involved in the transfer of agrobacterial T-DNA into the plant hosts (21,22). VirD4 and VirB4 proteins of other T4SS have been implicated in the extrusion of protein virulence factors or cell surface structures in an ATP-dependent manner (23,24).

The solution of the crystal structure of the TrwB protein from the conjugative plasmid R388 revealed that these proteins form a hexameric ring, which is similar to the tertiary structures of a number of other P-loop ATPases, such as those of the AAA⁺ and the RecA/DnaB-like classes (25,26). A general model for the functioning of these ring ATPases has been suggested whereby the substrate (e.g. DNA) is threaded through the central pore of the ring and the ATPase activity facilitates pumping of the substrate (26,27) and/or (dis)assembly of other symmetric structures on the face of the ATPase ring.

Recently, we and others have shown that all archaea and some bacteria encode a highly conserved homolog of FtsK, TrwB and VirB4/VirD4 named HerA (the name used hereinafter) (18) or MlaA (28). The HerA protein from *Sulfolobus acidocaldarius* has been shown to have a bi-directional (3'-5'and 5'-3') DNA helicase activity (18). Electron microscopic studies have shown that MlaA from *Pyrococcus furiosus* forms hexameric rings similar to those formed by TrwB (28). The archaeal HerA proteins define a new family of FtsK-related ATPases, which includes additional divergent paralogs of HerA encoded in most archaeal genomes, as well as homologs from several phylogenetically distinct bacterial lineages (18). Examination of gene neighborhoods of the *herA* gene in archaeal genomes revealed strict co-occurrence in the same predicted operon with the *nurA* gene, which encodes an archaeal $5' \rightarrow 3'$ nuclease (29). In addition, the *herA* and *nurA* genes often co-localize with the genes coding for components of the highly conserved DNA repair complex comprised of the archaeal orthologs of the eukaryotic Mre11 (a nuclease of the calcineurin-like phosphoesterase fold) and Rad50 (a P-loop ATPase of the ABC class) (18,28). Given that conserved gene neighborhoods in prokaryotic genomes are strong predictors of functional and physical interactions (30–33), it seems likely that these four proteins interact to form a DNA processing complex involved in DNA repair, replication and/or segregation during cell division.

While both prokaryotic superkingdoms, bacteria and archaea, share certain similarities in their cell division (34), little is known of the chromosome segregation process in archaea. The strict conservation of HerA in the archaea with sequenced genomes parallels the nearly ubiquitous presence of FtsK in bacteria, suggesting that HerA might have a biological role similar to that of FtsK in maintaining genome integrity and facilitating chromosomal separation during cell division. Furthermore, given the bacterial-archaeal split in the distribution of the HerA/FtsK ATPases, reconstruction of their evolutionary history might shed light on the origins of cell division and associated DNA processing, and the nature of these processes in the last universal common ancestor (LUCA) of cellular life forms. With this objective, we performed a detailed computational sequence analysis of the FtsK/HerArelated proteins, identified novel members and explored their evolutionary relationships as well as their relationships with other P-loop ATPases. In addition, we employed a comparative genomic approach to extract contextual information for the HerA/FtsK superfamily, which led to the identification of previously unrecognized probable functional partners of these ATPases, including nucleases and transmembrane proteins. These leads allow us to predict the structure of the chromosome separation and cell division apparatus of the archaea and several bacteria that lack FtsK. We propose that HerA and FtsK, along with several families of ATPases encoded by plasmids, conjugative transposons and viruses, constitute a superfamily of ATPases descending from an ancestral DNA-pumping enzyme of LUCA. Members of this superfamily appear to have been repeatedly used as ATP-dependent pumps, for the partitioning of DNA into the daughter cells during division, extruding proteins into the extracellular space and packaging DNA into viral capsids.

MATERIALS AND METHODS

The non-redundant (NR) database of protein sequences (National Center for Biotechnology Information, NIH, Bethesda, MD) was searched using the BLASTP program. Iterative database searches were conducted using the PSI-BLAST program (35) with either a single sequence or an alignment used as the query, with the position-specific scoring matrices (PSSM) inclusion expectation (E) value threshold of 0.01 (unless specified otherwise); the searches were iterated until convergence. For all searches with compositionally biased proteins, the statistical correction for this bias was employed. Multiple alignments were constructed using the

T_Coffee (36) or PCMA (37) programs, followed by manual correction based on the PSI-BLAST results. All large-scale sequence analysis procedures were carried out using the SEALS package (38). Transmembrane regions were predicted in individual proteins using the TMPRED, TMHMM2.0 and TOPRED1.0 programs with default parameters (39). For TOPRED1.0, the organism parameter was set to 'prokaryote' or 'eukaryote' depending on the source of the protein.

Protein structure manipulations were performed using the Swiss-PDB viewer program (40) and the ribbon diagrams were constructed using the MOLSCRIPT program (41). Protein secondary structure was predicted using a multiple alignment as the input for the PHD program (42). Similarity-based clustering of proteins was carried out using the BLASTCLUST program (ftp://ftp.ncbi.nih.gov/blast/documents/README.bcl).

Phylogenetic analysis was carried out using the maximum likelihood, neighbor-joining and minimum evolution (least squares) methods (43–45) Gene neighborhoods were determined by searching the NCBI PTT tables with a custom-written script. These tables can be accessed from the genomes division of the Entrez retrieval system.

RESULTS AND DISCUSSION

The FtsK-HerA superfamily

In order to identify all members of the FtsK-HerA superfamily, we performed PSI-BLAST searches (35) of the NR database with PSSMs for the bacterial FtsK orthologs and archaeal HerA orthologs. These searches were run with E-value thresholds in the range from 10^{-4} to 10^{-7} to avoid inclusion of Ploop ATPases of other families with highly conserved Walker A and B motifs into the PSSM. As a result of these controlled searches, we collected a divergent set of FtsK-HerA homologs which contain several specific sequence motifs defining this superfamily, to the exclusion of other groups of P-loop NTPases. Reciprocal searches with newly detected members of the superfamily employed as queries were carried out to eliminate false positives. Only those newly detected sequences that reciprocally recovered other HerA/FtsK-related proteins as the best hits and/or contained the conserved motifs characteristic of this superfamily were included in the PSSMs for subsequent iterations. Exhaustive, transitive searches with the newly detected superfamily members were also employed to identify more divergent homologs.

The searches initiated with the FtsK and HerA PSSMs readily detected each other as the best hits (*E*-value of 10^{-6} – 10^{-8} at the point of first recovery in iterations 2-3) and also recovered the VirB4-, TrwB- and VirD4-like proteins $(10^{-6}-10^{-9})$ in iterations 3-5), and numerous uncharacterized proteins from diverse bacteria. Interestingly, these searches also recovered, with E-values in the range of 10^{-4} – 10^{-5} , the packaging ATPases of a variety of DNA viruses, including doublestranded DNA bacteriophages such as P9, large nucleocytoplasmic DNA viruses (NCLDV) (typified by the vaccinia virus A32 protein) and single-stranded DNA phages, such as F1 and M13 (gP1). All these proteins contain a unique set of conserved residues found only in the bona fide HerA and FtsK homologs, and none of them showed specific affinities to any other previously characterized group of NTPases (see below). Reciprocal searches with most viral proteins produced poor results due to their extreme sequence divergence. However, the A32 homologs from frog iridoviruses recovered the P9 packaging ATPase in iteration 2 (*E*-value = 10^{-6}) and the HerA proteins from the third iteration onwards. Reciprocal searches with the profiles for single-stranded DNA bacteriophages detected the HerA family as the best hits in the third iteration with *E*-value $\sim 10^{-2}$. Additionally, these searches recovered the Zonula occludens toxins (ZOT) from y-proteobacteria, such as Vibrio and Pseudomonas, suggesting that these proteins are derivatives of the phage packaging enzymes. With the sole exception of a small orthologous set of HerAlike proteins from filamentous ascomycete fungi, none of these searches recovered any eukaryotic cellular members. Taken together, the results of these searches suggest that HerA/FtsK homologs form a large superfamily, which is distinct from all previously described groups of P-loop NTPases.

The sequence and structural signatures of the FtsK–HerA superfamily and relationships with other P-loop ATPases

Using the Gibbs sampling procedure (46), we identified three highly significant conserved motifs in the FtsK-HerA superfamily proteins; these motifs served as anchors for constructing a complete multiple alignment of the entire superfamily. The alignment was further refined by taking into account the secondary structure elements derived from the crystal structure of TrwB (PDB : 1e9r) (25). Superposition of the TrwB structure over the multiple alignment showed that the conserved core of the FtsK–HerA superfamily is a seven-stranded β -sheet with a 7615423 topology. The last strand in the sheet is antiparallel to the rest of the strands (Figure 1). The first conserved block includes the Walker A motif and encompasses the first strand, the P-loop and the following helix. In most FtsK-HerA superfamily members, the P-loop has the canonical form (GX4GK[TS]), but some of the phage packaging enzymes and the ZOT deviate from this pattern (Figure 2) (47). With the exception of the viral proteins, most members of this superfamily have a conserved histidine at the beginning of the Walker A-associated strand (Figure 2). In TrwB and, by inference, in other FtsK-HerA superfamily proteins, this histidine packs against a conserved hydrophobic residue at the C-terminus of the helix located immediately downstream of the Walker B motif (Figure 2). The Walker B motif defines the second conserved block, which has the consensus sequence hhhh[DE]E (where h is any hydrophobic residue). The first acidic residue, which is conserved in all P-loop NTPases, coordinates the Mg^{2+} cation involved in NTP hydrolysis; the second acidic residue that is present only in a subset of the P-loop enzymes primes a water molecule for the nucleophilic attack on the gamma-phosphate. The third conserved motif includes strand 4, which contains a polar residue, most often glutamine, at the C-terminus and the distinct helix with a highly conserved arginine that precedes this strand (Figures 1 and 2). The conservation pattern associated with this motif helps in distinguishing the FtsK-HerA superfamily from all other P-loop ATPase groups. In the three-dimensional (3D) structure of TrwB, strand 4 is positioned in between the Walker A and Walker B strands (Figure 1) within the core β -sheet. The C-terminal polar residue of strand 4 is structurally equivalent to the polar residue in the so-called sensor-1 motif



Figure 1. Topology diagram of the ASCE ATPases showing the putative higher order relationships of the FtsK–HerA superfamily. Strands are shown as arrows with the arrowhead at the C-terminus, helices are shown as cylinders. Strands and helices conserved across the ASCE group are numbered, and colored yellow and blue, respectively. The C-terminal β -hairpin synapomorphic to the RecA–ABC clade and the helix-strand unit synapomorphic to the RecA clade are colored pink. This hairpin is secondarily lost in most helicases. STAND is a large clade of NTPases that include the previously described AP-ATPases and NACHT NTPases, as well as several uncharacterized ATPase lineages predicted to participate in signal transduction (D. D. Leipe, Eugene V. Koonin and L. Aravind, unpublished data). Non-conserved secondary structural elements are colored white. Abbreviations: WA, Walker A; WB, Walker B and Sen1, sensor-1. The dotted connecting lines in the topology diagrams represent regions of the protein where insertions are observed. Broken lines in the clade supported by the broken line.

of the AAA⁺ ATPases and the corresponding motif III of the SFI and SFII helicases (48–50). This residue appears to be required for sensing the triphosphate moiety of the bound nucleotide to trigger its hydrolysis.

Examination of the crystal structure of TrwB shows that the highly conserved arginine in the short helix upstream of strand 4 projects into the ATP-binding active site of the preceding protomer in the hexameric ring (25). Thus, this arginine is analogous to the arginine finger of the AAA⁺ superfamily, which is located at the C-terminus of the helix following the sensor-1 strand of the AAA⁺ ATPase domain (48,49,51). As in the AAA⁺ ATPases, this conserved residue of the FtsK–HerA superfamily is likely to function as an arginine finger promoting inter-protomer cooperation in ATPhydrolysis by binding the terminal phosphate of the substrate. Analogous arginine fingers supplied by adjacent subunits are implicated in cooperative ATP hydrolysis by ring ATPases of the RecA-like class, such as RecA, DnaB and ATP synthase (52–54). However, in this case, the arginine is located at the C-terminus of the ATPase domain, on a β -hairpin unique to the RecA-like ATPases (55) (Figure 1). Thus, the presence of a conserved arginine in P-loop NTPases is a good predictor of ring formation and inter-protomer cooperation. By this criterion, all FtsK–HerA superfamily ATPases are predicted to form hexameric rings. By contrast, in the PilT/VirB11 family, the arginine finger is supplied by a distinct domain, which is located in the same polypeptide, N-terminal of the ATPases, in which the arginine finger is supplied by an external GTPase-activating protein (56–59).

The presence of the additional catalytic glutamate after the conserved acidic residue in the Walker B motif and an intervening strand between Walker A- and B-associated strands place the FtsK–HerA superfamily into the additional strand conserved \underline{E} (ASCE) division of P-loop NTPases, which also includes the RecA-like, SFI/II helicase, ABC, AAA⁺,

PilT/VirB11, KAP and STAND (a large clade of NTPases that include the previously described AP-ATPases and NACHT NTPases, as well as several uncharacterized ATPase lineages predicted to participate in signal transduction; D. D. Leipe, Eugene V. Koonin and L. Aravind, unpublished data) clades (49,60,61). Structural comparisons show that the FtsK-HerA superfamily shares a C-terminal hairpin, formed due to the presence of a terminal anti-parallel strand, with many other NTPases of the ASCE division, namely, the ABC, RecA-like and SFI/II helicase classes, and the PilT/VirB11 superfamily (Figure 1). Furthermore, the latter three groups have an additional common feature with the FtsK-HerA superfamily, namely, an additional strand 'to the right' of the core P-loop fold (Figure 1). Clustering based on DALI Z-scores (62) suggests grouping of the PilT/VirB11 and FtsK-HerA superfamilies into a single, higher order cluster. However, no definitive sequence or structural synapomorphies unifying these two groups were detected. Additional 3D structures from diverse representatives of each of these groups should help in assessing the validity of this higher order clustering.

The FtsK-HerA and the PilT/VirB11 superfamilies are both traceable to LUCA (see below) and appear to have diverged from each other at an even earlier stage of evolution. Additionally, two other superfamilies of the ASCE division, the adenoviral packaging ATPases and the terminases of diverse DNA viruses, such as the herpes viruses and Mu-like bacteriophages (63), also have an additional strand to the 'right' of the core P-loop domain, suggesting an evolutionary relationship with the FtsK-HerA and PilT/VirB11 superfamilies. However, the conserved sequence motifs characteristic of any of the latter superfamilies are not detectable in the viral ATPases.

		Str1	Helix-1	Str-2	Hell.1	H1.2	Str-2.1	Helix	-2	
Sec.structure		.EEEEEEE	HHHHHHHHHhhhh	EEEEEE	НННННН	.HHH	.eeEEEe	hhhhhhhhh	IHHHh-	
MJ1429_Mj_15669620	122	REFAILSI-TGG	GKSNTASVLCRELAK.	GTVIMIDPHG	EYISLY (129)EFEIG	_KINIVNLSGLEI	QMVTFVGF1	KHLL	
APE0107_Ape_14600455	167	RELAILAV-TGG	GKSNTVCVITRRLVG.	VTTVIFDRHG	EYGTLG (137)IIEPG	KATIMDLSMMED	VADAVASHY	RRLL	
MG02890.4_Mgi_38100534	189	SSVFICGS-QGS	GKSHSLSCLLENCLI	YDT FV SDNGGEI	PCEAAYLAS (134) GVAPK	KVKARKNEQGTD	LFNICLSLF	EQD-	
SN9427.2_Asni_40747673	92	SSIFICGS-QGS	GKSHTLSCLLEGCLI	YDT FIGD QGGSH	PCEAAFLAT (123)ASTSK	KGKKAVNDAGTS	WSPKVF	EQD-	
TM1257_Tma_15644013	1/1	ALINVSGQSGVA	AK SYTTFLVKSMIE	ARY IIFNVKG	ESLLFL (204)WNVPG	RVTVIDISKLRE	AQAFVVGAI	SEVMenn	
aq_1682_Aae_15606//9	150	ALISISGMSGVA	TRUSYALFLLYSIFQ	IHGIIFNVKG	KDLLW1 (1/6) DWRSH	QLNVIDISDLHG	AKMEVVGS1	KK1Fsmm	
Npun6235_Npu_23129935	193	RSNGVFGK-SGT	GKSFLTRLLLAGVIR	AVNLIFDMHS	EYGWEA (13/	CLESG	-KNVVIEFGSQSN	SYMLVTNMI	RRIH	
S110284_Syn_16331876	190	RSNGVFGA-SGT	GROF LIKILLISGIIH	AVNIMEDMHS	DI TOUR (130	/CLEAG	MIDIPOSINDEDA	SIMLAINMI	RRIHem	
1)gR_EC_10132005	54	PICTUACA-WOW	CHEVELOI TREOLSA	VPVEMADVKG	DLIGVA(129) FFGEP	FIDDEDITEDWERTDA	COCTICITO	ERLINE	CONCILCULATION OF
h111925 Bian 27377036	178	PUT AVI ON-TOP	CROCTURCE TRUCKE	ARETUI DENC-	FVADAR(131	ACASHC	OTAUUNT CLUDS	TVUTVUEUT	DIVE	
ag 1852 Aae 15606891	306	M MAULOT-TOS	GKTTEVKKTLENEKE	SEVITEDIYG	EVAOEL (111	INSEE-	BURVENEREVDT.	TKUNLTGLT	KETE	
MA0204 Mac 20089102	59	HAVLICGK-RGY	GROYTMGCMLEELAF	LASLI IDTMG	IFWTMS (130	LISGG	BTTVLDVSPLEN-	VSAAAVSTL	GRLY	Contra Contra
APE2093 Ape 14601838	214	G IGVFGS-TGS	GKTTTLATIACGAAE.	LPVVVADWHG	EYORLV(104) PIGGR	GLYIVDTGSIGN	LRKTYSILL	AYLO	
SS00283 Sso 15897226	233	R IGIFGS-TGS	GKTNTAILLASOLNA.	VKVIILDWHG	EYDQLL (112) DYLT-	SSTIINLGFISN	LRKLYSLFL	KLIV	
CT1915 Chte 21674727	246	QKSALFGM-TRT	GKSNTTKI IA KSVFE	IGQIIFDPNG	EYANEN (194	WADED	LKKIIGIFQYSN	TRKIGKAAE	HSAD	
HH1039 Hehe 32266538	221	RRTAFFGM-TRT	GKSNTIKIIISAIEN	IGQIIFDVNG	EYTFSN (179) FNYLH	SLKGENDYKKDI	SALREGKII	VDLS.cmi and	
CagE Hp 15611559	591	G TLILGS-TGS	GKTVF <mark>M</mark> SMTLNAMGQ	LTMVYMDKDY	GAYGNIV(134)LDFSK	TIIGVDGSSFLD	NDVSPFICF	LFAR	
lvhB4 Lepn 19919310	465	GLAAIFGG-NNA	GKTTLVNFLDAQMGR	GRSFFIDRDE	SSKIYIL(116) LNLDF	DKVGFDVTYLMD	SVIATPVYL	LLHR	-
RP103_Rp_15603980	452	GUTLIGP-TGA	GKTVLMNFLCAEAQK	-PRMFFFDKDR	GAEIFIR(110) DLQRA	RVFGFDMTELLK	PVSLAPVLL	IFHRom	
BMEII0028_Bme_17988372	451	GNTRIIGQ-SGA	GKTVLMNFCLAQAQK	MCNVFFDKDQ	GAKGTIL(121) DFNTH	SNYGFDGTDFLD	ADVRTPISM	LLHR	
VirB4_Cje_32469876	467	GTMIIGG-TGA	GKTTLAQF LM CNLYK	IDIFSMDKLR	GMYNFAT (102) LSFNK	QLSILNMDSILK	PTLASLTAS	IFHRomm	
Icmb_Lepn_7465640	483	WIDLVYAR-PGS	GKSVLSNALNLALCL	PRIAI IDIGP	SSSGLIS(193)FDIGD	RVVSLDLDEVAK	ESLNNVPEO	KEYHammer-	
YddE_Bs_16077561	464	PHCLITGD-TGN	GKSYLAKLIFNYISM	"IKSLYIDPKK	EMRKWIQ(133)ALSLE	RITILEVENMDL	SQLKSSAVM	ALGKenna	
CP81_Pae_37955734	544	A GFIFGP-TGS	GKSASLTNLISQMLA.	PRMFVAEAGN	SFGLLAD(164) TPWPE.	DLTVVDFATYAR	IAYISLLNT	NNIA	
TrhC_St_10957214	456	SYNMVVGATSCA	GKSFWVAYIINNYLG	AQIFVVDVGR	SYQGISE(142) PINFD	RFIVLELEELKG	TVVLMSIIQ	AQHAem	
Reut5675_Rme_22980948	468	G LFVIGP-IGA	GKSVFLNFLVSQADR	ARRIRFDKDR	STRIPTL (107) FAVS	DNLCIECGELFQ	PRAAALFTD	AFYR	
RP293_Rp_15604162	154	Q ALLFAP-TGS	GEGVGFVIPNLLFWT	DSVVVHDIKL	ENHNLTS (158) FKKVK	TTVYVGLTPDNI	RLQKLMQVF	QQAT	
V1rD4_W01_8885501	157	QALLFAP-TGS	GRGVGEVIPNLLFWT	DSVIVHDIKL	ENYEITS (158) FKKKK	ITVYVGLTPDNL,	RLRPLMQVF	QQAT	
TIWB_BC_1084123	100	CNECT HOT - YOR	GREVELKELKITGLL	DRMVI VDPNG	DMLSKE (138) DPNGG	MI FICONSDEUS	ALKPLISAW	DVVChinam	
Trap_EC_9507649	224	CNFCLHGT-VGA	CHARLEN IARO	DEVVIIDRSG	EFVESI(145) DOKNG	WUPISSNAUTHA	SLRPVISPW	SIAImust	
CMa0020 Sma 16262055	224	T MLEFAC-SCC	VETTONUUDTALDVT	CPI TCI DPST-	-EVAPMVS (165	TUCCK	KOVFINIDACI	SVPGTCPUT	CSLT	-
Trek Phen 10956651	167	FUCTAVAC-ARM	TRACTAT PATCHAD	CATTVASNBC	-DUVSHTV (195) FUART	DTMYAMSKDGDG	SCCATCCAT	GOTI	
SC06849 Scoe 32141315	147	DVAVATMA-PRS	GKTTSLATPSMLGAP	GPVLLTSNKA	AGDAFTT (185	FUTST	DTLFLISKDGGG	GASALTAAC	DSVM	
SE1414 Sau 27468332	854	PUALTAGA-TOS	SKSVCINSILMSLLY	LELLIDPKM	VELAPYN	LPHLV	SPVITDVKAATO	EEMEKRYKL	AOYH	in the second
ftsK Sty 16764321	1007	P LLVAGT-TGS	GKSVGVNAMILSMLY.	VRFIMIDPKM	LELSVYE (1)	IPHLL_	-TEVVTDMKDAAN	NEMERRYKL	SALG	
TP0999 Tp 15639983	481	PLLIAGA-TGS	GKSVCVNALILSILY	TKLLIDPKI	VELKLYN	IAHLL	TPVITEPKRALO	CEMERRYAL	EOLE	Constant of the
FtsKL Fnu 15426256	205	PUGLIAGS-TGS	GKTYFLNYIICNLLA	ADITFIDPKS	ADIKAVG (1)	LVNP-	QKTACTENQIAK	EEMEARQKI	GKSG	iteres
YdcQ Bs 16077553	228	PMLIAGG-TGG	GKTYF M LT II KACVG.	ADVRILDPKN	ADLADLE	EVLP	KKVYSQKNGILM	DGMMERMDE	KQMS	-
Chte0289 C1th 23020216	693	PEGLVAGT-TGS	GKSELLQS FI ISLAI	VVFVLIDYKG	GGMANAF (2)	LPHLV.	TITNLGGNQTTR	SELKRRQTI	AQYG	
YueA_Bs_16080238	395	PEGLLAGT-TGS	GKSEF <mark>L</mark> QT YI LSLAV	AAFLLIDYKG	GGMAQPF	RNIPH-	TITNIEGSKNFS	SELKKRORL	DQYQ	
P9_PM2_27923906	19	VEHLVVGA-TGS	GKSAFIRDQVDFKG-	ARVLANDVDE	DYRLPR (14)	FGAIR	CALTVEPTE	ENFERFCQL	FAIS	
MCP_PR4_215736	10	QRILVLGK-TGT	GKTCAAVWHLSQKD-	FKRKAWIVLN	HKGDDL (17)	KPGLY	IYHPIPD	-VDDAEVTQ	LWDI	
ORF13_Bthu_32398322	8	E VFIAGK-TGS	GKSFLAEVYLAGYE-	HVVMLD	TKGQCL (19)	VLVET	LEEVAEARTKKI	EQDEEHYDA	MKWV	
VPA0903_Vpar_28900758	142	G VIYVAG-TGG	GKTSAVKHLGLVPK	AAQAVFFDPY	RNYAGA (20)	VMARK	RGKSFKLAYIPK	ELEFFSAAV	AVGN.	
BCp0009_Bce_29899120	10	EVFLSGQ-TGT	GKSFTAEVYLSGVT-	DVDVVKLDTK	GEVHER (17)	TVVER	LADIDDVETRKI	EQEQEYYDK	MKWV	
B164_St1V_46360638	4	DIV VI IGR-KRS	GKSYLIKHYFIPVLK.	KISYIIDDHNL-	LRSGSEYS-	KFGYN	VVTLSDIVSKQY	VYDREKNDV	FEKL	
075L_CIV_15078788	30	SKIVVIGK-AGT	GKSTLIRYLLFLKR	SIIPVGMVVS	GTEDS-	NCFY	SDIFPPLFIH	DEYDEEI	KKF1-mmm	- (D
LSV_1_26_LSV_13242498	32	CNAVVIGE-RNT	GREVVIAELLIILNK	QKVPRACVFS	ATEES-	NREE-	CRHIPDSFIPD-	-EKNVELKL	EIVEemaan	1000
AS92R PBCV 9031900	1.0	ALVGVVGR-ROS	GROVIERDEDIIRE	CEVERTSI TO	CCEV	NGIE	VEVELVENUUEV	PNAUGLEAL	ERQ- mm	-
ARVIDU AMERY 3304404	40	FRANT TOC-SOS	TYTICI POTIL	WWWWITETET-	DUV	NDDVD	CYTHONUTNEY_		TTD	
Mimiuri rue	30	NTSUUCE-TOC	CMPTUTONTOPTU-	TIN	5.4.1	NCDT-	EUTTL#SNDET-	-VENUDOCY	ALSOT	
Tir 68p Teth 18481479	21	SSTLIAGO-RRS	GKTVLCINLLIVITE.	VSYOSTILES	DTAGIET	NSA	FDFIDKNFIF	-ESROLDET	PKIM	
CBG24838 Chrig 39579232	27	TRACTICK-SCC	GKINLEMNLLDKFE	LDYDSLYIES	TTLEOP (9)	NGLSK.	TRECEKNONEMV	-EKKDKOET	THTS	
C6 Rsol 10954642	6	PITLITAT-PGG	GKTALAVOMMK-AAV.	RPLEVMGIPE	LKLPYI		TPAVSDWTELRE	PENPGMMLP	FTF-	
Xfas01947 Xfas 22997913	2	PIHVITAL-PGG	GKTALMVEMLO-AEA.	RPLFAAGIDG	LOPGL-		AIVLDD	PSO	NAKD	
ORF301 Pf3 9626322	1	MITLITAV-PGS	GKTLYAIGLIE-AAL	GREVETNISG	LVKDK-	SN	PHLLSD	APDD	RD	
NMA1799 Nm 15794690	з	EICLITGT-PGS	GKTLKMVSMMA-NDE	RRKVFTNIKG	LKIPH-	YIETD	-LPKSTDEQLS	AHDMYE	IKKP	
NE0893 Neu 30248897	2	SITLITAA-PGA	GKTIFAVWNIIKPAV-	RVVYTAGIPE	LKLPA-	SLSYS	LVEVENPSGIP-	IPDDEKPSR	QN	
YP02279_Yp_16122503	2	AISAYIGI-PGS	GKSYEAVCNVIIPAF-	GRRVVTNIYG	LQKDK-	TERYP	EIIVVDNDDVLK	FPFKGGEGS	CQ	
Zot_VsKK_17975157	28	MIYAIVGR-PRS	GKSYESVVYHIIPAI.	GRKVITNIP	LNIPM-	EKVFG-	-LIKVIDAQFTEY-	SKVEHYLDD	RDG-	
d1025076_Pae_3237271	1	MINLILGQ-PGG	GKSHEAVVYHVVPAL	.GRKVITNLA	LDMDK-	KAFFP	_LIELRDSTVEVF	SRVDHYADP	RHPD	_
ZOT_Vc_282084	2	SIFIHHGA-PGS	YKTSGALWLRLLPAI	GRHIITNVRG	LNLER-	AKYLK	-SIEFIDTDHPDG	ARFWH	AR	
gI_M13_17426226	2	AVYFVTGK-LGS	GKTLVSVGKIQ-DKI	GCKIATNLD	LRLQN-	PQVGRim	VLRIPDKPSI	SDLLA	GRGN	
Mp_fs2_9630769	з	SVYFVTGK-LGS	GKSLIAVSRIR-DAL	GVPVATNLN	INLKE-	-MIGR.	-LYRLPDKPTV	EDIEI	GYAI	
Desu6419_Deha_23120453	9	GVVLYFAN-IGT	GKTTYLSKLVQ-AEL	YRYIVSNAV	ISGV			RALLKI	IAA	
P3_NANDHV_112602	26	KLEAVVGS-KGS	GKSLYMSRVAD-KWL	KGFIYSNMG	IGYD		LEPEY	WKQT	AP	
LPIG.16_BSph_38639872	2	nertrogee-Ler	GRALPHSILAHYL	NVELFSNYD	LKES		TTMLD=====	YTDW	NVA-	and Summer
consensus/sus		nn.uSGS	oronnn	nn.s			· · · · · · P · · · · · ·			

	Str-3	Hel-3	Helix-3	Strand-4	Helix-4	Strand-5		
Sec. Structure	EEEEEE.	hhhHHH	ННННН. ННННН	EEEEEEE	ЕННННННННН			
MJ1429_Mj_15669620	VTKPVLLIVEE	AHIFIPVNE.	-SASLWLGK IA RE-G	KFGVGLGLVS	RPK-QLHPDVLSQTNT		442\Classical HerA \	HERA
APE0107_Ape_14600455	YPVPVITIEE	AHVLIPRDE	LTKRWAARIARE-G	KFGVGLVIVS	RPK-KLDVDVLSQTNK	KIILKMVEPQDISYVR	4751	1
MG02890.4_Mgi_38100534	TKIGRVVALDE	AHKYMGGGT	TLTENLLQT I RL-Q	HQACRIFIST	EPTVSPALLDLCSV		510 \fungal hera	6
5N9427.2_Asni_40747673	TNVGRVVALDE	AHKYMKDS	IFTETLLSSVRL-Q	HLATRIVIST	EPTISADLLSLCSV	TIVHRFSSPAWLRALQ	3831/	6
TM1257_Tma_15644013	FTQPVFIFLDE	LNKYAPRHG	ALANIFRDVAER-G	SFRVILIGAE	TAS-EVDYRVITQAAI	VVVGRQKGAELIKPEY	5501	1
aq_1682_Aae_15606779	PYPKI FVVLDE	LNKYAPKDK-	PIKEILLDIAER-G	SLGVILIGAQ	TAS-EIEKRIVANAAV	KVTGRLDSSEVLSKEY	4931	1
Npun6235_Npu_23129935	RPTPIMITIEE	AHRFLDPA	VQSTIFGTIARE-L	RYFVTLLVVD	RPS-GIDNEVMSQIGI	RITALLNDEKDIDAIF	502 \s110284	1
s110284_Syn_16331876	RPRQLVITIEE	AHRFLDSA	VHQTIFGTIARE-M	KYFVTLLVVD	RPS-GIDNEVMSQIGI	RITCLLNDEKDIEAIF	505//	8
r]gR_EC_16132085	EKPKLVFFFDE	AHLLENDAP.	VELOKIEQVIRL-1	SKGVGVWEVS	NPS-DIPDNVLGQLGNP	-VQHALRAFTPKDQKAVK	335\Y]gR	£.
RV2510C Mtu 15609647	ORT DESIGNATION	AHLLFTDAS	AFLEQVEQTVRL-T	SKGVGVFFCT	DPT-DLPNDVLSQLGAP	- IQHALKAF TPDDHKALK	500/ 617\b111035	1
1923_Bjap_2/3//030	DEVERT TUTER	ANIE VERGE	UN FUVAFUTAME_C	NT GLGLVLSS	RPS-ELSPIVLAQCNI		61/(D111920	8
44_1032_Ade_15000091	FEDRUMTETOR	AUTEIDACD	ACEUT INDOLO	OPCICIUIAT	RFA-NLSKF 115QLNI		373	
APE2093 Ape 14601838	ELCPLINUTE	AHNMEDGE-	==ESEDSIMMAE=S	KEGLVTALAT	NPH-LIPLRAVSNTNT		474\\$\$00283	6
5500283 5so 15897226	TNOOTLITLDE	AONYFNRE-	-GNEFTDRLASE-T	KYNTGLOFTT	SPS-LLSONVLKNTNT		502/	1
CT1915 Chte 21674727	NTPETLYYVEE	AHNTLPAGN	DISDIWVRTAKE-GS	SKYRIGMVYAT	EVS-STOKNTLKNTAN	WFISHLNNTDETKELC	660\CT1915	6
HH1039 Hebe 32266538	EPEFIOMYFEE	AHNIFPKDD	DLKNIYNRLAKE-GA	AKLKIGISYST	EVS-SIAPSILKNTON	WFISHLNNKDEIKALE	589/	1
CagE Hp 15611559	DGRRFVLDIDE	AWKYLGDPK		KRNAIVRLAT	SITDLLACPIADTIREQC	PTKIFLENDGGNLSDYOR	900\ClassicalVirB4\\	JirB4
lvhB4 Lepn 19919310	DGRLTSFIIAE	AWOLFASPE.	EKALREWLPT-I	KKNGHFIFDT	SPKTITDSPIKHIVLDNI	ATLIAFPNPLADRETYME	7451	1
RP103 Rp 15603980	DGQKTMIVLDE	AWALIDNPV.	APKIKDWLKV-L	KLNTF VIFAT	SVEDAAKSSISDTLIQQT	ATOIFLPNLKATDI-YRS	7231	6
BMEII0028 Bme 17988372	DGRRFIYWMDE	AWKWVDDEA	-FSEFANNKQLT-I	KONGLEVFAT	MPSSLLNSKVASALVQQV	ATEIYLPNPKADYHEYTD	7401	£ .
VirB4 Cje 32469876	KKRGF FCFI DE	LKDFLMDEN	-MRESILESILE-V	KIGGVMCMGF	NLSFFDDIPKGASFLENI	ANY	734/	É .
Icmb Lepn 7465640	REDHKRIVYDE	FHRTAKSSA	-VREQVIIDMRE-G	KWKVQISLLS	AVDDFDPVMIDF	ATAIYVMDAGPSQAVEKT	867>IcMB	C .
YddE Bs 16077561	QDEQTVEFIDE	AWIFTTSQQ	-GKKVERQMRRI-G	SYNNAEYFIS	STKDALKEEDSGNE	-GVAFAFDEPNEREEVLK	768>YddE	£L.
CP81_Pae_37955734	KGRPIVKITDE	GHIITKHPL	-LLPYAMKITKM-W	KLGAWFWLAT	NIDDIPASGAPMLNM	11EWWLCLNMPPDEVEKIS	875>CP81	£.
FrhC_St_10957214	DGRRRLFILDE	AWEYIRPDN	FFSSFLEAAWRR-F	KINCAGICIT	SFEDYFTSSVGRALTANS	PWKIIMKQEKESIEAMKV	804>TrhC	1
Reut5675_Rme_22980948	GIRYTVIEVEE	CGFFFQNER.	YKRFEDWITT-I	KLNGAIWAAT	SLRQIARVANFEILKENI	ANWIYLPNSQAKTSTDLY	738 /	Berner
RP293_Rp_15604162	EPYGVMFLLDE	FPTLGKMDT,	KAGIAY-F	GYRVRLFLII	DTQQLKGTYEDAGMNSFL-SNS	TYR <mark>I</mark> TFAANNYETANLIS	478\ClassicalVirD4\\	/irD4
virD4_Wo1_8885501	EPYGVLFLMDE	FPTLGKMEQ.	QTGIAY-F	GYRVRLFLIV	DTEQLKGIYEEAGMNSFL-SNS	TYRITFAANNIETANLIS	480/	6
FrwB_Ec_1084123	PKRRLWLFIDE	LASLEKLAS	LADALTK-G	KAGLRVVAGL	STSQLDDVYGVKEAQTLR-ASF	RSLVVLGGSRTDPKTNED	425\TrwB	8
FraD_Ec_9507649	RNRRVWFFCDE	LPTLHKLPD	LVEILPE-A	KFGGC YVF GI	SYAQLEDIYGEKAAATLF-DVM	INTRAFFRSPSHKIAEFAA	494/	6
Trag_Atu_13990963	LEGRALFLLDE	VARLGYMRI	ETARDA-G	RYGITLIMIY	SIGQMRETYGGRDASSKW-FES	ASWISFAANDPETADYIS	553\trag	81
SMa0929_Sme_16262955	FRRRALFMLDE	VDLLGYMRL	EEARDR-G	RIGISMMLLI	SLGQLERHF-GRDGAVSW-IDG	CAPASYAAVKALDTARNI	54//	8
115K_Kneq_10956651	I DEPMIAVIDE	AANICKLED.	PRWISHEGG	GILLITEL	SPSQGEKVWGREAFDAMVDACS	WIWIGGNVKDDAIISGLV	DI9\TESK	10
SC00049_SC00_S2141515	PMPRIVITUTOR	LADIMMMAP	DVFOSTARTACK-A	ACCTHMINAT	B D C VNUTTCI TKANI	PTRTAFMUSSSUDSPTTI.	1034\Classical Ftek \	Frek
ftek Sty 16764321	KLPYTVVIVDE	FADIMMTVG	KVEELTARLAOK-A	AAGTHIVIAT	RPSVDVITGLIKANI	PTRTAFTVSSKIDSRIIL	12121	LELAN
TP0999 Tp 15639983	PLPFIVITIDE	FADLMVASG	ELETSVARLCAM-S	AVGTHINLAT	RPSTDVITCLIKANI	PSRTAFMVSSKMDSRTTL	665/	1
FtsKL Fnu 15426256	GMKPOFLIFDE	LAAFKAGVE	SVENOLKKIILM-G	STGNEVILVA	OPNAEVVETGIBDOI	GLEVAFONTENELBLMMF	379\Ydc0	22
(dc0 Bs 16077553	GLKPVFIFFDE	YVAFMDLLD	EALSYMKOLVML-G	OAGYFLVLGA	RPDAKYLADGIRDOF	SFRVSLGLMSDTGYGMMF	402/	íč.
Chte0289 Clth 23020216	PLPHLVIIADE	FAELKSEOP	DFMRELVSTARV-G	SLGVHLILAT	KPA-GVVDDOIWSNA	RFRICLKVOGPODSODVI	876\YueA	6
YueA Bs 16080238	AMPHL FLISDE	FAELKSEEP	DFIRELVSAARI-G	SLGVHLILAT	KPG-GIIDDQIWSNS	RFKVALKVQDATDSKEIL	580/	1
P9 PM2 27923906	AGAPMVVIVEE	LADVARIG-	KASPHWGQLSRK-G	KYGVOLYVAT	SPQ-EIDKTIVRQCNF	KFCGALNSASAWRSMA	178\dsDNA phage	A32
MCP_PR4 215736	AMGNIGVYVDE	GYMIP	NRDPAFQALLTQ-G	SKRIPMIILS	RPV-WLTRFAISESDF	FQIFQLGDQRDRQTVQ	165	0.000
ORF13_Bthu_32398322	ERENTILWIDE	LMQVAPSPT	KYPFHLRALMTR-G	SKEATWACT	RPA-TIPIDVFNDSSH	FFIFDLNIPADRDRIV	178	1
VPA0903_Vpar_28900758	DADQLHVIIEE	LASCVETSG	KLKGKAGE LW RG-G	QYG LVL HSIF	RGQ-EVPKTVTEQSPV	WWIGAVNSMADARWLADK	318	6
BCp0009_Bce_29899120	MRENTILWIDE	LMEVCPSHF	KYPPYLKG LM TR-G	SKEATVWACT	RPA-EIPAIVMGNATH	FFIFDMNLPQDREKIA	182	6
B164_StiV_46360638	KWGTS VLII DE	AYYHFKYRQ	KVTPAIDE AL HA-N	HAGIGLILST	RVY-DLMPIVYKQADI	,	164/	15
075L_CIV_15078788	ENPWAVLLLDD	CTEDKKIFS	SKWQQS LF KN-G	HWKLLYILSL	HAT-DIPPAIRTNVDG	V F IFRETNENNLKNIY	180\NCLDV	1
EsV_1_26_ESV_13242498	TDLRVAIVLDD	MGYNRKVLT		HYDITLIVAI	HVM-QLTPALRSNTDY	VICLKEGNKNVMRNLY	1911	8
A392R_PBCV_9631960	NMSKV FVVL DD	LAFDTSIMK	QPVMRYIFMN-G	HLNIFLIFSS	YVADLGPPAIRANIDI	LLVCREAIQANRWRLY	166	0
AMV150_AMEPV_9964464	ENFRILVIYDD	IGKDTKDK	LSNFTNV-C	HSLVSNIFLV	IRLE-HLDTTTRDSLSY	HVINSESENMD-LIPC	1721	6
A32L_VacV_6969825	KSAHFLLIFDD	GDRLSKCN	TLIEFLNF-G	HENTSIILLC	TYR-HVPILGRAN1TH	IFCSFNISISDA-ENML	2001	£.
Mimivirus	PONKKILVIDD	THUNDOW	LIDOLNCK C	ACTUTATION	CFV-NFIPLIKNNLTI	LEINNSVASSDIKSMI	1731	
DC24020 Chair 20570222	DEFERENCE	INDINKRSK-	LLDDLISK-S	CANTOCRATE	ISKIVIINIIKSNIDI		107/	1
CBG24030_CDE19_39379232	- PPNGT TVI DE	AODVFDUDT	SKUPDHUAAFET_U	HINNIDCEIIS	NDT	V-CONVELEDACITCEM-	150\SS DNA virue cla	de
Vfae01947 Vfae 22097913	VPDCSLIEVDE	AWKWEGMKY	KTTPDHVRALAF-H	HRGLOFUNTT	ONHKOLVSFVHGI	T-CRHTFIKERFCTRFI.D	1461	ncie
DRF301 Pf3 9626322	TPECSLVVYDE	ACCAHLYPS	PUTDERLTAMET-H	HTGHDLVFTT	APTFVHHHTBKI	M-CLHTHLYRSRGLOAAS	1371	
NMA1799 Nm 15794690	ENIGSIVIVDE	AODVWPARS.	SKIPENVOWLNT-H	HOGIDIFVLT	GSKLLDONLBTI	V-REHYHTASNEMGMRT-	1561	
NE0893 Neu 30248897	ITEGALIVIDE	VOYLWPASG	REPGEDIKYLTK-H	HHGLEFVLIT	APOLIHKNVLAV	V-DKHIHMLSDWHGRK	1561	
YP02279 Yp 16122503	FGDLIVIDE	AWRIFGSDK	-SFIAEHRHFTHP-ET	TGISCOLVIVN	SLSDK	I-ETTYRMRKLKALGLNN	1601	
Zot VsKK 17975157	KNRAPLYVIDE	AHMVIPTRL	-GDPKILEFYSM-HO	GHYGIDIIILT	NLRKIHADIRAM	II-EMTYYCAKNTAFGSKK	1851	
d1025076 Pae 3237271	EGFGPLYVIDE	CHLSIPLRG	-TPVPVEEWYSL-H	HELADVLLIT	SYGKINRAIRDI	V-QVVYRCKKATAFGTND	1661	
ZOT Vc 282084	KDAFLFIDE	CGRIWPPRL	DRPESFEVAFDM-H	HHGWDICLTT	NIAKVHNMIREA	A-EIGYRHFNRATVGLGA	1691	
gI M13 17426226	ENKNGLLVLDE	CGTWFNTRS-	KERQPIIDWFLH-A	KLGWDIIFLV	DLSIVDKQARSAI	A-EHVVYCRRLDRITLPF	1561	
Mp_fs2_9630769	TSKDGLIVLDE	CGTWFNSRT.	KNRQALLDRFLH-I	KLGWDVIFIV	NISMVDKQAREGI	A-EHVVHCKRLDRMQIPY	1551	
Desu6419_Deha_23120453	-MADTLLLIDE	GSIEYNNRK.	NLTEHEIR YL KL-I	HYKSTIIVVS	SHDDVDVTLRRI	Y-TQIYLLRYLPFFTLIQ	1431	
P3_NaNbHV_112602	DSLILIDE	IGVLHSNRD,	AMPREAVE FF KM-Q	KYHLTIVVSS	TMDFDKKIRDI	C-DRIYLCNRIGWFCRLT	1501	
LP1G.16_Bsph_38639872	EADTSIICWDE	AQVVFNNRA.	KFGQGIATE V AMYT	KLRSIQIYAT	PNVGNVDSRIRDI	I-EVVVTMRKDNKGYHLY	137/	
Consensus/80%	hhhh	hh		Rph.hhhhhs	2pp.	h		

Figure 2. Multiple alignment of the FtsK-HerA superfamily. Proteins are denoted by their gene names, species abbreviations and gi numbers, separated by underscores. Amino acid residues are colored according to their side-chain properties and conservation in the multiple alignment. The coloring reflects 80% consensus and is shown underneath the alignment. The secondary structure shown above the alignment, is derived from the crystal structure of TrwB and secondary structure prediction programs. E and H represent a strand and helix, respectively. The consensus abbreviations and coloring scheme are as follows: h, hydrophobic residues (ACFILMVWY) shaded yellow; s, small residues (AGSVCDN) and u, tiny residues (GAS) colored green; o, alcohol group containing residues (ST) colored blue; p, polar residues (STEDKRNQHC) -, acidic residues (DE) and +, basic residues (HRK) colored purple. The conserved histidine in the Walker A strand, the arginine finger and the glutamine in sensor-1 are shaded red. Secondary structure elements that are conserved across the ASCE fold are numbered as integers. Species abbreviations are as follows: Aae, A. aeolicus; AMEPV, Amsacta moorei entomopoxvirus; Ape, Aeropyrum pernix; Asni, Aspergillus nidulans; Atu, Agrobacterium tumefaciens; Bce, Bacillus cereus; Bjap, Bradyrhizobium japonicum; Bme, Brucella melitensis; Bs, B. subtilis; Bsph, Bacillus sphaericus; Bthu, B. thuringiensis; CIV, Chilo iridescent virus; Cbrig, Caenorhabditis briggsae; Chte, C.tepidum; Cje, Campylobacter jejuni; Clth, C.thermocellum; Deha, Desulfitobacterium hafniense; ESV, Ectocarpus siliculosus virus; Ec, E.coli; Ec, Plasmid R100; Fnu, Fusobacterium nucleatum; fs2, V.cholerae filamentous bacteriophage fs-2; Hehe, H. hepaticus; Hp, Helicobacter pylori; Lepn, Legionella pneumophila; M13, Enterobacteria phage M13; Mac, Methanosarcina acetivorans; Mgi, Magnaporthe grisea; Mj, Methanococcus jannaschii; Mtu, Mycobacterium tuberculosis; NaNbHV, non-A, non-B hepatitis-associated virus; Neu, Nitrosomonas europaea; Nm, Neisseria meningitidis; Npu, Nostoc punctiforme; PBCV, Paramecium bursaria Chlorella virus 1; PM2, Alteromonas phage PM2; PR4, Bacteriophage PR4; Pae, Pseudomonas aeruginosa; Pf1, Pseudomonas phage Pf1; Pf3, Pseudomonas phage Pf3; Rheq, Rhodococcus equi; Rme, Ralstonia metallidurans; Rp, Rickettsia prowazekii; Rsol, Ralstonia solanacearum; Rsph, Rhodobacter sphaeroides; Scoe, Streptomyces coelicolor; Sep, Staphylococcus epidermidis; Sau, Staphylococcus aureus; Sme, Sinorhizobium meliloti; StiV, Sulfolobus turreted icosahedral virus; Sso, Sulfolobus solfataricus; St, Salmonella typhi, Sty, Salmonella typhimurium; Syn, Synechocystis sp.; Tel, Thermosynechococcus elongatus; Teth, Tetrahymena thermophila; Tma, Thermotoga maritima; Tp, Treponema pallidum; VacV, Vaccinia virus; Vc, V.cholerae; Vpar, Vibrio parahaemolyticus; VsKK, Bacteriophage VSKK; Vvul, Vibrio vulnificus; Wol, Wolbachia sp.; Xfas, Xylella fastidiosa and Yp, Yersinia pestis.

