Open Access

Research article

Natural history of S-adenosylmethionine-binding proteins Piotr Z Kozbial^{*1} and Arcady R Mushegian^{1,2}

Address: ¹Stowers Institute for Medical Research, 1000 E. 50th St., Kansas City, MO 64110, USA and ²Department of Microbiology, Molecular Genetics, and Immunology, University of Kansas Medical Center, Kansas City, Kansas 66160, USA

Email: Piotr Z Kozbial* - pzk@stowers-institute.org; Arcady R Mushegian - arm@stowers-institute.org * Corresponding author

Published: 14 October 2005

BMC Structural Biology 2005, 5:19 doi:10.1186/1472-6807-5-19

This article is available from: http://www.biomedcentral.com/1472-6807/5/19

© 2005 Kozbial and Mushegian; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<u>http://creativecommons.org/licenses/by/2.0</u>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: 21 July 2005 Accepted: 14 October 2005

Abstract

Background: S-adenosylmethionine is a source of diverse chemical groups used in biosynthesis and modification of virtually every class of biomolecules. The most notable reaction requiring S-adenosylmethionine, transfer of methyl group, is performed by a large class of enzymes, S-adenosylmethionine-dependent methyltransferases, which have been the focus of considerable structure-function studies. Evolutionary trajectories of these enzymes, and especially of other classes of S-adenosylmethionine-binding proteins, nevertheless, remain poorly understood. We addressed this issue by computational comparison of sequences and structures of various S-adenosylmethionine-binding proteins.

Results: Two widespread folds, Rossmann fold and TIM barrel, have been repeatedly used in evolution for diverse types of S-adenosylmethionine conversion. There were also cases of recruitment of other relatively common folds for S-adenosylmethionine binding. Several classes of proteins have unique unrelated folds, specialized for just one type of chemistry and unified by the theme of internal domain duplications. In several cases, functional divergence is evident, when evolutionarily related enzymes have changed the mode of binding and the type of chemical transformation of S-adenosylmethionine. There are also instances of functional convergence, when biochemically similar processes are performed by drastically different classes of S-adenosylmethionine-binding proteins.

Comparison of remote sequence similarities and analysis of phyletic patterns suggests that the last universal common ancestor of cellular life had between 10 and 20 S-adenosylmethionine-binding proteins from at least 5 fold classes, providing for S-adenosylmethionine formation, polyamine biosynthesis, and methylation of several substrates, including nucleic acids and peptide chain release factor.

Conclusion: We have observed several novel relationships between families that were not known to be related before, and defined 15 large superfamilies of SAM-binding proteins, at least 5 of which may have been represented in the last common ancestor.

Background

S-adenosylmethionine (SAM or AdoMet) is a conjugate of nucleotide adenosine and amino acid methionine, two

ubiquitous biological compounds that almost certainly were present in the common ancestor of living cells and may have been found in the prebiotic environment on Earth, predating the origin of Life itself [1]. SAM is an essential metabolic intermediate in every studied cellular life form, and each cellular organism has several SAM-utilizing enzymes. One relatively well-understood biological role of SAM is to donate methyl groups for covalent modification of different substrates – from as simple as oxidized arsenic, chloride, bromide, and iodine ions [2-4], to as complex as rRNA, tRNA, and essential proteins, whose methylation status can serve as a regulatory signal for maturation and control interactions with other macromolecules ([5-7] and references therein).

Methyl transfer is but one of many biochemical processes requiring SAM. Enzymatic reactions that involve interaction of proteins with SAM or its structurally similar derivatives include transfer or methylene, aminoalkyl, ribosyl, and 5'deoxyadenosyl groups; formation of 5'deoxyadenosyl radical, which can be used as a redox intermediate in many reactions; SAM decarboxylation; and *de novo* synthesis of SAM from adenosine and methionine. There are also numerous interactions between SAM and non-enzymatic proteins, where SAM serves as a ligand triggering a regulatory change in the effector protein.

Despite the interest in this amazing variety of functions associated with SAM, and the known three-dimensional structures for representatives of almost every class of SAMdependent enzymes, the structural, functional, and evolutionary relationships between the SAM-binding domains remain not well understood. Do all or some of the SAMbinding proteins share common evolutionary ancestry? How many distinct structural modes of interaction between SAM and protein are there? Is there strong or weak correlation between conservation of sequence and structure, the mode of SAM binding, and the chemical reaction facilitated by the enzyme? Finally, what may have been the repertoire of SAM-binding proteins in the ancestral organisms - in particular, in LUCA, the Last Universal Common Ancestor of the three present-day domains of Life - Bacteria, Archaea, and Eukarya?

We sought to address these questions by comparing sequences and structures of various groups of SAM-binding domains recognized in proteins. We describe several previously unsuspected relationships between some of such groups, predict novel members for many of them, and conclude that LUCA may have had more than a dozen of SAM-binding proteins, belonging to several distinct folds.

Results and discussion

We have adopted the iterative comparison strategy, using the known or suspected SAM-binding protein domains as the queries in increasingly sensitive probabilistic methods of sequence modeling and database searching. In many

cases, a SAM-binding part of the protein constitutes only part of the polypeptide chain. For example, methyltransferases typically consist of well-conserved SAM-binding portions and highly variable substrate-binding regions, sometimes further supplemented with portable domains also found in otherwise unrelated proteins, such as chromo domain interacting with methylated histone tails in eukaryotes, or PUA domain that probably interacts with RNA [8]. In this work, we are concerned with the protein moieties that bind SAM, so we neither examine these other domains, not consider methyltransferases that utilize other sources of methyl groups, like folate or methyl-We cobalamin derivatives. did not describe isoprenylcysteine carboxyl methyltransferase (ICMT), an integral endoplasmic reticulum membrane protein with unknown structure [9,10] (reviewed in ref. [11]).

The phylogenetic relationships inside of several recognized groups of SAM-binding proteins, especially within Rossmann-fold SAM-dependent methyltransferases, have been reviewed recently [12,13]. Although we summarize and extend their observations, our main focus is on the analysis of more distant, previously unexamined, relationships.

Versatile lphaetaeta architectures adapted for SAM binding

Rossmanoids: ancient and ubiquitous SAM-dependent transferases The majority of SAM-dependent methyltransferases belong to a large class of enzymes with the Rossmann-like fold, one of the more common arrangements of protein spatial structure, observed in dozens of diverse families of enzymes [14]. SAM-dependent methyltransferases are a large group of enzymes within the Rossmanoid class, and they account for a substantial fraction of all proteins in completely sequenced genomes; for example, with 1.7% of genes in Helicobacter pylori J99 coding for known or predicted SAM-dependent methyltransferases, this group makes the list of 10 most commonly used sequence and structure families in that species [15].

In the most basic arrangement, the Rossmann-like fold consists of alternating β -stranded and α -helical regions, with all strands forming a central relatively planar β -sheet, and helices filling two layers, one on each side of the plane. As with many other Rossmann-like folds, the N-terminal β -strand of methyltransferases is located in the middle of the sheet, and the strand topology is 3214576, with the 7th strand antiparallel to all other strands (Figure 1a). Yet another typical feature of Rossmanoid enzymes is that the functionally important, conserved residues are often located in the C-termini of the β -strands or in the adjoining loops [14]. Some methyltransferases conform to this plan quite well, with an occasional addition of an extra helix or a β -hairpin [16], or, rarely, deletion of one or both



Figure I

Fold and topology of SAM binding proteins. Corresponding fragments of cartoon and topology representations of selected structures were rainbow colored from N-terminal (blue) to C-terminal (red) end. Less significant fragments of secondary structure were left white in topology diagrams. Reference to other representative structures are provided (in parentheses) as SCOP sunid numbers (i.e. ref. [139]). (a) lej0A, Rossmann-fold methyltransferase (53335 excluding: 102555, 69556, 69557, and 69560); (b) IIIeA, cyclopropane fatty acid synthase (69560); (c) ImjfA, spermidine synthase (69557); (d) IvhvA, porphyrin C-methyltransferase (53789); (e) Ij6rB, Met synthase reactivation domain (methyltransferase; 56506); (f) ImxiA, SPOUT methyltransferase (89629, 75218); (g) Ir30A, SAM dependent radical enzyme (102114); (h) ImvhA, SET domain methyltransferase (82199); (i) IrqpB, 5'-fluoro-5'-deoxyadenosine synthase (102521, 101851); (j) Im7y, I-aminocyclopropane-I-carboxylate (ACC) synthase (53441, 64130, and similar but different: 53439); (k) IvkyA, tRNA-ribosyl transferase-isomerase (111338); (l) Io9tB, methionine adenosyltransferase (55972); (m) Icmc, MetJ – methionine repressor (dimer; 100972); (n) ImsvA, SAM decarboxylase (56275); (o) IpbjA, CBS domain (dimer; 54630 – not all CBS domains bind SAM); (p) IkzfA, acylhomoserine lactone synthase (75508).

of strands 6 and 7 [17]. Most methyltransferases, however, contain additional domains appended or inserted into the basic Rossmann fold [16].

Notwithstanding the insertions of additional domains and structural elaborations, comparative sequence analysis of the Rossmann-fold methyltransferases identifies the set of five highly conserved regions of the SAM-binding region, each centered on one or more nearly-invariant residues (Figure 2). They correspond to motifs I-V from motifs initially proposed for DNA:m5C MTases by Posfai et al. [18] (reviewed in refs. [12,19]), but some of the conserved residues highlighted in this work have not been pointed out before (see below). Each motif has a clear counterpart at the structural level. Five motifs are arranged in the same linear order in almost all known methyltransferases, with a notable exception of several groups of DNA- and RNA-methyltransferases, where circular permutation of the sequence results in a main chain fission after motif II, while the spatial structure of the domain and mode of SAM binding remain virtually unperturbed (discussed in more detail by Bujnicki [20]).

The first conserved sequence block (Motif I) includes in its C-terminal part the consensus GxGxG, considered the hallmark SAM-binding site of the Rossmann-fold SAM-dependent methyltransferases. None of three glycine residues is universally conserved, but the replacements are typically by the residues with small side chains, or with propensity of bending the main chain. This agrees with the structure data, indicating that the consensus is located in a loop connecting the first β -strand and the α -helix in the Rossmann fold core. The complete β -strand and part of the preceding loop are also part of Motif I. In the middle of β -strand 1, there is an exceptionally well conserved acidic residue (D or E); one or more conserved positively charged residues are found close to the N-terminus of this strand (Figure 2).

Motif II encompasses β -strand 2 and adjoining turn. A partially conserved acidic residue is common at the C terminus of this strand. Motif III corresponds to β -strand 3, located at the edge of the β -sheet in the Rossmann fold. An acidic residue is partially conserved close to the C-terminus of this strand, too. Whenever the substrate (SAM), its analogs, or reaction product (SAH) are co-crystallized, they are found close to the invariant residues in Motifs I-III (Figure 2 and see below).

Motif IV consists of β -strand 4 and the flanking loops. In this motif again, there is a well-preserved D/E/N residue, located at the extreme N-terminus of the strand, i.e. at the side of the fold that is not involved in substrate binding. Motif V corresponds to the helix following the strand with motif IV. In some Rossmann-fold methyltransferases, it

serves as a scaffold for large hydrophobic or aromatic side chains that stabilize the adenine moiety of AdoMet, but it many cases it has been shows these residues are not essential for the MTase activity [21]. Finally, Motif VI corresponds to Strand 5 of the β -sheet, and the preceding tight turn with a nearly-invariant glycine residue.

Several residues from Motifs I-V are known to make direct contact with SAM. In particular, one or more residues in "GxGxG" loop are in contact with the carboxypropyl moiety of SAM, while conserved acidic residue in Motif II forms hydrogen bonds with the ribose hydroxyls (Figure 2; ref. [22]). Variable residues at the C-termini of strand 2 (Motif II) and conserved acidic residue in Strand 3 (Motif III) interact with the nitrous base, while variable residues C-terminal to Strand 4 (Motif IV) appear to contact the amino and sulfonium groups of the methionine moiety of SAM [22]. Residues from motif IV, VI, VIII, and sometimes X are associated with the catalytic pocket, where residues from motif V and VII are important mostly for the structural stability [19,23].

The roles of other conserved residues in SAM-dependent methyltransferases are less well understood. Near-omnipresence of the D/E residue in Motif I suggests that it has an important role. It has been noted [24], that in FtsJ RNA methyltransferase this residue coordinates SAM through a water molecule. In fact, in all 3-D structures of methyltransferases where solvent molecules are present (i.e. PDB structure 1EI0, 1KYW, and 2ADM), the oxygen atoms in the carboxyl group of this D/E residue make direct contacts with two water molecules, one of which is capable of forming a hydrogen bond with the side chain of methionine moiety of SAM. In some ribose 2'-O-MTases, D/E amino acid conserved in motif I is substituted by tyrosine (Figure 2 and ref. [25]), and it has been proposed that this residue could be used to directly (not via the water molecule) coordinate the amino-carboxyl end of SAM (J.M. Bujnicki, personal communication).

These observations are of interest for understanding the mechanism of methyl transfer by Rossmann-fold methyltransferases. Two best-studied groups of transferases that have Rossmann-like fold and use a nucleotide derivative as a cofactor, namely ATPase-like kinases and nucleoside diphospho-sugar transferases, appear to require a divalent metal cation for polarization of water molecule that can then attack a scissile phosphoester bond [26-29]. Methyl-transferases, on the other hand, need to work on a C-S⁺ bond in SAM, but do not seem to have any metal ion bound in the appropriate position (even though divalent cations have been included in some crystallization media). The proposals for reaction mechanisms of different classes of SAM-dependent methyltransferases include nucleophilic catalysis, with the identity of nucleophile

				4	motii	1	<u> </u>		motif II	&	mot	.11 111	<u> </u>	OTIT IV		tir v mo	tir VI
gi numbe	r,name,species		55	sss	hl	hh-hhł	h-hh ss	s-sss	hhhhhhhhhh	hh-	-hh ss-	-ss	sss	SSSS	hhhhhh	hhh ss	ssss h sssss
17374834	DNMT2_HUMAN	1	MEPLR	LE <mark>LY</mark> SGV	-G	MH-HA <mark>I</mark>	R-ESCIPAQV	A-AID	VNTVANEV <mark>Y</mark> KYN <mark>I</mark>	PH <mark>T</mark> Q-	-LAKT-	- <mark>I</mark> EGI-TL	-EE(5)FD	MILMSPPc(17)nsFLH <mark>IL</mark> D	01 <mark>L</mark> PRLQKLPKY <mark>1</mark>	LLENvkgfeVSSTRD
7299276	ART4 DROME	176	VDFQDKI	LDVGAGS	-I	LS-FF <mark>7</mark>	V-QAGAAK <mark>V</mark>	Y-AIE	ASN-MAQYAQQL <mark>V</mark>	(4) VQHK <mark>I</mark> S-	-VIPGK-	- <mark>I</mark> EEI-EL	-PEKVD	VIISEPMg(5)ermLET <mark>YL</mark> H	IARKW-LKPQGK	YPTH(14) SEQYNKA
7299364	ART1 DROME	88	HLFQGKT	LDVGCGT	-1	LS-MF <mark>Z</mark>	A-KAGAAOV	I-AVD	CSN-IIEFARQV	(4)LQDVIT-	- <mark>v</mark> vkgk-	IEEI-EL	-PN(4)VD	IIISEWMg(5)esmLDT <mark>VL</mark> Y	ARDKWLKKDGM	FPDRgtlYITAIE
20093722	PCM METKA	79	DPRPGHK	LEVGAGS	-Y	HA-AV	A-ELVKPDGRV	I-TVE	RIPELADFARNNI	(4) YDRFVK-	-VLVGD-	GTKG-YP	-PEAPY-DI	RILVTAG	APDVPE	SLLEOLKPGGK	VIPV(17)eIRRRRH
21465733	GidB ECOLI	62	PYLOGER	TDVGTGP	-L	PG-IPI	S-IVRPE-AH	T-LLD	SLGKRVRFLROV	HELKLENIE-	-PVOSR-	VEEF-PS	-EPPFD	JVISRAFa-	slndmvs	WCHHLPGEOGR	YALK(17)eSVVKLO
17552558	KOG3420 CAEEL	44	IGFEGKK	TDTGCGC	-M	M-TTZ	A-TMYEL-ETV		TDDEALKICSEN	(4) VODRCE-	-T.L.OAD-	TLDP-ES	-DL(4)FD	VAVINPPf(5)agiDMOFVC	TGLOMVRPGGS	FSLH(4) rdYILKNA
15643511	Gcd14 THEMA	97	DVKEGDR	TDTGVGS	-A	MC-AVI	A-RAVGSSGR	F-AYE	KREEFAKLAESNI	(4)LIERVT-		TSEG-FD	-EK(2)-D	ALFLDVP	dPWNYTE	KCWEALKGGGR	ATVC(5) VOETLEKE
14521938	COG2263 PYRAB	44	GDIEGKV	ADLCAGT	-v	LS-YGZ	L-SLGAKKV	Y-AVE	VDEDAVEILKTNI	KK-FEGKFE-	-VFIGD-	VSSF-KV	-RVD	TVTMNPPf(4) khaDRPFLI	KAFEISDVV	YSTH(19) ITTHRLT
17865746	Homk SALTY	107	PVKT-CR	T.D.C.T.CT		T	A-CERPD-CEN	T-AVD	RMPDAVALATRNZ	EHLATENVE-	-TLOSC-	WFSA-T.S	GOOFD	TUSNPPU (28) adt.THTT	N	LLEH(7) WRAVERRS
24212091	DEAM1 ADATH	270	NIKROW	UDVCCCT		DEV	A FRED VUL	V CTD	T CUMMTCENTED	TO TECEVE	PEVA D		DDMC FD	TVCPDTil	h-igdypar	T PEVM VDCCVI	TTCD(49) FMOUT KP
23200202	CONT MEDGA	107	OBBOL VO				W CKNDT TYC	T NIT	DUITED DUITED	DOVE	-FEVAD-		VN D		A) Jahor War	W OVEN DDWOR	DISD(49) INQUER
23200293	COMI_MEDSA	170	GFEGERSI	DVGGG1	-A	C TI	V-SKIPI-ING	I-NPL	LPRVIEDAPSI	PGVE-	-nvGGD-		KAD	AVPMKWIC(4) dencikrii	NCIERTPDNGK	IVAL(30) KERIQKE
14520947	Met-IO+_PIRAB	1/9	MAQEGEV	FDMFAGV	-P		A-KKA	F-ACD	1NPWAIKIBEEN	KLNKVNNVV-	-PILGD-	-SREI-EV	-KADI	RIIPINLPK-	yAndrur	AISCINDGOVI	HIIG(20) EVLGKRV
21431825	PIMT_ARATH		HLKPGMR	LDVGSGT	- ¥	LT-ACE	A-VMVGTEGRA	I-GVE	HIPELVASSVKNJ	(11) RSLAV-	-HV-GD-	GRQG-WA	-EFAPY-D	AIHVGAA	APEIPE	ALIDQUKPGGRI	VIPV(5) digvvDKN
15669418	COG2521_METJA	131	RVKKGEK	LDTCMGL	-Y	TA-167	Y-RRGAB	I-TIE	KNPNVLELAKINE	(5)FKGGIK-	-IILGD-	- <mark>A</mark> YDV-1K	-RF(4)FD	VVIHDPPr(2) UTASEELI	EIFRV KPGGR	FHIV(16)MERLRKV
13473166	COG0500_RHILO	41	LSLEGKT	LDIGCGS	-G	IT-LHI	V-ERHGA-ARA	T-GFD	VEQPVIEAARRR	(4)LSDRAS-	-FVQA	P-PG	-AL(6)FD	AVFSKDA	LLHVPD	OK(6)IFRVLKPGGVE	AASN(15) YIAAEGL
15613250	TrmA_BACHD	307	NLTGSET	/ID <mark>AY</mark> CGI	-T	IS-LF <mark>I</mark>	A-QQAKHV	Y-GVE	IVPEAISDAKRNA	RLNGFANVQ-	- <mark>F</mark> AVGD-	- <mark>A</mark> EKV-MP	-WW(7)-D <mark>1</mark>	<mark>VIV</mark> VDPPrk	gcDEALL	T	YVSCnpATLARD
731019	TEMT_MOUSE	52	GGVGGDVI	JIDIGSGP1	r-1	YQ-LLS	SA-CEVFREI	I-VTD	YTPQN <mark>L</mark> QE <mark>L</mark> QKW <mark>I</mark>	(34)RTV <mark>T</mark> R-	- <mark>V</mark> LRCD-	- <mark>V</mark> TKT-PP	-LG(5)LAI	D <mark>CV</mark> LTFLa (7)idtYRA <mark>AL</mark> F	RR <mark>L</mark> AGL <mark>L</mark> KPGGH <mark>I</mark>	VTLV(20) KEVVEKA
22001590	FlpA_METKA	73	PFEEGSK	/LY <mark>LG</mark> AAA	-T	TP-SH <mark>V</mark>	S-DIIKESGVE	Y-CVE	FASRMMQELIPV(-E-KRPN <mark>M</mark> I-	-PILGD-	- <mark>A</mark> TKP-HG	-YA(5)VD <mark>V</mark>	VIY QDIAqp	kqAEV <mark>VA</mark> E	DN <mark>A</mark> EAF <mark>L</mark> RPGGY <mark>V</mark>	<mark>IV</mark> AIka-rsIDVTKE
16079790	YrrM_BACSU	51	SVKQPKK	ILE <mark>IG</mark> TAI	-Y	SA-IR <mark>N</mark>	A-LELPS-AB	Y-TIE	RNEKR <mark>H</mark> EE <mark>A</mark> VNN <mark>J</mark>	(4)LDDR <mark>I</mark> H-	- <mark>V</mark> FYGD-	- <mark>A</mark> LEL-AD	-AV(5)YD <mark>V</mark>	VIFIDAAk-	gq¥QN <mark>FF</mark> H	IL <mark>Y</mark> EPM <mark>L</mark> SPDGV <mark>I</mark>	ITDN(20)rLVAKID
15237618	KOG2904 ARATH	197	EWFKKDI	ADLGTGS	-A	IA-IG	A-KVLGSRGR <mark>V</mark>	I-ATD	LSPVA <mark>I</mark> AV <mark>A</mark> GHN <mark>V</mark>	(4)LEGM <mark>I</mark> E-	-VREGS-	-WFEP-LK	-DL(4)V- <mark>(</mark>	GLVSNPPy (29)dsLFH <mark>LC</mark> H	IG <mark>A</mark> SRM <mark>L</mark> QPGGF <mark>F</mark>	VFET(5)skMIVDYM
732118	RsmC ECOLI	193	TPHTKGK	LDVGCGA	-v	LS-VA	A-RHSPK-IR	T-LCD	VSAPA <mark>VEAS</mark> RAT <mark>I</mark>	AA-NGVEGE-	-VFASN-	VFSE-VK	-GRFD	MIISNPPf(7)ldaAQT <mark>LI</mark> F	RG <mark>A</mark> VRH <mark>L</mark> NSGGE <mark>I</mark>	RIVA(15) fHEVIAQ
15899066	ChiT SULSO	37	RIKKGDK	LDIGCGT	-s	IT-VEZ	S-LLVGNSGRV	Y-GID	KEEKAINLTRRN	(4)VLNNIV-	-LIKGE-	APAI-LS	-TINEKFD	RIFIGGGs-	ekIKE <mark>II</mark> S	ASWEINKGGR	VIDAil-leTVNNAL
16764439	YCCW SALTY	224	RYVENOR	UNCEST	-G	FA-VSZ	L-M-GGC-ROV	v-svp	TSODALDIARON	(5) DLSKAE-	-FVRDD-	VFKL-LR	AY(6)FD	TTMDPPk (12) rgYKDTNM	LATOL NPGGI	LTFS(11)kITADAA
455655	PrmA ECOLI	155	DLTG-KT	TDFGCGS	-1	A- TAZ	L-KLGAAKA	T-GTD	T DPOATOASRDNA	(4) VSDRLE-	-T.YLPK-	DOPE-EM	-KAD	VVVANTLa-	GDLRELAP	PLTSVLPVSGGL	GLSG(29) TITGRKN
1/5/0150	FDDT WIMAN	156	NTERCAR	U VI CAAC	-	T V CUT	E DIVORDOL	V AUP	P	AV VODILL	DUTE	N DUD UV	VP (E) VD	TRADUAG	DOTRIN	I	UTET(10) SEAVENE
12220065	PERT DEFAR	150	TIPRCMT	UDI CA AD	<u>c</u>			T-ACD	T T T T T T		-FVIED-	P TED AV	-FA(11)D	UTEDMan(14) wicet AL		TTVT(12) *FMEDVU
16700305	COGA122 LIGTN	45	DIDK ONI	DIGAAF		3-00		T-ABD			TTEN D				14) yicebab		DIRI(15)IEFFDRV
16/99285	COG4123_LISIN	45	PIRK-GK	LDLCSGN	-1	1P-LLI	S-TRTKAQI	V-GVE	IQERLADMAKRS	(4)LEDQIE-	-11EXD-	NHI-TD	-LI(5)-D	LVTCNPPY(23) CELEDTIE	(VAASLUKQGGKA	NFVH(36)IVEGIKD
3023819	ID/6_METJA	43	VIKPGKI	LDLGCAP	-G	wM-Qv <mark>4</mark>	AR-EIVGDKGFV	T-GTD	LQPVKPFEI	DNVV-	-AIKGD-	- <mark>F</mark> TLE-EN	-LN(13)D	VISDASP(14) divitanç	21 <mark>ATEML</mark> KERGNE	VAKV(34)VIVIAKK
15669643	COG4076_METJA	31	VVDEDDV	FDLGTGS	-I	LA-MI <mark>Z</mark>	A-KKAKK <mark>V</mark>	Y-AIE	LDPFT <mark>Y</mark> DYAKEN <mark>I</mark>	KVNGFNNIE-	- <mark>IIEGD</mark> -	- <mark>A</mark> STY-NF	-KEKAD <mark>I</mark>	VVIAELLd(8)vkvMNS <mark>II</mark> E	R <mark>G</mark> FLKEDVK	IPAKAISTIQ
15615070	SUN_BACHD	249	APRPGER	LD <mark>AC</mark> AAP	-G	KS-TH <mark>N</mark>	A-ERMDGEGTI	V-SLD	MHRHK <mark>V</mark> KLIEEQ <mark>7</mark>	(5) DNIK <mark>A</mark> Q-	- <mark>T</mark> LDARk	(<mark>A</mark> REV-FA	-DEPFDI	R <mark>IL</mark> VDAPc(25) nvQLD <mark>IL</mark> S	SS <mark>V</mark> AKL <mark>V</mark> KPGGL <mark>I</mark>	IYST(6)enEQVIAE
21362981	TAM_RHILO	28	PVDFPRR	/VDIGCGP	-N	ST-EL <mark>I</mark>	V-QRWPD-AQ <mark>V</mark>	S-G <mark>F</mark> D	<mark>T</mark> SPDM <mark>I</mark> EK <mark>A</mark> RVR <mark>I</mark>	PNVS-	- <mark>F</mark> DLAD-	- <mark>A</mark> STW-QP	-AEPVN	VIF ANAVfq	w-lptHPA <mark>VF</mark> Ç)R <mark>L</mark> MGF <mark>L</mark> APGGA <mark>I</mark>	AVQM(43)sPLADRL
21363063	UbiG RHIME	61	QPLEGLR	LDIGCGG	-L	LS-EPN	A-RMGAD <mark>V</mark>	V-G <mark>A</mark> D	<mark>A</mark> SEKN <mark>I</mark> GI <mark>A</mark> RTH <mark></mark>	(4)VSVD <mark>Y</mark> R-	-AVTAE-	- <mark>A</mark> LAE-AG	-ES(4)LN	EVVEHV	aDVE <mark>FF</mark> M	ITtCAHMVRPGGL	FVATin-rtLKAAAL
20139025	MraW LISIN	17	EVKPDGI	VDATLGG	A-G	HS-EYI	L-NKLNEQGH	F-AFD	QDQTA <mark>I</mark> DN <mark>A</mark> KIK <mark>I</mark>	AE-YSDRVT-	-FIKSN-	FRYM-KE	-AL(7)VD	GILYDLGV (115) aVEDSLE	N <mark>ALTL</mark> KPGGR	SVIT(56)rARSAKL
3122431	NodS RHITR	77	AOGPIAN	LEVGCAA	-A	FT-EKI	A-PHCORL	T-VID	VVPEAIDRTRRR	-N-KPAHIS-	-wvvsD-	v00F-SS	-EELFDI	LIVVAEVI	5)iaeMRMAVG	NLLRMLAPGGH	VFGS(16)tVIAILT
15219724	KOG1271 ARATH	166	KYLSSWN	U.D.GTGN	-L	IHOI	A-KEGESDI	T-GTD	YSDGAVELAOHLS	SORDGE PNTR-	-FMVDD-	TLDT-KL	-EO(11)D	ATGLHPDop	vkRVMVWD	SVSKLVAPGGI	VTTS(72) ±VAFLRK
12642920	TRM1 SCHRO	120	DNIACET	TRALCAT		р с тр <mark>у</mark>	A-VET DNUVDT	T-AND	T TENAVETERNI	(A)VEDTUT-	DNKC	NAV MH	MNKEHVD	TDI DRYg	CAAPPER	AVOSUSEDCI T	CTTC(29) bEOAVEN
15222076	VOC2179 APATH	212	CEEDUNIT	UDVCCCN	- -		T CEVER TYC	U-NPD	T AOVI TOA BEN	DOVE	WVCC D	M FUE VD	KC D	VPMPWT1/	A) dodCTWTTW		TTVE
16766959	When Colory	213	DUDGUDAU				CDVN NO			AMT KANNAD			ODCEDUU	AVPPARATI(helt prot		TIVEFVIFRE
10/00050	INNF_SALII	**	PVMVDAH	LDCFAGS	-A	LG-LEA	L-SRIAAQA	1-665	MDRAVSQQLQKNI	ATLKANNAR-	-VVNTN-		-QPGIPHH	VVFVDPPII	KGILEEIL	TI-GIOGMUNDOWL	IVESEVENGL
1/236921	KOG2899_CAEEL	148	DWFEHKQ	LDIGCNA	-F	LT-LS	A-KDFSP-RKI	T-GTD	IDEHLIGVARKNI	(46)KENYV-	- <mark>LESDE</mark> -	-wrdw-16	-PEFD	VILALSIT (8) aagmer <mark>ff</mark> f	(R <mark>AYAQL</mark> HPGGRL	ILEP(14) CLKANYS
17988577	Ubie_BRUME	78	SKRPGWTS	LDVA GGT	-D	IA-FRI	V-EASGROAN	T-ILD	INGSMLGVGRER	(4)LIDNLE-	-FVEAN-	AEEL-PF	-EDNS-FD	AYTIAFGir	n-vphIDKALS	SE <mark>A</mark> YRVLKPGGRE	LCLE(4)elPVLDKV
12721903	p103602_PASMU	54	PYIVDAT	CLDCFAGS	-A	LG-FE <mark>7</mark>	L-SRQAKQV	T-FLE	LDKNVATQLSKNI	(5) NAEQAN-	- <mark>VINQN</mark> -	SLQF-LN	-QP(5)FD	IVFIDPP±h	iglaqQTVD <mark>L</mark> L	ARHNWLKPDAL	YVET(13)1LKEKIT
7296118	KOG2198_DROME	287	QVKPGER	/LDACASP	-Gk	SL-LMI	Q-TLHLDH	V-CND	IQESRLNKLRKVN	(14)R <mark>L</mark> I-	- <mark>L</mark> SQSD-	- <mark>A</mark> RNL-DQ	-YEQFDI	K <mark>IL</mark> VDVPc(27)elQAG <mark>IL</mark> A	AN <mark>C</mark> LRL <mark>L</mark> RPGGSI	VYST(12) hMALQKV
15964087	COG3963_RHIME	51	DPESGLP	LE <mark>LG</mark> PGT	-v	IT-KA <mark>I</mark>	L-ERGIEPEK <mark>I</mark>	V-SIE	YSTDF <mark>Y</mark> KQ <mark>L</mark> KAH <mark>I</mark>	DG <mark>V</mark> H-	- <mark>F</mark> INGD-	AFDL-SR	-TL(7)FDS	S <mark>VI</mark> SAVP11	nfpMHRR <mark>V</mark> E	L <mark>I</mark> EDL <mark>L</mark> SR	PFGRPVVQIS
17232307	COG0270 ANASP	4	REQPI	IDLFAGA	-G	FG-LG	E-MAGFSVF	L-SVE	IDTWACDTLRYNF	RPDST-	-VIQND-	IGNF-ST	-EN(10)D	IIIGGPPc(18) glFIN <mark>FA</mark> C	W <mark>I</mark> KF- <mark>L</mark> -EPKA <mark>F</mark>	VMENvkqLLSRKN
30580330	CaMT4 ARATH	88	KLVNAKN	MEIGVYT	-Y	SL-LA	A-LALPEDGK	L-AMD	VNREN <mark>Y</mark> ELGLPI	(4)VAHKID-	-FREG(4)V-	LDEI-VA	-DE(5)YD	FIFVDADk-	dnYIN <mark>YH</mark> K	RLIDLVKIGGV	GYDN (25) VLELNKA
68127775	DREV LEIMA	111	SDFOFDS	LDIGAGD	-G	VT-AN	A-PLFKKV	Y-VTE	FSVSMRWRLRRR	YE-VLPHDD-	-PFHMN-	TAEO-LL	-DRRY-FD	VIACNNV1d	r-admPET <mark>LL</mark> F	REMRDSLKPNGLL	VLAV(50) fEVVRWT
12620207	KOG1253 HIMAN	272	FERKPLE	TDARGAT	- T	MG_1.0	A-KHLONAVEN		LNENSVTL TOEN	(29) NTKVT_	-KMDAN-	VT.MH_T.R	SFD	THLDPF-	+ SVNVL	S	TVSV(12) GHVARRH
7295917	KOG3191 DROME	39	DRLOPSL	VELOSOS	-V	TT-TAT	A-KKLAG-FSI	CIATO	TNPKACNATPRT	TRNGAR-	-T.DSTP-	CSLA-DA	LEPRSVD	T.T.FNPPy(35) rrVTDTL	K	VI.I.I.(5) poET TKVI.
4506567	DIMT WIMAN	10/	VVVPDTT	UDICCCK	· ·		W WORT NUT	V CTD	T ADVENTOCOOP	(11) TECNE	PTTA D	C CVP II	TOVER DI	POMCEDIC	12) og a DMML	N ACEPTERCOVE	TOTT (22) + WEFORE
4506567	RNMI_HUMAN	194	KKKKDII	LDLGCGK	-6	DL-LK	K-KGRINKL			(11)11546-	-FITAD-	-SSKE-LL	-IDKFK-DI	PONCFDIC(13) eqaDMML	INACER SPGGI	IGIT(25) UVRPORK
63206/3	KMT2_YEAST	238	EATNSAT		-1	ID-TFI	Q-ARKPYR	Y-ICE	AHPDVLAKMKM	(6)DNVV-	-ILEGK-	WQDT-LN	-NL(7)FD	arribtes-	enropitio	DLYAVVVGLIKPEGV	SFFNGLGADKS
25407988	KOG16//_ARATH	200	DLGPDTL	FDVCCGT	-T	IG-LTI	A-HRVGMV	I-GIE	MNASAVADAERNA	(106)DN		-SSSL-EP	-EK(8)V-	VALVDPPrs	ginpavik <mark>al</mark> e	THPRLKRL	ILSC(25) NIRGREE
418431	YE05_YEAST	34	DGER-KL	JVDVGCGP	-T	ATLQN	A-QELKPFEQI	I-GSD	<mark>l</mark> Satm <mark>i</mark> kt <mark>a</mark> e-V <mark>i</mark>	(17)SDDFK-	- <mark>F</mark> LGAD-	-SVDK-QK	-IDMIT <mark>2</mark>	AVECAHWI-	dfek <mark>f</mark> Qf	RS <mark>A</mark> YAN <mark>L</mark> RKDGTI	AIWG(77)pLLIRKQ
12230777	TRM1_PYRAB	48	NVLKPRI	/LD <mark>AL</mark> SAT	-Il	RG-IR <mark>E</mark>	A-LETPA-EE <mark>I</mark>	W-MND	<mark>I</mark> SEDA <mark>Y</mark> NL <mark>M</mark> KKN <mark>V</mark>	(18)SEK <mark>T</mark> L-	- <mark>V</mark> VNHD-	DANR-LM	-AE(4)FH	FIDLDPFg-	sPME <mark>FL</mark> D)Ta <mark>L</mark> RSVKRKGIL <mark>G</mark>	ITAT(36)gVVARYA
18202486	AB140_YEAST	430	KDAEPVT	FE <mark>IG</mark> CGA	-N	TF-FP <mark>]</mark>	L-KDN(4)LR	I-AAD	<mark>F</mark> APRA <mark>V</mark> EL <mark>V</mark> KNSE	E(7)HATV <mark>W</mark> D-	- <mark>L</mark> ANPD-	- <mark>G</mark> NLP-DG	-VEPHSVD	<mark>IAV</mark> MIFVf(4)pnqWDQ <mark>AM</mark> D	DN <mark>L</mark> HKI <mark>L</mark> KPGGK <mark>I</mark>	IFRD(6)tqVRFKKN
14521755	CheR_PYRAB	100	KGSRSLK	WSAACST	-Qepys	IA-MS <mark>I</mark>	Y-EVL(4)-SG	F-RVS	<mark>I</mark> (5)DKEA <mark>L</mark> QI <mark>A</mark> MKGE	C(33)KRL <mark>V</mark> T-	- <mark>F</mark> RWFN-	LLSS-VY	-PKGFD	VIFIRNV1(4)keaQEE <mark>IF</mark> F	RK <mark>L</mark> YDS <mark>L</mark> EDHGY <mark>I</mark>	ILGKtetILGKSA
15240639	PIMT ARATH	119	ERDGGKF	ALMFVGT	-Y	LT-GC	A-LMVGAEGRV	V-G <mark>V</mark> D	HIPELVDMSIKN	(10)KGSLS-	-LHVGD-	GRKGW	-QEFAPYD	AIHVGAA	ASEIPC	P <mark>LLDQL</mark> KPGGR	VIPLgtyFQELKV
20137529	ANMX HUMAN	62	RGOK-AL	LDIGTGT	-L	LS-MMZ	V-TAGADFC	Y-AIE	VFKPMADAAVKI	(4)FSDKIK-	-VINKH-	STEV-TV	-GPEGD	MPCRANIL(27) enCEAVPH	Ra-tVYAOLVESGR	WSWNkl-fpIHVOTS
15895346	COG1189 CLOAB	79	LGLKDKV	FDIGAST	-G	FT-DCN	L-ONGASKV	F-ATD	VGYGOFAWKLE	-T-DPRVVC-	-MERTN-	VRYV-TP	-ED(3)	FGNFAS	4) fisLKKVTF	VVINLUKDDGET	VALI (22) DIEVINT
116930	Cog3 YEAST	123	DKRPEVS	UDVGCGG	-1	LS-EST	A-RLKWV-KNV	0-GTD	LTRDCTMVAKEH	(5) LEGKTN-	-YECKA-	LEDV-TG	-OFD	TTCMEMLe	h-vdmPSETLE	HCWSRUNPEKG	LFLSTINRDL
534839	CheR RHIME	121	KNCCRVR	WSAACSD		T	I.SLL(7) FPT	L-ATD	T DPKTLALARAC	(32) DRVKR_	TTTEN-	EINTMA	0W(5) FD	TECRNVy (RRAGUIDNCCH	VTGH(15) aTTTVRH
1252092	CogE VEAST	100	NETTOIN	TDVACCE	P P		T DUAR ST	DTR		(15) DPPUP	PT VC N	C FVI FF	TDEDE VI	TYTUEFai	rnftdtoret	AVRU VRCCT	VOIE(6) DIMDERVO
27151402	COQJ_IERSI	100	AL DOWDIN			m ww	T DEAL SKI	G-DIE		(1) DFRVR-	-FLV3N-	U IVE DI	DER N	Den aviant		ID - TED C AT	VEOD -ENTRY
15025700	Cide CENT	23	DENODI	CDVCPGT	- A	D C T	W THERA	T THE	C T NUPTORI	(4) OT OTHE		A	CW(A) W	CVANLPY (a) prorkubb	Rprrk-CAll	WAT V (20)
15925700	GIGB_STAAM	66	DFNQPIS	CDVGAGA	-r	PS-1P1	K-IMPPQ-LKV	T-1VD	SLNKRIQFLN-HI	(4)QLQDVS-	-FIHDK-	AETF-GK	-GV(4) XD	VVTARAV		SEICLPLIKKGGQF	VALK(38) MF11DKK
3310801	RTMZ_CAEEL	38	1LKGVRR	A DECAAP	-8	s-QVI	S-KKL(6)AK	V-AID	u(a)INGAIÖFÖCDI	SVD-	-TANQV-	KHF-SG	-5KSD	LVICUGAP(14) el ILA <mark>AF</mark> N	1TSHVLKEGGNE	LAKIFRSRNS
6323315	Nntl_YEAST	79	ELIKGKT	LE <mark>LG</mark> AAA	A-L	PS-VI <mark>C</mark>	A-LNGAQMV	V-STD	YPDPDL <mark>MQNI</mark> DYN <mark>I</mark>	(18)GNDYS-	-PLLAH-	- <mark>I</mark> EKI-GN	-NNGK-FD	<mark>LII</mark> LSDLvf	n-hteHHK <mark>LL</mark> Ç	QT <mark>T</mark> KDL <mark>L</mark> AEKGQA	LVVF(14) fELAKNE
6325100	Tgs1_YEAST	94	CMPNAER	LDVFCGG	-Gl	NT-IQ	A-MQFPY	r-GAD	YSIEHIYCTAKN	(4) VDDRIW-	-LKRGS-	WKKL-VS	-KQ(6)YD	VFGSPPw(19) mgITK <mark>ML</mark> B	SFLKLSPNV	MFLP(22) kVLYVKE
33301727	TrmA_VIBVU	207	TQESQGD	LELYCGN	-N	FS-LA <mark>I</mark>	A-QNFER <mark>V</mark>	L-A <mark>T</mark> E	LAKPS <mark>V</mark> ES <mark>A</mark> QYN <mark>I</mark>	AANKIEN <mark>V</mark> Q-	- <mark>1</mark> IRMS-	AEEF-TE	-AM(11)N <mark>(</mark>	JIDLKSYn (7)pprSGMD <mark>V</mark> D	DT <mark>C</mark> KMVQGYERI <mark>I</mark>	YISCnpETLKEN
1723920	YGN6_YEAST	69	KNRTDQR	LD <mark>LG</mark> YAP	-A	WS-QV <mark>Z</mark>	<mark>A</mark> R-QRSSPNSM <mark>I</mark>	L-G <mark>V</mark> D	<mark>I</mark> (5) PHGVNS IQAN I	(54)DDI <mark>Y</mark> E-	- <mark>T</mark> PNTN-	STLI-ER	-EKFPV-D <mark>V</mark>	VIISDMYe(35)dlCDA <mark>AL</mark> V	/T <mark>A</mark> IDL <mark>L</mark> RPLGS <mark>F</mark>	VCKL(36) yIGLKKK
1169870	Gcd14_YEAST	107	NCSPHSR	TE <mark>AG</mark> TGS	-S	FS-HA <mark>F</mark>	A-RSVGH	F-SFE	FHHIR <mark>Y</mark> EQ <mark>A</mark> LEE <mark>I</mark>	(5)IDDN <mark>V</mark> T-	- <mark>I</mark> THRD-	• <mark>V</mark> CQG-GF	-LIKKG-D	TTSYEFGn (41)spCIEQ <mark>V</mark> E	0K <mark>T</mark> LDV <mark>L</mark> EKY <mark>C</mark>	WTDV(7)rqYESRRQ
7298980	KOG1541_DROME	48	DDDESRL	ILDIGCGS	-L	SG-SV <mark>I</mark>	E-DSEHMW	I-GID	<mark>I</mark> SKSM <mark>L</mark> DI <mark>A</mark> VERE	EVA <mark>G</mark> D-	- <mark>V</mark> ILGD-	- <mark>M</mark> GEG-MP	-FKPGTFD <mark>C</mark>	<mark>GAI</mark> SISAl(107)aCRE <mark>A</mark> RG	K <mark>A</mark> PKKSRDW	LAKK(14) kYTARKR
585375	KsgA_BACSU	45	EVTEKTG	/IE <mark>IG</mark> PGI	-A	LT-EQI	A-KRAKK <mark>V</mark>	V-AFE	<mark>I</mark> DQRL <mark>L</mark> PI <mark>L</mark> KDT <mark>I</mark>	-S-PYENVT-	-VIHQD-	-VLKA-DV	-KS(12)- <mark>2</mark>	WVVANLPy-	yvTTP <mark>II</mark> M	K1-1EEHLPLKGIV	MLQKEVAERM
14335102	pf03141 ARATH	204	ENGTVRT	LDTGCGV	4-S	WG-AY	W-SRNV	R-AMS	FAPRDSHEAQVQ	-A-LERG <mark>V</mark> P-	-AVIGV-	LGTI-KL	-PY(4)FD	MAHCSRCli	pwganDGM <mark>YL</mark> M	E <mark>V</mark> DRV <mark>L</mark> RPGGY	ILSG(74)dVWYKKM
17017991	FtsJ3 HUMAN	41	FLQKARA	LDLCAAP	-G	WL-QV <mark>Z</mark>	A-KFMPVSSL	V-GVD	LVPIKPLPN	(4)EDIT <mark>T</mark> E-	-RCRQA-	LRKE-LK	-TWKVD	VVLNDGAp (14) hlTLMALF	RL <mark>A</mark> CDFLARGGS	ITKV(70)eLVTKKK
16126496	COG4798 CAUCR	59	GLKPKOT	TEVSPGG	-Y	WT-AII	A-PYAKAT	G-GTY	VAG-VADLANPKI	SEGAR-	-KGRAD-	FEAR-FA	-DEAKYGK	POFANFGD (27) gyADK <mark>VF</mark> A	DFFAVLKPGGVL	AVEE(20) tVIAIAE
15964298	COG4976 RHIME	144	RFRH	TDLCCGT	-L	FG-EP	R-DRADFT	E-GE	LSANMLAKAEAK	VVD-	-RLAO	DLSL-AP	-ED 90	SVEGALEC	15) gnJ.EGVEU	IADRITAPOCT	AFSV(6)doFVLRPS
127447	MTE1 HERAII		MOOFP	TDLEACT	-6	RR-LOT	E-AVGGVCV	3-92		PTDGVD	-HNLG	TTOT-00	LPAHD	VI.VGGVPc (16) ggLWADW	R	TRENVKCT VD
6320522	Omel VEAST	239	MKHCOCDI	LEVSCOT	-R		D_MSRTNOT	TEL	SSENMMETTUVV	(5) KYKKWA	-FVVGV	A(13)LEN_EV	ENOVEYD	TTVEAFC1	s-hodpvk	INRCKI.KPDCP	TLLE(5) WARTNET
10112200	COG1041 SCUPO	224	MAAPORT	VDPFUC		D T V	C PEC	T CE	T (E) POWNOPST	(5) CDFF	TETC	THC	DVNITTO	TUCDBR	26) dmuyn	E ACOPT VOCCO	WTWT (20)+UWCDDT
17200500	SUGIOTI SCRPU	2.34	CHIMIT	A DEP VOT			WADU -	-050	- (J) NORNORDINON	US BOOK			TRALLUNA	CDPPY(11)		COL
11389500	IT DIM_HUMAN	32	GNLINAY	LEVGFGP	-6	r-KS	-NADVABL		KDIRF IPGLOMLS	-APGKLR-	- vngD-	LIF-KV	- 5 A (3) LKI	ALMEDDAD (II) VSTPLII		VIGREQ-MELTFOKE
1/552504	NUGISSI CAEEL	45	QUSAGSI		1-K	1TS	NOC SHU	1-GFD	CSEVLSSSKK	DDID-	-LCLAD-	INI-PI	-KD(8)VS	LHHLATt-	argrovLQ	LECSKCLRIGGO	LLIA(132)MLGRRL
21112457	TPMT_XANCP	34	QLPAHAR	VPLCGK	с- Б]	uL-HW <mark>I</mark>	A-AQGHRV	H-GVE	LSPLAVTQ	(21)-GPIE-	-IICGD-	AFTL-DA	-SVLGDCT	AVIDKAAI (aaregryLE	arviaRLPAGCRO	TT(56)VYRLQRD
19115526	COG3129_SCHPO	70	GQKDKKR	GLDIGT	-A	SC-IYE	L-LGCR	Y-SYD	r (5) DKFS <mark>FETA</mark> KSN <mark>I</mark>	(6)SQIKIV-	-LRSKQ-	-uCLL-PD	TE(4)FT	rvmCNPPf-	yeffEED <mark>F</mark> I	NFKQNPPSG	CTGV(11)gFANKIL
7292846	KOG1540_DROME	97	GPTHGMR	LDMAGGT	dI	TF-RYI	R-YLN(9)SHV	T-VSD	INQHMLNVGEER	(11)NCTVA-	-WQCAD-	AEKL-PF	-PD(5)-Y	FIAFGIRnc	thVDKVLS	SE <mark>AYRVL</mark> QPGGR <mark>E</mark>	MCLE(4)tnETMQWL
127420	MtbA_BACAR	1	MKIK	VDLFAGI	-G	IR-IG <mark>I</mark>	E-RAAKRFE	L-ETE	C(6)DKKACETYALN <mark>E</mark>	KE-	-EPQGD-	- <mark>I</mark> HEI-TS	-FPEFD	FLLAGFPc (20)feVER <mark>VL</mark> F	RDnrpKAFLL-ENVR	LVTH(6)ktIISKLE
14023803	MetW_RHILO	17	LIPPQSR	LD <mark>VG</mark> SGD	-S	LL-EL <mark>I</mark>	Q-DTKQVD <mark>Q</mark>	R-GLE	LSQRG <mark>V</mark> NE <mark>C</mark> VA	RG <mark>L</mark> S-	- <mark>V</mark> IQGD-	ADKD-LE	-FY(4)FD	FVVLSQT1-	qATRNPS	Lv <mark>l</mark> del <mark>l</mark> rignr	IVSF(10)sLFVKGR
15828869	COG2189_MYCPU	472	GTEKNDL	LDFHLGS	-T	TTAV	H-KMGRK	I-GIE	QMDYIQDITIE	RMKK-	-VIDGE-	QGGI-SK	-NV(6)-S	FVYFELLe (17)reRAE <mark>TI</mark> S	ST <mark>I</mark> KNE <mark>I</mark> YNDER <mark>I</mark>	VPYI(15)nTEEKKK
27151628	NAS3 ARATH	121	GFVP-KS	AFIGSGPI	-P	LTsIVI	A-SFHLKDTI	H-NFD	IDPSANSLASLL	(5) ISQRMF-	-FHTVD-	IMDV-TE	-SLKS-FD	VVFLAALvg	mnkeeKVK <mark>VI</mark> E	EH <mark>L</mark> QKHMAPGAVI	MLRS(35) vVISKKH
2507512	Y116 HAEIN	186	GWQVGSP	VDPMCGS	-T	LL-IE	A-QME(8) RL	W-GFD	F(4)NQSAWEKVKNE	(6)KQREIQp	h <mark>F</mark> YGFD-	LDHR-VL	-KK(6)-N	AGVSHLIk-	wqqadva	ALKN-PRLNEV	TVIC(10)tPALIAL
7291671	KOG3987 DROME	122	SGQEPVT	LDIGAGD	-E	IS-LR	A-NTV(8)LRV	F-ATE	ASWTMRDRLKK	LNF-	-NVITE-	IDGL-QN	-LE	LILCLNV1d	r-cfdSFKLLE	DIHLALAPNGR	VVAL(4)mhYVETNT
1170712	KsgA HAEIN	36	OPNOFI	VEIGPGI.	-A	LT-FPL	G-ELVDHT	T-VVR	LDRDLAERLRHH	P-F-LHOK	-VIETD-	AMOF-DF (5)	KE(8)I-	RVFGNLPv(4)plmFHLF	KYHDVI-ODMH	MLOKEVVKRI.
2497109	YMA5 YEAST	272	PDICNND	VDLYAGT	-Y	FT-FS	L-TKGART	F-AFR	LNPWSVEGLKRGI	KAN-	-GFNKS-	GNCH-VF	-OE(2)-F	CVORLTO	4)npgFRLRTF	H INL GLI	PSSKggwPLATKL
68223981	Skb1 LEIMA	697	PDERILH	VLLCCCP	-P	LideC-LHZ	V-SAL(4)-PT	F-ATE	KNRLAAAFTP-ME	(14) GHTLE-		GRTIaTA	AE(7)AD	FGLCDLIV	10)elSPECT	AFHAOLEDIC	SRGIAFNPH
7387561	MTS1 RHIME	220	STKPODU	TDPFFC	-T	TC-3177	K-RLGP	V-CTV	REODYTDaaaro	(4) PLCKAT	-LSVM	GKKA-EP	-RVAFN	TLVESC		I KPCTV	
71900202	KOG3672 CAPPY	297	REPTEN	ADVCACE	-6-	PC PV	LWRKGEVNA	R-CP	L M VDRWY		-1550 -	FPFT_PV	GTKD	DVMDBUr -	del Paure	RCTSCT	MMAD(12) OTT SET
CONSERVE	D:			* * * *		5-51		*					*	*			(/ 2000000

