FOR THE RECORD

The IclR family of transcriptional activators and repressors can be defined by a single profile

TINO KRELL, ANTONIO JESÚS MOLINA-HENARES, AND JUAN LUIS RAMOS Estación Experimental del Zaidín, Consejo Superior de Investigaciones Científicas, 18008 Granada, Spain (RECEIVED October 13, 2005; FINAL REVISION JANUARY 26, 2006; ACCEPTED JANUARY 28, 2006)

Abstract

In the last decade enormous advances in life sciences have been possible due to the information obtained from DNA sequencing projects. The optimal interpretation and analysis of genome sequence data requires the precise annotation and classification of proteins deduced from open reading frames, which is usually done with the help of family-specific signatures. Here we report a novel profile for the IcIR type of transcriptional activators and repressors. In contrast to profiles for other families of transcriptional regulators, the new IcIR profile is located outside the helix-turn-helix DNA-binding motif. We provide evidence that the new profile is more specific than any of the existing signatures for this family of regulators. More than 500 representatives of this family were identified with this profile. A database on bacterial regulators (http://www.bactregulators.org) was built to compile and regroup the sequences with the aid of the new profile.

Keywords: IclR; transcriptional regulator; family profile

Recent developments in functional genomics and the availability of bacterial DNA chips have revealed that microorganisms are able to alter its transcriptome pattern in response to changing environmental conditions. This involves a series of adaptive responses that are mainly triggered by regulatory proteins (Ramos et al. 2001).

The most recurrent DNA-binding motif for the binding of regulators to their corresponding promoters is a conserved DNA recognition motif that consists of an α -helix, a turn, and a second α -helix (referred to as HTH). The latter helix, termed the "recognition helix," was shown to fit into the DNA major groove (Pabo and Sauer 1992). Among HTH transcriptional regulators, families have been proposed based on common 3D structural motifs, conserved domains, and primary sequences (Nguyen and Saier 1995; Gallegos et al. 1997; Rigali et al. 2002; Ramos et al. 2005). Comparative studies have led to the determination of a specific signature for some families of bacterial regulators, and these signatures have made it possible to detect and classify new family members (Schell 1993; Gallegos et al. 1997; Rigali et al. 2002; Busenlehner et al. 2003).

One of the families of bacterial transcriptional regulators is termed IcIR, which has been named after the *Escherichia coli* IcIR protein. This protein controls the glyoxylate shunt and represents the best-characterized member of the family (Nègre et al. 1992; Yamamoto and Ishihama 2003). The specific functions regulated by members of the IcIR family are diverse and include, for example, carbon metabolism in enterobacteriaceae (Yamamoto and Ishihama 2003), degradation of aromatic compounds by soil bacteria (Gerischer et al. 1998), solvent tolerance in *Pseudomonas* (Guazzaroni et al. 2004), inactivation of quorum sensing signals in *Agrobacterium* (Zhang et al. 2004), plant virulence by certain enterobacteriaceae (Reverchon et al. 1991), and sporulation in *Streptomyces* (Jiang and Kendrick 2000).

Reprint requests to: Juan L. Ramos, Estación Experimental del Zaidín, Consejo Superior de Investigaciones Científicas, C/ Prof. Albareda 1, 18008 Granada, Spain; e-mail: jlramos@eez.csic.es; fax: +34-958-135740.

Article published online ahead of print. Article and publication date are at http://www.proteinscience.org/cgi/doi/10.1110/ps.051857206.

Interpro (Mulder et al. 2003) assigns proteins to the IclR family according to the PROSITE profiles PS51077 and PS51078 (Hofmann et al. 1999), the SMART domain SM00346 (Schultz et al. 2000), or the Pfam Hidden Markov Model (HMM) PF01614 (Bateman et al. 2002). The signatures used by PROSITE, SMART, and Pfam to identify IclR differ significantly and are located in different parts of the protein sequence. Pfam PF01614 HMM and PROSITE PS51078 are based on a very large segment of the proteins, the former comprising residues 82-269 and the latter comprising residues 87-272 in the IclR primary sequence, and do not consider the HTH motif. This contrasts with the SMART domain SM00346 and PROSITE profile PS51077. The former uses the HTH region of the protein and a large adjacent fragment up to residue 114 in the E. coli IclR primary sequence, whereas the latter is located between amino acids 24 and 86 in the primary IclR sequence. These differences between signatures do not guarantee an unequivocal identification of family members. Therefore, efforts were made to define a precise profile for the recognition of members of the IclR family of transcriptional regulators, which is reported here. This profile has allowed the identification of >500 members of the IclR family of transcriptional regulators (as of August 2005), which were found to be widely distributed in bacteria. In addition, data on IclR proteins were collected and deposited in our database of bacterial regulator proteins (http://www.bactregulators.org).

Results and Discussion

The first step in the development of the new signature for IclR family members was the selection of a seed containing 53 sequences based on the following two criteria: (1) InterPro entry IPR005471 identifies the protein unequivocally as an IclR family member; (2) the proteins were similar in size, i.e., 240–280 amino acids. BLASTCLUST analysis showed that each of the 53 proteins could be clearly distinguished from each other. The sequences were subsequently aligned with CLUSTAL (http://www.clustalw. genome.jp), which revealed three regions that were particularly well conserved in the multialignments (Fig. 1). One of the conserved regions comprises the HTH DNA binding motif located at the N terminus, a second region covers part of the N-terminal portion of the proteins toward the central region, and the third one corresponds to a segment from the central region of the protein toward the C terminus (see Fig. 1). The conserved regions were progressively extended in both directions until the global score of the multialignment diminished. The resulting alignments of these three regions were used as a seed to construct different conventional profiles, each covering a conserved region (available at http://www.bactregulators.org/docs.php). The profiles were built with the "pfmake" program available at the Swiss Institute of Bioinformatics (http://www.isrec.isb-sib. ch/ftp-server/pftools) (Bucher et al. 1996). The different profiles were confronted against all entries in the SWISS-PROT and TREMBL databases (released July 2005). We found that the profile covering the central region toward the C-terminal end (amino acids 151-229 in E. coli IclR) identified all IclR members recognized as such by PROSITE PS51078 and Pfam PF01614, whereas the profiles based on other segments of the protein had a reduced discriminatory capacity, and identified not only IclR family members but also regulators unequivocally ascribed to other families. A profile based on the combination of any of the conserved regions was found to be less precise than the profile that was based solely on the central region toward the Cterminal end of the multialignment. We thus considered that members of the IclR family are best identified by a profile that does not include the HTH domain of this set of proteins, and that covers a significant portion of the C terminus of the proteins. This contrasts with findings for the AraC/XylS (Gallegos et al. 1997), TetR (Orth et al. 2000; Schumacher et al. 2002; Ramos et al. 2005), and GntR (Rigali et al. 2002) families, which are best defined by a specific profile that includes the HTH DNA binding domain.

The IcIR profile, available at the BacTregulators database (http://www.bactregulators.org/docs.php), was confronted against all prokaryotic proteins in the SWISS-PROT and TrEMBL (SPTR) databases (release 13-8-05) using the "pfsearch" program available at http://www.isrec.isb-sib.ch/ftp-server/pftools (Bucher et al. 1996). The program, which proposes a tentative threshold N-score of 8.5 to consider a protein as member of the IcIR family, selected 546 proteins as putative members of the IcIR family, of which 34 were encoded by plasmids.