Evolutionary classification and phyletic spread of the FtsK–HerA superfamily

We analyzed the relationships between the members of the FtsK-HerA superfamily using a combination of approaches. To identify the major clades within the superfamily, the multiple alignment was examined for distinct sequence signatures characteristic of subsets of the superfamily members. Clustering by sequence similarity using the BLASTCLUST program was employed to identify subgroups and orthologous lineages. Finally, at the level of high sequence similarity, such as within an orthologous group or a closely related cluster of paralogs, conventional phylogenetic tree analysis using maximum-likelihood, neighbor-joining and minimum evolution methods was performed to decipher the evolutionary history of each such group. As a result of this analysis, we identified six major clades within the FtsK-HerA superfamily: (i) HerA, (ii) VirB4, (iii) VirD4/TrwB, (iv) FtsK, (v) ssDNA phage packaging ATPases and (vi) A32-like dsDNA viral packaging ATPases. Table 1 shows the classification of the FtsK-HerA superfamily.

The HerA clade is defined by several synapomorphies including a small residue (typically, glycine) after strand 2, a hydrophobic residue in the α -helix after strand 2, an aspartate in the α -helix immediately after strand 5. This family also contains a large helical insert immediately after the second strand-helix unit of the P-loop domain. The HerA family proper, which includes the experimentally characterized HerA protein of *Sulfolobus*, consists of a core orthologous group of archaeal proteins that are encoded in a conserved operon with the Mre11 and Rad50 orthologs, many additional, more diverged paralogs from each archaeal genome and numerous homologs scattered over a wide range of bacteria (Table 1). The simplest interpretation of this phyletic pattern is that the HerA family emerged in archaea, followed by horizontal gene transfers (HGTs) to and between bacteria.

Most HerA family members are encoded in a conserved operon with a gene for a NurA nuclease (18,28,29). The HerA family proteins contain a distinct N-terminal β -barrel domain which is homologous to the N-terminal domain of F1/ F0 ATP synthases and is fused to the N-terminus of the P-loop domain (Figures 1 and 3); we named this domain the HASbarrel (HerA-ATP Synthase barrel). The HAS-barrel is likely to form an independently folding toroidal structure stacked on one surface of the central ring formed by the P-loop domain of HerA. The presence of several shared residues between the HAS barrels of ATP synthases and those of the HerA family (Figure 3), and an analogous location at the N-terminus of the P-loop, suggest that these domains have similar functions. In ATP synthases, this domain is implicated in the assembly of the catalytic toroid and docking of accessory subunits, such as the δ subunit of the ATP synthase complex (64). Similar roles in docking of the functional partner, the NurA nuclease, and assembly of the HerA toroid complex appear likely for the HAS-barrel of the HerA family.

The HerA clade also includes several additional families with substantial differences in domain/operon organizations and phyletic patterns (Figures 4 and 5, Table 1). For example, a distinct family of HerA homologs, found primarily in proteobacteria (typified by bll1925 from *Bradyrhizobium*), has a specific form of the HAS barrel only weakly similar to that of the HerA family. The CT1915 family includes a divergent group of proteins with a distinct N-terminal domain that appears to be unrelated to any previously characterized domains. This family has an unusual phyletic pattern, with representatives from *Chlorobium tepidum*, *Helicobacter hepaticus*, *Chloroflexus aurantiacus* and *Methanosarcina mazei*, suggesting a high degree of lateral mobility.

The prototype of the VirB4 clade is the VirB4 ATPase which is a component of the T4SS in numerous bacteria (65,66). This clade is unified by several distinctive sequence signatures, which include conserved patterns in the α -helical insert located after the second β/α unit of the conserved core of the P-loop domain (Table 1). Most VirB4-type ATPases also have a long N-terminal extension that is less conserved than the ATPase portion but is likely to form a distinct globular domain. This large globular domain probably mediates interactions with other components of T4SS or the conjugative apparatus of transposons and plasmids. There are several distinct families within this clade, the best studied of which is the classical VirB4 family that is encoded by the mobile T4SS gene clusters of diverse proteobacteria. Other families are encoded by conjugative plasmids and conjugative transposons [e.g. Tn916 of Gram-positive bacteria (67)] from various bacterial taxa (Figure 4 and Table 1). Consistent with this, the genes coding for VirB4-type ATPases show little evidence of vertical inheritance across genomes and appear to have been disseminated widely by these mobile elements.

The VirD4 clade is typified by the T4SS component VirD4 which has a large insert in the ATPase domain in the same position as the HerA and VirB4 clades. This insert shows several unique sequence motifs characteristic of this clade. Additionally, most members of this family contain a small membrane-spanning domain N-terminal of the ATPase domain; this domain probably functions as a membrane anchor. In addition to the family typified by VirD4 proper, there are several smaller families within this clade, which function as DNA pumps of diverse conjugative plasmids from various bacterial lineages (19) (Figure 4; Table 1). Not unexpectedly, the evolutionary pattern of this clade seems to mirror that of the VirB4 clade.

The FtsK clade consists of proteins that have no inserts in the ATPase domain, unlike the HerA, VirB4 and VirD4 clades. Most of the FtsK-like proteins contain N-terminal membranespanning segments that probably function as anchors. The main family in this clade, the FtsK proteins proper, is represented by conserved orthologous ATPases involved in cell division in the great majority of bacteria. However, in spite of its essential function, FtsK is missing in several bacterial lineages, such as Thermotoga, Aquifex, Chloroflexus and cyanobacteria. Phylogenetic analysis of the FtsK family suggests a predominantly vertical pattern of inheritance, with the tree topology resembling those of other proteins with an apparent dominant vertical component (e.g., ribosomal proteins and RNA polymerase subunits) (data not shown; Supplementary Material). Certain plasmids and phages, especially those from actinomycetes and Gram-positive bacteria, encode divergent variants of FtsK, which probably function in cis as DNA pumps for transmission of the respective plasmids during cell division or packaging of the phage DNA. Additionally, this clade includes a few smaller distinct families which consist, principally, of proteins encoded in conjugative transposons from various bacteria (Figure 4; Table 1). One notable

Table 1. Classification of the FtsK-HerA superfamily

Post-strand 2 α-helical insert superclade

Large α helical domain inserted after strand 2

HerA clade

- A small residue after strand 2; a hydrophobic residue in the helix after strand 2; a conserved aspartate in the helix immediately after strand 5.
- Classical Hera family (A > FF, Cya, Cau, Aq, Thth, Dr, Tma, Bha, Chth, Lme, Ruxy, Smu): Fused to a HAS barrel at the N-terminus; several members are in the neighborhood of NurA; fungi lack the HAS barrel
- bll1925 family (Pr, Fnu, Efae): Fused to a divergent HAS barrel at the N-terminus; several members are in the neighborhood of a Sir2-like predicted nuclease
- YjgR family (Pr, Spy, Chut, Act, Trde): Fused to a distinct α -helical domain at the C-terminus that has a conserved glutamate; glycine after conserved histidine before strand 1; PEhGD motif before Walker B strand; histidine in strand 5
- SSO0283 family (Sul, Ape, Pyae): Distinct N-terminal domain with two membrane-spanning helices
- CT1915 family (Chte, Hehe, Cau, Mma): Fused to a distinct N-terminal domain; Chte, Hehe and Mma are operonic with a methylase
- Ta1216 family (Tac, Sul): pNOB8 like conjugative plasmids

VirB4 clade

Distinct N-terminal domain with a conserved proline before the core ATPase domain; hxPh (x: any, h: hydrophobic) motif in post-strand 2 α -helical insert; alcohol or small residue prior to glutamine of sensor I strand

- Classical VirB4 family (Pr): Asparagine in strand 5; associated with type IV secretory system in proteobacteria
- CTnDOT-TraG family (Bdes): Present in conjugative transposons of Bacteroides and Porphyromonas
- CP81 family (Pr): EAG motif in strand 2; aspartate before helix 4; acidic residue after strand 5
- YddE family (LGC Gm +ve): PxxE motif after strand 2; encoded by Tn916-like conjugative transposons.
- TrhC family (Pr): aN (a: aromatic) motif before strand 1; IE (I: aliphatic) residue in helix 3; encoded by antibiotic resistant plasmids like R-27s
- R64-TraU family (Pr): Sporadically present in proteobacterial R64-like plasmids and genomes
- all8046 family (Cya): Distinct N-terminal domain with membrane-spanning helices that can exist as a solo domain; bN (b: big) motif before strand 1; DGosT (o: alcohol, s: small) motif after strand 2; hhGRI (h: hydrophobic) motif in strand 5; present on cyanobacterial plasmids
- TN5252 ORF26 family (LGC Gm +ve, Blon): Mostly present in conjugative plasmids and transposons
- VirD4 clade

Proline in post-strand 2 helical insert; aromatic residue in post-strand 2 α-helical insert; glutamine N-terminal to the helix following strand 4; serine after strand 5; most members also have two to four N-terminal membrane-spanning domains

- Classical VirD4 family (Pr): Associated with type IV secretory system in proteobacteria
- CtnDOT-MobC (Bdes): Present in conjugative transposons of Bacteroides and Porphyromonas
- TrwB family (Pr, Pchm): RxW motif in strand 3; mainly present on proteobacterial plasmids
- TraG family (α Pr): Present on several plasmids of α proteobacteria, often in the vicinity of TraA
- TrsK family (Act): Serine in place of conserved asparate in strand 2; conserved asparagine after strand 2; asparagine after Walker B strand; GS motif before strand 4; present on actinobacterial plasmids
- alr7539 family (Cya): Aspartate between helix 1 and strand 2; many members are in the neighborhood of a TrwC-like RCR superfamily nuclease; present on cyanobacterial plasmids
- TN5252 ORF21 family (LGC Gm +ve, Cgl, Smal): Mostly present in conjugative plasmids and transposons

Clades without insert after strand 2

FtsK clade

N-terminal membrane associated region; a variable coiled coil central region; C-terminal ATPase; glycine before core ATPase domain; Kh

- (h: hydrophobic) motif after strand 2; hxxR motif in helix 2; arginine or lysine at the beginning of strand 5; glycine after core ATPase domain
- Classical FtsK family (most bacteria): Histidine in helix before strand 2.1 and strand 4; RlsQ (l: aliphatic, s: small) motif in helix 3; DSR motif at the beginning of strand 6; SxhQR motif after core ATPase domain
- YueA family (LGC Gm +ve, Act, Cau): Three tandem ATPase domains often found in the vicinity of ESAT-6; some YueAs in Clostridium are fused to one to two FHA domains at their N-terminus; GGs (s: small) motif after strand 2 in first ATPase; WLPPL motif between the first two ATPase domains
- YdcQ family (LGC Gm +ve, Fnu): Alanine in Walker B strand; RD motif at the end of helix 4; present on Tn916-like conjugative transposons
- A32 clade
- ds DNA phage packing ATPase family (V, prophages)
- A32 ATPase family (V, Nem, Teth): Aspartate in Walker B strand and in the C-terminal strand after the core ATPase domain —TIr transposon subfamily (Nem, Teth); transposons in Tetrahymena are often associated with a TrwC/TraA-like superfamily I helicase (TIr 8Rp)
- -A32 ATPases of NCLDV viruses
- SS DNA virus clade (V)

Asparagine or glycine at the end of strand 2; C-terminal transmembrane region

A, Archaea; Act, actinobacteria; Ape, *Aeropyrum*; Aq, *Aquifex*; Bdes, Bacteroides; Bha, *Bacillus halodurans*; Blon, *Bifidobacterium longum*; Cau, *Chloroflexus*; Cgl, *Corynebacterium glutamicum*; Chte, *Chlorobium*; Chth, *Clostridium thermocellum*; Chut, *Cytophaga*; Cobu, *Coxiella*; Cya, cyanobacteria; Dr, *Deinococcus*; Efae, *Enterococcus*; FF, filamentous fungi; Fnu, *Fusobacterium*; Hehe, *Helicobacter hepaticus*; Lepn, *Legionella pneumophila*; LGC Gm +ve, low GC Gram positive bacteria; Lme, *Leuconostoc mesenteroides*; Nem, Nematodes; Pchm, Parachlyamydia; Teth, *Tetrahymena thermophila*; Mma, *Methanosarcina mazei*; Pr, proteobacteria; Pyae, *Pyrobaculum*; Ruxy, *Rubrobacter*; Smal, *Stenotrophomonas maltophilia*; Smu, *Streptococcus pyogenes*; Sul, *Sulfolobus*; Tac, *Thermoplasma*; Thth, *Thermus*; Tma, *Thermotoga*; Trde, *Treponema denticola*; V, viruses; > indicates lateral transfer.

family of the FtsK clade, typified by the YueA protein, is restricted to Gram-positive bacteria and actinomycetes and includes proteins with three tandem ATPase domains in the same polypeptide. These proteins are likely to dimerize and form toroidal structures with a total of 6 ATPase domains. The YueA-like proteins are implicated in the secretion of the unique extracellular peptides of Gram-positive bacteria and actinomycetes (17). Packaging ATPases of single-stranded DNA bacteriophages comprise another distinct clade in the FtsK–HerA superfamily. These proteins, which are encoded by gene 1 of filamentous enterobacteriophages (e.g., F1 and M13) consist of an N-terminal, cytoplasmic ATPase domain, followed by a membrane-spanning region and an extracellular domain. Proteins with similar architectures are encoded by a variety of filamentous phages infecting several proteobacteria such as

		Strl	Str2		Strand-3		Strand-4		Str5		Str6			
Secondary Structure		EEEEE	EEEEE		.EEEEEEEE		EEEEEEEEEE		EEEEE		EEE			
PH0932_Ph_14590784	9	IGIVRGESSF-	-INYE FSV NP	2	NISFGEFVVTKNR	1	GEWVLGVVRSVKN	1	ENEEVVATV		RILGKVDG	100\HAS	fused t	o HerA
AF1030_Afu1_11498635	4	VGLVMGKSSI-	-TDFS FAV NP	2	IPKFGEYVTAINR	1	GEEVIGIVREISN	1	KNDVI V ATA		TVIGVVKD	951		
MJECL08_Mj_10954499	9	VGTVVASKNV-	-NEFE FVI EN	4	KIKKGEFVITKNT	1	GDYLLSKITKIVS	1	NSSKFLASA		KILGVINN	1041		
Faci0361_Fac_22405468	11	IGYIIGENTS-	-GRFQ F VISD	2	NIKKWE YV YLKIK		NEIVIGRIEEIKS		NDFVNICIS		TILGKLKD	961		
aq_aa31_Aae_10957065	31	IKEE Y SE S FGD	DLLGFTVNP	14	DIGLHSYVEVKLQ		EGIVLGKITSIFA		DDEWKKAAI	16	DVIGILKD	155		
PAE2903_Pae_18313675	3	IGYI <mark>V</mark> AASTP-	-FEFIATLDP	2	PISLYDYVAVDHV	1	YDPSRGELINVRL	1	ILEVQIAKV		KVLGYVEG	971		
Chte1457_Cth_23021377	5	IGKLIGNTGNE	NDLK IAL EN	2	SAKRGEFVKIKHR	6	DTYVLGRIVSISR	6	TGETLFGTI		ELVGYRDN	991		
MJ1565 Mj_15669760	7	IGYTIGETRI-	DELTFLAKE		APKVGDYVKINYD		DSELLGMVESTIQ		SSYYI <mark>L</mark> GKI		KVLGDIRD	911		
APE0107_Ape_14600455	13	IGVI <mark>I</mark> GESSYS	SYSTILLERD	3	KVIVGSHVVSTLN		GRCVLGIVESIRS		VESTRYHVA	2	RWVSYLET	1061		
MTH307_Mth_15678335	4	VGRCYGETSP-	-WRVSFVSRE		MPGVGEYVVMEYD		GRRI LGMV ES <mark>L</mark> LR		GRQYVRGTV		RILGDVET	881		
HerA Sac_37665381	3	IGYIIGSATI-	-NEATAILEQ		KIRAGY YV ILEYD		GDKILGLITNVYT		P-FFIKARI		KLLCKLDG	891		
t112095_The1_22299638	6	LGIVVQGSLT-	-QGLE V RLSG	5	ELRVGQFLVVQGR		RSRFFCLLTDVTL		GTFATLSVA	3	MILDDEQE	101		
sll0284 Syn 16331876	6	LGSVTQGSLS-	-KGLEVRLHA	5	EMRVGKFLVIQGR		RSRFFCLLTDVSL		GTYGTLELA	37	YQANSNAD	1351		
glr4405 Glvi 37523974	18	IGTVVQGSLS-	-EGLEVRLSP	5	EMRVGKFCVVYGR		RTRFFSMLTDVTL		STFGTVQLT	5	ETVNGQSE	115		
Tery4431 Tery 23043894	7	LGSVIQGSLS-	-QGLEVRLHP	5	DMRVGKFLVVQGV		RAHFFCMLTDVLL		STYGTIELA	38	AQSSSQIK	1371		
Npun6235 Npun 23129935	7	LGSVIQGSLT-	-EGLE V RLHP	5	DMRVGKFLVVQGM		RSRFFCMLTDVAL		GTYGTINLA	35	PQTSTTME	134		
MTH542 Mth 15678570	5	AGQIIGGETA-	AVLIRQKA	2	PIELGDLLVAEGE		GYTIL-QVKDLRY		YGREVELLE	11	RPILHVRD	961		
SSO2200 Sso 15898975	9	IGIV <mark>L</mark> QKSEA-	NEMQGLIRA	2	EISVGQLLLVD		DSEKLSLVRVENY		EILDMNTII		KATLHLIK	941		
APE0080 Ape 14600433	39	MGEVVGRVTRY	SPVTTSPGS	15	GVRIGDYLCIVDP	2	LHIILGVVSTIKR	2	TSPESILTS	5	RLLLEADP	1441		
SSO2285 Sso 15899052	25	LGDLVGKVSRY	IPNKLDEEN	17	LGKIGIFLGAIDI	2	LYFVLLRVIGYER	2	SLITNVTLR	1	EMLTKVDF	1321		
DR0837 Dr 15805863	15	IGMVLGTEDVI	PTVFWFAVS	3	SVGLDDLVVVETR	5	PVRFYGLVDNVRK	5	LPASVSYAA		RVLVTRVD	1031		
ag 1682 Aae 15606779	3	VGIVLGTKPS-	-NPLEFWVG-		-VEKGKFLQLDDV	3	NSKIDGTNEEIKF	3	GVVPANLAY	2	RVSVTRIE	921		
TM1257 Tma 15644013	6	IGVVTGIFQS-	-SPYEFFVRM	1	AEKPGEAYKVFAQ	14	TVVTYGMIVDION	14	IYIAKVKVT	3	LKEGNKLL	1091		
tlr0250_The1_22297794	11	VGIVKGPGDS-	-GSEYVFITA	3	PVRIGEFVYYELS	6	SKSPAGAVHQVLG	6	LIGFTCEPA	7	EVIGEFHR	116		
PAE2998 Pyae 18313750	4	IGVVIKSPSI-	HYYI F RPFR	2	ELDVGA FVIA EVD		GVRVISRVTAIRH		VLYYTEAKA		VVLGARRG	971		
SMc01432 Smel 15965868	28	LGRVVACNGS-	RATIAAVAE	8	LWSVGKLVSISVG		TNRV VALV YSMQT		NNPFRIEVE		-LMGEVHV	1051		
Atu2038 Atu 17935924	28	LGRVIACNGA-	HATIAAETE	8	LWSVGRLISIEMG		TSRVTALVFSMRT		PNRLLIDVE		-LVGEVYR	105		
mlr1445 Mlo 13471466	28	LGNVVQCDGA-	RATISAYAD	8	LWTVGKMISINLG		TTRTVGLVYGIGK		QNAIEVSIE		-LIGEVRD	1051		
FNV2193 Fnu 34762304	6	IGRVISVDSF-	-KIMIELDEN	14	VAKVNSYVIVPIG		SDKIVALITRVKT		GIRFSKSKR	4	TMLGTITE	961		
EF2348 Efae 29376849	10	VALVVEVNGI-	RCKAITFDD	14	NLSVNS FV VIRQN		FIKIIGRINSESI		TIKRILDIQ		-IIGYISE	1031		
aq 1852 Aae 15606891	148	VGYAVSSRSP-	-KEAEIVLLE		DLKEQTYLGIQ		-GEEFFLCRLSNV		ALFAHLYER	8	EILGEYEK	239/		
Chlo0592 Cau 22970556	10	IGEVIESSTI-	HFVAATYEL	2	SPPFGSLVRATTT	2	GLHVYGLIYDIHT	2	DLSVVLQTE	3	LIVGYTLH	109\sold	HAS	
sll1318 Syn 16329340	18	IAEVIETSTT-	-GFLAQCLEP	7	MPAFGSWVKATDE	2	GNTIFAVVSYATT	2	QIFAMLTTE	3	AIVGFQSR	115		
tlr0637 Thel 22298179	11	FAEIIQTATD-	-HCIAQCHEP	7	VPALGSWVRIPEG	1	-RVIYGVVAYVT	1	HIFAMLKTE	3	AIAGFQER	1061		
Cwat205701 Cwat 45526704	18	IAEVIETSTT-	-QFLAQCLEP	7	MPPFGSWLKSLDE	2	GNKI IAVVTYATT	2	QIFAMLKTE	З	TIVGFESY	115/		
FliI Hp 15646029	16	LSPRYGSVKKI	MPNIVYADG	1	NPSVGDVVKIEKS	1	GSECVGMVVVAEK	1	EQFGFTPFN	5	RAGDKVLF	85\HAS	fused t	o FOF1
1COWA Bota 1827809	24	DLEETGRVLSI	GDGIARVHG	1	RNVQAEEMVEFS-		-SGLKGMSLNLEP		DNVGVVFG	5	KEGDIVKR	901	ATPase	S
1SKY Thth 114531	24	QVSDVGTVIQV	GDGIARAHG	2	NVMSGEAVEFA		-NAVMGMALNLEE		NNVGIVILG	5	KEGDEVRR	901		
F1ATP Caro 11466328	45	QPVLLGEVEKV	KDGVAFVTR	2	NVRFSELVSFIPA	13	NLIVEGMVVGIEQ	13	DYISVIIFG	5	KVGDRVRP	1271		
1MAB Rno 6729934	24	DLEETGRVLSI	GDGIARVHG	2	NVQAEEMVEFS		-SGLKGMSLNLEP		DNVGVVFG	5	KEGDIVKR	901		
AtpA Bacs 114531	24	QVSDVGTVIQV	GDGIARAHG	2	NVMSGEAVEFA		-NAVMCMALNLEE		NNVGIVILG	5	KEGDEVRR	901		
AtpA Spol 114527	25	KVVNTGTVLQV	GDGIARIHG	2	EVMAGELVEFE		-EGTIGIALNLES		NNVGV V LMG	5	QEGSSVKA	911		
AtpA Ec 15804334	24	EAHNEGTIVSV	SDGVIRING	2	DCMQGEMISLP		-GNRYAIALNLER		DSVGAVVMG	5	AEGMKVKC	901		
AtpB Af 11498767	2	KMKEYKTITQV	AGPLVFVEK	2	PVAYGELVTITLP	1	GSTRRGQVLDTSK	1	DVVVVQVFE	5	DTSSTVRF	721		
VATP TVO 13540884	2	PKLTYKSVSEI	SGPLLFVEN	2	NAAYNEMVDIELD	1	GETROGOVLDTRK	1	GLAIVQIFG	б	TEGTRVKF	731		
Vma2p Sc 6319603	24	PRLNYNTVSGV	/NGPL VIL EK	2	FPRYNEIVNLTLP	1	GTVRQGQVLEIRG	1	DRAIVQVFE	6	VKKTTVEF	95/		
Consensus/80%		.shs	h.hp.		sphh.h		hublh				s			

Figure 3. Multiple alignment of the HAS-barrel domain. The coloring reflects 80% consensus. The coloring scheme, consensus abbreviations and secondary structure representations are as in Figure 2. Additionally, big residues (LIYERFQKMW) are shaded gray. Species abbreviations are as follows: Af, Archaeoglobus fulgidus; Ape, A.pernix; Aae, A.aeolicus; Atu, A.tumefaciens; Bacs, Bacillus species; Bota, Bos Taurus; Caro, Cafeteria roenbergensis; Cau, C.aurantiacus; Cth, C. thermocellum; Cwat, Crocosphaera watsonii; Dr, D.radiodurans; Ec, E.coli; Efae, Enterococcus faecalis; Fae, Ferroplasma acidarmanus; Fnu, F. nucleatum; Glvi, Gloeobacter violaceus; Hp, H.pylori; Mj, M.jannaschii; Mlo, Mesorhizobium loti; Mth, Methanothermobacter thermautotrophicus; Npun, N.punctiforme; Ph, Pyrococcus horikoshii; Pyae, Pyrobaculum aerophilum; Rno, Rattus norvegicus; Sac, S.acidocaldarius; Sc, Saccharomyces cerevisiae; Smel, S.meliloti; Spol, Spinacia oleracea; Sso, S.solfataricus; Syne, Synechocystis sp.; Tery, Trichodesmium erythraeum; Thel, T.elongatus; Thth, Thermus thermophilus; Tma, T.maritima and Tvo, Thermoplasma volcanium.

Vibrio, Pseudomonas, Neisseria, Nitrosomonas and Ralstonia, as well as the actinomycete *Propionibacterium*. The ATPase domain does not contain any inserts and seems to correspond to the minimal conserved core of the FtsK-HerA superfamily. The mechanism of these ATPases has not been studied in detail but, by analogy to FtsK and TrwB, it seems likely that they associate with the bacterial membrane and act as ATPdependent DNA pumps, which load the phage DNA into the capsids. The ZOT of Vibrio cholerae is the packaging ATPase of the integrated phage CTX Φ (47,68). ZOT has been shown to associate with the outer membrane through its single transmembrane region. The extracellular portion is cleaved off and binds to intestinal cells triggering a signaling cascade that leads to the disassembly of tight junctions (69). The potential role of the ATPase domain in the localization of the pro-toxin remains to be experimentally investigated.

Packaging ATPases of eukaryotic double-stranded DNA viruses, typified by the vaccinia virus A32R gene product,

comprise a distinct clade of the FtsK-HerA superfamily; the similarity between the ATPase domains of these proteins and the ssDNA bacteriophage packaging enzymes has been noticed previously (70). The A32R-like ATPases comprise one of the several orthologous protein sets that unify poxviruses, asfarviruses, iridoviruses and phycodnaviruses into a monophyletic lineage of large NCLDV (71). Subsequently, an orthologous ATPase was also detected in the large mimivirus, an ameba virus (72), which also probably belongs to the NCLDV (Figure 2). Furthermore, homologous proteins are also encoded by the Tlr transposons of the ciliate Tetrahymena (73) and a distinct group of nematode transposons, which were discovered as part of this work (Table 1). In the present work, we also found that these proteins are also related to the packaging ATPases of the Bacillus thuringiensis phage Bam35c, enterobacteriophage PRD1 and the Alteromonas phage PM2. Notably, we found that a predicted ATPase of this family is also encoded in the recently sequenced genome of the



Figure 4. Major lineages of the FtsK–HerA superfamily. The horizontal lines show temporal epochs corresponding to two major transitions in evolution, namely, the LUCA and the divergence between the archaeo-eukaryotic lineage and the bacterial lineage. Solid lines indicate the maximum depth in time to which a particular lineage can be traced. The broken lines indicate uncertainty with respect to the exact point of origin of a lineage. Bacterial lineages are colored in red, archaeal in blue and viral in green. Black lines indicate lineages with representatives from more than one of the three major superkingdoms, bacteria, archaea or eukaryotes. In such mixed lineages the phyletic distribution is shown in brackets with A denoting archaea; B, bacteria; FF, filamentous fungi; Nem, Nematodes; Pl, plants; Teth, *T.thermophila* and > lateral transfer.

turreted icosahedral archaeal virus (74). This observation, taken together with the proposed common origin of the capsid proteins of several distinct DNA viruses (74), favor an early recruitment of these ATPases in viral DNA packaging. However, more recent dissemination of this family via HGT between different viral groups cannot be entirely ruled out either. The finding that viral packaging ATPases comprise a family of the FtsK-HerA superfamily suggests that they catalyze dsDNA pumping into viral capsids similarly to the function of FtsK, TrwB and other members of the superfamily in bacterial and plasmid DNA pumping. The function of the homologous ATPases in eukaryotic transposons is less obvious. They could have been recruited for an alternative function in DNA transposition but, given that these transposons have other uncharacterized ORFs, it cannot be ruled out that they are packaged into virus-like particles that are released from the cells.

Implications of the phyletic patterns and higher order relationships of the FtsK–HerA superfamily. Cladistic-type analysis provides for a reconstruction of the likely evolutionary history of the FtsK–HerA superfamily, even though the level of sequence conservation is insufficient for traditional phylogenetic analysis (Figure 4). The HerA clade is unified into a higher order lineage with the VirB4 and VirD4 clades on the basis of the presence of a shared, predominantly α -helical insert after the second conserved β/α unit of the ATPase domain. These three families, in turn, join the FtsK clade on the basis of several shared sequence features, such as the aspartate at the end of second core strand. The two viral clades, which lack these features, appear to lie outside of this assemblage of predominantly cellular proteins (Figure 4) with the packaging proteins of the double-stranded DNA viruses being closer to the cellular proteins as they share with the latter a conserved histidine at the N-terminus of the Walker A strand; however, it cannot be ruled out that this deep branching of the viral ATPases is an artifact of their extreme divergence.

The HerA clade, which includes a core, pan-archaeal orthologous set, appears to have originated in the common ancestor of the archaea, whereas the FtsK clade similarly can be inferred to have evolved in the ancestral bacterium. The clear-cut archaeo-bacterial complementarity in the distribution of the HerA and FtsK orthologs implies that LUCA encoded



Figure 5. Domain architectures, conserved gene neighborhoods and contextual network graph for the FtsK–HerA superfamily. (A) Domain architectures of proteins containing a FtsK–HerA like ATPase. SSO0283-N, CT1915-N and VirB4-N are conserved N-terminal regions found in the SSO0283, CT1915 and VirB4 families respectively. Transmembrane regions are labeled TM. (B) Genes that have a conserved neighborhood are shown as boxed arrows. A representative gene, the species in which it is present and its gi number are shown below the boxes. The phyletic distribution of a particular gene context is shown in brackets. Species abbreviations are as in Figure 2. The dotted lines bounded by brackets indicate that the genes bounding the bracket are in the general neighborhood and do not show a close operonic association. Genes that are poorly characterized are repesented as white boxed arrows. (C) Contextual network graph for the FtsK–HerA family. Each vertex represents a domain and the edges represent a contextual association. Domain combinations are shown as black arrows, with the arrow pointing from the N-terminus to the C-terminus of the multi-domain protein. Circular arrows indicate multiple copies of the same domain. Operonic and neighborhood are shown as red arrows with an O at the tail and the direction of the arrows point from the 5'–3' direction of the coding sequence. Lines with O at both ends indicate that the genes bounding the line are in not operonic but in close vicinity of each other. The blue arrows with the boxed tails represent experimentally observed functional associations. The green arrow with the feathered tail indicates an insertion of a Zn-ribbon with the arrow head pointing to the location of the insertion in NurA. Additional species abbreviations not in Figure 2. Aful, *A.fulgidus*; Ana, *Anabaena* sp.; Cab, *Clostridium acetobutylicum*; Cau, *C.aurantiacus*; Cwat, *C.watsonii*; Glvi, *G.violaceus*; Mth, *M.thermautotrophicus*; Pfu, *P.furiosus*; Suae, *S.acidocaldarius*; Tac, *Thermoplasma acidophil*

the common ancestor of these families, from which the HerA and FtsK clades diverged concomitantly with the split between the archaeal and bacterial lineages. This archaeo-bacterial dichotomy is similar to that in some families of proteins involved in DNA replication, such as PCNA/DNA polymerase III β subunit and ATP/NAD-dependent DNA ligase (75,76). In each of these cases, the fundamental separation among the conserved members of these families, which share only limited sequence similarity, corresponds to the split between the bacterial and archaeo-eukaryotic lineages. Moreover, those bacteria that lack FtsK always encode a HerA protein that belongs to the conserved core of this family, which is predominantly found in archaea. These bacterial genomes, without exception, also encode the nuclease partner of HerA, NurA (29) (Table 1 and Supplementary Material, Table S1). This complementarity in the phyletic distributions of the core orthologous set of HerA proteins and the FtsK family, even within the bacterial kingdom, along with the co-occurrence with NurA, suggests that HerA and FtsK are responsible for the same function, namely DNA pumping during cell division. It should be emphasized that, although there is no experimental data on the biological functions of various HerA paralogs, the strict conservation of the 'main' *herA* gene in archaea, in terms of both the ubiquitous presence and the sequence itself, implies that it is this gene that has an essential function in cell division rather than any of the extra *herA* paralogs present in some of

the archaea. The archaeal HerA–NurA system appears to have displaced FtsK in most bacterial extremophiles and cyanobacteria, which might have been facilitated by their ecological proximity with archaea. These observations imply that LUCA already had a DNA pumping system similar to those in the extant prokaryotes (Table 1 and Supplementary Material, Table S1).

Although not demonstrated experimentally, it seems likely that the pumping process could introduce double-strand breaks in the DNA. If this were the case, the bidirectional DNA helicase activity that has been detected *in vitro* in the purified *Sulfolobus* HerA protein (18) might be involved, together with MRE11 and Rad50, in double-strand break repair (see below). Alternatively, it cannot be ruled out that the helicase activity is unmasked in the *in vitro* assay as a result of the absence of other subunits which are associated with HerA *in vivo*.

The observed phyletic distributions suggest that VirB4 and VirD4 families might have been derived from DNA pumps of ancestral plasmids that were not part of the core cellular genomes. The relationship of the FtsK-HerA proteins with the viral packaging proteins suggests that DNA pumping activity in diverse systems, both cellular and viral, has an ancient common origin. The emergence of the FtsK-HerA ATPase might have marked the origin of structures in which copies of DNA were compartmentalized after replication. These primordial compartments, into which DNA was packaged by the ancient members of the FtsK-HerA superfamily, could have been the evolutionary precursors of cells and viral capsids. The absence of this superfamily in eukaryotes, with the exception of the apparent late HGT into filamentous ascomycetes, is consistent with the dramatic difference between the mechanisms of chromosome segregation in eukaryotes and prokarvotes. The emergence of eukaryotic cytoskeletal components facilitated segregation through the mitotic process, which involved chromosome translocation by ATPase motors, such as dynein and kinesin (77,78). This radically different segregation mechanism appears to have rendered the ancestral HerA-like DNA pump superfluous or even deleterious, thereby favoring its elimination through gene loss at an early stage of eukaryotic evolution.

Contextual information from gene fusions, domains architecture and conserved operons: functional implications for the FtsK–HerA superfamily

Conserved operons, gene fusions and domain architectures are useful in extracting functional information for otherwise uncharacterized proteins based on the principle of 'genomic context' or 'guilt by association'. Products of genes cooccurring in the same operon in multiple, sufficiently distant genomes (conserved gene neighborhoods) or undergoing gene fusions tend to interact physically and functionally (30–33). Accordingly, we systematically surveyed the genomic context information for the proteins of the FtsK–HerA superfamily. In Figure 5, this information is represented as domain architectures (Figure 5A), gene organizations (Figure 5B) and a graph where the nodes are the connected proteins and the edges denote different types of contextual connections (Figure 5C).

The archaeal herA operons and the widespread herA-nurA gene pairs. The largest conserved operons including genes for

FtsK-HerA superfamily ATPases are those that contain the highly conserved archaeal HerA proteins of the core orthologous set. These operons typically encode four proteins, namely, HerA, orthologs of Mre11 and Rad50, and the NurA nuclease; the same gene order (HerA-Mre11-Rad50-NurA) is conserved in six genera of euryarchaea and crenarchaea (Figure 5B). Variants of this order are seen in Aeropyrum (Mre11-Rad50-NurA-HerA) and Thermoplasma (NurA-HerA-Mre11-RAD50). In Methanothermobacter, the herA gene in this operon is apparently split into two genes and the gene encoding the C-terminal half is fused to the gene for Mre11. In Methanosarcina, Halobacterium and Methanococcus, the operon is split into separate HerA-NurA and Mre11-Rad50 (predicted) operons. In the genome of Nanoarchaeum, the operon is completely disrupted, although all four genes are present. A parsimonious evolutionary scenario places the complete operon consisting of the four genes in the typical order into the genome of the common ancestor of archaea, with partial disruption of the operon during subsequent evolution of individual lineages. There is no trace of this conserved operon in any of the currently available bacterial genomes, with the sole exception of Bradyrhizobium, which shows a linkage of a HerA protein of the bll1925 family with SbcD, the bacterial ortholog of Mre11 (Figure 5B).

In addition to this core orthologous lineage, archaea have several additional paralogs of HerA, most of which are encoded next to a conserved ORF of approximately the same size as NurA. Comparison of these sequences to the NurA PSSM showed that the HerA-associated ORFs were divergent paralogs of NurA. Further iterations of these searches, with PSSMs that included the newly detected NurA homologs resulted in the detection of numerous additional divergent members of the NurA family from archaea and bacteria. Remarkably, these searches showed that the phyletic distribution of the NurA family is nearly identical to that of the HerA family (Supplementary Material; Table S1). In most of the archaeal genomes, the newly detected NurA proteins are encoded next to herA genes. Among bacteria, nurA genes of Bacillus halodurans, Clostridium thermocellum, Deinococcus radiodurans and Aquifex aeolicus are adjacent to genes for HerA homologs whereas, in the rest of bacteria, the nurA and herA genes are located in different parts of the chromosome. In Chloroflexus, the NurA homolog is encoded next to a gene encoding a stand-alone HAS-barrel (Figure 5B). These observations suggest that the nuclease NurA and the ATPase HerA are not only functionally linked, but also tend to be horizontally transferred among archaea and bacteria as a gene pair. Furthermore, the situation in Chloroflexus supports the prediction that interaction between NurA and HerA is mediated by the HAS-barrel. The tight functional association between HerA and NurA mirrors the functional connections between FtsK and the ParCD topoisomerase or the Xer recombinases, suggesting that NurA has a function similar to the functions of these bacterial enzymes in DNA processing during chromosome segregation (14,16).