Figure 2

Multiple sequence alignment of Rossmann-fold methyltransferases and nicotianamine synthase. Sequences are denoted by NCBI gi number, short protein name (when available, otherwise COG/KOG/Pfam number was used), and abbreviated species name (as in UniProt Knowledgebase [140]). Nicotianamine synthase is marked by a blue box. Conserved motifs are labeled above the alignment. Conserved residues are marked by asterisk. Consensus positions of the secondary structure elements are shown above the alignment. Numbers in parentheses indicate number of residues omitted for clarity. Residues are highlighted according to the amino acid properties. **Gray shading** indicates conservation of single residue. **Red font** indicates conservation of acidic residues (D, E). **Cyan font** indicates conservation of Ser/Thr (S, T). **Blue gray font with yellow shading** indicates conservation of aliphatic residues (I, L, V). **Dark blue font** indicates conservation of basic residues (H, K, and R). **Green font** indicates conservation of tiny residues (A, G, and S). **Blue font with yellow shading** indicates conservation of small residues (A, C, D, G, N, P, S, T, and V). **Bright blue font** indicates conservation of big residues (E, F, H, I, K, L, M, Q, R, W, and Y). **Black font with yellow shading** indicates conservation of big residues (E, F, G, H, I, L, M, T, V, W, and Y).



Figure 3

Structural alignment of cyclopropane fatty acid synthase and Rossmann-fold methyltransferase. Red color – YecO methyltransferase (PDB structure <u>IIM8</u>:A) **Yellow** color – ligand in YecO methyltransferase. **Green** color – mycolic acid cyclopropane synthase CmaA2 (PDB structure <u>IKPI</u>:A). **Blue** color – three ligands in CmaA2 (PDB structure <u>IKPI</u>:A). Motifs I-VI conserved in Rossmann-fold methyltransferases are labeled above the alignment.

ranging from moderately conserved residues scattered across the SAM-binding domain to bound water molecule [30], as well as $S_N 2$ reaction, which would require initiation by concerted action of several side chains, or, perhaps, by the amino group of the substrate itself [31]. The highly conserved D/E residue in motif I may, however, provide a unifying theme in the catalysis, by polarizing a water molecule that is close to the methyl group of SAM. The water molecule could either serve as a nucleophile, or aid bond displacement between the sulfonium ion and methyl group in some other way.

Finally, we noticed that the conserved basic residue at the beginning of Motif I and nearly-invariant acidic residue at the beginning of motif IV are typically located within a short distance (3Å or less) of each other, potentially forming a salt bridge that may be important for locking other elements of the Rossmann fold in place (Figure 2).

Rossmann-fold SAM-binding proteins that do not have methyltransferase activity

I. Methylene transferases

Formation of the cyclopropane ring in unsaturated fatty acids by cyclopropane fatty acid synthase [EC: 2.1.1.79] has been studied extensively in bacteria. The reaction involves transfer of a methylene group from SAM to the double bond of an unsaturated acyl chain [13,32].

Crystal structures of mycolic acid cyclopropane synthases CmaA1, CmaA2, PcaA, and MmaA2 have fold similar to Rossmann-fold methyltransferases (Figure 1b), with the conserved position of SAM and very similar pattern of interactions with the cofactor [32]. A hallmark of methylene transferases is the presence of the carbonate ion (CO₃ ²⁻) at the active center, which probably enables the formation of carbocation intermediate required for completion of the reaction (Figure 3) and ref. [32]. Conserved residues involved in carbonate ion binding (Cys35/Ser, His167/ Gln, and Tyr232/Phe – numbered as in PDB structure <u>1L1E</u>) appear to distinguish methylene transferases from Rossmann-fold methyltransferases.

II. Amino alkyl transferases: nicotianamine synthase and spermidine synthase

Nicotianamine synthase (S-adenosyl-L-methionine: Sadenosyl-L-methionine: S-adenosyl-L-methionine 3amino-3-carboxypropyltransferase, EC: 2.5.1.43) catalyses direct condensation of three molecules of SAM, followed by the formation of an acetidine ring, to yield one molecule of nicotianamine, a chelator of various transition metals ubiquitously present in higher plants. In graminaceous plants, nicotianamine is the precursor of phytosiderophores that are secreted from the roots to solubilize ferric iron in the soil. Reduced levels of endogenous nicotianamine affect the development of plant reproductive organs and seeds' maturation [33].

Protein structure of nicotianamine synthase is not known, but sequence similarity searches indicate a significant similarity between nicotianamine synthase and Rossmannfold methyltransferases. A PSI-BLAST search, using with Arabidopsis thaliana NAS [GenBank:O80483] as a query, matched Pseudomonas syringae ubiE/COQ5 methyltransferase [GenBank:YP_233497] with E-value = 1e-21 and score = 105 at 4th iteration; PDBBLAST top match is to (N5)-Glutamine Methyltransferase [PDB:1T43]. Similarity is the highest in conserved motifs II-IV and VI (Figure 2), and motif I is also conserved, in a modified form (F-x-G-S-G-P-x-P). Interestingly, related sequences with the same modification of Motif 1 are found in archaea (Methanothermobacter thermautotrophicus, [Gen-Bank:NP_275817], bacteria (Pseudomonas aeruginosa, [GenBank:NP_253523], and fungi (Neurospora crassa, [GenBank:XP_330777]. The replacement of the conserved D/E/N residue in motif I (see above) may partially explain the change in the functional group that is transferred from SAM: unlike the related Rossmann-fold methyltransferases, nicotianamine synthases lack negatively charged residue in Motif I, so the aminoalkyl moiety of SAM is not bridged to the enzyme by a water molecule and is free to leave in the course of the reaction.

Spermidine synthase (EC: 2.5.1.16) catalyzes the transfer of the aminopropyl group from decarboxylated SAM to putrescine to form spermidine. Putrescine, spermidine and spermine, formed from spermidine, are polyamines essential for the regulation of cell proliferation and differentiation in most species, and gram-negative bacteria outer membrane permeability in response to the acid stress [34,35]. Spermidine synthase is an oligomeric enzyme, each monomer consisting of a C-terminal domain with a Rossmann-like fold and an N-terminal tetramerization β -stranded domain [36]. Spermidine synthase has very high sequence similarity (approx. 70% identity) to putrescine N-methyltransferase. It has been shown that aminopropyl moiety of spermidine synthase inhibitor - AdoDATO (a compound containing both substrate and product moieties) binds in a similar orientation to the homologous part of SAM in Rossmann-fold methyltransferases. However, the binding site in spermidine synthase contains invariant residue Asp101 (PDB structure <u>1IQ3</u>), located in the middle of glycine-rich loop (corresponding to motif I of Rossmannfold methyltransferases) making binding cavity too small to accommodate the carboxyl group of SAM, that prevents SAM binding and enables specific binding of decarboxylated SAM [36]. The polyamine moiety of AdoDATO is oriented toward putrescine binding cleft. Invariant Asp170 (corresponding to D/N residue located at the end of β-sheet in motif-IV of Rossmann-fold methyltransferases) is most likely candidate to deprotonate putrescine, enabling it to perform a nucleophilic attack on methylene carbon of decarboxylated SAM [36].

III. Acalcynomycin-10-hydroxylase

Aclacinomycin 10-hydroxylase RdmB is a methyltransferase homolog that catalyses a SAM-dependent hydroxylation of the C-10 carbon atom of 15-demethoxy- ϵ rhodomycin, a step in the biosynthesis of the polyketide antibiotic β -rhodomycin. In order to activate molecular oxygen, the enzyme uses SAM instead of cofactors usually associated with hydroxylase activity, such as flavins, 2oxoglutarate, or metal ions. It has been proposed that positive charge of the SAM cofactor plays a role in delocalisation of electrons into the anthraquinone core of the substrate [37]. The C-terminal domain of RdmB has an α / β Rossmann-like fold, and contains the conserved signature DLGGGxG in motif I.

The enzyme lacks methyltransferase activity due to the positioning of SAM in which the methyl group points in a direction unfavorable for a S_N^2 type methyl transfer to the substrate [37]. The difference in SAM-substrate positioning is not well understood, but might be related to differential interactions between SAM binding C-terminal domain and substrate binding N-terminal domain or differences in the conserved loop (amino acids 292–298) [37].

Non-catalytic Rossmannoids

The lack of detectable SAM binding motifs in several Rossmann-fold methyltransferases suggests that they may be recruited for a new function. For example, the threedimensional structure of sc-mtTFB (Saccharomyces cerevisiae mitochondrial transcription factor B) bears strong resemblance to ribosomal RNA adenine dimethylases (i.e. KsgA and ErmC'). However, several residues required for interaction with SAM are not conserved in sc-mtTFB; in particular, the glycine-rich loop (motif I) contains bulky Tyr residue, and motif IV is poorly conserved.

Human co-orthologs of sc-mtTFB (h-mtTFB1 and h-mtTFB2) have rRNA N6-adenine methyltransferase activity (in an *Escherichia. coli* assay), but mutational analysis of h-mtTFB1 indicates that this activity is not required for transcriptional activation [38].

In Gcd10p/Gcd14p complex – tRNA(1-methyladenosine) methyltransferase of *S. cerevisiae*, the lack of SAM binding was observed in Gcd10p that directs binding of tRNA, where Gcd14p binds the required cofactor *S*-adenosylmethionine [39,40].

In another case, Kar4p (pheromone induced, karyogamyspecific transcription factor) does not bind SAM, where similar (circularly permuted) Ime4p is SAM-binding methyltransferase [41].

In bacterial rRNA:m2G methyltransferases RsmC and RsmD the inactivated domain and the catalytic domain are fused together in one polypeptide [42].

The *de novo* methyltransferase-like protein, DNMT3L, is required for methylation of imprinted genes in germ cells. Although enzymatically inactive, human DNMT3L accelerates DNA and SAM binding to *de novo* DNA methyl-transferases [43-45].

Rossmann-like domain of bacterial fluorinating enzyme

Actinomycete Streptomyces cattleya is able to produce C-F bonds using inorganic fluoride. The fluorinating activity requires SAM, and the primary product of the reaction is SAM derivative, 5'-fluoro-5'-deoxyadenosine [46]. The protein, 5'-fluoro-5'-deoxyadenosine synthase FlA, belongs to a conserved sequence family represented in most archaea and in a subset of bacteria [46].

The structure of FIA consists of two domains - a larger Nterminal domain with α/β fold, and a smaller C-terminal β-barrel. Both domains interact with SAM and with reaction products [46]. FlA is a hexamer in solution and trimer in crystal, and three SAM molecules are bound by a trimer, between the N-terminal domain of one subunit and the C-terminal domain of the adjoining subunit. This arrangement, however, appears to be dependent on a long (24 amino acids) loop in the N-terminal domain, which is missing from the closely related sequences in all other species. On the other hand, the linker connecting two domains in a monomer is long enough to allow significant domain motions, and it is plausible that two domains may interact in other oligomeric arrangements and perhaps even within a monomer. Therefore, we speculate that SAM binding by FlA-like proteins from other species may occur in the crevice formed by the N- and Cterminal domains of the same molecule, and the following discussion does not consider the oligomerization state.

The N-terminal domain makes contacts mostly with methionine, ribose, and fluoride ion, and C-terminal domain contacts methionine and adenine ring. The α/β N-terminal domain exhibits several features similar to other enzymatic domains with Rossmann-like topology, namely: three-layer $\alpha/\beta/\alpha$ architecture; the planar central, mostly parallel β-sheet filling "inside-out" (strand topology 2135467), and concentration of the substrate-binding and catalytic residues in the loops following the C-termini of strands. More specifically, the loop after strand 1 contains Asp16 (numbered as in PDB structure <u>1ROP</u>) hydrogen-bonded to both hydroxyls of ribose, Leu17 that may be involved in water-mediated interaction with methionine amino group, Asp21 and Ser23, both of which can form hydrogen bonds with the same amino group. Loop following strand 2 contains Trp50 that is able to contact one ribose hydroxyl and perhaps to have Van der Waals interactions with the adenine ring. Loop between strands 3 and 4 contain two ribose interactors, Thr76 and Tyr77. Loop after strand 6 hosts Thr155, which is part of hydrogen-bond network linking two domains via two water molecules and methionine carboxyl group, as well as catalytic Ser158 that is expected to make two polar contacts with deprotonated fluoride ion [46]. Although most of these interactions are provided by residues located in loops at the edge of β -sheet, there is no specific sequence similarity between Rossmann-like domain in fluorination enzyme and in Rossmann-fold methyltransferases. There is also no similarity to other SAM-utilizing enzymes.

Rossmann-like fold in SPOUT methyltransferases?

A distinct superfamily of SAM-dependent methyltransferases, SPOUT, which includes families specified by bacterial SpoU, TrmD, and TrmH, proteins, as well as many uncharacterized proteins in all three domains of Life, have been shown to share a set of conserved sequence elements and an α/β -type fold [47]. Trm10, a recently characterized tRNA m¹G₉ methyltransferase, is also predicted to have this fold [48,49]. All experimentally characterized members of this large superfamily are DNA or RNA methyltransferases. A unique structural feature of this α/β fold is a trefoil knot of two crossing loops in the C-terminal region [47].

Several hallmarks of Rossmann-like structure are evident in the SPOUT fold. There are three main layers, with a central β -sheet sandwiched between two helical layers; the β sheet is formed "inside-out", with the first and one of the last strands in the center of the sheet; and the SAM ligand interacts mostly with the conserved residues located in the loops at the C-termini of β -strands [50]. There is, however, no sequence similarity between SPOUT-fold methyltransferase and any other Rossmann-fold SAM-binding protein.

Structural similarity between SPOUT-fold and Rossmann-fold methyltransferases (i.e. DALI Z-score = 3.1 for 88 aa with RMSD = 3.7 and sequence identity = 17% for PDB structures <u>1QAO</u> and <u>1MXI</u>) is confined mainly to the N-terminal half of those folds. There is no similarity in the C-terminal part, where strands 4 and 5 are rearranged.

SAM binding inside and outside of β -barrels

SAM-radical enzymes: recruitment of ancient enzymatic TIM barrel A $(\beta/\alpha)_8$ fold, also known as triose phosphate isomerase (TIM)-like barrel, is one of the largest classes of protein structures, exceeding even Rossmann-like fold in omnipresence and versatility [51]. Most of TIM-barrel proteins are enzymes, belonging to almost all of the major EC classes [52]. A well-known version of a TIM barrel is a $(\beta/\alpha)_6$ "semi-barrel," in which the inner layer of slanted β -strands does not form a complete cylinder, but has a lateral opening (PDB structure <u>1OLT</u>).

Recently, several structures of proteins from a large sequence family of "SAM radical" enzymes (ref. [53]; Figure 1g) have been determined, and it became evident that members of this family have $(\beta/\alpha)_6$ and $(\beta/\alpha)_8$ folds. SAM-radical enzymes utilize non-covalently linked Fe-S cluster and a SAM molecule, in a reductive cleavage reaction that produces methionine and 5'-deoxyadenosyl radical, that can be used to generate further glycyl or thiyl radicals on the same protein molecule or on a coupled enzyme [54]. It has been noted that SAM-radical sequence family is very large, diverse, but can be recognized by a hallmark CxxxCxxC signature close to the N-terminus, followed by another conserved "GG" motif [55].

We detected more than 2000 non-redundant sequences from SAM-radical family in the sequence databases. Interestingly, when the region containing the three characteristic cysteines was deleted from the queries, the searches resulted in almost the same collection of sequences as with full-length domain, indicating strong evolutionary signal along the stretch of 200-250 residues to the C-terminal side of the CxxxCxxC signature. Multiple alignment of many representative sequences identified four regions with high sequence similarity and three weaker conserved motifs (Figure 4). Comparison of the alignment with the known structures of biotin synthase (PDB structure 1R30), coproporphyrinogen III oxidase (PDB structure 10LT) and molybdenum cofactor biosynthesis enzyme MoaA (PDB structure 1TV7) suggests structural and functional correlates for these regions and for the most conserved residues within them. The best-conserved motifs correspond to the β -strands of the inner barrel and their C-terminal loops, while the regions of additional partial conservation correspond to the outer-shell α -helices.

Motif 1 includes the most N-terminal β -strand in the (β/α)₆ barrel (strand 1). Three invariant cysteine residues in the adjoining loop (Figure 4) coordinate the [Fe-S]₄ cluster, which is present in a similar configuration in all protein structures resolved thus far. One iron atom has no contact with the cysteine side chains, and is instead ligated by the N and O atoms from the amino- and carboxy groups of SAM. Also highly conserved is aromatic or heterocyclic residue (Y, F, or H) preceding the last of the three cysteines; the main chain of this residue seems to form a hydrogen bond with the adenosyl moiety of SAM, but the significance of the side chain conservation is unclear; perhaps it contributes to the non-polar milieu of the bound Fe-S cofactor, preserving it from oxidation.

The second prominent motif does not contain any invariant amino acids, but includes several residues with small side chains, most often two or three glycines in a row (Figure 4). This motif corresponds to the second strand in the barrel and the tight turn after the strand. The main chain of this turn is within a contact distance from the amino group of the methionine part of SAM. The third motif also corresponds to the strand-turn structure. A signature T/S-N-G that follows strand 3 is well conserved; as a rule, residues in this turn form hydrogen bonds with the carboxyl group of methionine in SAM. The fourth motif consists of strand 4 and the loop with a highly conserved acidic or amide residue (D, E, N, or Q). Typically, this residue is within hydrogen-bonding range from both 2'- and 3'hydroxyl groups of the ribose ring of SAM. The strands 5 and 6 followed by loops provide one or more residues that form hydrogen bonds with the amino group of adenosyl; however, sequence conservation in these regions is moderate.

The heterogeneity of the SAM radical protein superfamily is most pronounced in their C-terminal regions, which are responsible for binding of substrates and auxiliary cofactors. On the other hand, the structure and sequence of the N-terminal, SAM-binding region of SAM-radical proteins is well conserved, analogously to the Rossmann-fold methyltransferases. The SAM-binding region is essentially an incomplete (β/α)₆ "semi-barrel," which is typically modified by evolutionarily diverse elements (commonly consisting of α -helices, but sometimes also containing β hairpins or small sheets) that serve substrate-binding and regulatory roles.

Rossmann folds and TIM-barrels in fact have quite similar β/α architectures. This becomes especially evident in the case of incomplete barrels. The primary difference is lack