To evaluate the specificity (false positives) and sensitivity (false negatives) of the new IclR profile, we used an in-house developed tool termed "Provalidator." Provalidator is a PHP-based tool that assists in the automation of profile construction and validation, and will be available free of charge at http://www.bactregulators.org. Our analysis revealed no apparent false positive proteins. A search in Interpro (Zdobnov and Apweiler 2001), a database containing all currently available classification methods for IclR proteins, assigned 587 proteins to the IclR family. The 41 proteins assigned to the IclR family by Interpro, although not identified with the new profile constructed in this study, were considered as incorrectly assigned to the family. In fact, among these 41 proteins there were three truncated polypeptides (Table 1, proteins 30, 31, and 35) and six polypeptides of reduced size (71– 137 amino acids, namely proteins 23, 26, 30, 32, 34, 36, and 37 in Table 1), which made it unlikely for the latter being part of the IclR family, since our analysis revealed that these polypeptides do not possess an HTH DNA binding domain. The remaining 32 proteins assigned by

Interpro (listed in Table 1) were divided into two groups according to their score with the new profile developed here. A group of 25 proteins (Table 1, proteins 8-39, not considering the above-mentioned small or truncated proteins) yielded N-score values between 2.1 and 6.4. Alignment of these 25 proteins with IclR family members revealed substantial sequence conservation at the HTH DNA binding domain, but less sequence conservation in the C-terminal where the new profile is located (not shown). The reason why Interpro assigns these proteins to the IclR family is because PROSITE PS51077 and SMART SM00346 include the nondiscriminatory HTH region. In agreement with this observation is that, with the exception of protein Q57K18 that is exclusively recognized by SM00346, all proteins listed in Table 1 are recognized by PS51077, which is the other profile including the HTH sequence. Therefore, these proteins should be considered as incorrectly assigned to the IclR

family, since a profile based on the HTH as PROSITE PS51077 lacks the necessary discriminatory potential.

The second group consisted of seven proteins with N-score values between 8.46 and 7.71 (Table 1, proteins 1-7). The alignment of these proteins to IclR family members revealed significant sequence conservation in the fragment spanning the new profile, and thus, it cannot be ruled out that these proteins are IclR family members. However, the N-score threshold of 8.5, as proposed by the "pfsearch" program, cannot be lowered in order to avoid the inclusion of non-IclR proteins. The zone between N-scores of 8.5 and 7.5 is an empirically determined buffer zone where it is recommended to consider the assignment with caution. Sequence annotation is rarely a clear-cut issue, and the purpose of this zone is to prevent the detection of false positives. We consider precision in avoiding false positives more important than the possible exclusion of any family member. Experimental characterization of these proteins will provide support for