Detection of the new, diverged NurA homologs provided for a better characterization of the conserved structural elements and the active site of the NurA family nucleases (Figure 6). Secondary structure prediction combined with examination of the alignment suggests that NurA has an $\alpha + \beta$ -fold with a central, conserved β -sheet formed by at least eight strands. NurA has at least five conserved α -helices, with the last three forming a characteristic triple helical unit. These patterns do not bear any obvious resemblance to previously characterized folds found in nucleases or other proteins. The predicted active site is comprised of six charged/ polar residues, which include two characteristic aspartates at the ends of core strands 1 and 5, a conserved glutamate in the first core helix, a basic residue and acidic residue after strand 8, and a polar residue (usually histidine or aspartate) in the C-terminal helical unit (Figure 6). These residues might coordinate a metal cation as observed for the restriction endonuclease fold enzymes. The NurA family appears to be a rapidly diverging group, with a low level of sequence similarity between paralogs and even within orthologous groups. There are several inserts of variable size in different members including a small Zn-cluster in the cyanobacterial and Chloroflexus NurA homologs, and a Zn-ribbon in the NurAs associated with the CT1915-like proteins of the HerA clade. The extreme sequence divergence in the NurA family is reminiscent of restriction endonucleases, suggesting the possibility that, similar to restriction enzymes, different NurA family members recognize specific target sequences in DNA. The analogy could extend even further in that the NurA-HerA pairs might form mobile elements similar to restriction-modification system operons. Consistently with this hypothesis, a distinct subfamily of nurA genes (typified by HH1040 of H.hepaticus that associate with the distinctive CT1915 family HerA) comprises a predicted operon with a gene for a DNA methylase, which is related to methylases encoded by several restrictionmodification operons (Figure 5B) (79). The organization of this predicted operon closely resembles that of several restrictionmodifiation systems with the NurA gene taking the place of the endonuclease, and the HerA gene that of the accessory helicase or ATPase subunit (79). Thus, this family of NurA family is predicted to function as a bona fide restriction endonuclease. In addition, in cyanobacteria, the nurA genes co-occur with a gene for a distinct predicted hydrolase of the HAD (haloacid dehalogenase) superfamily (80); this particular orthologous set of predicted HAD hydrolases was detected only in cyanobacteria and plants (Figure 5B, Supplementary Material). Many proteins of the HAD superfamily have phosphatase activity, and some of them, such as the DNA 3' phosphatase, Tpp1p, cooperate with endonucleases in strand break repair (81). Hence we speculate that, in cyanobacteria, this particular HAD protein cooperates with NurA in DNA repair, probably as a polynucleotide phosphatase.

8		Str-1 Str2		53		54		Hl		S5	
Secondary Structure	17225	EEEEE EEEE		EEEEE		EEEEEE	530	нинининининини	05	EEEEEE	-
PF1168_Pfu_18977540	43	EKSSIYAVDGSRSVSRLSG-TVIYFL		SALAV		GSGKQLRLSYANAIKSNYGTSD	1	IVRMOMETLENMLGYLAYRK	4	KRAILMDGTLTGSLVRP	18
PAB0813_Pab_14521426	35	KKSKAYAIDGSRVARRLSG-TIVYFL		TACAV		GSGRGYSLSYANAMQYNYAVSE	1	MIRMOMETLENMIGYLSWRA	4	PKLILMDGTLTGSLIRP	18
PH0928_Pho_33359344	43	ERSS VY AVDGSRSVTRLSG-TIVYFL		SAIAI		GSGRSYKLFYANAMQYNHGVSD	1	VVRMOMETMENMIGYLAEKM	4	KKLILMDGTLTGSLIRP	18
MJECL07_Mj_10954498	46	VEGVLCGVDGSRGKVEFCS-GIVYGL		SSYAI		GKNIEKGMFELGVLPFFKEEDR		-VRRLMMTLEYRLATLVSKN		VDLILLDGTLSGALIMP	13
AF1033_Afu1_11498638	41	VRCRICGVDGSRGIERLSG-LVFYIV		SAASV		GEDIREMHEVTTLKPHMHIEER		-IRLHMHTSEYRLGSVAD		EEIVLMDGTLRGAIIRP	14
MM1869_Mma_21227971	70	PYPVTYACDSGSTNPKTYDSGLFVDF	2	CGLAA	37	GLGRAKLVTIAPDELKRKAPDM		VHSFAMYLAESEHLLFMKDR	3	ESFFIMDGPI YPKQLMY	12
Mbur020200_Mebu_41720377	70	PFDITYSCDSGSTNPMAFRSGLYVDI	2	GGIAS	37	GQGRSKIVRIGADLLKTRVDRM		VHNVALYLCESEHILWMMES	3	KG FFIM DGPIYPKQLMY	12
MA0708_Mac_20089593	70	PYPVTYACDSGSTNPRTYDSGLFVDF	2	CGLAA	37	GLGRAKLVTIAPDELKRKAPDM		VHSFAMYLAESEHMLFMKDM	3	ESFFIMDGPLYPKQLMY	12
Meth1917_Mba_23050405	70	PYPFTYACDSGSTNPRTYDSGLFVDF	2	CGLAS	37	GFGRAKLVTIAPDELKRKAPDI		VHSFAMYLAESEHILFMKDR	3	ES FFIM DGPIYPKQLMY	12
Vng1891h_Halsp_15790782	72	AFDTAHGLDSGTINPRTFTNGLVLDL	2	AAMSA	35	GYWRGRIVHTTPLARDQERVV-		-HGLALYLAESHHAREHAHR	1	GD LLLLDGPV YPKQLVN	13
t111039_The1_22298583	94	MEVTTVAVDGSQMVPSEELFLPVGVV	1	AGWFL	3	RRDRPYAKEIALELITPAELRQ	16	YIHLRRFELELKTLRERMAE	3	PALLLFDGSFVATFAES	1
Tery0911_Tery_23040220	79	KSHT VF ATDGSQIAPSQHEIAYCYLL	1	IGRVM	3	GESRYPLLDSIPEIFYEPGDLY	11	WMGYRRTISEALVLAQLGCE	11	PILAMVDGSLIYWFLES	1
alr3864_Ana_17231356	79	KVHS VI ATDGSQIAPSHHEIAYCYLL	1	IGRVV	3	GQNRYPLLDSLPEVFYRPEDLY	11	WMSFRRTASEAIVLAELACS	10	PT LAMV DGSLIYWFLEQ	1
Npun045_Npun_23124071	79	KIHT VI ATDGSQIAPNHHEIAYCYLL	1	IGRVV	3	GQNRHPLLDSLPEVFYRPEDLY	11	WMSFRRTASETTVLAELACA	5	PALAMVDGSLIYWFLEQ	1
sll0294_Syn_16331246	79	VNHS VF ATDGSQIAPSHHEIAYCYLL	1	IGRVM	3	GQNLHPLLDSVPEVYYRPEDLY	11	WLGHRRTVLEAERLAALACR	10	PNLAMYDGSLIYWFLEN	1
glr4039_Glvi_37523608	79	PVHT VI ATDGSQIAPSHHEIAYCSLI	1	VGKVC	3	GTGERPLLDSQPEVFYREEDLY	9	SLALRRSRAEMEVLADVALD	6	PAVALVDGSLIYWQLEQ	1
t111668_The1_22299211	79	RAHTIFATDGSQIAPSHHEIAYCYLI	1	VGRVA	3	GQSQRPHLDSVPQLVYDPDELE	11	WLRYRRTQAEMLALVELLHT	5	PSLALVDGSLIFWAWES	1
Chlo0591_Cau_22970555	80	TDYAVVATDGSHIDVDRHGEVSCYLI	1	IGQVY	3	GAHPEAYLESSPRLYFSEDDLY	13	VLSVRRDVEEGVALGRFAAM	7	PRLALQDGTLIRWALAN	1
CT1914_Ctep_21674726	58	P-DLVLAIDGSNLAAKAENGFPGA	1	FGYIT	1	ASVLIDLKLIGELEKKEFVEPK	28	KSSLRRALFEELRSNTIFSD	103	RVAFIMDGPLAVFSTSS	6
Chlo0996_Cau_22971012	58	PVELVLAIDGSNYEAALDDEIPST	1	LGYLK	1	GAILFSLKDVSKLREGRFVDPF	29	RDSFRAALDEQLLSEKTRFD	112	KI AFFI DGPLAIFGASA	6
HH1040_Hehe_32266539	59	CIERVITIDGGYQEVNINDNFPSQ	1	LCYYN	1	GILMFSVKDLEVVERQQTINPS	27	ISTFRKTLYEEVF-LKNYLS	119	KI LFI KDGPLALFSRLD	7
MM0197_Mma_21226299	63	PISLFITIDGGYTEVFIRREFPSC	1	LNFFQ	1	GALIFELKDLKTLSNKPFIDPE	27	TNSVRNTFYEFFTENKENFI	107	ETLFIKEGPLAFFSVTA	б
MJ1262_Mj_15669448	66	KDMGFAGGDGSCNKLDYI-SFSFYGV	1	AVSFI	1	GRGEKVKKAKEEYIFDITHPLD	3	RIRRYMLTLELKTALYVLKN	2	ID YYIF DGSLFSLLIFT	46
NurA_Sac_21388538	49	HSKKLLAIDGGMWVKETR-QGVIFIV		NAKAI		VFEGINEINSEGKVLVHIFSPG	5	RIELLMQLLELQLALKLVEN		VDYVLLDGSFSKKLGRH :	26
ST2109_Stok_15922435	49	KLST VL AIDGGMWIKELR-SGIVYIV		NAEIV		KAEGFNVTPIDSKALIGVLRPG	5	RVSLLMQLLELKLGLKHGDK		AEYILFDGSIVKKIGKH (27
SS02248_Sso_15899021	51	-TCKFVAIDGGSFGRPMR-IGIVYAV		GAESV		IGDNKGVKTLSEDGQIGIFKPG	5	RISLLMEALELSLALRDGSK		GDYILMDGSLSKKIGNK	28
MTH306_Mth_15678334	45	DSLSTAAADGGSGFKEFSGLILAA		INTAV	1	SDSGEIVEGSFLDILGPQSGI-	2	RVRNLTAICEFKNSIIAMED	2	PDLLILDGSLIGTILRP	38
SSO2284_Sso_15899051	63	SQSLIYSLDGSSRSFISS-KGIVSLA		SVVVS	1	TISPILGVYPPISGFPELDLKK	42	IETEIRTILETEALKKIPN-		DGSIILDGPLIPPLI	
ST2248 Stok 15922580	57	ENEVIYAVDGSSRSLISA-GGIISIN		TLAIS	1	STYPIIGVYPSLFGLPSLPIKK	44	IETELRSILETEGLKITKD-		KGLTVIDGPLFPSHI	
ST1039_Stok_15921290	63	ENVSIASIDSSSRYLRDP-SVNMVFV	1	LGVYS	1	IKGIKVGPFDIDINFMAIGTFE	30	IADELRLEAENVGLKENRNT		HDLVIVDGPLFPTPLEL	6
PAE0246_Pyae_18311806	52	GPPVDYAVDGSLRTITTPTFRLFL	1	VGVAY	1	KLGILTIPGFRDVKHVALQSDD	32	VGKVVREAVEIALIKEAK		-GTALIDGPLYYGIR	
SSO1325 Sso 15898167	63	NDKRIASIDSSSRYLRDF-SVNTCLI	1	LSVYS	1	REGFIDGPFTVDIPYIGISSYR	35	IADELRLEAENVGLKRVIQN		HDIVIIDGPI YPTPLEL	7
APE0082_Ape_14600434	62	SLGDVYAIDSSSRTVETPLHLILI	1	AVSLS	12	FEKPLVSLNIRPYLVLDRLNLA	19	ENAMLSYLLERLIEDRYPED	1	ASAVIIDGPILNGSILA	5
NEQ336 Naeq 41615125	37	KEGALPAIDSSIRVIPFRGNRLAL	1	VGAVI		VNKDNDYYNSNTMLTDVEDEN		VKTLAMKYTEWQLINEY		DGT III DGSPNIDKRLL	
PAE2154_Pyae_18313138	53	RSLD VY AVDSSYGSPPLE-LIGGVFT	1	VAYGY	1	GISKGVQDRFLTGALYFSDGRE	10	EKRLAARLLEAK-IRGEKQ-		LD LLVL DGE I AIHPLPF	4
APE0109_Ape_14600456	60	GMDNIVAVDGGNRSRDYR-EFTIYLA		RAWAG		GGGREALLHSLGVVVPPSDAEA		RISIYREILEAVVAKEAVPG	2	RG LLLL DGSPSTALKWW	24
PAE2902 Pyae 18313674	41	PSFNVAGVDGGIGVVKLA-NGHQVVL	2	AAAVG		SDFIEREFIADIASVDSASLP-		WAYLVIVESLVGMRALEK	2	VDVLLMDGSLYAKAVRL	17
ORF22 Unk 42557712	79	EPSIVCGTDSSCIKIAETGTEDLYAV	1	CGVVF		CLSMQVVLHLRIGPLLFYITES	21	AKRMIRINVERLIQFELSKL	2	GSVILVDGALKASI	
PAE2249 Pyae 18313211	38	DKPDYYAVDSGYVVQKIG-SFDVLIQ		SIVAV		GREVRRRFLIKKVVE		DVHTEARRREIRFAESID		AD IVLV DGPLTPYVGVT	4
MK0693 Mkan 20094130	73	DPEVTFGVDGGEGMREYQGVVLYV	1	RAAAW		SEADVLSSWDFGVLSRTRTPQM		RVAARRVKLESDVATRAVER	1	GE VVML DGPIVPHHDLK	
MTH543 Mth 15678571	51	EKRLMAFIDGGNSPIVEG-PAISVQI	1	RVALG	22	KFHPEDGTYRFQLHPLREEYRE	6	NLQLELKDAERAEESDLKGL	19	GD ILV RDG SL RASFERE	5
Faci0362 Feac 22405469	58	GKIDAAATDSSEFLRMLY-NGKNIVI	1	RAFTL	1	NNEVTSDFHADLVAVDPVDQRN		FTILLMEHSEHKSMLKFLDE	2	PEYLFVDGSLKGRLSHR	9
Ta0160 Tac 16081320	57	KSVSISATDSSEFSRELY-SGPFFIL	1	RSYSK	1	QGRVYVDFKSSIASVTPEEVKR		ETIIMMESSEHRSLMKLLAS	2	PDIALVDGSLMGRIGRC	6
TVN0230 Tvo 13541061	57	ENLKISATDSSEFSRELY-SGNFFIL	1	RGYTR	1	GDSIYNKLEVSFLSIGPEDVKK		ETMIMMENVEHLSMLECLKK	2	PDFMLVDGS1LSRILRC	6
SSO2197 Sso 15898973	53	SEHIACAVDGSKYEIELSDVTLI	1	ARAVK	1	LGKKKDKKSVPSEIAEDFKIIE	10	KSILFMLTLETSLIEKCEE-		CD VIFI DGPIVDPPTYY	8
ST0765 Stok 15920999	50	GEHIACAIDGGKLEVDLGDSYLII	1	KAVSV	1	GKYGETKEIPPTIVRDFKIVSD	9	RSIILMLTLETNLLSQAN		CDK IFI DGPLIDPPVYD	2
PAE0122 Pyae 18311728	48	PPGDVHGLDSHTAVVEFEGVSVIV	1	TGALV	2	CSAFVPGLSASWLGVRLNFKGG	29	RDEVRYHVEEALARAWDG		GGVLLLDGPVFRAL	
BH0051 Bha 15612614	60	DNKQIGAVDGSVNQTKGSHVPPLYLF	1	ALAKT		LQGVETKKTDVYSPLLDSPEDE	5	WRAHILSKLELKAALDLMDQ	2	LAILMMDGALYHYRIDA	
MM2560 Mma 21228662	1	MRASYDSFIGEISNLNSLL	1	RVEEL		GSSPQYLPIKAISSKSAIIMSS	15	EQINEMNISEEYIDDKMWKT	6	KALEYIDGKKLEAD	
ST0180_Stok_15920359	21	WNFTFTAIDGSFHEVQYISGNLAYVV		VGKVE		GEVKQNKILTLKIDYDEKICEN	1	KAEDEMREMEYEKARNAS		SD IIFL DRK I SMDDFNE	7
SS00467 Sso 15897396	28	RELTFAAIDGTFSDIILEGEKGGYIV		IGIIK		GRIFKDFKFTIEDISVDDELCI	1	EAEKRMRELEYYNIKENE		VDLIFFDRKLSFDKSLG	15
ag 977 Aae 15606289	54	ENLKIAFIDGVRRTENIVYLED	7	EGAFV		SIGAGAMVLSYGK-VNSLEDSLKK	42	YVNELMANLELWVAQQVYKA	2	ADLMI TOGTLHYVAKAK	1
TM1739 Tma 15644485	41	ILDQ II FVDGKRRSFVRITTDE	2	TGIFA		ELCVGAVIWDREGGTKTLFSPDKP	37	SIDLYMRSLEIEEVRKHMDK	1	NILIVKDGPAARELPFE	2
DR0836_Dr_15805862	45	SVREVLVVDGKPRMEARLLMDD	б	LAAFG		AYVVGAVSLCPHG-TRQAELLDVR	45	AVQKLMLQDEQKLSRQLASP	12	ESLVLQDGPVRLGGGG-	
Consensus/80%		hhs.Dush		s.h.				bbEb		hhhhDGsl	

	H2	56		S7		S8		H3	H4	H5	
Secondary Structure	ннннннннннннннн	EEEEE		EEEEEEEE		EEEE		НННННН	НННННН	EEHHHHH	
PF1168 Pfu 18977540	ESDFENLLNEFLEKLRDHYRKVE	EHLEKNGNYDSP	172	IDKGIHL AY VRF	2	GDVIY-MLQST		TNIEKIL	PLILHHKAGGYLRP	-QLAHHGVKIS	417
PAB0813 Pab 14521426	EHDFDNLIEEYKELLDEHYKEVE	EELEKKGKSDAP	172	AERGVYVAYVRL	2	GDVIY-MLQST		KKIRDIL	PLVLSHKSGG Y IEP	-RIAHNTVKIS	409
PH0928 Pho 33359344	INDFDNLIYEYVKMLEEHYREVR	RKLKDEGTSDAP	171	AERGVNIGYVRF	2	GDVIY-MVQST		KKIEDIL	PLILHHRVGGYIRP	-QLAHDGAKIS	416
MJECL07 Mj 10954498	PDLAEDLGWKFIKSLDNFWDEV-	-LENLDGNIYDN	141	PFEMIPKTYVRF	2	SSPIL-ALEVP	4	KSIEEVI	SLLIPYSKLGYPRY	-KDAHNKAKIS	380
AF1033 Aful 11498638	LYDLEGVIEDFIEVLEEWYKEIT	-EDVKAGMARKN	108	DKAEIYG AF VRF	2	SGSIY-MLEST	1	PIEEDTI	AKLCNYEADGYLIP	-IHAHRHAEIK	339
MM1869 Mma 21227971	ONPDARKILONYIDVMDHFLEM-	-KIPVIGFVKNP	110	YALCFFMLYVPS		KDVVF-KVEAP	7	FLRMOIT	1 KVLFDVSLHGFPLT	-TKADHLAKIR	418
Mbur020200 Mebu 41720377	EDPVAKNILONYIDIMDYOMKK-	-GMPLIGFVKNP	81	YAITFFMLYVPM		LDVLF-KVEAP	7	SVRDLIT	1 KVLFDISVNGIPLT	-SKVDSLAKIG	389
MA0708 Mac 20089593	RNNDARKILQNYIDIMDHFLEK-	-KRPVIGFVKNP	107	YALCFFMLYVPS		TDVVF-KVEAP	7	FLRMQIT	1 KVLFDLSLHGFPLT	-TKADHLAKIR	415
Meth1917 Mba 23050405	QNNDARKILQNYIDIMDHFLKN-	-RQPVIGFVKNP	105	YALCFFM LY VPY		KDVVF-KVEAP	7	SLRMOMT	1 KVLFDLSLYGFPLT	-TKADHLAKIR	413
Vng1891h Halsp 15790782	QNELVGEVLENYVRLVEDFADR-	-GVPLAGFVKSP	78	YEVAF'FV VY DPR		TDLVH-RVEVP	8	CRAAVER	1 VTSQVAAEAGPPKP	-GKADELARIS	384
tll1039 Thel 22298583	TPQWQGRYVAALCQTLEASRQQ-	-RIPLLAYIDGS	56	KGIGFCYLKAHQ		GYPVRVELP	б	GQLNAVI	2 LCGELITGQGYPYA	-EVADQVAVLQ	357
Tery0911 Tery 23040220	PSEARDRILNPIMEAWNKLKLA-	-GIPLVGYLSAS	76	GDNTIYF CY VNV		GREVA-RIDMP	7	EMLELVL	2 ILAQVQKGYGYPVA	-AEAHNQAVVR	367
alr3864 Ana 17231356	PVDARNCILPPILEAWEQMRAA-	-KIPLMGYVSAS	69	EDQAIYF CY VHV		GTEIA-RIEVP	7	EMFDOAL	2 MLAQVQKGYGYPVA	-AEAHNQAVVR	359
Npun045 Npun 23124071	PMEARDRILPPILEAWQQMRDA-	-QIPLMGYLSAS	69	GDQTIYF CY VHV		GTEIA-RIEVP	7	TMLEQAL	2 MLAQVQKGYGYPVA	-AEAHNQAVVK	354
s110294_Syn_16331246	PQEARNQILEPILAAWETLRQA-	-RIPLMSYISAP	73	PEQRVCF CY VQG		SSEVA-RVEFP	7	ELLDQSL	2 VLSQIEKGFGYPVA	-AEAHNQAVVR	363
g1r4039_G1vi_37523608	STKHQQAILEPMLAAWERLRER-	-RIPVAGYISSS	70	GPHRVHF CY LHV		GSEIA-RVEMP	7	ELRERVL	2 CLAQVEKGFGYPVA	-AEAHNRAVVQ	354
t111668_The1_22299211	PSMVREELLRPILAAWDQLRAK-	-RVPLVGYVSAS	77	GAHHIYV AY LHG		GSEVV-RLEVP	7	DLWQRAV	2 TAAQIQKGRGYPVA	-AEAHNLAVVR	362
Chlo0591_Cau_22970555	DEFIRDHFLGQYLNYLEQMRKI-	-GIPVASYISRT	77	GDHQIHF FF LRI		GRELA-RIEIP	7	DMVAQVH	2 VYDQAMRGQGYPVVI	-QRAHEQAVIR	367
CT1914_Ctep_21674726	SHELTRLNDLQRKINGQDLLII-	-GIEKSGTFFNH	38	KPYGQDT YF		GRKLFYKAASG	28	ADVMNLL	2 LVSSRYPNSVSPLVS	-AHAEAAIPLN	436
Chlo0996_Cau_22971012	MIYLHEVNTRLAKYHQPPLLII-	-GLQKKGQIVDY	33	GGFGFET YY		GQDFIFKTKQG	34	PTAVQLL	2 LQTELYRDAVIPIA	-AHRYTAISTQ	448
HH1040_Hehe_32266539	RPFIQFLYAKSLNENISYINLV-	-GLDKSGMFVEH	30	SVFGENT YF		GIKMFVRKEKE	32	KNILEVL	2 LKCDLYEKSFIPIA	-VNKLVSLSNI	449
MM0197_Mma_21226299	RDLSNYLLEKYDLNLV-	-GVEKSGSFVEH	33	NPYARTS YY G	•	GKLIFKSRDE	26	DSILQNL	2 LKCDMYDNSLFPVA	-ANKLVSLSAH	432
MJ1262_Mj_15669448	NKKILVEHVEYILTLTKLINEF-	-KDRIIGISKTS	45	LSSVVKGINFIK	2	PF-YA-KIDTA	20	KIDKEVL	SSLKEISING Y PYI	-KKSHETVEIT	360
NurA_Sac_21388538	DNMLRFLIAENQLVLSELVSRY-	-KDKLLFISKNS	38	LSRKASK LL SGL	2	YFTNL-RLEPS	11	DKIFEYL	KVLKPVSLKG Y PYP	-IKVHKDVRVG	306
ST2109_Stok_15922435	ELMHKYLVAENQLVMSALISKY-	-KGKLVWISKNS	49	ASFYSFYTRLKE		GEKIL-KIEMF	1	NEIENII	SILSPISIKGYPYPI	-LKVHTDVKVS	305
SS02248_Sso_15899021	RKMRDLLMLLNQFLVSKIIEEY-	-DGNVLWISKVS	49	MEYTTFYTRLDN		GKRVI-RVDIV	5	KIVKEIM	DRLSGVSIKGYPFP	-LKAHMDVRFS	312
MTH306_Mth_15678334	DTIRDALRSERPVVYLEGIERM-	-VSIWKLLKKSR	48	PVLDAFFSSLEF	3	TVFYA-KLEEN	14	AEIGEYL	EGLRASSADGYPHI	-RRVKEDVRIT	323
SSO2284_Sso_15899051	FLRSKPRDDILSLRLQAIKG-	RNVIGIVKRL	60	NTPVYVN YL IYP	6	KFAIL-RVESL		NNDSGVV	2 ISSLKFTKDGIPFI	-AIADRTAKEI	343
ST2248_Stok_15922580	YLPEKVREYIVKERIKVLDE-	KYVGIVKRL	59	KIRYYVN <mark>YL</mark> VIP	6	KFSIL-RIESL	i Sat	NKNAPAI	VASLPFSNDGIPKI	-SIADKTAKEL	335
ST1039_Stok_15921290	SEGRKRHQEAYLQLTKDRISIL-	-RNNVVGVVKRL	54	DAPTRYAYYLIL	5	MSSYF-RIESL	2	DSLEELT	1 YIVNRISERLIPTY	-EIADNLSKRV	338
PAE0246_Pyae_18311806	GFPEIDWWRVEALWDKR	VVAVVKRL	60	PLPERYLWYIYS		ASSIF-RVEAF	2	EYAEEVL	1 TLLRNRVVPGLPYH	-ALVDKLAKDL	310
SS01325_Ss0_15898167	TEARRKHRLAYAKLVNERIELL-	-NNNVIGVVKRL	67	KVPTKYAYYVII	5	PPTFL-RIEST		NKNLDIS	PVLSRLTERLMPTY	-EIVDRRSKRI	354
APE0082_Ape_14600434	IGDVWSELYRERQRIIASIEAL-	-GVPVIGYVKRI	60	GDEKLFEYLILP	9	KRRVY-RIEYT	10	GFRPYHI	1 LSETLLRGSLEPVT	-ARSDRRARSI	354
NEQ336_Naeq_41615125	EKYDLPIVNEYRDLIKAN-	-LEKTIFFARRI	38	RDIIFFNKKFKT	1	LAYFG-RNPSL	4	PNNMDLL	3 ANYTNYYGVYEPIV	-ADTIARQYSK	249
PAE2154_Pyae_18313138	KGGKYEEVNKVVNRLLRWAIESR	TSVVALAKRVRS	101	LGDVEVVYYVPP	1	HRTAV-RVEVE	3	NIGVDKI	3 LASTASSITGYPQI	-DAVDQYVRVS	352
APE0109_Ape_14600456	IRGQTLIPDCVEGGCVETLLEK-	ATERPASNEAAM	104	DVAGLRKFYENL	2	VSTYA-RLAPG	1/	GDHLGLL	4 SRLANFSPSGYPVP	-LIAHEKARVS	389
PAE2902 Pyae 18313674	ELATALYTLSRLIKTARERGAR-	LVFVSKDHNYKL	136	TEPKIYEEYIKE	2	DESPL-LVEIP	18	ARVDDVV	Z LREQERDPVHYNTW	-WIAHSVASEK	393
DE22_UNK_4255//12	FENRHINVKKILENCALHK	NSLLGIGKNIKF	30	NIFGINIMVKFA	2	SLVLRADVA	1	KUNEEVI	3 IRGNDLLPMGIPEC	-SMAHHISCES	311
PAE2249_Pyae_18313211	VSKDPRLVRIGPRIEDDATRDL-	FVRLSKRWGERE	10	LPVDIGGLRGTF	2	SEWVL-IVEFP		DEARCTY	1 CREIEVCOVOIDI DI	-RVAHHLAKIG	223
MRU695 MKan 20094150	GANEDSPNRRDIWGRIEDAR-	TRANCISSION	22	DDITTI	2	ISPPV-RVEIP	0	MEUNORI	1 UNAPPACEDOVPVC	-VMADELARVG	329
Faci0362 Forc 22405469	ENEMVEVMENT VET TEVEVEVC-	TISMSLLAAV	70	TOTNYUCEDTID	2	VDI DM_FUOVI	14	NUDERTI	1 LNARDACFPGIFIG	-VDVDFRARVS	221
Tac10302_reac_22403403	TOADEDVCMCLDELLEAAAVV	TILIFMANSS	5.4	MVEDIMUCDITE	2	FDUDT-PERF	14	VDEETUI	2 MMYCYTCYVU Y NTM	-COTOMORYCER	201
TVN0230 TWO 13541061	I DRENDVCI MI DELLEORKKON-	FATAFUSKSS	64	-I DESEDITOTEV	2	TCNPD-PFFDL	.4	VEELIVL	MEWSYCCYKUWIT	-SEVINIVER	302
SS02197 Sec 15898973	TISLOKIALVESTULERIKGKN-		78	DSLSTVSLYOV	2	TSPTV-PTDVI.	6	FTPTKYI.	NTWAMOSTREVTLL	-SEVERISETN	330
ST0765 Stok 15920999	IEEVELVPSNVTKP-	KEPIGIVKNE	72	TGLKTYSSYFOL	2	TSPVV-RIDTL	1	PSGLOFT	KIWGIEGEKE V TIJ	-KIADSLAKVR	301
PAE0122 Pyae 18311728	DVLKRGQVVAFLYSETYKRR-	YELEGGRRVV	44	CDLAKYMYYVCV	4	CTRVL-RAFAF	2	SLARFAD	1 WLCSLADSSCVPVP	-SAADRLARRI.	297
BH0051 Bba 15612614	PEEWEVLENBALER-	-GTLLVGVSEET	30	HERESLYTESTO	5	RSVWM-RASSS	16	NOVADEL	ATMTPEDGRGTPLE	-DTVDREVRLS	287
MM2560 Mma 21228662	YEKVVEVYSESERKNEKI	-NKPLLVKEAES	20	SKIFENTEFOLE	5	FGDET-RUTET	2.0	LKSFT	2 RNCTAHGYSGIPTPS	TOEVOEYTKEL	211
ST0180 Stok 15920359	KDPSERISINGETPWILPTYEK-	-GKIRGAYFKLF	2.0	PESWVELVE	1	TLNMN-WSETL		HTLYF	1 GSEPTPEALGYNYP	-FLADEVAKEY	200
SS00467 Sso 15897396	IVKRENLNNVENPPWLVINEKH-	-DETIYGYFKLF		PSSWVFYIE	2	VEVEN-PEELL		SLIYN	1 GREPIPEALGYNYP	-FLADKLVKYY	216
ag 977 Aae 15606289	LPFIGYVKKHRKIYLYPEHIKV-	LOELKVGORT	16	SFDKYTWYYKLN	4	ITSVA-RLEVP	16	YLMPKEA	3 FNDKRAPONLIPIK	-LENYLERRLG	317
TM1739 Tma 15644485	VGPIGLVKNIGVTELSKEDFKK-	LRFLKKGKRS	9	PLKKVGAYVKLI	4	IRGLV-RLETY	18	KTLPHLT	2 LPIPRLPENILPIO	-LEENLSYYLT	288
DR0836 Dr 15805862	SAVVGYVKTLHTDYLGADRIGL-	LSSLKCGERT	19	REQRETWYYRLC	8	LAGIM-RLEMH	20	ALLSKLG	3 HKDSRAPQNLIPTA	-LEQAMNRSMG	329
Consensus/80%		p		hh		hh.bh		h	hs	scp.s.h.	
										-	

Figure 6. Multiple alignment of the NurA superfamily. The coloring reflects 80% consensus and the coloring scheme, consensus abbreviations and secondary structure representations are as in Figures 2 and 3. Species abbreviations are as follows. Aae, *A.aeolicus*; Aful, *A.fulgidus*; Ana, *Anabaena* sp.; Ape, *A.pernix*; Bha, *B.halodurans*; Cau, *C.aurantiacus*; Ctep, *C.tepidum*; Dr, *D.radiodurans*; Feac, *F.acidarmanus*; Glvi, *G.violaceus*; Halsp, *Halobacterium* sp.; Hehe, *H.hepaticus*; Mac, *M.acetivorans*; Mba, *Methanosarcina c*; Mebu, *Methanococcoides burtonii*; Mj, *M.jannaschii*; Mkan, *Methanopyrus kandleri*; Mma, *M.mazei*; Mth, *M.thermautotrophicus*; Naeq, *Nanoarchaeum equitans*; Npun, *N.punctiforme*; Pab, *Pyrococcus abyssi*; Pfu, *P.furiosus*; Pho, *P.horikoshii*; Pyae, *P.aerophilum*; Sac, *S.acidocaldarius*; Sso, *S.solfataricus*; Stok, *Sulfolobus tokodaii*; Syn, *Synechocystis* sp.; Tac, *T.acidophilum*; Tery, *T.erythraeum*; Thel, *T.elongatus*; Tma, *T.maritima*; Tvo, *T.volcanium* and Unk, Uncultured crenarchaeote.

The bacterial orthologs of Mre11 and Rad50 are, respectively, SbcD and SbcC proteins, which typically are encoded in a conserved operon present in most major bacterial lineages. Most likely, the bacterial SbcD-SbcC operon and the orthologous archaeal Mre11-Rad50 operon descended from an ancestral nuclease-ATPase operon of LUCA. Since both HerA-NurA and Mre11-Rad50 operons are much more common that the complete four-gene operon, it appears likely that the latter evolved in the common ancestor of crenarchaea and euryarchaea as a result of fusion of the two gene pairs. The available information on the functions of the eukaryotic Mre11 and Rad50 proteins provide hints regarding the possible functional significance of the genomic linkage of these four genes. The ABC ATPases of the SMC-family, which includes Rad50, are involved in chromatin dynamics associated with chromosome condensation and segregation (82,83). In particular, Rad50 bridges the double-strand breaks in DNA and facilitates end processing by the Mre11 nuclease (84,85). Therefore, in archaea, the Rad50 and Mre11 orthologs could function in a complex with HerA to repair double-strand breaks, which could potentially arise during the process of chromosomal segregation. Furthermore, Rad50 could also function in reorganizing the higher order chromatin structure during segregation. Archaeal kleisins, which are predicted to be functional partners of Rad50 proteins (86), are also likely to participate in this process. The predicted HerA-Rad50-Mre11-kleisin repair system might also function in double-strand break repair during archaeal DNA replication. However, in view of the structural, functional and evolutionary relationships between HerA and FtsK discussed above, it seems most likely that the principal, essential role of this system is linked to chromosomal segregation.