				_1	motif	-1							n	notif-2	2						motif-3
gi#	COG/KOG#	<u>name</u>	<u>specie</u>	s				hhhh	hhhh	hhhhh	hhhhh	hh	SSSS	SSS	hh	hhhhhh	hhhhh	hhhhh	hhh :	ssss	SSSSS
18409989	KOG4355		ARATH	206	LGA	T <mark>Y</mark> C	K(6)	HL	3S <mark>Y</mark> TV	DS <mark>L</mark> VEF	R <mark>VRTV</mark> I	SEG	VKE <mark>]</mark>	WLSS-	E <mark>D</mark> (14)LLN	A	I <mark>V</mark> KE <mark>L</mark> I	PSDQS	FMLR	IG <mark>M</mark> TNPPF
7299338	KOG2492		DROME	250	DNM	T <mark>Y</mark> C	I(11) PLAS	(<mark>v</mark> aev	KA <mark>L</mark> AEÇ	<mark>g</mark> vke <mark>v</mark>	TLL(6	5)YRDF	RTAQE (6)AT(14)GTP	F <mark>A</mark> AL- <mark>I</mark>	L <mark>RSV</mark> A	QA <mark>V</mark> PD	MRIR	FTSP <mark>H</mark> PKD
2501533	COG0621	¥019	HAEIN	156	<u>NKY</u>	T <mark>F</mark> C	V(11) PVDD'	/ <mark>lf</mark> ei	AQ <mark>L</mark> AEÇ	<mark>GVRE</mark> V	NLLG-	QNVI	IA <mark>Y</mark> RG	P T -	HDGQIC	S <mark>F</mark> AE-I	L <mark>LRL</mark> V	AS <mark>I</mark> DG	IDR	LR <mark>F</mark> TTSHP
20094331	COG1032		METKA	196	• PFR•	LFC	A(5)	-GPKW	R <mark>GK</mark> SP	DR <mark>VVEE</mark>	ISYL <mark>v</mark>	EELG-	VER <mark>I</mark>	EFVD-	DV-	FTAHKR	R <mark>V</mark> RE – 1	I <mark>C</mark> EK <mark>M</mark>	REE	G	ID <mark>VP</mark> WDCG
15896586	COG4277		CLOAB	63	IYD	Q <mark>Y</mark> C	V(5)	-DTPR	AS <mark>F</mark> TP	RE <mark>L</mark> ADI	. <mark>TMNF</mark> Y	RRNY-	IEG <mark>I</mark>	FLSS-	GV-	MKNPDY	T <mark>C</mark> EQ-1	MIEVL	RI <mark>L</mark> RY	EYN	FY <mark>G</mark> YIHAK
3913815	COG0635	HEMN	AQUAE	60	EQR	L <mark>Y</mark> C	G(10)IEEP	[LERV	CREMDI	' <mark>nsðat</mark>	DKDR-	-kviq <mark>i</mark>	HWGG-	GTP	NYLSPE	Q <mark>I</mark> KW- <mark>I</mark>	F <mark>MEE</mark> I	RKRFE	FGDNAE	IS <mark>IEL</mark> DPR
14520767	COG2100		PYRAB	124	NMR	IFC	S(9)	-RKLD	7 <mark>V</mark> VDI	DY <mark>L</mark> LKW	I <mark>F</mark> DWV <mark>A</mark>	KEK(4	l)EAH <mark>I</mark>	DAQG-		EPLLYP	F <mark>I</mark> VE- <mark>I</mark>	L <mark>VQA</mark> L	RE <mark>H</mark> PH	vsv	IS <mark>MQS</mark> NGV
11499591	COG0535	NirJ2	2_ARCFU	123	NLR	KHC	-(5)	-KPLQ	DE <mark>LTT</mark>	DE <mark>A</mark> LEI	'IDK-I	DRLG-	VTI	AFSG-	GE-	-PLVRR	D	I FELT	RY <mark>A</mark> AE	KGIY	VAIATNGT
15833984	COG0641	AslB	ECOLI	108	NLN	DYC	F(9)	-EKPV	C <mark>HM</mark> DD	DT <mark>L</mark> EAY	(VRHY1	AAS (4	l)EVA <mark>I</mark>	TWQG	GEP	TLLGLE	F <mark>Y</mark> RRAV	V <mark>A</mark> LQ <mark>A</mark>	KY <mark>G</mark> AG	RKIS	NS <mark>FQT</mark> -NG
549552	COG0820	YFGB	ECOLI	124	ALE	KFC	S(13)EIIG	2 <mark>VW</mark> RA	AK <mark>I</mark> VGA	AKVT (QRP	ITN <mark>N</mark>	VMMG	-MGE-	PLLNLN	N <mark>V</mark> VP-4	A <mark>MEIM</mark>	LDDFG	FG(4)R	VTLST-SG
15606957	COG2516		AQUAE	54	HAT	AYC	G(15) RVEW	P <mark>T</mark> VKT	DE <mark>IIE</mark> F	R <mark>TKQV</mark> G	H	AER <mark>I</mark>	-QISQ-	IT-	HPRAIR	D <mark>T</mark> KY-1	I <mark>L</mark> EKV	HRELG	DKIFIS	<mark>ILLNA</mark> -TG
1705463	KOG2900	BIOB	ARATH	93	SED	SYC	P(8)	-VKAQ	R <mark>LM</mark> SK	DA <mark>VIDA</mark>	AKK-A	KEAG-	STR	CMGA-	-AWRD	TIGRKT	N <mark>F</mark> SQ-1	ILEYI	KE IRG	MGME	VCCTLG
15606163	COG1060	_ThiH	AQUAE	66	IAG	KFC	A(6)	-DEDA	Z <mark>VL</mark> DM	DT <mark>ILQ</mark> K	(VKELV	EWGG-	TT <mark>I</mark>	LMQG-	GIN	PDLPLE	F <mark>Y</mark> ED-1	MISQI	KK <mark>H</mark> FP	2 v 0	IHCFSAPE
19705171	COG1509	KamA	FUSNN	126	SMY	RHC	T(6)	SSDI	D <mark>AM</mark> PM	DRIDKA	IEYIA	KTPQ-	VRD	LLSG-	GDA	LLVSDK	KLES-	IIQKL	RAIPH	VEI	IR <mark>I</mark> GSRTP
18313168	COG1313	_P±1X	PYRAE	158	NFR	VY	Q(6)	-FSES	JEVT.	AEALAA	IQAKI	REEG-	ARN	NWVG-	GE-	PTP	NIPY-	ILESL.	RILAR	RGVN	VPQLWNSN
730103	COG2896	MOAA	BACSU	26	NFR	TY .	M(11) LKKE	ELLSF	EELERI	ATLFV	TRFG-	VEK	RLTG-	GE-	-PLMRK	DMPE-I	LIKKL	ARIPG	IRD	IAMTTNGS
22027536	KOG2876	MOCS	L_HUMAN	79	NLR	QY	-(8)	-TPKA	LLTT	EEILTI	ARL-F	VKEG-	IDK	RLTG-	GE-	-PLIRP	DVVD-	IVAQL	QR <mark>L</mark> EG	RT	IGVTT-NG
14521276	COG2108		PYRAB	46	PRN	FY	P(11) ANER	P <mark>VR</mark> SE	KD <mark>I</mark> IEE	AKIQI	ARGA-		-GITG-	GD-	PLARLS	RTVE-	Y I RLL	KEEFG	KKFH	IHLYT-TG
22001767	COG0320	LIPA_	YERPE	93	TRR	PF	DV	-AHGR	VTPD.	ANEPER	LAQTI	QDMG-		VITS-	VDR	DDLRDG	GAQH-H	FADCI	SAIRA	KNPT	IKIETLVP
17865692	COG0502	BIOB	SALTY	52	PED	KY	P(8)	-LEAE	R <mark>LME</mark> V	EQVLDS	ARK-A	KNAG-	STR	CMGA-	-AW	RNPHER	DMPY-1		QGVKA	MGLE	TCMTLG
14520966	COGII80	PILA	PYRAB	81	NMH	KH	Q(6)	ADEN				KHYE-	CES	AYTY-		PTI	WYEF-V		KLAKK		NILVT-NG
134/6168	COG1533	_SDIR	RHILO	74	EHG	VY	F(14	DFET		PDAARL		SKEGY	QPRT	AIGT-	NT (4) PIEK	QYRI-I		EVLEA	KGHP	
15605904	COG1856		AQUAE	59	ELM	DH	ASKI	LWHMI	ATTP			KKKG-	IEG		GS-	DKNGFV		FFDAM	KYLKE	SLG	FLLTCHVG
15043710	000014	Nado	ACDE	110	DVD	A I	W(O)	CEVD				CT SOC		RISG-	CN CN	PILGER					
13691369		_NIG	AGRID	43		3 <mark>0</mark>	D(0)	-SEIR	JIWKP	M <mark>S</mark> VEA1	MOE AT	GT966	-MPLI <mark>N</mark>	<mark>апас-</mark> .	GIV-	P.	A <mark>IQP-I</mark>		HK <mark>G</mark> QG		ranta 1062
conserve	a				· ·																
			_mo	tif-4							_			me	otif-	5				mot <u>i</u>	<u>f</u> -6
10400000	hhhl	hh	_mo ssss	tif-4	hh	hhh	hhhh		hhh	hhhh—h	1h	hhhhh	1 59	m	otif-	5 hh	-hhhhi	hh1	hh s	mot <u>i</u> ssss	<u>f</u> -6
18409989	hhhl 1(5)EI <mark>A</mark>	hh AV(4)(<u>mo</u> ssss C <mark>V</mark> YT <mark>F</mark> LH	tif-4 sss VPVQS	hh GSD	-hhh -SV <mark>L</mark>	hhhh TAMN	REYTA	hhh SE <mark>F</mark> RT	hhhh-h V <mark>V</mark> DT	1h	hhhhh EL <mark>V</mark> PG	n se G <mark>M</mark> Ç	m ssssss IATDI	otif- I <mark>C</mark> -GF	5 hh PGETDE	-hhhhl D <mark>FSQT</mark>	hh	hh s YKFPQ	mot <u>i</u> ssss VHISQF	<u>f</u> -6 Ypr(208)
18409989 7299338	hhhl I(5)EIA $F(5)EVII$	hh AV(4)(RD-HPN	_mo ssss CVYT <mark>F</mark> LH VCKQLH	tif-4 sss VPVQS LPAQS	hh GSD GNT	-hhh -SVL -QVL	hhhh TAMN ERMR	REYTA: RGYSRI	hhh SE <mark>F</mark> RT SAYLE	hhhh-h V <mark>V</mark> DT L <mark>V</mark> QH	h LT IR	hhhhh EL <mark>V</mark> PG QF <mark>L</mark> PN	n se 3 <mark>M</mark> Ç 1 <mark>V</mark> C	m sssss lATDI LSSDF	otif- IC-GF IC-GF	5 hh PGETDE CGETEE	-hhhhł DFSQT EFQDT	hh /(5) /SL	hh s YKFPQ IQQVQ	mot <u>i</u> ssss VHISQF YNVAYL	<u>f</u> -6 Ypr(208) FAY(129)
18409989 7299338 2501533	hhhl I(5)EIA I(5)EVII I(7)DVYI 2(7)EVYI	hh AV(4)(RD-HP1 RD-TP1	<u>mo</u> ssss VYT <mark>F</mark> LH VCKQLH LVSFLH	tif-4 sss VPVQS LPAQS LPAQS	hh GSD GNT GSD	-hhh -SVL -QVL -RVL	hhhh TAMN ERMR TMMK	REYTA: RGYSRI RAHTAI	hhh SE <mark>F</mark> RT EAYLE LE <mark>Y</mark> KS	hhhh—h V <mark>V</mark> DT LVQH IIRK	ih LT LR LR	hhhhh EL <mark>V</mark> PG QF <mark>L</mark> PN AVRPD	n se GMG JVC DIG	m SSSSSS IATDI LSSDF ISSDF	otif- IC-GF IC-GF IV-GF	5 hh PGETDE CGETEE PGETAE	-hhhhi D <mark>FSQT</mark> EFQDT DFEQTI	hh] V(5)-; VSL; MNL;	hh s YKFPQ IQQVQ IAQVN	moti ssss VHISQF YNVAYL FDMSFS	<u>f</u> -6 YPR(208) FAY(129) FVY(135)
18409989 7299338 2501533 20094331	hhhl I(5)EIA F(5)EVI I(7)DVY A(7)ELA A(5)DT	hh AV(4)(RD-HPI RD-TPI RT(4)(sss VPVQS LPAQS LPVQA GAESA	hh GSD GNT GSD SN	-hhh -SVL -QVL -RVL -ETL	hhhh TAMN ERMR TMMK KRIN	REYTA: RGYSRI RAHTAI KGIT	hhhi SEFRT EAYLE LEYKS VQD	hhhh	1h LT LR LR	hhhhh EL <mark>V</mark> PG QF <mark>L</mark> PN AVRPD KR <mark>H</mark> G-	1 SS M V I I I I	m SSSSSS IATDI LSSDF ISSDF ISSDF	otif- IC-GF IC-GF IV-GF IL-GF	5 hh PGETDE CGETEE PGETAE PWEDRE	-hhhhi DFSQTY EFQDTY DFEQTH DVFRTJ	hh	hh s YKFPQ IQQVQ IAQVN LEPDY	moti ssss VHISQF YNVAYL FDMSFS VQFTVC	f-6 YPR(208) FAY(129) FVY(135) TPY(86)
18409989 7299338 2501533 20094331 15896586	hhhi I(5)EIA F(5)EVI I(7)DVY A(7)ELA A(5)DTL	hh AV(4)(RD-HPI RD-TPI RT(4)(ID(4)I		tif-4 sss VPVQS LPAQS LPVQA GAESA	hh GSD GNT SN PSQ	-hhh -SVL -QVL -RVL -ETL -KSL	hhhh TAMN ERMR TMMK KRIN KLLA	REYTA; RGYSRI RAHTA] KGIT PNKS ()	hhhi SEFRT EAYLE LEYKS VQD 5)PMG	hhhh—h VVDT LVQH IIRK VIACRK YIKN-K	h LT LR LR VA S	hhhhh ELVPG QFLPN AVRPD KRHG- ENTT (n ss GW NV OI C (13) <mark>A</mark> C	m SSSSSS IATDI LSSDF ISSDF LLLSF QSTQM	otif- IC-GF IC-GF IL-GF IL-GF	5 hh PGETDE CGETEE PGETAE PWEDRE PDTD (5	-hhhhi DFSQTY EFQDTY DFEQTI DVFRTI)LTESI	hh V(5) VSL I(5) LVSK VSK	hh s YKFPQ IQQVQ IAQVN LEPDY YKLKR	moti ssss VHISQF YNVAYL FDMSFS VQFTVC VFFSAY	f-6 YPR(208) FAY(129) FVY(135) TPY(86) MPV(181)
18409989 7299338 2501533 20094331 15896586 3913815	hhhi I(5)EIA F(5)EVI I(7)DVY A(7)ELA A(5)DTL YLTDEQI	hh AV(4)(RD-HPM RD-TPM RT(4)(ID(4)I KA(4)(FF(4))		tif-4 sss VPVQS LPAQS LPVQA GAESA VNIEL	hh GSD GSD SN PSQ LDP	-hhh -SVL -QVL -RVL -ETL -KSL -KVQ	hhhh TAMN ERMR TMMK KRIN KLLA QAVN	REYTA: RGYSRI RAHTAI KGIT PNKS() RVQP()	hhh SEFRT EAYLE LEYKS VQD 5) PMG 5) KME	hhhh—h VVDT LVQH IIRK VIACRE YIKN-E KLRE	h LT LR VA (S	hhhhh ELVPG QFLPN AVRPD KRHG- ENTT (AGFES	n se GNg NVC DIg (13)AC G	M SSSSSS IATDI LSSDF ISSDF ILLSF QSTQM INLDL	otif- IC-GF IC-GF IL-GF ILGAT IY-GL	5 hh PGETDE CGETEE PGETAE PWEDRE PDTD (5 PYQTKE	-hhhh DFSQTY EFQDTY DFEQT DVFRTI)LTESI SFEKTY	hh V(5) VSL I(5) I(5) LYSK V(7)	hh s YKFPQ IQQVQ IAQVN LEPDY YKLKR PDRIA	moti ssss VHISQF YNVAYL FDMSFS VQFTVC VFFSAY IYSFAY	f-6 PR(208) FAY(129) FVY(135) TPY(86) MPV(181) IPQ(208) RPV(110)
18409989 7299338 2501533 20094331 15896586 3913815 14520767	hhhi I(5)EIA F(5)EVI I(7)DVY A(7)ELA A(5)DTL YLTDEQI LLNDKLV	hh AV(4)(RD-HPI RD-TPI RT(4)(ID(4)I KA(4)(EE(4)(DELISING	hh GSD GSD SN PSQ LDP LD	-hhh -SVL -QVL -RVL -KSL -KSL -KVQ -PEK	hhhh TAMN ERMR TMMK KRIN KLLA QAVN AKML	REYTA: RGYSRI RAHTAI KGIT PNKS() RVQP() MGIK()	hhh SEFRT EAYLE LEYKS 5) PMG 5) KME 5) VLE	hhhh—h VVDT LVQH IIRK VIACRE YIKN-E KLRE MAEA	LT LT LR (VA (IS LV	hhhhh ELVPG QFLPN AVRPE KRHG- ENTT (AGFES NAGIE	$ \frac{1}{2} = -\frac{M}{2} $	M SSSSSS IATDI: LSSDF ISSDF ILLSF QSTQM INLDL IAPVI	otif- IC-GF IV-GF IL-GF I IG AT IY-GL MF-GV	5 hh PGETDE CGETEE PGETAE PWEDRE PDTD (5 PYQTKE NDDE (6	-hhhhh DFSQTY EFQDTY DFEQTY DVFRTJ)LTESI SFEKTY)FARRJ	hh V(5) MNL I(5) LYSK V(7) IGAG IGAG	hh s YKFPQ IQQVQ IAQVN LEPDY YKLKR PDRIA KRWPA	moti ssss VHISQF YNVAYL FDMSFS VQFTVC VFFSAY IYSFAY LGFQNY	<u>f</u> -6 FAY(129) FVY(135) TPY(86) MPV(181) IPQ(208) IPY(110) IPY(117)
18409989 7299338 2501533 20094331 15896586 3913815 14520767 11499591	hhhi I(5)EIA F(5)EVII I(7)DVYI A(7)ELAI A(5)DTL YLTDEQII LLNDKLVI LITEEIAI	hh AV(4)(RD-HPM RD-TPF RT(4)(ID(4)I KA(4)(EE(4)(KR(4)(tif-4 sss VPVQS LPAQS GAESA VNIEL LGVQD LSIHS ISLDG	hh GSD GSD SN PSQ LDP LD MK	-hhh -SVL -QVL -RVL -ETL -KSL -KSL -KSL -ETH	hhhh TAMN ERMR TMMK KRIN KLLA QAVN AKML EAFR	REYTA; RGYSRI RAHTAJ KGIT PNKS() RVQP() MGIK() GIR()	hhh SEFRT SAYLE LEYKS VQD 5)PMG 5)KME 5)KME 5)VLE SCFDK	hhhh	h IR LR VA IS IS IS IS IS	hhhhh ELVPG QFLPN AVRPD KRHG- ENTT (AGFES NAGID NAVKI	1 SS GW OI (13)A GV OVI GLF TG-LF	M SSSSSS IATDI LSSDF ISSDF ILLSF QSTQM INLDL IAPVI VNVSM	otif- IC-GF IC-GF IL-GF IIGAT IY-GL MF-GV TVTRY	5 hh PGETDE CGETEE PGETAE PWEDRE PDTD(5 PYQTKE NDDE(6 NYHEVP	-hhhhi DFSQTY EFQDTY DFEQTI JETESI SFEKTY JFARRI KVVELC	hh V(5) VSL I(5) LYSK V(7) IGAG CEK ZEK	hh s YKFPQ IQQVQ IAQVN LEPDY YKLKR PDRIA KRWPA LGVNW	moti ssss VHISQF YNVAYL FDMSFS VQFTVC VFFSAY IYSFAY LGFQNY FMHYNF	f-6 YPR(208) FAY(129) FVY(135) TPY(86) MPV(181) IPQ(208) IPY(110) IPY(110) IPY(1210)
18409989 7299338 2501533 20094331 15896586 3913815 14520767 11499591 15833984	hhhi I(5)EIAJ F(5)EVII I(7)DVYI A(7)ELAI A(5)DTL VLTDEQII LLNDKLVI LITEEIAI V-LLDDE V(4)DE	hh AV(4)(RD-HP RD-TPF RT(4)(ID(4)I KA(4)(EE(4)(KR(4)(WC(4)F		tif-4 sss VPVQS LPAQS GAESA VNIEL LGVQD LSIHS ISLDG LSLDG	hh GSD GSD SN PSQ LDP LD MK PAEI- BND-	-hhh -SVL -QVL -RVL -ETL -KVQ -PEK -ETH -HNQ	hhhh TAMN ERMR TMMK KRIN KLLA QAVN AKML EAFR YRVT	REYTA; RGYSRJ RAHTAJ KGIT PNKS(RVQP((MGIK() GIR KGGR-J BINWE	hhh SEFRT SAYLE LEYKS VQD 5) PMG 5) ME 5) VLE GCFDK STHKL	hhhh-h VVDT LVQH IIRK VIACRE YIKN-E KLRE MAEA TVEG VMRA SIAA	h LT LR VA IS LV LV LT VP	hhhhh ELVPG QFLPN AVRPD KRHG- ENTT (AGFES NAGID NAVKI L-LQK	1 SS GV OI (13)A GV CGL CGL CGL CGL CH	M SSSSS IATDI SSDF ISSDF ILLSF QSTQM INLDL IAPVI VNVSM VNVSM VNVSM	otif- IC-GF IC-GF IL-GF IIGAT IY-GL MF-GV TVTRY LV-CV	5 hh PGETDE CGETEE PGETAE PWEDRE PDTD (5 PYQTKE NDDE (6 NYHEVP NRTSAQ	-hhhh DFSQTV EFQDTV DFEQTI DVFRTJ)LTESI SFEKTV)FARRJ KVVELC QPLQV	hh V(5) VSL I(5) LYSK V(7) IGAG IGAG CEK YDFLC	hh s YKFPQ IQQVQ IAQVN LEPDY YKLKR PDRIA KRWPA LGVNW DAGVE DAGVE	mot <u>i</u> ssss VHISQF PDMSFS VQFTVC VVFFSAY LGFQNY FMHYNF FIQF	f-6 YPR(208) FAY(129) FVY(135) TPY(86) MPV(181) IPQ(208) IPY(110) IPY(117) IPT(177) IPV(210) NPF(70)
18409989 7299338 2501533 20094331 15896586 3913815 14520767 11499591 15833984 549552	hhhi I(5)EIAJ F(5)EVII I(7)DVYI A(7)ELAI A(5)DTL YLTDEQII LLNDKLVI LITEEIAI V-LLDDET V-LLDDET V(4)DKL H-TYEDL	hh AV(4)(RD-HP RD-TPF RT(4)(ID(4)I KA(4)(EE(4)(KR(4)(KR(4)(GD		tif-4 sss VPVQS LPAQS LPVQA GAESA VNTEL LGVQD LSIHS LSLDG LSLDG LSLDG ISLHA	hh GSD GSD SN PSQ LDP LD MK PAEI- PND	-hhh -SVL -QVL -RVL -RVL -KVQ -PEK -PEK -ETH -HNQ	hhhh TAMN ERMR TMMK KRIN KLLA QAVN AKML EAFR YRVT DEIV FEKI	REYTA; RGYSRJ RAHTAJ KGIT PNKS(MGIK((GIR(KGGR-) PINKK PINKK	hhh SEFRT EAYLE LEYKS 5) PMG 5) KME 5) VLE SCFDK PTHKL (NIET 5) WET	hhhh-h VVDT LVQH IIRK VIACRE YIKN-E KLRE MAEA TVEG VMRA FLAA VMR4	h LT LR LR LR LS LV LT VR VR	hhhhh ELVPG QFLPN AVRPE KRHG- ENTT(AGFES NAGIE NAVKI L-LQB RYLE(DIMCE	$ \begin{array}{l} 1 & ss \\ \overline{s} V \\ \overline{s} I \\ \overline{s} \\ \overline{s} I \\ \overline{s} \\ \overline{s} \\ \overline{s} V \\ \overline{s} \\ \overline{s} \\$	M SSSSSS IATDI LSSDF ISSDF ILLSF QSTQM INLDL IAPVI VNVSM VNVSM VTIEVVI VZIEVVI VZIEVVI	otif- IC-GF IV-GF IL-GF IIGAT IY-GL MF-GV TVTRY LV-CV MLDHV UV-CL	5 hh PGETDE CGETEE PGETAE PWEDRE PDTD (5 PYQTKE NDDE (6 NYHEVP NRTSAQ NDGT (5 GETEEE	-hhhhi DFSQTY EFQDTY DFEQTY JLTESI SFEKTY)FARRI KVVELC QPLQVY)LAELI -MYPT	hh V(5) VSL I(5) LYSK V(7) IGAG ZEK YDFLC LK LK	hh s YKFPQ IQQVQ IAQVN LEPDY YKLKR PDRIA KRWPA LGVNW DAGVE DTPCK DTPCK	moti ssss VHISQF VNVAYL FDMSFS VQFTVC VFFSAY TYSFAY LGFQNY FMHYNF FIQF INLIPW	f-6 YPR(208) FAY(129) FVY(135) TPY(86) MPV(181) IPQ(208) IPY(110) IPT(177) IPV(210) NPF(70) MPF(130)
18409989 7299338 2501533 2009431 15896586 3913815 14520767 11499591 15833984 549552 15606957 1705463	hhhl I(5)EIA F(5)EVII I(7)DVYI A(7)ELA A(5)DTL: YLTDEQII LITEEIA V-LLDDE V(4)DKL H-TYEDLI MIEKOOA	hh AV(4)(RD-HP RD-TPF RT(4)(ID(4)I KA(4)(C(4)F GDM ED(4)(C(4)F GDM		tif-4 sss VPVQS LPAQS LPVQA GAESA VNTEL LGVQD LSIHS LSLDG LSLDG USLHA VALDA	hh GSD GSD SN PSQ LDP LD MK PAEI PND AT SP	-hhh -SVL -QVL -RVL -KSL -KVQ -PEK -HNQ -EIR -PEL	hhhh TAMN ERMR TMMK KRIN KLLA QAVN AKML EAFR YRVT DEIV FEKL	REYTA: RGYSRI RAHTAJ KGIT PNKS(RVQP(MGIK(GIR(KGGR-1 PINKK RGRP(TPS	hhh SEFRT EAYLE LEYKS 5) PMG 5) KME 5) VLE SCFDK 201 ET 5) WET 5) WET	hhhh-h VVDT LVQH IIRK VIACRK YIKN-K KLRE MAEA TVEG VMRA FLAA YWKV PLAA YWKV	h LT LR LR LR LS LT VR LEWCA LS	hhhhh ELVPG QFLPN AVRPC KRHG- ENTT(AGFES NAGIE NAGFES NAGIE NAVKI L-LQK RYLE(DIMGE	$ \begin{array}{l} S \\ G \\ $	M SSSSSS LATDI LSSDF JSSDF JSSDF JSSDF JSSDF JSSDF JSSDF JSSDF JSSDF JSSSF JSS	otif- IC-GF IC-GF IL-GF IIGAT IY-GL MF-GV TVTRY LV-CV MLDHV IV-GL	5 hh PGETDE CGETEE PWEDRE PUTD (5 PYQTKE NDDE (6 NYHEVP NRTSAQ NDGT (5 GETEEE GETEEE	-hhhhi DFSQTY EFQDTY DFEQTY JLTESI SFEKTY)FARRI KVVELC QPLQVY)LAELI -MVRT	hh V(5) VSL I(5) LYSK V(7) IGAG ZEK ZEK LK LQEVR LK	hh s YKFPQ IQQVQ IAQVN LEPDY YKLKR PDRIA KRWPA LGVNW DAGVE DTPCK DLGGR SHPFS	moti ssss VHISQF YNVAYL FDMSFS VQFTVC VFFSAY TYSFAY LGFQNY FMHYNF FIQF INLIPW THLFSF VPINAL	f-6 YPR(208) FAY(129) FVY(135) TPY(86) MPV(181) IPQ(208) IPY(110) IPT(177) IPV(210) NPF(70) WPE(130) WPE(130)
18409989 7299338 2501533 2009431 15896586 3913815 14520767 11499591 15833984 549552 15606957 1705463	hhhl I(5)EIAJ F(5)EVII I(7)DVYI A(7)ELAI VLTDEQII LLIDEU VLTDEQII V(4)DKL(H-TYEDLI MIEKQQAI I(1)VI	hh AV(4)(RD-HP RD-TPF RT(4)(ID(4)I KA(4)(EE(4)(KR(4)(GDN ED(4)(LE(4)(PB(4)(SSSS CVITFLH CVCKQLH CLVSFLH CCRTVIV CCRTVIV CCRTM-S SFNRI-S SFNRI-S SLORV-N SUGIV CIDVALA ARETL-T SLTAY-N SLWSTPG	tif-4 sss VPVQS LPAQS LPVQA GAESA VNIEL LSINS ISLDG LSINS ISLDG LSLDG LSLDG HNLDA HNLDA	hh GSD GNT SN PSQ LDP LD MK PAEI- PND AT SR	-hhh -SVL -QVL -ETL -KSL -KVQ -PEK -ETH -HNQ -EIR -PEL -EYY	hhhh TAMN ERMR TMMK KRIN KLLA QAVN AKML EAFR YRVT DEIV FEKL PNVIS	REYTA: RGYSRI RAHTAJ KGIT PNKS(RVQP(MGIK() GIR(KGGR-1 PINKK RGRP() TTRS PGCC()	hhh SEFRT EAYLE LEYKS 5)PMG 5)VLE 5)VLE GCFDK 200 200 200 200 200 200 200 200 200 20	hhhh-h VVDT VQH IIRK VIACRK YIKN-K KLRE MAEA TVEG VMRA FLAA YWKV RLET VWFT	h IR LR IS IR IR LV LR	hhhhh ELVPG QFLPN AVRPD KRHG- ENTT (AGFES NAGID NAVKI L-LQK RYLE (DIMGD H-VR(P-LC-	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	IN ISSSS IATDI ISSDF SSTOM TNLDL IAPVI VNVSM VDYNV TIEYV VCSGG VCSGG	otif- IC-GF IV-GF IL-GF IIGAT IY-GL MF-GV TVTRY LV-CV MLDHV IV-GL II-GL	5 hh PGETDE CGETEE PWEDRE PWEDRE PDTD (5 PYQTKE NDDE (6 NYHEVP NRTSAQ NDGT (5 GETEEE GEAE (5 VFS1 (5	-hhhhi DFSQTY EFQDTY DFEQTY LTESI SFEKTY)FARRI (VVELC QPLQV) LAELI -MVRTI)LLHTI)LLHTI	hh V(5) VSL I(5) IVSK V(7) IGAG IGAG ZEK ZEK LGEK LGEK LATLP: VPFIO	hh s YKFPQ IQQVQ IAQVN LEPDY YKLKR PDRIA KRWPA LGVNW DAGVE DTPCK DLGGR SHPES	moti ssss WHISQF YNVAYL FDMSFSS VQFTVC VFFSAY LGFQNY FMHYNF FIQF INLIPW THLFSF VPINAL	f-6 YPR(208) FAY(129) FVY(135) TPY(86) IPQ(208) IPY(110) IPY(110) IPY(210) NPF(70) WPE(130) LAV(107) TPW(109)
18409989 7299338 2501533 20094331 15896586 3913815 14520767 11499591 15833984 549552 15606957 1705463 15606163 15606163	hhhi (5)EIA F(5)EVII I(7)DVYI A(7)ELAI VLTDEQII LLNDKLVI LITEEIAI V-LLDDET V(4)DKLC H-TYEDLI MIEKQQAI I(13)VII V	hh AV(4)(RD-HP RT(4)(ID(4)I KA(4)(EE(4)(KR(4)(GD ED(4)(LE(4)(RR(4)(PO(4))	 SSSS CVITFLH CVCKQLH CLVSFLH CCRTVIV CCRM-S SFRRI-S SLDRV-N SLARV-N SUSY-Q SNHFLVG MIDVALA SAETL-T JLTAY-N SLMSIPG CLCM	tif-4 sss VPVQS LPAQS LPVQA GAESA VNIEL LSINS ISLDG LSINS ISLDG LSLDG LSLDG CSL CSLDG LSLDG LSLDG CSL LSLDG	hh GSD GNT SN PSQ LDP LD MK PAEI- PND SR LSQ LSQ	-hhh -SVL -QVL -RVL -ETL -KSL -KVQ -PEK -PEK -HNQ -EIR -PEL -EYY -EYR	hhhh TAMN ERMR TMMK KRIN KLLA QAVN AKML EAFR YRVT DEIV FEKL PNVI DKIS LNTH	REYTA; RGYSRI RGIT PNKS(MGIK() GIR KGGR PINKK RGRP() TTRS PGKC() FNHP()	hhh SEFRT EAYLE LEYKS 5) PMG 5) VLE 5) VLE GCFDK 20 VLE 5) VLE 5) VLE 10 VE 10	hhhh-h VVDT VQH IIRK VIACRK YIKN-K KLRE MAEA TVEG VMRA FLAA YWKV RLET VMRT ACCH	h IR UR UR IS IR IR LR VR LEWCA LS AH UR	hhhhh ELVPG QFLPN AVRPD KRHG- ENTT (AGFES NAGID NAVKI L-LQK RYLE (DIMGD H-VR (R-LG- DAGVE	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	IN ISSSS IATDI ISSDF SSTOM TNLDL ISSDF TNLDL INLDL INV VNVSM VNVSM VNVSM VNVSM TEXV VCSGG TTATM	otif- IC-GF IV-GF IL-GF II-GF IY-GL VT-GL VT-GV NLDHV IV-CV NLDHV IV-GL II-GL MF-GH	5 hh PGETDE CGETEE PGETAE PWEDRE PDTD (5 PYQTKE NDDE (6) NUTEVP NRTSAQ NDGT (5 GETEEE GEAE (5 VESI (5 NDSVPV	-hhhhi DFSQTY DFEQTY DFEQTY JLTESI SFEKTY)FARRI (VVELC QPLQV) LAELI -MVRTI)LLEI)LLEI)HLER	hh V(5) VSL I(5) I(5) IQSK V(7) IGAG IQSK ZEK LGEK LQEK LQEK LQEK LQEK LQEK LQEK LQEK LQEK	hh s: YKFPQ IQQVQ IAQVN LEPDY YKLKR PDRIA KRWPA LGVNW DAGVE DIGGR SHPES DETGG HDIJM	moti ssss VHISQF PDMSFS VQFTVC VFFSAY UFFSAY GMHYNF F1QF INLIPW THLFSF VPINAL FTAF	f-6 YPR(208) FAY(129) FVY(135) TPY(86) IPQ(208) IPY(110) IPY(110) IPY(170) NPF(70) NPF(70) NPF(130) LAV(107) IPW(109) IPW(132)
18409989 7299338 2501533 2009431 15896586 3913815 14520767 11499591 15833984 549552 15606957 1705463 15606163 19705171 8313168	hhhi (5)EVI (5)EVI (7)DVY A(7)ELA (5)DTL VLTDEQI LLNDKLV LITEEIA V-LLDDE V(4)DKL H-TYEDL H-TYEDL MIEKQQAI (13)VI W-VVI	hh AV(4)(RD-HP RD-TPF RT(4)(ID(4)I KA(4)(EE(4)(WC(4)F GDN ED(4)(LE(4)(RR(4)(PQ(4)F GQ(4)F GQ(4)F	 SSSS VYTFLH VCKQLH ELVSFLH SCRTVYV SPNRI-S SLDRV-N SVGYV-Q NHFLVG (IDVALA AGETL-T SLTAY-N SLMSIPG ELCNM	tif-4 sss VPVQs LPAQS LPVQA GAESA VNIEL LSUG LSIDG LSLDG LSLDG LSLDG LSLDA TSLA VALDA HNLDT GGAEI LKK	hh GSD GSD SSN PSQ LDP LDP MK PAEI PND AT SR LSQ YH GN	-hhh -SVL -QVL -ETL -ETL -KSL -PEK -ETH -HNQ -EIR -PEL -EYY -EYR -PIW -DAH	hhhh TAMN ERMR TMMK KRIN KLLA QAVN AKML EAFR YRVT DEIV FEKL PNVI DKIS LNTH ALRY	REYTA; RGYSRI RAHTAI KGIT PNKS(MGIK() GIR KGGR PINKK RGRP() TTRS PGKC() FNHP() SVAP	hhh SEFRT EYKS O O O O O O O O	hhhh VVDT LVQH IIRK VIACRE YIKN KLRE MAEA TVEG VMRA FLAA VMRA FLAA VMRA VMRA VHRT ACEM TTRNFS	h IT IR IS IS IS IS IS IS LT VR LE WCA LS AH LA	hhhhh ELVPG QFLPN AVRPG ENTT(AGFES NAGID NAVKI L-LQB RYLE(DIMGD H-VR(R-LG- DAGVB	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	IN ISSSESS IATDI ISSDF ISSDF ILLSF QSTQM INLDL IAPVI VNVSM VVVSM VVVSM VVSGG TTATM NQVL	otif- IC-GF IL-GF IL-GF IIGAT IY-GL MF-GV TVTRY LV-CV MLDHV IV-GL II-GL MF-GH LR-GH	5 hh PGETDE CGETEE PGETAE PWEDRE PWDT0(5 PYQTKE NDDE(6 NYHEVP NRTSAQ NDGT(5 GETEEE GEAE(5 VESI(5 NDSVPV VECCTK	-hhhhh DFSQTY EFQDTY DFEQTY JLTESI SFEKTY JFARRI KVVELC QPLQVY JLAELI -MVRT JLLHTT JLLHTT JHLERY -MKRLY PVLRW	hh V(5) VSL VIL L(5) L(5) L(5) L(5) V(7) IGAG ZEK V(7) L(7) VCR L(7) L(7) VCR L(7) VCR L(7) L(7) L(7) L(7)	hh s YKFPQ IQQVQ IAQVN LEPDY YKLKR PDRIA KRWPA LGVNW DAGVE DTPCK DLGGR SHPES DETGG HDLVM AENCN	moti sss VHISQF YDASFS VQFTVC VFFSAY TYSFAY LGFQNY FMHYNF F-IQF INLIPW THLFSF VPINAL F-TAF MRVRPY HALVN	f-6 YPR(208) FAY(129) FVY(135) IPV(86) MPV(181) IPQ(208) IPY(110) IPY(110) IPY(210) NPF(70) MPE(130) LAV(107) IPW(109) YIY(132) MDQ(44)
18409989 7299338 2501533 2009431 15896586 3913815 14520767 11499591 15833984 549552 15606957 1705463 15606163 19705171 18313168 730103	hhhi (5)Ella F(5)EVI (7)DVYI A(7)ELA A(5)DTL LITDEQI LINDKLVI LITEELA V-LLDDE V(4)DKL H-TYEDL MIEKQQA I(13)VII VVLI M-VLTPE LIPUYA	hh AV(4)(RD-HP RD-TPF RT(4)(ID(4)I KA(4)(EE(4)(GD ED(4)(CRR(4)(CRR(4)(FQ(4)F GL(4)F GL(4)F GL(4)F	 SSSS VYTFLH VCKQLH LVSFLH CCTWYV CCTWSFUN SCTWS SUDRV-N SVGYV-Q NHFLVG MIDVALA SAETL-T JLTAY-N SLMSIPG ELCNM VIDIWL VLKWT	tif-4 sss VPVQs LPAQS GAESA VNIEL LGVQD LSIG USLG USLG USLG USLA VALDA HNLDT GGAEI LKK PDFKY	hh GSD GSD SN PSQ LDP LDP MK PAEI PND SR SR LSQ YH GN	-hhh SVL QVL ETL -KSL -KVQ -PEK -ETH HNQ -EIR -PEL -EYY -EVR -PIW -DAH	hhhh TAMN ERMR TMMK KRIN KLLA QAVN AKML EAFR YRVT DEIV FEKL PNVI DKIS LNTH ALRY FKKI	REYTA: RGYSR RAHTAJ KGIT PNKS (RVQP (MGIK (GIR PINKK RGRP (TTRS PGKC (FNHP (SVAP-) NGRG (hhh SEFRTE LEYKS O DMG O DM	hhhh-h VVDT LVQH IIRK VIACRE YIKN-E KLRE MABA TVEG VMRA FLAA FLAA FLAA FLAA FLAA FLAA TVRT ACEM TRNFS TRNFS TEAA	LT IR VA IS IR LV IR LT VR LA AH LA VI	hhhhh ELVPG QFLPN AVRPG- ENTT(AGFES NAGII NAVKI L-LQK RYLE(DIMGI H-VR(R-LG- DAGVF CKRRE	1 55 GK (1K (1K (1K (1K (K (K (K (K (K (K (K (K (K (K (K (K (K (K) (K (K (K) (K) (K (K) (K) (K (K)	TTATM	otif- IC-GF IV-GF IL-GF IIGAT IY-GL MF-GV TVTRY LV-CV IV-CU IU-GL II-GL MF-GH LR-GI VLPGH OK-GV	5 hh PGETDE CGETEE PGETAE PWDTD (5 PYOTKE NDDE (6 NYHEVP NRTSAQ NDGT (5 GETEEE GEAE (5 VESI (5 NDSVPV VECTKE VESU (5 NDSVPV	-hhhhi DFSQTY EFQDTY DFEQTY JLTESI SFEKTY JFARRI KVVELC QPLQYJ JLAELI -MVRTJ JLLHTI JLLHTI JHLERY -MKRLY PVLRWI -TLPM	hh V(5) VSL VSL L(5) L(5) LVSK V(7) IGAG ZEK VTLC LVSL IQRVR LATLP VRLQ VRLQ V LATLP VRLQ V	hh s IQQVQ IQQVQ IAQVN LEPDY YKLKR PDRIA KRWPA LGVNW DAGVE DAGVE DAGVE DIPCK SHPES DETGG HDLVM AENCN EKGHT	moti ssss VHISQF FDMSFS VQFTVC VFFSAY LGFQNY FMHYNF F-IQF INLIPW THLFSF VPINAL F-TAF MRVRPY HALVNI LAFLEN	f-6 YPR(208) FAY(129) FVY(135) IPY(86) MPV(181) IPQ(208) IPY(110) IPY(177) IPY(210) NPF(70) WPE(130) IPW(107) IPW(109) YIY(132) MDQ(44)
18409989 7299338 2501533 2009431 15896586 3913815 14520767 1149957 15833984 549552 15606957 1705463 19705171 18313168 730103 22027536	hhhi (5)Elar F(5)EVI (7)DVYI A(7)ELA A(7)ELA LTDEQII LLTDEQII LLTDELA V-LLDDE V(4)DKLC H-TYEDLI MIEKQQAI S(1)VII VVLI M-YLTPE L-LPVYAI S(1)POL	hh AV(4)(RD-HPI RD-TPI RT(4)(ID(4)I (KR(4)(CEE(4)(KR(4)(CEE(4)(CEE(4)(CEE(4)(CRR(4)(CEE(4)(CRR(4)(CEE(4)(CRR(4)(CEE(4)(CRR(4)(CEE(4)(CRR(4)(CEE(4)(CRR(4)(CEE(4)(CRR(4)(CEE(4)(CRR(4)(CEE(4)(C		tif-4 sss VPVQS LPAQS LPVQA GAESA VNIEL LGVQD LSIDG LSLDG LSLDG LSLDG VALDA HHLDT GGAEI LKK PDFKY ISLDS	hh GSD GSD SN PSQ LDP LDP PND PND SR SR LSQ YH GN LE LE	-hhh -SVL -QVL -RVL -ETL -KSL -FEK -PEK -PEK -PEK -PEK -PEK -PIW -DAH -DER -FF	hhhh TAMN ERMR TMMK KRIN AKML EAFR YRVT DEIV FEKL DKIS LNTH ALRY IVRR	REYTA: RGYSRJ RAHTAJ KGIT PNKS((MGIK() GIR(KGGR-) PINKK RGRP() TTRS PGKC(FNHP() SVAP-1 NGRG()	hhh SEFRT EAYLE EYKS -VQD 5) PMG 5) KME 5) VE SCFDK VIE VIE VIE VIE 1 VE 1 VE 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1	hhhh	h IT IR VA IS IS IL VR LEWCA LS AH LAH VI KQ TE	hhhhh ELVPG QFLPN AVRPE KRHG- ENTT(AGFES NAGIE NAVKI L-LQS RYLE(DIMGE H-VR(R-LG- DAGVF CKRRE AGLG- LGYNF	$ \begin{array}{l} S \\ G \\ $	TTATM NOTOTO	otif- IC-GF IV-GF IL-GF IIGAT IY-GL IY-GV VTVTY UV-CV IV-CV IV-GL II-GL MF-GH LR-GI VLPGH QK-GV	5 hh PGETDE CGETEE PGETAE PWEDRE PUDTO (5 PYQTKE NDDE (6 NYHEVP NRTSAQ NDGT (5 GETEEE GEAE (5 VDSV (5 NDSVPV VESI (5 NDSVPV VESCTK NEKEL NEDELL	-hhhhh DFSQTY EFQDTY DFEQTY DVFRTJ)LTESI SFEKTY)FARRI KVVELC QPLQVY)LABLI -MVRTJ)LLHTI)HLERY -MKRLY PVLRWI -ILPMZ DFAALY	hh V(5) VSL I(5)-: LYSK V(7) IGAG-1 CEK ZEK LK IQRVRI LATLP: VRKIQ VRKIQ VRKIQ V L ARYFKI TEGL-	hh s IQQVQ IQQVQ IQQVN LEPDY YKLKR PDRIA KRWPA LGVNW DAGVE DITPCK SHPES DETGG HDLVM AENCN EKGHI PLDVR	moti ssss VHISQF YNVAYL FDMSFS VQFTVC VFFSAY ISSFAY CGFQNY FMHYNF F-IQF INLIPW INLIPW IHLFSF VPINAL F-TAF MRVRPY HALVNI LRFIEF F-IEY	f-6 YPR(208) FAY(129) FVY(135) IPY(86) MPV(181) IPY(110) IPY(110) IPY(100) NPF(70) WPE(130) IPW(109) YIY(132) MDQ(44) MDY(133)
18409989 7299338 2501533 2009431 15896586 3913815 14520767 11499591 15833984 549552 15606957 1705463 15606163 15705171 18313168 730103 22027536	hhhi (5)EIA F(5)EVII I(7)DVYI A(7)ELA A(5)DTL: YLTDEQII LITEEIA V-LLDDE V(4)DKL(H-TYEDLI MIEKQQA I(13)VII VVLI M-YLTPE(L-LPVYA I(6)PQL(E(6)PQL(I(4)PAT)	hh AV(4)(RD-HPI RD-TPI RT(4)(ID(4)I KR(4)(EE(4)(KR(4)(GD	 SSSS VYTFLH VCKQLH LVSFLH CRTVYV CDRM-S SHRI-S SHFLVG NHFLVG NHFLVG NHFLVG NHFLVG LDCAM VIDIWL LCCM VIDIWL SLRV-T SLSAL-N LDEIR	tif-4 sss VPVQS LPAQS LPVQA GAESA VNIEL LSUPS LSUG LSLOG LSLOG LSLOG LSLOG CSLOG VALDA HNLDT GGAEI LKK PDFKY ISLDS ISLDS TSLDS	hh GSD GSD SN PSQ LDP LDP MK PAEI- PND AT SR YH GN LSQ YH CN LSQ YH CN	-hhh -SVL -QVL -ETL -KSL -KVQ -PEK -PEL -EYY -EVR -PIW -DAH -DER -FF	hhhh TAMN ERMR TMMK KRIN AKILA QAVN AKKL EAFR YRVT DEIV FEKL PNVI DKIS LNTH ALRY IVRR	REYTA: RGYSRI RAHTAI KGIT PNKS(GIR(KGGR-) PINKK RGRP(FNHP(SVAP-) NGRG(KGFH-) ENMK(KGFH-)	hhh SEFRT EAYLE LEYKS 5) PMG 5) PMG 5) KME 5) VLE GCF DK CNIET 5) WET 5) WET 5) WET 5) AKK CWEV 5) AKK CWEV 5) AKK	hhhh	h LT IR VA IS IL LT VR LS LS LA SVI KQ IE AV	bhhbh ELVPG QFLPN AVRPC KRHG- ENTT (AGFES NAGIC NAGIC NAVKI L-LQM RYLE (DIMGC H-VR (DAGVF CKRRE AGLG- LGYNF PGFF ($ \begin{array}{cccccccccccccccccccccccccccccccccccc$	TTATM TTATM TASSE TATDI SSSSF SST SST SST SST SST SST SST SST S	otif- IC-GF IL-GF IL-GF IL-GF IL-GF IL-GF IL-GL VTVTRY IV-CV MLDHV IV-CV MLDHV IV-GL II-GL VLPGH QK-GV MR-GL KFLNT	5 hh PGETDE CGETEE PGETAE PWEDRE PDTD (5 PYQTKE NDDE (6 NYHEVP NRTSAQ GETEEE GEAE (5 VESI (5 NDSVPV VECCTK NESPLI NEDELL	-hhhhi DFSQTY EFQDTY DFEQTY DFECTY)ITESI SFEKTY)FARRI KVVELC QPLQY)LAELI -MVRTI)HLERY -MKRLY PVLRWI -ILPMZ DFAALI	hh V(5) VSL VSL VSL V(5) I(5) I(3)- CEK ZEK ZEK ZEK IQRVRL LATLP VRKIQ V L ARYFKI TEGL-1 L(13)	hh s YKFPQ IQQVQ IAQVN LEPDY YKLKR PDRIA KRWPA LGVNW DAGVE DTPCK DLGGR SHPES DETGG HDLVM AENCN EKGHI PLDVR IKGSI	moti ssss FDMSFS VVQFTvC VFFSAY LGFQNY FMHYNF FIQF INLIPW FHLFSF VPINAL FTAF MRVRPY HALVNI LRFIEF FIEY EMGLKV	f-6 YPR(208) FAY(129) FVY(135) TPY(86) IPY(110) IPY(110) IPY(110) IPY(10) NPF(70) WPE(130) LAV(107) IPW(109) YIY(132) MDQ(44) MDV(137) MPF(133)
18409989 7299338 2501533 20094331 15896586 3913815 14520767 11499591 15833984 549552 15606163 19705171 18313168 730103 22027536 14521276	hhhi (5)EVI (5)EVI (7)DVY (7)ELAI (7)ELAI (5)DTL LINDKLV LITEEIAI V-LLDDET V(4)DKLC H-TYEDLI MIEKQQAI (13)VII VVLI MIEKQQAI (13)VII U-LPVYAI (16)PQL (16	hh AV(4)(RD-HP RD-TPF RT(4)(CEE(4)(KR(4)(GDh ED(4)(GDh ED(4)(GDh ED(4)(GDh ED(4)(GDh ED(4)(CA)(CA)(CA)(CA)(CA)(CA)(CA)(CA		tif-4 sss VPVQS LPAQS LPAQS GAESA VNIEL LSLG LSLG ISLHA VALDA HNLDT GGAEI -LKK PDFKY ISLDS ISLDS ISLDT FHPL	hh GSD GSD SN PSQ LDP LDP MK PAEI- PND AT SR LSQ LSQ LSQ LSQ LSQ LSQ LSQ LSQ LSQ LSQ SR	-hhh SVL QVL RVL ETL KSL PEK ETH HNQ EER PEV EVR PEV EVR DAH DER SKFF RVY	hhhh TAMN ERMR TMMK KRIN QAVN AKML EAFR YRVT DEIV FEKL PNVI DKIS LNTH ALRY IVRR EREI IVRR	REYTA; RGYSRI RAHTAJ KGIT PNKS(() KGGR-) PINKC() RGGR-() FINKP() SVAP-1 NGRG() KGFH-1 ENMK() PGAN-'	hhh SEFRT EAYLE EYKS -VQD 5) WG 5) VLE SCFDK 2 CFDK C CFDK C CF	hhhh-h VVDT LVQH IIRK VIACRK KLRE MAEA TVEG VMRA FLAA FLAA YWKV RLET VHRT ACEM IEAA IFRNFS IEAA IHKA GGEV-F KLLE-F	LL LT LR VA IS LV LT LT LS LS LA	bhhbh ELVPG QFLPN AVRPD KRHG- ENTT (AGFES NAGID NAVKI L-LQM RYLE (DIMGD H-VR (R-LG- R-LG- CKRRE AGLG- LGYNF PGFE (EAHPC	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	TATM SSSSS TATDI LSSDF SSSF SSSF TILLSF SST SST UVDYN TILSV UVDYN TISV UVGAHL UVCSGG TTATM SNQTVL IVRHL IVRHL IVRHL CINWV UVRV MQTVL IVRHL	otif- IC-GF IL-GF IL-GF IL-GF IL-GF IL-GF V-GL MF-GV VTVTRY LV-CV MLDHV IV-GL II-GL MF-GH LR-GI VLPGH QK-GV MR-GL KFLNI MV-GL	5 hh PGETDE CGETEE PDTD (5 PYQTKE NDDE (6 NJDE (6 NJDE (6 NJDE (5 GETEE5 VESI (5 VESI (5 VESI (5 VESI (5 VESI (5 NDSVPV VECCTK NEKD NELE (5 GETN (5	-hhhhi DFSQTY EFQDTY DFECTH DVFRTJ)LTESI SFEKTY)FARRI KVVEL(QPLQVS)LAELI -MVRTJ)LLHTT)HLERY -MKRLY PVLRWI -ILPMA DFAALI)LRALI)VMHDU	hh V(5) VSL VSL U(5) LYSK V(7) IGAG ZEK ZEK LGTLP VDFLC LATLP VRKIQ VRKIQ VRKIQ VRKIQ VRKIQ LATLP LATLP LATLP LATLP LATLP LATLP LATLP LATLP LATLP LATLP LATLP VRKIQ VRKIQ LATLP LATLP VRKIQ VRKIQ LATLP VRKIQ VRKIQ VRKIQ LATLP VRKIQ VRX VRKIQ VRX VRX VRX VRX VRX VRX VRX VRX VRX VRX	hh s YKFPQ IQQVQ IAQVN LEPDY YKLKR PDRIA KRWPA LGVNW DAGVE DIPCK SHPES DETGG HDLVM AENCN AENCN EKGHI FLDVR IKGSL	moti ssss VHISQF PDMSFS VQFTVC VFFSAX TYSFAY LGFQNY FMHYNF FIQF INLIPW THLFSF VPINAL FTAF MRVRPY HALVNI LRFIEF FIEY EMGLKV ZTLGOV	f-6 YPR(208) FAY(129) FVY(135) IPY(86) MPV(181) IPQ(208) IPY(110) IPT(177) IPY(210) NPF(70) MPE(130) LAV(107) IPW(109) YIY(132) MDQ(44) MDV(137) MPF(133) LEW(122)
18409989 7299338 2501533 2009431 15896586 3913815 14520767 11499591 15833984 549552 15606957 1705463 15606163 19705171 18313168 730103 22027536 14521276 22001767 17865692	hhhi (5)EVI (5)EVI (7)DVY A(7)ELA (5)DTL (1)DKLV LITEELA ULNDKLVI LITEELA U-LLDDE V(4)DKL H-TYEDL MIEKQQA (1)3)VI VVL MIEKQQA (1)3)VI UVL (1)(6)PQL (1)(6)PQL (1)(6)PQL (1)(6)PQL (1)(6)PQL (1)(6)DRA MLNESOA	hh AV(4)(RD-HPM RD-TPF RT(4)(EE(4)(EE(4)(KR(4)(GDM ED(4)(LE(4)(GDM ED(4)(LE(4)(CRR(4)(CRR(4)(CQKA(EK(4)(CQKA(EK(4))(CQK-A(CQK-A(CQK))) EX(4)(CQK-A(CQK)) EX(4)(CQK-A(CQK)) EX(4)(CQK) EX(4)(C		tif-4 sss VPVQS LPAQS LPVQA GAESA VNIEL LGVQD ISLDG LSING LSING LSING CSI SLDG ISLDG ISLDG ISLDG ISLDG ISLDG ISLDG ISLDG ISLDT FHPDL HNLEN HNLEN	hh GSD GNT GSD SN PSQ LDP LDP LDP PAEI- PAEI- PAEI- SR SR LSQ YH GN LSQ YH LSQ YH LSQ YH SR SR SR SR SR SR SR SR SR SR SR SR SR SR SR SR SR SR SR	-hhh SVL QVL RVL ETL KSL PEK ETH HNQ EER PEY EVR PEY EVR DAH DER SFFF F RVY EFY	hhhh TAMN ERMR TMMK KRIN KLLA QAVN AKKI VRVT PNVI PNVI DKIS LNTH ALRY FKKI IVRR EREI IVRR EREI I	REYTA; RGYSRI RAHTAI KGIT PNKS((MGIK((GIR(KGGR-1) PINKK(FNHP() FNHP() SVAP1 NGRG() KGFH-1 ENMK() PGAN-1 TTR	hhh SEFRT EAYLE EAYLE EAYLE EAYLE EAYLE EAYLE SOUTH SO	hhhh	h IT IR IR IS	hhhhh ELVPG QFLPN AVRPD KRHG- ENTT(AGFES NAGID NAVKI L-LQK RYLE(DIMGD H-VR(R-LG- DAGVF CKRRE CKRRE CKRE AGLG- LGYNF PGFF(EAHPD EAHPD	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	TATAL ISSSES IATAII ISSDF ISSDF ILLSF GOSTOM INDL IAPVI VVNSGG TTATM NQTVL IVGAHL IVCSGG TTATM NQTVL IVRL IVRL IVRL IVRL IVRL IVRL IVRL INVVV V	otif_ IC-GF IC-GF IL-GF IL-GF II-GT IY-GL V-CV MC-CV NLDU-CV IV-GL II-GL MF-GH IR-GH QK-GV MR-GL KFLNI MV-GL IV-GL	5 hh PGETDE CGETEE PDTD (5 PYQTKE NDDE (6 NDDE (6 NDGT (5 GETEEE GETEEE (6 GETE (5 GETN (5 GETN (5 GETV (5)	-hhhhi DFSQTV EFQDTV DFEQTV JLTESI SFEKTV JFARRI KVVELC QPLQVV JLAELI -MVRTJ JLAELI JLAELI HLERV -MKRLV PVLRWI -ILPMU DFAALI JLALI JLALI JLLIO	hh V(5) VSL VSL LYSK V(7) IGAG ZEK ZEK IQRVRI LATLP VRKIQ VRKIQ V LATLP VRKIQ V LG13 L LANLP LANLP	hh s YKFPQ IQQVQ IAQVN LEPDY YKLKR PDRIA KRWPA LGVNW DAGVE DTPCK DTPCK SHPES DETGG HDLVM AENCN EKGHI PLDVR IKGSL IKGSL HGVTM	moti ssss yHISQF yNVAYL POMSFS yQFTVC YFFSAY IYSFAY LGFQNY FMHYNF FIQF INLIPW THLFSF YILF HALVNI LRFIEF FIEY EMGLKV LTLGQY YPINML	f-6 YPR(208) FAY(129) FVY(135) IPY(86) MPV(181) IPQ(208) IPY(110) IPY(110) IPY(107) IPV(210) MPF(130) IPW(109) YIY(132) MDQ(44) MDV(137) MDY(133) LEW(122) LQP(53) VKV(118)
18409989 7299338 2501533 2009431 15896586 3913815 14520767 17499591 15833984 549552 15606957 1705463 15606163 19705171 18313168 730103 22027536 14521276 22001767 178656922	hhhi (5)EVII (7)DVY	hh AV(4)(RD-HPP RD-TPP RD-TPP ID(4)I EE(4)(KR(4)(GDM ED(4)(C(4)) GDM ED(4)(C(4)) GDM ED(4)(C(4)) GDM ED(4)(C(4)) C(4)) C(4) C(4)) C(4) C(4)) C(4) C(4		tif-4 sss VPVQS LPVQA GAESA VNIEL LGVQD LSIG LSLGC LSLDG ISLDG ISLDG ISLDG ISLDT GGAEI SLDS ISLDT SLDS ISLDT SLDS ISLDT HNLEN HNLDT ISLDT		-hhh SVL QVL ETL KSL ETL KSL ETH HNQ EIR EVR DAH CEYY CAH COR FEFY RFY EFY RFY	hhhh TAMN ERMR TMMK KRIN KLLA QAVN AKML EAFR YRVT DEIV PNVI DKIS LNTH ALRY FKKI IVRR EREI RQVR GNII	REYTA RGYSR RAHTAJ KGIT PNKS(RVQP(GIR PINKK GGR PINKK RGRP(FNHP(SVAP NGRG(KGFH ENMK(PGAN TTRT SVPN-	hhh BEFRT EAYLE EAYLE EAYLE EAYLE EAYLE EAYLE IN COPDK IN IN COPDK IN IN IN IN IN IN IN IN IN IN IN IN IN	hhhh-h VVDT LVQH IIRK VIACRK YIKN-K KLRE MAEA TVEG VMRA FLAA VMRA VMRA TIRNFS IEAA IIEAA IIEAA GGEV-F KLLE-F RLDT R	h LT IR IS IS IS LV IR LT LT VR LT VR LS VR LS VR LS VR LS VR VR LS NR VR LS NR	hhhhh ELVPCQFLPN AVRPD ENTT(NAVRD NAGFES NAGII NAVKT L-LQF RYLE (NAGFC H-VR (K-RC LGYNF PGFE (EAHPC EAHPC K-VR (KEFGI	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	TATDI JASSS JATDI JSSSF JISSF JISSF JILSF JSSTQM JILSF JSSTQM JILSF JSST JILSF JSST JILSF JSST JILSF JSST JSST JSST JSST JSST JSST JSST JS	otif- IC-GF IV-GF IIL-GF IIL-GF MF-GV WTVTV VLV-CV MLDHV-GL II-GL MF-GH II-GL MF-GH II-GL MF-GH KFLNI WLCGL KFLNI VLCGL	5 hh PGETDE CGETEE PDETDE PDTD (5 PYQTKE NDDE (6 NYHEVP NRTSAQ NDGT (5 GETEEE GEAE (5 VDSVPV VESI (5 NDSVPV VESI (5 NDSVPV VESI (5 SGETV (5 GETV (5 NDKE (5)	-hhhhi DFSQTV EFQDTV DFEQTI)LTESI SFEKTV)FARRI KVVELC QPLQVD)LABLI -MVRTI)LABLI -MVRTI)LABLI -MKRLV PVLRWI -ILPM2 DFAALI)LRALI)LRALI)LRALI)JFARW	hh	hh s YKFPQ IQQVQ IAQVN LEPDY YKLKR PDRIA KRWPA LGVNW DAGVE DTPCK DLGGR SHPES DTPCK DLGGR HDLVM AENCNI EKGHI PLDVR IKGSL IKGSL	moti ssss FDMSFS FDMSFS VWQFTVC WFFSAY IYSFAY LGFQNY FFMIYNF F-IQF INLIPW INLIPW F-TAF WRVRPY HALVNI LRFIEF F-IEYF EMGLKV LFLQYY VHFSRF	f-6 YPR(208) FAY(129) FVY(135) IPY(86) MPV(181) IPQ(208) IPY(110) IPT(177) IPY(210) NPF(70) WPF(130) IPW(109) YIY(132) MDQ(44) MDV(437) MDV(137) MDY(133) LEW(122) LQV(53) VPH(90) FPH(90)
18409989 7299338 2501533 2009431 15896586 3913815 14520767 11499591 15833984 549552 15606957 1705463 19705171 18313168 730103 22027536 14521276 22001767 17865692 14520966	hhhi (5)ElAI F(5)EVII (7)DVYI A(7)ELAI A(5)DTL: YLTDEQII LITEEIAI V-LLDDEI V(4)DKLC H-TYEDLI MIEKQQAI S(1)VII VVLI (1)VII M-YLTPEC L-LPVYAI S(6)PQLC (1)CDRAI MINESQAC (1)CDRAI (1)CCRAI (1)CC	hh AV(4)(RD-HP RD-HP RD-TP RD-TP RD(4)(ID(4)(ID(4)(CE(4)(GD(tif-4 sss VPVQS LPVQS CAESA VNIEL LGVQD LSIDG LSLDG LSLDG LSLDG LSLDG LSLDG TS		hhhh SVL QVL ETL KSL KVQ PEK ETH HNQ EEH HNQ EEY PEL EY PEL EY CAR FEF FRY RY RRL	hhhh ERMR TMMK KRIN QAVN AKML EAFR YRVT DEIV FEKL IVRR EREI IVRR EREI IVRR EREI MKIA	REYTA; RGYSR; RAHTAJ KGIT PNKS() RUQP() RGIR PINKK; RGRP() TTRS PINKK; RGRP() SVAP SVAP SVAP SVAP SVPN EPRA	hhh SEFRT AVLE -VQD S) MG S) VLE S) VLE CCFDK CCFD	hhhh-h VVDT LVQH IIRK VIACRK YIKN-K KLRE MAEA TVMRA FLAA YWKV RLEA I TRNFS IEAA IHKA RCEM RLEA RLEA	h IR VA IR IS VA IR V- VR LEWC2 AH VI VI VI VI VI VI V VA V V	hhhhh ELVPG QFLPN AVRPC ENTT AGFES NAGIL ENTT AGFES NAGIL L-LQU RYLE (DIMG H-VR (CKRRF AGLG CKRRF AGLG LGYNF PGFE (EAHPD K-VR (KEFGI) O-LS (1 SS 	TATDI LSSSS IATDI LSSDF ILSSF ILSSF ILLSF SGSTQM INLDL ILSF VNVSM VVVSM TILLY VTIEVVI VGSGG IVRHL IVRL INNVV VVCVV VVCVV VNCVV ILVRL VKSGL VKSGG ZELTYL	otif- IC-GF IIC-GF IIC-GF IIC-GF IIC-GF IIC-GF IIC-GT VTVTRY VTVTRY IV-GL IIC-GI VC-GL VC-GL VC-GL VC-GL VC-GL VC-GL VC-GL VC-GL VC-GL VC-GL VC-GL VC-GL VC-GF VC-	5 hh PGETDE CGETEE PDETDE PUEDRE PUEDRE PUTD (5 PYQTKE NDDE (6 NYHEVA NDGT (5 GETEEE GEAE (5 VDESI (5 NDSVPV VESI (5 MDSVPV VESI (5 GETN (5 GETN (5 GETN (5 MDKE (5) NDKE (5) NDKE (5)	-hhhhi DFSQTV EFQDTV DFEQTN DVFRTJ)LTESI SFEKTV)FARRI KVVELC QPLQVV)LABLI -MVRTJ)LLABLI -MKRLV PVLRWI -ILPMA DFAALJ)LRALI)LRALI)LRALI)LLQI)FARWI -MERI	hh W(5) WSL U(5) U(5) U(7) C(7) C(7) U(7) L(7) U(7)	hh s: YKFPQ IQQVQ IAQVN LEPDY YKLKR PDRIA KKWPA LGVNW LGVNW DIGKG DIFCK DIGKC DIFCK DIGKC HOLVM AENCN EKGHI PLDVR IKGSL HGVTM TTPPES GSDTP AAGAR	moti ssss PHISQF YNVAYL FDMSFS VVQFTVC VFFSAY IYSFAY LGFQNY FMHYNF F-IQF HALVNI HALVNI LRFIEF F-IEY HALVNI LRFIEF F-IEY EMGLKV VPINML VPINML VPINSF E	f-6 YPR(208) FAY(129) FVY(135) IPY(181) IPQ(208) IPY(110) IPY(110) IPY(107) WPE(130) LAV(107) YIY(132) MDQ(44) MDV(137) MDV(137) LEW(122) LQP(53) VKV(118) FPH(90) (99)
18409989 7299338 2501533 20094331 15896586 3913815 14520767 11499591 15833984 549552 15606163 19705171 18313168 730103 220027536 14521276 14521276 14522966 13476168	hhhi (5)EIA F(5)EVI (7)DVY A(7)ELA (5)DTL LLNDKLV LITEEIA V-LLDDE V(4)DKLC H-TYEDL MIEKQQA (1)3VI MIEKQQA (1)3VI UVL MIEKQQA (1)3VI U-LPVYA (1)6)PQL (1)6	hh AV(4)(RD-HPI RD-TPI RD-TPI RC(4)(LD(4)) CRR(4)(C		tif-4 sss VPVQS LPAQS LPVQA GAESA VNIEL LSVIS LSIDS LSIDS LSIDS LSIDS LSIDS LSIDS SLDT TDIKA HNLDT HNLDT HNLDT LSVTT LSVTT LSVTT	hh GSD GSD SSN PSQLDP LDP PAEI PND SR SR SR LSQ YH LUPPA LUPPA SP SP SP SP SP DVPA	hhh SVL QVL ETL KSL KVQ PEK ETH HNQ EIR PEL EYY DAH DER KFF RVY RFY RFY RFY EFTI	hhhh ERMR TMMK KRIN QAVN AKML EAFR YRVT DEIV FEKL IVRR EREI IVRR EREI IVRR EREI IVRR ALRY ARTM ARTM	REYTA; RGYSR; RAHTA] KGJT PNKS((RVQP(MGIK((GIR(KGGR-) PINKK KGGR-) PINKK FNHP(SVAP SVAP SVPN-(EPRA-3 KLPH	hhh SEFRT CAYLE CAYL	hhhh	h IT IR IS VR VR VR VR VR VR VR VR IS VR IS IS IR	hhhhh ELVFC QFLPN AVRPC ENTT AGFES NAGILI NAVKT L-LQN RVLE CKRRE AGLG- LGYNF CKRRE AGLG- LGYNF PGFE(EAFPC KEFGI Q-LS (L-LK (1 SS 	TATATI I ATATI I SSSSS I SSDF I LLSF QSTQM I LLSF QSTQM I NLDL VVNVSM VVNVSG VVNVSG I VRHL VCSGG I VRHL I NMVV CVCSGG VCSGG VCSGG VCSGG VCSGG VMVAP	otif- IC-GF IC-GF IC-GF IIC-GF IIGAT IV-GL II-GL MF-GV IV-GL IV-GL IV-GL IC-GL MF-GH MV-GL IV-GL II-GL I	5 hh PGETDE CGETEE PDTD (5 PYQTKE NDDE (6 NJDE (6 NJDE (6 NJDE (5 GETEE5 VESI (5 VESI (5 VESI (5 GETE (5 GETN (5 SIDKE (5 TDQE HYGK (6	-hhhhi DFSQT EFQDT DFEQT DVFRT])LTESI SFEKT 0FARR KVVELC QPLQV)LABLI -MVRT])LLBLI 0FARL 0FAALI 0FAALI 1LRALI)LRALI 0FAALI 1LLQI 0FAALI 1LLQI 0FAALI 1LLQI 0FAALI 1LLQI 0FAALI 1LLQI 0FAALI 1ALAN	hh	hh s. YKFPQ IQQVQ IAQVN LEPDY YKLKR PDRIA KRWPA LGVNW DDGVE DLGGR SHPES DETGG HDLVM AENCN EKGHI FLDVR KGSLL HKGSL HKGSL HKGSL HKGSL	moti ssss VHISQF PDMSFS PDMSFS VQFTVC VFFSAX TYSFAY LGFQNY FMHYNF FIQF INLIPW THLFSF YPINAL FTAF MRVRPY HALVNI LRFIEF F-IEY EMGLKV VFINAL VHSRF E	f-6 YPR(208) FAY(129) FVY(135) IPY(86) MPV(181) IPQ(208) IPY(110) IPT(177) IPY(210) NPF(130) LAV(107) IPW(109) YIY(132) MDQ(44) MDV(137) MPF(133) LEW(122) VKV(118) FPH(90) (99) MPY(99)
18409989 7299338 2501533 2009431 15896586 3913815 14520767 11499591 15833984 549552 15606163 19705171 18313168 730103 22027536 14521276 22001767 17865692 14520966 13476168 15605904 15643710	hhhi (5)ELIA F(5)EVI (7)DVYI A(7)ELA 5)DTL VLTDEQI LLNDKLVI LITEEIA V-LLDDEI V(4)DKLC H-TYEDLI MIEKQQAI I(13)VII VVLI I(6)PQLC I(6)PQLC I(4)EALI D(5)DRAI MLNESQAC Y-INEEPI L(5)DLLS L(6)EKLI L(6)EKLI L	hh AV(4)(RD-HPH RD-TPF RT(4)() ID(4)I KA(4)(EE(4)() WC(4)F GD(4)F GD(4)F GD(4)F GL(4)F GL(4)C CRR(4)()) CRR(4)() CRR(4)() CRR(4		tif-4 sss VPVQS LPAQS LPVQA GAESA VNIEL LGVQD LSLDG LSLDG LSLDG LSLDG LSLDG LSLDG LSLDG LSLDG LSLDG LSLDG LSLDG LSLDT TSLDS ISLDT TSLDS LSVTT LSVTT LSVTT LSVTT LSCTG	hh GSD GSD SN PSQ LDP AT AT SR LSQ VH GN LVPAK FQPNS VP FDD LD DN SL	hhh SVL QVL RVL ETL KSL KVQ EETH HNQ EIR PEL EVR DAH DER GEFF KFFF RVY RFY, RFY, RFY RFY	hhhh TAMN ERMR TMMK KRIN KLLA QAVN AKML EAFR YRVT DEIV FEKL INTH ALRY FKKI IVRR EREI IVRR GNII MKIA ARTM	REYTA; RGYSR; RAHTA] KGIT PNKS((RVQP(KGGR-] PINKK(KGGR-] PINKK KGGR- FNHP(SVAP-] NGRG(KGFH-] ENMK(FNHP(SVAP-] SVAP-] SVAP- SVAP- KLPH-] 	hhh SEFRT ALL SEFRT SI SI SI SI SI SI SI SI SI SI SI SI SI	hhhh—h VVDT LVQH IIRK VIACRK KLRE WABA TVEG VMRA FLAA VMRA FLAA VMRT ACEM IHKA GGEV-F KLLE-F RLDT RLDT RLAK-I RLEA YDRS	LT LR LR LR LR LR LR LT LT LE FK LR FK LR LR LR LR LR	hhhhh ELVPG QFLPH KRHG ENTT (AGFES NAGII NAVKT L-LQN MGL DIMGL CKRRE AGLG LGYNF PGFE EAHPI KEFGI Q-LS (L-FVN	1 SS 	TATAL ISSSS IATAII ISSDF ISSDF ILLSF GSTQM ILLSF GSTQM INDL IAPVI IVOYSGG TTATM IVGAHL IVGAHL IVCSGG TTATM IVGSGL IVCSGG INTATM IVCSGG IVCSGG INV IVC IVCSGG INV IVC IVC IVC IVC IVC IVC IVC IVC IVC IV	otif- IC-GF IC-GF IC-GF IL-GF IL-GF II-GF II-GF II-GF II-GL MF-GV MC	5 hh PGETDE CGETEE PDTD (5 PYQTKE NDDE (6 NDDE (6 NDDE (6 NDDE (5 VESI (5 VESI (5 VESI (5 NDSVPV VECCTK NEKD NEEDELL NEELE (5 GETN (5 GETV (5 NDKE (5 TDQE HYGK (6 DEES	-hhhhi DFSQTV EFQDTV DFEQTD DVFRTJ)LTESI SFEKTV)FARRI KVVELC QPLQVY)LAELI -MVRTJ)LAELI)LLHTI)HLERV -MKRLV PVLRWI -ILPMU DFAALJ)LRALI)VMHDI)TARLI (TARLI)TARLI)TARLI)TARLI (TARLI)TARLI)TARLI)TARLI (TARLI)TARLI)TARLI (TARLI)TARLI)TARLI (TARLI)TARLI)TARLI (TARLI)TARLI (TARLI)TARLI (TARLI)TARLI (TARLI)TARLI (TARLI)TARLI (TARLI)TARLI (TARLI)TARLI (TARLI)TARLI (TARLI)TARLI (TARLI (TARLI)TARLI (T	hh V(5) V(5) V(5) V(7) LYSK LYSK V(7) LGAG-1 LGAG-	hh s: YKFPQ IQQVQ ILQVV ILQVN VKLKR PDRIÅ KRWPA LGVNW DAGVE DTPCK KSHPES DTPGK RSHPES SSTPP SGSDTP AAGAR: YEPDA	moti ssss yHISQF yNVAYL POMSFS yQFTVC yFFSAY TYSFAYL GFQNY FMYNF FIQF HILFSF YPINAL LRFIEF FIEY HALVNI LRFIEF FIEY LTLGQY YPINAL YHFSRF E	f-6 YPR(208) FAY(129) FVY(135) IPV(86) MPV(181) IPQ(208) IPY(110) IPY(210) NPF(70) WPE(130) IPW(109) YIY(132) MDQ(44) MDV(137) MDY(133) LEW(122) LQF(53) VKV(118) FPH(90) (99) MPY(90) MPY(99) MPY(90)
18409989 7299338 2501533 2009431 15896586 3913815 14520767 11499591 15833984 549552 15606163 19705171 18313168 22027536 14521276 2200767 17865692 14520966 13476168 15605904 15643710	hhhi (5)ELIA F(5)ELIA F(5)ELIA (7)DVYI A(7)ELA S(5)DTL LITELA LITELA ULLDDEU V-LLDDEU V(4)DKL H-TYEDL I(13)VII VVLI MIEKQQA I(13)VII VVLI MIEKQQA I(13)VII VVLI MIEKQQA I(16)PQL (hh AV(4)(RD-HPH RD-TPH RD-TPH RD-TPH ED(4)(CA)(4)(CA)(4)(4)(4)(4)(4)(4)(4)(4)(4)(4)(4)(4)(4)		tif-4 sss VPVQS LPVQS LPVQA GAESA VNIEL LGVQD LSIDG LSLDG LSLDG ISLDG ISLDG ISLDS ISLDS ISLDS ISLDS ISLDS ISLDT PHPDL LSVTT LDIG CSPCP ISPCP	hh GSD GSD GSD PSQ LDP LDP LDP LSQ YH GN SR LE LUPAK FQPNS VP SD SL SL	hhh SVL QVL ETL EXSL KVQ PEK ETH EETH EVR EVR DAH DER FEF FSKFF RVY EFY RKY EFY RKY	hhhh TAMN ERMR TMMK KRIN QAVN AKML EAFR YRVI FEKL PNVI FEKL IVRR EREI RQVR FKKI IVRR EREI MKIA ARTM ARTM	REYTA; RGYSR; RAHTA] KGIT PNKS(RVQP(KGGR-1 PINKK(FNHP(SVAP-1 NGRG(KGFH-1 ENMK(PGAN-7 TTRS- SVPN-4 EPRA-3 KLPH-1 	hhh SEFRT AYLE LeYKS 5)PMG 5)PMG 5)VLE 5)VLE CFDK VLE 5)VLE 5)VLE 5)VLE 5)WET 5)AKK CCWV 5)LEG 5)WDV CDWSLE EPSR STPTK SVDD	hhhh-h VVDT LVQH IRK VIACRK YIKN-K KLRE MABA VMRA FLAA VMRA FLAA VMRA FLAA VMRA FLAA YMKU RLET ACEM IFRAA IFAA GGEV-F KLLDF RLEA YDRS - - FLAA GGEV-F KLLDF RIAK YDRS - - - - - - - - -	h IR IR IS IS IT LT VR AH LE AK IR IR IR IR IR IR IR VD IR IR IR VD IR VD IR VD IC	hhhhh ELVPG QFLPPI KRHG- ENTT (AGFES NAGIL - L- DY RYLE (DIMG RYLE (DIMG RYLE (DIMG CKRRE AGLG- CKRE K- VR (KEFGI Q-LS (L-LV)	1 SS 	TATM ILSS 2 IATDI ILSS 2 SSTOM ILLS 2 STATM SOUTHOUS COMPANIAN SOUTHOUS SUPPO SUPO SU	Ctif- C-GF IC-GF IC-GF IIC	5 hh PGETDE CGETEE PDTD (5 PYQTKE NDDE (6 NVHEVP NRTSAQ NDGT (5 GETEEE GEAE (5 VDSVFV VESI (5 NDSVPV VESI (5 NDSVPV VESI (5 MDSVPV VESI (5 TDQU HYGK (6 DEES DDADYA	-hhhhi DFSQTV EFQDTV DFEQTI)LTESI SFEKTV)FARRI KVVELC QPLQVV)LAELI -MVRTI)LAELI)LAELI -MVRTI)LAELI)LAELI 1)HLERV -MKRLV PVLRWI -ILPMA DFAALI)LRALI)LRALI)LRALI)JARW -MERII)AIDMI 	hh V(5) V(5) V(5) V(5) URAG- URA	hh s. YKFPQ IQQVQ IAQVN LEPDY YKLKR PDRIA KKWPA LGVNW DDGGVE DTPCK DTPCK DTPCK DTPCK INGSL HDLVM AENCHI EKGTI HGVTS GSDTP AAGAR YEPDA	moti ssss PHISQF YNVAYL FDMSFS VQFTVC WFFSAY IYSFAY LGFQNY FMHYNF F-IQF INLIPW INLIPW F-ITAF F-ISY HALVNI LRFIEF F-IEY EMGLKV LYPINML VHFSRF S LVLVVV 3ASGEY LQPGNH	f-6 YPR(208) FAY(129) FVY(135) IPY(86) MPV(181) IPQ(208) IPY(110) IPY(177) IPY(210) NPF(70) WPF(130) IPY(109) YIY(132) MDQ(44) MDV(137) MDV(137) MDV(137) MPF(133) LEW(122) LQV(53) VPH(90) (99) MPY(90) FPY(90) FPY(90) TPP(49)