	1	10	20	30	40	50	60	70	80	90	100	110
	+	++	+	++	+	-+++	+	+	++ H ₂ T	++ H ₃	·++-	++
P76268					. MANADLE	KOPDSVSSVL	VEGTLOALG.	EER.ETGT	TELSORVMMS	KSTVYRFLOT	MKTLGYV	AOEGESEKYSI
P77300						MPIIOSVEF	ALOILDLFN.	.EOATELKI	TDISKLMGLS	KSTLHSLLKT	LOLHGYI	DONPENGKYRI
Q8EMM2						MSNVQSIEF	AFTILKVLS.	.NNPDGLRI	TVLAENSSLT	KSTTHRITST	LVNLGYV	QQDPETEKYMI
Q9KAZ9						MVKSVDF	ALTIISLVS.	.EHKQGLGV	TDVAAKLSLT	KSSAYKLLAT	LVEHGFI	EQDEETKKYRI
Q9K6K4						MVQSVDF	RALMILDILK.	.QHPKGLGI	TELAHELNVA	KSTAHRLVS1	LETYDYV	TQDRETSRYSI
Q8EPA1						MSVKSAKF	RALDILDLLA.	.KHTNGLTI	KNIGDKLELP(QSSTFNLLKI	LHSEGYV	RQD.SMKLYYI
Q838M2					MK	KETKPYGTVLLF	CAAKIMDCLS.	.LQP.DQTL	QEIAKNTEMT	ASTAIKILET	(LTLIGYV)	QKD.TNKTYRI
Q9WXS0	• • • • • •		• • • • • • • • • •	• • • • • • • • • • • •		MNTLKK	AFEILDFIV.	.KNPGDVSV	SEIAEKFNMS	VSNAYKYMV	/LEEKGFV	LRK.KDKRYVE
P3/6/1			• • • • • • • • • •	MGKEVMGK	KENEMAQEK	ERPAGSQSLFF	GLMLIEILS.	.NYPNGCPL	AHLSELAGLN	KSTVHRLLQC	LQSCGYV	TTAPAAGSYRI
P42968	•••••	• • • • • • • • •	•••••	• • • • • • • • • • • • •		MOEKNKTVVK	SMALLNLFL.	.HKP.SLTL	SELVSLTGMP	KTSVHRMVSS	SLEEMGFL	SRDAS.GAYSI
Q9KBV5	• • • • • • •			• • • • • • • • • • • •		. MQEKNKTVIF	SMNLLNLFK.	TIP.TLTL	QEMVKQSGMP1	KTSVHRMINS	SLEEMGP L	QKRHD.GAILI
OSEL TA	• • • • • •		•••••			IDS DKC T DTLOF	STDIINCET.	FENCEI TI	TETSIVELIA	KISVIKLIGS KCTTTDII CI	PIERMINEV	CUNER. GUINI
P16528				MVAPTPAKR	GRKPAVATA	PATCOVOSLTE	CI.KI.I.FWIA	FSNGSVAL	TELACOACLE	NSTTHRLT	PMOOOGEV	BOUGELCHWAT
P77734				MTEVER	RGRPGOAEE	VAOKGAOALEE	GTATLOYLE	KSGGSSSV	SDISLNLDLPI	I.STTFRI.LKV	/LOAADEV	YODSOLGWWHT
088JP8						MAGSOIEF	AFSLVESLT.	.GEPOGLPL	OTLAERLDIP	KSAAHRMLTH	LVRLGYV	RONRENSRYOI
08U755					MDF	TINVPSGGVEF	TIAMLELLA.	.LAEEPLKL	SDVAHOLDIP	KSACHRILTS	SLIENGWA	WOSPESDCYAL
09RZ01						.MPRTLATVEC	AVRVLEYFD.	.ADHTEWTL	SDLARALALP'	TSTLHEOLST	ILTASGLL	RRTGR.GRYOI
Q9RV13						MLSLQF	ASNILGAFS.	.AEQPEWGV	RALSAHLGVP	RATAHAYLAC	JLTEAGFL	RRTPH.GKYRI
Q88E46					MAKASSPTE)TGKQK <mark>V</mark> RSAEV	GTDILKALA.	.ELSPSTSL	SRLAEHVQMP	ASKVHRYLQA	ALIASGFA	EQDAATNHYGI
Q87Z78				M	TQETESPAG	GRQQK <mark>V</mark> QAAEV	GLGVLKALA.	.ELSPSTSL	SKLAEHLGMPI	PSKVHRYLQ <i>P</i>	ALIASGFA	EQDAVNNH <mark>Y</mark> GI
Q52154	M	SDETLVNDF	PVNPEPARPA:	SAAMAPPIVAS	PAKRIQAFI	'GDPDF <mark>M</mark> TSLAF	RGLAVIQAFQ.	.ERKRHLTI	AQISHRTEIP	RAAVRRCLHI	[LIKLGYA	TSDGRTYSI
Q8U7A8	• • • • • •				MAV	/SERDMMGGLAF	GLRVIEAFS.	.AERPRLSI	SDAAEIAGLD	RATTRRCLLT	ILSELGYA	AYDGKFFT
Q88JP0	•••••		MECONN		MNTPAI	HPRDLIAGLOK	GLALMQLFS.	.AEQPRLSV	PQAARLSGLT	SSAVERFLLI	LVHEGFA	ETDSRDYWI
083046			MWSNM	DDKKVKEEKIL	AUDUCCEET	RHEDEVAGISE	GMAILDSEG.	. TDRHRLNI	TMAAEKTGMT	RAAARRHLLJ	LEILGIL.	ESDGHIFII
Q43992	• • • • • •		•••••	MEQHHQIL.	AULUSSEET VDALUSSEET		GLALLEAFG.	PHDI SI CN	OFITDETCI	KIAARRILLA KATUCOMTAT	FI WNI CVI	DIDERIPWI
OSEX66		MRGRT	CSDASBROF	RSAFMRKPTML	TPLEACVEE	FDROFVTALAR	GLEVI.BCET	PTENTLON	OFTAHKTCLP	KPTVSRI.TH	TURLCYL	RODALSCLYOT
07WBS5	•••••	WRPRPDAAT	AMSSGPPAA	RKPGVVOPALL	NMSAWLRPF	PGLTVVTPLVF	ALVI.TSAFS	POERWIGN	GOLAORTGLPI	PSTVSRTAOS	SLVHLGYL	LHDGAERKYRI
07WBT1				MLSR	FPDPAPAAC	SRRPTVAPFAF	ALSVLASES.	.SRDOWLSN	SEFVARTGLP	ASTVTRMTST	LVTLGYL	HCAPDTRRFCI
08FV53					MDKLHIE	PTAKRSRGLEF	AFEILDYLR.	.VSRAPANP	NEIAARIKAP	RSTVYELVNI	LLSNGVL	EYCDDEGRVY
Q881V2					MTED	LIKRRARGLDF	AFDILDFLK.	.EKARPMRP	NEIASGIGSP	KSTVYELVAS	LLERRIL	ESVGKDGHVY
Q8UAV3					MKA	AKRSRVSGIDF	RALQVIDYLY.	.ETGSPAGA	YAIAKAIKAPI	LSTVYVIVDI	DLVEKTML	ARN.ADGNIWI
P77732					M	ILESSK <mark>V</mark> PALTF	RAIDILNLIA.	.RIG.PCSA	ATIIDTLGIP	KSTAYL <mark>LL</mark> NE	LRRQRFL	SLDHQEN.FCI
P39360						MVRKGCNSLVF	RAEKILTHIA.	.WVG.MASY	MELLNKFQYPI	KSSLLNLLNV	/MVDCGFL	IKN.KNGYYSI
P15360						.MARNIQSLEF	RAAAMLRLLA.	.GGERRLGL	SDIASSLGLA	KGTAHGILRI	LQQEGFV	EQDDASGRYQI
Q7VTS3	• • • • • •		•••••M	SRIPNATARGP	GSHSDDGAG	GQHPIAIQVIDF	RAMRLLDALA.	.AQPEPVTL	KELSATTGLH	ASTAHRILNI	JLVVGRYV	ERVDN.GLYQI
Q/WBT2	• • • • • •		• • • • • • • • • •	MTS	LPDTVDEHL	DSESAALMTVEF	RGLKVLRAFR.	.SDRTPLSN	AELVRRTGLP	KATVSRLTST	I'LLQVGFL	RQVPGRREFEI
066964	• • • • • • •		• • • • • • • • • •		MEH	IQYKDIYIIQNV	HLALEVLFL.	. LEKKPYTF	CDIADCTUID	ROKLOKILTI	LLESEWI	SFDEENKLYRI
Q01C00	• • • • • •			MDDDVVANCCV	DAVDDOUC	VERNARCET CR	ALSIIKLUNS	INEREIWAL	TOWERCOD	V S I V RELLING	JINERGLV	TOTICIKOIKI
088484	• • • • • • •		•••••••••••	MENDERKANGGE	TERVITCE		CI CUUONUNS	CARDI	VETADI TOTT	DOTTUDI ACC	TUFFDVI	DHAADQI.IA
08171117			м	TRISVHYTAVV	PPTEGKSTT	NRDEGTOVI.RE	AVAALDETA	AEPGHLBL	VDLCERLGLA	KSTTRRLLV	LVEVGLV	SVDSHG REAL
08G4W3				MNNSLTESEPV	EDTTAKDDO	FILSGVGVLDE	TVKTLDALE.	SGPSTL	GOLVAATGLAI	RPTAHRLATA	LEBHREV	LEDOHG . REVI
0882U7				M	NNSADRNSA	PAEVGVGAVSF	LFAVLRCLG.	ECGEGGERV	TOLAORVGLS	OPTTHRLLRS	SLMDEGMV	EODLLSKRYRI
08U6I3				MA	TTTTTIPVE	KDLTGARAVDF	ALALLSMVG.	RHAERGVAL	SVIVTESGLN	KPTARRLLL	LIRAGLV	EODEETRRYYI
Q8FUQ1						MSTIGF	ALTLLDVLS.	.RLNTEAGL	TGIAQACGYDI	KATTRRFLVI	LEKHGEV	EQLAESRKYR
P77569	MIFYCAL	LSIGRVFSA	TIKTCPNVH	QVHHVVLTIEM	SINMQNNEQ	TEYKT <mark>V</mark> RGLTF	RGLMLLNMLN.	.KLDGGASV	GLLAELSGLH	RTTVRRLLET	LQEEGYV	RRSPSDDSFRI
Q7WBP5				MAADG	SKTRKEEGO	GAESGGSQVMTK	GLRVLSHVA.	.HCGGEIGV	RELARDVGLP	VAVVHRLVSS	JISELNYL	EKNPESGKYRJ
Q8G4X6		MS	SHASQQYAFT	EPRTVKRAQVK	PSEREHQTS	SSAPR <mark>V</mark> SLAPA	ADRTLDILEF	IASNGQTQA	ATLARELGI PI	RSTVYQLLEI	(LERHGLV	TRLAEQRAYGI
Q7WNC9			M	PSEPAKPTSLP	PPAHAVDGE	DSPLYVQAIAF	RAFQVLQAFD.	.GPRREITL	ADIAKAAGMGI	RSAVQRVVYI	lestgyr	ARVPDSKN <mark>Y</mark> RI
Q8G2T0					.MREPTLET	DDRYRAPALD	GLDILELLA.	.SVDGGLTQ	AEIAKHLDRSI	PNEFYRMLDF	XLVKRGYV	TKLDGD.RYSI
Q8G976					.MSDSEESS	SAKHGGIQVIAF	AASIMRALG.	.SHPQGLSL	AALAQVVDLPI	KSTVQRIINA	ALGAEHLV	EALGPSGGFRI

cont.