	Strl		Н	0.1 He	lix 0.2		Helix-1	Str2
Secondary Structure	EEEEE		HF	инннннн	ннинининини.			····EEEEE
XAC2443_Xca_21243176	67 FQSEHLSLLAGSGLTHAVH -	-YLATSKAAAGMDA	57 LTGILNAFSP	SILQSERGI	ATAEEARREQALNT		LVAFLMSFASRS	1 VRDRLNIFTTNYD
XAC2189 Xax 21242924	32 PHYHRCTTLICNCASMAUS	-PSECVCSLLOHAO	29 OASNUNKSIE	TPDDRTHOA	VAGEDPASEDAALG	3	-ALQSTLLSTASKA	1 IKCEDTVISLNVD
BRA0375 Brsu 23500128	12 VENYGCTILLGNGASISVN -	-PSESYTSLLOHAT	29 OASNVNRSLE	IPDERTHAT	YVALRDCLIOTVRD	3	EHGAVSAHLPTIYO	1 LKRFOTVISLNYD
NE2530 Neu 30250449	14 AE-GWSS <mark>LL</mark> LGNGASIAIH -	-KEFAYPTLHGIAD	28 YAEHVNGALO	TPSAAISAA	YEEVRTALIEAVHS	3	VHADVAADLQRVGA	1 ASAFPTVVSLNYD
Pf1u2135_Pf1_23059957	39 ATTDFSG LL LGNGASRAVW -	-DDFGYDSLFENAR	31 TTSRVNKAL	VSSAAPRNR	YYAIKEALINTVHA	4	WRLIEASTLTTLNS	1 LARYRT VFTTNY D
PP1406_Pput_26988140	16 ARHPCDALLLGNGASRAVW -	-KPFGYFSLFEQAQ	32 TTVRTNAALA	INSTAPLNR	YYSIKEALIHAVRT	4	WPLMPKATLATLNQ	1 LRGYRS VY TSN Y D
LA1794_Lepin_24214494	15 PENKKPHLLGNGFSISWN V	VNKESYQSLLDKAD	28 DASKVVKYYN	INKNLVDDLI.	YDANKLKETLAQTI	7	PSEIDRDSYEHCKK	1 LSYFKH IY TLNYD
LA1852_Lepin_24214552	1MKVAFLLGAGFSYDLG -	-MPLGIDLTNYFLN	54 ELASISKEGO	SVIKTSYRYL	LNLFYETIYSNLML	9	YKTNEDLYAGLKHV	1 NENETWFFTLNHD
NE2094_Neu_30250034	13 VYCKAPITICSCASAAHC	-MEGMEGLAQHLID	36 VSEELICRI	NSIWSLINS	EDAALFKNSLQNSS		-MEPLORLLEHMER	1 SLEKINIVITNID
EF2349 Efae 29376850	13 LSEKRTTFLFGAGASVPFF -	-SSLGKIEEYMTSK	29 KKGESENHLI	SGSSVMNOY	SRFIFNCV		EFLKLRNSR	1 SPRRTNIFTTNYD
VVA0689 Vivul 37676349	33 COLENVGVLLGAGASKSAG -	-GMVMKEIWLDFKS	40 VMDTWOODST	ALFNVLVEL	YRSVTKAALLVNSD	13	KHRNLLEKLLSNRO	1 GOAAPSIFTTNYD
VP1802 Vipa 28898576	90 DQANSVMLEKGYVLSQYYL -	-HVDLSALTVDILD	87 LGMALDAQHI	SLKAALKDI	WDKPDFLTLEQQQL	67	GSLPHAFETALVDF	1 QQTKSHVATLNYD
PG1107_Porgi_34540840	73 SNRQSIA <mark>IL</mark> LGAGFSAPKG -	-YPIGNDINKKIQN	80 DLINKSDDYI	NYLFNVVHI	YNQMVLYLLKDKEG	13	ADAYNEFLLYLSQQ	1 KECIINIHTLNHD
Sir2_Smel_16264652	12 LERRHAVLFVGAGVSMSVG I	LPSWKTLVDHILSE	10 PDISYQTIAE	CYRLQHGDI	ISLCEWMRQNWRV-	1	PERIERSELHRLIV	ELDFPV IY TTN Y D
AK024756.1_Hs_10437125	38 KKPRELVLVIGTGISAAVA 5 1	LKSWKGLIQALLDA	9 EESKKFQKCI	HEDKNLVHV	AHDLIQKLSPRTSN	11	EVFDGLESKMEDSG	10 MENGALVLTINFD
FLJ21103_Hs_13375721	38 KRPRELVLVIGTGISAAVA 5 I	LKSWKGLIQALLDA	9 EESKKFQKCI	HEDKNLVHV	AHDLIQKLSPRTSN	11	EVFDDLESKMEDSG	10 MENGALVLTINFD
1 71005 Tado (2510021	25 KOPOLLLOVIGTOVSAAVA 5 1	SCIEAVIEA	3 GDVAEFRRA	TRDRDLLVV	CKCMAONDEEXTKV	11	EVEDDLEQHIRSPL	3 INICATETTINE
RPA0819 Rhpa 39933896	64 VREKGAAVEFGAGLSIPCG	IPGWSOLLONFKIE	-RALLEDADI	RDDPLTMAE	LASKOVGHENLOET	5	OAATKFSVNHALLA	ALSCRIVITINYD
CV1059 Chvi 34496514	29 INSOKVTIFAGAGISTESR 1	IVLKNTFYDDVALE	9 VKELTFPNLM	OALKSOPNG	RIKLLEMITKRFSL	5	ELRKMATRFHRELG	1 FFPIKNIITTNWD
TDE0266 Tden 42525782	8 YNEKKVI LF VGAG <mark>V</mark> SKNLK I	LPTWSELIDQIAKD	5 DVFKTYGENI	ALAEYYRII	KGNIGPLRSWMDTN	4	GINISDSKIHELIA	KAKFPI IY TTN Y D
MBNC211202_Mesp_45916072	12 IKRRSAVLFVGAGVSMSVG I	LPSWDKLIRQMEQE	8 KSSRYSYQTI	AEYYRLKQG	SLGPLRSWMDRNWS	2	REKVKESLLHRLIV	ELDFPI IY TTN Y D
usg_Pepe_11322457	16 INEGYAAVFAGAGLSVASG Y	YVDWKTLVTPLAQE	LNLNIEEE	NDLTKVAQY	YRNEFKSRGDINQT	5	SKEKEPNENIDILT	RLPIYT YW ITN Y D
Aple047201_Acpl_32033900	14 LSANNLA IFAGAGL SVSAG F	FVNWKELLTDLAKE	2 LDINKEEN	DLVSLAQYY	VNSKCGKRSKINYL	5	SRKAEITENHRILA	RLPIDTEWTTNYD
BB0483_Bobr_33599473	16 LAEGNLAIFAGAGLSRAAG	FVDWKSLLKPIADD	1 DLDVEKEW	DLVTLAQYH	ANANATNRSKLNOM	5	SSGVAPTENHEILT	RLPIQTYWTTNYD.
BC1912 Bce 30020052	14 KINIAFF DE IGSGESKKIL I I	LEDWAGLMARE SNL	29 IWWREQUIER	SPERFRORT	SSRQSPLKVEVARI	8	GLUERNDREIDALK	OWIDGITTEM
BC1271 Bce 30019423	18 OLAISPVFFVSSRISRBYF 1 A	APDWESLLREVRDT	25 KLCDMYYEKI	ADEOL-EPN	KDKTYYFKKRITDI	8	MOSEDDNHETEAFK	KTSPSAVITTNYD
BK5-Tp30 bk5-t 14251154	9 DNNOFPIIFVGSGVTKRYF 2 0	GLKWEOLLLELWNL	34 MMAGILEEKI	NNAFYSDEL	NIDNLTLAOAHTEH	9	NTFSNLDRKKGFDE	8 LVKARFIVTTNYD
SAG1992_Saga 22538128	9 ENNOFPIIFVGSGITORYF 2 #	APTWEKLLKDIWLE	31 LLEKEVSKAN	INGNIQVDN	LDLKTAYELNISPF	5	NRFSNLKIREEKIE	8 LSKARI
plu4324Phlu_37528154	7 NFKNYPI VF IGSGMSKRYL 2 S	SPTWPELLEEYWNK	38 YIEKKFKELH	'NSGKISLEG	LTSKRVFDEDISPF	5	QRFKQDKLRQNIKK	9 LKNAKMIITTNYD
lici_Aful_14278228	21 AESKYLVALTGAGVSAESG 1	I PTFRGKDGLWNRY	-RPEELANP(AFAKDPEKV	WKWYAWRMEKVFN-		AQPNKAHQAFAELE	1 LGVLKCLITONVD
PAB0801_Pab_14521406	10 ASSKNAIAFTGAGISAESG V	/PTFRGKDGLWNKY	-RPEELATPE	AFARNPKLV	WEFYKWRINKILK-		AKPNPAHYALVELE	1 MGILRAVITONVD
APE1/82_Ape_14601621	14 ANSRFAVAPTGAGISAESG 1	IPTFRGKDGLWSRF	-DPRDLATPE	AFNRDPRLV	WEWYSWRIERVLA-		AKPNKAHRLLARLE	1 SGVLKAVITONVD
aq_2170_Ade_15607106	12 KNEWNIVIERGACUSTESC 1	I PIERGRUGLWINKE	A FTDFOLUCIO	FFFFVDFFF	NEWIDWERQUIAE-		CKDNDAUTATAKTE	1 MCKIKATUPONTO
TM0490 Tma 15643256	11 NESELTVTLTGAGISTESG 1	I PDFRGPNGTYKKY	SONVEDI	FFYSHPEEF	VRFAKEGIFPMLO-		AKPNLAHVILLAKLE	1 KGLIEAVITONID
HST1 Sc 6324504	198 RNAKKILVLTGAGVSTSLG	IPDFRSSEGFYSKI	5 EDPODVFNLI	IFLODPSVF	YNIAH-MVLPPE		NMYSPLHSFIKMLQ	1 KGKLLRNYTONID
Sir2p_Sc_6320163	252 HTARKIL <mark>VL</mark> TGAG <mark>V</mark> STSLG 1	IPDFRSSEGFYSKI	5 DDPQDVFNYN	IFMHDPSVF	YNIAN-MVLPPE		KIYSPLHSFIKMLQ	1 KGKLLRNYTONID
SSO2478_Sso_15899220	11 ISSSYTI <mark>AF</mark> TGAG <mark>I</mark> STASG 1	I PDFRGPQGLWKKY	SPELASIE	YFEKDPKNF	WGFYSLRMRGLFE-		AQPNKAHYSLAELE	1 MGIIKVIITQNID
cobB_Ec_26247264	38 MEKPRVLVLTGAGISAESG 1	IRTFRAADGLWEEH	-RVEDVATPE	GFDRDPELV	QTFYNARRRQLQQP	1	IQPNAAHLALAKLQ	2 LGDRFLLVTQNID
TTC1026_Thth_46199328	12 EEAKRVAVLTGAGISKPSG 1	I PTFRDAEGLWKNF	-NPLDYATPH	AYARDPERVI	WAWYAWRIQKVRE-		AKPNPAHYALVELE	5 RGGSFLLVTQNVD
Conceptus/80%	40 ARARHIVIISHAWSALSG	ch h	-QAQDLATPI	ATAINPERV	WEFINIKKEVPGS-		REPNAGRRAIAEUE	5 QGKKVVVITQNID
consensus/ ove	·p·p·····		· · · · P · · · · · · ·	********				· · · · · · · · · · · · · · · · · · ·
	H2		Str-3	Heli	x-3 5	tr4	He	14 Str5
Secondary Structure	Н2 ННННН		Str-3	Heli	х-3 5 НННННЕЕЕ	tr4 EEE	Не Е*.*ННННН	14 Str5 HHHHEEEEEEE
Secondary Structure XAC2443_Xca_21243176	H2 HHHHH LIEAGAELAGLHLLDRFLGNL	 TP 19 GEPRYLEG	Str-3 EEEEEEEE. VAR <mark>F</mark> TR <mark>L</mark> HGS <mark>V</mark>	Heli HHHHH 58 <mark>F</mark> RDL <mark>A</mark> AA	x-3 5 HHHHHHEEE I C RPNSTL V IY	tr4 EEEI G <mark>Y</mark> SI	He E*.*HHHHH FG <mark>PETINRVI</mark> RDM	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIS 1
Secondary Structure XAC2443_Xca_21243176 b111926_Brja_27377037	H2 HHHHH LIEAGAELAGLHLLDRFLGNL LIEHGCDFAGLRIIDRFVGAL	TP 19 GEPRYLEG	Str-3 EEEEEEEE. VAR F TR L HG SV VIR L TK L HG SL	Heli HHHHH 58 FRDLAAA 55 FRDFAAA	x-3 S HHHHHHEEE I C RPNS TLVTY A C RPNAVV VTY	tr4 EEEI GYSI GYGI	He E*.*HHHHH FG <mark>PE</mark> INRVIRDM FG <mark>P</mark> P <mark>V</mark> NRVIFDM	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIS 1 L-TIPSTHLVIMA 1
Secondary Structure XAC2443_Xca_21243176 b111926_Brja_27377037 XAC2189_Xax_21242924 DBA031F_Berry 321242924	H2 HHHHH LIEAGAELAGLHLLDRFLGNL LIEHGCDFGLNYQDGHRFKD LVYWMTYGLNYQDGHRFKD	TP 19 GEPRYLEG NP 19 GEPRFMEG CF 16 RGLYREQT	Str-3 EEEEEEEE. VARFTRLHGSV VIRLTRLHGSL NTLVFYPHGNL	Heli HHHH 58 FRDIAAA 55 FRDFAAA 56 ATVYREV	x-3 S HHHHHHEEE ICRPNSTIVTY ACRPNAVVVTY LTSHRSTLTLF	tr4 EEE GYS GYG GWG	He E*.*HHHHH FGIE INRVIRDM FGID VNRVIEDM IGEHOR ILKRWYGT	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIS 1 L-TIPSTHLVIMA 1 GIQRVAVSVPGGS CIND UMEVEDCD
Secondary Structure XAC2443_Xca_21243176 b111926_Brja_27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NF2530_Beu_30250449	H2 HHHHH LIEHGGBLAGLHLLDRFLGNL LIEHGCDFGLRIDRFVGAL LVYWTMTYGLNZDGHRFKD VYWAITYGLDZDGHRAFKD	TP 19 GEPRYLEG NP 19 GEPRFMEG CF 16 RGLYREQT CF 14 QPIRGERS	Str-3 EEEEEEEE. VARFTRLHGSV VIRLTKLHGSL NTLVFYPHGNL TSLVFYPHGSL	Heli HHHH 58 FRDLAAA 55 FRDFAAA 56 ATVYREV 56 STVYREV 59 TNVYFEV	x-3 S HHHHHHEEE ICRPNSTIVTY ACRPNAVVVTY LTSHRSTITLF LKSQRDTLVY LGAIGE-FSIVY	tr4 EEE GYS GYG GWG GWG	He E*.*HHHH FG'E INRVIRDM FG'D VNRVIEDM IGEH R LLKRMYGT FAEQ'I LLQRNGT FAEQ'I LLQRNGT	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIS 1 L-TIPSTHLVVIS 1 GIQRVAVSVFGGS GINRVAVSVFGGS GINRVAVSVFRGD DEVEMAVSVFRGD
Secondary Structure XAC2443_Xca_21243176 b111926_Brja_27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530_Neu_30250449 Pf1u2135_Pf1_23059957	H2 HHHHH LIEAGAELAGLHLLDRFLGNL LIEHGCDFAGLRIIDRFVGAL LVYWTMTYGLNUQDGHRFKD VYWAITYGLDIDDRHAFKD TLYWAMLLFNAANGSWFKDA LNYWALDFNAANGSWFKDA	TP 19 GEPRYLEG NP 19 GEPRFMEG CF 16 RGLYREQT CF 14 QPIRGERS FH 12 RPYGHAAG AS 1 DLCESHTG	Str-3 EEEEEEE. VARFTKLHGSV VIRLTKLHGSL NTLVFYPHGNL TSLVFYPHGSL ATLVFYPHGSL KPRLVVHGGL	Heli HHHH 58 FRDIAAA 55 FRDFAAA 56 ATVYREV 56 STVYREV 59 TNVYEEV 57 LSFCYDO	x-3 S HHHHHHEEE ICRPNSTLVTY ACRPNSTLVTY LTSHRSTLTLF LKSQRDTLVY LPALGESLVY LPALGESLVY	tr4 GYSI GYGI GWG GWGI GWGI GWGI GHA	He E*.*HHHHH FGIE INRVIRDM FGID WNRVIRDM IGHPR LIKRMYGT FAEQII LIQRMNGT FAEQII LIQRMNGT FDERIG VIAALAAN LGGOS LIVRABELA	14 Str5 HHHHEEEEEEE L-TIPSTHLVIS 1 L-TIPSTHLVIMA 1 GIQRVAVSVFGG GINRVAVSVFGG PFKRMAVSVFTGQ KFKTVATSIVPRS
Secondary Structure XAC2443 Xca 21243176 b111926_Brja 27377037 XAC2189_Xax_21242924 BRA0375_Brsu 23500128 NE2530_Neu 30250449 Pflu2135_Pf1_23059957 PPl406_Pput 2698140	H2 HHHHH LIEAGAELAGLHLLDRFLGAL LVYWTMTYGLNVQDGHRFKD VYWNAITYGLDIDDRHAFKD TLYWANLQHQGEVIDDLFNGFD LLYNNVQHQGEVIDDLFNGFD	TP 19 GEPRYLEG NP 19 GEPRYLEG CF 16 RGLYREGT CF 14 QPIRGERS FH 12 RPYGHAAG NAS 1 DLCESHTG YF DVRRTGSE	Str-3 EEEEEEE. VARFTKLHGSV VIRLTKLHGSL TSLVFYPHGSL ATLVFYPHGSL ATLVFYPHGSL KPRLLYLHGGL GTRVLLHGGL	Heli HHHHH 58 FRDIAAA 55 FRDIAAA 56 ATVYREV 56 STVYREV 59 TNVYEEV 57 LSFCYDQ 53 LSWCLQQ	x-3 S HHHHHHEE ICRPNSILTY ACRPNAVVIY LISHRSILTF LKSQRDTLVY LPALGESLVY LEHGDTLVY LLEHGBILCIP LAQENHGICLF	tr4 EEE GYS GYG GWG GWG GWG GWS GHA GQH	He E*.*.HHHHH FG E INRVIRDM FG D VNRVIRDM IGEH R LLKRMYGT FDER C VLALARNGT FDER C VLALAN LGEQ S LVRALRA LGSS R LLDAVROA	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIS 1 L-TIPSTHLVVIS 1 GIQRVAVSVFGG GINRVAVSVFRGD PPKRMAVSVFTGQ KPKTVAISIYPRS RPQQ-LSIAIRPLG
Secondary Structure XAC2443 Xca 21243176 b111926_Brja 27377037 XAC2189 Xax 21242924 BRA0375_Brsu 23500128 NE2530 Neu 30250449 Pf1u2135_Pf1_23059957 PP1406_Pput_26988140 LA1794_Lepin_24214494	H2 HHHHH LIEHGAELAGLHLLDRFLGNL LIEHGCDFGLNYQDGHRFKD VYYWAITYGLNYQDGHRFKD TLYWAMLD	TP 19 GEPRYLEG NF 19 GEPRFMEG CF 16 RGLYREGT FH 12 RPYGHAAG FH 12 RPYGHAAG YF DVRRTGSE Sy5 5 TWDIGNTN	Str-3 EEEEEEEE. VARFTKLHGSV VIRLTKLHGSL NTLVFYPHGSL ATLVFYPHGSL ATLVFYPHGSL GTRVLHLHGGL SGNIFFLHGAL	Heli HHHH 58 FRDIAAA 55 FRDFAAA 56 ATVYREV 59 TVYREV 59 TVYEEV 51 LSFCYDQ 53 LSWCLGQ 55 LNRGLRS	x-3 S HHHHHHEEE ICRPNSTLVIY ACRPNAVVIY LTSHRSTLVIY LKSQRDTLVYY LEHGESLVY LEHGHGICLP FAULTDFLIF	tr4 EEE GYG GYG GWG GWG GWG GWG GHA GQH GHS	He E*.*	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIS 1 L-TIPSTHLVIMS 1 GIQRVAVSVFGG GINRVAVSVFGG PFKRMAVSVFTGQ RFKTVATSIYFRS RPQQLSIAIRPLG KFKQLFISIYGNH
Secondary Structure XAC2443 Xca_21243176 b111926_Brja_27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530 Neu_30250449 Pf1u2135_Pf1_23059957 PP1406_Pput_26988140 LA1794_Lepin_242124592	H2 HHHHH LIEHGCDFGLRLUDRFLGNL LVYWTMTYGLNIDRFVGAL LVYWTMTYGLNIDDRHAFKD VYYMAITYGLDIDDRHAFKD LLYWALQHQGEVIDDLFNGPD LLPWAVQHAPSGFAELFDEQG LLYWIMQDEITPTFTCDDGFRNPD YLELLCIDYDIPATYGDTEVIKFPI	TP 19 GEPRYLEG NP 19 GEPRYMEG CF 16 RGLYREQT CF 14 QPIRGERS FH 22 RPYGHAAG AS 1 DLCESHTG VYF DVRRTGBE S0 5 TWDIGNTN DN 21 KAYFKNKF	Str-3 VARFTRIHGSV VARFTRIHGSV NTLVFYPHGNL TSLVFYPHGSL ATLVFYPHGSL KPRLLYLHGGL GTRVLEHGGL GANIVRLHGGL	Heli HHHH 58 FRDLAAA 55 FRDFAAA 56 ATVYREV 59 TNVYEEV 57 LSFCTDQ 53 LSWCLGQ 55 LNRCLRS 55 LNRCLRS 86 KLKLFED	x-3 S HHHHHHEEE ICRPNSTLJTY LTSHRSTLITP LKSQRDTLVVY LEHGSLUY LEHGDNLCIP LACENHGICLP FANLTDPLIIF VMKSVDKLIII	tr4 GYSI GYGI GWG GWGI GWGI GHA GQH GQH GHSI GYG	He E*.*HHHHH FGO I INRVIRDM FGO VNRVLFDM IGEH R LIKRMYGT FAEQI I LIQRMNGT FAEQI I LIQRMNGT FDER C. VLAAIAAN LOEQI S LVRABRLA LDSS R LIDAVRQA LTDS N ILKIMEKG FGOQ INFRINHQ	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIS 1 L-TIPSTHLVIMA 1 GIQRVAVSVFGGS GINRVAVSVFFGG PFKRMAVSVFFGG PFKRMAVSVFFG KFKQLSTAIRPLG KFKQLFISIVGNH LVKNEFTIEIVD
Secondary Structure XAC2443_Xca_21243176 b111926_Brja_27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530_Neu_30250449 Pf1u2135_Pf1_23059957 Pp1406_Pput_26988140 LA1794_Lepin_24214592 NE2094_Neu_30250034	H2 HHHHH LIEHGGAELAGLHLLDRFLGNL LIEHGCDFGLNIDRFVGAL LVYWTMTYGLNIDDRHAFKD TLYWAMITGLDIDDRHAFKD LYWALITPNAANGSWFKDA LYWALQHAPSGFAELFDEQG LLYWTIMQDEITPTFTCDGFRWDD LLYWTIMQDEITPTFTGDTEVIKFPI LLELLCID-YDIPATYGDTEVIKFPI LAEYACDQSRIHHYTGFTHGFFRQL	TP 19 GEPRYLEG NP 19 GEPRYLEG CF 16 RGLYREQT CF 14 QPIRGERS FH 12 RPYGHAAG AS 1 DLCESHTG TYF DVRRTGSE SD 5 TWDIGNTN DN 21 KAYFKNKF AT PDELTCSR	Str-3 VARFTKLHGSV VIRLTKLHGSV NILVFYPHGSL ATLVFYPHGSL ATLVFYPHGSL GTRVLHHGGL GTRVLHHGGL GONIVKLHGGL RVNIWKVHGSL	Heli HHHH 58 FRDLAAA 55 FRDFAAA 56 ATVYREV 59 TNVYEEV 57 LSFCYDQ 53 LSWCLQQ 55 LNRCLRS 86 KLKLFED 45 IINNADI	x-3 S HHHHHHEEEI ICRPNSTLVTY LTSHRSTLIE LKSQRDTLVYY LPALGESUVY LDALGESUVY LACENDNLCIF FANLTDPLIIF WKSYDKLIII AIMEAGSYLCI	tr4 GYG GYG GWG GWG GWG GHA GQH GYG GYG GYG	He E*.*HHHHH FG E INRVIRDM FG D VNRVLFDM IGEH RILLKRMYGT FAEQ I LLQRMNGT FDERIQ VLAAIAAN LCSS R LLDAVRQA LDSS R LLDAVRQA LDSS N ILKLMKG FG Q INFRINQ FN E VQPKLMAK	14 Str5 HHHHEEEEEEE L-TIPSTHLVINS 1 GIQRVAVSVFGG GINRVAVSVFGG FPKRMAVSVFTGQ KFKTVAISIYPRS RPQQ-LSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQGAPVTIITY
Secondary Structure XAC2443 Xca 21243176 b111926_Brja 27377037 XAC2189_Xax 21242924 BRA0375_Brsu 23500128 NE2530_Neu 30250449 Pflu2135_Pfl 23059957 PPl406_Pput 26988140 LA1794_Lepin_24214494 LA1852_Lepin_24214552 NE2094_Neu 30250034 seal0_Seen_38201747 Pr2346_Brc20027650	H2 HHHH LIEAGAELAGLHLLDRFLGAL VYWMTYGLNVQDGHRFKD VYWMITYGLVQDGHRFKD TLYWALQHQGEVIDDLFNGPD LLPWAVQHQGEVIDDLFNGPD LLPWAVQHQGEVIDDLFNGPD LLPWAVQSRIHHYGFTHGFRQL LAEYXCDQSRIHHYGFTHGFFRQL LAEYXCDQSRIHYTGFTBGFRQD LAEYXCDQ	TP 19 GEPRYLEG NP 19 GEPRYMEG CF 16 RGLYREGT CF 14 QPIRGERS FH 12 RPYGHAAG NAS 1 DLCESHTG FJ DVRRTGSE SD 5 TWDIGNTN DN 21 KAYFKNKF AT PDELTCSR FR 4 PNELTSSR	Str-3 VARFTKLHGSV VIRLTKLHGSV VIRLTKLHGSL MILVFYPHGSL ATLVFYPHGSL ATLVFYPHGSL GTRVLHLHGGL GTRVLHLHGGL GANIVKLHGGL RVNIKKVHGSL RVNIKKVHGSL	Hell HHHH 58 FRDIAAA 55 FRDIAAA 55 FRDIAAA 56 ATVYREV 56 STVYREV 57 LSFCYDQ 53 LSWCLGQ 53 LSWCLGQ 53 LSWCLGQ 54 LNRCLRS 86 KLKLFED 45 IINNADI 45 TINNADI	x-3 S HHHHHHEEU ICRPNXVVIY LTSHRSTLTLY LKSQRDTLVVY LPALGDTLVVY LDALGESLVVY LDALGESLVVY LAQENHGICLF FANLTDPLIIF VMKSVDKLIIT MKSVSYLCI AINAASYLCI	tr4 EEEE GYG GYG GWG GWG GHA GQH GYG GYG GYG GYG	He E*.*.HHHHH FG E INRVIRDM FG D VNRVIEDM IGHP R LIKRWGT FAEQII LLQRMNGT FDER C VLAAIAAN LGEQ S IVRALRA LDSS R LLDAVRQA LDSS N ILKIMEKG FG C INFRINHQ FN E VQPKMAK FN E VQPKMAK	14 Str5 HHHHEEEEEEE L-TIPSTHLVINS 1 L-TIPSTHLVIMA 1 GIQRVAVSVFGG GINRVAVSVFGG KFKTVAISIVPRS RPQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQQAPVTIITY CIRQNTPITIITY
Secondary Structure XAC2443 Xca 21243176 b111926_Brja 27377037 XAC2189_Xax 21242924 BRA0375_Brsu 23500128 NE2530_Neu_30250449 Pflu2135_Pfl_23059957 Pfl406_Ppt_2698140 LA1794_Lepin_24214494 LA1852_Lepin_24214552 NE2094_Neu_30250034 seal0_Seen_38201747 EF2349_Efae_29376850 VyA0688_Vivn1_37656349	H2 HHHHH LIEHGCDFAGLHLLDRFLGNL LIEHGCDFGLRIDRFVGAL LVYWMTYGLNVQDGHRFKD TLYWAITYGLNVQDGHRFKD LLYWALQHQGEVIDDLFNGPD LLPWAVQHQGEVIDDLFNGPD LLYWIMQDEITFTFTCDDGFRNPD LLYWIMQDEITFTFTCDDGFRNPD YLELLCIDYDIPATYGDFVIKFFI LAEYACDQGTHHYTGFSHGF FLEIAVEK-NIESNPKIFFNDGTNGTM LTEWSAFESUNIMCFSGH	TP 19 GEPRYLEG NP 19 GEPRFMEG CF 16 GLVREGT FH 12 RPYGHAAG FH 12 RPYGHAAG SS 5 TWDIGNTN DN 21 KAYFKNFF AT PDELTCSR FF 4 FNELTSAR KR 17 DNYANELP 97 7 GEABEGRY	Str-3 EEEEEEEE. VARFTKLHGSV VIRITKLHGSI NTLVFYPHGSI ATLVFYPHGSI ATLVFYPHGSI GTRVLHLGGL SONIFFLHGAL GANIVKUHGGI RVNIWKVHGSI TINLVKCHGSV TINLVKCHGSV	Heli HHHH 58 FRDIAA 55 FRDIAA 56 ATVYREV 59 TNVYREV 57 LSFCYDQ 53 LSWCLGQ 55 LNRCLRS 86 KLKLFED 55 TINNADI 45 TINNADI 45 TINNADI 45 TINNADI	x-3 S HHHHHHEE ICRPNSTUTY ACRPNAVVYIY LTSHRSTUTY LACRTUVY LEHGDNLIT FANLTDFLIF WMXSVDKLII AINAASSTCI LEKEQSULAF	tr4 GYS GYG GWG GWG GWG GWG GYG GYG GYG GYG GYG	He E*.HHHHH FG D INRVIEDM FG D VNRVLEDM IGEH R LLKRMYGT FAEQ I LLQRMNGT FDER Q VLAAIAAN LGEQ S LVRALRLA LDSS R LLDAYRQA LTDS N LLLKIMEKG FS Q INFRIMQ FN E VQPKIMAK FN E IQPNIMAK FL E ITDIVQRS	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIS 1 L-TIPSTHLVIMA 1 GIQRVAVSVFGGS GINRVAVSVFTGG PFKRMAVSVFTGG PFKRVATSIYPRS RPQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQGAPVTIITY CIRQNTPITIITY L-NN-PSLLVFIFCY 1 L-NN-PSHLTVFY
Secondary Structure XAC2443_Xca_21243176 b111926_Brja_27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530_Neu_30250449 Pflu2135_Pfl_23059957 PP1406_Pput_26988140 LA1794_Lepin_24214542 LA1794_Lepin_24214552 NE2094_Neu_30250034 seal0_Seen_38201747 EF2349_Efae_29376850 VVA0689_Vivul_37676349 VP1802_Vipa_28898576	H2 HHHHH LIEAGAELAGLHLLDRFLGNL LIEHGCDFGLRIUDRFVGAL LVYWWMTYGLRIUDRHAFKD VVYWAITYGLNUQDGHRFKD VYWAITYQGEVIDDENGFD LLYWNWALQHQGEVIDDLFNGFD LLYWIMQ-DEITFTFTCDDFRNPD YLELLCIDYDIPATYGDTEVIKFPI LAEYXCDQSGIHLYTGFTHGFFRQL LAEYXCDQSGIHLYTGFSHGF FLEIAVEK-NIESNPKIFFNDGTNGYM AIEMSAEESGINLINGFSGIH	TP 19 GEPRYLEG NP 19 GEPRYLEG CF 16 RGLYREQT CF 14 QPIRGERS NS 1 DLCESHTG VF DVRRTGSE S5 5 TWDIGNTN DN 21 KAYFKNKF AT PDELTCSR FR 4 PNELTSAR KR 17 DNYANELP SR 17 GEARFGHY LD 11 LERKYDNN	Str-3 EEEEEEEE. VARFTKIHGSV VIRITKIHGSI NTIVFYPHGSI TSIVFYPHGSI KPRLLYIHGGI GIRVILHGGI GANIVKIHGGI GANIVKIHGGI RVNIKVHGSI TINIVKCHGSV 1 NNYLYKIHGSI	Hell HHHH 58 FRDLÄAA 55 FRDTÄAA 56 ATVYREV 57 INVYREV 57 INVYREV 57 INSCIO 55 INSCI 55 INI	x-3 S HHHHHHEEE ICRPNSTLJTY ACRPNSTLJTP LISHRSTLTP LKSQRDTLVVY LDEHGDNLCIP LQENHGICLP YMKSVDKLIII AINAANSFLCI LEKEQSVLIAP LIKEQSVLIAP LKEQSVLIAP LKEQEIILP	tr4 EEEE GYS GYG GWG GWG GYG GYG GYG GYG GYG GYG GYG	He E*.HHHHH FGE INRVIRDM IGEH R LURMYGT FAEQU I LURMYGT FAEQU I LURMYGT FAEQU I LURMYGA LOEQUS IVRABRLA LDSS NI LURMYGA LDSS NI LURMYGA FGQ INFRINHQ FNE VQPKIMAK FLE ITDIVQRS FGY INRIIGA GFT I LMILKY	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIS 1 L-TIPSTHLVVIS 1 GIQRVAVSVFGGS GINRVAVSVFRGD PFKMAVSVFTGQ KPKTVAISIYPRS RPQQ-LSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQGAPVTIITY CCRQGAPVTIITY L-NN-PNLLVFIFCY 1 L-NN-PNLLVFIFCY 1 L-NN-PSLHVICY L-NTFSLHVICY
Secondary Structure XAC2443_Xca_21243176 bl11926_Brja_27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530_Neu_30250449 Pflu2135_Pfl_2305957 PP1406_Pput_26988140 LA1852_Lepin_24214592 NE2094_Neu_30250034 seal0_Seen_38201747 EF2349_Efae_29376850 VVA0689_Vivul_37676349 VP1802_Vipa_28898576 PG1107_Porgi_34540840	H2 HHHHH LIEHGCDFAGLHLLDRFLGNL LIEHGCDFGLRIJDRFVGAL LVYWMTYYGLNVQDGHRFKD VYYWAITYGLNVQDGHRFKD TLYWALLPNANGSWFKDA LNYWALQHQGEVIDDLFNGPD LLPWAVQHAPSGFAELFDEQG LLYWIMQBGTFHTTCDDGFRNPD YLELLCIDYDIPATYGDTEVIKFPI LAEYACDQSRIHHYTGFTHGFFRGL LAEYACDQGTHHYTGFSHGF FLEIAVEK-NIESNFKIFFNDGTNGFSGH LLYWSFISGINLINGFSGIH LLYNSFIDNDLVDGYNGSL LIESFNQTSNINGNISDGFDEYG	TP 19 GEPRYLEG NP 19 GEPRYLEG CF 16 RGLYREQT CF 14 QPIRGERS HI 22 RPYGHAAG AS 1 DLCESHTG VYF DVRRTGSE SD 5 TWDIGNTN DN 21 KAYFKNKF AT PDELTCSR FR 4 PNEITSAR FR 17 DNYANELP ISR 17 GEARFGHY JID 11 LERKYDNN SE 17 RYTGRYNT	Str-3 VARFTKLHGSV VARFTKLHGSL VIRITKLHGSL NTLVFYPHGNL ATLVFYPHGSL KPRLLYLHGGL GTRVLHHGGL GANIVKLHGGL GANIVKLHGGL RVNIWKVHGSL TINLVKCHGSV I NMYLYKLHGSI FGYYLHLHGSI	Hell HHHH 58 FRDIAA 55 FRDIAA 56 ATVYREV 56 STVYREV 57 LSFCYDQ 57 LSFCYDQ 57 LSFCYDQ 53 LSWCLGQ 55 LNRGLRS 56 KLKLED 45 IINNADL 45 TINNADL 45 TINNADL 45 TINNADL 45 TRRFSEF 76 FRFSEF 48 YWDYLQF 71 LFKKRN	x-3 S HHHHHHEEEI IGRPNSTLJTY LTSHRSTLTLF LKSQRDTLVVY LLSHRSTLTLF LKQCSLVY LLEHGDNLCIF FANLTDPLIFF WKSVDKLIII AINAASYLCI LEKEQSVLIF LKEQSVLIF LKEQSVLIF LKEQSULIF LKEQSULIF	tr4 EEEE GYS GYG GWS GWS GHA GQH GYG GYG GYG GYG GYG GYG GYG GYG	He E*.*	14 Str5 HHHHEEEEEEE L-TIPSTHLVINS 1 L-TIPSTHLVING 1 GIQRVAVSVFRGD GINRVAVSVFRGD PPKRMAVSVFTGQ KFKTVAISIYPRS RFQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQGAPVTIITY CIRQNTPITIITY CIRQNTPITIITY L-NN-PNLLVFIFCY 1 L-NN-PSLHVICY LKVTPLRVVEWSG FDYKNRPSFIIDK
Secondary Structure XAC2443 Xca 21243176 b111926 Brja 27377037 XAC2189_Xax 21242924 BRA0375 Brsu 23500128 NE2530 Neu 30250449 Pflu2135 Pfl 2305957 PP1406 Pput 26988140 LA1794 Lepin 24214494 LA1852 Lepin 24214552 NE2094 Neu 30250034 seal0 Seen 38201747 EF2349 Efae 29376850 VVA0669_V1vul 37676349 VP1802 Vipa 28898576 PG1107 Porgi 34540840 Sir2_Smel 16264652	H2 HHHHH LIEAGAELAGLHLLDRFLGAL VYWMTMTYGLNVQDGHRFKD TLYWMTMTYGLVQDGHRFKD TLYWAMLLPNANGSWFKDA LNYWALQHQGEVIDDLFNGFD LLPWAVQHAPGFAELFDEQG LLYWWIMQDEITPTFTCDDGFRNFD YLELLCIDYDIPATYGDTEVIKFPI LAEYACDQSRIHHYTGFHGFFRQL LAEYACDQSGTHHYTGFSGTH LLYNSFISGINLINGFNGYM AIEWSAEESGINLINGFSGIH LLYNSFIYGKRYAKISHV	TP 19 GEPRYLEG NP 19 GEPRYLEG CF 16 RGLYREQT CCF 14 QPIRGERS SHFH 12 RPYGHAAG AS 1 DLCESHTG VYF DVRRTGSE SSD 5 TWDIGNTN DN 21 KAYFKNKF AT PDELTCSR FR 4 PNELTSAR KR 17 DNTANELP SSR 17 GEARFGHY LID 11 LERKYDNN E1 7 RYTGRYNT RD -IASAPAG	Str-3 VARFTKLHGSV VIRLTKLHGSV VIRLTKLHGSL MILVFYPHGSL ATLVFYPHGSL KVRLLHGGL GTRVLHHGGL GTRVLHHGGL RVNIWKVHGSL RVNIWKVHGSL RVNIWKVHGSL NMYLYKLHGSL FGYYLHLHGSI VTRIKXHGDF	Hell HHHH 58 FRDIAA 55 FRDIAA 56 ATVYREV 59 TNVYEEV 57 ISFCYDQ 53 ISWCIGQ 53 ISWCIGQ 53 ISWCIGQ 54 INNADI 45 IINNADI 45 IINNADI 9 IRFISYE 57 FRRFSEF 48 YWDIGC 71 IFKKRN 24 VRFRSDA	x-3 S HHHHHEEU ICRPNXVVIY LTSHRSTLVIY AGCRPNAVVVIY LTSHRSTLVIY LAGCAESLVVY LDALGESLVVY LDALGESLVVY LDALGESLVVY LAGNANSFLCI LEKEQSVLIAP LEKEQSTLIAPN ALSEAEIILFVN ALSEANTLIAP	tr4 GYS GYG GWG GWG GWG GYG GYG GYG GYG GYG GYG	He E*.*	14 Str5 HHHHEEEEEEE L-TIPSTHLVIMS 1 GIQRVAVSVFGG GINRVAVSVFGG KFKTVATSIYPRS RPQQ-LSIAIRPLG KFKQ-LFISIYGNH LVKNEKFTIEIVD CQRQGAPVTIITY CIRQNTPITIITY L-NN-PRLHIVFIFCY 1 L-IN-PSLHIVICY FDYKNKPSFIDK MSIWRRSGHEKDR
Secondary Structure XAC2443 Xca 21243176 b111926_Brja 27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530_Neu_30250449 Pflu2135_Pfl_23059957 PPl406_Pput_26988140 LA1794_Lepin_24214494 LA1852_Lepin_24214552 NE2094_Neu_30250034 seal0_Seen_38201747 EF2349_Efae_29376850 VYA0689_Vivul_37676349 VP1802_Vipa_28898576 PG1107_Borgi_34540840 Sir2_Smel_16264652 AK024756.1_Hs_10437125	H2 HHHH LTEAGAELAGLHLLDRFLGNL LIEHGCDFAGLHLIDRFVGAL LVYWTMTYGLNVQDGHRFKD TLYWATTYGLNVQDGHRFKD LLYWALQHQGEVIDDLFNGPD LLPWAVQHPSANAGSWFKDA LLYWALQHQGEVIDDLFNGPD LLPWAVQHPGTHEVIKFPI LAEYACDQSRIHYTGFTHGFFRQL LAEYACDQSRIHYTGFTHGFFRQL LAEYACDQSRIHYTGFTHGFFRQL LAEYACDQSRIHYTGFTHGFFRQL LIWSFIDNDLVDGYNGSL LIWSFIDNDLVDGYNGSL LIEIAYEVQGKQLESLDLTDEKK	TP 19 GEPRYLEG NP 19 GEPRYMEG CF 16 GL/YREJ FH 12 RPYGHAAG NAS 1 DLCESHTG FF 12 RPYGHAAG SS 5 TWDIGNIN DN 21 KAYFKNKF AT PDELITCSR FR 4 PNEITSAR KR 17 DNIANELP FR 4 PNEITSAR KR 17 DNIANELP SI 7 GEARFGHY ID 11 LERKYDNN SE 17 RYTGRINT P - IASAFAG VL - EWAQEER	Str-3 EEEEEEEE. VARFTKLHGSV VIRITKLHGSV NTLVFYPHGSL ATLVFYPHGSL GTRVLHLGGL GGNVLHLGGL GANIVKLHGGL RVNIWKVHGSL TINLVKCHGSL TINLVKCHGSL FGYYLHLHGSP PIRLIXKLHGSP PIRLIXKLHGSP	Heli HHHH S8 FRDLAA S5 FRDLAA S6 ATVYREV S5 TNVREV S7 ISYCTO S3 LSWCLQO S5 LNRCLSS S4 KLKLFED 55 TINNADI 45 TINNADI 45 TINNADI 45 TINNADI 45 TINNADI 45 TINNADI 45 TINNADI 45 TINNADI 44 WDYLQF 7 FRRFSDA 44 WDYLQF 71 LFKKFRN	x-3 S HHHHHHEE ICRPNSTUTY ACRPNAVVYIY LTSHRSTUTY LKSQRDTLVYY LEHGDNLCTP FANLTDFLIF WMXSVDKLII AINEASYLCI AINEANSFLCI LEKEQSVLIAF AINEANELIIF NLSKANSLII LEKEQSVLIAF NLSKANSLII LECATVFFI NKSYLFLC	tr4 GYS GYG GWG GWG GWG GYG GYG GYG GYG GYG GYG	He E*.HHHHH FG D UNRVLFDM IGEH R LLKRMYGT IGEH R LLKRMYGT IGEH R LLKRMYGT IGEH R LLKRMYGT ILQRMNGT FAEQUI LLQRMNGT FAEQUI LLQRMGT FAEQUI LLQRMGT FOR C UNRALA ISSN 1200 FN E VQPKLMAK FL E ITDIVQRS FG Y INRIIGA GFIT INNLKPY CKDKGINEMIKDN MSDINTRILLBRI VDDTTPQALFLEA	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIS 1 L-TIPSTHLVVIS 1 GIQRVAVSVFGGS GINRVAVSVFTGG PFKRMAVSVFTGQ PFKRMAVSVFTGQ VERVVATSIYPRS RPQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CIRQNFPITIITY CIRQNFPITIITY L-NN-PSLHVVFTCY 1 L-NN-PSLHVVFTCY LKVTPLRVVEWSG FDYKNKPSFIIDK WSIWRSGEBKDR VKIWSDLEHHML
Secondary Structure XAC2443 Xca 21243176 b111926_Brja 27377037 XAC2189_Xax 21242924 BRA0375_Brsu 23500128 NE2530_Neu_30250449 Pflu2135_Pfl_23059957 PP1406_Pput_26988140 LA1794_Lepin_24214552 NE2094_Neu_30250034 seal0_Sean_38201747 EF2349_Efae_29376850 VVA0689_Viv_1_37676349 VP1802_Vipa_28898576 PG1107_Porgi_34540840 Sir2_Smel_16264652 AK024756.1_Hs_10437125 FLJ21103_Hs_13375721	H2 HHHHH LIEAGAELAGLHLLDRFLGNL LIEHGCDFGLRIDRFVGAL LVYWMTYGLRIDDRHAFKD VVYWAITYGLNUDGHRFKD LLYWAIUHQGEVIDDLFNGPD LLPWAVQHQGEVIDDLFNGPD LLYWIMQDEITFTFTCDDGFRNPD LLYWIMQDEITFTFTCDDGFNPD LLYWIMQSRIHHYTGFTHGFFRQL LAEYACDQSGIHHYTGFTHGFFRQL LAEYACDQSGIHLYTGFSHGF FLEIAVEK-NIESNPKIFFNDGTNGYM ATEWSAEEGGINLINGFSGIH LIYNSFIDNDLVDGYNGSL LISFNQTSNINGNISDGFDEVG NLEIAYEVQGKQLESLDITDEKK LLELYAADQGKQLESLDITDEKK	TP 19 GEPRYLEG NP 19 GEPRYLEG CF 16 RGLYREQT CF 14 QPIRGERS FH 22 RPYGHAAG NAS 1 DLCESHTG TYF DVRRTGSE SD 5 TWDIGNTN DN 21 KAYFKNFF AT PDELTCSR FK 4 PNEITSAR KKR 17 DNYANELP SSR 17 GEARFGHY ID 1 LERKYDNN SSE 17 RYTGRYNT RD -IASAPAG CU - EWAQEKR	Str-3 EEEEEEEE. VARFTKIHGSV VIRITKIHGSI NTIVFYPHGSI TSIVFYPHGSI ATIVFYPHGSI GIRVILHGGI GIRVILHGGI GIRVILHGGI GANIVKIHGGI TINIVKCHGSI TINIVKCHGSI TINIVKCHGSI TINIVKCHGSI TINIVKCHGSI PIFLIKIHGSP PIFLIKIHGSP VIRIIKYGDF KLSVLIHHGVY	Heli HHHH 58 FRDLAA 55 FRDLAA 56 ATVYREV 56 STVYREV 57 INVYREV 57 INVYREV 57 INVKEV 55 INRGRS 86 KLKLFED 45 IINNADI 19 IRFLSYE 57 FRFSEF 71 IFKKRN 24 YURVIOF 71 LFKKRN 24 EIQKIYE 24 EIQKIYE	x-3 S HHHHHEEE ICRPNSTUJTY ACRPNAVVTY LTSHRSTLTP LKSQRDTLVYY LEHGDNLCTF FANLTDFLIF YMXSVDKLII AINAANSFICI LEKEQSVLIF LKEQSVLIF LKEQSVLIF NLSKANSLII LEATVLFI NLSKANSLII LEATVLFI NKSFLFLCC	tr4 EEEE GYS GYG GWG GWG GYS GYS GYG GYS GYS GYS GYS GYT GWT GWT	He E*.*	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIS 1 L-TIPSTHLVIMA 1 GIQRVAVSVFGGS GINRVAVSVFRGD PFKRMAVSVFTGQ KPKTVAISIYPRS RPQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CCRQGAPVTIITY CCRQGAPVTIITY L-NN-PNLLVFIFCY 1 L-NN-PNLLVFIFCY 1 L-NN-PSLHIVICY LKVTPLHVVEWSG FDYKNKPSFIIDK WSIWRSGHEKDR VKHKSDLEHFML VKHKSDLEHFML
Secondary Structure XAC2443_Xca_21243176 bl11926_Brja_27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530_Neu_30250449 Pf1u2135_Pf1_2305957 PP1406_Pput_26988140 LA1852_Lepin_24214592 NE2094_Neu_30250034 seal0_Seen_38201747 EF2349_Efae_29376850 VVA0689_Vivul_37676349 VP1802_Vipa_28898576 PG1107_Porgi_34500840 Sir2_Smel_16264652 AK024756.1_Hs_10437125 FLJ21103_Hs_13375721 FLJ2103_Hs_13375721 FLJ220635_Hs_15489177 LU1095_Laio_42510921	H2 HHHHH LIEAGAELAGLHLLDRFLGNL LIEHGCDFGLRIUDRFVGAL LVYWTMTYGLNUQDGHRFKD VYWMITYGLDIDDHAFKD TLYWMLLPGGPLIDDLFNGPD LLPWNVQHAPSGFAELFDEQG LLYWIMO-DEITFTFTCDGFRNPD YLELLCIDYDIPATYGDTEVIKFPI LAEYXCDQSGIHLTGFSHGF FLEIAVEK-NIESNFKIFFNDGTNGYM AIEMSAEESGINLINGFSGIH LLYNSFUSGINLINGFSGIH LLYNSFUSGINLINGFSGIF LLEYAPCYGRYAKISHV LLELYADQGKQLESLDITDEKK LLELYADQKRQESLDLINDEKK	TP 19 GEPRYLEG NP 19 GEPRYMEG CF 16 RGLYREQT CF 16 RGLYREQT FH 22 RYTGHAAG NAS 1 DLCESHTG YFF DVRNTGSE SD 5 TWDIGNTN DN 21 KAYFKNKF AT PDELTCSR FR 4 PNEITSAR KR 17 DNYANELP ISR 17 GEARFGHY ID 11 LERKYDNN SE 17 RYTGRYNT TRD -IASAPAG VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR	Str-3 EEEEEEEE. VARFTKLHGSL VARFKLHGSL NTLVFYPHGNL TSLVFYPHGNL TSLVFYPHGSL GRNVLHGGL GRNVKLHGGL GRNVKLHGGL GNNVKLHGGL TINLVKCHGSV TINLVKCHGSV TINLVKCHGSV VTRIKXHGSF VTRIKXHGSF VTRIKXHGSF NTRLVKCHGSV KLSVLHHGVY KUSVLHHGVY KUSVLHHGVY	Hell HHHH 58 FRDLÄAA 55 FRDLÄAA 56 ATVYREV 59 TNVYREV 57 LSFCYDQ 53 LSFCYDQ 53 LSKCLGQ 55 LNRGLS 55 LNRGLS 55 LNRGLS 54 KLKLFED 45 TIHNADL 19 LRFLSYE 57 FRRFSEF 71 LFKKFRN 24 VLGKLYK 24 ELQKLYE 24 ELQKLYE 24 ELQKLYE	x-3 S HHHHHHEEE ICRPNSTLJTY LCRPNSTLJTP LXSQRDTLVVY LTSHRSTLTP LXSQRDNLCP LACENSTLTP NACSORIGICA FANLTDPLIF VMKSVDKLIII AINAANSFJCI LEKEQSVLAP LTKPQTLFVN NLSKASLIII LEATVLPI NKSFLFLCC TKSFLEV	tr4 EEES GYG GWG GWG GWG GHS GYG GYG GYG GYG GYG GYG GYG GYG GYG GY	He E*.*	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIMS 1 L-TIPSTHLVIMA 1 GIQRVAVSVFGGS GINRVAVSVFFGG PFKRMAVSVFTGQ KPKTVAISIYPRS RPQQLSIAIRPLG KFKQLFISIYGNH LFNV-PLFIVT L-NN-PNLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLHIVICY L&WTPLRVVEWSG FDYKNRSSFIIDK WSIWRRSGEEKDR VKHKSDLEHFML VSHKSDLEHFML VSHKSDLEHFML
Secondary Structure XAC2443 Xca 21243176 bl11926 Brja 27377037 XAC2189_Xax 21242924 BRA0375 Brsu 23500128 NE2530 Neu 30250449 Pflu2135 Pfl 23059957 PP1406 Pput 26988140 LA1794 Lepin 24214494 LA1852 Lepin 24214592 NE2094 Neu 30250034 seal0 Seen 38201747 EF2349 Efae 29376850 VVA0689_Vivul 37676349 VP1802 Vipa 28898576 PG1107 Porgi 34540840 Sir2 Smel 16264652 AK024756.1 Hs 10437125 FLJ2103 Hs 1337521 FLJ20635 Hs 15489177 LJ1095 Lajo 42519021 RPA0819 Rhba 3993896	H2 HHHHH LIEAGAELAGLHLLDRFLGAL LVYWTMTYGLRVQDGHRFKD VYWMITYGLVQDGHRFKD TLYWAMLLFNAANGSWFKDA LNYWALQHQGEVIDDLFNGFD LLPWAVQHAFGFAELFDEQG LLPWAVQHAFGFAELFDEQG LLYWTIMQDEITPTFTCDDGFRNPD YLELLCIDVDIPATYGDTBCFNGF LLEYACDQ-SRIHHYTGFTHGFFRQL LAEYACDQ-SRIHHYTGFTHGFFRQL LAEYACDQSGTHLIYTGFSGTF LLEIAVEK-NIESNFKIFFNGGTNGYM XIEWSFISGINLINGFSGIH LLYNSFISGINLINGFSGIH LLEYAADQGRQLESLDLTDEKK LLEYAADQGRQLESLDLTDEKK LLEYAADQGRQLESLDLTDEKK LLEYAADQRNPMSLDLKDKTK QIEKAYEATFQSRPNILKDIN	TP 19 GEPRYLEG NP 19 GEPRYMEG CF 16 RGLYREGT FF 12 RPYGHAAG NAS 1 DLCESHTG FF 12 RPYGHAAG ST DVRATGSE SD 5 TWDIGNTN DN 21 KAYFKNKF AT PDELTCSR FR 4 PNEITSSR KR 17 DNYANELP SR 17 GEARFGHY ID 11 LERKYDNN SE 17 RYTGRYNT RD -IASAFAG VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -ENAQEKR VL -ENAQEKR VL -ENAQEKR VL -ENAQEKR	Str-3 EEEEEEEE. VARFTKLHGSV VIRLTKLHGSL MTI.VFYPHGSL ATI.VFYPHGSL ATI.VFYPHGSL GANIVKLHGGL GANIVKLHGGL GANIVKLHGSL RVNIWKVHGSL RVNIWKVHGSL RVNIWKVHGSL FGYYLHLHGSP PIRLYKLHGSSI VIRIIKYHGSP KLSVLHHGVY KKSVLHHGVY KKSVLHHGVY DKSILHHGCY	Heli HHHH 56 FRDLAA 56 FRDLAA 56 TVYREV 59 TNVYEEV 57 ISFCTQO 55 LNRGLRS 86 KLKLEED 57 FRRFSEF 48 YWDYLOF 77 FRRFSEF 48 YWDYLOF 48 VRFRSDA 24 EIQKLYE 24 EIQKLYE 24 EIQKLYE 24 EIQKLYE 23 FFKKKS	x-3 S HHHHHEE ICRPNSTUTY ACRPNSTUTY LTSHRSTUTY LSHGDTLVVY LPALGESLVVY LDALGESLVVY LAGWHGICLF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDFLIF NUSANSFLIF LEKEQSVLIAF LEKEQSULIF NSSAFLFLCC TFLFLCC TFLFLCC TFLFLCC	tr4 EEES GYG GWG GWG GWG GYG GYG GYG GYG GYG GYG	He E*.*	14 Str5 HHHHEEEEEEE L-TIPSTHLVIMA 1 GIQRVAVSVFGG GINRVAVSVFTGQ KFKTVAISIYPRS RFQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQGAPVTIITY L-NN-PNLVVFIFCY 1 L-NN-PSLHVITICY L-NN-PSLHVVEWSG FDYKNRSSFIIDK WSIWRRSGHEKDR VKHKSDLEHFML LYSVPNKVDLEHYML SLNL-KKYALMQLRD IVY-ENKOSGRAP
Secondary Structure XAC2443 Xca 21243176 b111926_Brja 27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530_Neu_30250449 Pflu2135_Pfl_23059957 PPl406_Ppu1_26988140 LA1794_Lepin_24214494 LA1852_Lepin_24214552 NE2094_Neu_30250034 seal0_Seen_38201747 EF2349_Efac_29376850 VYA0689_Vivul_37676349 VP1802_Vivul_37676349 VP1802_Vivul_37676349 VP1802_Vivul_37676349 VP1802_Vivul_37676349 VP1802_Vivul_37676349 VP1802_Vivul_37676349 VP1802_F1_264652 AK024756.T_Hs_10437125 FLJ20635_Hs_15489177 LJ1095_Laj0_42519021 RPA0819_Rhpa_39933896 CV1059_Chvi_34496514	H2 HHHHH LTEAGAELAGLHLLDRFLGNL LIEHGCDFAGLHLIDRFLGNL LVYWTMTYGLNYQDGHRFKD TLYWAITYGLNYQDGHRFKD LYWAILQHGLNYQDGHRFKD LLYWALQHQGEVIDDLFNGPD LLPWAYQHPGTHEVIKFPI LAEYACDQSRIHHYTGFTHGFFRQL LAEYACDQSRIHYTGFTHGFFRQL LAEYACDQSRIHYTGFTHGFFRQL LAEYACDQSRIHYTGFTHGFFRQL LAEYACDQSRIHYTGFTHGFFRQL LEYACDQSRIHYTGFTHGFFRQL LEYACDQSRIHYTGFTHGFFRQL LEYACDQGTHLYTGFSGIH LLYNSFIDNDLVDGYNGSL LIEIAYEFGKQLESLDLTDEKK LLELYAADQGKQLESLDLTDEKK LLEIAYARATFQSRPNILKDIN LFFEKAFKEIHRGGKIPVVVNDADMQTN	TP 19 GEPRYLEG NP 19 GEPRYMEG CF 16 GLYREGT FF 12 RPYGHAAG NAS 1 DLCESHTG FF 12 RPYGHAAG SS 5 TWDIGNTN DN 21 KAYFKNKF TP DELTCSR FR 4 PNEITSAR KR 17 DNYANELP SS 17 GEARFGHY ID 11 LERKYDNN SSE 17 RYTGRYNT NC - IASAPAG VL - EWAQEKR VL - EWAQEKR VL - EWAQEKR VL - ENARGHM IAL - ENIDKLA EV - QDALANK QD - IAFWEAA	Str-3 EEEEEEEE. VARFTKLHGSV VIRITKLHGSL MTIVFYPHGSL ATIVFYPHGSL GTRVHLHGGL GTRVHLHGGL GANIVKHGGL RVNIKKYHGSL RVNIKKYHGSL RVNIKKYHGSL RVNIKKHGSL FGYYLHLHGSP PIRLYKLHGSP PIRLYKLHGSP KLSVLHIHGYY KLSVLHIHGYY KLSVLHIHGYY KLSVLHIHGYY KLSVLHIHGYY KLSVLHIHGYY KLSVLHIHGYY KLSVLHIHGYY	Heli HHHH S FRDLAA S FRDLAA S FRDLAA S FRDLAA S FRDLAA S FRDLAA S FRDLAA S FRDLAA S FRDLAA S TVYREV S INVEV S INVE S INVE	x-3 S HHHHHEE ICRPNSTUTY ACRPNAVVYIY LTSHRSTUTY LKSQRDTLVYY LEHGDNLCTF FANLTDFLIF WMXSVDKLII AINAANSFLCI LEKEQSYLCI AINAANSFLCI LEKEQSYLLIA NLSEANELIIF NLSEANELIIF NLSEAFLFLCC NKSFLFLCC QLIKQNA-WLFI LMEDHCVLFA LATQTIFFI	tr4 EEEE GYG GWG GWG GWG GYG GYG GYG GYG GYG GYG	He E*.HHHHH FG D UNRVLFDM IGEH R LLKRMYGT IGEH R LLKRMYGT IGEH R LLKRMYGT IGEH R LLKRMYGT ILQRMNGT FAEQUI LLQRMNGT FAEQUI LLQRMGT FAEQUI LLQRMGT FOR CONTRACTOR FOR CONTRACTOR FSCONTRACTOR FSONT LNMLKAPY CKDKGINEMIKDN MSDATFQALFLEA STATPQIFQALFLEA STATPQIFQLFSD LSQEVINISYTY	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIS 1 L-TIPSTHLVIMA 1 GIQRVAVSVFGGS GINRVAVSVFTGG PPKRMAVSVFTGQ NEKTVAISIYPRS RPQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQGAPVTIITY CLRQNFPITIITY L-NN-PSLHVFTCY 1 L-NN-PSLHVFTCY 1 L-NN-PSLHVFTCY 1 L-NN-PSLHVFTCY 1 L-NN-PSLHVFTCY 1 L-NN-PSLHVFTCY 1 L-NN-PSLHVFTCY 1 L-NN-PSLHVFTCY 1 L-NN-PSLHVFTCY 1 LSNT-NKYALMQLRD VKHKSDLEHFML SLNL-KKYALMQLRD IYDY-ENHQSKPAP
Secondary Structure XAC2443 Xca 21243176 b111926_Brja 27377037 XAC2189_Xax_21242924 BRA0375_Brsu 23500128 NE2530_Neu_30250449 Pflu2135_Pfl_23059957 Pl406_Put_2698140 LA1794_Lepin_24214552 NE2094_Neu_30250034 seal0_Sean_38201747 EF2349_Efae_29376850 VVA0689_Vivul_37676349 VP1802_Vipa_28898576 PG1107_Porgi_34540840 Sir2_Smel_16264652 AK024756.1_Hs_10437125 FLJ20635_Hs_15489177 LJ1095_Lajo_42519021 RPA0819_Rhpa_3993896 CV1059_Chvi_34496514 TDE0266_Iden_2252782	H2 HHHHH LIEAGAELAGLHLLDRFLGNL LIEHGCDFAGLRIIDRFVGAL LVYWMTYGLNYQDGHRFKD VYWMAITYGLNYQDGHRFKD LLYWMIQHQEVIDDLFNGPD LLPWAVQHQEVIDDLFNGPD LLPWAVQHQEVIDDLFNGPD LLYWIMQDEITFTFTCDDGFRNPD LLYWIMQDEITFTTCDDGFRNPD LLYWIMQDEITFTTCDDGFNCH AEYACDQSRIHHYTGFTHGFFRQL LAEYACDQSRIHHYTGFTHGFFRQL LAEYACDQSRIHHYTGFTHGFFRQL LLYNSFIORDLVDGYNGSL LLYNSFIORDLVDGYNGSL LLYNSFIQGRQLESLDLTDEKK LLELYAADQGRQLESLDLTDEKK LLEAYARDQGRQLESLDLTDEKK LLEAYARDQGRQLESLDLTDEKK LLEAYARDQKNPMESLDLKDKTK QIEKAFKEIHAGGKIFVVVNDADMQTN FFEECKAYSIAYTKIATV	TP 19 GEPRYLEG TP 19 GEPRYMEG CF 16 RGL/REQT CF 16 RGL/REQT CF 14 QPIRGERS 15 TF 12 RPYGHAAG 15 TUCS 15 TWIGNTN DN 21 KAYFKNFF AT PDELTCSR FR 4 PNEITSAR KR 17 DNYANELP 10 11 LERKYDNN ISE 17 RYTGRYNT RD -IASAPAG VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -ENIDKLA EV -QDALAMKA SD -ISKITDG	Str-3 EEEEEEEE. VARFTKIHGSI VIRITKIHGSI NTIVFYPHGSI TSIVFYPHGSI KPRLLYIHGSI GIRVIHLHGGI SONIFYIHGAI GANIVKIHGGI RVNIWKVHGSI TINIVKCHGSS 1 NMYLYRIHGSI TINIVKCHGSS PIRLYKIHGSI VTRIKXHGDF KLSVLIHHGVY KLSVLIHHGVY KLSVLIHHGVY KLSVLIHHGVY KLSVLIHHGVY KLSVLIHHGVY KLSVLIHHGVY KLSVLIHHGVY VKIKIHGSE VPCIVKCHGSE	Heli HHHH 58 FRDLAA 55 FRDLAA 56 ATVYREV 56 STVYREV 57 ISVCIGO 53 LSWCIGO 53 LSWCIGO 55 LNRCIRS 55 LNRCIRS 55 LNRCIRS 56 KLKLEPED 45 TIHNADL 19 LFKKFRDA 45 TIHNADL 19 LFKKFRDA 24 EIQKIYE 24 EIQKIYE 24 EIQKIYE 24 EIQKIYE 23 AKKKIET 23 FFKKKS 24 LGKKKET	x-3 S HHHHH5EE ICRPNSTUTY ACRPNSTUTY LTSHRSTUTY LTSHRSTUTY LEALGSIVY LEALGSIVY LAQENGSICI FANLTDFLIF VMXSVDFLIF VMXSVNSFICI LEKPQTUFV AINAANSFICI LEKPQTUFV MISSASILIF NISKASILIF NISKAKSLIF NISSFLFUC C NKSFLFUC UIKONAKSVLFI MEDCVLFA UIKONCVLFA	tr4 EEEE GYG GWG GWG GWG GYG GYG GYG GYG GYG GYG	He E*.HHHHH FGIE INRVIRDM IGEH R LIKRMGT FAEQII LLRAMGT FAEQII LLQRMGT FAEQII LLQRMGT FAEQII LLQRMGT FAEQIS LVRALRLA LDSS NILLRAMEG FGIQ INFRINHQ FNE VQPKLMAK FLIE ITDIVQRS FGIY INRILGA GFIT LNMLKPY CKDKGINEMIKDN MSDLNIRLLHRI VDDTTQALFLEA ETLHQIFQALFLEA ETLHQIFQALFLEA LSDYPRSIYTFV LSDYPRSIYTFV	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIS 1 L-TIPSTHLVIMA 1 GIQRVAVSVFGGS GINRVAVSVFRGD PFKRMAVSVFTGQ KPKTVAISIYPRS RPQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQGAPVTIITY CIRQNFITIITY L-NN-PNLLVFIFCY 1 L-NN-PNLLVFIFCY 1 L-NN-PSLHIVICY LKVTPLVVEWSG FDYKNKPSFIIDK WSIWRSGLEHFML LYSVPNKVDLEHYML SLNL-KKYALMQLRD IVDY-ENHQSGRPAP RKQM-RGLHKQAYIV
Secondary Structure XAC2443_Xca_21243176 bl11926_Brja_27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530_Neu_30250449 Pflu2135_Pfl_2305957 PP1406_Pput_26988140 LA1852_Lepin_24214592 NE2094_Neu_30250034 seal0_Seen_38201747 EF2349_Efae_29376850 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VP1802_Vipa_28898576 PG1107_Porgi_34540840 Sir2_Smel_16264652 AK024756.1_Hs_10437125 FLJ2103_Hs_13375721 FLJ20635_Hs_15489177 LJ1095_Lajo_42519021 RPA0819_Rhpa_3993896 CV1059_Chvi_34496514 TDE0266_Tden_42525782 MBNC211202_Mesp_45916072	H2 HHHHH LIEAGAELAGLHLLDRFLGNL LIEHGCDFGLRIUDRHAFKD VVYWMITYGLNUQDGHRFKD VVYWAITYGLNUQDGHRFKD VVYWAITYGLNUQDGHRFKD LIYWMLQHQGEVILDDENGFD LLYWIMQDEITFTFTCDDGFRNPD YLELLCIDYDIPATYGDTEVIKFPI LAEYACDQSGIHIYTGFTHGFFNQL LAEYACDQSGINLINGFSIH FLEIAVEK-NIESNPKIFFNDGTNGYM AIEWSAEESGINLINGFSIH LIYNSFLSGINLINGFSIH LLYNSFLSGINLINGFSIH LLYNSFLSGINLINGFSIH LLYNSFLGKQLESLDITDEKK LLELYADQGKQLESLDITDEKK LLELYADQGKQLESLDITDEKK LLELYADQKQLESLDITDEKK LLELAFGRRQNKPMSSLDLKDTK FFEKAFKEIHRGGKIPVVVNADMQTN FFEECCKAYSIAYTKIATV	TP 19 GEPRYLEG TP 19 GEPRYLEG CF 16 RGLYREQT CF 16 RGLYREQT CF 16 RGLYREQT FH 22 RPYGHAAG AS 1 DLCESHTG DYRTGSES ST UDUGNTN DN 21 KAYFKNKF AT PDELTSAR KR 17 DNYANELP SR 17 GEARFGHY TG EARFGHY LD 1 LERKYDNN SE 17 RYTGRYNT TRD -IASAPAG UL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR SD -IAFWEAA SD -IAFWEAA SD -IAFWITCH RD -IAFWEAA SD -ISRITDG RD -ISRIADG	Str-3 EEEEEEEE. VARFTKIHGSV VARFTKIHGSV MTLVFYPHGNI TSLVFYPHGSI MTLVFYPHGSI GIRVIHGGI GIRVIHGGI GIRVIHGGI GANIVALHGGI GANIVALHGGI RVNIMAVHGSI TINLVACHGSV 1 NMYLYKIHGSI VIRIIAYHGSI VIRIIAYHGSI VIRIIAYHGSI DESLIMHGYY KLSVIHHGVY KLSVIHHGVY KLSVIHHGVY KLSVIHHGVY KLSVIHHGVY KLSVIHHGVY MTGIACHGEF VIQIIXYHGDF	Hell HHHH 58 FRDLÄAA 55 FRDTAA 56 ATVYREV 57 INVYREV 57 INVYREV 57 INVYREV 57 INVKEV 53 LSKCIQ 55 LNKGLK 55 LNKGLK 55 INNADI 19 LRFLSYE 57 FRKTSES 76 KLKLFED 19 LRFLSYE 57 FRKTSES 44 YUDYLQF 71 LFKKFRN 24 EIQKLYE 24 EIQKLYE 24 EIQKLYE 23 AKKKET 23 FFKLKS 24 LGSKLKD 24 JCKFRA	x-3 S HHHHHHEEE ICRPNSTLJTY ICRPNSTLJTP LKSQRDTLVVY LTSHRSTLTP KSQRDNLCIP LACENBSLVY LLEHGDNLCIP MKSVDKLIII ANNANSTLII LEKEQSVLIAP NLSKASVLII LEKEQSVLIAP NLSKANSLIII LENTLFVC NKSFLFLCC NKSFLFLCC QLIKQNA-WLFI LMEDHCVLPA LACEN-FVLFI DALGRTILFI	tr4 EEEE GYG GYG GWG GWG GYG GYG GYG GYG GYG GYG	He E*.*	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIS 1 JTIPSTHLVIMA 1 GIQRVAVSVFGGS GINRVAVSVFRGD PFKRMAVSVFTGQ KPKTVAISIYPRS RPQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CCRQGAPVTIITY C.RQNFTTIITY L-NN-PNLLVFIFCY 1 L-NN-PNLLVFIFCY 1 L-NN-PSLHIVICY LKVTPLAVVEWSG FDYKNKPSFIIDK WSIWRSGHEKDR VHKSDLEHFML LSVPNKVDLEHYML SLNL-KKXHAMOLRD IYDY-ENHQSGKPAP RRQM-RGLHKQAYIV KKWRDSG
Secondary Structure XAC2443 Xca 21243176 bl11926 Brja 27377037 XAC2189_Xax 21242924 BRA0375 Brsu 23500128 NE2530 Neu 30250449 Pflu2135 Pfl 23059957 PP1406 Pput 26988140 LA1794 Lepin 24214494 LA1852 Lepin 24214494 LA1852 Lepin 24214522 NE2094 Neu 30250034 seal0 Seen 38201747 EF2349 Efae 29376850 VVA0668 Vivul 37676349 VP1802 Vipa 2898576 PG1107 Prorj 34540840 Sir2 Smel 16264652 AK024756.1 Hs 10437125 FLJ2103 Hs 1337571 FLJ20635 Hs 15489177 LJ1095 Lajo 42519021 RPA0819 Rhpa 3993896 CV1059 Chvi 34496514 TDE0266 Tden 42525782 MBNC21120 Mesp 45916072 usg Pepe 11322457	H2 HHHHH LIEAGAELAGLHLLDRFLGAL VYWMTTGLNVQDGHRFKD UYWMITGLNVQDGHRFKD TLYWMLQHQGVIDDHFNFD LLYWALQHQGVIDDLFNFD LLYWYIMQDEITPTFTCDDGFRNPD YLELLCIDYDIPATYGDTEVIKFPI LLYWTMQSRIHYTGFTHGFRQL LAEYXCDQSGTHLYTGFTGFRQT LAEYXCDQSGTHLYTGFSGTH LLYNSFISGINLINGFSGTH LLYNSFISGINLINGFSGTH LLYNSFIYGRFYKXISHV LLELYAADQGKQLESLDLTDEKK LLELYAADQGKQLESLDLTDEKK LLELYAADQGKQLESLDLTDEKK LLELYAADQKRYAKISHV LLELYAADQKRYAKISHV LLELYAADQKRVLSLUTDEKK LLEAFGRRATFQSPRILKDIN FFEKEKFHGGKLYVVVNDADMQTN FFEECKATFQSPRVILKDIN LLEVAFD	TP 19 GEPRYLEG NP 19 GEPRYMEG CF 16 RGLVREGT CF 14 QPIRGERS FH 12 RPYGHAAG MAS 1 DLCESHTG FF DVRRTGSE SD 5 TWDIGNTN DN 21 KAYFKNNF AT PDELTCSR FR 4 PNEITSAR KR 17 DNYANELP ST 7 GEARFGHY ID 11 LERKYDNN SE 17 RYTGRYNT TR - IASAFAG VL - EWAQEKR VL - EWAQEKR VL - EWAQEKR VL - EWAQEKR VL - ENAQEKR SD - IASHTDG D - IASHTAG SD - ISKITDG D - ISKITDG CR - ISKITDG	Str-3 EEEEEEEE. VARFTKLHGSV VIRLTKLHGSL MTI.VFYPHGSL ATI.VFYPHGSL ATI.VFYPHGSL GANIVKLHGGL GANIVKLHGGL GANIVKLHGGL RVNIWKVHGSL RVNIWKVHGSL RVNIWKVHGSL RVNIWKVHGSL FGYLHLHGSP PIRLYKLHGSI VIRIIKYHGDF KLSVLHIHGVY KLSVLHIHGVY KLSVLHIHGVY KKSVLHHGVY KKSVLHHGY KRVLKIHGSE VFQIVCHGDF DAILWXHGDF DAILWXHGDF	Heli HHHH Solver Heli Heli Heli Heli Heli Heli Heli Heli	x-3 S HHHHHEE ICRPNSTUTY AGRPNSTUTY LTSHRSTUTY LAGPNDTUVY LPALGESLVVY LAGGDUVY LAGYDUVY LAGYSTUTY AINASTUTY AINASTUTY AINASTUTY AINASTUTY NASEAEIIIF NLSKANSFLIY LEEATVLFI NKSFLFLCC TKSFLY QIKONA-HVLFI LKDKSFLY QIKONA-HVLFI LKDCVFA ILATQTIFF DLGKSVFI DLGKSFLY	tr4 EEEE GYG GYG GWG GWG GYG GYG GYG GYG GYG GYG	He E*.*	14 Str5 HHHHEEEEEEE L-TIPSTHLVIMA 1 GIQRVAVSVFGG GINRVAVSVFRGD PPKRMAVSVFTGQ KFKTVAISIYPRS RPQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQGAPVTIITY CIRQNTPITITY CIRQNTPITITY L-NN-PNLVVFIFCY 1 L-NN-PSLHIVICY LKVTPLRVVEWSG FDYKNRSGHEKDR WSIWRSGHEKDR WSIWRSGHEKDR LYSVPNKVDLEHFML LSVPNKVDLEHFML SLNL-KKYALMQLRD IDY-ENHQSGKPAP RKQM-RGLHKQAYIV KKWRANV WGTWRANV
Secondary Structure XAC2443 Xca 21243176 b111926_Brja 27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530_Neu_30250449 Pflu2135_Pfl_23059957 Ppl406_Pptl_26988140 LA1794_Lepin_24214454 LA1852_Lepin_24214552 NE2094_Neu_302250034 seal0_Seen_38201747 EF2349_Efae_29376850 VYA0689_Vivul_37676349 VYA0689_Vivul_37676349 VP1002_Vipa_28898576 PG1107_Porgi_34540840 Sir2_Smel_16264652 AK024756.1_Hs_10437125 FLJ21035_Hs_15489177 LJ1095_Lajo_42519021 RPA0819_Rhpa_39933896 CV1059_Chvi_34496514 TDE0266_Tden_42525782 MBNC211202_Mesp_45916072 usg_Pepe_11322457 Aple047201_Acpl_32033900 BP0483_Bcha_33594473	H2 HHHH LTEAGAELAGLHLLDRFLGNL LIEHGCDFGLNLJDRFLGNL LVYWMITYGLNLJDRHAFKD TLYWAIUGLNUDGHRFKD LYWMIUGLNUDGHRFKD LLPWAVGH	TP 19 GEPRYLEG NP 19 GEPRYMEG NCF 14 QPIRGERS FH 12 RPYGHAAG NAS 1 DLCESHTG SUS 5 TWDIGNTN DN 21 KAYFKNFF TP DELTCSR FR FR PNEITSAR IL SIS 17 RYEARFGRY JID 11 LERKYDNN SIS 17 RYEARFGRY JID 11 LERKYDNN SIS 17 RYEARFGRY JU 11 LERKYDNN SIS 17 RYEARFGRY VL -EWAQEKR VL VL -EWAQEKR SD VL -EWAQEKR SD SIS -ISRIDG SR SIG -ISRIDG SR GL -ISRIAG SU SIG -ISRIAG SU SIG -STQKYDR SU	Str-3 EEEEEEEE. VARFTKLHGSV VIRITKLHGSV VIRITKLHGSL MTIVFYPHGSL ATIVFYPHGSL GTRVHLHGGL GTRVHLHGGL GANIVKHGGL RVNIWKVHGSL RVNIWKVHGSL RVNIWKVHGSL RVNIWKVHGSL RVNIWKVHGSL FGYYLHLHGSP PIRLYKLHGSP PIRLYKLHGSP KLSVLHIHGVY KLSVLHIHGVY KLSVLHIHGVY KLSVLHHGVY KLSVLHHGVY KLSVLHHGVY DFVIRJHGDF DAILYKNHGDF DAILYKNHGDY DAVYKNHGDY	Heli HHHH S8 FRDLAA 55 FRDLAA 56 ATVYREV 55 SRDYREV 57 INVYEV 57 INVYEV 57 INFC 53 LSWCLQ 55 INRCLS 54 INNADI 45 IINNADI 45 IINNADI 45 IINNADI 45 IINNADI 45 IINNADI 45 IINNADI 45 IINNADI 44 VRFSDA 44 VLQKIYE 24 VLQKIYE 23 AKKKIET 23 FFKIKS 24 IIQKIYE 24 ILQKIYE 24 ILQKIYE 24 ILQKIYE 24 ILQKIYE 24 ILQKIYE 24 ILQKIYE 24 ILQKIYE 25 ILQKIYE 25 ILQKIYE 25 ILQKIYE 25 ILQKIYE 26 ILTARG 27 FFTALG 27 F	x-3 S HHHHHHEE ICRPNSTUTY ACRPNAVVVIY LTSHRSTUTY LKSQRDTLVYY LEHGDNLCTF FANLTDPLIF VMKSVDVLII IAQENHGICLP FANLTDPLIF UMSSANSFLCI LEKEQSVLIAF LEKEQSVLIAF NLSKANSLII LKEQFLFLCC NKSFLFLCC NKSFLFLCC QLIKQNA-HVLFI LMEDHCVLFA LMEDCVLFA IMEAVLFF DIGRTLFF DLISFLFLCC DISFLFLCC DISFLFLCC	tr4 EEEE GYS GWG GWG GWS GYS GYS GYS GYS GYS GYS GYS GYS GYS GY	He E*.HHHHH FGE INRVIRDM IGEH R LLKRMYGT IGEH R LLKRMYGT IGEH R LLKRMYGT IGEH R LLKRMYGT IGEN R LLQRMGT FAEQUI LLQRMGT FAEQUI LLQRMGT FAEQUI LLQRMGT FAEQUI LLQRMGT FOR CONTRACTOR FOR CONTRACTOR FNE LQPNIAN FNE IQPNIAN FNE IQPNIAN FNE IQPNIAN FUE ITDIVORS FGY INRIILGA GFIT LNMLLKPY CKDGTFQALFLEA DDTTFQALFLEA DDTTFQALFLEA SUNTRILLHRI FSDWNTRLLHRI FSDWNTRLLHRI FDPNINTYILSRI FDPNINTYILSRI FDPNINTYILSRI	14 Stt5 HHHHEEEEEEE L-TIPSTHLVVIS 1 L-TIPSTHLVIMA 1 GIQRVAVSVFGGS GINRVAVSVFGGS GINRVAVSVFTGQ PFKRMAVSVFTGQ KFKTLAISIYPRS RPQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQ-GAPVTIITY L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLHIVICY LKVTPLRVVEWSG FDYKSDLEHFML VKHKSDLEHFML SLNL-KKYALMQLRD INJV-ENNGSGENKDR NKMSDLEHFML SLNL-KKYALMQLRD INJV-ENNGSGENKPA RKQM-RGLHKQAYIV KKWRDSG RNIL-DENISNNYWL BIOF2-ADOPONYCI
Secondary Structure XAC2443 Xca 21243176 b111926_Brja_27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530_Neu_30250449 Pflu2135_Pfl_23059957 PPl406_Pput_2698140 LA1794_Lepin_24214552 NE2094_Neu_30250034 seal0_Seen_38201747 EF2349_Efae_29376850 VYA0685_Vivul_37676349 VP1802_Vipa_28898576 PG1107_Dorgi_34540840 Sir2_Smel_16264652 AK024756.I_Hs_10437125 FLJ2063_Hs_15489177 LJ1095_Laj0_42519021 RPA0819_Rhpa_3933896 CV1059_Chvi_34496514 TDE0266_Tden_42525782 MBNC211202_Mesp_45916072 usg_Pepe_1322457 Aple047201_Acpl_32033900 BB0483_Bobr_3359473 pX02-73 Ban_10956463	H2 HHHHH LIEAGAELAGLHLLDRFLGNL LIEHGCDFAGLHIDRFVGAL LVYWMITYGLNVQDGHRFKD TLYWAIUHGLNVQDGHRFKD LLYWALQHQGEVIDDLFNGPD LLPWAVQHQGEVIDDLFNGPD LLPWAVQHQGFAELFDEQG LLYWIMQDEITFTFTCDDGFRNPD LLELCIDYDIPATYGFTBGFTRQL LAEYACDQGITHYTGFSHGF FLEIAVEK-NIESNPKIFFNDGTNGYM AIEWSAEESGINLINGFSGIH LLYNSFIDNDLVDGYNGSL LLESFNQTSNINGNISDGFDEYG LLEIAVEK-NIESNNINGNISDGFDEYG LLEIAVERQGRQLESLDLTDEKK LLEJAADQGRQLESLDLTDEKK LLEJAADQGRQLESLDLTDEKK LLEAFGRATFQSRPNILKDIN FFEEECKATFQVDDADMQTN LLEGAFELASINATKIATV NLEVAFELASINATKIATV NLEVAFELAGKVVDVKHNVE LLESSLERNGKVPDVYIGOK	TP 19 GEPRYLEG TP 19 GEPRYLEG CF 16 RGLVREGT CF 14 QPIRGEN FH 12 RPYCHAAG SS 5 TWDIGNTN DN 21 KAYFKNFF AT PDELTCSR FF 4 PNEITSAR KR 17 DNYANELP ISR 17 GEARFGH TG EARFGH TG EARFGH CL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR SD -ISKITDG ISKITDG CR -ISKITDG CR -ISKITDG CR -STGKYDR CQ - ATTRSKR CQ - ATTRSKR CQ - ATTRSKR	Str-3 EEEEEEEE VARFTKIHGSI VARFTKIHGSI MTIVFYPHGSI MTIVFYPHGSI GTRVIHLGGL GTRVIHLGGL GONIFYIHGAL GANIVKUHGSI RVNIMKVHGSI RVNIMKVHGSI TINIVKCHGSV I NMYLYKIHGSI VTRIKXHGDF KISVLHHGVY KISVLHHGVY KISVLHHGVY KISVLHHGVY VTRIKXHGDF DZVIKTHGP DZVIKHGDY DAVVYKHGDY DAVVYKHGDY DAVVYKHGDS	Heli HHHH 58 FRDLAA 55 FRDLAA 56 ATVYREV 57 INVYREV 57 INVYREV 57 INVYREV 57 INVYREV 57 INVYREV 57 INVYREV 57 INVYREV 57 INVYREV 57 FRRFSEF 48 YWDTOF 71 LFKKRRN 48 YWDTOF 71 LFKKRN 44 EIQKLYE 44 EIQKLYE 44 EIQKLYE 44 EIQKLYE 44 VLQNLYR 23 AKKKIET 23 FFKIKS 44 ULQNLYR 24 FIGKLKET 24 FIGKLYE 24 FIGKLYE 24 FIGKLYE 25 FFALRG 22 FFTALRG 25 FITALSG 22 LAKKLT	x-3 S HHHHHHEE ICRPNSTUTY ACRPNSTUTY LTSHRSTUTY LTSHRSTUTY LTSHRSTUTY LEHGDNLCTF FANLTDFLIF VMXSVDFLIF VMXSVDFLIF VMXSVNSFICI LEKQQSVLIF AINAASSFICI LEKQQSVLIF AINAASSFICI LEKQSULIF NISKASILIF NISKASFLFU CLIKQNA-HVLFI LECCVIFA IMEACVIFA IMEACVIFA DLIGFLFUCC DLISFLFUCT DLISFLFUCT DLISFLFUCT DLISFLFUCT	tr4 EEES GYS GWS GWS GWS GYS GYS GYS GYS GYS GYS GYS GYS GYS GY	He E*.*HHHHH FGE INRVIRDM FGE INRVIRDM IGEH R.LIKRMGT FAEQI ILQRMGT FAEQI ILQRMGT FAEQI ILQRMGT FAEQI ILLRMEKG FGQ INFRINHQ FNE VQPKLMAK FLE ITDIVQRS FGY INRIIGA GFT INMLKY CKDKGNEMINDN MSDINRLLHRI VDDTTQALFLEA VDDTTQALFLEA VDDTTQALFLEA LSQUVRYLHKI L-SQUVRYLHKI L-SDMNRLLHRI FDPNIVYLLSRI FDPNIVILSRI FDPNIVILSRI FDPNIVILSRI	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIS 1 L-TIPSTHLVIMA 1 GIQRVAVSVFGGS GINRVAVSVFRGD PFKRMAVSVFTGC KPKTVAISIYPRS RPQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQGAPVTIITY CIRQNFITIITY L-NN-PNLLVFIFCY 1 L-NN-PSLHIVICY LKVTPLHVVEWSG FDYKNKPSFIIDK WSIWRSGHEKDR VHKKSDLEHFML LYSVPNKVDLEHYML SLNL-KKYALMQLRD IVDY-ENHQSGKPAP RKQM-RGLHKQAYIV KKWRANV WQTWRANV MQTWRANV RASY-SENQKEHYCI RASY-SENQKEHYCI RICP-AQDQRQHYCI
Secondary Structure XAC2443_Xca_21243176 bl11926_Brja_27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530_Neu_30250449 Pflu2135_Pfl_2305957 PP1406_Pput_26988140 LA1852_Lepin_24214542 NE2094_Neu_30250034 seal0_Seen_38201747 EF2349_Efae_29376850 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VP1802_Vipa_28898576 PG1107_Porgi_34540840 Sir2_Smel_16264652 AK024756.1_Hs_10437125 FLJ2103_Hs_13375721 FLJ20635_Hs_15489177 LJ1095_Lajo_42519021 RPA0819_Rhpa_39933896 CV1059_Chvi_34496514 TD50266_Tden_42525782 MBNC211202_Mesp_45916072 usg_Pepe_11322457 Ap1e047201_Acp1_32033900 BB0483_Bobr_3359473 pX02-73_Ban_10956463 BC1912_Bca_30020052	H2 HHHHH LIEAGAELAGLHLLDRFLGNL LIEHGCDFAGLRIIDRFVGAL LVYWMTYGLNYQDGHRFKD VVYWAITYGLNYQDGHRFKD LYWMMLQHQGEVIDDLPNGPD LLPWAVQHQGEVIDDLFNGPD LLPWAVQHQGEVIDDLFNGPD LLPWAVQHQGEVIDULFNGPD LLEUCIDYDIPATYGDFEVIKFPI LAEYACDQSRIHYTGFTHGFFRQL LAEYACDQSRIHYTGFTHGFFRQL LAEYACDQSRIHYTGFTHGFFRQL LYWSAEESGINLINGFSGIH LLYNSFIQGKQLESLDITDEKK LLEJYAADQGKQLESLDITDEKK LLEJYAADQGKQLESLDITDEKK LLEAFGRRNFFQSRPNILKDIN FFEKAFKEIHRGGKIPVVVNDADMQTN FFEECKYGRDFVKVANA LLEQGFKNNGKVPDVKTNK LLEQGFKNNGKVPDVKTNK LLEQIFEEQEMQVYIGK	TP 19 GEPRYLEG TP 19 GEPRYLEG CF 16 RGLYREQT CF 16 RGLYREQT CF 16 QPIRGERS NF DVRRTGSE SS 5 WDIGNTN DN 21 KAYFKNKF AT PDELTCSR FR 4 PNELTSAR KR 17 DNYANELP SS 17 GEARFGHY TG -IASAFAG UL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR SD -ISKITDG RD -ISKITDG RD -ISKITDG RD -ISKITDG RD -ISKITDG RL -STQKYDR RL -STQKYDR RL -LFSHPLE EL -LFSQPFE	Str-3 EEEEEEE. VARFTKIHGSI VARFKIHGSI MTIVFYPHGNI TSIVFYPHGSI GIRVILHGGI GIRVILHGGI GIRVILHGGI GANIVKIHGGI GANIVKIHGGI TINIVKCHGSI TINIVKCHGSI TINIVKCHGSI TINIVKCHGSI VIRIIKHGSI VIRIIKHGSI VIRIIKHGCS DAVYKHGDY DAVYXMHGDY DAVYXMHGDY DAVYXMHGDY MELYXIHGCS	Heli HHHH 58 FRDLÄAA 56 ATVYREV 55 FRDFÄAA 56 ATVYREV 57 INVYEEV 57 INVYEEV 57 INVKEV 53 LSKCIQ 55 LSKCIQ 55 LSKCIQ 55 LSKCIQ 55 LSKCIQ 57 INVÄD 57 INVÄD 57 FRRSEF 77 ILFKKFRN 24 EIQKIYE 24 EIQKIYE 24 EIQKIYE 24 EIQKIYE 24 EIQKIYE 23 AKKKET 23 FFTKIS 22 LDVKFRA 22 FTTALSG 22 FTTALSG 22 LAAKLIT 24 LAAKLIT 24 LAAKLIT 25 LAAKLIT	x-3 S HHHHHHEE ICRPNSTLVTY ACRPNSTLVTY LTSHRSTLTP LKSQRDTLVVY LEHGSLVY LEHGDNLCF FANLTDFLIF VMXSVDKLII AINAANSFLCI LEKEQSVLIAF LKEQSVLIAF LKEQSVLIAF LKEQSVLIAF LKEQSVLIAF LKEQSVLIAF LKEQSVLIAF LKEQSVLIAF LESSANSLII LECFFLCC NKSFLFLCC NKSFLFLCC NKSFLFLCC NKSFLFLCC DLIGKFLFL DLISKFFLFL DLIAKFFLFL DLIAKFFLFL DLIAKFFLFL	tr4 EEEE GYS GWG GWG GWG GYS GYS GYS GYS GYS GYS GYS GYS GYS GY	He E*.*	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIXS 1 L-TIPSTHLVIXA 1 GIQRVAVSVFGGS GINRVAVSVFFGG PFKRMAVSVFTGQ KPKTVAISIYPRS RPQQLSTAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQGAPVTIITY CCRQGAPVTIITY L-NN-PNLLVFIFCY 1 L-NN-PNLLVFIFCY 1 L-NN-PNLLVFIFCY 1 L-NN-PSLHIVICY LKTPLEVVEWSG FDYKNKPSFIIDK WSIWRSGHEKDR VHKSDLEHFML LSVDNKVDLEHYML SLNL-KKXALMQLRD IYDY-ENHQSGKPAP RRQM-RGLHKQAYIV KKWRDSG RNIL-DENISNNYWL RASY-SEQKEHYCI RIQF-AQDQRQHYCI TACL-DQDNIHKLKD TACL-DQONIHKLKD