Figure 4

Multiple sequence alignment of SAM-dependent radical enzymes. Sequences are denoted by NCBI gi number, conserved domain name (as in NCBI CDD database [141]), short protein name (if available), and abbreviated species name. Secondary structure elements extracted from PDB structure <u>IOLT</u> coordinates are shown above the alignment. Residues are highlighted according to the amino acid properties with designations as in Figure 2.

of one α -layer in TIM-barrels, and correlated changes in sheet curvature and strand orientation. The two classes of SAM-binding enzymes both use loops between strands and helices to interact with various moieties of relatively extended SAM molecules, but the details of this interaction are quite different (see below).

TIM barrel-like catalytic domain in QueA?

Queuosine is a hyper-modified nucleoside in bacterial and eukaryotic tRNAs, produced by a multi-step enzymatic pathway that includes a transfer, with simultaneous isomerization, of ribose moiety from SAM to a modified base in tRNA, called 7-(aminomethyl)-7-deazaguanosine, or preQ1. This step is performed by QueA protein, an Sadenosylmethionine:tRNA ribosyltransferase-isomerase. QueA homologs are found in most bacteria, but their sequence is not strongly similar to any other protein family, and high-resolution structure of QueA in complex with SAM is unavailable. We interrogated the fold recognition meta-servers with individual QueA sequences and with a probabilistic model of aligned QueA homologs. The highest 3D-Jury consensus score (69 units, indicating the upper level of the "gray zone" of provocative, if statistically insignificant, sequence similarities [56]) was to pyruvate kinases, a distinct class of proteins with three structural domains. The C-terminal, regulatory domain of pyruvate kinases has no counterpart in QueA. The other two domains are arranged in such a way that a smaller, β -barrel domain is inserted into the larger, α/β TIM-barrel domain but folds independently. Similar arrangement of two domains is predicted for QueA.

When this manuscript was under preparation, the structure of QueA from Thermotoga maritima was resolved (PDB structure 1VKY). In agreement with the fold recognition data, it shows an α/β domain with insertion of independently folding β -barrel (Figure 1k). The structure of the α/β domain shows one β -sheet, with preponderance of α -helices on one side (Figure 1k). This structure resembles a semi-barrel, given a strongly curved β-sheet, relative absence of α -helices on the concave side, and a lid-like irregular arrangement of elements that covers the cavity. There is an unresolved protein segment of 32 residues, which should be located close to the inner β -layer of the semi-barrel, and may in fact extend its wall. An unresolved ligand is placed in the proximity of the C-termini of several β -sheets, and if this is in fact SAM, its binding mode would be similar to what is observed in other SAMbinding proteins (see below).

β -barrels in QueA and fluorination enzyme

Both QueA and fluorination enzyme structures show fusions of a larger α/β domain and a smaller all- β domain with barrel-like topology. The role of all- β domain in QueA is unclear, but it is not very likely to be involved in interaction with SAM. In contrast, the β -barrel domain in FlA (which, in fact, is more similar to "smashed β-can," with one side caved in, producing a double-concave surface) makes many contacts with the ligand. Proceeding from the N- to C-terminus, the Asp210-His211 (as in PDB structure <u>1ROP</u>) dipeptide in the loop after the first strand bonds with the amino group of methionine; Asn215 bonds with the amino group of adenine; Ser269 and Arg270 after strand 5 can form 4 hydrogen bonds altogether, all with the carboxyl group of methionine; and at the C-terminus of strand 5, Arg277 and Ala279 provide additional interactions with adenine. While the catalytic mechanism of FIA is dependent on correct positioning of the fluoride atom with regards to ribose, which is mediated by Ser158 in the Rossmann-like domain (see above and reference [46]), the β -barrel domain appears to be essential for correct orientation of SAM, which serves as fluoride acceptor.

Between the sheets: double- β SAM-binding folds with a common theme of internal domain duplication Decarboxylase

S-adenosylmethionine decarboxylase (EC: 4.1.1.50, SAMDC) is a key enzyme in spermidine and spermine biosynthesis. It produces decarboxylated SAM (dcSAM), which then donates aminopropyl group to putrescine or spermidine, two essential intermediates in polyamine biosynthesis. Because polyamines link diverse pathways in cellular metabolic networks, and because chemical inhibitors of SAMDC display potent antitumor and antiparasite activities [57,58], structure-function relationships of SAMDC are of considerable interest.

SAMDC activities have been purified from all three domains of Life (bacterium E. coli, archaea Methanococcus jannaschii, and several eukaryotes), and certain common features of the enzymes have been noticed. All three enzymes are processed in vivo, forming a small subunit derived from the N-terminus and a large subunit accounting for the rest of the molecule; the N-termini of all large subunits contain a pyruvoyl group, produced from a serine residue by autoprocessing and required for the formation of the Schiff base during catalysis. All studied enzymes form multimers from the heterodimers of large and small subunits. There are also differences among bacterial, archaeal and eukaryotic SAMDC: the length of the precursor proteins in different species varies from 105 to 460 amino acids; mammalian enzymes require putrescine for full activity, E. coli enzyme requires Mg²⁺ cation, while archaeal and plant enzymes apparently do not require those factors. The bacterial enzyme is a tetramer of heterodimers, while eukaryal and archaeal enzymes are homodimerized heterodimers.

Comparative sequence analysis has revealed statistically significant sequence similarity between archaeal and bacterial SAMDC [59]. Multiple alignment of these two classes of SAMDC spans the complete length of the shorter (ca. 120 aa) archaeal enzymes, and also suggests that there are two types of bacterial enzymes – some are about the same size as archaeal SAMDC, and some are longer and phylogenetically distinct (reference [60], Figure 5a, and unpublished observations). No sequence similarity has been reported between these enzymes and eukaryotic SAMDC.

High-resolution structures of eukaryotic SAMDC from humans and plants in complex with substrate analogs and various inhibitors have been reported. The heterodimer folds as a sandwich of two β -sheets between α -helical regions, where the smaller subunit forms a half of one β -

Α

	SSSSSSS		SSSSSSSS-		-ss s	sssss		-S SSSSSS		SSSSSSS
human_N	TEKLLE <mark>VW</mark> FSR	24	QCSI <mark>I</mark> SVT	0	KTD K	KQE <mark>A</mark> Y <mark>VL</mark> –	0	S <u>ES</u> S <mark>M</mark> F <mark>V</mark> SK	0	RR <mark>FIL</mark> KT <u>C</u> G
leishm_N	PEKRLEVILRC	24	NAQI <mark>V</mark> SRE	0	SNE Y	INSY <mark>VL</mark> –	0	TESSLFVMK	0	NR <mark>IIL</mark> ITCG
yeast_N	PEKL <mark>L</mark> E <mark>IW</mark> FFP	26	KCEV <mark>L</mark> SMK	0	<mark>k</mark> tk e	E <mark>LDA</mark> FLL-	0	SESS <mark>L</mark> F <mark>V</mark> FD	0	HK <mark>LTM</mark> KTCG
	ssssssssssshhhhh sss		ssssss-		-s sss	sssssss		SSSSSSSSS		SSSSSSSSS
human_C	PDQTLEILMSELDPAVMDQFYM	19	IPGS <mark>V</mark> IDA	2	FNPC	Y <mark>SMNG</mark> MK	1	DGT <mark>YWTIHIT</mark> PEPE	0	FSY <mark>VSF</mark> EINL
leishm_C	DDTQLSMTMYGLDKEQTKHWFS	25	GTWT <mark>L</mark> HDL	1	FEPC	YS <mark>INAI</mark> R	0	DEE <mark>YQ</mark> TMHITPEDH	0	CSF <mark>ASYE</mark> INS
trp	NDTQLSMTMYGLDRTQTKHWFS	25	DSWK <mark>L</mark> HDL	1	FEPC	YS <mark>INTI</mark> R	0	GAE <mark>YQTIHIT</mark> PEDH	0	CSF <mark>ASYE</mark> INT
DCAM_YEAST	DDET <mark>FEVLM</mark> TE <mark>L</mark> DPEC <mark>ASKFV</mark> C	36	NSAQDSDL	8	FTPC	YSSN <mark>MI</mark> L	1	EKY <mark>YYT</mark> LHVTPEKG	0	WSY <mark>AS</mark> FESNI
DCAM_XENLA	PDQTLEILMSELDPEVMDQFYM	19	ITGS <mark>V</mark> IDA	2	FSPC	YS <mark>MNG</mark> MK	1	DGT <mark>YWTIHIT</mark> PEPD	0	FSY <mark>VSF</mark> EINV
DCAM_DROME	PDQT <mark>IE<mark>ILM</mark>QD<mark>L</mark>DPET<mark>M</mark>SIFY</mark> K	19	LPTM <mark>H</mark> IDD	2	FDPC	YS <mark>MNGI</mark> N	1	KGE <mark>YMTIHIT</mark> PENQ	0	FSY <mark>VSF</mark> EINV
s2	PDHTLEILMSDLDEEVLHKFTK	21	AGAD <mark>V</mark> HDE	1	FDPC	Y S MNA YM	2	TDQ <mark>Y</mark> AT <mark>IHVT</mark> PEKA	0	FSF <mark>ASF</mark> ETNQ
	hhhhhh		hhhhhhh?		_	SSSSSS		SS		SSSSSSS
SPEH_ARCFU	VGRHI <mark>I</mark> AE <mark>LY</mark> G	15	VEEV <mark>V</mark> DKA	11	FNPH	V <mark>T</mark> GIVLI	0	A <u>ES</u> HVS <mark>IHTW</mark> PEYG	0	LVN <mark>L</mark> DIF <u>IC</u> G
SPH1_AERPE	<mark>F</mark> GLHV <mark>Y</mark> GN <mark>FY</mark> C	15	VLEA <mark>A</mark> REG	10	KIGE	VS <mark>VVAI</mark> I	0	LES <mark>HITVHTW</mark> PEYR	0	FAT <mark>VDVYS</mark> CG
SPH2_AERPE	VGKHV <mark>Y</mark> GSLYG	15	VVRA <mark>A</mark> ESA	12	GEKG	VS <mark>VIVL</mark> V	0	LES <mark>HLALHTW</mark> PEYD	0	YAT <mark>F</mark> DIYICG
s3	LGKHL <mark>ILE</mark> LWG	15	LVDS <mark>V</mark> KAC	11	FSPQ	ATGVAVL	0	AES <mark>HIAIHTY</mark> PEYG	0	YAALDVFICG
SPEH_THEMA	LGRHL <mark>VAE</mark> FYE	15	MKQA <mark>A</mark> YES	11	FLPY	VS <mark>GVVV</mark> I	0	SES <mark>HLTIHTW</mark> PEYG	0	YAA <mark>IDL</mark> FICG
SPEH_THEVO	VGIHI <mark>I</mark> ADFYG	15	IEGA <mark>V</mark> KYG	11	FRPQ	AS <mark>GVVL</mark> L	0	AES <mark>HL</mark> SFHTWPEYG	0	LVTLDIYICG
s4	VGEHLLDLYG	15	LREA <mark>A</mark> DAL	21	EAGG	VTGVLLL	0	AES <mark>HLSIHTW</mark> PEHG	0	FAA <mark>ID</mark> AFMCG
SPEH_PYRAB	<mark>I</mark> GHHY <mark>IVE</mark> AAG	15	FLEA <mark>A</mark> KRG	11	FSPM	V <mark>SGVVI</mark> V	0	AES <mark>HISVHTW</mark> PEKG	0	YAA <mark>L</mark> D <mark>V</mark> YICG
SPH1 PYRAE	VGRHV <mark>Y</mark> GNLYG	15	VKEA <mark>V</mark> KVA	12	GPGG	LT <mark>VFAV</mark> V	0	AES <mark>HISIHTW</mark> PEHG	0	FAT <mark>VDV</mark> YICG
SPEH PYRHO	<mark>I</mark> GHHY <mark>IVE</mark> AAG	15	FLEA <mark>A</mark> KRG	11	FSPM	VS <mark>GVVI</mark> V	0	AES <mark>HISVHTW</mark> PEKG	0	YAALDVYICG
SPH1 SULSO	<mark>I</mark> GKHV <mark>F</mark> GN <mark>LY</mark> D	15	VLEA <mark>V</mark> DIA	13	GKKG	VS <mark>VIAL</mark> V	0	EES <mark>HIALHTW</mark> NEYN	0	YAT <mark>LDV</mark> YICG
SPH2_SULSO	VGKQV <mark>Y</mark> GSLYE	15	IKEA <mark>A</mark> DIG	10	KIGE	VS <mark>VVAI</mark> I	0	LES <mark>HITIHTW</mark> PEYR	0	FAT <mark>VDVYS</mark> CG
SPEH_AQUAE	LGLHILADLYG	15	LEGA <mark>V</mark> KYA	11	FQPH	ATGVVLL	0	AES <mark>HISIHTW</mark> PEHG	0	LAT <mark>VDVY</mark> ICG
SPH1 BACHD	MGRHV <mark>I</mark> AE <mark>LW</mark> G	15	FVDA <mark>A</mark> LKA	11	FAPQ	IS <mark>GVVI</mark> I	0	SES <mark>HLTIHSF</mark> PEHG	0	YAS <mark>IDV</mark> YICG
SPEH PSEAE	LGKQL <mark>VIE</mark> LFD	15	MLEA <mark>A</mark> RQA	11	FSPI	IS <mark>GVVV</mark> I	0	AES <mark>HLAIHTW</mark> PEYG	0	YAA <mark>VDVF</mark> ICG
SPED SALTY	LKLHG <mark>FNN</mark> LTK	33	RLTE<mark>I</mark>LSE	16	YEPQ	AS <mark>VTIL</mark> V	27	DKSHICVHTYPESH	8	RAD <mark>IEV</mark> SICG
SPED PSEAE	LKLHG <mark>FNN</mark> LTK	33	RLTQ <mark>I</mark> LTD	16	YDPQ	AS <mark>VTIL</mark> I	27	DKSHITVHTYPEIH	8	RVD <mark>IDV</mark> SICG
SPED_YERPE	LKLHG <mark>F</mark> NN <mark>LT</mark> K	33	RLTE <mark>I</mark> LSE	16	YDPQ	ASVTILV	27	DKSHICVHTYPESH	8	RADIEVSICG
SPED CLOAB	IKLHG <mark>F</mark> NN <mark>LT</mark> K	33	RLTE <mark>I</mark> LTN	16	YDPQ	ASVTILV	29	DKSHITVHTYPEYH	8	RAD <mark>IDV</mark> SICG
SPED_BUCAI	LKLYG <mark>F</mark> NN <mark>LT</mark> K	33	RLTK <mark>I</mark> LKK	16	YEPQ	ASVTILV	26	DKSHICVHTYPESH	8	RADIEVSICG
SPED_XYLFA	LRLQG <mark>F</mark> NN <mark>LT</mark> K	33	RLTQ <mark>I</mark> LTD	16	YDPQ	ASVTILI	24	DKSHITVHTYPETH	8	RADIDVA CG
SPED ECOLI	LKLHG <mark>F</mark> NN <mark>LT</mark> K	33	RLTE<mark>I</mark>LSE	16	YEPQ	ASVTILV	27	DKSHICVHTYPESH	8	RAD <mark>IEVS</mark> CG





Figure 5

Multiple sequence alignment of SAM decarboxylases. Sequences are denoted by gene and species name. **A:** Decarboxylase alignment – the residues underlined on the top line are those involved in SAM-binding, enzyme self-processing, and catalysis. Residues are highlighted according to the amino acid properties with designations as in Figure 2. **B:** Superimposition of the conserved residues in SAMdc from *T. maritima* (PDB structure <u>ITLU</u>), human (PDB structure <u>II7B</u>) and potato (PDB structure <u>IMHM</u>) are shown on cartoon representation of *T. maritima* structure (gray). **Blue** – Ser; **Red** – Ser and Ser converted into pyruvoyl group (or pyruvoyl group with covalently bound S-adenosylmethionine methyl ester); **Orange** – His; **Cyan** – Cys.

sheet, and the larger subunit completes this sheet and accounts for all the strands in the other sheet. The arrangement is unique among the known protein folds, but visual inspection and superposition of the two α - β halves of the molecule revealed their remarkable similarity and suggested the hypothesis of internal duplication [61,62]. The evolutionary origin and catalytic mechanism of prokaryotic SAMDC remained unclear.

Searches of sequence databases with the PSI-BLAST program and more involved probabilistic models of aligned SAMDC enzymes confirmed statistically significant sequence similarity between archaeal and bacterial enzymes, and also, intriguingly, produced several statistically insignificant local matches to one-half of eukaryotic SAMDC sandwich, in the area corresponding to the β strand 12 in the three-dimensional structure. This strand is positioned next to the active center of the enzyme, and contains residues important for catalysis and/or binding of the substrate (see below). Because both these residues appeared to be preserved in the BLAST output, we sought better statistical validation of this similarity using Metaserver [56]. When SAMDC homolog from archaea Archaeoglobus fulgidus was used as a query, the highest 3D-Jury consensus score (46-60) was reported to the set of the eukaryotic SAMDC structures; this score is at the top of the zone with borderline significance, where most of the non-trivial similarities are discovered [56]. The first false positive (bacterial luciferase) was associated with the sharp drop in the 3D-Jury scores (14.5).

Almost complete archaeal sequence can be aligned to the half of eukaryotic template, with just one short gap. Conversely, the aligned region of the template corresponds almost precisely to the C-terminal half of the double sandwich. We conclude that the archaeal enzyme may resemble a half of the eukaryotic SAMDC fold and may be directly related to the pre-duplication ancestor of that fold. Multiple sequence alignment of archaeal, bacterial, and eukaryotic enzymes strongly reinforces these observations (Figure 5a). The C-terminal halves of eukaryotic enzymes could be aligned to prokaryotic homologs directly and unequivocally; the structurally similar N-terminal halves had to be superimposed using the knowledge of secondary structure and information about a few conserved residues.

The functional and evolutionary implications of the alignment are provocative. In mammalian enzymes, SAM decarboxylase is active as a dimer in which each protomer contains one large and one small subunit, and each of the two halves of the sandwich contributes several residues to binding the substrate and actually performing the catalysis. In particular, Ser residue in β -strand 4 of the eukaryotic enzymes, which is converted into catalytic pyruvoyl

group, appears to be within a short distance of the carboxyl group of SAM forming a Schiff base adduct with it. Before product release, carbon of decarboxylated SAM is protonated by adjacent Cys (Figure 5b). This protonation regenerate the pyruvoyl group [63]. Also close to the active site is the side chain of the histidine residue in strand 12, which is believed to be responsible for abstraction of a proton from the α -carbon of the catalytic serine during proenzyme processing [64].

Two acidic residues contribute to binding of SAM: glutamic acid at the C-terminus of strand 3 contacts the base, and another glutamate, at the C-terminus of strand 12, interacts with both hydroxyl groups of the ribose ring. All these residues are conserved in eukaryotic SAMDCs – some in the N-terminal half of the sandwich, and others in the structurally equivalent C-terminal half. Interestingly, in archaeal and most bacterial enzymes, the pattern of conserved elements in the two halves of eukaryotic enzymes (Figure 5a), as if the bacto-archaeal enzyme is a homolog of one half of the eukaryotic enzyme, and the β -sandwich in the holoenzyme are made of two identical molecules.

When this manuscript was in preparation, the structure of ligand-free holoenzyme from bacterium *T. maritima* was deposited in the database (PDB structure <u>1TMI</u>). Analysis of this structure confirms this sequence-based prediction and suggests that the bacto-archaeal form is ancestral, and the eukaryotic form has been derived from it by domain duplication/fusion, followed by functional specialization of two halves (most notably, by mutating the C-half of the enzyme so that it no longer undergoes autoproteolysis – Figure 5b).

SET domain

Discovered as conserved domain shared by chromatin remodeling proteins Su(var)3–9, E(Z) (short for Enhancer of Zeste) and Trithorax, SET domains turned out to be a distinct class of SAM-dependent methyltransferases. All studied SET methyltransferases transfer methyl group to lysine within various nuclear proteins involved in chromatin function and regulation of transcription, such as histones, TAF10, tumor suppressor p53, but also in such diverse proteins as Rubisco and cytochrome C [65-68].

In SET-domain methyltransferases amine of the substrate lysine residue access the methyl donor (SAM) through a narrow channel connecting the substrate and SAM binding surfaces [69]. SAM binding site and the catalytic center of all studied SET domains seem to be constructed on the unusual but conserved, all- β , knot-like structure [70]. Adenosyl moiety of SAM interacts directly and indirectly, through water, with conserved histidine (PDB structure:

<u>109S</u>-His297; <u>1P0Y</u>-His243). This histidine may serve as a proton acceptor for the hydroxyl group of invariant Tyr (PDB structure <u>109S</u>-Tyr335). The -OH of this Tyr is within 4 Å of the presumptive location of the substrate Lys N ζ , and may be involved in Lys side chain deprotonation (deprotonated Lys is presumed to make a nucleophilic attack on the SAM methyl group). Positively charged amino nitrogen from SAM hydrogen bonds with the side chain of invariant asparagine (PDB structure <u>109S</u>-Asn296). This interaction may contribute to the compact conformation of the SAM molecule.

Phylogenetic analysis of the SET domain suggests that it is an evolutionary innovation in the eukaryotic lineage (with secondary lateral transfer to bacteria, archaea and viruses) [71]. SET domains (Figure 1h and SCOP superfamily: 82199) have a fold unique for SAM binding proteins – a substrate binding subdomain between two structural repeats, which may have evolved by duplication of 3-stranded unit with a generic ligand binding role [71,72]. Those repeats have a β -clip fold formed by double-stranded ribbons sharply bent in two places; the ribbon ends form incomplete barrel.

Similar duplication of a basic three-stranded unit containing the β -clip structural motif probably occurred also in related SAF and dUTPase superfamilies [72], which, however, tend to bind sugar and sugar derivatives [72]. There have been several other cases of adaptation of a generic ligand binding domain for SAM-binding, both in enzymes and in regulatory proteins without catalytic activity (see below).

SAM synthetase

S-adenosylmethionine synthetase (SAM synthetase, ATP:L-methionine S-adenosyltransferase, or MAT, EC: 2.5.1.6) is the main, or, possibly, the only enzyme of de novo SAM biosynthesis. SAM synthases from bacteria and eukaryotes are closely related at the sequence level and have very similar structures [73]. SAM synthases transiently interact with SAM prior to its release. The mechanism of reaction is thought to rely on conserved His14 (Figure 6), which acts as an acid to cleave the C5'-O5' bond of ATP, while simultaneously a change in the ribose ring conformation from C4'-exo to C3'-endo occurs, and the S of Met makes a nucleophilic attack on the C5' to form SAM [74].