Figure 1. (Continued on next page)

	120	130	140	150	160	170	180	190	200	210	220
P76268	TLKLFELGARA	LONVDLIRSA	DIQMRELSRLT	.ETIHLGALI	EDSIVY	IHKID	SMYNLRMYSRI	GRRNPLYSTA	IGKVLLAWR	DRDEVKQILE	GVEYK
Q8EMM2	GNQLIKLTSVM		EPYLVNLSRDVI	I.ETVHLCVES	SNGEVLY	VDKKE	KNQNIRMYSTI	GSRAPLYCTA	VGKVLLSGM	VPVYFEEVIN.	NMTFE
Q9K6K4	GLKFLEMQQVV	MENMDIVAVA	HPILEALTEECO	G.EISHLVMRI	DHFDVVY	IDKLE	AVSTIRIYSRT	GKRAPIHCTS	VGKVIMAYL VGKAIAAHF	DEPLLQEYVE.	QAAFQ
Q8EPA1 Q838M2	GTKLIRYANQS	VEQID <mark>LAEIA</mark>	KPYLADLMKAVI V <mark>P</mark> YLENLQNTVI	C.ETVFMAMLS D.ETIHLGILM	SKDELVY NNNEILY	IAKMD	ADRSIRTTAQP KNQTIRMSSKV	GYKKPIYCTG GITRPLYNSA	LGKTFLAFM MGKAVLAEF	SEEQVQQYLD.	TQTLI
Q9WXS0 P37671	GYKLIEYGSFVI TTKFIAVGQKA	LRRFNIRDIA LSSLNIIHIA	HDHLVDIMKRT(APHLEALNIAT(G.ETVHLILKI G.ETINFSSRE	DGFEGVY EDDHAIL	IDKVEG IYKLEP	E.QSIPMVSRL TTGMLRTRAYI	GMKVDLYSTA GQHMPLYCSA	SGKSILAFV MGKIYMAFG	PEKELKEYLK. HPDYVKSYWES	IVELK SHQHEIQ
P42968 09KBV5	GLVFLEFGQLVA GVVFLELGOLVS	ADRLDIRKIA SDRLDLROVA	KPVMEELCREVI FPIMEALRNOVI	D.EAVQLIMRI	DGNEAIY	VEKIE	GTQTVRLYTAI TLHPVRLFTKV	GRRSPLYAGA GRRAPLYAGA	CARSILSFL CORVILSFL	PREEIEAYIK(PEOEOTAYLE	QTELI RVELI
Q81NS7 08EL-14	GVVFLRFGQLVS	SQRLSVRNIA SOTFELNSLA	IPYMKELRDNLO	G.QAVNLIIQI	GKHRVC	VEKME	GVQPVRVYTAV	GRREPLYAGA	CPRILLSYF	FEEEKQKYVER	ETDFK
P16528	GAHAFMVGSSFI		HPILRNLMEES	G.ETVNMAVLI	QSDHEAII	IDQVQ	CTHLMRMSAPI	GGKLPMHASG	AGKAFLAQL	SEEQVTKLLH	RKGLH
Q88JP8	SAKLVALSFRY	LAS.SGADII	PILDRLAQDS	G.ELVRLGVII	OG. ARQTW	IAKSQG	ARSGLRYDPDM	GRDAPLFYTA	SGHAWLASL	DDEQALQMVL	ROG. IADPD
Q9RZ01	GWRLLKLSSAL	YGSLPWYSLA	HAEMERLARGT	G.LLAFVSVL	QGEQVIC	IARSVQ	GRGGAAVEGET	RFELPPHATA	SGKLLYAYA	GLP	
Q88E46	GREALRVGLAAI		AMPLSQLRDEL	I.ESCFIAVW	GNQGAT	VVSIEP	AVRAVTVVTQI	GSVLPLLSSS	TGLVFAAYL	PERETQELRDI	
Q87278 Q52154	GREALQVGLAS LPKVLTLGHAY	LSSTPLAISA	APWLASLRDELI QPYLDRISDQLI	I.EAANMATLE	GGDDI <mark>L</mark> Y	IARSAT	VERLISVDLSV	GSVLPLLSSS GGRLPAYCTS	MGRILLAAM	AQGETALLRE(DDTSLREY <mark>L</mark> GI	2DLKAR
Q80/A8 Q88JP0	TPKVLRLGTGCI TPKALRIGQAY	LATMPLPKIV VDSAQLPRMLI	QPLLDRLSEEI(R <mark>PIV</mark> EQVARQT(G.QSTSVSILI Q.EHVSVGTRI	DEAEIVY DGDEIIH	VARAA LVRSR	QQRVMSIALMP YSHVAS <mark>L</mark> SIRP	GSRLPAYCTS GSRVPMYCTA	MGRVLLAAQ SGRIWLAWL	PAERRRDILE <i>I</i> DEG <mark>E</mark> RDEYFAH	ASRLVAR RHPLRAL
083046 Q43992	TPKILKFSGSY THRVLRFSSSY	LGGAQLPKIS LSSAHLPKVA	2PLLNLLTTQT: 2SFLNLLCAQT:	S.LIYSVMVLI S.LTFSIVVLI	DGYEAIT DEHEVVP	IARSAAHQQT VARSYLPQQD	.DRVNPYGLHL NLRVSPYGMHL	GNRLPAHATS GNRLPAHATS	AGKILLAYL TGKVLLSVL	DDHAQQEWLNG DREVQIEWIEH	QYPLQRL KYGLKRL
Q8FUX8 Q8FX66	GPATIALGYSGI DIGILRLGYAMI	LSANVVVHIA LSNLMIRTVA	MPLMRKLAEKT(SPLMQVLADYAI	G.VAVAMGLRE (.AAVAMAARI	EQQEMVY DRLSMVY	IANAR	SENPVSLRLNV GEGNMTMRRQI	GSRLPIWKTA GSTLPLAGSS	MGLSYYAGM VGRACLAAM	EESQREALLEH PED <mark>E</mark> RTFILKH	RMLQCEP HIREREP
Q7WBS5 Q7WBT1	APAVLGLGYAA SPSVLALGYGA	IAHSAIQGLA AVDAEIHRPTI	GERMAAFARQHI NOLMRVFAEHH(.VH <mark>VCL</mark> AARI	DRLDL <mark>V</mark> V DRLDLVV	LECRRS MDSCLT	LECPVALPLHV DALPALLOPGV	GMRVGIAQSP GTRLGMASSA	MGWSMLAAL AGWALLACL	PELERCYLLDI PEAERDYLLRA	VVERRMA ARPHAPA
Q8FV53 0881V2	GRKLYFLGAAYI GROLYFLGOAHI	EDHFNFIREC	DRALSKVAEQTI EACLGDIVSOTI	R.ETAQFCMLI R.ETAOMCLLM	DENKYTV NGRKYTV	VRMHE	GARPFRISSDV GERHFRISSDI	GORVPIPWTA GENAPIPWTA	SGRLLLSHL	NDA <mark>EILAFI</mark> PI FDOOIVDLIDI	PEDFQLP
Q8UAV3 P77732	GARLYHYGLAY	ARSLDFMGVA	THEMHDLCRHA	G. ETVQLCGRI	GDHMLV	LAMAD	GPSHFQVASRV	GTRVPLNWTA	SGRLLVGHL	PETERLELFRE	RCARS
P39360 P15360	GIKNYELGCQA	LHRQNIFEVT	KRPMQELSLKS	.LVCHLGAME	SISAIY	LDKIES	PDSVPTSKSWI	GKKLELHITA	LGKALLAWK	TREELDYFLEA	ALTLT
Q7VTS3	GMRLLELGSLV	KGRLNVREAA	IGAMRSLHKQT	G.QTINLSVQQ	2GDEIVY	IDRAWS	ERSGMOVVRAI	GGRAPLHLTS	TGKLFLSTY	DTRLVRAYAL	RTGLA
066964	GIKNFELGVSY	LKHLDIKKASI	KPILEEVAKKVI	C.ENVYLTTR	/SYEVLY	IEKSE	VEREVMILSRY	GRVLPLYASA	SGKIYLSHF	DEKELENYFK	CVKW