Secondary Structure XAC2443 Xca 21243176 bl11926 Brja 27377037 XAC2189_Xax 21242924 BRA0375 Brsu 23500128 NE2530 Neu 30250449 Pflu2135 Pfl 23059957 PP1406 Pput 26988140 LA1794 Lepin 24214494 LA1852 Lepin 24214592 NE2094 Neu 30250034 seal0 Seen 38201747 EF2349 Efae 29376850 VVA0668_Vivul 37676349 VP1802 Vipa 2898576 PG1107 Porgi 34540840 Sir2 Smel 16264652 AK024756.1 Hs 10437125 FLJ2103 Hs 1337521 FLJ20635 Hs 15489177 LJ1095 Lajo 42519021 RPA0819 Rhpa 3993896 CV1059 Chvi 34496514 TDE0266 Tden 42525782 MBNC21120 Mesp 45916072 usg Pepe 11322457 Aple047201 Acpl 32033900 BB0483 Bobr 33599473 pX02-73 Ban 10956463 BC1912 Bce 3002052 BC1271 Bce 30019423	H2 HHHHH LIEAGAELAGLHILDRFLGAL LYYWTMTYGLRYQDGHRFKD UYWWMITYGLNYQDGHRFKD TLYWAMLLFNANGSWFKDA LNYWALQHQGEVIDDLFNGPD LLPWAVQHAPGFAELFDEQG LLYWTIMQDEITPTFTCDDGFRNPD YLELLCIDYDIPATYGGTHCFFRQL LAEYXCDQSGTHLYTGFTHGFFRQL LAEYXCDQSGTHLYTGFSGTH LLYNSFIYGRHYTGFSGTH LLYNSFIYGRUISDGFDEYG MLEIAVEK-NIESNENIFYNGTNGYM LLEIAVEKQGKQLESLDLTDEKK LLEYAADQGKQLESLDLTDEKK LLEYAADQGKQLESLDLTDEKK LLEYAADQGKQLESLDLTDEKK LLEYAADAFFQSRPNILKDIN FFEEECKATFQSRPNILKDIN FFEEECK	TP 19 GEPRYLEG NP 19 GEPRYLEG CF 16 GL/RECT CF 14 QPIGERS FH 12 RPYGHAAG NAS 1 DLCESHTO ST DURATGSE SD 5 TWDIGNTN DN 21 KAYFKNFF TP DELTCSR FR 4 PNEITSAR KR 17 DNYANLLP TR 17 GEARFGHY ID 11 LERKYDNN ID 11 LERKYDNN SE 17 RYTGRYNT TR - IASAPAG (VL - EWAQEKR VL - EWAQEKR VL - EWAQEKR VL - ENARGHM AL - SINIDG CR - ISKIDG RL - STQKYDR QL - ISKIDG RL - STQKYDR QL - FYSIHKR LE - LFSHPLE EL - LFSHPLE SL - LSTNLGA	Str-3 EEEEEEEE. VARFTKLHGSV VIRLTKLHGSL MTI.VFYPHGSL ATLVFYPHGSL ATLVFYPHGSL GANIVKLHGGL GANIVKLHGGL GANIVKLHGGL SONIFYLHGAL GANIVKLHGSL RVNIWKVHGSL RVNIWKVHGSL RVNIWKVHGSL FGYLHLHGSP FGYLHLHGSP FGYLHLHGSP KLSVLHLHGVY KLSVLHLHGVY KLSVLHLHGVY KLSVLHLHGVY KLSVLHLHGVY KLSVLHLHGV KRVKLHGSE VPQIVKHGDP DAVIXCHGDV DAVVXCHGDV DAVVXCHGDV INELYKIHGCS VNELYKIHGCS	Heli HHHH SFRDLAA FRDLAA FRDLAA SFRDLA	x-3 S HHHHHEEU ICRPNSTLVIY AGCPNAVVVIY LTSHRSTLVIY LAGCPNDTLVYY LEHGDNLCIF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDFLIF MKSNSFLCI LEKEQSVLAT AINAANSFLCI LEKEQSVLAT AINAANSFLCI LEKEQSVLAT NKSFLFLC NKSFLFLC NKSFLFLC QIKQNA-HVLFL UKDCVFA ILATQTIFF DLISKFLFJ DLISKFLFJ DLISKFLFJ DLISKFLFJ DLISSTFLFI DLISFFLFI DLISFFLFI DLISFFLFI DLISFFLFI DLISFFLFI DLISFFLFI DLISFFLFI DLIAFFLFI DLISFFLFI DLIA	tr4 EEEE GYS GWG GWG GWG GYS GYS GYS GYS GYS GYS GYS GYS GYS GY	He E*.*	14 Str5 HHHHEEEEEEE L-TIPSTHLVVNS 1 L-TIPSTHLVVNS 1 GIQRVAVSVFGGS GINRVAVSVFGGS GINRVAVSVFGG PFKRMAVSVFTGQ KFKTVAISIYPRS RPQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQGAPVTIITY L-NN-PSLIVFIFCY 1 L-NN-PSLIVVFIFCY 1 L-NN-PSLIVVFIFCY 1 L-NN-PSLIVVFIFCY 1 L-NN-PSLIVVFIFCY 1 L-NN-PSLIVVFIFCY 1 L-NN-PSLIVVFIFCY 1 LSVPNKVDLEHYML SLNL-KKYALMQLRD IVYY-ENHQSKPAP RRQM-RGLHKQAYIV KKWRDSG RNIL-DENISNNYWL RASY-SENQKEHYCI RCQF-AQQQCHYCI TRCL-DQCNIHKLKD TACL-DQCNIKLKN
Secondary Structure XAC2443 Xca 21243176 b111926_Brja_27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530_Neu_30250449 Pflu2135_Pfl_23059957 PP1406_Pput_26988140 LA1794_Lepin_24214592 NE2094_Neu_30225034 seal0_Seen_38201747 EF2349_Efae_29376850 VVA0689_Vivul_37676349 VP1802_Vipa_2898576 PG1107_Porgi_34540840 Sir2_Smel_16264652 AK024756.1_Hs_10437125 FLJ21635_Hs_15489177 LJ1095_Laj0_42519021 RPA0819_Rhpa_3993896 CV1059_Chvi_34496514 TDE0266_Tden_42525782 MBNC211202_Mesp_45916072 usg_Pepe_11322457 Aple047201_Acpl_32033900 BB0483_BbcF_3359473 pX02-73_Ban_10956463 BC1912_Ece_30019423 BK5-rp30_bK5-t_14251154	H2 HHHHH LTEAGAELAGLHLLDRFLGNL LTEHGCDFGLNYQDGHRFKD LYWMTTYGLNYQDGHRFKD YWYMAIUGLNYQDGHRFKD LYWMLUGLNYQDGHRFKD LYWMIQHQGEVIDDLFNGPD LLPNAYQHPGFAELFDEQG LLYWTIMQDEITPTFTCDDGFRNPD YLELLCIDYDIPATYGGTBUFFRQL LAEYACDQEGTHHYTGFSHGF FLEIAVER-NLESNFKIFFNDGTNGTM LEYASDQBGTHHYTGFSGIH LLYNSFIDNDLVDGYNGSL LIESNQTSINLGNISGFDEYG LLEIAVER-NLESNFKIFFNDGTNGTM LLEIAVER-NLESNFKIFFNDGTNGTM LLEIAVER-NLESNFKIFFNDGTNGTM LLEIAVERONDLVDGYNGSL LLEIAVER	TP 19 GEPRYLEG TP 19 GEPRYLEG CF 16 GGLYREGT FF 12 RPYGHAAG MAS 1 DLCESHTG FF 12 RPYGHAAG MAS 1 DLCESHTG FF 12 RPYGHAAG MAS 1 DLCESHTG FF 4 PNEITSAR FF 4 PNEITSAR FF 4 PNEITSAR FF 4 PNEITSAR FF 7 GEARFGHY ID 11 LERKYDNN ISE 17 RYTGRYNT MASAPAG FF 4 PNEITSAR FF 4 PNEITSAR	Str-3 EEEEEEEE. VARFTKLHGSV VIRITKLHGSL MTIVFYPHGSL ATIVFYPHGSL ATIVFYPHGSL GTRVLHLGGL GTRVLHLGGL GTRVLHLGGL GANIVKLHGGL RVNIWKVHGSL RVNIWKVHGSL RVNIWKVHGSL RVNIWKVHGSL FGYYLHLHGSP FGYYLHLHGSP FGYYLHLHGSP FGYYLHLHGSP DFVLFKLHGCS KLSVLHIHGVY KLSVLHIHGVY KLSVLHHGVY KLSVLHHGVY DFVLFKLHGCS VTQIIKYHGDF DAVYXMHGDY DAVYXMHGDY INEIYKIHGCS VMELYKIHGCS VMELYKIHGCS	Heli HHHH 58 FRDLAA 55 FRDFAA 56 ATVYREV 55 FRDFAA 56 ATVYREV 57 INVREV 57 INRCLRS 53 LSWCLQ 55 INRCLRS 54 INRCLRS 55 INRCLRS 54 INRCLRS 55 INRCLRS 54 INRCLRS 55 INRCLRS 55 INRCLRS 54 INRCLRS 55 INRCLRS 55 INRCLRS 55 INRCLRS 56 INR	x-3 S HHHHHHEE ICRPNSTUTY ACRPNAVVVIY LTSHRSTUTY LKSQRDTUVY LEHGDNLCTF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDPLIF ILXEQSYLCI AINAANSFLCI LEKEQSYLLIA NLXEANSFLCI LEKEQSYLFIC NKSFLFLCC NKSFLFLCC NKSFLFLCC NKSFLFLCC DICGKTUFI DIGKTUFI DIGKTUFI DIGKTUFI DLIAKTEFFI DLIAKTEFFI DLIAKTEFFI DLIAKTEFFI DLIAKTEFFI DLIAKTEFFI DLIAKTEFFI DLIAKTEFFI DLIAKTEFFI DLIAKTEFFI DLIAKTEFFI DLIAKTEFFI DLIAKTEFFI DLIA	tr4 EEES GYSG GYSG GYSG GYSG GYSG GYSG GYSG	He E*.HHHHH FGEINRVIRDM IGEHR LLKRMYGT IGEHR LLKRMYGT IGEHR LLKRMYGT IGEHR LLKRMYGT IGEN CLURRNGT FAEQUI LLQRMNGT FDERQC VLANIAN LGEQUS LVRALRA LDSS R LLDAYRQA LDSS R LLDAYRQA ILSS R LLQRMYGT ILST ILST ILST ILST ILST ILST ILST IST ICT IST ICT IST IST IST IST IST IST IST IS	14 Stt5 HHHHEEEEEEE L-TIPSTHLVVIX 1 L-TIPSTHLVIXA 1 GIQRVAVSVFGGS GINRVAVSVFGGS GINRVAVSVFTGQ PFKRMAVSVFTGQ KPKTVAISIYPRS RPQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQ-GAPVTIITY L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFFFCY 1 L-NN-PSLLVFFFCY 1 L-NN-PSLLVFFFCY 1 L-NN-PSLLVFFFLDK WSIWRSGEBKDR VKHKSDLEHFML VKHKSDLEHFML SLNL-KKYALMQLRD IYDY-ENHQSGFAPM KKLWRDSG RNIL-DENISNNYWL RASY-SENQKEHYCI RIQF-AQQQRHYCI TACL-DQDNIHKLKD TACL-DQENIDKLKN IXML-PQEKVNELSN
Secondary Structure XAC2443 Xca 21243176 b111926_Brja_27377037 XAC2189_Xax_21242924 BRA0375 Brsu_23500128 NE2530_Neu_30250449 Pflu2135_Pfl_23059957 PP1406_Ppt_2698140 LA1794_Lepin_24214552 NE2094 Neu_30250034 seal0_Seen_38201747 EF2349_Efae_29376850 VYA0689_Vivul_37676349 VP1802_Vipa_28898576 PG1107_Porgi_34540840 Sir2_Sme_116264652 AK024756.I_Hs_10437125 FLJ20615_Hs_15489177 LJ1095_Lajo_42519021 RPA0819_Rhpa_3933896 CV1059_CNi_34496514 TDE0266_Tden_42525782 MBNC211202_Mesp_4591672 usg_Pepe_11322457 Aple047201_Acpl_3203900 BB0483_BODr_33599473 pX02-73_Ban_10956463 BC1912_Ece_30019423 BK5-Tp30_k5-t_14251154	H2 HHHHH LIEAGAELAGLHLLDRFLGNL LIEAGAELAGLHLIDRFLGNL LVYWMITYGLNYQDGHRFKD TLYWMIQHGLNYQDGHRFKD LLYWALQHGEVIDDLFNGPD LLYWALQHQGEVIDDLFNGPD LLYWALQHQGYLDDLFNGPD LLYWIMQDEITFTFTCDDGFRNPD LLYWIMQBIHHYTGFTBGFFRQL LAEYACDQGIHHYTGFTBGF FLEILAVER-NIESNPKIFFNDGTNGYM AIEMSAEESGINLINGFSGIH LLYNSFIDNDLVDGYNGSL LLESFNQTSNINGNISDGFDEYG LLEIAYERQGRQLESLDLTDEKK LLEIYAADQGRQLESLDLTDEKK LLEIYAADQGRQLESLDLTDEKK LLEIYAADATFQSRPNILKDIN FFEEECKATFQSRPNILKDIN FFEEECK	TP 19 GEPRYLEG CF 16 GGLVREGT CF 16 GGLVREGT CF 14 QPIRGET FH 12 RPYCHAAG SSD 5 TWDIGNTN DN 21 KAYFKNFF AT PDELTCSR FF 4 PNEITSAR KR 17 DNYANELP ISR 17 GEARFGH ISR 17 GEARFGH ID 11 LERKYDNN SSE 17 RYTGRYNT CH - ISAPAG VL - EWAQEKR VL - EWAQEKR VL - EWAQEKR VL - EWAQEKR VL - ENAQEKR SD - ISKITDG ISR 17 GY CH STGKYDR ISR 17 GY ISR	Str-3 EEEEEEEE. VARFTKLHGSV VARFTKLHGSV NTLVFYPHGSL MTLVFYPHGSL GTRVLHLGGL GTRVLHLGGL SCNIFYLHGAL GANIVXLHGGL SCNIFYLHGAL RVNIWKVHGSL RVNIWKVHGSL TINLVXCHGSV I NMYLYXLHGSS I TINLVXCHGSV I NMYLYXLHGSS VTRILKYLGDF KLSVLHHGVY KLSVLHHGVY KLSVLHHGVY DFXILHGCS VPQIVKFHGDF DAVVXKHGDY DAVVXKHGDY DAVVXKHGDY DAVVXKHGDY DAVVXKHGCS VNELYXLHGCS VGELYXLHGCS	Heli HHHH 58 FRDLAA 55 FRDLAA 56 ATVYREV 56 STVYREV 57 INVYREV 57 INVYREV 57 INVYREV 57 INVYREV 57 INVYREV 57 INNADI 57 FRRFSEP 48 YWDTQF 71 IFKKREN 48 YWDTQF 71 IFKKREN 48 YWDTQF 71 IFKKREN 48 YWDTQF 71 IFKKREN 44 VLQNYR 24 EIQKLYE 24 EIQKLYE 24 EIQKLYE 24 EIQKLYE 24 FIQKLYE 24 FIQKLYE 24 FIQKLYE 24 FIQKLYE 25 FFTALRG 22 FFTALRG 22 IAXKLT 22 INAKLS 22 VNAKLS	x-3 S HHHHHHEE ICRPNSTUTY ACRPNAVVYIY LTSHRSTUTY LTSHRSTUTY LAGENHUVYY LEHGDRLIF FMLTDFLIF VMXSVDFLIF VMXSVDFLIF LAQENSTUTF MXSKANSFICI LEKEQSVLIF LKEQSULIF NLSKANSFICI LEKEQFLFLGC NKSFLFLGC NKSFLFLGC QLIKQNA-HVLFF LECKSTLF QLIKQNA-HVLFF LECSTLFU DLIGKVLFF DLIGKVLFF DLIGKVLFF DLIGKVLFF DLIGKVLFF DLIGKVLFF DLIGKVLFF DLIGKVLFF DLIG	tr4 EEES GYS GYS GWS GWS GYS GYS GYS GYS GYS GYS GYS GYS GYS GY	He E*.*HHHHH FGE INRVIRDM IGEH & LIKRMGT FAEQUI LLRAMGT FAEQUI LLQRMGT FAEQUI LLQRMGT FAEQUI LLQRMGT FAEQUI LLQRMGT FAEQUI LLQRMGT FAEQUI LLQRMGT FGOQ IVALALAA LDSS R LLDAYRQA LDSS R LLDAYRQA LDSS R LLDAYRQA ILQRMGT FGOQ INFRINHQ FNE VQPKLMAK FLE ITDIVQRS FGUY INRILIGA GFT LIMLLKPY CKDKGKNEMIKDN MSDLNRLLLHRI YDDTTQALFLEA VDDTTQALFLEA VDDTTQALFLEA LSQUVRYLYKL L-SQUVRYLYKL L-SDVNRYLYKL L-SDVNRYLYKL FDPRLYLSRV FDPRLYLSRV FDPRLYLSRV ISDRNVLSTISI LSDRNVLSTISI LSDRNVLSTISI LSDRNVLSTISI LSDRNVLSTISI	14 Str5 HHHHEEEEEEE L-TIPSTHLVINS 1 L-TIPSTHLVINA 1 GIQRVAVSVFGGS GINRVAVSVFRGD PFKRMAVSVFTGQ KPKTVAISIYPRS RPQQ-LSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQ-GAPVTIITY CIRQNFUTIITY L-NN-PSLHIVICY LKVTPLHVVEWSG FDYKNKPSFIIDK WSIWRSGHEKDR VHKKSDLEHFML LYSVPKVDLEHYML LYSVPKVDLEHYML LYSVPKVDLEHYML SINL-KYALMQLRD SINL-KYALMQLRD SINL-KYALMQLRD KKQM-RGLHKQAYIV KKWRANS RSY-SENQKEHYCI RASY-SENQKEHYCI RASY-SENQKEHYCI RICL-DQDNIHKLKN TACL-QQENIDKLKN IMML-PGEKVNELSN SENLPFLISES PON-
Secondary Structure XAC2443_Xca_21243176 bl11926_Brja_27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530_Neu_30250449 Pflu2135_Pfl_2305957 PP1406_Pput_26988140 LA1794_Lepin_24214552 NE2094_Neu_30250034 seal0_Seen_38201747 EF2349_Efae_29376850 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VP1802_Vipa_28898576 PG1107_Porgi_34540840 Sir2_Smel_16264652 AK024756.1_Hs_10437125 FLJ2103_Hs_13375721 FLJ20635_Hs_15489177 LJ1095_Lajo_42519021 RPA0819_Rhpa_39933896 CV1059_Chvi_34496514 TD50266_Tden_42525782 MBNC211202_Mesp_45916072 usg_Pepe_11322457 Ap1e047201_Acp1_3203900 BB0483_Bobr_3359473 pX02-73_Ban_10956463 BC1912_Bce_30020052 BC1271_Bce_30019423 BK5-Tp30_bK5-t_14251154 SAG1992_Saga_22538128 plu4324_Ph1u_37528154 lici_afu_1427205	H2 HHHHH LIEAGAELAGLHLLDRFLGNL LIEHGCDFAGLRIIDRFVGAL LVYWMTYGLRIDDRHAFKD VVYWAITYGLNUDGHRFKD LYWMIQHQEVIDDLFNGPD LLPWAVQHQEVIDDLFNGPD LLPWAVQHQEVIDDLFNGPD LLPWAVQHQEVIDDLFNGPD LLPWAVQHQEVIDDLFNGPD LLPWAVQHQEVIDDLFNGPD LLPWAVQHQEVIDDLFNGPGV LLEVACDQGETHYTGFTGFRGLI LAEYACDQSRIHHYTGFTHGFFRQL LAEYACDQGULINGFSGH FLELVEK-NIESNPKIFFNDGTNGYM ATEWSAEESGINLINGFSGH LLLYAADQGKQLESLDLTDEKK LLEAYAADQGKQLESLDLTDEKK LLEAYAADQGKQLESLDLTDEKK LLEAYAADQGKQLESLDLTDEKK LLEAYAADQGKQLESLDLTDEKK LLEAYAFGERAYSIAYTKIATV NLEVAFELAYSIAYTKIATV NLEVAFELAGKVVDVKRWE LIELGFKNSDIKAYIGQK FLEQUFADSDIKAYIGQK FLEQUFADSDIKAYIGQK FLEQUFAD	TP 19 GEPRYLEG TP 19 GEPRYLEG CF 16 RGLYREQT CF 16 RGLYREQT CF 14 QPIRGERS 15 CF 16 RGLYREQT VF DVRATGSE S5 5 WDIGNTN DN 21 KAYFKNFF AT PDELTCSR FR 4 PNEITSAR KR 17 DNYANELP ISR 17 GEARFGHY ID 11 LERKYDNN ISE 17 RYTGRYNT RD -IASAPAG (U - EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR C - ISKITDG RL -STQKYDR QL -FYSIHKR QL -LSSNLGA G- LFVSSND GG - LFLKSSD	Str-3 EEEEEEE. VARFTKIHGSI VARFTKIHGSI MTIVFYPHGSI TSLVFYPHGSI ATLVFYPHGSI GRVDHLHGGL GRVDHLHGGL GRVDHLHGGL GRVDHLHGSI TINLVKCHGSI TINLVKCHGSI TINLVKCHGSI TINLVKCHGSI TINLVKCHGSI TINLVKCHGSI TINLVKCHGSI TINLVKCHGSI TINLYKCHGSI VTRIIKYHGDF KLSVLHHGVY KLSVLHHGVY KLSVLHHGVY KLSVLHHGVY MKILHHGSI DAVVKMHCDY DAVVKMHCDY DAVVXMHCDY DAVVXMHCDY DAVVXMHCDY MELYKIHGCS VGELYKIHGCY MELYKIHGCS	Heli HHHH 58 FRDLAA 56 ATVYREV 55 FRDTAA 56 ATVYREV 57 INVYEEV 57 INVYEEV 57 INVKEV 57 INVKEV 55 INRGRS 86 KIKLFD 19 IRFLSYE 57 FRFSEF 77 ILFKRFN 24 VIQUIY 24 EIQKIYE 24 EIQKIYE 24 EIQKIYE 24 EIQKIYE 24 EIQKIYE 23 AKKKET 23 FFTKIKS 24 IQKIYE 24 FIJLQG 22 FFTAIG 22 FFTAIG 22 FFTAIG 22 INAKIIT 22 INAKIIT 22 INAKIIT 22 INAKIIT 22 INAKIIT	x-3 S HHHHH5EE ICRPNSTLVIY ACRPNSTLVIY LTSHRSTLVIY LSGRDUVVY LEHGDNLCTF FANLTDFLIF VWXSVDKLII AINAGSYLCI AINASTLFU LEKEQSVLAF MSSA	tr4 GYS GYG GWG GWG GYG GYG GYG GYG GYS GYS GYS GYS GYS GY	He E*.*HHHHH FGE INRVIRDM FGE INRVIRDM FGE INRVIRDM IGEN & LIKRWGT FAEQI ILQRWGT FAEQI ILQRWGT FAEQI ILQRWGT FAEQI ILQRWGT FAEQI ILQRWGT FAEQI ILQRWGT FGQ INFRINHQ FNE VQPKMAK FLEE ITDIVQRS FGY INRIIGA FDE ITDIVQRS FGY INRIIGA FDE ITDIVQRS FGY INRIIGA FDE ITDIVQRS FGY INRIIGA ETLHQIFQALFLAN VDDTTQALFLAN VDDTTQALFLAN VDDTTQALFLAN USDIVPRIVEDI L-SDVPNIVIISTFV L-SDVPNIVIISTFV LSDNVRYLIKLI FDPNICVILSRU FDPNICVILSRU ISDRVIVISTISI L-SDRVVISTISI LTDRVKKLITDF LTDRVKKLITDF	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIS 1 L-TIPSTHLVIMA 1 GIQRVAVSVFGGS GINRVAVSVFFGG PFKRMAVSVFTGQ KPKTVAISIYPRS RPQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQGAPVTIITY CIRQNFHTIITY L-NN-PNLLVFIFCY 1 L-NN-PNLLVFIFCY 1 L-NN-PSLHIVICY LKVTPLEVVEWSG FDYKNKPSFIIDK WSIWRSGHEKDR VKHKSDLEHFML LVSVPNKVDLEHYML SLNL-KKXHAMQLRD IVDY-ENHQSGKPAP RKQM-RGLHKQAYIV KKWEDSG RNIL-DENISNNYWL RASY-SENQKEHYCI TRCL-DQDNIHKLKD TACL-DQENIDKLKN IMML-PQEKVNELSN SENLPYEISEA AENSPFDISES
Secondary Structure XAC2443 Xca 21243176 bl11926 Brja 27377037 XAC2189_Xax 21242924 BRA0375 Brsu 23500128 NE2530 Neu 30250449 Pflu2135 Pfl 23059957 PP1406 Pput 26988140 LA1794 Lepin 2421452 NE2094 Neu 30250034 seal0 Seen 38201747 EF2349 Efac 29376850 VVA0689_Vivul 37676349 VP1802_Vipa_2898576 Pfl107 Progi 34540840 Sir2_Smel 16264652 AK024756.1 Hs 10437125 FLJ2103 Hs 1337571 FLJ20635 Hs 15489177 LJ1095 Laj0 42519021 RPA0819 Rhpa 3993896 CV1059 Chvi 34496514 TDE0266 Tden 42525782 MBNC211202 Mesp 45916072 usg Pepe 13322457 Aple047201 Acpl 32033900 BB0483 Bobr 33599473 pX02-73 Ban 10956463 BC1912_Bce 30019423 BK5-Tp30 kK5-t 14251154 SAG1992_Saga 22538128 plu3242_Ph1u 37528154 lici Aful 14278228 PAB0801 Pab 14521405	H2 HHHHH LIEAGAELAGLHLLDRFLGAL LVYWTMTYGLNVQDGHRFKD VYWMITYGLNVQDGHRFKD TLYWAMLLFNANGSWFKDA LNYWALQHQGEVIDDLFNGFD LLPWAVQHAFSGTAELFDEQG LLYWTIMQDEITPTFTCDDGFRNPD YLELLCIDVDIPATYGGTHEVIKFPI LAEYACDQSGTHLYTGFTHGFRQL AAEYACDQSRIHYTGFTHGFRGT LIEYAFLSGINLINGFSGIH LLYNSFISGINLINGFSGIH LLEYAADQGKQLESLDITDEKK LLEYAADQGKQLESLDITDEKK LLEYAADQGKQLESLDITDEKK LLEYAADQGKQLESLDITDEKK LLEYAADQGKQLESLDITDEKK LLEYAADQKQLESLDITDEKK LLEYAFEHRGGKLYVVVNDAMQTN FFEEECKATFQSRPNILKDIN LFEXAFKELHRGGKLYVVVNDAMQTN FFEEGFKN	TP 19 GEPRYLEG NP 19 GEPRYLEG CF 16 GL/RECT CF 14 QPIRGES FH 12 RPYGHAAG AS 1 DLCESHTG FF 12 RPYGHAAG SD 5 TWDIGNTN DN 21 KAYFKNFF AF PDELTCSR FR 4 PNEITSAR KR 17 DNYANLEP ST 7 GEARFGHY ID 11 LERKYDNN SE 17 RYTGRYNT TR - IASAPAG VL - EWAQEKR VL - EWAQEKR VL - EWAQEKR VL - ENAGEK SD - ISKITDG CR - ISKITDG CR - ISKITDG CR - ISKITAG CR - ISKITAG CR - LFYSIHKR GG - LFVKSDD GG - LFVKSDD GG - FFDDTIG 	Str-3 EEEEEEE. VARFTKLHGSV VIRITKLHGSL MTI.VFYPHGSL ATLVFYPHGSL ATLVFYPHGSL GANIVKLHGGL SONIFYLHGGL GANIVKLHGGL SONIFYLHGGL RVNIWKVHGSL RVNIWKVHGSL RVNIWKVHGSL RVNIWKVHGSL RVNIWKVHGSL FGYLHLHGSI VIRIKKHGSS KLSVLHHGLY DKSILHHGLY DKSILHHGLY DKSILHHGLY DKSILHHGCS VFQIVKHGDY DAVVXKHGDY DAVVXKHGDY DAVVXKHGDY DAVVXKHGDY DAVVXKHGDY DAVVXKHGDY INELYKIHGCS VGELXKHGCY YGELXKHGST	Heli HHHH Solver Heli Heli Heli Heli Heli Heli Heli Heli	x-3 S HHHHHEEE ICRPNSTLVIY AGRPNAVVVIY LTSHRSTLTLY LSAGPADTLVYY LEHGDNLCTF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDFLIF NUSSANSTLCI LEKEQSTLF NLSSANSTLF LESEQSTLF NLSSANSTLF LESEQSTLF NLSSANSTLF LESEQSTLF NLSSANSTLF LESEQSTLF DLSSTLF DLSSTLF DLSSTLF NTFEPVIFL DLSSFLF VFIEPVIFL LFEPVIFL DLSSTLF NLSSASTLF NLSSASTLF NLSSASTLF NLSSASTLF NLSSASTLF NLSSASTLF NLSSASFLF NLSSA	trif GYS GYS GYS GYS GYS GYS GYS GYS GYS GYS	He E*.*	14 Str5 HHHHEEEEEEE L-TIPSTHLVVNS 1 L-TIPSTHLVVNS 1 GIQRVAVSVFGGS GINRVAVSVFGGS FPKRMAVSVFTGQ KPKTVAISIYPRS RPQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQGRGAVTIITY L-NN-PSLLVFIFCY 1 L-NN-PSLHVVFFCY 1 L-NN-PSLHVVFFCY 1 L-NN-PSLHVFFCY 1 L-NN-PSLHVFFCY 1 L-NN-PSLHVFFCY 1 L-NN-PSLHVFFCY 1 L-NN-SDLEHFML SINL-KYALMQLRD IVDY-ENHQSKPAP RKQM-RGLHKQAYIV KKWRDSG RNIL-DENISNNYWL KK2WRDSG RNIL-DENISNNYWL RASY-SEQQCHYCI TRCL-QQCNHKLKD TACL-DQENINKLKD TACL-DQENINKLKN SENLPYEISEA AENSPFDISES
Secondary Structure XAC2443 Xca 21243176 b111926_Brja_27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530_Neu_30250049 Pflu2135_Pfl_23059957 PP1406_Pput_26988140 LA1794_Lepin_24214592 NE2094_Neu_302250034 seal0_Seen_38201747 EF2349_Efae_29376850 VVA0689_Vivul_37676349 VP1802_Vipa_2898576 PG1107_Porgi_34540840 Sir2_Smel_16264652 AK024756.1_Hs_10437125 FLJ21635_Hs_15489177 LJ1095_Laj0_42519021 RPA0819_Rhpa_3993896 CV1059_Chvi_34496514 TDE0266_Tden_42525782 MBNC211202_Mesp_45916072 usg_Pepe_11322457 Aple047201_Acpl_32033900 BB0483_BbcF_3359473 pX02-73_Ban_10956463 BC1912_Ece_30019423 BK5-Tp30_bK5-t_14251154 SAG1992_Saga_22538128 PL04324 Phu_37528154 lici_Acpl_14521406 APE1782_Ne_14521406 APE1782_Ne_14521406	H2 HHHHH LTEAGAELAGLHLLDRFLGNL LIEHGCDFGLATIDRFVGAL LVYWMTYGLATIDRFVGAL LVYWMITYGLATIDRFNGTA LYWMLQHGLATIDRFNED LLYWYIALQHQEVIDDLFNGPD LLPWAVQHPGTAFLFDEQG LLYWTIMQDEITPTFTCDDGFRNPD YLELLCIDYDIPATYGGTBUFFRQL LAEYACDQSRIHYTGFTHGFFRQL LAEYACDQSRIHYTGFTHGFFRQL LAEYACDQSRIHYTGFTHGFFRQL LAEYACDQSRIHYTGFTHGFSGIH LLYNSFIDNDLVDGYNGSL LLENSFQTSUNGNISDGFDEYG LLEIAVEK-NLESNEKIFFNDGTNGTM LLEIAYEQGKQLESLDLTDEKK LLEIYAADQGKQLESLDLTDEKK LLEIYAADQGKQLESLDLTDEKK LLEIYAADQGKQLESLDLTDEKK LLEAFKBFRNGRVPDVVNDADMQTN FFEEECKATFQSRPNILKDITN LEEGFKNNGRVPDVKINK LLEQFER	TP 19 GEPRYLEG TP 19 GEPRYLEG CF 16 GLYREGT CF 14 QPIGEAT FH 12 RPYGHAAG MAS 1 DLCESHTG FF 12 RPYGHAAG MAS 1 DLCESHTG FF 27 RPYGHAAG TP DURTGSE ST 5 TWDIGNTN DN 21 KAYFKNFF TP DELTCSR FR 4 PNEITSAR KR 17 DNYANELP TR 4 PNEITSAR KR 17 DNYANELP TR 4 PNEITSAR TR	Str-3 EEEEEEEE. VARFTKLHGSV VIRITKLHGSV VIRITKLHGSL MTIVFYPHGSL ATIVFYPHGSL GTRVLHLGGL GTRVLHLGGL GTRVLHLGGL SCNIFFLHGAL GANIVKUHGGL RVNIKKUHGSL TINLVKCHGSL TINLVKCHGSL TINLVKCHGSL FGYYLHLHGSP PIRLYKLHGSS VIRIKYHGCS KLSVLHIHGVY KLSVLHHGVY KLSVLHHGVY KLSVLHHGVY DAVVKHGDF DAVVKHGDF DAILYKHGCS VVCIKHGDF DAILYKHGCS VVELYKIHGCS VVELYKIHGCS VVELYKIHGCS SRVUHLHGSL SRVUHLGSL	Heli HHHH HS8 FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA HS7 HRS7	x-3 S HHHHHEE ICRPNSTUTY ACRPNAVVYIY LTSHRSTUTY KSQRDTUVYY LEHGDNLCTF FANLTDPLIF VMKSVDVLII AURASYLCI AINAANSFLCI LEKEQSVLIA AINAANSFLCI LEKEQSVLIA AINAANSFLCI LEKEQSVLIA AINAANSFLCI LEKEQSVLIA AINAANSFLCI LEKEQSVLIA AINAANSFLCI LEKEQSVLIA DISEA	trees GYSG GWSG GWSG GWSG GYSG GYSG GYSG GYSG	He E*.HHHHH F-GEINRVIRDM IGEHR LLKRMYGT IGEHR LLKRMYGT IGEHR LLKRMYGT IGEN VIRVLEDM IGEN VIRVLEN IGEN VIRALRA LDSS IR LLOAVROA LDSS IR LLOAVROA LDSS IR LLOAVROA ILSS IR LLOAVROA ILSS IR LLOAVROA ILSS IR LLOAVROA FALSE FOR INTELLE INTEL INTEL INTEL ICT INTELLE INTEL IST INTELLERI V-DDTTPQALFLEA 	14 Stt5 HHHHEEEEEEE L-TIPSTHLVINS 1 L-TIPSTHLVINA 1 GIQRVAVSVFGGS GINRVAVSVFRGD PFKRMAVSVFTGQ RPKQLSIAIRPLG KFKQLFISIVGNH LVKNEKFTIEIVD CQRQ-GAPVTIITY CLRQNTPITIITY L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLVVEWSG KVHKSDLEHFML SINJ-KKYALMQLRD IXDY-ENHQSGKPAP RKQM-RGLKQAYIV KKWRDSG RNIL-DENISNNYWL RASY-SENGKEHYCI RIQF-AQDQRQHYCI TACL-DQDNIHKLKD SSQLPYEISEA AENSPTEISES SSQLPEDKRKS IVKQ-RGGAILEINP IVKE-SGGTVIEVNV
Secondary Structure XAC2443 Xca 21243176 b111926_Brja_27377037 XAC2189_Xax_21242924 BRA0375 Brsu_23500128 NE2530_Neu_30250449 Pflu2135_Pfl_23059957 PP1406_Ppt_2698140 LA1794_Lepin_24214552 NE2094 Neu_30250034 seal0_Seen_38201747 EF2349_Efae_29376850 VYA0689_Vivul_37676349 VP1802_Vipa_28898576 PG1107_Porgi_34540840 Sir2_Sme_116264652 AK024756.I_Hs_10437125 FLJ20635_Hs_15489177 LJ1095_Lajo_42519021 RPA0819_Rhpa_3933896 CV1059_Cvija_345946514 TDE0266_Tden_42525782 MBNC211202_Mesp_45916072 usg_Pepe_11322457 Aple047201_Acpl_3203900 BB0483_B0br_33599473 pX02-73_Ban_10955463 BC1912_Ece_30019423 BK5-Tp30_bK5-t_14251154 SAG1992_Saga_22538128 plu4324_Ph1u_37528154 lici_Aful_14278228 PA0601_Pab_14521406 APE1782_Ape_14601621 aq_2170_Aae_15607106	H2 HHHHH LTEAGAELAGLHLLDRFLGNL LIEHGGEFAGLHIDRFVGAL LVYWMTYGLDIDDRHAFKD TLYWAILQHGLOIDDRHAFKD LLYWALQHQEVIDDLFNGPD LLPWAVQHAPSGFAELFDEQG LLYWIMQDEITFTFTCDDGFRNPD YLELLCIDYDIPATYGFTBGFFRQL LAEYACDQSRIHHYTGFTHGFFRQL LAEYACDQSRIHYTGFTHGFFRQL LAEYACDQSRIHYTGFTHGFFRQL LAEYACDQSRIHYTGFTHGFRQL LLYNSFIDNDLVDGYNGSL ILESFNQTSNINGNISDGFDEYG ILEIAYER-NIESNPKIFFNDGTNGYM LLELAYERQGRQLESLDLTDEKK LLEJYAADQGRQLESLDLTDEKK LLEAFGRAATFQSRPNILKDIN FFEEECKAFKBIHHGGKIFVVVNDADMGTN NLEKAFFDATFQSRPNILKDIN LEGFKFR	TP 19 GEPRYLEG CF 16 GGLVREGT CF 14 QPIRGET FH 12 RPYGHAAG SSD 5 TWDIGNTN DN 21 KAYFKNFF AT PDELTCSR FF 4 PNEITSAR KR 17 DNYANELP ISR 17 GEARFGH ID 11 LERKYDNN SSE 17 RYTGRYNT CH - ISAPAG VL - EWAQEKR VL - EWAQEKR VL - EWAQEKR VL - EWAQEKR VL - EWAQEKR SD - ISKITDG GD - ISKITDG GL - STYKTPG QL - PVSIHKR QL - ATTRSKR QL - PVSIHKR QL - ATTRSKR GG - LFVKSND GG - LFVKSND GG - LFVKSND GG - FFDDTIG GG - CH STYCR CH - STYCR	Str-3 EEEEEEEE. VARFTKIHGSV VARFTKIHGSV MTLYYPHGSL MTLYYPHGSL GTRVLHLGGL GTRVLHLGGL GTRVLHLGGL GANIVXIHGGL GANIVXIHGGL RVNIMKVHGSL RVNIMKVHGSL TINLVKHGSS 1 NMYLYKLHGSS 1 NMYLYKLHGSF PIRLIXLHGSF PIRLIXLHGSY VTRIKYHGDF DEVLFXIHGCY VTRIKYHGDF DEVLFXIHGCY VTGLIXHGSV VTGLIXHGSV VTGLIXTHGCS VNELXKHGCY VGELYKIHGCS VGELYKIHGSL TRNLIEHGNI SRRVIEHGNV SRRVIEHGNV	Heli HHHH 58 FRDLAA 55 FRDLAA 56 ATVYREV 57 INVYREV 57 INVYREV 57 INVYREV 57 INVYREV 57 INVYREV 57 INNADI 45 IINNADI 45 IINNADI 45 IINNADI 45 IINNADI 45 IINNADI 45 IINNADI 47 INFRSPA 48 YWDTQF 71 IFKKREN 48 YWDTQF 71 IFKKREN 44 VLQNYR 44 VLQNYR 44 VLQNYR 45 ILQKIYE 45 ILQKIY	x-3 S HHHHHSEE ICRPNSTUTY ACRPNAVVYIY LTSHRSTUTY LTSHRSTUTY LTSHRSTUTY FALGBILT VWXSVDFLIF VMXSVDFLIF VMXSVDFLIF VMXSVNSTICI LEKEQSVLIF LEKEQSVLIF LKEQSVLIF LKEQSTUF NISKANSILI LKEQFLFLGC NKSFLFLGC NKSFLFLGC VICOKSVLFI DLIGKSTLFU DLIGKSTLFU DLIGKSTLFI DLIGKSTLFI DLIGKSTLFI DLIGKSTLFI DLIGKSTLFI DLIGSVLFI DLIGKSTLFI DLIGSVLFI DLIGSVLFI DLIGSVLFI DLIGSVLFI DLIGSVLFI DLISSVLFI DLISSVLFI DLIS	tref GYSG GYSG GWSG GWSG GWSG GYSG GYSG GYSG	He E*.*HHHHH FGE INRVIRDM FGE INRVIRDM IGEH & LIKRMGT FAEQI ILQRMGT FAEQI ILQRMGT FAEQI ILQRMGT FAEQI ILQRMGT FAEQI ILQRMGT FGQ ILQRMGT FGQ INFRINHQ FNE VQPKLMAK FLE ITDIVORS FGI INNILKY CKDKGKNEMIKDN MSDINRLILHRI VDDTTQALFLEA VDDTTQALFLEA VDDTTQALFLEA HKDFEVSRUEDI LSDMNIRLLHRI FDPRINIVTAL ESDMNIRLLHRI FDPRINIVTISI ISDNVQILDSI LSDRVISTISI ISDRNVLSTISI LTDRNKKLTDF ISGLYPAALIPY ISGLYPAALIPY ISGLYPAALIPY ISGLYPAALIPY	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIS 1 L-TIPSTHLVINA 1 GIQRVAVSVFGGS GINRVAVSVFRGD PFRMAVSVFRGC PFRMAVSVFRGC KPKTVAISIYPRS RPQQ-LSIAIRPLG KFYQLFISIYONH LVKNEKFTIEIVD CQRQ-GAPVTIITY CIRQNFITIITY L-NN-PNLLVFIFCY 1 L-NN-PNLLVFIFCY 1 L-NN-PSLHIVICY LKVTPLHVVVEWSG FDYKSDLEHFML LVSVPKVDLEHYML LVSVPKVDLEHYML SUNL-KKYALMQLRD KKQM-RGLHKQAYIV KKLWKANV WQTWRDSG RNIL-DENISNNYWL RASY-SENQKEHYCI RICL-DQDNIHKLKN TACL-QQENIDKLKN IMML-PGEKVNELSN SSQLPFDISES SSQLPFDISES SSQLPFDISES
Secondary Structure XAC2443_Xca_21243176 bl11926_Brja_27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530_Neu_30250449 Pflu2135_Pfl_2305957 PP1406_Pput_26988140 LA1794_Lepin_24214494 LA1852_Lepin_24214552 NE2094_Neu_30250034 seal0_Seen_38201747 EF2349_Efae_29376850 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VP1802_Vipa_28898576 PG1107_Porgi_34540840 Sir2_Smel_16264652 AK024756.T_H8_10437125 FLJ2103_JBs_13375721 FLJ20635_H8_15489177 LJ1095_Laj0_42519021 RPA0819_Rhpa_39933896 CV1059_Chvi_34496514 TDE0266_Tden_4252782 MBNC211202_Mesp_45916072 usg_Pepe_11322457 Ap1e047201_Acp1_32033900 BB0483_Bobr_3359473 pX02-73_Ban_10956463 BC1912_Bce_30020052 BC1271_Bce_30019423 BK5-Tp30_0K5tc_14251154 SAG1992_Saga_22538128 plu4324_Ph10_37528154 lici_Aful_14278228 PAB0801_Pab_14521406 APE1782_Ape_14601621 aq_2170_Aae_15607106	H2 HHHHH LIEAGAELAGLHLLDRFLGAL LVYWMTTGLRUJDRFVGAL LVYWMTTGLRUJDCHRFKD TLYWAMLLFNAANGSWFKDA LNYWALQHQGEVIDDLFNGFD LLYWWIMQDEITPTFTCDDGFRNPD YLELLCIDYDIPATYGDTEVIKFPI LAEYACDQSRIHHYTGTHGFFRQL LAEYACDQSRIHHYTGTHGFFRQL LAEYACDQSRIHHYTGTHGFFRQL LAEYACDQSRIHHYTGTHGFFRQL LLYNSFISGINLINGFSGIH LLYNSFISGINLINGFSGIH LLYNSFIYGRVDKYKK LLEJAADQGRQLESLDITDEKK LLEJAADQGRQLESLDITDEKK LLEJAADAFFQSRPNILKDIN FFERECK	TP 19 GEPRYLEG TP 19 GEPRYLEG CF 16 RGLYREQT CF 16 RGLYREQT CF 14 QPIRGERS ID LCESHTG IF 12 RYTGHAAG ID LCESHTG IF 2 RYTGHAAG ID LCESHTG IT A POELTSAR KR 17 DNYANELP ISR 17 GEARFGHY ID 11 LERKYDNN ISE 17 RYTGRYNT RD -IASAPAG (L -EWAQERR VL -EWAQERR VL -EWAQERR VL -EWAQERR VL -EWAQERR VL -EWAQERR VL -EWAQERR ISR 17 RD -IASAPAG C -IASAPAG ISR 17 RD -ISRIDG RL -STQKYDR ISR 10 ISR 17 RD -ISRITGG RL -LFSQFE ISL -LFSNELA GG -LFVSSND GG -FFDDTIG 	Str-3 EEEEEEE. VARFTKIHGSI VARFTKIHGSI MTIVFYPHGSI TSLVFYPHGSI KPRLYIHGGI GRVDHLHGGI GRVDHLHGGI GRVDHLHGGI GRVDHLHGGI GRVDHLHGSI TINIVKHGSI TINIVKHGSI TINIVKHGSI TINIVKHGSI TINIVKHGSI TINIVKHGSI PIFLYKHGSI VTRIIKYHGDF MKIVHHGVY KLSVLHHGVY KLSVLHHGVY KLSVLHHGVY KLSVLHHGVY MKIKHGDF DAVYKHGDF DAVYKHGDF UTQIIKYHGDF DAVYKHGDY DAVYKHGDY DAVYKHGDY DAVYKHGDY MEIXIHGCS VGEIXIHGCS VGEIXIHGSI SRVUELHGNI SRVUEHGNI	Heli HHHH 58 FRDLÄAA 56 FRDLÄAA 56 ATVYREV 57 STVXEEV 57 ISFCTOQ 53 LSKCLGQ 53 LSKCLGQ 54 ISKCLGQ 55 LNKCLK 55 LNKCLK 56 KLKLFDD 45 TIHNADL 19 LFKKRNA 45 TIHNADL 19 LFKKRNA 44 SUDVIQ 46 YUNIQ 47 VRFSDA 44 EIQKLYE 44 EIQKLYE 44 EIQKLYE 44 EIQKLYE 44 EIQKLYE 45 ZLONKRA 45 ZLONKRA 45 ZLONKRA 45 ZLONKRA 45 ZLONKRA 45 ZLONKRA 45 ZLONKRA 45 ZLONKRA 46 LPPULD 48 LPREELD 48 LPPULS 49 LPUNS 40 LPUNS 40 LPUNS 40 LPUNS 40 LPUNS 40 LPUNS 41 LDTSLE 41 LPUNS 41 LPUNS 41 LDTSLE 41 LPUNS 41 LPUNS 41 LDTSLE	x-3 S HHHHHEEE ICRPNXVVTY LTSHRSTUTY ACRPNXVVTY LTSHRSTUTY LSHGDIVY LAGENBILIF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDPLIF TAUNANSFLIF LEKEQSVLIAF LKEQTTLFVN ALSEAEIILFVN NLSKANSIIF LEFATVLFF NKSFLFUCC TKSFLFV QLIKQNA-HVLFF DLGKSFLFV QLIKQNAKSFLFV QLIKQNAKSFLFV DLGKSFLFV DLGKSFLFV DLGKSFLFV DLGKSFLFV DLGKSFLFV NEFEFVFFL LFLEFVFFL NTFESFLFFF SMIKSSIIFI RAMREVE-RADVIT RAMREVE-RADVIT RAMREVE-RADVIT RAMREVE-RADVIT RAMELSREAHVFT GID-AISKADTL	tref GYS GYS GWS GHA GHS GYS GYS GYS GYS GYS GYS GYS GYS GYS GY	He E*HHHHH FGEIINRVIRDM FGEIINRVIRDM FGEIINRVIRDM IGENRILLKRMYGT FAEQIIILQRMNGT FAEQIIILQRMNGT FAEQIIILQRMNGT FAEQIIILQRMNGT FGQIINRIHGKG FGQIINFRIHG FNEIQPINKAK FLEEITDIVQRS FGYINRILGA FGYINRILGA FGYINRILGA FGYINRILGA FGYINRILGA FGYINRILGA CKDKGINEMIKDN MSDINIRLLGA VDDTTPQALFLAS VDDTTPQALFLAS VDDTTPQALFLAS VDDTTPQALFLAS VDDTTPQALFLAS VDDTTPQALFLAS VDDTTPQALFLAS ISDYDFNSIVFVI L-SDYNVRYIXIKL L-SDNNRILLIHRI FDPRINIQILSAI FDPNIQUILSAI FDPNIQUILSAI FDPNIQUILSAI ISDRNVLSTISI LTDENIRKLTDF LTDRNVKKLTDF TSGAVDPAGLIPL TSGVVPAAELPF TSGVVPAAELPF	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIS 1 L-TIPSTHLVIMA 1 GIQRVAVSVFGGS GINRVAVSVFFGG PFKMAVSVFFGG PFKMAVSVFFGG KPKTVAISIYPRS RPQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQGAPVTIITY CIRQNFHTIITY L-NN-PNLLVFIFCY 1 L-NN-PNLLVFIFCY 1 L-NN-PSLHIVICY LKVTPLHVVEWSG FDYKNKPSFIIDK WSIWRSGHEKDR VKHKSDLEHFML LVSVPKVDLEHYML SLNL-KKXHMQLRD IVDY-ENHQSGKPAP RKQM-RGLHKQAYIV KKWENSG RNIL-DENISNNYWL RASY-SENQKEHYCI TRCL-DQDNIHKLKD TACL-DQENIDKLKN IKML-PQEKVNELSN SENLPYEISEA AENSPFDISES SQLPKEDKKS IVKQ-RGGAILEINP VAKE-SGATLINVNP VAKE-NGAQVIEVNP VAKE-NGAQVIEVNP
Secondary Structure XAC2443 Xca 21243176 bl11926 Brja 27377037 XAC2189_Xax 21242924 BRA0375 Brsu 23500128 NE2530 Neu 30250449 Pflu2135 Pfl 23059957 PP1406 Pput 26988140 LA1794 Lepin 2421452 NE2094 Neu 30250034 seal0 Seen 38201747 EF2349 Efae 29376850 VVA0668_Vivul 37676349 VP1802_Vipa_2898576 Pfl107 Prorj 34540840 Sir2_Smel 16264652 AK024756.1 Hs 10437125 FLJ2103 Hs 1337521 FLJ20635 Hs 15489177 LJ1095 Laj0 42519021 RPA0819 Rhpa 3993896 CV1059 Chvi 34496514 TDE0266_Tden 42525782 MBNC21120 Zesp 45916072 usg Pepe 1322457 Aple047201 Acpl 32033900 BB0483 Bobr 33599473 PX02-73 Ban 10956463 BC1912_Bce 30019423 BK5-Tp30 kK5-t 14251154 SAG1992_Saga 22538128 plu4324_Phlu 37528154 lici Aful 14278228 PA08010 Hab 14521406 APE178_Ape_14601621 aq 2170 Aae_15607106 CPE0256 Chep 1830928	H2 HHHHH LIEAGAELAGLHILDRFLGAL LIEHGCFAGLRIJDRFVGAL LVYWMITYGLNVQDGHRFKD TLYWAMLLFNANGSWFKDA LNYWALQHQGEVIDDLFNGPD LLYWYIMQDEITPTFTCDDGFRNPD LLYWTIMQDEITPTFTCDDGFRNPD LLYWTIMQSRIHYTGFTHGFFRQL LAEYACDQSGTHLYTGFTHGFRQL LAEYACDQSGINLINGFSGIH LLYNSFIDNDLVDGYNGSL LIESFNQTSGINLINGFSGIH LLYNSFIYGRYKKISHV LLELYAADQGKQLESLDLTDEKK LLELYAADQGKQLESLDLTDEKK LLELYAADQGKQLESLDLTDEKK LLELYAADQGKQLESLDLTDEKK LLELYAADQGKQLESLDLTDEKK LLELYAAD	TP 19 GEPRYLEG TP 19 GEPRYLEG CF 16 GL/RECT CF 14 QPIGERS FH 12 RPYGHAAG AS 1 DLCESHTC FF 12 RPYGHAAG SD 5 TWDIGNTN DN 21 KAYFKNFF TP DELTCSR FR 4 PNEITSAR KR 17 DNYANELP TR 7 GEARFGHY ID 11 LERKYDNN SE 17 RYTGRYNT TD - IASAPAG VL - EWAQEKR VL - EWAQEKR VL - EWAQEKR VL - EWAQEKR VL - ENAQEKR VL - ENAQEKR VL - ENAQEKR CH - STQKYDR GD - IAFWEAA SD - ISKITDG GC - LFYSNPLE LEL - LFSNPLE ELL - LFSNPLE ELL - LFSNPLE IG - LFVKSND IG - LFVKSND IG - LFVKSND IG - LFVKSND IG - LFVKSND IG - LFVCSNP IS - LGTNTCA IS - CONTACT IS - C	Str-3 EEEEEEEE. VARFTKLHGSV VIRLTKLHGSL MTI.VFYPHGSL ATLVFYPHGSL ATLVFYPHGSL GANIVKLHGGL GANIVKLHGGL SONIFYLHGAL GANIVKLHGGL RVNIWKVHGSL RVNIWKVHGSL RVNIWKVHGSL RVNIWKVHGSL RVNIWKVHGSL FGYYLHLHGSI YTRIIKYHGDF PIRLYKLHGSI VTRIIKYHGDF DAILYKHHGYY KLSVLHHGVY KLSVLHHGYY KLSVLHHGY DAILYKHGCS VPCIUKHGDF DAILYKHGCS VMELYKIHGCS VMELYKIHGCS SRNVFLHGSI SRNVFLHGNI SRRVLELHGNI SRKVFLHGSV	Heli HHHH HS FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA HS FRDLAA FRSLA	x-3 S HHHHHEEE ICRPNSTLVIY AGRPNAVVVIY LTSHRSTLVIY LKSQRDTLVYY LEHGDNLCIF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDFLIF FANLTDFLIF NUSSANSILI LKEQSVLTA AINEASTLFIC NKSFLFLCC NKSFLFLCC NKSFLFLCC ILATCVIFI DLISKSUFY DLISKSUFY DLISFLFLC DLISFLFF DLISFFLFI DLISFFLFI DLISFFLFI DLISFFLFI DLISFFLFI DLISFFLFI DLISFFLFI DLISFFLFI DLISSFIFI RAMEVE-SPIFI NUVESFIFI RAMEVE-RADVIT RAMEVE-RADVIT EAFG-LARSDVMI RAYELSREAHVFI EAFG-LSSEASLMI	tree GYGG GYGG GYGG GYGG GYGG GYGG GYGG G	He E*.*HHHHH FGD VNRVLFDM IGED R. LLKRMYGT FAEQI I LLQRMNGT FDEROG VLAAIAAN LCSS R LLAVROA LDSS R LLQRMNGT FGQ ILLLRMKG FGQ INFRINQ FNE VOPKAMAK FLE ITDIVQRS FGY INRILGA GFT INMLKYPY CKDKGINEMIKNN MSDINIRLLHRI VDDTTPQALFLEA VDDTTPQALFLEA VDDTTPQALFLEA VDDTTPQALFLEA USDYDFNSIYTFV LSDNNRLLHRI FDPNLEYILSRV ISDNNRLLHRI FDPNLEYILSRV ISDNNQLINAAT FDPNLEYILSRV ISDNNQLINAAT FDPNLEYILSRV ISDNNQLINAAT FDPNLEYILSRV ISDNNQLINAAT FDPNLEYILSRV ISDNNQLINAAT FDPNLEYILSRV ISDNNQLINAAT FDPNLEYILSRV ISDNNQLINAAT FSUVVYPAASLPL ISGVYPAASLPL TSGVYPAAELPF	14 Stt5 HHHHEEEEEEE L-TIPSTHLVINS 1 L-TIPSTHLVINS 1 GIQRVAVSVFGGS GINRVAVSVFGGS PFKRMAVSVFTGQ KPKTVAISIYPRS RPQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQGAPVTIITY L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFFFCY 1 L-NN-PSLLVFFFCY 1 L-NN-PSLLVFFFCY 1 L-NN-PSLLVFFFCY 1 L-NN-PSLLVFFFCY 1 LSVPNKVDLEHYML SLNL-KYALMQLRD IYDY-ENHQSGFAP RNIL-DENISNNYWL KKWRDSG RNIL-DENISNNYWL KK2PLISSNYWL TACL-DQENIDKLKM TML-PQEVNELSN SSQLPYEISEA AENSPFISES SSQLPKEDKKS IVNQ-RGGAIIEINP IVKE-SGATLINVNP VAKE-NGAQVIEVNP YFF-RGKLVLNLK
Secondary Structure XAC2443 Xca 21243176 b111926_Brja_27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530_Neu_30250049 Pflu2135_Pfl_23059957 PP1406_Pput_26988140 LA1794_Lepin_24214592 NE2094_Neu_302250034 seal0_Seen_38201747 EF2349_Efae_29376850 VVA0689_Vivul_37676349 VP1802_Vipa_2898576 PG1107_Porgi_34540840 Sir2_Smel_16264652 AK024756.T_Hs_10437125 FLJ2103_Hs_13375721 FLJ20635_Hs_15489177 LJ1095_Laj0_42519021 RPA0819_Rhpa_3993896 CV1059_Chvi_34496514 TDE0266_Tden_42525782 MBNC211202_Mesp_45916072 usg_Pepe_11322457 Aple047201_Acpl_32033900 BB0483 Bobr_3359473 pX02-73_Ban_10956463 BC1912_Ece_30019423 BK5-Tp30_bK5-t_14251154 SAG1992_Saga_22538128 PH04324_Ph1u_37528154 lici_Aful_14278228 PAB0801_Pab_14521406 APE1782_Ape_14601621 aq_2170_Aae_15607106 CPE0256_C1pe_18309238	H2 HHHHH LTEAGAELAGLHLLDRFLGNL LIEHGCDFGLATIDRFVGAL LVYWMTTYGLATIDRFVGAL LVYWMITYGLATIDRFVGAL LVYWMILQHQEAUTDDHAFKD LYWMIQHQEAUTDDHFNGPD LLPNAVQHPGTHEYTGTBURFPI LAEYACDQEGTHHYTGFTHGFFRQL LAEYACDQEGTHHYTGFSGH LLYNFIDNDLVDGYNGSL LLYNFIDNDLVDGYNGSL LLYNFFIDNDLVDGYNGSL LLYNFFIDNDLVDGYNGSL LLENAVEQEKQLESLDLTDEKK LLELYAADQGKQLESLDLTDEKK LLELYAADQGKQLESLDLTDEKK LLELYAAD		Str-3 EEEEEEEE. VARFTKLHGSV VIRITKLHGSV VIRITKLHGSL MTIVFYPHGSL ATIVFYPHGSL GTRVLHLGGL GTRVLHLGGL GTRVLHLGGL GTRVLHLGGL KIRVKHGSL TINLVKCHGSL TINLVKCHGSL TINLVKCHGSL TINLVKCHGSL TINLVKCHGSL FGYYLHLHGSP PIRLYKLHGSP PIRLYKLHGSP MYLYKHGDF DFVLFKLHGCY MSLLHHGVY MSLLHHGVY MSLLHHGVY MSLLHHGSV MSLLYKLHGSV YGELYKLHGSV YGELYKLHGSV YGELYKLHGSV SKKVIELHGNV SKKVIELHGNV SKKVIELHGNV	Heli HHHH HS8 FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA FRDLAA HS7 FRFSP FRFSP FRFSP FRFSP FRFSP FRFSP HWDYLOF FRFSP FRF	x-3 S HHHHHEE ICRPNSTUTY ACRPNAVVYIY LTSHRSTUTY KSQRDTUVYY LEHGDNLCTF FANLTDPLIF VMKSVDVLII AURASYLCI AINAANSFLCI LEKEQSVLIA AINAANSFLCI LEKEQSVLIA AINAANSFLCI LEKEQSVLIA AINAANSFLCI LEKEQSVLIA AINAANSFLCI LEKEQSVLIA AINAANSFLCI LEKEQSVLIA DISFLFLCC VICOFLFLCC VICOFTIFI DLIACTIFI DLIACTIFI DLIAKTFLFI DLIAKTFLFI DLIAKTFLFI DLIASVLIFI IFIEFVIFI DLIASVLIFI RAFELAEKADAVI RAFELAEKADVII RAFELAEKADVII RAFELAEKADVII RAFELAEKADVII TIRA-ISRAMUFI GAIDAISSAN	ti4 EEEE GYGGGYGGGYGGGYGGGYGGGYGGGYGGGYGGGY	He E*.HHHHH FGE INRVIRDM IGEH R LLKRMYGT IGEH R LLKRMYGT IGEH R LLKRMYGT IGEH R LLKRMYGT IGEN C UNRVLFUM IGEN C UNRVLFUM IGEN C UNRVLFUM IGEN C UNRVLFUM FALSEN FGO INRIILLIGN FNE UPVRIMENT FNE UPVRIMENT FNE UPVRIMENT FSINTRILLIGN GFI INNLIFT V-DDTTPQALFLEA GFI INNLIFT VDDTTPQALFLEA CEXCNONSIVITY LSDNVRILLIGN FDPVSRIVEDI LSDNVRILLIGN FDPNLINUSSI FDPNLINUSSI FDPNLINUSSI FDPNLINUSSI FDPNLINUSSI FDPNLINUSSI FDPNLINUSSI ISDNVQILSSI LSDNVQILSSI LSDNVQILSSI LSDNVQILSSI LSDNVQILSSI LSDNVQILSSI LSDNVQILSSI LSCVYPASLPL T-SGLVTPASLPL T-SGLVTPASLPL T-SGLVTPASLPL T-SLVYPASLPL T-SLVYPASLPL T-SLVYPASLPL T-SLVYPASLPL T-SLVYPASLPL T-SLVYPASLPL T-SLVYPASLPL T-SLVYPASLPL T-SLVYPASLPL T-SLVYPASLPL T-SLVYPASLPL T-SLVYPASLPL T-SLVYPASLPL T-SLVYPASLPL T-SLVYPASLPL T-SLVYPASLPL T-SLVYPASLPL	14 Stt5 HHHHEEEEEEE L-TIPSTHLVIXS 1 L-TIPSTHLVIXS 1 GIQRVAVSVFGGS GINRVAVSVFRGD PFKRMAVSVFTGQ RPKQLSIAIRPLG KFKQLFISIVGNH LVKNEKFTIEIVD CQRQ-GAPVTIITY L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-SGIBKDR WSIWRSGEBKDR VKHKSDLEHFML SLNL-KKYALMQLRD IXDY-ENHQSGKPAP RKQM-RGLKQAYIV KKWRDSG RNIL-DENISNNYWL RASY-SENGKEHYCI RICL-DQCNIHKLKD SSQLFYEISEA ARKS-SGINIENNYW VKQ-RGGAILEINP IVKQ-RGGAILEINP IVKQ-RGGAILINNY VKKNGAQVIEVNP XKF-NGAQVIEVNP YFF-RGKNLVLINK IVFN-SGKUVINLNK MVP-SHVPQILINR
Secondary Structure XAC2443_Xca_21243176 bl11926_Brja_27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530_Neu_30250449 Pflu2135_Pfl_2305957 Prl406_Pput_26988140 LA1852_Lepin_24214552 NE2094_Neu_30250034 seal0_Seen_38201747 EF2349_Efae_29376850 VVA0689_Vivu1_37676349 VV1802_Vipa_28898576 PG1107_Porgi_34540840 Sir2_Smel 16264652 AK024756.I Hs_10437125 FLJ21103_Hs_13375721 FLJ20635_Hs_15489177 LJ1095_Laj0_42519021 RPA0819_Rhpa_39933896 CV1059_Crwi_34496514 TDE0266 Tden_42525782 MBNC211202_Mesp_45916072 usg_Pepe_11322457 Aple047201_Acpl_32033900 BB0483_B0br_33599473 pX02-73_Ban_10956463 BC1912_Ece_3002052 BC1271_Bce_30019423 BK5-Tp30_K5-t_14251154 SAG1992_Saga_22538128 plu4324_Phlu_37528154 lici_Afu1_1427828 PA0801_Pab_14521406 APE1782_Ape_14601621 aq_2170_Ae_15607106 CPE0256_Clpe_18309238 TM0490_Tma_15643256 HS71_Sc_6324504 Sir21_Sc_6320163	H2 HHHHH LTEAGAELAGLHLLDRFLGNL LIEHGCEFGLNUDDGHRFKD LVYWMITYGLDIDDRHAFKD LYWMIQHGLOIDDRHAFKD LYWMIQHQGEVIDDLFNGPD LLPWAVQHQFGPAELPDQG LLYWINQDEITFTFTCDDGFRNPD LLEWAVQHBGTHHYTGFFHGFFRQL LAEYXCDQEGTHHYTGFFHGFFRQL LAEYXCDQ	TP 19 GEPRYLEG CF 16 GLVREOT CF 14 QPIRGET CF 14 QPIRGET FH 12 RPYCHAAG SSD 5 TWDIGNTN DN 21 KAYFKNFF AT PDELTCSR FF 4 PNEITSAR KR 17 DNYANELP SSD 5 TWDIGNTN DN 21 KAYFKNFF TG PIRSTAR TG PIRSTAR TG PIRSTAR CL EWAQEKR VL EWAQEKR VL EWAQEKR VL EWAQEKR VL EWAQEKR CL ENIDKLA SD ISKITDG ISKITDG ISKITDG ISKITDG CL STYKYPR QL ATTRSKR QL ATTRSKR QL STRIAG SG LFSHPLE EL LFSHPLE EL LFSHPLE EL LFSHPLE CG STRIAG GG LFVKSND GG FFDDTIG GG STRIAG GG STRIAG CG STRIAG	Str-3 EEEEEEEE VARFTKLHGSV VARFTKLHGSV VARFTKLHGSL MTLVFYPHGSL ATLVFYPHGSL GTRVLHLGGL GTRVLHLGGL GTRVLHLGGL GANIVKVHGSL RVNIWKVHGSL RVNIWKVHGSL TINLVKCHGSV TINLVKCHGSV TINLVKCHGSV TINLVKCHGSV VTRIKYGDF KLSVLHHGSV VTRIKYGDF KLSVLHHGVY KLSVLHHGVY KLSVLHHGVY DAVVKHGDF DAVVKHGDF DAVVKHGDF DAVVKHGDF DAVVKHGDF DAVVKHGDF MKELYKLHGS VGELYKLHGSS VGELYKLHGSS VGELYKLHGSS SRNVILHGSL SRNVIHGSL SRNVILHGSL SRNVILHGSL SRNVILHGSL SRNVILHGSL SRNVILHGSL SKVFELHGNV SKKVELHGNV SKKVELHGNV SKVCHGSF TDKLVCCHGSF TDKLVCCHGSF	Heli HHHH 58 FRDLAA 55 FRDLAA 56 ATVYREV 56 STVYREV 57 INVYREV 57 INVYREV 57 INVYREV 57 INVYREV 57 INVYREV 57 INNADI 45 IINNADI 45 IINNADI 45 IINNADI 45 IINNADI 45 IINNADI 45 IINNADI 45 IINNADI 47 INFRSDA 44 VRFSDA 44 VRFSDA 44 VLQLYR 45 INNKLET 23 AKKKIET 24 IIQKIYE 24 IIQKIYE 25 INAKLIT 22 INAKLIT 22 INAKLIT 22 INAKLIS 22 INAKLIS 22 INAKLIS 22 INAKLIS 22 INAKLIS 22 INAKLIS 24 ILPUDIS	x-3 S HHHHHEE ICRPNSTUTY ACRPNSTUTY LTSHRSTUTY LSSRDTLVYY LSHGDVVYY LEHGNICTF FANLTDFLIF VMXSVDFLIF VMXSVDFLIF VMXSVDFLIF VMXSVNSFICI LEKEQSVLIF LEKEQSVLIF ILKEQSTUF NLSKANSFICI LEKEQSTUF NLSKANSFICI LEKEQFLFUC VISSFLFUC UISSFLFUC QLIKQNA-HVLFI LMEDHCVFF DLISKSTUF DLISKSTUF DLISKSTUF DLISKSTUF DLISKSTUF DLISKSTUF DLISKSTUF LFLSIIFI SMIKSIIFI NLTESPIFI SMIKSIIFI MARVE-FADVIT RAFK-LAEKADVIT RAFK-LAEKADVIT RAFK-LAEKADVIT AGAID-AISKADTI CAID-AISKADTI CAID-AISKADTI SIRE-DILCCULI STRE-DILCCULI	tit EEES GYGG GYGGG GWGC GYGGG GYGGG GYGGG GYGGG GYGGG GYGGG GYGGG GYGGG GYGGG GYGGG GYGGG GYSGG GY GY GY GY GY GY GY GY GY GY GY GY G	He E*.*HHHHH FGE INRVIRDM FGE INRVIRDM IGEH R. LIKRMGT FAEQI I LIQRMGT FAEQI I LIQRMGT FGO I LIQRMGT FGO I LIQRMGT FGO INFRINHQ FNE VQPKLMK FLE ITDIVORS FGI Y INRIIGA GGI Y INRIIGA GGI Y INRIIGA GGI Y INRIIGA FDDTIQALFLAN VDDTIQALFLAN VDDTIQALFLAN VDDTIQALFLAN VDDTIQALFLAN CSDMNIRLLHRI FTDPNIVRYLHKL L-SDMNIRLLHRI FTDPNIQIISI ISDMNIRLLHRI FTDPNIQIISI ISDMNIRLLHRI FTDPNIQIISI ISDMNIRLLHRI FTDPNIQIISI ISDMNIQIISI ISDNVVSITSI ISGLYPAASLPL TSGLYPAASLPL TSLKVAPVSEIVN TSLKVAPVSEIVN	14 Str5 HHHHEEEEEEE L-TIPSTHLVINA 1 GIQRVAVSVFGGS GINRVAVSVFRGD PFRMAVSVFRGC FPKPMAVSVFRGC PFRMAVSVFRGC RPQQ-LSIAIRPLG KFKQLFISIVGNH LVKPQ-LSIAIRPLG KFKQLFISIVGNH LVKNEKFTIEIVD CCRQGAPVTIITY CIRQ-NPENITIITY LVN-PSLHIVICY LVKTSDLEHFML LSVPRKVDVEWSG FDYKMKPSFIIDK WSIWSDLEHFML LSVPRKVDLEHYML KKYSDLEHFML LSVPRKVDLEHYML RSY-SENQKEHYCI RILD-DENISNYWL RASY-SENQKEHYCI RASY-SENQKEHYCI TACL-QQENIDKLKN IMML-PQEKVNELSN SSNLPYEISES SSQLPKEDKKS SSQLPEDSKS SSQLPEDSKS SSQLPEDSKS SSQLPEDSKS SQLPKEDKKS VKE-SGGTVIEVNV AAKE-SGATLINVNP VKE-SGGVIEVNVNL WPSHVPQULI
Secondary Structure XAC2443_Xca_21243176 bl11926_Brja_27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530_Neu_30250449 Pflu2135_Pfl_2305957 PP1406_Pput_26988140 LA1794_Lepin_24214494 LA1852_Lepin_24214552 NE2094_Neu_30250034 seal0_Seen_38201747 EF2349_Efae_29376850 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VVA0689_Vivul_37676349 VP1802_Vipa_28898576 PG1107_Porgi_34540840 Sir2_Smel_16264652 AK024756.T_H8_10437125 FLJ2103_H8_1337521 FLJ2103_H8_1337521 FLJ2063_H8_15489177 LJ1095_Laj0_42519021 RPA0819_Rhpa_39933896 CV1059_Chvi_34496514 TDE0266_Tden_4252782 MBNC211202_Mesp_45916072 usg_Pepe_1322457 Aple047201_Acpl_32033900 B0483_Bobr_3359473 pX02-73_Ban_10956463 BC1912_Bca_30020052 BC1271_Bce_30020052 BC1271_Bce_30020052 BC1271_Bce_30020052 BC1271_Bce_30020052 BC1271_Bce_30020052 BC1271_Bce_3020052 BC1271_Bce_3020052 BC1271_Bce_3002052 BC1271_Bce_3002052 BC1271_Bce_30252 BC1271_Bce_30252 BC1271_Bce_30252 BC1271_Bce_30252 BC1271_Bce_30252 BC1271_Bce_30252 BC1271_Bce_30252 BC1271_Bce_30252 BC1271_Bce_30252 BC1271_Bce_30252 BC1271_Bce_30252 BC1271_Bce_30252 BC1271_Bce_30252 BC1271_Bce_30252 BC1271_Bce_30252 BC1271_Bce_30252 BC1271_Bce_30252 BC1271_Bce_30252 BC1271_Bce_30252	H2 HHHHH LIEAGAELAGLHLLDRFLGAL VYWMITYGLRVUDGHRFKD TLYWAMLLFNAANGSWFKDA TLYWAMLLFNAANGSWFKDA LNYWALQHQGEVIDDLFNGFD TLYWAMLD	TP 19 GEPRYLEG NP 19 GEPRYLEG CF 16 GLYREGT FH 12 RPYGHAAG MAS 1 DLCESHTG FF 12 RPYGHAAG SD 7 TWDIGNTN DN 21 KAYFKNRF AT PDELTCSR FR 4 PNEITSSR KR 17 DNYANELP SIR 17 GEARFGHY ID 11 LERKYDNN SE 17 RYTGRYNT RD -1ASAFAG VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -EWAQEKR VL -ENAQEKR VL -ENAQEKR SG -IASTRANG GG -LFVKSND GG -LFVKSND GG -LFVKSND GG - FFDDTIG 	Str-3 EEEEEEE. VARFTKIHGSI VARFTKIHGSI WARFKGSI MTIVFYPHGSI TSLVFYPHGSI ATLVFYPHGSI GANIVKIHGGI GANIVKIHGGI GANIVKIHGGI GANIVKIHGGI RVNIKVGHGSI TINLVKCHGSV INMYLYKIHGSI TINLVKCHGSV VTRIKYHGDF KLSVLHHGVY KLSVLHHGVY KLSVLHHGVY KLSVLHHGVY KLSVLHHGVY KLSVLHHGVY MYGIKHGDF DAVIKHGDF DAVIKHGDF DAVIKHGDY DAVVXKHGDY DAVVXKHGDY DAVVXKHGDY DAVVXKHGDY DAVVXKHGDY DAVVXKHGDY DAVVXKHGDY MELXIHGSS SRVIHHGSI SRVIHHGSI SRVIHHGSI SRVIHHGSI SRVIHHGSI SRVIHCSN SKVICLHGNY SKVICHGNY SKVICHGN	Heli HHHH 58 FRDLAA 56 FRDLAA 56 FRDLAA 56 TVVREV 59 TNVYEEV 59 TNVYEEV 57 ISFCTQO 55 LNRGLRS 86 KLKLEED 51 LNRGLRS 86 KLKLEED 19 LRFLSYE 84 WUDYLOF 7 FRRFSEA 48 YWDYLOF 7 FRRFSEA 48 YWDYLOF 40 VRFRSDA 44 EIQKLYE 44 EIQKLYE 42 EIQKLYE 42 EIQKLYE 42 EIQKLYE 42 EIQKLYE 43 FFTKLSS 42 IGSKLKD 42 FFTALG 42 FFTALG 42 FFTALG 42 FFTALG 42 FFTALG 43 LDTSLLE 43 LDTSLLE 43 LDTVNIK 44 EPVKNIY 44 EVKNIY 44 EVKNIY 45 CM	x-3 S HHHHH5EE ICRPNXVVTY LTSHRSTUTY ACRPNXVVTY LTSHRSTUTY LSHGDTLVYY LPALGESLVYY LASORSTUTY ANNASYLCI ANNASYLCI AINAANSFLCI LEKEQSVLIAF LKEQTTLFVN ALSEAEIILFVN ALSEAEIILFVN MKSLFUCC TKSFLFV QLIKON-HVLFI ILATQTIFI DLGKSFLFV QLICONKSVLFI DLGKSFLFV QLICON-FFLFL DLGFFLFL DLGFFLFL DLGFFLFL DLGFFLFL SIRQTIFI DLGFFLFL SIRSPIFFL SIRSPIFFL SIR-LSREADVI EAG-LSREADVI SIRE-DILECDLLI SIRE-DILECDLLI SIRE-DILECDLLI SIMSSILFI FAMBURE	ti4 EEEE GYSG GYGG GYGG GYGG GYGG GYGG GYGG	He E*.*HHHHH FGE INRVIRDM FGE INRVIRDM FGE INRVIRDM FGE INRVIRDM IGHDR LIKRWGT FAEQI ILQRWGT FAEQI ILQRWGT FAEQI ILQRWGT FAEQI ILQRWGT FGQ ILQRIAN FDE IQPNIAMS FGQ INFRINHQ FNE IQPNIAMS FGU INRIIGA FGU INRIIGA FGU INRIIGA FGU INRIIGA FGU INRIIGA FGU INNILGA FGU INNILGA FSGAU FAALIPA TSLVY FAALIPA FSLVY FAALIPA FSGU FAALIPA FSGU FAALIPA	14 Str5 HHHHEEEEEEE L-TIPSTHLVVIS 1 L-TIPSTHLVIMA 1 GIQRVAVSVFGGS GINRVAVSVFFGD PFKRMAVSVFTGQ KPKTVAISIYPRS RPQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQGAPVTIITY CIRQNFHTIITY L-NN-PNLLVFIFCY 1 L-NN-PNLLVFIFCY 1 L-NN-PSLHIVICY LKVTPLHVVEWSG FDYKNKPSFIIDK WSIWRSGHEKDR VKHKSDLEHFML LYSVPNKVDLEHYML SLNL-KKXHAMQLRD IVDY-ENHQSGKPAP RKQM-RGLHKQAYIV WGTWEDSG RNIL-DENISNNYWL RASY-SENQKEHYCI TRCL-DQDNIHKLKD TACL-DQENIDKLKN IML-PEKYNELSN SENLPFDISES SSQLPKEDKKS IVK2-SGGTLIENNP YKE-SGGTLIENNP YFFRGKNLVLINK ITVR-SGGKLVIVNL MVP-SHVPQULINR MVP-SHVPQULINR
Secondary Structure XAC2443 Xca 21243176 bl11926 Brja 27377037 XAC2189_Xax 21242924 BRA0375 Brsu 23500128 NE2530 Neu 30250449 Pflu2135 Pfl 23059957 PP1406 Pput 26988140 LA1794 Lepin 2421452 NE2094 Neu 30250034 seal0 Seen 38201747 EF2345 Efae 29376850 VVA0689_Vivul 37676349 VP1802_Vipa_2898576 Pfl107 Progi 34540840 Sir2_Smel 16264652 AK024756.1 Hs 10437125 FLJ2103 Hs 1337571 FLJ20635_Hs_15489177 LJ1095 Laj0_42519021 RPA0815 Rhpa 3993896 CV1059 Chvi 34496514 TDE0266_Tden 42525782 MBNC211202 Mesp 45916072 usg Pepe 1322457 Aple047201 Acpl 32033900 BB0483 Bobr 33599473 pX02-73 Ban 10956463 BC1912_Bce 30019423 BK5-Tp30 kK5-t 14251154 SAG1992_Saga 22538128 Plu4324_Phlu 37528154 lici Aful 14278228 PA08010 Hab 14521406 APE1782_Ape_14601621 aq 2170_Aae_15607106 CPE0256 Clpe 18309238 TM0490_Tma 15643256 HS1126_S6320163 SS02478 Ss0_15899220 cobb_Ec_26247264	H2 HHHHH LIEAGAELAGLHILDRFLGAL LIEHGCFAGLRIIDRFVGAL VYWMITYGLNUQDGHRFKD TLYWAMLLFNANGSWFKDA LNYWALQHQGEVIDDLFNGPD LLPWAVGHAFSGFAELFDEQG LLYWTIMQDEITPTFTCDDGFRNPD YLELLCIDYDIPATYGGTBUFFRQL LAEYACDQSGTHLYTGFTBGFRGF LLYWFIGSRIHYTGFTBGFSGTH LLYNSFISGINLINGFSGTH LLYNSFISGINLINGFSGTH LLYAADQGKQLESLDLTDEKK LLELYAADQGKQLESLDLTDEKK LLELYAADQGKQLESLDLTDEKK LLELYAADQGKQLESLDLTDEKK LLELYAADQGKQLESLDLTDEKK LLELYAADQGKQLESLDLTDEKK LLELAFGR	TP 19 GEPRYLEG CF 14 QPIRGES CF 14 QPIRGES FH 12 RPYGHAAG MAS 1 DLCESHTO ST DLCESHTO TP ST DLCESHTO TP ST TP DELTCSR FR 4 PNEITSAR KR 17 DIAPATSAR SID 11 LERKYDNN SID 11 LERKYDNN SID 11 LERKYDNN SID 13 RAPGRY VL -EWAQEKR VL VL -EWAQEKR QD SID -ISKIDNG RL SID -ISKIDNG RD GL -STQKYDR SID GL -LFSHIDKA GG GG -LFVSNDD GG GG -LFVSTINLGA GG GG -LFVSTINLGA GG GG -LFVKSND GG <	Str-3 EEEEEEEE. VARFTKLHGSV VIRLTKLHGSL MTI.VFYPHGSL ATLVFYPHGSL ATLVFYPHGSL GANIVKLHGGL GANIVKLHGGL SONIFYLHGAL GANIVKLHGGL SONIFYLHGAL RVNIWKVHGSL RVNIWKVHGSL RVNIWKVHGSL RVNIWKVHGSL RVNIWKVHGSL FGYYLHLHGSF FGYYLHLHGSF FGYYLHLHGSF KLSVLHHGVY KLSVLHHGVY KLSVLHHGVY KLSVLHHGVY KLSVLHHGVY KLSVLHHGV DAVYKHGDF DAILYKHGCS VPCIVKHGDF DAILYKHGCS VMELYKIHGCS WELYKIHGSS SKNVFLHGNI SKKVFLHGNI SKKVFLHGNI SKKVFLHGNI SKKVFLHGNI SKKVFLHGNI SKKVFLHGSV SKKVFLHGSV SKKVFLHGSV	Heli HHHH Stephenson FRDJAA Se FRDJAA Se FRDJAA Se ATVYREV Se TIVYREV Se INNGLS Se STVYREV Se INNGLS Se KLKDED Se INNGLS SE INNADI SE TIHNADI SE TIHNADI S	x-3 S HHHHHEEE ICRPNSIVIY AGRPNAVVVIY LTSHRSTLTLF HKSQRDTLVYY LEHGDNLCTF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDPLIF FANLTDFLIF FANLTDFLIF FANLTDFLIF NESCSTLT NESCSTLF NESCSTLF NESCSTLF NESCSTLF DLISKSFLF QLIKON-HVLF NESCSTLF DLISKSFLF DLISKSFLF DLISKSFLF DLISFFLF SIKSSIFF FAREX-LSREADVI EAGL-ARSDVMI EAGL-SRASLMT TRK-DILECDLI EAGL-ARSDVMI EAGL-ARSDVMI EAGL-SRASLMT TRK-DILECDLI EAGL-ARSDVMI E	ti4 EEEE GYSG GYGGGWSGGWSGGGWSGGGGGGGGGGGGGGGGGG	He E*.HHHHH FGEINRVIRDM IGEHRALLKRMYGT IGEHRALLKRMYGT IGEHRALLKRMYGT IGEHRALLKRMYGT IGEHRALLKRMYGT IGENCIC IGENCIC IGENCIC IGENCIC IGENCIC IGENCIC ICON IC	14 Stt5 HHHHEEEEEEE L-TIPSTHLVINS 1 L-TIPSTHLVINS 1 GIQRVAVSVFGGS GINRVAVSVFGG PFKRMAVSVFTGQ KPKTVAISIYPRS RPQQLSIAIRPLG KFKQLFISIYGNH LVKNEKFTIEIVD CQRQ-GAPVTIITY CIRQ-NTPITIITY L-NN-PSLUVFIFCY 1 L-NN-PSLUVFIFCY 1 L-NN-PSLUVFIFCY 1 L-NN-PSLUVFIFCY 1 L-NN-PSLUVFIFCY 1 L-NN-PSLUVFIFCY 1 L-NN-PSLUVFIFCY 1 L-NN-PSLUVFIFCY 1 L-NN-SGEHEXR WKHKSDLEHFML ISVPNKVDLEHYML SLNL-KKYALMQLRD IYDY-ENHQSGKPAP RNUWRDSGHEXR NUCWRDSG RNIL-DENISNNYWL KKSDLEHFML IYDY-SENQCHIVCI TRCL-DQENIDKLKN TML-PQEVNELSN SSQLPFEISEA AENSPFISES SSQLPFEDISES SSQLPFEDRKSS IVNP-RGGAIIEINP IVKE-SGGTVIEVNV AKE-SGATLINVNP VAKE-NGAQVIEVNP YFF-RGKLVLINK TVR-SGGKLVINL MYP-SHVPQULINR MYP-SHVPQULINR
Secondary Structure XAC2443 Xca 21243176 bl11926_Brja_27377037 XAC2189_Xax_21242924 BRA0375_Brsu_23500128 NE2530_Neu_30250049 Pf1401-Ppt1_26988140 LA1794_Lepin_24214552 NE2094_Neu_30250034 seal0_Seen_38201747 EF2349_Efa=29376850 VVA0689_Vivul_37676349 VP1802_Vipa_2898576 PG1107_Porgi_34540840 Sir2_Smel_16264652 AK024756.1_Hs_10437125 FLJ20635_Hs_15489177 LJ1095_Lajo_42519021 RPA0819_Rhpa_3993896 CV1059_Chv1_34496514 TDE0266_Tden_42525782 MBNC211202_Mesp_45916072 usg_Pepe_1322457 Aple047201_Acpl_32033900 BB0483 Bobr_33599473 px02-73_Ban_10956463 BC1912_Bce_30019423 BK5-Tp30_hK5+t_14251154 SAG1992_Saga_22538128 Ph43034_Ph1u_37528154 lici_Afu1_14278228 PAB0801_Pab_14521406 APE1782_Ape_14601621 aq_2170_Aee_15007106 CPE0256_Clpe_18309283 TM0490_Tma_15643256 HS11_Sc_6320163 SS02478_Sso_15899220 cobB_Ec_26247264 TSC_056_Hs15899220 cobB_Ec_26247264 TSC_056_Hs169328 SIRT5_Hs_6912664	H2 HHHHH LTEAGAELAGLHLLDRFLGNL LTEHGCPEGLNUQDGHRFKD VYWMITYGLNUQDGHRFKD TLYWAMLLFNANGSWFKDA LNYWALQHQGEVIDDLFNGPD LLPNAVQHPGTABLFDEQG LLYWTIMQDEITPTFTCDDGFRNPD YLELLCIDYDIPATYGGTBUFFRQL LAEYACDQEGTHHYTGFTHGFFRQL LAEYACDQEGTHHYTGFSGH LLYMSFIDNDLVDGYNGSL LIESFNQTSDINLINGFSGTH LLYAADQGKQLESLDLTDEKK LLEJYAADQGKQLESLDLTDEKK LLEJYAADQGKQLESLDLTDEKK LLEJYAADQGKQLESLDLTDEKK LLEAFKGRRNRKVDVKCRET FFEEECKATFQSRPNILKDIN FFEEECKNRKVDVKCRET VIETALKESDIKYTGKAYKISHV LLELGFKNNRKVDVKCRET LIESSFRP		Str-3 EEEEEEEE. VARFTKLHGSV VIRITKLHGSV VIRITKLHGSL MTIVFYPHGSL ATIVFYPHGSL GTRVLHGGL GTRVLHLGGL GTRVLHGGL GTRVLHGGL GTRVLHGGL TINLVKCHGSL TINLVKCHGSL TINLVKCHGSL TINLVKCHGSL FGYYLHLHGSL FGYYLHLHGSP FGYYLHLHGSP FGYYLHLHGSP FGYYLHLHGSP DFVLFKLHGCS KKSVLHGDF DFVLFKLHGCS VTQIKXHGDF DAVYXHHGDY DAVYXHHGDY DAVYXHGDF DAVYXHGDF DAVYXHGDF DAVYXHGDF MSELYKLHGSS VGELYKLHGSV YGELYKLHGSV SKRVILHGNI SRRVIELHGNI SKRVIELHGNI SKRVIELHGNI SKRVIELHGNI SKRVIELHGNI SKRVIELHGNI SKRVIELHGNI SKNVIELHGSL	Heli HHHH 58 FRDLAA 56 FRDLAA 56 FRDLAA 56 ATVYREV 57 INVYEV 57 INVYEV 57 INVSEV 57 INRGLS 53 LSWCLQ 55 INRGLS 53 LSWCLQ 55 INRGLS 54 INRADL 54 INRADL 54 INRADL 54 INRADL 24 VIQUIYS 24 VIQUIYS 24 VIQUIYS 24 VIQUIYS 24 VIQUIYS 24 VIQUIYS 24 VIQUIYS 24 VIQUIYS 24 INKKIS 22 INKTS 22 INKTLS 22 INATLS 24 INATLS 24 INATLS 25 INATLS 25 INATLS 25 INATLS 26 INATLS 26 INATLS 27 INATLS 27 INATLS 28 INATLS 28 INATLS 29 INATLS 20 INATLS 20 INATLS 20 INATLS 20 INATLS 20 INATLS 20 INATLS 21 INATL	x-3 S HHHHHEE ICRPNSTUTY ACRPNSTUTY LTSHRSTUTY LSACGDTUVY LEHGDNLCTF PANLTDPLIF PANLTDPLIF PANLTDPLIF PANLTDPLIF PANLTDPLIF TAOENNSFLCI LEKEQSVLIAF ILSEANSFLCI LEKEQSVLIAF ILSEANSFLCI LEKEQSVLIAF NESKANSLII LEKEQSVLIAF NESKANSLII NESKANSFLCI NESKANSFLCI NESKANSFLCI DISFIFLCC NKSFIFLCC NKSFIFLCC DISFIFLCC DISFIFLCC DISFIFLFI DISFIFLFI DISFIFLFI DISFIFLFI DISFIFLFI DISFIFLFI DISFIFLFI DISFIFLFI DISFIFLFI DISFIFLFI DISFIFLFI DISFIFLFI DISFIFLFI DISFIFLFI DISFIFLFI NEVESPIFL NEVESPIFLFI NEVESPIFL NEVE-SSIFI RAMER-LSEADANI SIRE-DIECDLI SIRE-DIECDLI ELS-ANSDIN	ti4 EEES GWG GWG GWG GWG GYG GYG GYG GYG GYG GYG	He E*.*HHHHH FGE INRVIRDM IGEH R. LIKRMYGT IGEH R. LIKRMYGT IGEH R. LIKRMYGT IGEH R. LIKRMYGT IGEN C. LIKRMYGT ILQRMNGT FAEQUI LLQRMNGT FAEQUI LLQRMNGT FAEQUI LLQRMNGT FOR C. UNRILL FNE IQPNIMAK F	14 Stt5 HHHHEEEEEEE L-TIPSTHLVINS 1 L-TIPSTHLVINA 1 GIQRVAVSVFGGS GINRVAVSVFRGD PFKRMAVSVFTGQ KPKTVAISIYPRS RPQQLSIAIRPLG KFKQLFISIGNH LVKNEKFTIEIVD CQRQ-GAPVTIITY CLRQNFPITIITY L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFIFCY 1 L-NN-PSLLVFFCY 1 L-NN-PSLLVFFCY 1 L-NN-PSLVVEWSG KNISDLEHFML XSIWKKYALMQLRD IYDY-ENHQSGHEXDR KKSDLEHFML SINL-KKYALMQLRD IYDY-ENHQSGKPAP RAQM-RGLHKQAYIV KKWRDSG RNIL-DENISNNYWL RASY-SENQKEHYCI RIQF-AQDQRQHYCI RIQF-AQDQRQHYCI RAKE-SGGTVIEVNV AAKE-SGGTVIEVNV VKQ-RGGAIIEINP IVK2-RGKLVIVNL MYPSHVPQULINR MYP-SHVPQULINR EAKL-HGAHTVELNL IAFA-SGGXLVEVNP