The fold of bacto-eukaryal SAM synthetase is unique; each protein chain is based on a β - α - β - β - α - β module that folds into a wedge-like shape. A polypeptide chain consists of three such tandemly repeated modules, so that the complete SAM synthetase fold looks like a three-slice cream pie with topping made of β -sheets (Figure 1L). The active form of the enzyme appears to consist of two pies, with β -

layers facing each other. Two SAM molecules are bound between the sheets of this dimer. In the E. coli enzyme, both subunits contribute many residues to SAM binding (Komoto et al. [74], Figure 6, and PDB structure: 1P7L and 1RG9) In particular, adenosine binds to Asp163, Arg229, Phe230 on one subunit and to Ser99 on another subunit, and interacts with many additional amino acids on both subunits via water-mediated hydrogen bond network. Methionine binds to Gln98 and Asp238 on one subunit, to Glu55 on another, and likewise makes many additional water molecule-mediated contacts. Bacto-eukaryal SAM synthetase is an evolutionary unique sequence and structural family. Even SAM decarboxylase, which is superficially similar in that it also sandwiches SAM between two β-sheets, has no detectable sequence or structure similarity to SAM synthetase.

SAM synthetases from Archaea have been isolated on the basis of their biochemical activity [75]. We performed sensitive searches of the conserved domain database, and found clear evidence for common ancestry of all SAM synthetases (Table 2). Multiple sequence alignment and secondary structure prediction indicate that archaeal enzymes share the same three-dimensional structure as their eukaryotic and bacterial homologs. All known SAM synthetases have conserved GHPD signature containing the main catalytic residue (His14 in E. coli gi:46015497) [75]. Despite high sequence divergence between archaeal and bacto-eukarval enzymes, the complement of substrate binding residues is well-preserved (Figure 6). The apparent common origin of this unique enzyme in all major divisions of Life is of great interest for reconstruction of the repertoire of SAM-binding protein in the ancestral life forms (see below).

SAM-binding modules derived from generic ligand-binding domains

ACC synthase

ACC synthase (S-adenosyl-L-methionine methylthioadenosine lyase, EC: 4.4.1.14, KOG0256) catalyses the ratelimiting step in biosynthesis of plant hormone ethylene by the α , γ -elimination of methylthioadenosine from SAM to produce 1-aminocyclopropane-1-carboxylate (ACC) [76]. ACC synthases require pyridoxal phosphate (PLP) for activity, and are related in sequence and structure to a large, diverse group of PLP-dependent transferases. The shared catalytic domain of this fold is of $\alpha/\beta/\alpha$ type, with mixed central β -sheet of 7 strands (order 3245671), where strand 7 is antiparallel to the rest (SCOP fold: 53382; Figure 1j). Several residues are essential for the substrate binding (reviewed by Jakubowicz; numbered as in PDB structure <u>1B8G</u>): Glu47 is responsible for putative ionic interaction with SAM; Ala46 and Arg407 interact with carboxypropyl moiety of SAM; Arg150 interacts with ribose moiety; and Ser18, Tyr19, Phe20, and Pro146 form hydro-