Q8UHH6	GLRLLQLAAKA	WSGNDLRNVA	APHLSALQEAT(2.ESVILNIPI	IGGQVTY	LDKMEG	.RHTLRMHSQV	GKTSPAYCTG	VGKAALSLL	PVDELAELAA	GMEFH
Q884B4 Q8VJW7	GERLLGFGS	VTGAY IAAAFI	RPTUERVARATI	DTVHLAVRI DGETVDLSVLI	RGQRMWF	VDQIES	.SYRLRAVSAV	GLRFPLNGTA	NGKALLLDS	DDADAEAALCH	R
Q8G4W3 Q882U7	SIEFFALAAAAA	GEDRLLTAAG GNTGNLRDVVI	RPSLLRLSASL	G.DSLFLLARS	5GFDAIC	LAAVER	PYPIRTFTGDI	GAMLSMEAGS GGRVALGVGQ	GSLAILAFL	DSDRLHQGLR. PED <mark>E</mark> RETVIA	(NLP
Q8U6I3 Q8FUQ1	GEETYVLGSLS: GAEPLRLARIRI	SRRFGLLQMA EARYPFLGTA	QDGLTRISRRSH L <mark>PFV</mark> RQLAERT(E.DSSFLSVRF G.ETVHLSEFS	RDTFALC SNGRLSTIH	LYREEG VEESPR	TWPVRTHALQA AHRVFINV	GFEHPLGIGA GMILPFHATA	GSLAMLAAL SGLAYLAAC	PDAEVESIIA/ ADDVVEAT <mark>V</mark> TH	ANSGL (SLD
P77569 Q7WBP5	TIKVRQLSEGFI GYEAFEVGRTYI	RDEQWISALA LRSARIETCA	APLLGDLLREV PPVLRNVVDAHI	/WPTDVSTLDV R.LNAFLGVME	7DAMVV RDLSVIY	RETTHR LITIQD	FSRLSFHRAMV .YGHYDIRIAP	GRRLPLLKTA GSEVPLATTA	SGLTWLAFC	PEQDRKEL <mark>I</mark> EN ADD <mark>E</mark> ILEK <mark>L</mark> G(4LASRPGDDY QYAEK
Q8G4X6 Q7WNC9	GLKTFELGSAY: TSRLLQFSYNY	SRQHRVSQVA VRNNELIAKA	HPVLARLVDET LPYLQELNRTFI	G.ENGHLAVLE D.ETINLQELI	HGNEIIY DGTEIVL	VIEERA VARFLS	.AHRPPLVSGV .RHLMNIEVAV	GVRLPSHLTA GSRLPAFCTA	TGRAILSAL SGTAILSRL	PRN <mark>Q</mark> VRALYPN PEKECDAILRA	NRQAFT ASRRV
Q8G2T0 Q8G976	TLKLFGLAQLHA GPAFGRLIT	APVRR <mark>L</mark> ASFA QAQTDIISLV	TPFMRELADRSH RPHLIALSEQV	(.QANQLAVFI (ESTCLLSLS)	DRGSV <mark>V</mark> V GEKIYV	IAQQEAP	DYWGISIRV .ERELRVVFPI	GSHISLFDTG GIHVPATAVS	SGHVLLAFR: GGKVLLAEL	SPE <mark>E</mark> REMMIAA SEEAQQALLPI	АНL
	230	240	250	260	270	280	290	300	310	320	330 334
P76268 P77300	230 -++ RSTERTITSTEA TFTPATLASREA	240 ++ ALLPVLDQVRI ALMSALAOTRI	250 ++- EQ <mark>GY</mark> GEDNEEQI EOGYALDSEENI	260 EGLRCIAVP	270 /FDRFGVVI /WNHESRVI	280 ++ A.GLSISFPT A.ALSLSTLT	290 ++ LRFSEERL.QE SRVDDAEL.AN	300 + YVAMLHTAAR FREOLOOAGL	310 + KISAQMGYH ALSRALGYP	320 ++ DYPF	330 334 +
P76268 P77300 Q8EMM2 09KAZ9	230 -++ RSTERTITSTEA TFTPATLASREA LRTALTITSQVI PHTERTITDKE	240 ++ ALLPVLDQVRI ALMSALAQTRI KLRKEIEKVK FEMKELEKIR	250 ++ EQGYGEDNEEQI EQGYALDSEENI LQGYALDNVENI LGGYGYEMEENI	260 EEGLRCIAVPY OGVRCVAVPY EEGIRCIASPI	270 /FDRFGVVI /WNHESRVI /YDYKGNII	280 ++ A.GLSISFPT A.ALSLSTLT A.SFSISGPT A AVSISGPS	290 ++ LRFSEERL.QE SRVDDAEL.AN NRVSIDRVNLE TRISKERL HE	300 + YVAMLHTAAR FREQLQQAGL LVDKVNMTSK LEPLITAIGK	310 + KISAQMGYH ALSRALGYP LISEQFGYI KISOBLGYO	320 + DYPF A KKESI	330 334 ++
P76268 P77300 Q8EMM2 Q9KAZ9 Q9K6K4 Q8EPA1	230 RSTERTITSTE/ TFTPATLASRE/ LRTALTITSQVI PHTERTITDKE: PLTRHTISEPE/ KITANTITDRS:	240 ALLPVLDQVRI ALMSALAQTRI KLRKEIEKVK IFMKELEKIR AFKAALQKVL' OLEEOLOLEO	250 ++- EQGYGEDNEEQI EQGYALDSEENI LQGYALDNVENI LEGYGYEMEENI TDGYAEDNEEHI KOGYSIDDEENI	260 EEGLRCIAVP OGVRCVAVP EEGIRCIASPI EEGIRCIAPI EEGIRCIAPI	270 /FDRFGVVI /WNHESRVI /DYKGNII IFDYQGAIT IFNHLGQVP IFDSOBOLL	280 A.GLSISFPT A.ALSLSTLT A.SFSISGPT A.AVSISGPS Y.AISITGPL A.AVSVAGPK	290 ++ LRFSEERL.QE SRVDDAEL.AN NRVSIDRVNLE IRLSKERL.HE ARMTDARL.QT DRMLKOKE.T	300 ++ YVAMLHTAAR FREQLQQAGL LVDKVNMTSK LRPLIIAIGK LIPRLKQAAA VITNLRDTAK	310 KISAQMGYH ALSRALGYP LISEQFGYI KISQRLGYQ AISRNIGYR LISNGMGYV	320 DYPF KKESI X	330 334
P76268 P77300 Q8EMM2 Q9KAZ9 Q9K6K4 Q8EPA1 Q838M2 Q938M2	230 -++ RSTERTITSTE/ TFTPATLASRE/ LRTALTITSQVI PHTERTITDRS/ PLTRHTISEPE/ KITANTITDRS/ PKTPNTITNDE/ PKTPNTITNDE/	240 ALLPVLDQVRI ALMSALAQTR KLRKEIEKVK FFMKELEKIR AFKAALQKVL QLEEQLQLFQ RLKKELKQVQ	250 +++ EQGYGEDNEEQI EQGYALDSENI LOGYALDNVENI LEGYGYEMEENI TDGYAEDNEEHI KQGYSIDDEENI 2TGVAYDDEEII	260 EEGLRCIAVP COVRCVAVP EEGIRCIASP EEGIRCIAPP EEGIRCIAPP EEGLYCLAAP EEGLYCLAAP	270 /FDRFGVVI /WNHESRVI IYDYKGNII IFDYQGAIT IFNHLGQVP IFDSQRQLL MKD.GEIA	280 ++ A.GLSISFPT A.ALSLSTLT A.SFSISGPT A.AVSISGPS Y.AISITGPL A.AVSVAGPK G.AFSVSMPK G.AFSVSMPK	290 ++ LRFSEERL.QE SRVDDAEL.AN NRVSIDRVNLE IRLSKERL.HE ARMTDARL.QT DRMLKQKET YRLTEENK.TT SKFTFEKIFF	300 + YVAMLHTAAR FREQLQQAGL LVDKVNMTSK LRPLI IAIGK LIPRLKQAAA VITNLRDTAK INQALLATKA VSDVLKEKAE	310 KISAQMGYH ALSRALGYP LISEQFGYII KISQRLGYQ AISRNIGYR LISNGMGYV AIEAKL FISBKLGY	320 ++ DYPF. A. KKESI. DI. X.	330 334