	H5		56		Helix-6				
Secondary Structure	НННН		EEE		-нннннннннннн				
AC2443 Xca 21243176	PQISLMIGPA	29	QRWHTAP		PVQKAEGDGSKDDLGI	460\Predicted	Sir2	like	nuclease
oll1926 Brja 27377037	AQVSLLIGPH	29	HRPGDPK		LNDPATVPLSATATGP	4331			
AC2189 Xax 21242924	QAYCNYAYQV	7	VRVDFFD		CESQGCWMRATPPPLQ	3591			
RA0375 Brsu 23500128	QAYCNRVFQI	8	VEVEFFN		SESPGCWIHPPHQEPR	3381			
IE2530 Neu 30250449	QAFCLQVLKA	7	TEVTFFD		SRSPGCWNNP	3371			
2flu2135 Pfl 23059957	AAFIQHQKRH	9	VELRFFD		SKSHALGHPKLSVPVE	3581			
P1406 Pput 26988140	EASVINGKOH	10	TAVHFFD		ASTHPLGQAALAIEVP	3321			
A1794 Lepin 24214494	NORLINRAEK	10	LELNFFN		AETANVWG	3381			
A1852 Lepin 24214552	PSFVEQFDYD	28	MKKIYSQ		REKIRSTVRKSYFK	4191			
E2094 Neu 30250034	DSTKKLILGG	15	QSVVYSS		LSSSSFTVEKNIWSLE	3301			
seal0 Seen 38201747	DAARKLILDG	15	OSIVYSS		LDKTPLIVEKNIWSLE	3301			
F2349 Efae 29376850	GANAVPSNIV	54	AFNRIIE		GNLSEKYIDSOYTEVV	4551			
VA0689 Vivul 37676349	ELAELCANSO	41	KPVLFPK		NNAAKVIAEAIODLGS	4161			
/P1802 Vipa 28898576	AGEOVO		RETYWAN		CLGRSVVVERMDNISD	512/			
G1107 Porgi 34540840	AGASVETEAO		KINAKIT		KESTERLDKTWFV	469			
ir2 Smel 16264652	PPSYVEMYRP	8	RNWGVTV	9	OALLAFTOGLEAFLAG	272			
K024756.1 Hs 10437125	VRRGDVDEFK	10	TKVISYG	9	FKRLTCEISTRGTSAG	329			
T.121103 Hs 13375721	VRRGDVDEFK	10	TKVISYG	9	FKRLTCEISTRGTSAG	329			
T.T20635 Hs 15489177	VIKENEDHEE	10	TKVVSVG	9	VODLATOICKOOSPDA	314			
.T1095 Laio 42519021	DISEDOANLY	9	TETTWYG	9	LOKMNKDISDKYOEKH	336			
PA0819 Rhpa 39933896	RFFSLOFOML	10	GIVALOP	9	SESLSLANSLGELIAR	330			
V1059 Chvi 34496514	TPFENETEKE	10	DGTYFIK	à	SHEIKDSFIEFSARLL	294			
DE0266 Tden 42525782	YGGVOPKSFF	10	ATLEOWG	9	NPOKALEVELAKLTT-	263			
IBNC211202 Mesp 45916072	YEKDRPPSEV	10	AVLARWG	9	SAOFALGTFLKHLLOM	269			
ISG Pepe 11322457	EKRVTKPSDI	10	EVDRTKO	9	YGIOTVYIDSYEEITE	273			
ple047201 Acpl 32033900	LRKVOKEDNE	10	KOOLEIS	9	LLIDEYSEITEILBEV	273			
B0483 Bobr 33599473	LRKASRMEGE	10	KEDLFKH	9	VYVDEFSEITDILRET	274			
X02-73 Ban 10956463	RLIEVERAGO	10	LTIGKIT	9	DYEKIYNALAONKEKE	304			
3C1912 Bce 30020052	RLIFVERAGG	10	TTSNRVT	9	NYELVYKALSKNKRKF	304			
3C1271 Bce 30019423	RIWFVRRNEH	10	NLGEGLE	9	DESNLYNTLSTNKTOK	301			
3K5-Tp30 bk5-t 14251154	AARIGVVEYT	10	SNIPDLG	9	NYKKIYDEISOIEOGY	312			
AG1992 Saga 22538128	AOKIGVVEYL	10	SSLPDLS	9	NETNIYRLISKINOGE	305			
1u4324 Phlu 37528154	AERITLIEFK	10	RTDDKSO	9	NYKKTYDEISNIDEGL	312			
ici Aful 14278228	DETPL/TPIA-	2.0	-DYSLRG		KAGEVMDELVRHVRKA	254			
AB0801 Pab 14521406	EESATTPIA-		-DFFLRG		RAGEVLPRVVHEVRRI	248			
PE1782 Ape 14601621	EPNRYSGVA-		-DIELRM		RAVEFAERLSRAMGID	246			
g 2170 Aae 15607106	EETPITKIA-		-DMHFKE		KASTGLKKVYDYLREK	234			
PE0256 Clpe 18309238	SSTOADSKA-		-DLVIND		STGKVLGKVID	244			
M0490 Tma 15643256	GETPEDDIA-		-TLKYNM		DVVEFARRVMEEGGIS	246			
IST1 Sc 6324504	DMVTHAEF		-DLNLLG		FCDDVASLVAKKCHWD	468			
sir2p Sc 6320163	DPVKHAEF		-DLSLLG		YCDDTAAMVAOKCGWT	527			
SO2478 Sso 15899220	EETPLDSTA-		-DYVVRE		PVETSIPKTLENVROK	244			
obB Ec 26247264	EPSOVGNEF-		-AEKYYG		PASOVVPEEVEKLLKG	272			
TC1026 Thth 46199328	EPTPLTPLA-		-HLSLRT		GAVEGMALLLPPSPED	247			
SIRT5 Hs 6912664	ETTPATNEF-		-RFHFOG		PCGTTLPEALACHENE	307			
Consensus/80%			11111111111		h				
Contraction of Miles 1, Miles of Miles									