$ \frac{d_1 f}{d_1 + d_1 + $					46					5			55							
<pre>4015497 ECOLI #HLEFESUSE HUE ADD SDATDALDE ==-CE FKAR #CET == YUKCG+==TL & ETESATO== TERMURE IGNES 400245 HUMAN TFLEFESUSE HUE ADD SDATDALDE ==-CE FKAR #CET ==-YUKCG+==TL & ETESATO== TERMURE IGNES 1583694 X1EA SULFESUSE HUE ADD SDATDALDU ==-CHERCH ==-CETATONE ==-CHERCH ==</pre>	gi#	species	SSSSSS	SSSS	hhhl	hhhh	hhhhhh	hhh		SSSSSS	sss s	ssss	sssss	hh	hhl	h	hhhh	hhhhh	1	
400245 HUMAN TILTES SVGEN FERK CRO SDATDAREOOFPERA : CTVARTCHLARE TERMADKALERKKYUNDKAKERKKYUND	46015497	ECOLI	KH <mark>LFT</mark> S	ESVSE	GHPDK		SDA <mark>VL</mark> D	A <mark>IL</mark> E-	QD PKA	RVACET-	<mark>YVKT</mark> G <mark>I</mark>	1 <mark>VLV</mark>	GEIT	TSA <mark>W</mark> VD	I EI	E	ITRN	TVRE I	G <mark>y</mark> ve	IS
1883694 XYLFA SYLFASSUSE HORAADO SOATDALIVOPPARE CTYVECUALA ETTSAVDLAALARC TETSAVDLAALARC TETSAVDLAALAALARC TETSAVDLAA	400245	HUMAN	TF <mark>LFT</mark> S	ESVGE	GHPD <mark>K</mark>		SDA <mark>VL</mark> D	a <mark>hl</mark> q-	QDPDA	KVACET-	<mark>VAKT</mark> G <mark>I</mark>	4 <mark>ILL</mark>	ACEIT	SRA <mark>A</mark> VD	––– <mark>Y</mark> QI	K	V VRE.	<mark>AV</mark> KHI	G <mark>y</mark> de	s
601653 CAELE KILTES SUBER DRACED SOATDARLKDEPKAR CETVERTEGVALCETTESKAPOXQVLVERKIEKIG-TOS 2009528 [UNN FYTTES SUGAP DRACED SOATDARLKDEPKAR CETCURIS CETVALCETTESKAPO	15836994	XYLFA	SY <mark>LFT</mark> S	e sv s e	GHPDK		SDA <mark>VL</mark> D	a <mark>il</mark> v-	QDPRA	RVACET-	<mark>MVKT</mark> G <mark>V</mark>	/ <mark>AII</mark>	ACEIT	TSA <mark>W</mark> VD	<mark>L</mark> Ей	A	LTRK	VI VDI	G <mark>y</mark> ns	ss
2205528 [USNN PTYFFSERVER HUDKESDG SDA LDACKDEPARSE CEVPOTOG ETTISTIDVO.GTVRKETDELCERP 230766 MYCEN WGRYAAAVGH DICCOLG SDA LDACKOSPISE CEVLANELVV. BETTHAAVOVV.GTARKITEPLCINES 240666 MYCEN WGRYAAAVGH DICCOLG ADS LDECKOSPISE CEVLANELVV.G ELLINGYDDVV.GTARKITEPLCINES 2406472 CHLE OPPEIVERKGI HDICCOL ADS LDECK	6016553	CAEEL	KF <mark>LFT</mark> S	ESVSE		4CDOI	SDA <mark>VL</mark> D	a <mark>hl</mark> k-	ODPNA	KVACET-	<mark>VTKTG</mark>	1 <mark>VML</mark>	CEIT	SKA <mark>V</mark> VD	<mark>Y</mark> 0	v	LVRR	<mark>vi</mark> eki	G <mark>F</mark> TD	s
13357974 UREPA KKITISSISVAAHD DITCICL SDA LDECLAOSINSE CEVLANELTU, BETTHEAVDWATARETTEPLOSINE 230666 M/VPP WGRVAANGREH DISCICL AASUSTALCK (13) TOVEL GGUA (6) MVPPTYTLIS RATHETDIX (4) VIK (5) TAXIAAGEA-FLIKK 2164672 CHETE QPEEVERKGI H DISCICL AASUSTALCK (13) TOVEL GGUA (6) MVPPTYTLIS RATHETDIX (4) VIK (5) TAXIAAGEA-FLIKK 2164672 CHETE QPEEVERKGI H DISCICL AASUSTALCK (13) TOVEL GGUA (6) MVPTYTLIS RATHETDIX (4) VIK (5) TAXIAAGEA-FLIKK 2164672 CHETE QPEEVERKGI H DISCICL AASUSTALCK (13) TOVEL GGUA (6) MVPTYTLIS RATHETDIX (4) VIK (5) TAXIAAGEA-FLIKK 2164672 CHETE QPEEVERKGI H DISCICL AASUSTALCK (13) TOVEL GGUA (7) FILE PARTYTORG (4) VIC TAEAAARCA-FLIKE 2164672 CHETE QPEEVERKGI H DISCICL AASUSTALCK (13) TOVEL GGUA (7) FILE PARTYTORG (4) VIC TAEAAARCA-FLIKE 2164672 CHETE QPEEVERKGI H DISCICL AASUSTALCK (13) TOVEL GGUA (7) FILE PARTYTORG (4) VIC TAEAAARCA-FLIKE 2164672 DM (3) VIEAA CKO PEI NOCODE (5) OLG COLOR TITLE DISLAMAA TITYARH VIC QAATEXIC-COTFFUERPOSTOR (10) (10) NUT (22095828	FUSNN	FT YFT S	EFVSP		ISDOI	SDATLD	a <mark>cl</mark> k-	DPNS		<mark>FCTT</mark> G	<mark>vvv</mark>	GEIT	TSTY ID	<mark>v</mark> õi	D	IVRK	K <mark>IDE</mark> I	. G <mark>y</mark> re	2
230066 WYCEN WGRYDARAGE H DITCICL ADS LOFUKOSPTSH CEVFASHITM CERLINGYUDWOTGRYDARGETSH 21646472 CHIE WGRYDARGH DITCICL ADS LOFUK	13357974	UREPA	KK <mark>IIT</mark> S	ESVGA			SDATLD	e <mark>cl</mark> a-	ODKNS			<mark>IVI</mark>	ACEIT		v vi	K	TAWE	ii kpt		IN
<pre>3334228 META RETEIVERGUE SIGGLAES (SALGE (13) TOYEL GUE (0) WSPITLIASEATHETDK (4) VIE (5) TAVEAARCULKE 21646472 CHLT OPFELVERGUE STOCKS MEA CIDEC (13) IKCLL GESDEVEG (7) KLT PDRATTEROK (4) VGELAEAARCULKE p P + m</pre>	2500686	MYCPN	WGRYV	EAVGR	HPDK	CDOT		E <mark>CI</mark> K-	OSPTS	HVACEV-	FAS KN	TMV	GEIL	TTGYVD	v v)	TGWK	VLNRI	G <mark>Y</mark> TF	IN
21646472 CHLTE QPPELVERKGIGHERT CDS MEA GIDLCR (13) INGUL GUSTPRTG (7) KLTPIDEATYTONG (4) VGETAEAAAKRGIRE P P P P P P P P P P P P P P P P P P P	3334428	METITA	RPTET	RKGI	HPDS	CDG	AESVSR	ALCK (13) TD OVE	GHA			SCRATI	METLOK	(4) V TI	£ K(5)	TAVK	AAKE-		ĸ
p p	21646472	CHLTE	OPERIA	RRKGT	HPDT		MEAVCT	D <mark>LC</mark> R (13) TOKGI	LU GRS		(7) KT.T	FDRA	TYTCNG	(4) VGI	E	TAEA			ज
<pre>sssss hhbhhbhbh ssss ss hhbhbhbhb sssss ss hbbhbhbhbhbhbhbhbhbhbhbhbhbhbhbhbhbh</pre>	210101/2	011212	2-	n	P +			2 (10,1	mK		()]]	M	••••	(1).	_				_
<pre>sssss sssss sssss shhhhhhhhh sssssss sss hhhhhh</pre>				Р	P								-		* n 4	2				
sssss ssssss show-hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh					1					P					16	2				
46015497 DM (8) VLSA GKOG PD INGOUR (5) OLICE OCIMES AT THE DULM PAPTYLARH, VORDA SYRKNGTLEWIDE PAKSOUTECY (5) VGIDA WY 46015497 SK (8) VINA LEQC PD LAGGVHL (5) DLAGG OCIMES ATTHE DESUMPAITLARH, VARLABLERNGTLEWIDE PAKSOUTECY (10) RWHTY 1836994 DV (8) VINA LEQC PD LAGGVHL (5) DLAGG OCIMES ATTHE DESWIPAITLARH, VARLABLERNGTLEWIDE PAKSOUTECY (10) RWHTY 2005828 GM (7) TLSC HAO PD LAGGVHL (5) DLAGG OCIMES ATTHE DESWIPAILISHK, VEOQALTRKKNSPLEWIDER SKOTTECY (10) RWHTY 2005828 GM (7) TLSC HAO PD LAGGVHL (5) DLAGG OCIMES ATTHE DESWIPAILISHK, VEOQALTRKKNSPLEWIDER SKOTTECY (10) RWHTY 2005828 GM (7) TLSC HAO PD LAGSVNK (5) TLGG OCIVES ACDE POYMETT DULARELLKKIERORRNKEFVKIGADMKSOVINT (6) DLY DIN 200582 DF (2) -LSC HAO PD LAGSVNK (5) TLGG OCIVES ACDE POYMETT TULAHELLKKIERORRNKEFVKIGADMKSOVINT (4) VOCESN 3334428 VL (11) DCR GOCO DD VDVERK (5) PLINT TSFGV. DESS TERIVIERTENDSE INNETPA VGRDTKINGENKKI (1) TVAAVDRXY 21646472 NL (11) DCR GOCO DD VDVERK (5) PLINT TSFGV. DFS TERIVIERTENDSE INNETPA VGRDTKINGENKKI (1) TVAAFVDRY 21646472 NL (11) DCR GOCO DD VDVERK (5) PLINT TSFGV. DFS TERIVIERTENDSE INNETPA VGRDTKINGENKKI (1) TVAAFVDRY 21646472 NL (11) DCR GOCO DD VDVERK (5) PLINT TSFGV. DFS TERIVIERTENDSE INNETPA VGRDTKINGENKI (1) TVA-OGARH-GGC F GND 7K 200686 SSS hhhhhhhhh hhh 2005486 VIC DESTING KEVIKAV-VPAKY TSATKFF INFT REVIGEN DCI IT THIL DTY-OGARH-GGC F GND 7K 2005486 VIC DESTING KEVIKAV-VPAKY TSATKFF INFT REVIGEN DCI IT THIL DTY-OGARH-GGC F GND 7K 2005486 VIC DESTING KEVIKAV-VPAKY T-SATKFF INFT REVIGEN DCI IT THIL DTY-OGARH-GGC F GND 7K 2005486 VIC DESTING KEVIKAV-VPAKY T-SALDESTVFH INFC FIVIGEN DA VTKILI DTY-OGGARH-GGG F GND 7K 2005486 VIC DESTING KEVIKAV-VPAKY KIKY NPTR REVIGER DI VTKILI DTY-OGGARH-GGG F GND 7K 2005486 NSOC DESTING KEVIKAV-VPAKY NSOK NATH I THIL DTY-OGGARH-GGG F GND 7K 2005486 NSOC DESTING VIC DV-V-L-V-V-RASI DSDP (4) EGVIYI TVI GFSARM DD SV FKNR NGL ITFRENNING ST INFO 400445 RSAAY			SSSSSS	3				SSSS	SSS	hh	hhhhhhh	hhhhhh	hhh		sssss	ssss	SSSS	SS	ssss	s
400245 SK (8) VLVALEQQ PD AQCVAL (6) IE GOQC MFS TTDE BECMPT VLARKINALAGERNA-GETEPTIRPDS AS TTO YT VY (10) RVHTYV 1838994 DV (8) VLVALEQQ PE AAGVAU (6) Y GOQC MFS TTDE BETMPT HILSHKINALAGERNA-GETEPTIRPDS AS TTRY (5) SADAVV 6016553 SI (8) VLVALEQQ PE AAGVAU (6) Y GOQC MFS TTDE BETMPT IVLSRE LVKLTNMKKSSPTPLWRPDS SQVTLEY (10) RVHTYV 22095828 GM (7) TLSC HAQ PD AMGVDIGUGGQ GOME AF RE VED BETMPT IVLSRE LVKLTNMKSSPTPLWRPDS SQVTLEY (10) RVHTYV 22095828 DFTIISNVNKQ VD ASSVNK (5) IE GOQC IVG ADE BEIMPLA IVLSRE LVKLTNMKSS-NELENARSSVS UNY (4) (0) CV ETK 2500686 DF (2) -LSC INOG SE INOAV SSNDEIL GOQC IVG ADE SELTIVLETER INSDELKNET PAVEL REGK KIT IT AMAVORYV 21646472 NL (11) ONE IFFC PE TDAFARK-VI NOTSVGV ADE SETERIVLETER INSDELKNET PAVEL REGK KIT IT AMAVORYV 21646472 NL (11) ONE IFFC PE TDAFARK-VI NOTSVGV ADE SETERIVLETER INSDELKNET PAVEL REGK KIT IT AMAVORYV 21646472 NL (11) ONE IFFC PE TDAFARK-VI NOTSVGV ADE SETERIVLETER INSDELKNET PAVEL REGK KIT IT AMAVORYV 21646472 NL (11) ONE IFFC PE TDAFARK-VI NOTSVGV ADE SETERIVLETER INSDELKET PAVEL NOT GOGGAH GGC F CKD YK SSSSS Nhhhhhhhh Nhh Nhh 2005826 VSVORSEDDYENERE I KEP- VIPERW T-SATKEFINPT REFUGER COLITCE KIT DTY-GGAAR-GGC F CKD YK 18336994 LST ONESELDQKSLGEAVMEE I KEP- VIPSKW HKGTKFFINPT REFUGER COLITCE KIT DTY-GGAAR-GGC F CKD YK 2005826 VSVORDEDTIFERENTILAD IKK-VIPSKW HKGTKFFINPT REFUGER TO TUTEKIT DTY-GGAAR-GGC F CKD YK 200686 MSVG TQDVILEKUREK KWE UFSKW HKGTKFFINPT REFUGER TO TUTEKIT DTY-GGAAR-GGC F CKD YK 200686 MSVG TQDVILEVAREE I KPVIPSKW HKGTKFFINPT REFUGER TO TUTEKIT DTY-GGAAR-GGC F CKD YK 200686 MSVG TQDVILEVAREE I KPVIFSKW HKGTKFFINPT REFUGER TO TUTEKIT DTY-GGAAR-GGC F CKD YK 200686 MSVG TQDVILEVAREE I KWE UFKW MFY TY TABC TGG VVG CUTKKIT IN TTA GGAAR-GGC F CKD YK 200686 MSVG TQDVILEVAREE I KF VIFKW MFY KWE VFG GFVT TUTEKIT DTY-GGAAR-GGC F CKD YK 200686 MSVG TQDVILEVAREE I KWE LAKGY MFY KWE MABC TGG VVG	46015497	DM (8)) V <mark>ls</mark> a <mark>i</mark> (GKQ <mark>SPD</mark>	I NQG <mark>V</mark> I	DR(5)	Q <mark>GRGDQ</mark>	G <mark>LMF</mark> C	XAT NE IDV	/L <mark>MPA</mark> P	<mark>IT</mark> Y AH RI	VQ <mark>RQ</mark> AE	V RKN	-GTLP <mark>W</mark>	LRPDA	KSQV	TFQY	(5)VG	JDA <mark>v</mark>	I V
18836994 DV (8) VLNL GKQ PD NQGVDR (5) C GOC MES ATME DSYMPADHISBR.VEQOAK EKKSNS PEPFWIR PDA AGOTT RY (5) SADAYU 2016553 SI (8) VLVALEQO PD AAGVAV (6) VG COC MES ATME DSYMPADHISBR.VEQOAK EKKSNS PEPFWIR PDA AGOTT RY (5) SADAYU 20195828 GM (7) TLSC HAO PD AAGVAV (6) VG COC MES ATME TO THE PETMPIT HISBR.VENAL-CELEWR PROSSOUTLAY (6) DHUDS IV 13357974 DFTIISN NKQ VD AQSUNK (5) TH GOC VFS ACDE PQ/MPIT UV LSRE: LVKLTNMKSNELEWARPDOKSOUTLAY (6) DHUDS IV 1335428 VL (11) DCR GQC MD VDVER (5) PL NDTSFGV APLS TERLVLETERINGLAV VQAARAF	400245	S<mark>K</mark> (8) V <mark>lv</mark> a <mark>l</mark> e	EQQ <mark>SPD</mark>	I AQG <mark>V</mark> I	L(6)	IGAGDQ	G <mark>LMF</mark> C	YAT DE TEE	CMPLT	<mark>IV</mark> LAHK	NAKLAE	LRRN	-G TL P <mark>W</mark>	LRPDS	KTQV	T <mark>V</mark> QY	(10)R	NT <mark>T</mark> TH <mark>V</mark>	v
601653 SI (8) VLVA EQCOPELAACHV (6) VC GOCIMEO TDE PETMETTLILSHK NARLHKLRNGTLFVKPDCKSOVTEX (10) R HTVV 22095828 GK (7) TISC (HAC) PD AMGYDICGC GDOCIMEO TDE PETMETALULSKE LIVELTMAKS -NEIEHARPDCKSOVTEAX (6) DHOSIV 2357974 DP -TISNNNKCYDI ACSYNK (5) TC GOCGIVEC ACCE POYMPLTTVLAREL LKEIEKORNKEPVK LADDKSOVS DY (5) -V ETML 2500666 DP (2) -LSC NSC SE TMOAVCSNDEIG GOCIVE ACCE POYMPLTTVLAREL LKEIEKORNKEPVK LADDKSOVS DY (5) -V ETML 2500666 DP (2) -LSC NSC SE TMOAVCSNDEIG GOCIVE ACCE POYMPLTTVLAREL LKEIEKORNKEPVK LADDKSOVS DY (5) -V ETML 2500666 DP (2) -LSC NSC SE TMOAVCSNDEIG GOCIVE ACCE PETVE ACEO -SUVAQA VORAAFIIDCYPFTKHDKSOVS DY (14) VOCES U 21646472 NL (11) ONE KPG PEITDAFARK-VIE NDTSVGVC PLS TERLIVLETERTINSPA KERFPEAGEDVKIMCGRNGRLDIT AVAFVDRFT M Haa R P SSSS hhhhhhhhh hhh SSS SS SS hhhhhhhhhh	15836994	D V (8) V <mark>ln</mark> l <mark>i</mark> (GKQ <mark>SPD</mark>	I NQG <mark>V</mark> I	R(5)	Q <mark>G7</mark> GDQ	G <mark>LMF</mark> C	YATNE IDS	SY <mark>MPA</mark> A	<mark>IH</mark> ISHR	VEQ <mark>Q</mark> AK	IRKKSI	NSP <mark>l</mark> P <mark>W</mark>	L <mark>R</mark> PDA	KAQI	TLRY	(5)SA	IDA <mark>v</mark>	I V
22095828 Gd (1) TLSC HAC PD LANGYDIGE GD GI MF G YKE PEIMPLAKULSRE LVKLTMMYKS-NEIGARDCKSOVIJAK (6) DH VSTWI 13357974 DFTIISNUNKO YD LAOCYNK (5) IG GD GI VG YCE PEIMPLGKULSRE LVKLTMMYKS-NEIFYK DADKSOVS DY (6) VJETMI 250686 DF (2) -LSC NSO SEI NOAD ONDEI G GD GI VC YCE SEI DE DUPLGCIVADA VQRAARIIDYFFIKIDKSOVIJAK (4) VQCESVI 334428 VI (11) DCR GG CH VD VDER (5) PL NDTSFEVG APLS TERLVLATERFINSDE KNEIFY YGED	6016553	S I (8) V <mark>lv</mark> a <mark>l</mark> e	EQQSPE	I AAG <mark>V</mark> I	V(6)	V <mark>GZGD</mark> Q	G <mark>IMF</mark> C	YATDETEE	TMPLT	<mark>LI</mark> LSHK	NAELHK	LRRN	-G <mark>TL</mark> P <mark>W</mark>	VR PDS	KSQV	TIEY	(10)R	VTHV	W
13357974 DE-TIISNUNKC, VD. AQSWNK (5) I ENGOG, VF OCDE POYMETATVLARELLKETEKORNKEPVKI CADMKSOVSLDY (5) -VETTLL 250086 DF (2) - LSCINSO, SE NOAVQSNDE I ENGOG, VF OCDES POYMETASTVAQALVQRAARIIDOPPETKHDMKSOVVLNY (4) VQCESVJ 334428 VI (11) DCR CQC MD VDVERK (5) PL ND TSRVS - PLS, TERLVLETERFUNDE INNET PAGEDIVK/MGCRKKLTITIZAMAVD RRV VI (11) DCR CQC MD VDVERK (5) PL ND TSRVS - PLS, TERLVLETERFUNDE INNET PAGEDIVK/MGCRKKLTITIZAMAVD RRV 21646472 NL (11) DCR CQC MD VDVERK (5) PL ND TSRVS - PPSETERIVLATEOFINSPA KERPPEAGEDVKIMGCRNORKLUIT VAVAFD RPI maa R P SSSSS hhhhhhhhh hhh SSSSS hhhhhhhhh hh SSSSS hhhhhhhhh hh SSSSS hhhhhhhhh hh SSSSS bhhhhhhhh hh SSSSS bhhhhhhhh hh SSSSS SSS SSS SSS SSSS SSSS SSSS SSSS	22095828	G<mark>M</mark> (7) T <mark>ls</mark> C <mark>i</mark> f	IAQ <mark>SPD</mark>	I AMG <mark>V</mark> I	DI	G <mark>GAGDQ</mark>	G <mark>IMF</mark> C	GVRETEE	LMPLA	<mark>LV</mark> LSRE	LVK <mark>L</mark> TN	MMKS	-NEIE <mark>W</mark>	ARPDQ	KSQV	TLAY	(6)DH	I <mark>V</mark> DS	v
2500666 DF (2) -LSCINSC SE NOAVOSNDEIG/GDOCITVO ICSE POIMPLGSTVAQALVQRARE-I-I-DOYPETKHDMKSOVUNY (4) VQCESYJ 333428 VI (11) DCR LOGG MD VDVER (5) PI NDTSFOV APLS TERIVLATEOFINSDELKNEIPAGEDIKVMGCRORKATTICAMAVDDRV VI (11) DCR LOGG MD VDVER (5) PI NDTSFOV APLS TERIVLATEOFINSDELKNEIPAGEDIKVMGCRORKATTICAMAVDDRV NI (11) DCR LOGG MD VDVER (5) PI NDTSFOV APLS TERIVLATEOFINSDELKNEIPAGEDIKVMGCRORKATTICAMAVDDRV NI (11) DCR LOGG MD VDVER (5) PI NDTSFOV APLS TERIVLATEOFINSDELKNEIPAGEDIKVMGCRORKATCOTTAWAFVDRPI P	13357974	D <mark>F</mark> '	TI ISN<mark>V</mark>N	iko <mark>svd</mark>	IAQS <mark>V</mark> I	K(5)	IGAGDQ	GIVE	ACDETPC	YMPLT	<mark>TV</mark> LAHE	LKEIEK	QR <mark>RN</mark> -	-KEFVK	IQADM	KSQV	SIDY	(5)-V	∕ <mark>IET</mark> M	E L
<pre>3334428 VL (11) DCR.GQGGMD.VDVEER (5) PL NDTSFGV APLS TERLVLETERFLNSDE.KNEIFAVGED IKVMGLREGKTTTTAAAVDRYV 21646472 NL (11) DRE.KRG PELTDAFARK-VIT NUTS GV PFSETERIVLATEOFINSPA KERPFAGED VKIMGCRIGRELD TVAVAFVDER P</pre>	2500686	DF (2) – LSCIN	ISOSSE	INOAV	SNDE	ICAGDO	GITV	CSETEC	LMPLG	s <mark>ivao</mark> aj	VORAAR	I I	-DOYPF	IKHDM	KSOV	VLNY	(4) VC	CESV	TL.
21646472 NL (11) ONE TKPG* PEI TDAFARK-VITENDTS'GV PFSETERIVLATEOFLNSPALKERFPEAGEDVKIMGCRNGRKLOTVAVAFVDRFI Mathia R R P OS SS 46015497 LSTOPSEEDOKSLOEADMEE IKPILPAEW TSATKFF NPT RFVIGEM OC LTORKII DTV-GGMARH-GG FS KDLS-K D 15836994 LSTOPSEEDOKSLOEADMEE IKPVIPAEW TSATKFF NPT RFVIGEPM OC LTORKII DTV-GGMARH-GG FS KDLS-K D 15836994 LSTOPSEEDOKSLOEADMEE IKPVIPAEW DEDTIVILOPS RFVIGEPC OC LTORKII DTV-GGMARH-GG FS KDLS-K D 15836994 LSTOPSEEDOKSLOEADMEE IKPVIPAEW DEDTIVILOPS RFVIGEPC OC LTORKII DTV-GGMARH-GG FS KDLS-K D 209582 VSVDHDEDTKHDEIKKIVIEKEV-VIPASK DESTVFH NPCCTFIVGEM DATVTORKII DTV-GGMARH-GG FS KDLS-K D 209582 VSVDHDEDTHDEIEKIVIEKVY-VIPASK DESTVFH NPCCTFIVGEM DATVTORKII DTV-GGMARH-GG FS KDLS-K D 209582 VSVDHDEDTHDEIEKIVIEKVY-VITEYK NS-SDN IKYY NPT GRUNGPKOTTORKII DTV-GGMAH-GG FS KDLS-K D 2105666 MSVDTDDVSLDOLROTTNOVILPVIKYN S-SDN IKYY NPT GRUNGPKOTTORKII DTV-GGVAH-GG FS KDLTK D 21646472 PNANYFER-KSALRHELLSFIESQSCNI (5) DINS DDP (4) EGVY ITVT TSAEM CD GC QV RCNR NGI ITPFRENSME ACKN NYNH G a a A AA mRK PP 400245 RSAYAFWY KNIVAAGIADRCE QVSYA GVAE TSIMUETFOTEKV-PSEQITIVE REKINDRENGTUEKK (5) TAAYG (15) KY 18	3334428	VL (1)	, 1) d c r <mark>1</mark> (GOGSMD	LVDVF	R(5)	PLANDT	SF <mark>GV</mark>	VAPLSTTE	R LVL ETH	ER <mark>FL</mark> NSDE	KNEIPA	VGED-	-IKVMG	LREGK	KITI	TIAM	AV		v
Main Date P OR R P 46015497 ISSO S hhhhhhhhh hhh SSS SS	21646472	NL (1)	1) ONE IF	PGSPE	LTDAF	ARK-V	ICANDT	s <mark>vgv</mark>	PFSETE	RIVLATE	eo <mark>fl</mark> nspai	KERFPE	AGED-	-VKIMG	CRNGR	KLOI	TVAV	AF		'I
P R		- ·	· ~	Mm			maa				~				R	P				
46015497 SSSS hhhhhhhh hhh SSS SS hh 46015497 LSTORSEELDQKSLQEAVMEETIKPL-PARW T-SATKFF NPT RFVIGGPMCDCLTRKII DTY-GGMARH-GGG F GKD S-K D 400245 ISVQ HDEEVCLDEMRDALKEKVIKAVPARY D-EDCTIYH.0PS RFVIGGPODALTRKII DTY-GGMARH-GGG F GKD S-K D 15836994 LSTORDFGKQKDLVEAVREETIKPUPASU D-ESTYFH NPT KFVIGGPKODA VT.RKII DTY-GGMARH-GGG F GKD S-K D 22095828 VSVQ HDEDTHHDETEKIVIEKVVRETULEAV. NSVG HDEDTHHDETEKIVIEKVKS-TIMDQZAIKYG N-TDFKKI NSS RFVIGGPLOT VT.RKII DTY-GGWGAH-GGG F GKD S-K D 23357974 VSIG HDEDYNIEYFNKKVS-TIMDQZAIKYG N-TDFKKI NSS RFVIGGPLOT VT.RKII DTY-GGYFKH-GGG F GKD F-K D 334428 KNIEEYKEV-IEKVRKV-TEDLAKKA (7) HINTADDYERFKKI NSS RFVIGGPLOT ST.RKII DTY-GGYGHN-GGG F GKD F-K D 334428 KNIEEYKEV-IEKVRKV-EDLAKKA (7) HINTADDYERESVYI TVT TSAEM DD SVCRGNR NGL IFPFRM-SME A-GKNEVNH G a a a AA a mK P 40015497 RSAAYA RYYAKN VAAGLADRCE QVSYA GVAE TSIMFETFGTEKV-PSEQ TILVREFFDLRYG IQMLDLLH (5) TAYY (23) LK K M M 12646472 PNANHYFER-KSALRHELLSF ESQSCNI (5) DINS DDP (4) EGYI TVL TSAEGG QV RGNR NGL IAFNRNOTHE AAG RWEVNH G M M M M M M M M M											•د	*				-				_
hhhhhhhhhh hhh sss ss hhhhhhhhhh hhh sss ss hhhhhhhhhh				a			P				Ô	8	άţ					5	െ	2
46015497 LSCCHSEEIQKSLCEAWMEEIKPLLPAEW TSATKFFINPTERFVIGGPM DC LITERKII DTY-GGMAH-GGGF GKD[S-K D 15836994 LSCCHDEGKQALVEAVEEILKPVLPAKY DEDTIYHINPTERFVIGGPV DC LTERKII DTY-GGMAH-GGGF GKD[S-K D 15836994 LSCCHDEGKQALVEAVEEILKPVLPAKY DEDTYFHINPTERFVIGGPV DC LTERKII DTY-GGMAH-GGGF GKD[S-K D 22095828 VSVCHDEDTHDELEKIVIEKVKPVLPASL DESTYFHINPTERFVIGGPH DA VTGRKII DTY-GGWAH-GGGF GKD[S-K D 2305774 VSTCHSEDYTTEKLRETILADVIKKVIPASL DESTYFHINPTERFVIGGPH DA VTGRKII DTY-GGWAH-GGGF GKD[S-K D 2500686 MSVQHDDDTHDELEKIVIEKVKPVLEKYN S-SDNIKYYINPTGFVIGPH DT VTGRKII DTY-GGWAH-GGGF GKD[S-K D 23334428 KNIEEYKEV-IEKVKKV-EDLAKKIA (7) HINTADDYERESVYITYTTSAEM DD SVCKORN NGLITPFRPM-SMEASCKNFVNH G a a AA a mRK PP A hhhhhhhhhhhhhh sssss ss ss hhhhhhhh hhh 46015497 RSAAYARWYIKNIVAGLADRCEIQ'SYA GVAEFTSIMVETFGTEKV-PSEQLTLIVREFFDLRPYGIQMIDLIH (5) TAYC (23) LK RSAAYARWYIKNIVAGLADRCEIQ'SYA GVAEFTSIMVETFGTEKV-PSEQLTLIVREFFDLRPYGIIQMIDLIH (5) TAYC (15) KY 253694 RSAAYARWYIKNIVAGLADRCEIQ'SYA GVAEFISISIFHYGTSC-SERELLEIVKNFFDLRPYGIIKDLIH (5) TAYC (15) KY 253694 RSAAYARWYIKNIVAGLADRCEIQ'SYA GVAEFISISIFHYGTSCK-SERELLEIVKNFFDLRPYGIIKDLIH (5) TAAYC (23) LK RSAAYARWYIKNIVAGLADRCEIQ'SYA GVAEFISISIFHYGTSCK-SERELLEIVKNFFDLRPYGIIKDLIH (5) TAAYC (15) KY 22095828 RSAAYARWYIKNIVAGLADRCEIQ'SYA GVAEFISISIFHYGTSCK-SERELLEIVKNFFDLRPYGIIKDLIH (5) TAAYC (15) KY 22095828 RSAAYARWYIKNIVAGLADRCEIQ'SYA GVAEFISISITFYGTSK-SENELLEIVKNFFDLRPYGIIKDLIH (5) TAAYC (15) KY 22095828 RSAAYARWYIKNIVAGLADRCEIQ'SYA GVAEFISISITFYGTSK-SENELLEIVKNFFDLRPYGIIKDLIH (5) TAAYC (15) KY 22095828 RSAAYARWYIKNIVAGLADRCEIQ'SYA GVAEFISISITFYGTSK-SENELLEIVKNFFDLRPYGIIKDLIH (5) TAAYC (15) KY 22095828 RSAAYARWYIKNIVAGLADRCEIQ'SYA GVPKVVKN DTFGTSKI-SENEKLERYSFDIKTFINDLEUKT (5) TAAYC (24) EL 13357974 RSAAYARWYIKNIVAAGLANCCEIQ'SYA GVPKVVSKNDTFGTSKV-DENKISFACKNFFLSPCI TYGZIKTIKLDUG (5) YATYC (23) SK 2334428 KIYMIFRWIAKNINYAAKLANCCEIQ'AFA GE-PQUVSKNN TFKTNLV-DEKKIFAANKSF				a			P				230	238	244 245					265	269	Ż
400245 ISVCHDERVCLDEMRDALKEKVIKAVVPAKY DEDTIYH.OPSERFVIGEPCDALTCRKII.DTY-GGWCAH-GG,FSGKDYT-K D 15836994 ISTCHDPGKQKDLVEAVREEILKPVLPSKW HKGTKFHINPTGKFVIGGPVDCLTCRKII.DTY-GGWCAH-GG,FSGKDY-K D 22095828 VSVCHDEDTHDETEKIVIEKVVKPVLESKU HKGTKFHINPTGKFVIGGPH DACVTCRKII.DTY-GGWCAH-GG,FSGKDY-K D 22095828 VSVCHDEDTHDETEKIVIEKVVKPVLEKYI S-SDNIKYYINPTGRFVIGGPH DTCVTCRKII.DTY-GGWCAH-GG,FSGKDY-K D 2305774 VSICHDEDYNIEYFNKKVS-TIMDCIAIKYG NTDFKKIINSSERFVIGGPI.DTCLTCRKII.DTY-GGYGKH-GG,FSGKDY-K D 2500686 MSVCHDEVSLDQLRQTINQVILPVLTHYG N-DPKIKHLVNFGSFVUGGPH DTCLTCRKII.DTY-GFYANH-GGGSFSGKDY-K D 2334428 KNIEYYKSV-IEKVRKV-EDLAKKIA(7)HINTADDYERESVYITVTGSAEM DDCSVCRGNR NGLITFFRPM-SME ASGKNEVNH G 21646472 PNAN YFER-KSALRHEILSFIESQSCNI(5)DINS DDP(4)EGVYITVLGTSAEM DDCSVCRGNR NGLITFFRPM-SME ASGKNEVNH G a a AA a mRK PP E K K hhhhhhhhhhhhhhh ssssss ss hhhhhhhh hhh 46015497 RSAXARYVAKNIVAACIADRCEICYSYA GVAEFTSIMVETFGTEKV-PSECITLUVREFFDLRPYGIQMUDLIK(5)TAAYC (23)LK 15836994 RSAXARYVAKNIVAACIADRCEICYSYA GVSFELSISITFFYGTSV-SERELEIVKKNFDLRPGUVRDLDLKK(5)TAAYC (15)KY 15836994 RSAXARWVAKSIVKGGLCRRVLVQISYA GVSFELSISITFFYGTSV-SERELEIVKKNFDLRPGUVRDLDLKK(5)TAAYC (15)KY 22095828 RSAXARWVAKSIVKGGLCRRVLVQISYA GVAEFISISVTFFGTEKV-PSECITLUVKKFDLRPGVIVRDLDLKK(5)TAAYC (15)KY 22095828 RSAXARWVAKSIVKGGLCRCVLVQISYA GVAFFISISVTFFGTCKI-SDDKIEKLERQHPDLRPYGINKLDLIH(5)SASYC (39)KL 6016553 RSAXARWVAKSIVKGGLAKRVLVQISYA GVAFFISISVTFFGTSKV-DEDKISEXVKVFDLSPRGIEKALELRE(7)IAAFC (24)EK 22095828 RSAXARWVAKSIVKAGLAKRVLVQISYA GVPFPVSVKVDTFGTSKV-DEDKISEXVVFDLSPRGIEKALELRE(7)IAAFC (24)EK 13357974 RSASYARWVAKSIVAACLADRCEICISYA GVPFPVSVKVDTFGTSKV-DEDKISEXVVFDLSPRGIEKALELRE(7)IAAFC (24)EK 22095828 RSAXARWVAKSIVAACLADRCEICISYA GVPFPVSVKVDTFGTSKV-DEDKISEXVVFDLSPRGIEKALELRE(7)IAAFC (24)EK 13357974 RSASYFRWINKWVAACLADRCEICISYA GVPFPVSVKVDTFGTSKV-DEDKISEXVFVDLSPRGIEKALELRE(7)IAAFC (24)EK 13357974 RSASYFRWINKWVAACLADRCEICISYA GVPFPVSVKVDTFGTSV-DEKKI		SSSS	s ł	a hhhhhh	hhhh	_	P	hhh		SSS	230	3 38 3	245* 245*	_				265	269	27. h
15836994 LSTQHDPGIKQKDLVEAVREEILKPVIPSKW HKGTKFHINPTCKFVIGGFVODC LTERKII DTY-GGSARH-GGC FSCKDFS-K D 6016553 VSTQHDPDIKLKLRETILADVIKKVIPASL DESTVFHINPCCTFIVGGPMCDACVTCRKII DTY-GGWGAH-GGC FSCKDFT-K D 13357974 VSIQHDEDIKNUKKVS-TIMQUIEKYINS-SDNIKYINPTCRFVIGGPHCDT VTERKII DTY-GGWGAH-GGC FSCKDFT-K D 2500686 MSVQHTQDVSLDQLRQTIINQVILPVITEYG N-DPKIKHLVNPGCSFVVGGPMADT LTERKII DTY-GGYFKH-GGC FSCKDFT-K D 2500686 MSVQHTQDVSLDQLRQTIINQVILPVITEYG N-DPKIKHLVNPGCSFVVGGPMADT LTERKII DTY-GGYFKH-GGC FSCKDFT-K D 2500686 MSVQHTQDVSLDQLRQTIINQVILPVITEYG N-DPKIKHLVNPGCSFVVGGPMADT LTERKII DTY-GGYFKH-GGSFSCKDFT-K D 2500686 MSVQHTQDVSLDQLRQTIINQVILPVITEYG N-DPKIKHLVNPGCSFVVGGPMADT LTERKII DTY-GGYFKH-GGSFSCKDFT-K D 2500686 MSVQHTQDVSLDQLRQTIINQVILPVITEYG N-DPKIKHLVNFGSFVVGGPMADT LTERKII DTY-GGYFKH-GG FSCKDFT-K D 2500686 MSVQHTQDVSLDQLRQTIINQVILPVITEYG N-DPKIKHLVNFGSFVVGGPMADT LTERKII DTY-GGYFKH-GG FSCKDFT-K D 2500686 MSVQHTQDVSLDQLRQTIINQVILPVITEYG N-DPKIKHLVNFGSFVVGGPMADT LTERKII DTY-GYFKH-GG FSCKDFT-K D 21646472 PNANHYFE-KSALRHELLSFIESQSCNI (5) DINS DDP (4) EGVIITVLGTSAEG DG GV GR NGL ITFFRMV ME AA KNEVNH G a a AA a mRK P E M Y hhhhhhhhhhhhhh ssssss sss hhhhhhhhh hh 46015497 RSAAYARWVAKSIVKGGLCRRVLVQSYA GVAEFTSIMVETFGTEKV-PSEQ TILVREFFDLRPYGI IQMLDLH (5) TAAYG (23) LK 400245 RSAAYARWVAKSIVKGGLCRRVLVQSYA GVAEFISISIFHYGTSGK-SERELEIVKNFDLRPGV VRDLDLKK (5) TAAYG (15) KY 15836994 RSASYAWYYZKNVAAGLADRCEVQYSA GVAEFISISVTFFGTGKI-SDDKIEKLIRQHFDLRPYGI IKMLDLIH (5) SASYG (39) KL 2005828 RSAAYARWVAKSIVKGLARRVLVQISYA GVAEFISVLVAFGTSPL-TDELHQIVEDSFDLTPGKI IKMLDLIH (5) SASYG (39) KL 2005828 RSAAYARWVAKNTVAAELADKCE IQSAA GVAEFISVLVAFGTSPL-DELKISASVKVDDLSPRGT EKALELRK (7) TAENG (15) KT 2334428 KIYNII ARWVAKNTVAAELADKCE IQSAA GVPEVSVKVDTFGTSKV-DEDKISEASKVFDLSPRGT EKALELRK (5) TAENG (15) KT 2334428 KIYNII ARWVAKNTVAAELADKCE IQSAA GVPEVSVKVDTFGTSV-DIALKEFSTDIKTFINDLELWT (5) VATYG (23) SK 2500686 RTGAYFARFIAKH	46015497	ssss L <mark>ST</mark> Q	s i HSE <mark>EI</mark> DÇ	a hhhhhi 2 <mark>KSLQE</mark>	hhhh A <mark>V</mark> ME <mark>E</mark>	I <mark>I</mark> KP-	P - <mark>IL</mark> :	hhh PAEW	TS A	sss -TK <mark>FFINI</mark>	ÖR RFVIG	SD <mark>D M</mark> GD	з 24544 24544	I <mark>I V</mark> DTY	- GG<mark>M</mark>AI	R <mark>H-G</mark>	GG F	a <mark>3285</mark> 765	S-К	н 12 h
<pre>6016553 VSTOHSEDVTLEKLRET LAD/TKKVIPASL DESTVHINPCCFFUGGPM DACVTCRKII DTY-GGVGAH-GGC FSCKDFT-K D 22095828 VSVOHDEDTHDEIEKIVIEKVVKPVIEKN S-SDNIKYYINPTCRFVIGGPLDTSUTCRKII DTY-GGVGAH-GGC FSCKDFS-K D 2500686 MSVOHTODVSLDQLRQTINQVILP-VITEYG N-DFKIKHLVNPGCSFVUGGPMADTSITCRKII DTY-GGYGHH-GGCFFSCKDFT-K D 3334428 KNIEEYKEV-IEKVKKV-ED AKK-IA (7) HINTADDYERESYYITTTTSAEM DD SVCRGNR NGLITPFRPM-SME ASCKNPVNH G 21646472 PNANHYFER-KSALRHEILSFESQSCNI (5) DINS DDP (4) EGYYITVLGTSAEM DD SVCRGNR NGL-IAFNRNOTME AACKNPVNH G a a AA a mRK PP</pre>	46015497 400245	ssss L <mark>ST</mark> Q I SV Q	s ł HSE EI DÇ HDE EV CI	a nhhhhi 2KSLQE DEMRD	hhhh A <mark>V</mark> MEE ALKEK	<mark>II</mark> KP- /IKA-	P - IL : - <mark>VV</mark> :	hhh PAEW PAKY	TSA DED	sss -TK <mark>FFI</mark> NI -TI <mark>YHL</mark> QI	ÔC PTORFVIG PSORFVIG	s SP <mark>M DC</mark> SPQ DA	rz P LCB K S 2454 2454 2454 2454	I IVDTY I IVDTY	- GG<mark>M</mark>AI -GGWG2	R <mark>H-G</mark> A H-G	GGAF GGAF	SCKD5 SCKD7 SCKD7	596 S-K T-K	<mark>в в</mark> ч 27'
<pre>22095828 VSVQ DEDTTHDELEKU VEKVVRPVLEKYN S-SDNIKYYINPT RFVIGGFHCDT VTCRKII DTY-GGYFKH-GGGF CKD S-K D 13357974 VSIOHDEDYNIEYFNKKVS-TIMDQIAIKYG NTDFKKIINSS RFVIGGFLOTT LTCRKII DTY-GGYGH-GGGF CKD FT-K D 2500686 MSVQHTQDVSLQLQTINQVILPVLTEYG N-DPKIKHLVNPG SFVVGGFMADT LTCRKII DTY-GGYANH-GGGF CKD FT-K D 3334428 KNIEEYKEV-EKVRKV-EDLAKKIA (7) HINTADDYERESVYITVT TSAEGOG OV CRGNR NGL TPFRPM-SMEAACKNEVNH G 21646472 PNANHYFER-KSALRHEILSFIESQSCNI (5) DINS DDP (4) EGVYITVL TSAEGOG OV CRGNR NGL -IAFNRNQTME AACKNEVNH G a a AA a mRK PP E E V M hhhhhhhhhhhhhh ssssss sss hhhhhhhh hh 46015497 RSAAYAARVYAKN VAAGLADRCEIQ SYA GVAE TSIMVETFGTEKV-PSEQ TLLVREFFDLRPYGLIQMLDLH (5) TAAY (23) LK 400245 RSAAYAARVYAKN VAAGLADRCEIQ SYA GVAE TSIMVETFGTEKV-PSEQ TLLVREFFDLRPYGLIQMLDLK (5) TAAY (23) LK 5836994 RSASYAARWYAKSLVKGGLCRRVLVQ SYA GVAE TSIMVETFGTGKI-SDDKLEKLIRQHFDLRPYGIIKMLDLIH (5) SASY (39) KL 22095828 RSAAYAARWYAKSLVKGGLARRCEVQ SYA GVAE TSISVTFFGTGKI-SDDKLEKLIRQHFDLRPYGIIKMLDLIH (5) SASY (39) KL 33377774 RSAYARWYAKSLVKGGLCRRVLVQ SYA GVPF VSVKVDTFGTSFL-DEELHQIVEDSFDLTPGKI KELDLKR (5) TAAY (24) EL 3357774 RSASYARWYAKSLVKGGLCRRVLVQ SYA GVAE TSISVTFFGTGKI-SDDKLEKLIRQHFDLRPYGI IKMLDLIH (5) SASY (39) KL 3337774 RSASYARWYAKSLVKGGLCRVLVQ SYA GVPF VSVKVDTFGTSFL-DEELHQIVEDSFDLTPGKI KELDLKR (5) TAAY (24) EL 3357974 RSASYARWYAKSLVKGGLCRVLVQ SYA GVPF VSVKVDTFGTSFL-DEELHQIVEDSFDLTPGKI KELDLKR (5) TAAY (24) EL 3357974 RSASYFARWIAKNVVAAELADKCEIQI SYA GVPF VSVKVDTFGTSFLV-DEKK SEAVSKVFDLSPRGT EKALELRE (7) LAAFG (24) EL 3357974 RSASYFARWIAKNVVAAELADKCEIQI SYA GV-PF VSVKVDTFGTSFLV-DEKK SEAVSKVFDLSPRGT EKALELRE (7) LAAFG (24) EL 33428 KIYNIEARWYAKNIVAAELADKCEIQI SYA GV-PK VSVKVDTFGTSKV-DEDKX SEAVSKVFDLSPRGT EKALELRE (7) LAAFG (24) EL 33428 KIYNIFARFIAKHVSLGWAEECEVS SWVFSQ-PLE OSIQVKCFNINKEFSEQL NQVISQYFNWSVAKIAKLKLLD (6) YAYY (22) KN 334428 KIYNILINLINNDIAKLEGVKECYNR LSQ GREIME KALDIE ITTEDSY-DIKD EPKAKEIANKWLDANIKEYQKMTVE</pre>	46015497 400245 15836994	ssss L STQ I SVQ L ST Q	s h HSEEldg HDEEVCI HDPG <mark>I</mark> KG	a hhhhhi 2KSLQE DEMRD 2KDLVE	hhhh A <mark>V</mark> MEE ALKE K AVREE	I <mark>IK</mark> P- /IKA- I <mark>L</mark> KP-	P - IL : - VV : - VL :	hhh PAEW PAKY PSKW	TS A DED H K G	sss TK FFINI TI YHL QI TK <mark>FHI</mark> NI	Ö PT GRFVIG PS GRFVIG PT GKFVIG	s SP <mark>McDc</mark> SPQ cDA SPQ cD A	S LTCRK: LTCRK: LTCRK:	I IVDTY I IVDTY I IVDTY	-GG <mark>M</mark> AI -GGWG2 -GGSAI	R H-G A H-G R H-G	GGAF GGAF GGAF	SCKDP SCKDY SCKDP	697 S-K T-K S-K	оо и 27'
<pre>13357974 VSIQHDEDYNIEYFNKKVS-TIMDQIAIKYG NTDFKKIINSS RFVIGGPI DTGLTGKII DTY-GGIGHH-GGG F5 CKD T-K D 2500686 MSVQHTQDVSLDQLRQTIINQVILPVLTEYG N-DFKIKHLVNPG SFVVGGPMADTGLTGKII DTY-GPYANH-GGSF5 CKD T-K D 334428 KNIEYKEV-IEKVRKKV-EDIAKKIA (7) HINTADDYERESVYLTVT TSAEMCDDSVCRGNR NGLITPFRPM-SMELASGKNFVNH G 21646472 PNANHYFER-KSALRHELLSFIESQSCNI (5) DINS DDP (4) EGVY TVL TSAEG DG QV RGNR NGL IAFNRNQTME AACKNFVNH G a a AA a mRK PP E E V M hhhhhhhhhhhhhh sssss ss ss hhhhhhhhh hhh 46015497 RSAAYALRYVIKNTVAAGLADRCEIQVSYA GVAETSIMVETFGTEKV-PSEQLTLLVREFFDLRPYGLIQMLDLLH (5) TAAYG (23) LK 400245 RSAAYALRYVIKNTVAAGLADRCEIQVSYA GVSHCLSISIFHYGTSQK-SERELLEIVKKNFDLRPGV.VRDLDLKK (5) TAAYG (23) LK 15836994 RSASYALRYVIKNTVAAGLADRCEIQVSYA GVSHCLSISIFHYGTSQK-SERELLEIVKKNFDLRPGV.VRDLDLKK (5) TAAYG (23) KL 22095828 RSAAYALRWVAKSUKGGLCRRVLVQUSYA GUAETSISVTFGTGKI-SDDKIEKLIRQHFDLRPYGIIKMLDLH (5) TAAYG (24) EL 3357974 RSASYALRWVAKSUKAGLARKVLVQLSYA GUPKYVSVKVDTFGTSKV-DEDKISENVKVFDLRPGKIKELDLKR (5) TAAYG (24) EL 3357974 RSASYFARWIAKNIVAAGLADRCEIQUSYA GVPKYVSVKVDTFGTSKV-DEDKISENVFDLSPRGIEKALELRE (7) LAAFG (24) EL 3357974 RSASYFARWIAKNIVAAGLADRCEIQUSYA GUPKYVSVKVDTFGTSKV-DEDKISENVFDLSPRGIEKALELRE (7) LAAFG (24) EL 3357974 RSASYFARWIAKNIVAAKLAQCEIQLAFA GE-POQ VAMFINTFKTNLV-DEKKIFEAINKSFNFDIKTFINDLEUWT (5) VATYG (23) SK 2500686 RTGAYFARFINKHIVSLGWAECEVSISWVFSQPL; OSIQVKCFNINKEFSEQLINQVISQYFNWSVAKI IAKLKLLD (6) YAVYG (22) KN 334428 KIYNILENLINDTAKLEGVKECYURILSQ_GKYFNIKKALDIGTTGTGTY-DIKDIEFKAKEIANKWLDNIMEVQKMIVEG KVT-TF</pre>	46015497 400245 15836994 6016553	ssss L <mark>ST</mark> Q I SV Q L ST Q V ST Q	s ł HSEEIDO HDEEVCI HDPGIKO HSPDVII	a hhhhh DKSLQE DEMRD DKDLVE LEKLRE	hhhh AVMEE ALKEK AVREE T <mark>I</mark> LAD	IIKP- /IKA- ILKP- /IKK-	P -IL: -VV: -VL: -VL:	hhh PAEW PAKY PSKW PASL	TSA DED HKG DES	SSS TK FFINI TI YHL QI TK FHINI TV <mark>FHI</mark> NI	ÕR PTGRFVIG PSGRFVIG PTGKFVIG PCGTFIVG	SP <mark>MCDC</mark> SPQCDAC SPQCDC SP <mark>VCDC</mark> SP <mark>VCDC</mark>	S S LT RK: LT RK: LT RK: VT RK:	I IVDTY I IVDTY I IVDTY I IVDTY I IVDTY	- GGM AI - GGW GJ - GG SAI - GG<mark>W</mark>GJ	RH-G AH-G RH-G AH-G	GGAF GGAF GGAF GGAF	502 SCKDP SCKDP SCKDP SCKDP	S-K T-K S-K T-K	в в в в ч 27'
<pre>2500686 MSVQHTQDVSLDQLRQTIINQVLPVLTEYG N-DPKIKHLVNPG SFVVGGPMADTCLTEKII DTY-GPYANH-CGGSF CKDET-K D 3334428 KNIEYKEV-IEKVRKKV-EDLAKKLA (7) HINTADDYERESVYLTVT TSAEM DDGSV GRNR NGL ITPFRPM-SME A CKNEVNH G 21646472 PNANHYFER-KSALRHELLSFIESQSCNI (5) DINS DDP (4) EGVYLTVL TSAEG.DG QC V RGNR NGL IAFNRNQTME A CKNEVNH G a a AA a mRK PP</pre>	46015497 400245 15836994 6016553 22095828	ssss L st o I sv o L st o V st o V st o	s h HSE EI DG HDE EV CI HDPGIKG HSP DV TI HDE DT TH	a hhhhh KSLQE DEMRD KDLVE EKLRE IDE <mark>IEK</mark>	hhhh A <mark>V</mark> MEE ALKEK AVREE TILAD I <mark>V</mark> IEK	IIKP- /IKA- ILKP- /IKK- /VKP-	P -IL: -VV: -VL: -VI: -VI:	hhh PAEW PAKY PSKW PASL EKYN	TSA DED HKG DES S-SDN	sss TK FFINI TI YHLQI TK FHINI TV FHINI IK <mark>YYI</mark> NI	PTGRFVIG PSGRFVIG PTGKFVIG PCGTFIVG PTGRFVIG	SPMCDC SPQCDA SPQCDC SPMCDA SPMCDA SP <mark>HCD</mark> T	*** S LTCRK: LTCRK: LTCRK: VTCRK: VTCRK:	II DTY II DTY II DTY II DTY II DTY II DTY	- GGM AH - GGW GA - GG SAH - GGW GA - GG<mark>W</mark>FH	RH-G AH-G RH-G AH-G KH-G	GGAF GGAF GGAF GGAF	SCKD F SCKD F SCKD F SCKD F SCKD F	S-K T-K S-K S-K S-K	0000 400 427
3334428 KNIEEYKEV-IEKVRKKV-EDLAKKIA (7) HINTADDYERESVIITVTGTSAEM DDCSV EGNR NGLITPFRPM-SME AS CKNEVNH G 21646472 PNANHYFER-KSALRHELLSFIESQSCNI (5) DINS DDP (4) EGVYLTVLGTSAEG DGCOVERON NGL-IAFNRNOTME AACKNEVNH G a a a Aa a mRK PP E E M E k + p M hhhhhhhhhhhhhh ssssss sss hhhhhhhhh hhh hhhhhhhhhhhhhhh ssssss sss hhhhhhhhh	46015497 400245 15836994 6016553 22095828 13357974	SSSS LSTQ ISVQ LSTQ VSTQ VSVQ VSLQ	s h SEEIDG HDEEVCI HDEGIKG HSPDVTI HDEDTTH HDEDT	a hhhhh KSLQE DEMRD KDLVE EKLRE IDE IEK IDE IEK	hhhh AVMEE ALKEK AVREE TILAD IVIEK K <mark>V</mark> S-T	IIKP- /IKA- ILKP- /IKK- /VKP- IMDQ-	P -IL -VV -VL -VL -VL -IA	hhh PAEW PAKY PSKW PASL EKYN IKYG	TSA DED HKG DES S-SDN NTD	SSS TK FFIN I TI YHL QI TK FHINI TV FHINI IK <mark>YYINI</mark> FKK <mark>IIN</mark> S	PTGRFVIG PSGRFVIG PTGKFVIG PTGRFVIG PTGRFVIG SSGRFVIG	S SPMCDCC SPQCDAC SPVCDCC SPMCDAC SPHCDTC SPHCDTC SPLCDTC	S LTCRK: LTCRK: LTCRK: VTCRK: VTCRK: LTCRK:	IIDTY IIDTY IIDTY IIDTY IIDTY IIDTY IIDTY	- GGM AI - GGW GJ - GG SAI - GGW GJ - GGY FI - GGI GI	RH-G AH-G RH-G AH-G KH-G HH-G	GGAF GGAF GGAF GGAF GGAF GGAF	SCKD P SCKD Y SCKD P SCKD P SCKD P SCKD P	S-K T-K S-K S-K S-K T-K	0000000000000000000000000000000000000
21646472 PNANHYFER-KSALRHELLSFTESQSCNT (5) DINS DDP (4) EGVYLTVLTSAEG DG QV F GNR NGL-IAFNRNQTME AACKNEVNH G a a AA a mRK PP E E M k p M hhhhhhhhhhhhhh ssssss sss hhhhhhhh hh 46015497 RSAYAERYVAKNIVAAGLADRCE QVSYA GVAE TSIMVETFGTEKV-PSEQLTLLVREFFDLRPYG IQMLDLLH (5) TAAY G (23) LK 400245 RSAAYAERYVAKNIVAAGLADRCE QVSYA GVAE TSIMVETFGTEKV-PSEQLTLLVREFFDLRPYG IQMLDLLK (5) TAAY G (23) LK 400245 RSAAYAERYVAKNIVAAGLADRCE QVSYA GVAE TSIMVETFGTEKV-PSEQLTLLVREFFDLRPYG IQMLDLLK (5) TAAY G (23) LK 400245 RSAAYAERWVAKSUKGGLCRRVLVQVSYA GVAE TSIMVETFGTGKI-SDDKIEKLIRQHFDLRPYG IXMLDLIH (5) SASY G (39) KI 5836694 RSASYAERWVAKSUKGGLCRRVLVQLSYA GVAE TSISVTTFGTGKI-SDDKIEKLIRQHFDLRPYG IXMLDLIH (5) SASY G (39) KI 6016553 RSAAYAERWVAKSUKAGLARRVLVQLSYA GVAE TSISVTTFGTGKI-SDDKIEKLIRQHFDLRPYG IXMLDLIH (5) TAAY G (23) K 22095828 RSAAYAERWVAKSUKAGLARCE QUSYA GVPK-VSVKVDTFGTSKV-DEELKJEXSKVFDLSPRG IEKALELDEK (7) LAAF G (24) EL 33357974 RSASYFARWIAKNVAAGLADCCE QLSYA GVPK-VSVKVDTFGTSKV-DEDKISEAVSKVFDLSPRG IEKALELRE (7) LAAF G (24) EL 23357974 RSASYFARWIAKNVAAKLAKQCE OLSYA G-PL QSIQVKCFNINKEFSEQLINQVISQYFNWSVAKI IAKLKLLD (6) YAVY G (22) KN 334428 KIYNILNLINDTAKLEGVKECYVRILSQ G KPINE KALDIE ITTEDSY-DIKDIEFKAKE IANKWLDNIMEVQKMIVE KVT-TF	46015497 400245 15836994 6016553 22095828 13357974 2500686	SSSS LS <mark>T</mark> Q IS V Q LSTQ VSTQ VS V Q VS V Q MS V Q	S P HSEELDG HDEEVCI HDPGIKG SPDVTI HDEDTTH HDEDYNJ HTQDVSI	a hhhhhh KSLQE DEMRD KDLVE EKLRE IDE IEK EYFNK DQLRQ	hhhh AVMEE ALKEK TILAD IVIEK KVS-T TILNQ	IIKP- /IKA- /IKA- /IKK- /IKK- /VKP- IMDQ- /ILP-	P -IL -VL -VL -VL -IA -VL	hhh PAEW PAKY PSKW PASL EKYN IKYG TEYG	TSA DED HKG DES S-SDN NTD N-DPK	SSS TK FFI NI TI YHL QI TK FHINI TV FHINI IK <mark>YYI</mark> NI FKK IIN IK <mark>HLV</mark> NI	PT RFVIG PS RFVIG PT KFVIG PT KFVIG PT RFVIG SS RFVIG PG SFVVG	SPMCDCC SPQCDAC SPQCDAC SPMCDAC SPMCDAC SPHCDTC SPICDTC SPMADTC	S LTCRK: LTCRK: LTCRK: VTCRK: VTCRK: LTCRK: LTCRK:	IIDTY IIDTY IIDTY IIDTY IIDTY IIDTY IIDTY IIDTY	-GGMAI -GGWGA -GGSAI -GGWGA -GGYFI -GGIGI -G GY AI	RH-G AH-G AH-G AH-G KH-G HH-G NH-G	GGAF GGAF GGAF GGAF GGAF GGAF	SCKD P SCKD P SCKD P SCKD P SCKD P SCKD P SCKD P	S-K T-K S-K T-K T-K T-K	0000000000000000000000000000000000000
a a AA a mRK PP K p M hhhhhhhhhhhhhh ssssss sss hhhhhhhh hhh 46015497 RSAAYARWYAKIVAACIADRCE OVSYA GV-AE TSIMVETFGTEKV-PSEQITLIVREFFDLRPYG LQMLDLLH (5) TAYG (23) LK 400245 RSAAYARWYAKSIVKGGLCRRVLVOVSYA GV-AE TSIMVETFGTEKV-PSEQITLLVREFFDLRPYG LQMLDLLH (5) TAYG (23) LK 400245 RSAAYARWYAKSIVKGGLCRRVLVOVSYA GV-SH LSISIFHYGTSQK-SERELLE IVKKNFDLRPGVTVRDLDLKK (5) TAYG (15) KY 15836994 RSASYARWYAKSIVKGGLCRRVLVOVSYA GV-SH LSISIFHYGTSQK-SERELLE IVKKNFDLRPGVTVRDLDLKK (5) TAYG (15) KY 15836994 RSASYARWYAKSIVKGGLCRRVLVOSYA GV-SH LSISIFHYGTSQL-TDEELHQIVEDSFDLTPGK IKHLDLH (5) SASY (39) KL 6016553 RSAAYARWYAKSIVKGGLCRRVLVOISYA GVAE ISISVTTFGTGKI-SDDKIEKLIRQHFDLRPYG IKMLDLH (5) TAYG (15) KT 22095828 RSAAYARWYAKSIVKACIAKRUVOISYA GV-PK VSVKVDTFGTSKV-DEELK ISAVSKVFDLSPRG IEKALEIDEK (7) LAAF (24) EI 13357974 RSASYFARWIKNIVAAELAACCET OLSYA GVPK VSVKVDTFGTSKV-DEDK ISEAVSKVFD ISFT INDLELWT (5) VATYG (23) SK 2500686 RTGAYFARFICKHIVSLGWAEECEVSISWFSQ-PL OSIQVKCFNINKEFSEQL NOVISQYFNWSVAKI IAKLKLLD (6) YAVYG (22) KN 333428 KIYNILANLINDIAKLEGVKECYVRILSQ GKPINE KALDIE ITTEDSY-DIKDIEPKAKEIANKWLDNIMEVQKMIVE KVT-TF	46015497 400245 15836994 6016553 22095828 13357974 2500686 3334428	SSSS LS <mark>TO</mark> IS VO LSTO VSTO VSTO MSTO KN <mark>I</mark> E	s h HSEEIDG HDEEVCI HDPGIKG HSPDVTI HDEDTTH HDEDTNJ HTQDVSI EYKEV-J	a hhhhhh KSLQE DEMRD KDLVE EKLRE IDE IEK LEY FNK DQLRQ EKVRK	hhhh AVMEE ALKEK AVREE TILAD IVIEK KVS-T TIINQ KV-ED	IIKP- /IKA- /IKA- /IKK- /IKK- /VKP- IMDQ- /ILP- IAKK-	P - IL	hhh PAEW PAKY PSKW PASL EKYN IKYG TEYG HINTA	TSA DED DES S-SDN NTD N-DPK DDYER	SSS TKFFINI TIYHLQI TKFHINI TVFHINI IKYYINI FKKIINS IKHLVNI ES <mark>VYLT</mark> Y	PT RFVIG PS RFVIG PT KFVIG PT KFVIG PT RFVIG SS RFVIG PG SFVVG TT TS	s SPC 2 SPC	S S LTCRK: LTCRK: VTCRK: VTCRK: LTCRK: LTCRK: SVCRGI	II DTY II DTY II DTY II DTY II DTY II DTY II DTY NR NGL	-GGMAH -GGWGA -GGSAH -GGWGA -GGYFH -GGIGH -GPYAH ITPFRH	RH-G AH-G AH-G AH-G KH-G HH-G NH-G PM-S	GGAF GGAF GGAF GGAF GGAF GGSF MEZA	SCKD F SCKD F SCKD F SCKD F SCKD F SCKD F SCKD F SCKD F SCKN F	697 S-K S-K S-K S-K T-K T-K VNH	<u> </u>
K + p M hhhhhhhhhhhhhh sssss ssss hhhhhhhh 46015497 RSAAYARYVAKNTVAAGLADRCET QVSYALGVAE TSIMVETFGTEKV-PSEQLTLLVREFFDLRPYGLIQMLDLLH(5) TAAYG (23) LK 400245 RSAAYARWVAKSLVKGGLCRRVDVQVSYALGVAE 15836994 RSASYARWVAKSLVKGGLCRRVDVQVSYALGVAE 15836994 RSASYARWVAKSLVKGGLCRRVDVQVSYALGVAE 15836994 RSASYARWVAKSLVKGGLCRRVDVQVSYALGVAE 15836994 RSASYARWVAKSLVKGGLCRRVDVQVSYALGVAE 15836994 RSASYARWVAKSLAKGLARCEVQVSYALGVAE 15836994 RSASYARWVAKSLAKGLARCEVQVSYALGVAE 15836994 RSASYARWVAKSLAKGLARCEVQVSYALGVAE 15836994 RSASYARWVAKSLAKQLARCEVQVSYALGVAE 15836994 RSASYARWVAKNIVAAGLARCEVQVSYALGVAE 15836994 RSASYARWVAKNIVAAGLARCEVQVSYALGVAE 15836994 RSASYARWVAKNIVAAGLARCEVQVSYALGVAE 12095828 RSAAYARWVAKNIVAAGLARCEVQLSYALGVPK-VSVVVPTGTSKV-DEDKISEAVSVFDLSPRGLEKALELRE(7) LAAFG (24) EL 13357974 RSASYFARFINKHVSLGWAEECEVSISWVFSQPL 2500686 RTGAYFARFINKHVSLGWAEECEVSISWVFSQPL 2500686 RTGAYFARFINKHVSLGW	46015497 400245 15836994 6016553 22095828 13357974 2500686 3334428 21646472	SSSS LSTO ISVO LSTO VSTO VSVO MSVO KNIE PNAN	s f SEEIDC DEEVCI DPGIKC SPDVTI DEDTTF DEDYNJ TQDVSI EYKEV-J YFER-F	A A A A A A A A A A A A A A	hhhh A <mark>V</mark> MEE ALKEK AVREE TILAD IVIEK KVS-T TIINQ KV-ED ELLSF	IIKP- /IKA- /IKK- /VKP- /VKP- IMDQ- /ILP- LAKK- IESQS	P - IL	hhh PAEW PAKY PSKW PASL EKYN IKYG TEYG HINTA DINS	TSA DED DES S-SDN NTD N-DPK DDYER DDP(4)	SSS TK FFI NI TI YHL OI TKFHINI TVFHINI IK <mark>YYI</mark> NI FKKIINS IKHLVNI ES VYLT EG VYLT	PT RFVIG PT RFVIG PT KFVIG PT RFVIG	SPAC DC SPQ CAC SPQ CAC SPV CC SPM CAC SPH CC SPH CC SPAC SPMAC CC SPMAC CC SPC CC SPC CC SPC CC SPC CC SC SC SC SC SC SC SC SC SC SC SC SC S	TCRK: LTCRK: LTCRK: VTCRK: LTCRK: LTCRK: SVCRGI QVCRGI	II DTY II DTY II DTY II DTY II DTY II DTY II DTY NR NGL	-GGMAI -GGWG2 -GGSAI -GGWG2 -GGYF1 -GGYAI -GPYAI ITPFR1 -IAFN1	RH-G AH-G AH-G AH-G HH-G NH-G NH-G RNQI	GGAF GGAF GGAF GGAF GGAF GGSF MEAA MEAA	SCKDF SCKDF SCKDF SCKDF SCKDF SCKDF SCKDF SCKNF ACKNF	697 S-K S-K S-K S-K T-K T-K VNH	00000000000000000000000000000000000000
p MM hhhhhhhhhhhhhh hh 46015497 RSAAYARYYTKNTVAAGLADRCEI QVSYALGVAE: TSIMVETFGTEKV-PSEQLTLLVREFFDLRPYGLIQMLDLLH (5) TAAYG (23) LK 400245 RSAAYARWYTKNTVAAGLADRCEU QVSYALGVSH: LSISIFHYGTSQK-SERELLEIVKKNFDLRPGV: VRDLDLKK (5) TAAYG (23) LK 15836994 RSASYARWYTKNTVAAGLADRCEU QVSYALGVSH: LSISIFHYGTSQK-SERELLEIVKKNFDLRPGV: VRDLDLKK (5) TAAYG (15) KY 15836994 RSASYARWYTKNTVAAGLADRCEU QVSYALGVSH: LSISIFHYGTSQK-SERELLEIVKKNFDLRPGV: VRDLDLKK (5) TAAYG (39) KL 16016553 RSAAYARWYTSLVKAGLARKVLVQLSYALGIAK: ISVTTFGTGKI-SDDKIEKLIRQHFDLRPGKI IKKLDLIKK (5) TAENG (15) KI 22095828 RSAAYARWYTSLVKAGLAKKVLVQLSYALGIAK: ISVLVYAFGTSPL-TDEELHQIVEDSFDLTPGKI IKELDLKR (5) TAENG (15) KI 13357974 RSASYFARWIG KNVVAAKLAKQCEQ LAFALGE-PQ: VAMFVNTFKTNLV-DEKKIFEAINKSFNFDIKTFINDLELWT (5) VATYG (23) SK 2500686 RTGAYFARFICKHIVSLGWAEECEVSISWYFSQ-PL: OSIQVKCFNINKEFSEQLINQVISQYFNWSVAKI IAKLKLLD (6) YAVYG (22) KN 333428 KIYNILENLINDTAKLEGVKECYURI LSQ. (KPINE: KALDIEITEDSY-DIKDIEFKAKEIANKWLDNIMEVQKMIVEG KVT-TF	46015497 400245 15836994 6016553 22095828 13357974 2500686 3334428 21646472	SSSS LSTO ISVO LSTO VSTO VSTO MSVO KNIE PNAN a	S P SEEIDG DEEVCI DPGIKC SPDVTI HDEDTTF HDEDYNJ HTQDVSI EYKEV-J	A A A A A A A A A A A A A A	hhhh AVMEE ALKEK AVREE TILAD IVIEK KVS-T TIINQ KV-ED ELLSF	IIKP- /IKA- ILKP- /IKK- /VKP- IMDQ- /ILP- LAKK- IESQS	P - IL - VU - VI - IA - IA (7)1 CN I (5)1	hhh PAEW PAKY PSKW PASL EKYN IKYG IKYG HINTA DINS	TSA DED DES S-SDN NTD N-DPK DDYER DDP(4)	SSS TK FFI NI TI YHL OI TKFHINI TVFHINI IK <mark>YYI</mark> NI FKKIINS IKHLVNI ES VYLT Y EG <mark>VYLT</mark> Y a	R R R R R R R R R R R R R R	SPC C SPQ CA SPQ CA SPQ CA SPM CA SPH	**************************************	II DTY II DTY II DTY II DTY II DTY II DTY II DTY NR NGL	-GGMAI -GGWGJ -GGSAI -GGWGJ -GGYFI -GGIGI -GPYAI ITPFRI -IAFNI	RH-G AH-G RH-G AH-G KH-G NH-G NH-G RN-S RNQT	GGAF GGAF GGAF GGAF GGAF GGAF MEAA MEAA	SCKDF SCKDF SCKDF SCKDF SCKDF SCKDF SCKDF SCKNF ACKNF ACKNF	697 S-K S-K S-K T-K T-K VNH	
M hhhhhhhhhhhhh sssss ss hhhhhhhh hh 46015497 RSAYARWYAKNIVAACLADRCEI QVSYAI GV-AEI TSIMVETFGTEKV-PSEQLILLIVREFFDLRPYGLIQMLDLLH (5) TAAY G (23) LK 400245 RSAAYARWYAKSIVKGGLCRRVLVQVSYAI GV-AEI TSINVETFGTEKI-SDDKIEKLIRQHEDLRPYGI VKDLDLKK (5) TAAY G (15) KY 15836994 RSASYARWYAKSIVKGGLCRRVLVQVSYAI GV-AEI TSISVITFGTGKI-SDDKIEKLIRQHEDLRPYGI IKMLDLH (5) SASY G (39) KL 6016553 RSAAYARWYAKSIVKAGLAKRVLVQISYAI GI-AKI ISVLVYAFGTSPL-TDEELHQIVEDSFDLTPGKI IKELDLKR (5) TAAY G (15) KY 22095828 RSAAYARWYAKSIVKAGLAKCEI QLSYAI GV-PEY VSVKVDTFGTSKV-DEDKI SEAVSKVFDLSPRGIEKALELRE (7) LAAF G (24) EL 13357974 RSASYFARWIAKNIVAACLADCEE ULAFA GE-POP VAMFWNTFKTNLV-DEKKI FFAINKSENFD IKTFINDLELWT (5) VATY G (23) SK 2500686 RTGAYFARFIAKHIVSLGWAEECEVSI SWVFSQ-PL: QSIQVKCFNINKEFSEQL INQVI SQYFNWSVAKI IAKLKLLD (6) YAVY G (22) KN 334428 KIYNILINLIND TAKLEGVKECYURI LSQI GKPINE; KALDIE ITTEDSY-DIKDIEFEAKEIANKWLDNIMEVQKMIVEG KVT-TF	46015497 400245 15836994 6016553 22095828 13357974 2500686 3334428 21646472	SSSS LSTO ISTO VSTO VSTO VSTO MSTO MSTO RNTE PNAN a	s h SEEIDG DEEVCI DPGIKG SPDVTI DEDTTF DEDTTF DEDYNJ TQDVSI EYKEV-J YFER-F	a hhhhhh XSLQE. DEMRD. XKDLVE. EKLRE DE IEK EY FNK DQLRQ EKVRK SALRH	hhhh AVMEE ALKEK AVREE TILAD IVIEK KVS-T TIINQ KV-ED ELLSF	IIKP- /IKA- ILKP- /IKK- /VKP- IMDQ- /ILP- LAKK- IESQS	P - VI	hhh PAEW PAKY PSKW PASL EKYN IKYG TEYG HINTA DINS	TSA DED DES S-SDN NTD N-DPK DDYER DDP(4)	SSS TK <mark>FFI</mark> NI TIYHLQI TKFHINI TVFHINI IKYYINI FKKIINS IKHLVNI ESVYLTV EG <mark>VYLT</mark> Y	PT RFVIG SRFVIG PT KFVIG C TFIVG SRFVIG SRFVIG SFVVG TT TS1 A AA a	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	**************************************	II DTY II DTY II DTY II DTY II DTY II DTY II DTY II DTY NR NGL	-GGMAI -GGWGJ -GGSAI -GGWGJ -GGYFI -GGIGI -GPYAI ITPFRI -IAFNI	RH-G AH-G AH-G KH-G HH-G NH-G RN-S RNQI	GGAF GGAF GGAF GGAF GGAF GGSF MEAA MEAA	502 SCKDF SCKDF SCKDF SCKDF SCKDF SCKDF SCKDF SCKDF SCKDF SCKNF ACKNF E	S-K S-K S-K T-K T-K T-K VNH	+
hhhhhhhhhhhhhh sssss ssss hhhhhhhh hhh 46015497 RSAYARYYAKNIVAACLADRCE QVSYA GVAE TSIMVETFGTEKV-PSEQLTLLVREFFDLRPYG LQMLDLLH (5) TAAY G (23) LK 400245 RSAYARWYAKSUVKGGLCRRVLVQVSYA GVSH LSISIFHYGTSQK-SERELLEIVKKNFDLRPGVIVRDLDLKK (5) TAAY G (15) KY 15836994 RSAYARWYAKSUVKAGLADRCE QVSYA GVAE TSISVTTFGTGKI-SDDKIEKLIRQHFDLRPYG IKMLDLIH (5) TAAY G (23) LK 6016553 RSAAYARWYAKSUVKAGLARRVLVQLSYA GVAE TSISVTTFGTGKI-SDDKIEKLIRQHFDLRPYG IKMLDLIK (5) TAAY G (15) KI 22095828 RSAAYARWYAKSUVKAGLAKRVLVQLSYA GVPK-VSVKVDTFGTSKV-DEELKJEVGUSSKVFDLSPRG TEKALEJRE (7) LAFG (24) EL 13357974 RSASYFARWIAKNVVAAKLAKQCE QLAFA GEPQQ VAMFVNTFKTNLV-DEKKIEFAINKSFMFDIKTFINDLEUW (5) VATY G (23) SK 2500686 RTGAYFARFIAKHIVSLGWAEECEVSISWVFSQPL:QSIQVKCFNINKEFSEQLINQVISQYFNWSVAKIIAKLKLLD (6) YAVY G (22) KN 333428 KIYNILANLINDTAKLEGVKECYURILSQ.GKVFNINE ALD EITEDSY-DIKDIEPKAKEIANKWLDNIMEVQKMIVE GVXT-TF	46015497 400245 15836994 6016553 22095828 13357974 2500686 3334428 21646472	SSSS LSTO ISVO LSTO VSTO VSVO VSIO MSVO KNIE PNAN a	s h HSEELD HDEEV HDPGIK HSPDVIT HDEDTIT HDEDTIT HDEDVI HTQDVSI EYKEV-J HYFER-F	a hhhhhh XSLQE. DEMRD. XKDLVE. EKLRE DE IEK EY FNK DQLRQ EKVRK SALRH	hhhh AVMEE ALKEK AVREE TILAD IVIEK KVS-T TIINQ KV-ED ELLSF	IKP- /IKA- IKF- /IKK- /VKP- IMDQ- /ILP- LAKK- IESQS	P 	hhh PAEW PAKY PSKW PASL EKYN IKYG TEYG HINTA DINS	TSA DED DES S-SDN NTD N-DPK DDYER DDP(4)	SSS TKFFINI TIYHIQI TKFHINI TVFHINI TKYYINI FKKIIN: ESYYIT EGYYLT a	PT RFVIG PS RFVIG PT RFVIG PT RFVIG PT RFVIG PT RFVIG PT RFVIG PT TS	S SPM DC SPQ DA SPV DC SPM DA SPM DA	S LTCRK: LTCRK: LTCRK: LTCRK: LTCRK LTCRK SVCRGI QVCRGI QVCRGI PP	II DTY II DTY II DTY II DTY II DTY II DTY II DTY NR NGL NR NGL	-GGMAI -GGWG/ -GGSAI -GGWG/ -GGYFI -GGIGI -GPYAI ITPFRI -IAFNI	RH-G AH-G RH-G KH-G HH-G NH-G RN-S RNQI	GGAF GGAF GGAF GGAF GGSF MEAA MEAA P	SCKD F SCKD F SCKD F SCKD F SCKD F SCKD F SCKD F SCKD F SCKN F ACKN F ACKN F	S-K S-K S-K S-K T-K VNH VNH	+ 7 1 1 1 1 1 1 1 1 1 1
Infinite	46015497 400245 15836994 6016553 22095828 13357974 2500686 3334428 21646472	SSSS LSTO ISVO USTO VSTO VSIO MSVO KNIE PNAN a	s h HSEELDQ HDEEVC HDECK SSPDVTI HDEDTT HDEDTT HDEDTY HTQDVSI EYKEV-J HYFER-F	ahhhhh XSLQE DEMRD XDLVE EKLRE IDE IEK EYFNK DQLRQ EKVRK (SALRH	hhhh A <mark>VMEE</mark> ALKEK ALKEK TILAD IVIEK KVS-T TIINQ KV-ED ELLSF	IKP- /IKA- IKF- /IKK- /VKP- IMDQ- /ILP- LAKK- IESQS	P -VV -VI -VI -IA -VL' -IA(7)1 CNI(5)1	hhh PAEW PAKY PSKW PASL EKYN IKYG TEYG HINTA DINS	TSA DED DES S-SDN NTD N-DPK DDYER DDP(4)	SSS TKFFINI TIYHLQI TKFHINI IKYYINI IKYYINI IKHLYNI EG <mark>VYLT</mark> A	PT RFVIG PT RFVIG PT RFVIG PT RFVIG PT RFVIG SS RFVIG PT RFVIG SS RFVIG PT RFVIG SS RFVIG PT RFVIG PT RFVIG SS RFVIG PT RFVIG	SPM.DC SPQ DA SPQ DA SPQ DA SPM DA SP	S LTCRK LTCRK LTCRK VTCRK LTCRK LTCRK SVCRGI QVCRGI QVCRGI PP	II DTY II DTY II DTY II DTY II DTY II DTY II DTY NR NGL NR NGL	-GGMAI -GGWG/ -GGSAI -GGVG/ -GGVGI -GGYAI -GPYAI ITPFRI -IAFNI	RH-G AH-G RH-G KH-G HH-G NH-G NH-G RNQI	GGAF GGAF GGAF GGAF GGSF MEZA MEZA	Sec DP SekDP SekDP SekDP SekDP SekDP SekDP SekDP SekNP SekNP SekNP SekNP	S-K S-K S-K S-K T-K VNH VNH	+ 7 9 9 9 9 9 9 9 7,
4001497 RSAA1ARRUARN VAACHADROELOVSIA CVSHILSINVEHEDIER/PSEUDIELVKEFDDLRPIGIUVRDLDLR(5) TAAIG(25) EK 400245 RSAAYARRUVKSLVKGGLCRVLVQVSYA CVSHILSISIFHYGTSQK-SERELLEIVKKNFDLRPIGIUVRDLDLK(5) TAAYG(15) KY 15836994 RSASYAVRVVAKUVVAGGLARRCEVQVSYA CVAETISISVTFGTGKI-SDDKIEKLIRQHFDLRPIGIIKMLDLH(5) SASYG(39) KL 6016553 RSAAYARRVVAKUVVAGLARRVLVQISYA GIAK;ISVLVYAFGTSPL-TDEELHQIVEDSFDLTPGK IKELDLR(5) TAEG(15) KI 22095828 RSAAYARRVVAKUVVAGLARRVLVQISYA GIAK;ISVLVYAFGTSPL-TDEELHQIVEDSFDLTPGK IKELDLR(5) TAEG(15) KI 2357974 RSASYFARVVAKUVVAGLARRVLVQISYA GVPK;VSVKVDTFGTSKV-DEDKISEAVSKVFDLSPRGIEKALELRE(7) LAAFG(24) EL 13357974 RSASYFARVIKKVVAAKLAKQCETQLAFA GEPQ:VAMFVNTFKTNLV-DEKKIFEAINKSFNFDIKTFINDLELWT(5) VATYG(23) SK 2500686 RTGAYFARFIKHVSLGWAEECEVSISWVFSQPL:QSIQVKCFNINKEFSEQLINQVISQYFNWSVAKIIAKLKLLD(6) YAVYG(22) KN 3334428 KIYNILANLINNDIAKLEGVKECYVRILSQ/GKPINE-KALDIEITTEDSY-DIKDIEPKAKEIANKWLDNIMEVQKMIVEGKVT-TF	46015497 400245 15836994 6016553 22095828 13357974 2500686 3334428 21646472	SSSS LSTO LSTO VSTO VSTO MSTO KNIE PNAN a	s h HSEEIDQ HDEEVC HDECIKQ HSPDVTI HDEDTTH HDEDTYN HTQDVSI EYKEV-J HYFER-F	ahhhhh KSLQE DEMRD. KDLVE EKLRE IDE IEK EYFNK DQLRQ EKVRK (SALRH	hhhh AVMEE ALKEK AVREE TILAD IVIEK KVS-T TIINQ KV-ED ELLSF	IIKP- /IKA- ILKP- /IKK- /VKP- IMDQ- /ILP- IAKK- IESQS	P -IL -VL -VL -IA -IA -IA(7)1 CNI(5)1	hhh PAEW PAKY PSKW PASL EKYN IKYG TEYG HINTA DINS	TSA DED DES S-SDN NTD NDPK DDYER DDP(4)	SSS TKFFINI TIYHLQI TKFHINI IKYYINI FKKIIM: ESYYLTY EG <mark>YYL</mark> T a	PT RFVIG PT RFVIG PT RFVIG PT RFVIG SS RFVIG SS RFVIG PG SFVVG /T TS1 /L TS1 a AA a	SPM-DC SPQ-DA SPQ-DA SPQ-DA SPM-DA SP	S LTCRK LTCRK LTCRK VTCRK LTCRK LTCRK LTCRK LTCRK LTCRK QVCRGI QVCRGI PP	II DTY II DTY II DTY II DTY II DTY II DTY II DTY NR NGL	-GGMAI -GGWGJ -GGSAI -GGWGJ -GGYAI -GGYAI ITPFRI -IAFNI	RH-G AH-G RH-G KH-G KH-G NH-G RN-S RNQI	GGAF GGAF GGAF GGAF GGSF MEZA MEZA P	Sex Sex Sex Sex Sex Sex Sex Sex Sex Sex	S-K T-K S-K T-K T-K VNH	+ 10 0 0 0 0 0 0 0 1 1
400243 RSAAIARWYARSIVAGUCKRUJOVSIA GV-SHIDSISTERIGISGR-SEKELEIVKANDDERPGV VKDEDERK(5) TAAIS(13) KI 15836994 RSASYAVRYVANGLARCEVOVSIA GV-SHIDSISTERIGISGR-SEKELEIVKANDDERPGV VKDEDERK(5) TAAIS(13) KI 15836994 RSAAYARWVARSIVAGUARCEVOVSIA GV-SHIDSISTERIGISGR-SEKELEIVKANDDERPGV VKDEDERK(5) TAAIS(13) KI 15836994 RSAAYARWVARSIVAGUARCEVOVSIA GV-AE: ISISTERIGISGR-SEKELEIVKANDDERPGV VKDELDER(5) TAES(15) KI 22095828 RSAAYARWVARNIVAAELADKCE OISYA GV-PROVSVKVDTFGTSKV-DEELHQIVEDSFDLTPGK IKELELRE(7) LAAFG (24) EL 13357974 RSASYFARWIKKNVVAAKLAKQCEIOLSYA GE-POPVAMFVNTFKTNLV-DEKKIFEAINKSFNFDIKTFINDLELWT (5) VATYG (23) SK 2500686 RTGAYFARFINKHIVSLGWAEECEVSISWVFSQ-PL: OSIQVKCFNINKEFSEQLINQVISQYFNWSVAKIIAKLKLLD (6) YAVYG (22) KN 333428 KIYNILINLINDIAKLEGVKECYVRILSQ-GKPINE-KALDIEITTEDSY-DIKDIEPKAKEIANKWLDNIMEVQKMIVECKVT-TF	46015497 400245 15836994 6016553 22095828 13357974 2500686 3334428 21646472	SSSS LSTO ISTO VSTO VSTO VSTO MSTO KNIE PNAN a	s h SEEID HDES K C HDEG K C SSP U T HDED T HDED T HDED T HDED T HDED T H HDED T H HEC V S H S S H S H S H S H S H S H S H S H	ahhhhh KSLQE DEMRD KDLVE EKLRE IDE IEK EY FNK DQLRQ EKVRK SALRH	hhhh AVMEE ALKEK AVREE IVIEK KVS-T IINQ KV-ED ELLSF	IIKP- /IKA- /IKK- /VKP- IMDQ- /ILP- IESQS	P -VL -VL -VL -IA -IA(7)1 CNI(5)1	hhh PAEW PAKY PSKW PASL EKYN IKYG TEYG HINTA DINS	TSA HKG DES S-SDN NTD N-DPK DOYER D-DY(4)	SSS TKFFINI TIYHLQI TKFHINI TKYYINI IKYYINI FKKIINS IKHLVNI ES <mark>VYLT</mark> T a	C TFIVG C TFIVG C TFIVG C TFIVG C SFVIG C SFVVG C SFVVG C SFVVG C TFIS1 a AA a	SPM DC SPC DA SPC DA SPC DC SPM DA SPM DA SP	S LT CRK LT CRK LT CRK LT CRK VT CRK VT CRK LT CRK SV CRGI QV CRGI QV CRGI PP	II DTY II DTY II DTY II DTY II DTY II DTY II DTY II DTY NR NGL	-GGMAI -GGWG2 -GGSAI -GGWG2 -GGYF1 -GGIGI -GPYAI ITPFR1 -IAFN1	RH-G AH-G AH-G KH-G HH-G NH-G RNQT	GG2F GG2F GG2F GG2F GG2F GG3F ME2A ME2A P	502 SCKDP SCKDP SCKDP SCKDP SCKDP SCKDP SCKDP SCKDP SCKDP SCKNP SCKNP SCKNP SCKNP SCKNP		+ 1 000000
15050594 RSASIAVRIVARNVVAACHADRCEVUVSIA UVAETISISVITEGIGKI-SDDKLEKLIKQHPDLRPYGIIKMLDLH(5) SASYC (39) KL 6016553 RSAYARRWYATSIVKACLAKRVLVQISYA CIAK: ISISVITEGIGKI-SDDKLEKLIKQHPDLRPYGIIKMLDLH(5) SASYC (39) KL 22095828 RSAYARRWYAKNIVAACLAKRVLVQISYA CIAK: ISVVYAFGTSPL-TDEELHQIVEDSFDLTPGKIIKELDLKR(5) TAENG (15) KI 13357974 RSASYFARWIAKNIVAAELADKCE: QISYA (2VPK: VSVKVDTFGTSKV-DEDKISEAVSKVFDLSPRGIEKALELRE(7) LAAF: (24) EL 13357974 RSASYFARWIAKNVVAAKLAKQCEIQLAFA (EPQ: VAMFVNTFKTNLV-DEKKIFEAINKSFNFDIKTFINDLELWT(5) VATYG (23) SK 2500686 RTGAYFARFIAKHIVSLGWAEECEVSISWVFSQPL: QSIQVKCFNINKEFSEQLINQVISQYFNWSVAKIIAKLKLLD (6) YAVYG (22) KN 3334428 KIYNILENLINNDIAKLEGVKECYVRILSQ (KPINE; KALDIEITEDSY-DIKDIEFKAKEIANKWLDNIMEVQKMIVEGKVT-TF	46015497 400245 15836994 6016553 22095828 13357974 2500686 3334428 21646472 46015497	SSSS LSTO ISTO VSTO VSTO VSTO VSTO MSTO RSTA hhhhh RSTA	s h HSEELD HDEEV HDEGIK HDEDT HDEDT HDEDT HDEDY HTOV S HYFER-F HYFER-F hhhh Yhh	ahhhhh KSLQE DEMRD KDLVE EKLRE DE IEK EYFNK DQLRQ EKVRK (SALRH hhhhhh	hhhh AVMEE ALKEK AVREE TILAD IVIEK KVS-T TIINQ KV-ED ELLSF hhh IVAAG	IIKP- /IKA- ILKP- /IKK- /VKP- IMDQ- /ILP- LAKK- IESQS	P -VV -VI -VI -VI -IA (7)1 CNI (5)1 ssssss EIQVSY1	hhh PAEW PAKY PSKW PASL EKYN IKYG TEYG HINTA DINS	TSA HKG DES S-SDN NTD N-DPK DDYER D-DY(4)	SSS TKFFINI TIYHQI TKFHINI TKYYINI KKIIN: KKIIN: EGVYLT EGVYLT a a SSS	PT RFVIG SRFVIG PT RFVIG PT RFVIG CTFIVG SRFVIG SRFVIG SFVVG TTS1 A AA a	SPM DC SPC DA SPC DA SPM DA SPM DA SPM DA SPM DT SPM DT SP	S LTCRK LTCRK LTCRK LTCRK LTCRK LTCRK LTCRK LTCRK LTCRK SVCRG QVCRG QVCRG PP	II DTY II DTY II DTY II DTY II DTY II DTY II DTY NR NGL NR NGL	-GGMAH -GGWGJ -GGSAI -GGSAI -GGYFI -GGIGI -GPYAI ITPFRI -IAFNI -IAFNI hh	RH-G RH-G RH-G KH-G KH-G PM-S PM-S RNQT	GG2F GG2F GG2F GG2F GG2F GG3F ME7A ME7A ME7A P	SCKDF SCKDF		+ 100000 + 27
0010000 KSAAIARKWURTSLVKACHAKKUUULSYA UIAKILSYLVIAEGTSPL-TDEELHQIVEDSPDLTPGKILKELDLKK(5) TAENC(15) KI 22095828 RSAAYARKWURKNVAAELAAKCETOLSYA UV-PKUVSVKVDTFGTSKV-DEDKISEAVSKVFDLSPRGTEKALELRE(7) LAAFG(24) EL 13357974 RSASYFARWIRKNVVAAELAACCETOLSYA UV-PKUVSVKVDTFGTSKV-DEDKISEAVSKVFDLSPRGTEKALELRE(7) LAAFG(24) EL 13357974 RSASYFARWIRKNVVAAELAACCETOLSFA GE-PQUVAKVNTFKTNLV-DEKKIFEAINKSFNFDIKTFINDLELUKT(5) VATY (23) SK 2500686 RTGAYFARFICKHIVSLGWAEECEVSISWVFSQ-PLPOSIQVKCFNINKEFSEQLINQVISQYFNWSVAKIIAKLKLLD(6) YAVY (22) KN 3334428 KIYNILANLIANDIAKLEGVKECYVRILSQIGKPINE-KALDIEIITEDSY-DIKDIEPKAKEIANKWLDNIMEVQKMIVECKVT-TF	46015497 400245 15836994 6016553 22095828 13357974 2500686 3334428 21646472 46015497 400245	SSSS LSTO ISTO VSTO VSTO VSTO MSTO MSTO RSTO AN A A A A A A A A A A A A A A A A A A	s h SEEID DEEV DEEV SPDVTI BEDTT HDEDT HDEDT HDEDT SE SE SE SE SE SE SE SE SE SE SE SE SE	ahhhhh KSLQE DEMRD KDLVE EKLRE EKLRE EYFNK DQLRQ EKVRK (SALRH hhhhh YVAKN WVAKS	hhhh AVMEE AIKEK AVREE TILAO IVIEK KVS-T TIINQ KV-ED ELLSF hhh IVAAG IVAAG	IIKP- /IKA- ILKP- /IKK- /VKP- IMDQ- /ILP- /ILP- SQS s s LADRC LCRRV	P -IL -VI -VI -IA -IA(7)] CNI(5)] SSSSSSS EIQVSY, IVQVSY,	hhh PAEW PAKY PSKW PASL EKYN IKYG TEYG HINTA DINS	TSA DED DES S-SDN NTD NDPK DDYER DDP(4)	SSS TKFFINI TIYHQI TKFHINI TKYINI KYYINI FKKINS FKKINS FKKINS A SSS TEFGTH FFGTH	PT RFVIG PT RFVIG PT KFVIG PT KFVIG PT RFVIG PT RFV	SPM DC SPC DA SPC DA SPC DA SPM DA SPM DA SPM DT SPH DT SPH DT SPH DT SPH DT SPH DT SPM DT SPH DT SPM DT SPM DT SPH DT SPM DT SP	* 6 S LTCRK LTCRK LTCRK LTCRK LTCRK LTCRK LTCRK SVCRG QVCRG QVCRG QVCRG PP	II DTY II DTY II DTY II DTY II DTY II DTY II DTY NR NGL NR NGL	-GGMAI -GGWGJ -GGSAI -GGYGI -GGYGI -GGYGI -GPYAI ITPFRI -IAFNI hh MLDLLI DLDLLI	RH-G AH-G RH-G KH-G KH-G PM-S PM-S RNQT H(5)) K(5)	GGAF GGAF GGAF GGAF GGSF MEAA MEAA P TAAY	SCKDF SCKDF SCKDF SCKDF SCKDF SCKDF SCKDF SCKDF SCKNF ACKNF P	S-K S-K S-K T-K T-K VNH VNH	+ 5 000000000000000000000000000000000000
22095828 KSAAIARKWYRKNIVAAELADKCELQLSYA UVPKYVSYKVDIEGTSKV-DEDKISEAYSKVEDLSPRGIEKALELER(/)LAAF (24)EL 13357974 RSASYFARWIRKNIVAAKLAKQCELOLAFA GEPQ:VAMFVNTFKTNLV-DEKKIFEAINKSENFDIKTFINDLELWT (5)VATY (23)SK 2500686 RTGAYFARFIAKHIVSLCWAEECEVSISWVFSQPL:QSIQVKCFNINKEFSEQLINQVISQYFNWSVAKIIAKLKLLD (6)YAVY (22)KN 333428 KIYNILANLIANDIAKLEGVKEYVRILSQ GKPINE-KALDIELITERS GKPINE-KALDIELITERS	46015497 400245 15836994 6016553 22095828 13357974 2500686 3334428 21646472 46015497 400245 15836994	SSSS LSTO ISVO LSTO VSTO VSVO VSIO MSVO KNIE PNAN a hhhhh RSAA RSAA RSAA	s h HSEEID HDEEV SPDVTI HDEDTTH HDEDTTH HDEDTYN HTQDVSI EYKEV-1 HYFER-F HYFER-F HYFER-F HYFER-F	ahhhhh KSLQE DEMRD KDLVE EKLRE IDE IEK CYFNK CQLRQ EKVRK CSALRH Nhhhhh YVAKN WVAKS	hhhh AVMEE ALKEK AVREE TILAD IVIEK KVS-T TIINQ KV-ED ELLSF Hhh IVAAG LVKGG VVAAG	IIKP- /IKA- ILKP- /IKK- /VKP- IMDQ- /ILP- IESQS S S LADRC CRRV LADRC	P -VI -VI -VI -VI -VI -IA(7)) CNI(5)] SSSSSSS EIQVSY, LVQVSY, LVQVSY,	hhh PAEW PSKW PASL EKYN IKYG TEYG HINTA DINS	TSA DED DES S-SDN NTD N-DPK DJYER D-DP(4) DDP(4) SS AEFTSIM SHPLSIS AEFISIS	SSS TKFFINI TIYHQ TKFHINI IKYYINI FKKINS FKKINS SYLT GVYLT a a SSS VETFGT	PT RFVIG PT	SPM.DC SPQ.DAC SPQ.DAC SPM.DC SPM.D	Ahh FFDLR NFDLR A	II DTY II DTY II DTY II DTY II DTY II DTY II DTY NR NGL NR NGL PYGLIQ PGVIVR PYGIIK	-GGMAH -GGWG -GGWG -GGWG -GGYGI -GGYGI ITPFRI ITPFRI ITPFRI ITPFRI ITPFRI ITPFRI ITPFRI ITPFRI ITPFRI ITPFRI ITPFRI ITPFRI ITPFRI ITPFRI	RH-G RH-G RH-G KH-G NH-G NH-S RNQT H(5) H(5)	GGAF GGAF GGAF GGAF GGAF MEAA E TAAY TAAY SASY	SCKDE SCKDE	S-K S-K S-K S-K T-K S-K T-K VNH VNH VNH	+ 1 00000 +
1335/9/4 KSASIFERMURKNVVAALLAKQCELQLAFALEEPQLVAMFVNTEKTNLV-DEKKIFEAINKSENEDIKTFINDLELMT(5)VATY(23)SK 2500686 RTGAYFARFIAKHIVSLCWAEECEVSISWVFSQPLEQSIQVKCENINKEFSEQLINQVISQYENWSVAKIIAKLKILD(6)YAVY(22)KN 3334428 KIYNILANLIANDIAKLEGVKEYVRILSQ GKPINEEKALDIEIITEDSY-DIKDIEPKAKEIANKWLDNIMEVQKMIVECKVT-TF	46015497 400245 15836994 6016553 22095828 13357974 2500686 3334428 21646472 46015497 400245 15836994 6016553	SSSS LSTO ISTO VSTO VSTO VSIO KNIE PNAN a hhhhi RSAA RSAA RSAA RSAA	s h SEEIDC HDEEVCI HDEGTK BSPDVII HDEDTH HDEDVNI HTQDVSI EYKEV-J HYFER-F HYFER-F YA-F YA-F YA-F	ahhhhh KSLQE DEMRD RDLVE EKLRE IDE IEK CYFNK DQLRQ EKVRK SALRH Ahhhhh XVAKN WVAKS XVAKN WVAKS	hhhh AVMEE ALKEK AVREE TILAD KVS-T TIINQ KV-ED ELLSF LVAAG VVAAG VVAAG	IIKP- /IKA- ILKP- /IKK- /ILP- IAKK- IESQS	P -IL -VI -VI -IA -IA -IA(7)1 CNI(5)1 SSSSSS EIQVSY1 LVQVSY1 LVQVSY1 LVQVSY1 LVQUSY1 LVQUSY1 LVQUSY1	hhh PAEW PAKY PSKW PASL EKYN IKYG TEYG HINTA DINS	TSA DED BSS NTD N-DPK DDYER DDYER DDP(4) SS -AEFTSIM SHFLSIS AEFISIS AEFISIS AEFISIS	SSS TKFFINI TIYHLQI TKFHINI FKFHINI FKKIIMI IKHLYNI EGVYLT GVYLT A SSS VETFGTI FFHYGT SFFHYGT VTFFGTQ	PT RFVIG PT RFV	SPM.DC SPQ.DA SPQ.DC SPM.DA SPM.DA SPH.DT SPM.DA SPH.DT SPM.DA SPH.DT SPM.DA SPH.DT SPM.DA SPH.DT SPM.DA SPH.DT SPM.DA SPH.DT SPM.DA SPH.DT SPM.DA SPH.DT SP	**************************************	II DTY II DTY II DTY II DTY II DTY II DTY II DTY NR NGL NR NGL PYGLIQ PGVIVR PYGLIQ PGVIVR	-GGMAH -GGWGJ -GGWGJ -GGYGI -GGYAH ITPFRI -IAFNI -IAFNI hh MLDLLI DLDLKI MLDLI	RH-G RH-G RH-G AH-G KH-G MH-G PM-S RNQT H(5) K(5) H(5) R(5)	GGAF GGAF GGAF GGAF GGAF GGAF GGAF E TAAY TAAY SASY SASY TAAY	SCKDF SCF SCF SCF SCF SCF SCF SCF SCF SCF SC	S-K S-K S-K S-K T-K S-K T-K T-K VNH VNH VNH	+ 1 000000000000000000000000000000000000
2500686 RIGAIFERFIARHIVSLOWAEECEVSISWVFSQPLPQSLQVKCENINKEFSEQLINQVISQYFNWSVAKIIAKLKLLD(6)YAVY (22)KN 3334228 KIYNILANLIANDIAKLEGVKEYVRILSQ GKPINEFKALDIEIITEDSY-DIKDIEFKAKEIANKWLDNIMEVQKMIVECKVT-TF	46015497 400245 15836994 6016553 22095828 13357974 2500686 3334428 21646472 46015497 400245 15836994 6016553 22095828	SSSS LSTO ISTO VSTO VSTO VSTO VSTO MSTO KNIE PNAN a A RSAA RSAA RSAA RSAA	s h HSEEIDC HDEGVCI HDEGVCI HDEDT HDEDT HDEDT HTOVSI EYKEV-J HYFER-H HYFER-H YAH YAH YAH YAH YAH	ahhhhh KSLQE DEMRD KDLVE EKLRE IDE IEK EYFNK DOLRQ EKVRK SALRH NHhhhh YVAKN WVAKS WVAKS WVAKS	hhhh AVMEE ALKEK AVREE TILAD IVIEK KVS-T TIINQ KV-ED ELLSF LVSG LVSG LVAAG LVKAG IVAAG	IIKP- IKP- IKP- IMDQ- IMDQ- ILP- AKK- IESQS CCRV LADRC LADRC LADRC	P -VL -VL -VL -IA-(7)1 CNI (5)1 SSSSSS EIQVSY1 LVQVSY1 LVQVSY1 EVQVSY1 LVQLSY1 EIQLSY1	hbh PAEW PAKY PSKU PSKU EKYN IKYG HINTP DINS A GV- A GV- A GV- A GV- A GV- A GV-	TSA HKG DES S-SDN NTD N-DPK DDYER DDYER DDP(4) SS -AEFTSIM -AFTSIM -AFTSIM -AFTSIM -AFTSIM -AFTSIM -AFTSIM -AFTSIM	SSS TKFFINI TIYHLQI TKFHINI TKYYINI IKYYINI FKKIIN: ESVYLTY EGVYITT A SSS VETFGTI IFHYGTS VTTFGIC	PT RFVIG SRFVIG PT KFVIG CTFIVG CTFIVG SRFVIG GSFVVG TTS1 A AA a CKV-PSEQ SQK-SERE SKI-SDC SPL-TDEE SKV-DEDK	SPM DAC SPV DC SPV DC SPM DAC SPM DAC	APPENDENCE APPEND	II DTY II DTY II DTY II DTY II DTY II DTY NR NGL NR NGL PYGLIVR PYGLIVR PYGLIKI PRGIEK	-GGMAI -GGWGJ -GGSAI -GGWGJ -GGYFI -GGYAI ITPFRI -IAFNI MLDILI DLDLKI ALELRI ALELRI	RH-G AH-G RH-G KH-G HH-G PM-S RNQI H(5) K(5) H(5) R(5) C (5)	GGAF GGAF GGAF GGAF GGAF GGSF MEBA MEBA E TAAY TAAY SASY TAAY SASY TAEN LAAF	555 5 CKD F 5	S-K S-K S-K S-K T-K T-K VNH VNH VNH VNH K K K K K K K K K K K K K K K K K K K	+ +
3334428 KIYNILONLIANDIAKLEGVKECYVRILSQUEKPINE KALDIEIITEDSY-DIKDIEFKAKEIANKWIDNIMEVQKMIVECKVT-TF	46015497 400245 15836994 6016553 22095828 13357974 2500686 3334428 21646472 46015497 400245 15836994 6016553 22095828 13357974	SSSS LSTO ISTO VSTO VSTO VSTO VSTO VSTO NSTO RSTO RSAA RSAA RSAA RSAA RSAA RSAA	s h SEELD HDEEV HDEGIK SSPUTI HDEDTTH HDEDTTH HDEDY SYFER-F HY	ahhhhh KSLQE DEMRD KDLVE EKLRE EKLRE EKLRE EKKRK CSALRH Ahhhhhh XVAKN WVAKS WVAKS WVAKN WVAKN WVAKN WVAKN	hhhh AVMEE AIKEK AVREE TILAD IVIEK KVS-T TIINQ KV-ED ELLSF LVKGG VVAAGI VVAAGI VVAAGI VVAAGI VVAACI	I KP- I KA- I KA- VKP- VILP- I LP- I LP- LP- LP- LP- LP- LP- LP- LP- LP- LP-	P -UL -VI -VI -IA -IA(7)1 CNI(5)1 SSSSSSS EIQVSY1 EVQVSY1 EVQVSY1 EIQLSY1 EIQLAF;	hhh PAEW PASY PSKW IKYG TEYG HINTA DINS A CV- A CV- A CV- A CV- A CV-	TSA HKG S-SDN NTD N-DPK DDYER DDYER DDYER DDP(4) SS -AEPTSIN -SHPLSIS -AEPISIS -AEPISIS -AEPISV 	SSS TKFFINI TIYHLQI TKFHINI TKYYINI FKKIIN: FKKIN: SSS VETFGTI IFHYGT: VTTFGTC VTFFGT	PT RFVIG	SPM DC SPM DC SPM DAC SPM DAC SPM DAC SPM DAC SPM DT C SPM DT SPM DT C SPM DT SPM	s LTCRK LTCRK LTCRK LTCRK LTCRK LTCRK LTCRK LTCRK LTCRK QVCRGI QVCRGI PP hh FFDLR NFDLR HFDLR SFDLT VFDLS SFNFD	IIDTY IIDTY	-GGMAI -GGWGJ -GGSAI -GGYGJ -GGYGI -GPYAI ITPFRI -IAFNI MLDLLI DLDLKI MLDLLI ALELRI DLELKI DLELRI	RH-G AH-G RH-G KH-G HH-G PM-S RNQT H(5) H(5) H(5) R(5) E(7) I(5)	GGAF GGAF GGAF GGAF GGAF GGAF CCSF E CSSF E E TAAY TAAY TAAY SASY TAEN LAAF VATY	S S K C K S K C K S K C K S K C S K K S K K S K K S K K S K K S K K S K K S K K S K K S K K S K K S K K S K K S K S K S K S K S K S K S K S K S K S K S K S K S K S K S K S K S K S K S K S K S K S K S K S K S K S K S K S K S K S K S K S S K S S K S S S S S S S S	697 S-K S-K S-K S-K S-K T-K VNH VNH VNH K K K K L K K L K K L S K	+ 1 000000 +
	46015497 400245 15836994 6016553 22095828 13357974 2500686 3334428 21646472 46015497 400245 15836994 6016553 22095828 13357974 2500686	SSSS LSTO ISTO VSTO VSTO VSTO VSTO MSVO KNIE PNAN a hhhhi RSAA RSAA RSAA RSAA RSAA RSAA RSAA	s h SEEID HDEEV HDEGK SSPDVTI HDEDT HDEDT HDEDY HTQDV SI EYKEV-J HYFER-P HYFER	ahhhhh KSLQE DEMRD KDLVE EKLRE DEIEK EYFNK DQLRQ EKVRK CSALRH NHHHH YVAKN WVAKS WVAKS WVAKS WVAKS WVAKS WVAKS WVAKS WVAKS	hhhh AVMEE AIKEK AVREE TILAD IVIEK KVS-T TIINQ KV-ED ELLSF LUSF LUSF VAAG IVAAG IVAAG IVAAG IVAAG IVAAG IVAAC IVAAK	LIKP- IKA- IKK- VKP- VILP- AKK- LESQS S CRRV LADRC CRRV LADRC LAKRV ADKC LAKRV MAEEC	P -IL -VV -VI -IA -IA -IA(7)) CNI(5)] SSSSSS EIQVSY, IVQVSY, EVQVSY, EVQVSY, EVQUSY, EIQLSY, EIQLAF; EVSISW	hhh PAEW PASU PSKW PSKW PSKW EKYN IKYG TEYG HINTA DINS A CV- A CV- A CV- A CV- A CV- A CV- A CV- A CV- CV- CV- CV- CV- CV- CV- CV- CV- CV-	TSA DED S-SDN NTD N-DPK DDYER D-DYER DDP(4) DP(4) DP(4) 	SSS TKFFINI TIYHQI TKFHINI TKYINI FKKINS FKKINS FKKINS FKKINS FKKINS TFFGT TFFGT TFFGT VTFFGT VTFFGT VTFFGT VTFFGT	PT RFVIG PT RFV	SPM DC SPM DC SPM DAC SPM DAC SPM DAC SPM DAT SPH DT SPH DT SPH TT SPH DT SPH DT SPH DT SPH DT SPH DT SPH DT SPH DT SPH DT SPH SPH SPH SPH SPH SPH SPH SPH SPH SPH	**************************************	II DTY II DTY II DTY II DTY II DTY II DTY II DTY II DTY II DTY PGTIQ PGY PGY II PGGIIK PGGIIK PGGIIK PGGIIK PGGIIK PGGIIK PGGIIK	-GGMAH -GGWGJ -GGWGJ -GGYGI -GGYGI GPYAI ITPFRI -IAFNI MLDLLI DLDLKI MLDLLI LLDLKI ALELRI DLELWI KLKLLI	RH-G RH-G RH-G KH-G PM-S RNQT H(5) K(5) H(5) R(5) F(5) F(5) C(6)	GGAF GGAF GGAF GGAF GGAF GGSF MEAA MEAA MEAA MEAA MEAA MEAA MEAA MEA	55 5 CKD F 5 CKD F	697 S-K S-K S-K S-K T-K S-K VNH VNH SK KN SK KN	+ 5 000000000000000000000000000000000000