P76268 P77300 Q8EMM2 Q9KAZ9 Q9K6K4 Q8EPA1 Q838M2 Q9WXS0 P37671 P42868	230 	240 ALLPVLDQVR ALMSALAQTRI KLRKEIEKVK FFMKELEKIR AFKAALQKVL QLEEQLQLFQ KLKKELEKIR AMFDELAHIR AMFDELAHIR	250 ++ CGYGEDNECQI CGYALDSEENI LGYALDNUENI LGYGYEMEENI TOGYAEDNEENI KRGYAVDNEENI ESGAAMDREENI SGAAMDREENI	260 EGIRCIAVP COGVRCVAVP EGIRCIASP EGIRCIASP EGIRCIGAP EGIYCIAAP EGIYCIAAP EGIYCIAP EGIYCIAP ELGVSCIAVP ELGVSCIAVP	270 /FDRFGVVI WNHESRVI YDYKGNII IFDYQGAIT IFNHLGQVP FDSQRQLL MKD.GEIA IFDHNGYPV FDIHGRVP	280 ++ A.GLSISFPT A.ALSLSTLT A.SFSISGPT A.AVSISGPS Y.AISITGPL A.AVSVAGPK G.AFSVSMPK A.GVSISGVA Y.AVSISLST C.TSTACE	290 ++LRFSEERL,QE SRVDDAEL,AN NRVSIDEVNLE IRLSKERL.HE ARMTDARL,QT DRMLKQKE.T RKFTEEKI.EE SRLKQVGE.KN	300 ++ FREQLQQAGL LVDKVMMTSK LRPLIIAIGK LIPRLKQAAA VITNLRDTAK INQALLATKA YSDVLKEKAE LLKPLRETAQ LTEKVKDAAI	310 KISAQMGYH ALSRALGYP LISEQFGYI KISQRLGYQ AISRNIGYR LISNGMGYV AIEAKL EISRKLGY. AISNELGFT	320 DYPF. A. DI. DI. X. VRDDLGAIT.	330 334
P76268 P77300 Q8EMM2 Q9KAZ9 Q9K6K4 Q82PA1 Q838M2 Q9WXS0 P37671 P42968 Q9KBV5 O91N67	230 RSTERTITSTE/ TTTPATLASRE/ LRTALTITSO// PHTERTITDKE/ PLTRHTISEPE/ KITANTITDRS/ PYTENTITNPLI PKTPNTITNPLI PLTRNTITELP/ SIGSGTITDPEI PLADGTITDFU	240 ALLPVLOVER ALMSALAQTRI ALMSALAQTRI ALMSALAQTRI ALMSALAQTRI TEKKELEKIR AFKALQVQU ULKELEKIR AMFDELAHIR AMFDELAHIR AMFDELAHIR ULQEIDASVU	250 t	260 DEGLECIAVP DEGIECIASPI DEGIECIASPI DEGIECIASPI DEGIECICAPI DEGLECIAPPI DEGLECIASPI DEGLECSCIAVPI DINTATIASPI DINTATIASPI	270 FFDRFGVVI WWHESRVI YDYKGNII FPDYQGAIT IFNHGQVP FFDSQRQLL MKD.GEIA FFDHNGYPV FFDIHGRVP FFNHERQVA YDDHSGTIV	280 ++ A.GLSISFDT A.ALSLSTLT A.SFSISGPT A.AVSISGPT A.AVSVAGPK G.AFSVSMPK A.GVSISGVA A.GVSISGVA A.GISIAGFE A.GISIAGFE A.GISIAGFE	290 +++ LRFSEERL.QE SRVDDAEL.AN NRVSIDRVNLE IRLSKERL.HE ARMTDARL.QT DRMLKQKE.T RKFTEEKI.EE SRLKQVCE.KN ARFTEDRL.PY IRVTEERL.ST	300 YVAMLHTAR FREQLQQAGL LIDKVNMTSK LIPLIAISK LIPLKQAAA VINLRDTAK INQALLATKA YSDVLKEKAE LLKPLRETAQ LLKPLRETAQ LEKVKDAAL LKVTLKNAAK	310 KISAQMGYH ALSRALGYP KISQRLGYQ AISRNIGYR LISNCMGYV AIEAKL EISRKLGY. AISNELGFT QISRK SISKQLGWL	320 	330 334
P76268 P77300 Q8EMM2 Q9KAZ9 Q9K6K4 Q8EPA1 Q838M2 Q9WXS0 P37671 P42968 Q9KBV5 Q81NS7 Q81NS7 Q81NS7	230 RSTERTITSTE TETPATLASRE URTALTITSOU PHERTITORS PLTRHTISEPEJ KITANTITDRS PTRNTITELP SIGSGTITOPEI PLADGTITORS QFADGTIVDKK KSKVI	240 ALLPVLDQVR ALMSALAQTR KURKEIEKVK KURKEIEKVK AFKAALOKVL JLEEUGLFO ULKKELKQVQ VLKRELEKIR AMFDELAHIR AMFDELAHIR SLLQEIDASV LLKEISETQ DLLEVLQMAK	250 CGYGEDNEEQI CGYALDSENI LGYALDRVENI LGYALDRVENI TDGYAEDNEEHI KGYASIDDEENI SGAAMDREENI DNGYTVSYSELI DTGVAISSELI DTGVAISSELI DTGVAISSELI	260 EGELRCIAVP DGVRCVAVP EGERCIAPP EGITCIAPP EGITCIAPP EGITCIAPP EGITCIAPP EGITCIAPP EGVSCIAVP ELGVSCIAV	270 FFDFGVVI WNHESKVI FDYQGAIT FFNGQVF FFNLGQVP FFNLGQVP FFNLGVP FFNLGVP FFNLGVP FFNLGVV YDHSGTIV FANDGTVV FFNDGVV	280 ++ A.GLSISFPT A.ALSLSTLT A.AVSISGPS Y.AISITGPL A.AVSISGPS Y.AISITGPL A.AVSISGPS Y.AVSISGLA A.GISIAGPE A.GISIAGPE A.GISISGLA A.GISISGLA	290 +	300 ++ FREQLQQAGL LVDKVNMTSK LIPLIIAICK LIPLIAICK LIPLLKQAAA VIINLRDTAK LLKPLRETAQ LLKPLRETAQ LKVLKCAAL LKVLKNAAK LKVKCAAL LKVKCAAL	310 KISAQMGYH ALSRALGYP, LISEQFGYI KISQRLGYQ AISRNIGYR LISNCMGYV AISALGFT QISRK SISKQLGWL RISKELGFL EISKNLGYI	320 DYPF. A. KKESI. DI. X. VRDDLGAIT. A. SNRAHEN.	330 334