Figure 7. Multiple alignment of the predicted Sir2-like nuclease. The coloring reflects 80% consensus and the consensus abbreviations, coloring scheme and secondary structure designations are as in Figures 2 and 3. The histidine and aspartate residue conserved in the predicted nucleases are shaded red. Secondary structure elements are numbered according to their position in the core Rossmann fold. Helix 0.1 and 0.2 reflect helices that are synapomorphic to the Sir2-clade. Species abbreviations are as follows: Aae, *A.aeolicus*; Acpl, *Actinobacillus pleuropneumoniae*; Aful, *A.fulgidus*; Ape, *A.pernix*; Ban, *Bacillus anthracis*; Bce, *B.cereus*; bk5-t, Lactococcus phage bk5-t; Bobr, *Bordetella bronchiseptica*; Brja, *B.japonicum*; Brsu, *Brucella suis*; Chvi, *Chromobacterium violaceum*; Clpe, *Clostridium perfringens*; Ec, *E.coli*; Efae, *E.faecalis*; Hs, *Homo sapiens*; Lajo, *Lactobacillus johnsonii*; Lepin, *Leptospira c*; Mesp, *Mesorhizobium* sp.; Neu, *N.europaea*; Pab, *P.abyssi*; Pepe, *P.ediococcus pentosaceus*; Pfl, *Pseudomonas fluorescens*; Phlu, *Photorhabdus luminescens*; Porgi, *Porphyromonas gingivalis*; Pput, *Pseudomonas pulida*; Rhpa, *Rhodopseudomonas palustris*; Saga, *Streptococcus agalactiae*; Sc, *S.cerevisiae*; Seen, *Serratia entomophila*; Smel, *S.meliloti*; Sax, *Xanthomonas axonopodis* and Xca, *Xanthomonas campesteris*.