Figure 6

Multiple sequence alignment of SAM synthetases. Essential amino acids for the first (*) and the second substrate binding subunit are numbered as in gi:46015497 (PDB structure <u>IP7L</u> and <u>IRG9</u>). Aligned sequences represent protein from Archaea (gi:3334428 – *M. jannaschii*), Eukaryota (gi:400245 – *H. sapiens*, gi:6016553 – *C. elegans*), and Bacteria (gi:46015497 – *E. coli*, gi:15836994 – *X. fastidiosa*, gi:22095828 – *F. nucleatum*, gi:13357974 – *U. parvum*, gi:2500686 – *M. pneumoniae*, gi:21646472 – *C. tepidum*). Residues are highlighted according to the amino acid properties with designations as in Figure 2. Substrate binding is annotated below the alignment as follows: **small letters** – water mediated interactions; **inverted colors** – interactions with ligand from the second subunit of the homodimeric protein ; **A**/a – adenosyl moiety; **R** – ribosyl moiety, **M**/m – methionine moiety; (+) – Mg²⁺; **K** – K⁺; **P** – PPNP/Phosphate moiety. Consensus positions of the secondary structure elements are shown above the alignment. Numbers in parentheses indicate number of residues omitted for clarity.

Table 1: Classes of SAM binding proteins

Reaction summary	Trivial name, EC number	Sequence families and spatial folds (summary)	Evolutionary roots and status in LUCA
Methyl transfer	SAM-dependent methyltransferase, EC:2.1.1	Five classes: I, Rossmann fold; II, reactivation domain of methionine synthase; III, "corrinoid-like" MTases; IV, SPOUT domain; V, SET domain. Classes I-IV are α/β folds, class V is a β -clip. The Rossmann-fold MTases are the largest class of SAM-dependent enzymes. The folds of Classes II and III are unique. Trm I0 and TrmH families of RNA MTases appear to be the modified versions, of, respectively, class IV fold [47], [49] and class I fold fused to PP-superfamily ATPase (this study). GTP MTase of Sindbis-like viruses may belong to α/β class, but specific fold prediction is unavailable	Several distinct Rossmann-fold methyl transferases in LUCA.
Methylene transfer	Cyclopropane fatty acid synthase, EC:2.1.1.79	Rossmann-fold methyltransferase family	Derived from an ancient enzyme; not in LUCA
Aminoalkyl transfer ¹	Nicotianamine synthase, EC:2.5.1.43	Rossmann-fold methyltransferase family with permuted order of sequence motifs	Derived from an ancient enzyme; not in LUCA
	ACC synthase, EC:4.4.1.14	PLP-dependent aminotransferase fold; the SAM-binding domain is derived from generic substrate-binding cleft	Derived from an ancient enzyme; not in LUCA
	Acyl-homoserine lactone synthase, EC:6.1	GNAT-type acetyltransferase fold; the SAM- binding domain is derived from generic substrate-binding cleft	Derived from an ancient enzyme; not in LUCA
Aminopropyl transfer	Spermidine synthase, EC:2.5.1.16	Rossmann-fold methyltransferase family, but the substrate is decarboxy-SAM	Probably in LUCA
Ribosyl transfer	tRNA-ribosyl transferase- isomerase, EC:5	QueA family; smaller β -barrel N-terminal domain and a larger C-terminal domain with α/β fold, distantly related to a TIM-barrel	Bacterial invention; not in LUCA
5'deoxyadenosyl transfer	5'-fluoro-5'-deoxy-adenosine synthase, EC:2.5.1.63	Two-domains; larger N-terminal domain has distant similarity to Rossmann-fold methyltransferases, smaller C-terminal domain is a β-barrel	Bacterial invention; not in LUCA
5'deoxyadenosyl radical synthesis	SAM radical enzymes	TIM-like α/β barrel with additional inserted elements. May have distant sequence similarity to TIM barrel of corrinoid methyltransferase (see text)	Probably in LUCA
SAM decarboxylation	SAM decarboxylase, EC:4.1.1.50	$\alpha/\beta/\beta/\alpha$ sandwich in eukaryotes, apparently produced by duplication of a half-unit; stand- alone half-units exist in many bacteria and archaea	Probably in LUCA
De novo SAM synthesis	Methionine adenosyl transferase, EC:2.5.1.6	Unique fold: repeat of 3 $\beta\text{-}\alpha\text{-}\beta\text{-}\beta\text{-}\alpha\text{-}\beta$ units	Probably in LUCA
Regulatory binding of SAM	Methionine repressor	All- α SAM-binding domain is derived from generic small molecule-binding domain	Bacterial innovation
	CBS domain	Mostly-β SAM-binding domain is derived from generic small molecule-binding domain	Not in LUCA
	Transcription factor mtTFB	Rossmann-fold methyltransferase family member that has lost catalytic activity	

I An additional reaction in this class is synthesis of $acp^{3}U$, a modified base in some tRNAs and rRNAs. This amino alkyl transfer requires SAM, but responsible protein has not been identified.

phobic pocket for the adenine ring of SAM [77]. Invariant Tyr85 is involved in the substrate recognition, and interacts with active-site Lys273 from the adjacent subunit (Lys273 forms a covalent Schiff base with PLP cofactor) [78,79]. Structure of ACC synthase is similar to other PLP-dependent transferases, such as transaminating aminotransferases, β -eliminating lyases, and cystathionine synthase. The only other enzyme in this group that binds SAM is 7,8-Diaminopelargonic acid (DAPA) synthase, which utilizes SAM in a different way than ACC syn-

COG	Phyletic pattern parsimony ¹	Tree topology ²	Protein product name and function	Fold class
			SAM-binding proteins in LUCA – best-supported C	OGs
COG0030	1/0	+	KsgA, Dimethyladenosine transferase (rRNA methylation)	Rossmann fold
COG0144	1/0	+	Sun, tRNA and rRNA cytosine-C5-methylases	Rossmann fold
COG0192	0/1 3	+ 3	MetK, S-adenosylmethionine synthetase (releases SAM as product)	SAM synthase
COG0421	1/0	+	SpeE, Spermidine synthase	Rossmann fold
COG0500	1/0	+	SmtA, SAM-dependent methyltransferases	Rossmann fold
COG0621	1/0	+	MiaB, 2-methylthioadenine synthetase	Rossmann fold
COG1180	0/1	+	PfIA, Pyruvate-formate lyase-activating enzyme	SAM radical (TIM barrel)
COG1586	0/1 3	+ 3	SpeD, S-adenosylmethionine decarboxylase	SAM decarboxylase
COG2890	1/0	+	PrmC (HemK), Methylase of polypeptide chain release factors	Rossmann fold
			SAM-binding proteins in LUCA – tentatively suppo	rted COGs
COG0007	1/0	?	CysG, Uroporphyrinogen-III methylase	SAM radical (TIM barrel)
COG0220	1/0	?	Predicted S-adenosylmethionine-dependent methyltransferase	Rossmann fold
COG0293	0/1	?	Fts], 23S rRNA methylase	Rossmann fold
COG0320	0/1	?	LipA, Lipoate synthase	SAM radical (TIM barrel)
COG0502	0/1	?	BioB, Biotin synthase and related enzymes	SAM radical (TIM barrel)
COG0565	0/1	?	LasT, rRNA methylases	SPOUT
COG0685	1/0	?	MetF, 5,10-methylenetetrahydrofolate reductase (binds SAM as inhibitor)	TIM barrel
COG1189	0/1	?	Predicted rRNA methylase	Rossmann fold
COG2226	1/0	?	UbiE, Methylase involved in ubiquinone/menaquinone biosynthesis	Rossmann fold

Table 2: Repertoire of SAM-binding proteins in the Last Universal Common Ancestor

¹ Placement of COGs in LUCA based on weighted parsimony analysis of phyletic patterns [138]. 1/0 indicates that this gene has been tracked back to LUCA genome under all tested conditions; 0/1 indicates tentative placement of gene in LUCA under some but not all conditions.

² Phylogenetic analysis of each COG. Plus sign indicates tree topology that is fully consistent with gene presence in LUCA, and the question mark indicates a tentative support for the LUCA hypothesis (see Materials and Methods)

³ The common ancestry and phylogeny of archaeal, bacterial, and eukaryotic homologs was established in this work (see text for details).

thase, namely as amino group donor in aminotransferase reaction. Both ACC synthase and DAPA synthase have similar active site residues involved in PLP cofactor binding. The interactions between SAM and DAPA synthase have not been studied in detail.

It is likely that ACC synthase and DAPA synthase evolved from other aminotransferases with different, perhaps broad, specificity, by accumulating changes in the ligandbinding region that increased its specificity towards SAM [80]. The evolutionary heritage of ACC synthase is manifest in the retained ability of the enzyme to catalyze slow transamination of substrates such as alanine [81]. Structural similarity between ACC-, DAPA-synthase and some other PLP-dependent enzymes (i.e. cystine C-S lyase – PDB structure <u>1ELQ</u>; cystathionine β -lyase – PDB structure <u>1CL2</u>) indicate that SAM binding in this case may have originated from ancestor with PLP-dependent binding of amino group of various sulfur containing amino acids or amino acid derivatives.

AHL synthase

Synthesis and detection of acyl-homoserine lactones (AHLs) enables many gram-negative bacteria to engage in quorum/diffusion sensing, an intercellular signaling mechanism that activates differentiation towards virulence and biofilm formation [82-84].

The AHL synthases (COG3916) catalyze acylation and lactonization of SAM, where the acyl group is provided by acylated acyl carrier protein (acyl-ACP) [85,86]. AHL synthases (SCOP family: 75508; Figure 1p) has acyl-CoA N-acyltransferase ("GNAT-like") fold: $\alpha/\beta/\alpha$ sandwich with highly twisted β -sheet (SCOP fold: 55728). The conserved N-terminal residues: Arg23, Phe27, and Trp33 (numbered as in PDB structure <u>1RO5</u>) form putative SAM binding pocket and undergo a dramatic conformational rearrangement upon acyl-ACP binding. This conformational change brings conserved residues of the putative SAM binding site in close proximity to the catalytic site [87]. Position of conserved β -bulge formed by Ser103 and Arg104 in β 4 (numbered as in PDB structure <u>1RO5</u>) dis-

tinguishes SAM-binding from other proteins with acyl-CoA-N-acyltransferase fold [86,87]. There is no detectable sequence or structural similarity between AHL synthases and other known SAM binding proteins, indicating independent origin of SAM binding in this fold. As with ACC synthase and DAPA synthase, the most likely mechanism of adaptation was by selecting relatively small changes in a generic ligand-binding region that increased relative affinity to SAM.

Met repressor

The E. coli MetJ repressor (Figure 1m; SCOP family: 100972; COG3060) uses SAM as a co-repressor to regulate the production of methionine. MetJ is a homodimeric, DNA-binding protein with ribbon-helix-helix fold. Corepressor (SAM) binds to each monomer of the protein dimer at sites that lie on the opposite side of the protein from the DNA-binding motif. Binding of co-repressor affects DNA affinity, but apparently not specificity of MetJ [88-91]. Affinity of MetJ DNA binding is affected primarily by the positive charge associated with the ternary sulfur atom in co-repressor (SAM), which creates a region of positive electrostatic potential on the DNA binding surface overlapping the adjacent phosphodiester backbone in the region of the operator [92-94]. The SAM's adenine ring inserts itself deeply inside a hydrophobic pocket, consisted of side chains from both monomers. The positively charged sulfur of the SAM is greatly attracted by the net negatively charged C-terminal end of the β -helix, hence docking the SAM molecule in place. Electrostatic properties of SAM and its ability to serve as a regulatory feedback molecule in the common metabolic pathway of methionine synthesis probably played an important role in the emergence of this unique mode of SAM binding by MetJ.

MetJ is the only known SAM-binding representative of evolutionarily ancient ribbon-helix-helix (RHH) class of DNA-binding proteins [92,95]. Evolution of SAM cofactor binding in this protein was feasible because SAM adenosyl moiety fit into the cleft formed by both monomers, and its sulfonium center conformation was able to adapt to non-catalytic electrostatic interactions with MetJ repressor.

CBS

CBS-domains (COG0517) are widely distributed in all divisions of Life, in the form of fusions with various unrelated proteins, where they usually form tandem pairs. Binding of the adenosyl-containing molecules, such as ATP, AMP, and SAM by CBS-domains is important for their function as energy or redox status-sensing modules [96-98]. Some CBS-domains also binds single stranded nucleic acids [99]. In general, a tandem of CBS domains (encoded by ~120 aa) folds into one domain with a β -sandwich and 4 α -helices extending from one edge (Figure 10). CBS domains within each pair are asymmetric. CBS-domain is common in multidomain proteins (i.e.: 15 in Bacteria and 9 in Archaea [100]) and is probably derived from generic small molecule-binding domain. The mode of SAM binding to CBS domains remains unknown.

Rare folds

Porphyrin C-methyltransferase

Porphyrin C-methyltransferases are a family of proteins involved in biosynthesis of tetrapyrroles, which are used in chelation of metal ion. Representative structures of Cbif and CysG with bound reaction product (SAH) are known. They have the same fold, which consists of two dissimilar α/β domains. Domain 1 has parallel sheet of 5 strands (order 32415) sandwiched between 3 α -helices; domain 2 has mixed sheet of 5 strands (also sandwiched between 3 α -helices), order 12534; strands 4 & 5 are antiparallel to the rest; (Figure 1d) [101,102]. The reaction product is bound in a large pocket between the N and C-terminal domains. Binding pocket contains conserved glycine-rich loop GAGPGD, similarly to Rossmann-fold methyltransferases, but in crystal structures of porphyrin C-methyltransferases (i.e. PDB structure 1CBF, 1PIQ, and 1S4D) glycine-rich loop does not contact methionine moiety of SAH, at least in the absence of the precorrin substrate. Instead, Pro30 from the glycine-rich loop (numbered as in PDB structure 1CBF) forms hydrogen bond with adenosyl moiety. Other conserved residues participate in (SAH) binding: adenosyl moiety hydrogen-bond with conserved Ala213, ribosyl moiety hydrogen bound with conserved hydrophobic Leu184 and Ala241; carboxyl group bind Asp103 from conserved Gly-Asp-Pro motif and also Tyr131; the amino group hydrogen-bond with conserved hydrophobic Met106. Near the entrance of SAM binding cleft there is conserved aromatic residue (Tyr107) positioned next to the sulfur of SAH.

Despite structural similarity to Rossmann-fold methyltransferases and local sequence similarity in the glycinerich loop, porphyrin C-methyltransferases have distinct mode of SAM binding. The evolutionary relationships between porphyrin methyltransferases and other SAMbinding (or indeed any other) proteins remain unclear.

Met synthase activation domain

 B_{12} -dependent methionine synthases are large multidomain proteins, which catalyze a multistep reaction of the transfer of a methyl group from 5-methyltetrahydrofolate to homocysteine. The C-terminal domain of Met synthase is involved in reactivation of spontaneously oxidized coenzyme $B_{12'}$ and therefore is required for the catalysis. This activation domain (pfam02965 and C-terminal part of COG1410) has an unusual fold: $\beta \cdot \alpha(2) \cdot \beta(2) \cdot \alpha(2) \cdot \beta \cdot \alpha_{-\beta}$; antiparallel β -sheet; order 12354 (Figure 1e; SCOP fold: 56506). Large conformational changes are required for reactivation reaction to occur [103-105].

This domain binds SAM in a shallow groove along the edges of the β -strands. There are several conserved residues involved in SAM binding (numbered as in PDB structure <u>1MSK)</u>: Asp946 and Glu1101 bind carboxypropyl moiety; Arg1134 and Tyr1130 (later via water mediated interactions) bind ribosyl moiety; Tyr1139 and Tyr1189 participate in stacking interactions with adenosyl moiety; Tyr1190 main chain (or Phe in other known sequences) hydrogen bond with adenosyl moiety of SAM. This unusual domain is found in bacterial and metazoan proteins. Interestingly, Met synthases of thermophilic bacteria lack the SAM-binding/activation domain altogether. Presumably, the lack of the transferase domain activity is compensated for by the methyltransferase protein also involved in the same reaction [106]. This particular adaptation in thermophiles seems to be partly due to requirement for greater thermal stability of the enzyme - making large conformational changes required for reactivation less favorable.

Divergence and convergence in evolution of SAM binding

The known SAM-binding proteins belong to 15 distinct superfamilies/folds, which comprise at least 20 more sequence families (Figure 1 and Table 4), that is: 1 – Rossmann fold (methyltransferase, methylene transferase, nicotianamine synthase, spermidine synthase, spermine synthase, and acalcynomycin-10-hydroxylase), 2 – Met synthase activating domain, 3 – porphyrin C-methyltransferase, 4 – SPOUT methyltransferase, 5 – SET domain methyltransferase, 6 – isoprenylcysteine carboxyl methyltransferase, 7 – fluorinating enzyme, 8 – tRNA ribosyltransferase-isomerase (QueA), 9 – SAM decarboxylase, 10 – SAM synthetase, 11 – ACC synthase and DAPA synthase, 12 – N-acyl-homoserine lactone synthase, 13 – Met repressor, 14 – CBS domain, 15 – SAM-dependent radical.

These folds belong to all four large structural classes [107], though there is a distinct excess of α - β and especially α/β folds.