P76268 P77300 Q8EMM2 Q9KAZ9 Q9K6K4 Q8EPA1 Q838M2 Q9WXS0 P37671 P42968 Q9KEV5 Q81NS7 Q8ELJ4 P16528 P77734	230 RSTERITISTE TFTPATLASRE LRTALITSOVI PUTRTITSOVI PUTRTITSEPE KITANITISPE FKTPNITTELP PLAPGITTOPEI PLAP	240 ALLPVLDOVR ALMSALAQTR KLRKEIEKVK KLRKEIEKIR AFKAALOKUL DLEEQLOLFO LLKKELKOVO VLKRELEKIR KLLQEIDASV LLKELGATR TLLKDLQATR TLLKDLQATR	250 CGYGEDNEEQ CGYALDSEENI CGYALDSEENI CGYALDSEENI CGYAEDNEENI CGYADDEENI CGYADDEENI SGAAMDREENI SGAAMDREENI SGAAMDREENI SGAAMDREENI KGYAVSIEL CGYADASSEEL CGYADAS	260 Control Control C	270 FDRFGVVI WNHESRVI IYDYKGNII FDYQGAIT FDHLGQVP IFDSQRQLL MKD.GEIA MKD.GEIA MKD.GEIA YFDIHGRVP FNHERQVA FDHRGYP IYDHSGTIV FANDGTVV FNFQNEVV FDEHREPF IYDDVGSVV	280 ++ A.GLSISFFT A.ALSLSTLT A.AVSISCPS Y.AISITGPL G.AFSVSMEK A.GVSISCVA G.AFSVSMEK A.GUSISCT A.GISIAGFE A.GISIAGFE A.GISISGLA A.USVSCPA A.AISISCPI A.AISISCPS	290 +LFSEERL.QE SRVDDAEL.AN NRVSIDRVNLE IRLSKERL.HE ARMTDARL.QT JRLTEENK.TT VRLTEENK.TT KFTEEKLSE SRLKQVGE.KN IRVTEERL.ST IRVTEERL.ST IRVTEERLST SRLKPREW.PE SRLTEDRV.TE	300 ++ FREQLQQAGL LVDKVNMTSK LIPLIIAIGK LIPLIKQAAA VITNLRDTAK UINDLLATKA YSDVLKEKAE LLKPLRETAQ LIKPLRETAQ LTEKVKDAAL LKVTLKNAAK QGELVRDTAR QGELVRDTAR	310 KISAQMGYH ALSRALGYP, LISEQFGYI KISORLGYQ AISRNIGYR AISRELGFT QISRK SISKQLGWL KISKELGFL EISKNLGYI EISKNLGYL EISKNLGYL EISKNLGYL DISTALGLK	320 DYPF. A. XKESI. DI. X. VRDDLGAIT. AT. A. ARAHEN. R.	330 334
P76268 P77300 Q8EMM2 Q9KAZ9 Q9K6K4 Q838M2 Q9WXS0 P37671 P42968 Q9KBV5 Q81NS7 Q8ELJ4 P16528 P77734 Q8JP8 Q8U755	230 	240 ALLPVLDOVRI ALLSVLDOVRI ALLSVLDOVRI ALLSVLDOVRI ALLSVLDEVK IFMKELEKIR AFKAALOKVL ULEEOLOLFO ULEVLOLFKIR MFDELAHIR MFDELAHIR MFDELAHIR MFDELAHIR ULKRELDAT ILLKDLEOAR ILLKDLEOAR ILLSVLGAR	250 CGYGEDREEQ CGYALDSEENI CGYALDSEENI CGYALDSEENI CGYAEDNUENI CGYAEDNEENI CGYAEDNEENI SGAAMDREENI SGAAMDREENI SGAAMDREENI SGAAMDREENI SGAAMDREENI CGYAAVSIESI EGYAAVSEENI EGYAAVEENI CGYAA	260 CEGLRCIAVP COCVACVAVP CEGLRCIASP CEGLRCIASP CEGLRCIASP CEGLYCIAAP C	270 FPDRFGVVI WNHESRVI YDYKGNII FPDYGGAIT FPHLGQVP IFDSQRQLL MKD.GEIA FFDHNGYPV FPDHGRVP FNHERQVA YDHSGTIV FNFQNEVV FNFQNEVV YRPQSDEV VRPQSDEV R.QGKDV	280 A.GLSISFFT A.ALSLSTLT A.SFSISGFT A.VSISGFT A.VVSISGFT A.VVSISGLT A.GVSISGVA A.GVSISGVA A.GVSISGVA A.GISIAGFE A.GISIAGFE A.GISISGLA A.JLSISGFI A.ALSISGFI A.ALSISGPS IGVLSIAGFS VGTMSVAAPL	290 +LFSEERL.QE SRVDDAEL.AN NRVSIDRVNLE IRLSKERL.HE ARMTDARL.QT JRLTEENK.TT VRLTEENK.TT KFTEEKLI.EE SRLKQVGE.KN ARFTEDRL.PY IRVTEERL.ST IEVSESNI.PY SRLKPREW.PE SRITDDRV.TE SRLTEDRF.VS ARLAQSRLAAL	300 	310 KISAQMGYH ALSRALGYP LISEQFGYI KISQRLGYQ AISRNIGYR LISNGMGYV AISNELGFT QISRK SISKQLGWL RISKELGFL EISKNLGYI EISKNLGYL EVTLAYGGM DISTALGLK LSAASPASE: MEIAW	320 YPFF A. KKESI DI VRDDLGAIT. A. SNRAHEN. R. HP FA.	330 334
P76268 P77300 Q8EMM2 Q9KA29 Q9KA24 Q8EPA1 P42968 Q9KBV5 Q9KBV5 Q9KBV5 Q8LJ4 P16528 P7774 Q8ELJ4 P16528 P7774 Q8U755 Q9R201 Q9RV15	230 	240 ALLPVLDOVR ALLPVLDOVR ALLPVLDOVR ALLPVLDOVR ALLPVLDOVR FWKELEKIK FWKELEKIK AFKALOKVL DLEVLOKL DLEVLOK DLEVLOK DLEVLOMAK DLEVLO	250 CGYGEDNEEQ CGYALDSEEN CGYALDSEEN DGYAEDNEEH KQGYSIDDEEN TGYAEDNEEH SGAAMDREEN SGAAMDREEN SGAAMDREEN SGAAMDREEN EGGYTSYSEL CGYCAVSIEH ERGYADDEEH ERGYAVSEAN ARELAATEDEW LGYCAVSEAN	260 CEGLRCIAVP COVRCVAVP EGCRCIASP EGIRCIASP EGIRCIASP EGIRCIGAP EGIYCIAAP UCLICSOV3 UCLIC	270 VEPREGVU WNHESRVI YDYKGNII FPDYGAIT FPNLGQVP FPDSCROLL MKD.GEIA MKD.GEIA MKD.GEIA FDDINGVV FPNDGVV FPNDGVV FNPCOSVV RNFQSDEV RQGKDV RAAGGVL MRG.SVV	280 A.GLSISFFT A.ALSLSTLT A.SFSISGFT A.VSVASGPK G.AFSVSMPK G.AVSVAGPK G.ASVSVAGPK G.ASVSNPK A.GVSISGVA A.GVSISGVA A.GISIAGFE A.GISIGLA A.ISISGFI A.AISISGFI GVLSIAGPS VGTMSVAAPL AAGVSLETP A.GISVQMSAG	290 +	300 	310 KISAQMGYH ALSRALGYP LISEOFGYI KISORLGYQ AISRNIGYR LISNGMGYV AISRNIGYR LISNKGGYU AISRNIGYR KISKELGFL EISRKLGYL BISKELGFL EVTLAYGGM DISTALGLK LSAASPASE EVTLAYGGM DISTALGLK LSAASPASE ELG	320 YPFF A. KKESI DI. VRDDLGAIT. A. SNRAHEN. R. AHP EGLVSPMH.	330 334