Many members of the FtsK–HerA superfamily contain membrane-spanning regions that probably anchor them to the cell membrane during DNA pumping. No such membranespanning regions are present in the core orthologous set of archaeal HerA proteins. The contextual association (albeit weak) of HerA and the highly conserved small membrane proteins typified by MJ1617 (COG2034) implicates these proteins as potential candidates for the role of a membrane tether for HerA. In the case of other HerA ATPases, additional, poorly conserved membrane proteins might function as their partners. However, the absence of membrane-spanning regions in HerA proteins themselves or conserved genes for membrane proteins in the predicted *herA* operons raises the possibility of fundamental functional differences between HerA proper and the rest of the FtsK–HerA superfamily ATPases.

Additional nuclease connections of the FtsK–HerA superfamily and prediction of a novel nuclease with the Sir2 fold. Several previously described conserved gene neighborhoods of VirB4, VirD4 and FtsK encode components of the T4SS of proteobacteria or the ESAT-6 system of Gram-positive bacteria and actinomycetes (17,87). However, in other conserved gene neighborhoods, FtsK-HerA superfamily ATPases are encoded together with nucleases involved in DNA processing. In particular, genes for ATPases of the TrwB/TrsK and the TraG families of the VirD4 clade are found in operons that also contain genes for conjugative relaxases of the TrwC and TraA families, respectively (Figure 5B) (65,88). These relaxases have an N-terminal nuclease domain of the rolling circle replication (RCR) fold combined with a C-terminal SF-I DNA helicase domain. The TraA relaxases belong to the RCR superfamily proper, with a HXH active site motif and a catalytic tyrosine (89,90), whereas the TrwC relaxase domain shows an evolutionarily distant, circularly permuted version of the fold [(91); L. Aravind, unpublished data]. Thus, at least on two independent occasions, VirD-like ATPases appear to have been combined with distinct members of the RCR nuclease fold in conserved operons.

The VirB4-like ATPases of the YddE family encoded by conjugative TN916 transposons often co-occur with genes for a large membrane protein with six transmembrane regions (YddG), a hydrolase of the NlpC/P60 superfamily (92), a smaller membrane protein with a single transmembrane region and a catalytic tyrosine containing relaxase of the pT181-Rep domain superfamily, which is unrelated to the RCR superfamily relaxases (Figure 5B). Using iterative sequence database searches with the PSSM for this relaxase family, we showed that they are homologous to the nicking enzyme of the filamentous bacteriophages, such as M13 and f1. In these phages, the nicking enzyme functions in conjunction with the packaging ATPase that also belongs to the FtsK-HerA superfamily. Thus, as in the case of the RCRs with the HXH motif, the pT181-Rep relaxases have also formed multiple, independent functional associations with ATPases of the FtsK-HerA superfamily. NlpC family hydrolases encoded by the adjacent ORFs in these transposons are likely to facilitate local degradation of the cell wall, whereas the transmembrane proteins are likely to be components of the conjugation tube through which the DNA of the conjugative transposons is pumped by YddE after it is processed by the associated relaxase (Figure 5; Supplementary Material). Thus, persistent operonic associations with several unrelated nucleases are prevalent in different clades of the FtsK-HerA superfamily.

This contextual theme was exploited to predict previously uncharacterized nucleases with probable functional links to FtsK-HerA ATPases. Several members of the proteobacterial bll1925 family of the HerA clade, which contains a divergent version of the HAS barrel (Table 1, Figure 5), are encoded next to a conserved, co-directional ORF. This ORF, bll1926, is unrelated to NurA, but iterated database searches using PSI-BLAST showed that it defines a previously undetected protein family, which is distantly related to the Sir2 proteins. Despite their high sequence divergence, the bll1926-like proteins contained all the hallmarks of the Sir2 fold (a variant Rossmann fold), such as the glycine-rich loop at the N-terminus, the central NhD motif (where h is any hydrophobic residue) and the C-terminal HG motif. However, the bll1926 family proteins lack the Zn-ribbon insert characteristic of the Sir2 family and contain a distinct, C-terminal DXH motif which is absent in Sir2 (Figure 7). Members of the Sir2 family deacetylate acetyl-lysines in a variety of protein substrates, a reaction that utilizes NAD and produces 2'-O-acetyl-ADP-ribose (93–96). A superposition of the conservation pattern of the bll1926 family onto the crystal structure of the Sir2 catalytic domain (94-97) suggests that, despite the conservation of the active site residues, the surface involved in peptide interaction in Sir2 is not conserved between the two proteins (Figure 7). Furthermore, the additional DXH motif of the bll1926 family, together with the conserved histidine of the HG motif, forms a potential di-histidine active site configuration similar to those in nucleases or phosphoesterases of the RNAse A and 2H superfamilies (98,99). Given the persistent linkage of the FtsK-HerA ATPases with nucleases, we predict that the bll1926 family proteins are nucleases rather than deacetylases like the Sir-2 proteins. Hence, two very different catalytic activities appear to have emerged within the same fold as a result of recruitment of partially different sets of conserved residues for the active center of Sir2 and bll1926. The apparent horizontal mobility of the bll1925-bll1926 pair in proteobacteria mirrors that of the HerA–NurA gene pair. This observation suggests a close functional parallel between these systems and supports the prediction of the nuclease function for the bll1926 family of Sir2 homologs.

Despite the close functional association with various nucleases, as indicated by the presence of conserved operons, HerA ATPases do not form fused genes with any of these nucleases. There might be a single exception to this trend. A protein from A.aeolicus, aq 1852, consists of a HerA domain and an N-terminal HKD domain, the catalytic module of numerous phosphohydrolases, such as phospholipase D, eukaryotic tyrosine-DNA phosphodiesterases and certain DNAses, such as Nuc (100-102). Thus, the HKD domain fused to the HerA ATPase in aq 1852 could be a DNAse, polynucleotide phosphatase or a tyrosine-DNA phosphodiesterase. The (near) absence of HerA-nuclease fusions is somewhat unexpected because, on many occasions, genes that are part of the same operon in some genomes are fused others (103). The absence of such fusions suggests that FtsK-HerA superfamily ATPases and the associated nucleases might be present in the respective functional complexes in non-stoichiometric amounts.

GENERAL EVOLUTIONARY CONSIDERATIONS AND CONCLUSIONS

The majority of experimentally characterized members of the FtsK–HerA ATPase superfamily are involved in pumping substrates, particularly DNA, through membrane-spanning pores. The two primary clades of this superfamily, HerA and FtsK, show nearly perfect complementarity in their phyletic patterns: predominantly archaeal HerA (the core orthologous set) versus mostly bacterial FtsK. Together with the evolutionary relationship between these proteins discussed here, this suggests that HerA and FtsK perform analogous functions in DNA pumping during cell division. The operonic organization of HerA, NurA, MRE11 and Rad50 that is conserved in most archaea suggests additional players in this process and points to the potential importance in it of doublestrand break repair. The bacterial orthologs of MRE11 and Rad50 do not form operons with FtsK and so far have been implicated only in recombinational repair pathways (104,105). It appears likely that functional association of HerA-NurA with Rad50-MRE11 is an archaeal innovation.

While at least one representative of the FtsK–HerA families is present in each prokaryotic genome, they are practically absent in eukaryotes except for some fungal forms which probably were acquired via relatively late HGT. Given that eukaryotes evolved a mechanism of chromosome segregation that is radically different from the prokaryotic one, this observation lends further support to the conjecture that FtsK and HerA are functionally equivalent enzymes, which are ancestral in the bacterial and archaeal lineages, respectively. Eukaryotes probably lost HerA and its nuclease partner, NurA, concomitantly with the advent of the new segregation mechanism, whereas their functional partners, MRE11 and Rad50, have been retained as essential repair enzymes.

Under the most parsimonious evolutionary scenario, FtsK and HerA descended from a single ancestral ATPase pump that was present in LUCA, along with several other P-loop ATPases. It seems likely that separation of the FtsK–HerA critical early stage in the evolution of life, the origin of a specialized, active mechanism for segregation of daughter genomes during cell division. It is also notable that viral packaging ATPases comprise two of the early branching lineages of the FtsK–HerA superfamily. Thus, DNA packaging into capsid-like structures might have evolved roughly synchronously with chromosomal segregation.

Other conserved bacterial cell division proteins, such as FtsA, MreB and FtsZ (106,107), are not universally represented in all prokaryotic lineages. To date, FtsA is absent in almost all archaea, MreB is absent in all crenarchaea and several euryarchaea and FtsZ is absent in all crenarchaea [V. Anantharaman and L. Aravind, unpublished data; (108)]. These phyletic patterns raise the possibility that the cell septation apparatus in LUCA lacked some of the key extant components. At least part of the septation apparatus, along with the cell wall, might have evolved later than the putative DNApumping complex that included the prototype FtsK-HerA ATPase. Hence, the proto-cells, prior to and including LUCA, probably were relatively simple structures that did not possess a complex apparatus for septation that is seen in extant cells and principally depended on DNA-pumping for daughter genome segregation. Despite functionally similar associations of the FtsK-HerA superfamily ATPases with nucleases, none of the nuclease partners of the FtsK-HerA superfamily ATPases can be traced to LUCA. Given that even the smallest plasmids, conjugative transposons and phages have a nuclease or topoisomerase that functions along with the pumping ATPase of the FtsK-HerA superfamily, the ancestral nuclease might have been displaced during evolution of large cellular genomes. The mechanisms for decatenation of replication products of the larger chromosomes appear to have evolved independently in the archaeal and bacterial lineages, resulting in the independent recruitment of NurA and Xer/ ParCD enzymes, respectively.

Thus, using computational analysis of proteins sequences and structures along with genome context analysis, we predict the central components and the possible mechanism for chromosomal segregation in archaea. The observations described here may help in designing further experiments aimed at dissection of two of the most fundamental biological processes, chromosomal segregation and cell division.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We thank C. Elie and F. Constalinesco for their contributions in the early stages of this work and thank P. Forterre and D. D. Leipe for helpful discussions.

REFERENCES

- Pogliano,K., Pogliano,J. and Becker,E. (2003) Chromosome segregation in eubacteria. Curr. Opin. Microbiol., 6, 586–593.
- Gerdes, K., Moller-Jensen, J., Ebersbach, G., Kruse, T. and Nordstrom, K. (2004) Bacterial mitotic machineries. *Cell*, 116, 359–366.
- Romberg,L. and Levin,P.A. (2003) Assembly dynamics of the bacterial cell division protein FTSZ: poised at the edge of stability. *Annu. Rev. Microbiol.*, 57, 125–154.

- 4. Margolin,W. (2003) Bacterial division: the fellowship of the ring. *Curr. Biol.*, **13**, R16–18.
- 5. Margolin,W. (2001) Spatial regulation of cytokinesis in bacteria. *Curr. Opin. Microbiol.*, **4**, 647–652.
- Chen,J.C. and Beckwith,J. (2001) FtsQ, FtsL and FtsI require FtsK, but not FtsN, for co-localization with FtsZ during *Escherichia coli* cell division. *Mol. Microbiol.*, **42**, 395–413.
- Errington, J., Daniel, R.A. and Scheffers, D.J. (2003) Cytokinesis in bacteria. *Microbiol. Mol. Biol. Rev.*, 67, 52–65.
- Yu,X.C., Weihe,E.K. and Margolin,W. (1998) Role of the C terminus of FtsK in *Escherichia coli* chromosome segregation. *J. Bacteriol.*, 180, 6424–6428.
- Aussel,L., Barre,F.X., Aroyo,M., Stasiak,A., Stasiak,A.Z. and Sherratt,D. (2002) FtsK is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases. *Cell*, **108**, 195–205.
- Sherratt, D.J., Soballe, B., Barre, F.X., Filipe, S., Lau, I., Massey, T. and Yates, J. (2004) Recombination and chromosome segregation. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, **359**, 61–69.
- Barre,F.X., Soballe,B., Michel,B., Aroyo,M., Robertson,M. and Sherratt,D. (2001) Circles: the replication-recombination-chromosome segregation connection. *Proc. Natl Acad. Sci. USA*, 98, 8189–8195.
- 12. Donachie, W.D. (2002) FtsK: Maxwell's demon? Mol. Cell, 9, 206–207.
- Errington, J., Bath, J. and Wu, L.J. (2001) DNA transport in bacteria. *Nature Rev. Mol. Cell. Biol.*, 2, 538–545.
- Ip,S.C., Bregu,M., Barre,F.X. and Sherratt,D.J. (2003) Decatenation of DNA circles by FtsK-dependent Xer site-specific recombination. *EMBO J.*, 22, 6399–6407.
- Massey,T.H., Aussel,L., Barre,F.X. and Sherratt,D.J. (2004) Asymmetric activation of Xer site-specific recombination by FtsK. *EMBO Rep.*, 5, 399–404.
- Espeli,O., Lee,C. and Marians,K.J. (2003) A physical and functional interaction between *Escherichia coli* FtsK and topoisomerase IV. *J. Biol. Chem.*, 278, 44639–44644.
- 17. Pallen, M.J. (2002) The ESAT-6/WXG100 superfamily—and a new Gram-positive secretion system? *Trends Microbiol.*, **10**, 209–212.
- Constantinesco, F., Forterre, P., Koonin, E.V., Aravind, L. and Elie, C. (2004) A bipolar DNA helicase gene, herA, clusters with rad50, mre11 and nurA genes in thermophilic archaea. *Nucleic Acids Res.*, 32, 1439–1447.
- Moncalian,G., Cabezon,E., Alkorta,I., Valle,M., Moro,F., Valpuesta,J.M., Goni,F.M. and de La Cruz,F. (1999) Characterization of ATP and DNA binding activities of TrwB, the coupling protein essential in plasmid R388 conjugation. *J. Biol. Chem.*, 274, 36117–36124.
- Hamilton,C.M., Lee,H., Li,P.L., Cook,D.M., Piper,K.R., von Bodman,S.B., Lanka,E., Ream,W. and Farrand,S.K. (2000) TraG from RP4 and TraG and VirD4 from Ti plasmids confer relaxosome specificity to the conjugal transfer system of pTiC58. *J. Bacteriol.*, 182, 1541–1548.
- Dang, T.A., Zhou, X.R., Graf, B. and Christie, P.J. (1999) Dimerization of the Agrobacterium tumefaciens VirB4 ATPase and the effect of ATP-binding cassette mutations on the assembly and function of the T-DNA transporter. *Mol. Microbiol.*, 32, 1239–1253.
- Fullner, K.J., Stephens, K.M. and Nester, E.W. (1994) An essential virulence protein of *Agrobacterium tumefaciens*, VirB4, requires an intact mononucleotide binding domain to function in transfer of T-DNA. *Mol. Gen. Genet.*, 245, 704–715.
- Berger,B.R. and Christie,P.J. (1993) The Agrobacterium tumefaciens virB4 gene product is an essential virulence protein requiring an intact nucleoside triphosphate-binding domain. J. Bacteriol., 175, 1723–1734.
- 24. Kado,C.I. (1994) Promiscuous DNA transfer system of *Agrobacterium tumefaciens*: role of the virB operon in sex pilus assembly and synthesis. *Mol. Microbiol.*, **12**, 17–22.
- Gomis-Ruth,F.X., Moncalian,G., Perez-Luque,R., Gonzalez,A., Cabezon,E., de la Cruz,F. and Coll,M. (2001) The bacterial conjugation protein TrwB resembles ring helicases and F1-ATPase. *Nature*, 409, 637–641.
- Egelman, E.H. (2001) Structural biology. Pumping DNA. *Nature*, 409, 573–575.
- Laskey, R.A. and Madine, M.A. (2003) A rotary pumping model for helicase function of MCM proteins at a distance from replication forks. *EMBO Rep.*, 4, 26–30.

- Manzan,A., Pfeiffer,G., Hefferin,M.L., Lang,C.E., Carney,J.P. and Hopfner,K.P. (2004) MlaA, a hexameric ATPase linked to the Mre11 complex in archaeal genomes. *EMBO Rep.*, 5, 54–59.
- Constantinesco, F., Forterre, P. and Elie, C. (2002) NurA, a novel 5'-3' nuclease gene linked to rad50 and mre11 homologs of thermophilic *Archaea. EMBO Rep.*, 3, 537–542.
- Aravind,L. (2000) Guilt by association: contextual information in genome analysis. *Genome Res.*, 10, 1074–1077.
- Galperin, M.Y. and Koonin, E.V. (2000) Who's your neighbor? New computational approaches for functional genomics. *Nat. Biotechnol.*, 18, 609–613.
- Huynen, M., Snel, B., Lathe, W. and Bork, P. (2000) Exploitation of gene context. *Curr. Opin. Struct. Biol.*, 10, 366–370.
- Huynen, M.J. and Snel, B. (2000) Gene and context: integrative approaches to genome analysis. Adv. Prot. Chem., 54, 345–379.
- Margolin, W. (2000) Themes and variations in prokaryotic cell division. FEMS Microbiol. Rev., 24, 531–548.
- Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25, 3389–3402.
- Notredame, C., Higgins, D.G. and Heringa, J. (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol., 302, 205–217.
- Pei, J., Sadreyev, R. and Grishin, N.V. (2003) PCMA: fast and accurate multiple sequence alignment based on profile consistency. *Bioinformatics*, 19, 427–428.
- Walker, D.R. and Koonin, E.V. (1997) SEALS: a system for easy analysis of lots of sequences. *Proceedings of the Fifth International Conference* on Intelligent Systems for Molecular Biology (ISMB-97), Halkidiki, Greece, pp. 333–339.
- Ikeda, M., Arai, M., Lao, D.M. and Shimizu, T. (2002) Transmembrane topology prediction methods: a re-assessment and improvement by a consensus method using a dataset of experimentally-characterized transmembrane topologies. *In Silico Biol.*, 2, 19–33.
- Peitsch, M.C. (1996) ProMod and Swiss-model: internet-based tools for automated comparative protein modelling. *Biochem. Soc. Trans.*, 24, 274–279.
- Kraulis, P. (1991) A program to produce both detailed and schematic plots of proteins. J. Appl. Crystallogr., 24, 946–950.
- Rost,B., Sander,C. and Schneider,R. (1994) PHD—an automatic mail server for protein secondary structure prediction. *Comput. Appl. Biosci.*, 10, 53–60.
- Felsenstein, J. (1996) Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol.*, 266, 418–427.
- Hasegawa, M., Kishino, H. and Saitou, N. (1991) On the maximum likelihood method in molecular phylogenetics. J. Mol. Evol., 32, 443–445.
- Wolf,Y.I., Rogozin,I.B., Grishin,N.V., Tatusov,R.L. and Koonin,E.V. (2001) Genome trees constructed using five different approaches suggest new major bacterial clades. *BMC Evol. Biol.*, 1, 8.
- Neuwald,A.F., Liu,J.S., Lipman,D.J. and Lawrence,C.E. (1997) Extracting protein alignment models from the sequence database. *Nucleic Acids Res.*, 25, 1665–1677.
- Koonin,E.V. (1992) The second cholera toxin, Zot, and its plasmidencoded and phage-encoded homologues constitute a group of putative ATPases with an altered purine NTP-binding motif. *FEBS Lett.*, **312**, 3–6.
- Neuwald,A.F., Aravind,L., Spouge,J.L. and Koonin,E.V. (1999) AAA⁺: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.*, 9, 27–43.
- Iyer,L.M., Leipe,D.D., Koonin,E.V. and Aravind,L. (2004) Evolutionary history and higher order classification of AAA⁺ ATPases. *J. Struct. Biol.*, 146, 11–31.
- Gorbalenya,A.E. and Koonin,E.V. (1993) Helicases: amino acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Biol.*, 3, 419–429.
- Putnam,C.D., Clancy,S.B., Tsuruta,H., Gonzalez,S., Wetmur,J.G. and Tainer,J.A. (2001) Structure and mechanism of the RuvB Holliday junction branch migration motor. *J. Mol. Biol.*, **311**, 297–310.

- Sawaya,M.R., Guo,S., Tabor,S., Richardson,C.C. and Ellenberger,T. (1999) Crystal structure of the helicase domain from the replicative helicase-primase of bacteriophage T7. *Cell*, **99**, 167–177.
- Singleton, M.R., Sawaya, M.R., Ellenberger, T. and Wigley, D.B. (2000) Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell*, **101**, 589–600.
- Nadanaciva,S., Weber,J., Wilke-Mounts,S. and Senior,A.E. (1999) Importance of F1-ATPase residue alpha-Arg-376 for catalytic transition state stabilization. *Biochemistry*, 38, 15493–15499.
- 55. Leipe,D.D., Aravind,L., Grishin,N.V. and Koonin,E.V. (2000) The bacterial replicative helicase DnaB evolved from a RecA duplication. *Genome Res.*, **10**, 5–16.
- Hillig,R.C., Renault,L., Vetter,I.R., Drell,T.,4h, Wittinghofer,A. and Becker,J. (1999) The crystal structure of rna1p: a new fold for a GTPase-activating protein. *Mol. Cell*, 3, 781–791.
- Bourne,H.R. (1997) G proteins. The arginine finger strikes again. *Nature*, 389, 673–674.
- Ahmadian, M.R., Stege, P., Scheffzek, K. and Wittinghofer, A. (1997) Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP-hydrolysis reaction of Ras. *Nature Struct. Biol.*, 4, 686–689.
- Leipe, D.D., Wolf, Y.I., Koonin, E.V. and Aravind, L. (2002) Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol.*, 317, 41–72.
- Leipe,D.D., Koonin,E.V. and Aravind,L. (2003) Evolution and classification of P-loop kinases and related proteins. *J. Mol. Biol.*, 333, 781–815.
- Lupas, A.N. and Martin, J. (2002) AAA proteins. Curr. Opin. Struct. Biol., 12, 746–753.
- Holm,L. and Sander,C. (1995) Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.*, 20, 478–480.
- 63. Mitchell,M.S., Matsuzaki,S., Imai,S. and Rao,V.B. (2002) Sequence analysis of bacteriophage T4 DNA packaging/terminase genes 16 and 17 reveals a common ATPase center in the large subunit of viral terminases. *Nucleic Acids Res.*, **30**, 4009–4021.
- Bakhtiari, N., Lai-Zhang, J., Yao, B. and Mueller, D.M. (1999) Structure/ function of the beta-barrel domain of F1-ATPase in the yeast Saccharomyces cerevisiae. J. Biol. Chem., 274, 16363–16369.
- Llosa,M., Gomis-Ruth,F.X., Coll,M. and de la Cruz Fd,F. (2002) Bacterial conjugation: a two-step mechanism for DNA transport. *Mol. Microbiol.*, 45, 1–8.
- Ding,Z., Atmakuri,K. and Christie,P.J. (2003) The outs and ins of bacterial type IV secretion substrates. *Trends Microbiol.*, 11, 527–535.
- 67. Burrus, V., Pavlovic, G., Decaris, B. and Guedon, G. (2002) Conjugative transposons: the tip of the iceberg. *Mol. Microbiol.*, **46**, 601–610.
- Fasano, A., Fiorentini, C., Donelli, G., Uzzau, S., Kaper, J.B., Margaretten, K., Ding, X., Guandalini, S., Comstock, L. and Goldblum, S.E. (1995) *Zonula occludens* toxin modulates tight junctions through protein kinase C-dependent actin reorganization, *in vitro*. *J. Clin. Invest.*, 96, 710–720.
- Di Pierro, M., Lu, R., Uzzau, S., Wang, W., Margaretten, K., Pazzani, C., Maimone, F. and Fasano, A. (2001) *Zonula occludens* toxin structurefunction analysis. Identification of the fragment biologically active on tight junctions and of the zonulin receptor binding domain. *J. Biol. Chem.*, 276, 19160–19165.
- Koonin,E.V., Senkevich,T.G. and Chernos,V.I. (1993) Gene A32 product of vaccinia virus may be an ATPase involved in viral DNA packaging as indicated by sequence comparisons with other putative viral ATPases. *Virus Genes*, 7, 89–94.
- Iyer,L.M., Aravind,L. and Koonin,E.V. (2001) Common origin of four diverse families of large eukaryotic DNA viruses. J. Virol., 75, 11720–11734.
- La Scola,B., Audic,S., Robert,C., Jungang,L., de Lamballerie,X., Drancourt,M., Birtles,R., Claverie,J.M. and Raoult,D. (2003) A giant virus in amoebae. *Science*, **299**, 2033.
- Wuitschick, J.D., Gershan, J.A., Lochowicz, A.J., Li, S. and Karrer, K.M. (2002) A novel family of mobile genetic elements is limited to the germline genome in *Tetrahymena thermophila*. *Nucleic Acids Res.*, 30, 2524–2537.
- 74. Rice,G., Tang,L., Stedman,K., Roberto,F., Spuhler,J., Gillitzer,E., Johnson,J.E., Douglas,T. and Young,M. (2004) The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. *Proc. Natl Acad. Sci. USA*, **3**, 3.

- 75. Edgell,D.R. and Doolittle,W.F. (1997) Archaea and the origin(s) of DNA replication proteins. *Cell*, **89**, 995–998.
- Leipe, D.D., Aravind, L. and Koonin, E.V. (1999) Did DNA replication evolve twice independently? *Nucleic Acids Res.*, 27, 3389–3401.
- Andrews, P.D., Knatko, E., Moore, W.J. and Swedlow, J.R. (2003) Mitotic mechanics: the auroras come into view. *Curr. Opin. Cell Biol.*, 15, 672–683.
- Heald,R.W. (2004) Cell division: burning the spindle at both ends. *Nature*, 427, 300–301.
- Roberts, R.J., Vincze, T., Posfai, J. and Macelis, D. (2003) REBASE: restriction enzymes and methyltransferases. *Nucleic Acids Res.*, 31, 418–420.
- Aravind, L., Galperin, M.Y. and Koonin, E.V. (1998) The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold. *Trends Biochem. Sci.*, 23, 127–129.
- Karumbati,A.S., Deshpande,R.A., Jilani,A., Vance,J.R., Ramotar,D. and Wilson,T.E. (2003) The role of yeast DNA 3'-phosphatase Tpp1 and rad1/Rad10 endonuclease in processing spontaneous and induced base lesions. J. Biol. Chem., 278, 31434–31443.
- Strunnikov,A.V. and Jessberger,R. (1999) Structural maintenance of chromosomes (SMC) proteins: conserved molecular properties for multiple biological functions. *Eur. J. Biochem.*, 263, 6–13.
- Jessberger, R. (2002) The many functions of SMC proteins in chromosome dynamics. *Nature Rev. Mol. Cell Biol.*, 3, 767–778.
- van den Bosch, M., Bree, R.T. and Lowndes, N.F. (2003) The MRN complex: coordinating and mediating the response to broken chromosomes. *EMBO Rep.*, 4, 844–849.
- Borde, V., Lin, W., Novikov, E., Petrini, J.H., Lichten, M. and Nicolas, A. (2004) Association of Mre11p with double-strand break sites during yeast meiosis. *Mol. Cell*, 13, 389–401.
- Schleiffer, A., Kaitna, S., Maurer-Stroh, S., Glotzer, M., Nasmyth, K. and Eisenhaber, F. (2003) Kleisins: a superfamily of bacterial and eukaryotic SMC protein partners. *Mol. Cell*, 11, 571–575.
- Christie, P.J. (2001) Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines. *Mol. Microbiol.*, 40, 294–305.
- Lanka, E. and Wilkins, B.M. (1995) DNA processing reactions in bacterial conjugation. *Annu. Rev. Biochem.*, 64, 141–169.
- Ilyina, T.V. and Koonin, E.V. (1992) Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic Acids Res.*, 20, 3279–3285.
- Datta,S., Larkin,C. and Schildbach,J.F. (2003) Structural insights into single-stranded DNA binding and cleavage by F factor Tral. *Structure* (*Camb.*), **11**, 1369–1379.
- Guasch, A., Lucas, M., Moncalian, G., Cabezas, M., Perez-Luque, R., Gomis-Ruth, F.X., de la Cruz, F. and Coll, M. (2003) Recognition and processing of the origin of transfer DNA by conjugative relaxase TrwC. *Nature Struct. Biol.*, **10**, 1002–1010.

- Anantharaman, V. and Aravind, L. (2003) Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. *Genome Biol.*, 4, R11.
- Moazed, D. (2001) Enzymatic activities of Sir2 and chromatin silencing. *Curr. Opin. Cell Biol.*, 13, 232–238.
- Min, J., Landry, J., Sternglanz, R. and Xu, R.M. (2001) Crystal structure of a SIR2 homolog-NAD complex. *Cell*, **105**, 269–279.
- Zhao, K., Chai, X. and Marmorstein, R. (2004) Structure and substrate binding properties of cobB, a Sir2 homolog protein deacetylase from *Escherichia coli. J. Mol. Biol.*, 337, 731–741.
- Zhao, K., Chai, X. and Marmorstein, R. (2003) Structure of the yeast Hst2 protein deacetylase in ternary complex with 2'-O-acetyl ADP ribose and histone peptide. *Structure (Camb.)*, **11**, 1403–1411.
- Finnin, M.S., Donigian, J.R. and Pavletich, N.P. (2001) Structure of the histone deacetylase SIRT2. *Nature Struct. Biol.*, 8, 621–625.
- Deshpande, R.A. and Shankar, V. (2002) Ribonucleases from T2 family. Crit. Rev. Microbiol., 28, 79–122.
- Mazumder, R., Iyer, L.M., Vasudevan, S. and Aravind, L. (2002) Detection of novel members, structure-function analysis and evolutionary classification of the 2H phosphoesterase superfamily. *Nucleic Acids Res.*, 30, 5229–5243.
- Koonin, E.V. (1996) A duplicated catalytic motif in a new superfamily of phosphohydrolases and phospholipid synthases that includes poxvirus envelope proteins. *Trends Biochem. Sci.*, 21, 242–243.
- 101. Zaremba, M., Urbanke, C., Halford, S.E. and Siksnys, V. (2004) Generation of the BfiI restriction endonuclease from the fusion of a DNA recognition domain to a non-specific nuclease from the phospholipase D superfamily. J. Mol. Biol., 336, 81–92.
- Pouliot, J.J., Yao, K.C., Robertson, C.A. and Nash, H.A. (1999) Yeast gene for a Tyr-DNA phosphodiesterase that repairs topoisomerase I complexes. *Science*, 286, 552–555.
- Yanai, I., Wolf, Y.I. and Koonin, E.V. (2002) Evolution of gene fusions: horizontal transfer versus independent events. *Genome Biol.*, 3, research0024.
- Leach, D.R., Okely, E.A. and Pinder, D.J. (1997) Repair by recombination of DNA containing a palindromic sequence. *Mol. Microbiol.*, 26, 597–606.
- Pan,X. and Leach,D.R. (2000) The roles of mutS, sbcCD and recA in the propagation of TGG repeats in *Escherichia coli*. *Nucleic Acids Res.*, 28, 3178–3184.
- van den Ent,F., Amos,L. and Lowe,J. (2001) Bacterial ancestry of actin and tubulin. *Curr. Opin. Microbiol.*, 4, 634–638.
- 107. Amos,L.A., van den Ent,F. and Lowe,J. (2004) Structural/functional homology between the bacterial and eukaryotic cytoskeletons. *Curr. Opin. Cell Biol.*, **16**, 24–31.
- Anantharaman, V. and Aravind, L. (2004) The SHS2 module is a common structural theme in functionally diverse protein groups, like Rpb7p, FtsA, GyrI, and MTH1598/TM1083 superfamilies. *Proteins*, in press.