Comparison of fold classes and molecular functions reveals a broad picture of complex interplay between sequence divergence and functional convergence in the evolution of SAM-binding proteins and SAM-dependent molecular functions. On the whole, the assemblage of SAM-binding proteins is extremely heterogeneous. There are huge, apparently monophyletic, superfamilies, which in turn belong to even larger superfolds (Class I methyltransferases from a Rossmann-like superfold is an example), and there are small families with unique folds. There are molecular functions confined to just one superfamily, as in the case of SAM radical formation, which thus far is known to be performed only by enzymes belonging to large and diverse, yet apparently monophyletic, TIM-barrel-like SAM-radical fold. On the other hand, there is SAM-dependent methylation, performed by at least five classes of enzymes, which represent two completely different fold classes and may have been "invented" three or four times (ref. [12] and this study).

Duplication of protein domains is relatively common in both prokaryotes and eukaryotes, for example at least 58% of the domains in Mycoplasma [108] and 98% of the domains in humans [109-111] are duplicates. Several SAM-binding proteins appear to have evolved by ancient domain duplications. Examples include SET-domain methyltransferase [71], SAM synthase, where triplication of a basic wedge-shaped module is likely, and eukaryal SAM decarboxylase (this study). Here again, there is no strict correlation between that mechanism of protein emergence and its further evolution trajectory: SAM synthase and SAM decarboxylase have most likely persisted in evolution as single- or low-copy, vertically transmitted genes since the Last Universal Common Ancestor, whereas SET methyltransferases appear to be an eukaryotic innovation that experienced lineage-specific expansions, significant diversification of substrate specificity, and occasional horizontal transfer to prokaryotes.

Examination of the largest, most diverse set of SAMdependent enzymes unified by common biochemical function, namely, SAM-dependent methyltransferases, shows more of the same trend. All methyltransferases belong to the same EC class (EC: 2.1.1.-), but they comprise five structural families (reviewed in ref. [12]), which appear to lack clearly discernible common ancestor (with the possible exception of Class I and SPOUT methyltransferase, see above). Three classes, I, IV, and V, are large, found across broad groups of genomes, and, at least in the cases of classes I and V, in multiple genomic contexts, including various protein fusions. In contrast, class II is restricted to just one specialized enzymatic system, methionine synthase, with very specific domain composition. Class I displays a huge variety of substrate specificities, whereas the substrates of other classes are much more narrowly defined (tetrapyrroles for class III, mostly rRNAs and tRNAs for class IV, and short list of proteins for class V). Interestingly, each of these classes of substrates are also targeted by class I enzymes. For example, although most of precorrin methyltransferases are class III enzymes, precorrin-C6 methyltransferase CbiT is a typical Class-I Rossmannoid [112]. Similarly, though most histone lysine Nmethyltransferases are class V (SET-domain proteins), the Dot1 histone H3-Lys79 N-methyltransferase belongs to Class I [113]. In at least one case, exactly the same base in

Table 3: Counts of each family in selected genomes

Protein	family nai	me and nu	mber of g	enes (proto	eins) per g	enome	Additional information about counts of each family in selected genomes.
HUMAN	ARATH	YEAST	PSEAE	мүсти	SULSO	HELPY	
		L. Ros	smann fold	MTase			
49 (64)	70 (81)	30	55	37	31	26	Rossmann fold methyltransferase (MTase) makes the list of 10 most commonly used sequence and structure families [15]. In analyzed genomes, 4.6% of Rossmann fold MTase does not have motif-I (Gly-rich) and may lack MTase activity. Fold: Rossmann- fold.
	-	2. Met synt	hase activat	ing domain	-	-	
I	0	0	 MT	1	0	0	Present in many bacteria and metazoan. Fold: Met synthase activation domain-like.
	2 (1)	3.	o o o	5 III Z	7		Human ganama angadag protain similar to diphthing synthese but
I	3 (4)	Z	0	0	7	I	not other Class-III MTases (i.e. from tetrapyrrole biosynthesis pathway) identified in analyzed genomes. Fold: Tetrapyrrole- methyltransferase.
		4. 9	SPOUT MT	ase			
3	5	2	4	6	I	2	Every analyzed here genome has at least one putative SPOUT rRNA MTase, in archaea the same base of tRNA is methylated by Rossmann-fold instead of SPOUT MTase. Fold: α/β knot.
25 (24)	24 (41)	5. SE	I domain P	llase	•	•	
25 (34)	34 (41)	5	0	U	U	0	Present in all Eukaryota with sporadic lateral transfer to bacteria and archaea (but not identified in prokaryotic genomes analyzed here). In S. cerevisiae all selected sequences similar to SET-domain MTases are conserved within smart00317 domain (in contrast to SET-domain proteins other then MTases). Fold: β -clip.
		6. met	hylene tran	sferase			
0	4	0	I 	10	0	I	This enzyme performs cyclopropane fatty acid synthesis and is important for <i>Mycobacterium</i> survival. Fold: Rossmann-fold.
0	4	I. NICO	tianamine s	ynthase	•	•	The state of the s
0	4	0	U	0	0	0	function infrequently distributed in bacteria, archaea and fungi (discussed earlier). Fold: Rossmann-fold (predicted).
	8. spe	rmidine syn	thase & 9. s	permine syn	thase		
I	3	2	3	I	I	I	The mammalian enzyme is highly specific but the bacterial enzyme can use other acceptors then SAM and can synthesize spermine. Spermidine synthase but not spermine synthase is essential for survival of <i>Arabidopsis</i> and <i>S. cerevisiae</i> [142-144]. Fold: Rossmann-fold.
		10. aclacyno	omycin-10-l	nydroxylase			
0	0	0	0	0	0	0	Was found in <i>Streptomyces purpurascens</i> only. Fold: Rossmann-fold.
2	h	II. Isopre	enyicysteine	U-IMI ase	0	0	ICMT entrymes and present in all culture stic enternisms [145]
Z	2	י 12 לוע	U Iorinating e	J	0	0	Fold : unknown (predicted: $all-\alpha$).
0	0	0	0	0	0	0	Gene encoding fluorinating enzyme was identified in <i>Streptomyces</i> <i>cattleya</i> but not yet in plants synthesizing fluorinated metabolites. Fold: Rossmann-like and β-barrel.
			13. QueA				
0	0	0	Ī	0	0	I	QueA (tRNA ribosyltransferase – isomerase) homologs are found only in bacteria. Fold: QueA-like (TIM-barrel and β- barrel).
		14. SA	M decarbo	xylase			
Ι	4	I	2	?	2	?	SAM decarboxylase activity has been purified from all three domains of Life. However <i>M. tuberculosis</i> and <i>H. pylori</i> lack clearly identifiable homologue of this enzyme. Fold: SAM-decarboxylase.
2		15.9	SAM synthe	tase			
2	4 (5)	Z	I	I	I	I	at the sequence level and have very similar structures [73]. SAM synthase is involved in development and abiotic stress tolerance in plants and have complex expression pattern [146, 147]. Fold:

SAM-synthetase.

		16.	ACC syntha	ase			
2	12	0	0	0	0	0	ACC is the precursor of important plant hormone – ethylene. Human homologs have different function. Fold: PLP-dependent transferases.
	1 7. N	N-acyl-hom	noserine lac	tone syntha	se		
0	0	0	2	0	0	0	Found in some bacteria only. Fold: GNAT.
		18.	Met repress	sor			
0	0	0	0	0	0	0	Found in some enterobacteria and gamma proteobacteria only. Fold: ribbon-helix-helix.
		19.	CBS doma	in			
5	10	5	7	4	9	3	Numbers of SAM-binding CBS domains presented here are approximate because of strong similarity to CBS domains binding other adenosine derivatives. Fold: CBS-domain.
		20.	DAPA synth	ase			
0	Ι	I	6	2	0	I	Biotin biosynthesis is unique to plants, some fungi and most bacteria. Fold: PLP-dependent transferases.
		21. SAM	-dependent	radical			•
9 (13)	13 (14)	5	18	11	24	11	Human lacks SAM-dependent radical enzyme from biotin and thiamine biosynthetic pathways. Those enzymes generate highly oxidizing 5'-deoxyadenosyl radical in an anaerobic reducing environment, and utilize this radical as catalytic and stoichiometric oxidant in many different enzymatic reactions [148]. Those enzymes are essential for anaerobic growth. Fold: TIM-barrel.

Table 3: Counts of each family in selected genomes (Continued)

Number of genes and proteins encoded by selected genomes were counted for each SAM-binding protein family. **SULSO** – Sulfolobus solfataricus (Archaea), **PSEAE** – Pseudomonas aeruginosa (Gamma proteobacteria), **HELPY** – Helicobacter pylori J99 (Proteobacteria), **MYCTU** – Mycobacterium tuberculosis (Actinobacteria), **YEAST** – Saccharomyces cerevisiae (Eukaryota), **ARATH** – Arabidopsis thaliana (Eukaryota), **HUMAN** – Homo sapiens (Eukaryota).

tRNA is methylated by Class I methyltransferase in bacteria and by class IV enzyme in archaea [114], suggesting either parallel evolution of different molecular solutions for the same task, or functional takeover by an unrelated gene.

Phyletic distribution and phylogeny of SAM-binding proteins suggests multiple roles for SAM in the Last Universal Common Ancestor

SAM-binding proteins perform an unprecedented variety of chemical reactions, and belong to about 22 distinct sequence and structural families of proteins. It has been noted that some of these groups of proteins (most notably, Rossmann-fold methyltransferases) are extremely ancient and were more than likely represented by multiple paralogs in the life forms predating the divergence of bacteria, archaea, and eukarya [5]. On the other hand, some of the SAM-binding proteins appear to have been invented later in evolution. Such scenarios have been documented, for example, for SET-domain methyltransferase family, which is thought to have emerged in early eukaryotes and then passed to a few bacteria by lateral gene transfer, while greatly expanding in size in multicellular eukaryotes [71], and are also likely for Met repressor, which is a member of the large ligand-binding family essentially confined to bacteria [115]. A complex picture of gains, losses, and lineage-specific expansions of genes coding for SAM-binding proteins is illustrated in Table 3.

We performed a more detailed examination of evolutionary trajectories of SAM-binding proteins, using information from the NCBI COG database. Each COG is a set of orthologous genes in completely sequenced genomes, along with lineage-specific paralogs. A COG is characterized by phyletic pattern, which is the set of genomes that has at least one member of this COG, and by sequencebased phylogenetic tree of COG members [116]. This information can be used, in conjunction with the consensus phylogenetic tree of the completely sequence genomes, to infer the presence or absence of a COG in an ancestral life form. We used a relatively conservative estimate, allowing for occasional horizontal gene transfer and demanding a complete agreement between species' tree, gene family tree, and phyletic pattern (see Methods for more detail). Under these conditions, the SAM-binding complement of LUCA proteins consists of 9 ancestral COGs (Table 2).

Although 6 of 9 COGs in LUCA represent proteins with Rossmann-like fold, the remaining three folds are all different. It should be noted that the set of 9 COGs is almost certainly an underestimation of the SAM-binding proteome in LUCA. Evolutionary model allowing more frequent horizontal gene transfer and/or non-orthologous gene displacements, as well as slight disagreement between different lines of evidence, will increase the list of COGs that can be placed in LUCA genome. The specificity of these additional enzymes varies, although the modification of translation apparatus continues to figure prominently in the increased set; the diversity of folds, however, appears to increase only slightly, mostly due to addition of the SPOUT fold.

An already-diverse group of Rossmann-like SAM-dependent transferases, a variation of TIM-barrel, and two unique $\alpha\beta\beta\alpha$ architectures thus appear to represent the best-supported ancestral set of SAM-binding proteins. Although less diverse than the present-day variety of SAM-binding proteins, this set is far from simple. Interestingly, it consists of proteins with α/β architecture and is depleted of all- α and all- β proteins, as seems to be the case for other categories of ancestral enzymes [117] and perhaps nonenzymatic proteins too [118].

The substrates and molecular functions of many of these enzymes are hard to ascertain. Ancient enzymes may have had broader specificities than their present-day descendants [119], but several pathways nevertheless emerge from the analysis of the SAM-binding proteome of the LUCA (Table 2). The common ancestor of bacteria, archaea and eukarya appears to have been able to synthesize SAM de novo, from ATP and methionine; to use it for methylation of RNA bases and, probably, proteins such as translation factors; to decarboxylate SAM; and to synthesize polyamines with the aid of dcSAM. In addition, LUCA had the capacity for generating SAM radicals.

Slightly less restrictive evolutionary model enlarges the set of SAM-binding proteins in LUCA, mostly by increasing the number of paralogs in Class I MT and SAM-radical families, but also by supplementing the set with SPOUT methyltransferases. Further sequence and structure comparison may provide for more detailed understanding of these ancestors, perhaps even to the point of reconstructing the ancestral sequences and studying the ancient SAMbinding proteins in the laboratory.

Conclusion

There are 15 distinct superfamilies of SAM-binding proteins, at least 5 of which may have been represented in the last common ancestor.

Methods

Analysis of multiple sequences in the batch mode was handled using the SEALS package [120].

Iterative database searches with position-specific weight matrices (PSSMs) were performed using the PSI-BLAST

program, with the expectation value for inclusion into the PSSM (-h parameter) set at 0.01, unless otherwise indicated [121]. Additional profile searches were carried out using hidden Markov models generated from alignments of protein domains using the hmmsearch program of the HMMER2 package [122].

Multiple alignments of protein sequences were constructed in an iterative fashion, alternating between sequence and structure alignments [123]. Structural alignments of representative structures from the SCOP families [107] were produced using CE-MC [124] and DALI [125]. The muscle program [126] was used to refine all alignments.

The 3-dimensional structures of proteins were manipulated using the Rasmol program and ribbon diagrams were drawn using the PyMOL program [127].

The Metaserver approach [56] was used to interrogate the network of programs that perform secondary structure prediction and tertiary fold-recognition of proteins by a variety of probabilistic matching and energy calculations (here we used predictions from: PSIPRED [128], PROFsec [129], SAM-T02 [130], FFAS03 [131], 3DPSSM [132], BIOINBGU [133], ORFeus [134], FUGUE [135], and Pcons server [136]).

Topology diagrams were created by using TOPS [137].

The inference of the ancestral presences/absences of the COGs has been done by Mirkin et al. [138]; their model was used in this study, with modifications described in Mushegian [118].

List of abbreviations used

SAM: S-adenosylmethionine

PDB: Protein Data Bank

SCOP: Structural Classification of Proteins

Authors' contributions

ARM: initiated and supervised the study, wrote parts of the manuscript and performed the analysis of phyletic distribution and phylogeny. PZK: participated in the design of the study, wrote parts of the manuscript and performed the large scale sequence and structure comparison. Both authors contributed to the writing of the manuscript, interpreted all data, and approved the final version.

Acknowledgements

We thank G. Glazko for help with phyletic patterns analysis, M. Coleman and D. Thomasset for programming help. Additionally, the authors thank J.M. Bujnicki who as a non-anonymous reviewer provided insightful comments and suggestions. This work was supported by Stowers Institute for Medical Research

24. Bugl H, Fauman EB, Staker BL, Zheng F, Kushner SR, Saper MA, Bardwell JC, Jakob U: RNA methylation under heat shock control. Mol Cell 2000, 6:349-360.

References

- Waddell TG, Eilders LL, Patel BP, Sims M: Prebiotic methylation and Ι. the evolution of methyl transfer reactions in living cells. Orig Life Evol Biosph 2000, 30:539-548.
- Thomas DJ, Waters SB, Styblo M: Elucidating the pathway for 2. arsenic methylation. Toxicol Appl Pharmacol 2004, 198:319-326.
- Wuosmaa AM, Hager LP: Methyl chloride transferase: a carbo-3 cation route for biosynthesis of halometabolites. Science 1990, 249:160-162.
- 4. Saxena D, Aouad S, Attieh J, Saini HS: Biochemical characterization of chloromethane emission from the wood-rotting fun-Appl Environ Microbiol 1998, gus Phellinus pomaceus. 64:2831-2835.
- 5. Anantharaman V, Koonin EV, Aravind L: Comparative genomics and evolution of proteins involved in RNA metabolism. Nucleic Acids Res 2002, 30:1427-1464.
- Hopper AK, Phizicky EM: tRNA transfers to the limelight. Genes 6. Dev 2003, 17:162-180.
- 7. Kouzarides T: Histone methylation in transcriptional control. Curr Opin Genet Dev 2002, 12:198-209.
- Aravind L, Koonin EV: Novel predicted RNA-binding domains 8. associated with the translation machinery. J Mol Evol 1999, 48:291-302
- Romano JD, Michaelis S: Topological and mutational analysis of 9. Saccharomyces cerevisiae Stel 4p, founding member of the isoprenylcysteine carboxyl methyltransferase family. Mol Biol Cell 2001, 12:1957-1971.
- 10. Anderson JL, Frase H, Michaelis S, Hrycyna CA: Purification, functional reconstitution, and characterization of the Saccharomyces cerevisiae isoprenylcysteine carboxylmethyltransferase Stel 4p. J Biol Chem 2005, 280:7336-7345
- 11. Peterson YK, Winter-Vann AM, Casey PJ: Icmt. AfCS-Nature Molecule Pages 2005, : [http://www.signaling-gateway.org/molecule/ query?afcsid=A001154]. doi:10.1038/mp.a001154.01
- Schubert HL, Blumenthal RM, Cheng X: Many paths to methyl-12. transfer: a chronicle of convergence. Trends Biochem Sci 2003, 28:329-335
- 13. Fontecave M, Atta M, Mulliez E: S-adenosylmethionine: nothing goes to waste. Trends Biochem Sci 2004, 29:243-249.
- Lesk AM: NAD-binding domains of dehydrogenases. Curr Opin Struct Biol 1995, 5:775-783.
- Wolf YI, Brenner SE, Bash PA, Koonin EV: Distribution of protein 15. folds in the three superkingdoms of life. Genome Res 1999, 9:17-26.
- 16. Martin JL, McMillan FM: SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. Curr Opin Struct Biol 2002, 12:783-793
- 17. Bujnicki JM: Comparison of protein structures reveals monophyletic origin of the AdoMet-dependent methyltransferase family and mechanistic convergence rather than recent differentiation of N4-cytosine and N6-adenine DNA methylation. In Silico Biol 1999, 1:175-182.
- 18. Posfai J, Bhagwat AS, Posfai G, Roberts RJ: Predictive motifs derived from cytosine methyltransferases. Nucleic Acids Res 1989, 17:2421-2435.
- Fauman EB, Blumenthal RM, Cheng X: Structure and evolution of 19. AdoMet-dependent MTases. In In S-Adenosylmethionine-dependent methyltransferases: structures and functions Edited by: Cheng X and Blumenthal RM. Singapore, World Scientific Inc.; 1999:1-38.
- 20. Bujnicki JM: Sequence permutations in the molecular evolution of DNA methyltransferases. BMC Evol Biol 2002, 2:3.
- Sankpal UT, Rao DN: Mutational analysis of conserved residues 21. in Hhal DNA methyltransferase. Nucleic Acids Res 2002, 30:2628-2638.
- 22. Cheng X: Structure and function of DNA methyltransferases. Annu Rev Biophys Biomol Struct 1995, 24:293-318.
- Kumar S, Cheng X, Klimasauskas S, Mi S, Posfai J, Roberts RJ, Wilson 23. GG: The DNA (cytosine-5) methyltransferases. Nucleic Acids Res 1994, 22:1-10.

- Feder M, Pas J, Wyrwicz LS, Bujnicki JM: Molecular phylogenetics 25. of the RrmJ/fibrillarin superfamily of ribose 2'-O-methyltransferases. Gene 2003, 302:129-138.
- 26. Liu J, Mushegian A: Three monophyletic superfamilies account for the majority of the known glycosyltransferases. Protein Sci 2003, 12:1418-1431.
- 27. Cheek S, Zhang H, Grishin NV: Sequence and structure classifi-
- cation of kinases. J Mol Biol 2002, **320**:855-881. 28. Unligil UM, Zhou S, Yuwaraj S, Sarkar M, Schachter H, Rini JM: **X-ray** crystal structure of rabbit N-acetylglucosaminyltransferase I: catalytic mechanism and a new protein superfamily. EMBO / 2000, 19:5269-5280.
- 29. Qasba PK, Ramakrishnan B, Boeggeman E: Substrate-induced conformational changes in glycosyltransferases. Trends Biochem Sci 2005, 30:53-62.
- 30. Nes WD, Marshall JA, Jia Z, Jaradat TT, Song Z, Jayasimha P: Active site mapping and substrate channeling in the sterol methyltransferase pathway. J Biol Chem 2002, 277:42549-42556.
- 31. Zubieta C, He XZ, Dixon RA, Noel JP: Structures of two natural product methyltransferases reveal the basis for substrate specificity in plant O-methyltransferases. Nat Struct Biol 2001, 8:271-279
- 32. Huang CC, Smith CV, Glickman MS, Jacobs JWR, Sacchettini JC: Crystal structures of mycolic acid cyclopropane synthases from Mycobacterium tuberculosis. J Biol Chem 2002, 277:11559-11569.
- Takahashi M, Terada Y, Nakai I, Nakanishi H, Yoshimura E, Mori S, 33. Nishizawa NK: Role of nicotianamine in the intracellular delivery of metals and plant reproductive development. Plant Cell 2003, 15:1263-1280.
- 34. Dela Vega AL, Delcour AH: Polyamines decrease Escherichia coli outer membrane permeability. J Bacteriol 1996, 178:3715-3721
- 35. Nikaido H: Molecular basis of bacterial outer membrane permeability revisited. Microbiol Mol Biol Rev 2003, 67:593-656.
- Korolev S, Ikeguchi Y, Skarina T, Beasley S, Arrowsmith C, Edwards A, Joachimiak A, Pegg AE, Savchenko A: The crystal structure of 36. spermidine synthase with a multisubstrate adduct inhibitor. Nat Struct Biol 2002, 9:27-31.
- Jansson A, Koskiniemi H, Erola A, Wang J, Mantsala P, Schneider G, 37. Niemi J: Aclacinomycin 10-hydroxylase is a novel substrateassisted hydroxylase requiring S-adenosyl-L-methionine as cofactor. J Biol Chem 2005, 280:3636-3644.
- McCulloch V, Shadel GS: Human mitochondrial transcription 38. factor BI interacts with the C-terminal activation region of h-mtTFA and stimulates transcription independently of its RNA methyltransferase activity. Mol Cell Biol 2003, 23:5816-5824.
- Bujnicki JM: In silico analysis of the tRNA:mIA58 methyltrans-39 ferase family: homology-based fold prediction and identification of new members from Eubacteria and Archaea. FEBS Lett 2001, 507:123-127.
- Anderson J, Phan L, Hinnebusch AG: The Gcd10p/Gcd14p complex is the essential two-subunit tRNA(I-methyladenosine) methyltransferase of Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 2000, 97:5173-5178.
- 41. Bujnicki JM, Feder M, Radlinska M, Blumenthal RM: Structure prediction and phylogenetic analysis of a functionally diverse family of proteins homologous to the MT-A70 subunit of the human mRNA:m(6)A methyltransferase. J Mol Evol 2002, 55:431-444.
- Bujnicki JM, Rychlewski L: RNA:(guanine-N2) methyltrans-42. ferases RsmC/RsmD and their homologs revisited-bioinformatic analysis and prediction of the active site based on the uncharacterized Mj0882 protein structure. BMC Bioinformatics 2002, 3:10.
- Chen ZX, Mann JR, Hsieh CL, Riggs AD, Chedin F: Physical and 43. functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. J Cell Biochem 2005, 95:902-917.
- 44. Gowher H, Liebert K, Hermann A, Xu G, Jeltsch A: Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B

DNA-(cytosine-C5)-methyltransferases by Dnmt3L. J Biol Chem 2005, 280:13341-13348.

- 45. Bourc'his D, Bestor TH: Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature 2004, 431:96-99.
- Dong C, Huang F, Deng H, Schaffrath C, Spencer JB, O'Hagan D, Naismith JH: Crystal structure and mechanism of a bacterial fluorinating enzyme. *Nature* 2004, 427:561-565.
- Anantharaman V, Koonin EV, Aravind L: SPOUT: a class of methyltransferases that includes spoU and trmD RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases. J Mol Microbiol Biotechnol 2002, 4:71-75.
- Jackman JE, Montange RK, Malik HS, Phizicky EM: Identification of the yeast gene encoding the tRNA mIG methyltransferase responsible for modification at position 9. RNA 2003, 9:574-585.
- Ginalski K, von Grotthuss M, Grishin NV, Rychlewski L: Detecting distant homology with Meta-BASIC. Nucleic Acids Res 2004, 32:W576-81.
- Ahn HJ, Kim HW, Yoon HJ, Lee BI, Suh SW, Yang JK: Crystal structure of tRNA(mIG37)methyltransferase: insights into tRNA recognition. EMBO J 2003, 22:2593-2603.
- Gerstein M: A structural census of genomes: comparing bacterial, eukaryotic, and archaeal genomes in terms of protein structure. *J Mol Biol* 1997, 274:562-576.
- Nagano N, Orengo CA, Thornton JM: One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. J Mol Biol 2002, 321:741-765.
- 53. Sofia HJ, Chen G, Hetzler BG, Reyes-Spindola JF, Miller NE: Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. Nucleic Acids Res 2001, 29:1097-1106.
- Ollagnier-de Choudens S, Sanakis Y, Hewitson KS, Roach P, Munck E, Fontecave M: Reductive cleavage of S-adenosylmethionine by biotin synthase from Escherichia coli. J Biol Chem 2002, 277:13449-13454.
- 55. Nicolet Y, Drennan CL: AdoMet radical proteins-from structure to evolution-alignment of divergent protein sequences reveals strong secondary structure element conservation. *Nucleic Acids Res* 2004, **32:**4015-4025.
- Ginalski K, Elofsson A, Fischer D, Rychlewski L: 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics 2003, 19:1015-1018.
- 57. Marton LJ, Pegg AE: Polyamines as targets for therapeutic intervention. Annu Rev Pharmacol Toxicol 1995, 35:55-91.
- 58. Gerner EW, Meyskens FLJ: **Polyamines and cancer: old molecules, new understanding.** *Nat Rev Cancer* 2004, **4:**781-792.
- 59. Kim AD, Graham DE, Seeholzer SH, Markham GD: S-Adenosylmethionine decarboxylase from the archaeon Methanococcus jannaschii: identification of a novel family of pyruvoyl enzymes. J Bacteriol 2000, 182:6667-6672.
- Sekowska A, Coppee JY, Le Caer JP, Martin-Verstraete I, Danchin A: S-adenosylmethionine decarboxylase of Bacillus subtilis is closely related to archaebacterial counterparts. *Mol Microbiol* 2000, 36:1135-1147.
- Ekstrom JL, Mathews II, Stanley BA, Pegg AE, Ealick SE: The crystal structure of human S-adenosylmethionine decarboxylase at 2.25 A resolution reveals a novel fold. Structure Fold Des 1999, 7:583-595.
- Toms AV, Kinsland C, McCloskey DE, Pegg AE, Ealick SE: Evolutionary links as revealed by the structure of Thermotoga maritima S-adenosylmethionine decarboxylase. J Biol Chem 2004, 279:33837-33846.
- 63. Lu ZJ, Markham GD: Catalytic properties of the archaeal S-adenosylmethionine decarboxylase from Methanococcus jannaschii. J Biol Chem 2004, 279:265-273.
- Bennett EM, Ekstrom JL, Pegg AE, Ealick SE: Monomeric S-adenosylmethionine decarboxylase from plants provides an alternative to putrescine stimulation. *Biochemistry* 2002, 41:14509-14517.
- 65. Trievel RC, Beach BM, Dirk LM, Houtz RL, Hurley JH: Structure and catalytic mechanism of a SET domain protein methyltransferase. *Cell* 2002, 111:91-103.

- Yokoyama A, Wang Z, Wysocka J, Sanyal M, Aufiero DJ, Kitabayashi I, Herr W, Cleary ML: Leukemia proto-oncoprotein MLL forms a SET I-like histone methyltransferase complex with menin to regulate Hox gene expression. Mol Cell Biol 2004, 24:5639-5649.
- Kouskouti A, Scheer E, Staub A, Tora L, Talianidis I: Gene-specific modulation of TAF10 function by SET9-mediated methylation. Mol Cell 2004, 14:175-182.
- Chuikov S, Kurash JK, Wilson JR, Xiao B, Justin N, Ivanov GS, McKinney K, Tempst P, Prives C, Gamblin SJ, Barlev NA, Reinberg D: Regulation of p53 activity through lysine methylation. Nature 2004, 432:353-360.
- Xiao B, Jing C, Wilson JR, Walker PA, Vasisht N, Kelly G, Howell S, Taylor IA, Blackburn GM, Gamblin SJ: Structure and catalytic mechanism of the human histone methyltransferase SET7/9. Nature 2003, 421:652-656.
- Manzur KL, Farooq A, Zeng L, Plotnikova O, Koch AW, Sachchidanand, Zhou MM: A dimeric viral SET domain methyltransferase specific to Lys27 of histone H3. Nat Struct Biol 2003, 10:187-196.
- Aravind L, Iyer LM: Provenance of SET-domain histone methyltransferases through duplication of a simple structural unit. Cell Cycle 2003, 2:369-376.
- 72. Iyer LM, Aravind L: The emergence of catalytic and structural diversity within the beta-clip fold. *Proteins* 2004, **55**:977-991.
- Sanchez-Perez GF, Bautista JM, Pajares MA: Methionine adenosyltransferase as a useful molecular systematics tool revealed by phylogenetic and structural analyses. J Mol Biol 2004, 335:693-706.
- Komoto J, Yamada T, Takata Y, Markham GD, Takusagawa F: Crystal structure of the S-adenosylmethionine synthetase ternary complex: a novel catalytic mechanism of S-adenosylmethionine synthesis from ATP and Met. *Biochemistry* 2004, 43:1821-1831.
- Graham DE, Bock CL, Schalk-Hihi C, Lu ZJ, Markham GD: Identification of a highly diverged class of S-adenosylmethionine synthetases in the archaea. J Biol Chem 2000, 275:4055-4059.
- Huai Q, Xia Y, Chen Y, Callahan B, Li N, Ke H: Crystal structures of I-aminocyclopropane-I-carboxylate (ACC) synthase in complex with aminoethoxyvinylglycine and pyridoxal-5'phosphate provide new insight into catalytic mechanisms. J Biol Chem 2001, 276:38210-38216.
- 77. Jakubowicz M: Structure, catalytic activity and evolutionary relationships of I-aminocyclopropane-I-carboxylate synthase, the key enzyme of ethylene synthesis in higher plants. Acta Biochim Pol 2002, **49**:757-774.
- Capitani G, Eliot AC, Gut H, Khomutov RM, Kirsch JF, Grutter MG: Structure of I-aminocyclopropane-I-carboxylate synthase in complex with an amino-oxy analogue of the substrate: implications for substrate binding. *Biochim Biophys Acta* 2003, 1647:55-60.
- 79. Tsuchisaka A, Theologis A: Heterodimeric interactions among the I-amino-cyclopropane-I-carboxylate synthase polypeptides encoded by the Arabidopsis gene family. Proc Natl Acad Sci U S A 2004, 101:2275-2280.
- Eliot AC, Kirsch JF: Avoiding the road less traveled: how the topology of enzyme-substrate complexes can dictate product selection. Acc Chem Res 2003, 36:757-765.
- Feng L, Geck MK, Eliot AC, Kirsch JF: Aminotransferase activity and bioinformatic analysis of I-aminocyclopropane-I-carboxylate synthase. *Biochemistry* 2000, 39:15242-15249.
- 82. Fuqua C, Parsek MR, Greenberg EP: **Regulation of gene expres**sion by cell-to-cell communication: acyl-homoserine lactone quorum sensing. Annu Rev Genet 2001, **35**:439-468.
- Redfield RJ: Is quorum sensing a side effect of diffusion sensing? Trends Microbiol 2002, 10:365-370.
- 84. Lerat E, Moran NA: The evolutionary history of quorum-sensing systems in bacteria. *Mol Biol Evol* 2004, 21:903-913.
- Val DL, Cronan JEJ: In vivo evidence that S-adenosylmethionine and fatty acid synthesis intermediates are the substrates for the LuxI family of autoinducer synthases. J Bacteriol 1998, 180:2644-2651.
- Watson WT, Minogue TD, Val DL, von Bodman SB, Churchill ME: Structural basis and specificity of acyl-homoserine lactone signal production in bacterial quorum sensing. *Mol Cell* 2002, 9:685-694.

- 87. Gould TA, Schweizer HP, Churchill ME: Structure of the Pseudomonas aeruginosa acyl-homoserinelactone synthase Lasl. Mol Microbiol 2004, 53:1135-1146.
- He YY, Stockley PG, Gold L: In vitro evolution of the DNA binding sites of Escherichia coli methionine repressor, MetJ. J Mol Biol 1996, 255:55-66.
- Saint-Girons I, Belfaiza J, Guillou Y, Perrin D, Guiso N, Barzu O, Cohen GN: Interactions of the Escherichia coli methionine repressor with the metF operator and with its corepressor, S-adenosylmethionine. J Biol Chem 1986, 261:10936-10940.
- 90. Cooper A, McAlpine A, Stockley PG: Calorimetric studies of the energetics of protein-DNA interactions in the E. coli methionine repressor (MetJ) system. FEBS Lett 1994, 348:41-45.
- 91. Rafferty JB, Somers WS, Saint-Girons I, Phillips SE: Three-dimensional crystal structures of Escherichia coli met repressor with and without corepressor. *Nature* 1989, **341**:705-710.
- Somers WS, Phillips SE: Crystal structure of the met repressoroperator complex at 2.8 A resolution reveals DNA recognition by beta-strands. *Nature* 1992, 359:387-393.
- 93. Phillips K, Phillips SE: Electrostatic activation of Escherichia coli methionine repressor. *Structure* 1994, **2**:309-316.
- Parsons ID, Persson B, Mekhalfia A, Blackburn GM, Stockley PG: Probing the molecular mechanism of action of co-repressor in the E. coli methionine repressor-operator complex using surface plasmon resonance (SPR). Nucleic Acids Res 1995, 23:211-216.
- Aravind L, Koonin EV: DNA-binding proteins and evolution of transcription regulation in the archaea. Nucleic Acids Res 1999, 27:4658-4670.
- Scott JW, Hawley SA, Green KA, Anis M, Stewart G, Scullion GA, Norman DG, Hardie DG: CBS domains form energy-sensing modules whose binding of adenosine ligands is disrupted by disease mutations. J Clin Invest 2004, 113:274-284.
- 97. Kemp BE: Bateman domains and adenosine derivatives form a binding contract. J Clin Invest 2004, 113:182-184.
- Banerjee R, Zou CG: Redox regulation and reaction mechanism of human cystathionine-beta-synthase: a PLP-dependent hemesensor protein. Arch Biochem Biophys 2005, 433:144-156.
- McLean JE, Hamaguchi N, Belenky P, Mortimer SE, Stanton M, Hedstrom L: Inosine 5'-monophosphate dehydrogenase binds nucleic acids in vitro and in vivo. *Biochem J* 2004, 379:243-251.
- Koonin EV, Wolf YI, Karev GP: The structure of the protein universe and genome evolution. Nature 2002, 420:218-223.
- 101. Stroupe ME, Leech HK, Daniels DS, Warren MJ, Getzoff ED: CysG structure reveals tetrapyrrole-binding features and novel regulation of siroheme biosynthesis. Nat Struct Biol 2003, 10:1064-1073.
- 102. Vevodova J, Graham RM, Raux E, Schubert HL, Roper DI, Brindley AA, Ian Scott A, Roessner CA, Stamford NP, Stroupe ME, Getzoff ED, Warren MJ, Wilson KS: Structure/function studies on a S-ade-nosyl-L-methionine-dependent uroporphyrinogen III C methyltransferase (SUMT), a key regulatory enzyme of tetrapyrrole biosynthesis. *J Mol Biol* 2004, 344:419-433.
- 103. Dixon MM, Huang S, Matthews RG, Ludwig M: The structure of the C-terminal domain of methionine synthase: presenting S-adenosylmethionine for reductive methylation of B12. Structure 1996, 4:1263-1275.
- 104. Bandarian V, Pattridge KA, Lennon BW, Huddler DP, Matthews RG, Ludwig ML: Domain alternation switches B(12)-dependent methionine synthase to the activation conformation. Nat Struct Biol 2002, 9:53-56.
- 105. Jarrett JT, Huang S, Matthews RG: Methionine synthase exists in two distinct conformations that differ in reactivity toward methyltetrahydrofolate, adenosylmethionine, and flavodoxin. Biochemistry 1998, 37:5372-5382.
- 106. England JL, Shakhnovich BE, Shakhnovich El: Natural selection of more designable folds: a mechanism for thermophilic adaptation. Proc Natl Acad Sci U S A 2003, 100:8727-8731.
- 107. Murzin AG, Brenner SE, Hubbard T, Chothia C: SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995, 247:536-540.
- 108. Teichmann SA, Park J, Chothia C: Structural assignments to the Mycoplasma genitalium proteins show extensive gene duplications and domain rearrangements. Proc Natl Acad Sci U S A 1998, 95:14658-14663.

- 109. Gough J, Karplus K, Hughey R, Chothia C: Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. J Mol Biol 2001, 313:903-919.
- 110. Madera M, Vogel C, Kummerfeld SK, Chothia C, Gough J: The SUPERFAMILY database in 2004: additions and improvements. Nucleic Acids Res 2004, 32:D235-9.
- 111. Muller A, MacCallum RM, Sternberg MJ: Structural characterization of the human proteome. Genome Res 2002, 12:1625-1641.
- 112. Keller JP, Smith PM, Benach J, Christendat D, deTitta GT, Hunt JF: The crystal structure of MT0146/CbiT suggests that the putative precorrin-8w decarboxylase is a methyltransferase. Structure (Camb) 2002, 10:1475-1487.
- 113. Min J, Feng Q, Li Z, Zhang Y, Xu RM: Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase. *Cell* 2003, 112:711-723.
- 114. Christian T, Evilia C, Williams S, Hou YM: Distinct origins of tRNA(mIG37) methyltransferase. J Mol Biol 2004, 339:707-719.
- 115. Perez-Rueda E, Collado-Vides J, Segovia L: Phylogenetic distribution of DNA-binding transcription factors in bacteria and archaea. Comput Biol Chem 2004, 28:341-350.
- 116. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA: The COG database: an updated version includes eukaryotes. BMC Bioinformatics 2003, 4:41.
- 117. Anantharaman V, Aravind L, Koonin EV: Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. Curr Opin Chem Biol 2003, 7:12-20.
- 118. Mushegian A: Protein content of minimal and ancestral ribosome. RNA 2005, 11:1400-1406.
- 119. Copley RR, Bork P: Homology among (betaalpha)(8) barrels: implications for the evolution of metabolic pathways. J Mol Biol 2000, 303:627-641.
- Walker DR, Koonin EV: SEALS: a system for easy analysis of lots of sequences. Proc Int Conf Intell Syst Mol Biol 1997, 5:333-339.
- 121. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997, 25:3389-3402.
- 122. Eddy SR: Profile hidden Markov models. Bioinformatics 1998, 14:755-763.
- Sadreyev R, Grishin N: COMPASS: a tool for comparison of multiple protein alignments with assessment of statistical significance. J Mol Biol 2003, 326:317-336.
- 124. Guda C, Lu S, Scheeff ED, Bourne PE, Shindyalov IN: CE-MC: a multiple protein structure alignment server. Nucleic Acids Res 2004, 32:W100-3.
- 125. Holm L, Sander C: Searching protein structure databases has come of age. Proteins 1994, 19:165-173.
- 126. Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004, 32:1792-1797.
- 127. DeLano WL: The PyMOL Molecular Graphics System. 2002, http://www.pymol.org.
- 128. McGuffin LJ, Bryson K, Jones DT: **The PSIPRED protein structure prediction server.** *Bioinformatics* 2000, **16:**404-405.
- Rost B, Yachdav G, Liu J: The PredictProtein server. Nucleic Acids Res 2004, 32:W321-6.
- 130. Karplus K, Karchin R, Draper J, Casper J, Mandel-Gutfreund Y, Diekhans M, Hughey R: Combining local-structure, fold-recognition, and new fold methods for protein structure prediction. Proteins 2003, 53 Suppl 6:491-496.
- 131. Rychlewski L, Jaroszewski L, Li W, Godzik A: Comparison of sequence profiles. Strategies for structural predictions using sequence information. Protein Sci 2000, 9:232-241.
- Kelley LA, MacCallum RM, Sternberg MJ: Enhanced genome annotation using structural profiles in the program 3D-PSSM. J Mol Biol 2000, 299:499-520.
- 133. Fischer D: Hybrid fold recognition: combining sequence derived properties with evolutionary information. Pac Symp Biocomput 2000:119-130.
- 134. Ginalski K, Pas J, Wyrwicz LS, von Grotthuss M, Bujnicki JM, Rychlewski L: ORFeus: Detection of distant homology using sequence profiles and predicted secondary structure. Nucleic Acids Res 2003, 31:3804-3807.

- 135. Shi J, Blundell TL, Mizuguchi K: FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. J Mol Biol 2001, 310:243-257.
- Lundstrom J, Rychlewski L, Bujnicki J, Elofsson A: Pcons: a neuralnetwork-based consensus predictor that improves fold recognition. Protein Sci 2001, 10:2354-2362.
- 137. Michalopoulos I, Torrance GM, Gilbert DR, Westhead DR: TOPS: an enhanced database of protein structural topology. Nucleic Acids Res 2004, 32:D251-4.
- 138. Mirkin BG, Fenner TI, Galperin MY, Koonin EV: Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. *BMC Evol Biol* 2003, 3:2.
- 139. SCOP Superfamily: S-adenosyl-L-methionine-dependent methyltransferases. cgi?sunid=53335]
- 140. UniProt Knowledgebase: list of organism identification codes. [http://www.expasy.org/cgi-bin/speclist]
- 141. Marchler-Bauer A, Anderson JB, DeWeese-Scott C, Fedorova ND, Geer LY, He S, Hurwitz DI, Jackson JD, Jacobs AR, Lanczycki CJ, Liebert CA, Liu C, Madej T, Marchler GH, Mazumder R, Nikolskaya AN, Panchenko AR, Rao BS, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Vasudevan S, Wang Y, Yamashita RA, Yin JJ, Bryant SH: CDD: a curated Entrez database of conserved domain alignments. Nucleic Acids Res 2003, 31:383-387.
- 142. Imai A, Matsuyama T, Hanzawa Y, Akiyama T, Tamaoki M, Saji H, Shirano Y, Kato T, Hayashi H, Shibata D, Tabata S, Komeda Y, Takahashi T: Spermidine synthase genes are essential for survival of Arabidopsis. *Plant Physiol* 2004, 135:1565-1573.
- 143. Imai A, Akiyama T, Kato T, Sato S, Tabata S, Yamamoto KT, Takahashi T: Spermine is not essential for survival of Arabidopsis. FEBS Lett 2004, 556:148-152.
- 144. Chattopadhyay MK, Tabor CW, Tabor H: Spermidine but not spermine is essential for hypusine biosynthesis and growth in Saccharomyces cerevisiae: spermine is converted to spermidine in vivo by the FMS1-amine oxidase. Proc Natl Acad Sci U S A 2003, 100:13869-13874.
- 145. Sapperstein S, Berkower C, Michaelis S: Nucleotide sequence of the yeast STE14 gene, which encodes farnesylcysteine carboxyl methyltransferase, and demonstration of its essential role in a-factor export. *Mol Cell Biol* 1994, 14:1438-1449.
- 146. Gomez-Gomez L, Carrasco P: Differential expression of the Sadenosyl-L-methionine synthase genes during pea development. Plant Physiol 1998, 117:397-405.
- 147. Sanchez-Aguayo I, Rodriguez-Galan JM, Garcia R, Torreblanca J, Pardo JM: Salt stress enhances xylem development and expression of S-adenosyl-L-methionine synthase in lignifying tissues of tomato plants. *Planta* 2004, 220:278-285.
- 148. Jarrett JT: The generation of 5'-deoxyadenosyl radicals by adenosylmethionine-dependent radical enzymes. Curr Opin Chem Biol 2003, 7:174-182.