P76268 P77300 Q8EMM2 Q9KA29 Q9KA29 Q8EPA1 Q838M2 Q9WXS0 P37671 P42968 Q9KBV5 Q8ELJ4 Q8ELJ4 Q88DF8 Q8EL528 P77734 Q88L9 Q802755 Q9R2013 Q88Z758	230 RSTERITISTE TFTPATLASRE URTALITISOV PHTERITIONE KITANITISPE KITANITISPE KITANITISPE KITANITINE PLANTITOPE PLANGTICKL OFADGTICVKK KSKVU OFTPTIVUKW OFGPNAPRSTDI QQAKAINTIEL EPLITRTCSSI CQAKAINTIEL ETPRLSADQLH	240 ALLPVLDOVR ALLPVLDOVR ALST ALLPVLDOVR ALST ALLPVLDOVR AFKAALOKVL DIEEOLOIGO UKRELEKIR MFDELAHIR MFDELAHIR ALLVLDAVR DLEVLQMAK	250 CGYGEDNEEQ CGYALDSEEN LGGYALDSEEN DGYAEDNEEH KGGYSIDDEEN TGVAYDDEEI TGVAYDDEEN TGVAYDDEEN TGVAYDDEEN TGVAYDDEEN LGYAYSIEL KGYANVSIEEN KGYSFDDEEH ELGYTVDEEH KGYSFDDEEH LGYAYSIEW KGLHWRGLI LGYAYSIEW KGLHWRGLI	260 CEGLRCIAVP COVRCVAVP EGGLRCIASP DEGIRCIASP DEGIRCIASP DEGIRCIGAP DEGIVCIASP DEGIVCIASP DEGIVCIASP DEGIVCIASP DEGIVCIASP DEGINCISSA NHTAALASP DEGVGAIAVA DEGVGAIAVA DEGVGAIAVA DEGVGAIAVA DEGVGAIAVA	270 FDRFGVVI WINESRVI TDYKGNI FDYQGAT FDNLGOV FDNLGOV FDLNGFV FDLNGFV FDLNGFV FDLNGFV FDLNGFV FNERQVA FNERQVA RNFQSDV R. QGKDV RAAGGVL RAAGGVL FNAGATV FNAG	280 A.GLSISFFT A.ALSLSTLT A.SFSISGFT A.SFSISGFT A.VSVAGK G.AFSVSMPK A.GVSISGVA Y.AVSISLS A.GVSISGVA Y.AVSISLS A.GISIAGFE A.GISIAGFE A.GISIAGFE A.GISIAGFE M.ALSISGFI A.ALS	290 +LFSEERL.QE SRVDDAEL.AN NRVSIDRVNLE IRLSKERL.HE AMTDARL.OT DRMLKQKE.T JRLTEENK.TT SRLKQVE.KN ARFTEDRL.PY IRVTEERL.ST IRVTEERL.ST SRLTDRV.TE SRLTDRV.TE SRLTDRV.TE SRLTAGRALMAN TRLTDERVAEI TRLTDERVAEI TRLTDERVAEI TRLTDERVAEI TRLAGGG.Q	300 	310 KISAQMGYH ALSRALGYP, LISEQFGYI KISQRLGYO, ALSRNLGYR LISNGMGYV ALEARL ALSRNLGFT QUSRK SISKQLGWL SISKLGYI EISRKLGYI EISRKLGYI LISASPASE MEIAW ALGARGENGG ELG TYSWEMGYA ALSERMGSY	320 YPFF A. KKESI DI. VRDDLGAIT. A. NRAHEN. A. SNRAHEN. B. SOLVSPMH. SNPHCASTSSY JURS.	330 334
P76268 P77300 Q8EMM2 Q8EAM29 Q9K6K4 Q98K6K4 Q98K8V5 Q9WX50 Q9KU5 Q8KBV5 Q8KBV5 Q8KBV5 Q8KBV5 Q8K2V1 Q88JP8 Q80755 Q9RV13 Q88Z154 Q87Z78 Q87Z78	230 -STERITISTE TFTPATLASRE URTALITISOV PHTERITICK ULTALITISOV PLTRHITSEE KITANITICRS PLTRHITSEE KITANITICRS PLTRHITSEE PLADGTITCKL OFADGTICVKK KSKVU OFTPTTLVCMP OGAAKAINTIE EPLITRTCSSI UQAAKAINTIE ELALQOTAS.I ETPRLSADQLH TSRTLHDPE EKITLOMD	240 ALLPVLDOVR ALLSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR AFKAALOKVL DEEGOLOFO RLKKELKOVO VLKRELEKIR MFDELAHIR MFDELAHIR ALLKEDSVS MLEEAAHVR ALLSEIVTG ALLSEIVTG	250 CGYGEDNEEQ CGYALDSEEN LGGYALDSEEN LGGYALDVEN LGYGYENEEN KGGYSIDDEEN TGVAYDDEEI TGVAYDDEEI TGVAYDDEEI TGVAISSEL LGYAYSIEL RGYAVSIER RGYAVGETSI RLGYAYSIEW RGGYAVGETSI RLGYAYSIEW KGCHHVRLLI AGUACYDCEI TGVHQQCMLI AGUACYDQEI RGYAUGANI CGMI CGMI CGMI CGMI CGMI CGMI CGMI CGM	260 CEGLRCIAVP COVRCVAVP EGCRCIASP EGIRCIASP EGIRCIASP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIASP IGIMCIGA EGIRCIASCA EGIRCIASCA EGIRCIASCA EGIRCIASCA EGIRCIASCA EGIRCIASSA EGIRCIASSA EGIRCIASSA EGIRCIASSA EGIRCIASSA EGIRCIASCA EGIRCIASSA EG	270 FDRFGVVI WINHESRVI YDYKGNI FDYQGAT FPNLGQVP FDSQRQLL FDNGFVV FDLGRVP FDLGRVP FDLGRVP FDLGRVP FNLERQVD FNLERQVD RANGGVL RANGGVL RANGGVL RANGGVL RANGGVL FNCAGU RANGGVL FNCAGU RANGGVL FNCAGU FNCAGU FNCAGU RANGGVL KITKGQVL XTVRGQVL XTVRGQVL XTVRQVL	280 A.GLSISFFT A.ALSLSTIT A.SFSISGFT A.SFSISGFT A.VSISGFS A.VSISGFS A.GVSISGVA A.GVSISGVA A.GVSISGVA A.GVSISGVA A.GISIAGFE A.GISIAGFE A.GISIAGFE A.GISIAGFE GUSLSIGAGES VGTMSVAAFL A.GISVAGES GUSVAAFL A.GISVAGES A.GI	290 +	300 	310 KISAQMGYH ALSRALGYP, LISEQFGYI KISQRLGYQ ALSRNLGYR LISNCMGYV ALSRNLGYT GUSRK SISKQLGWL SISKQLGWL SISKQLGWL DISTALGLK DISTALGLK DISTALGLK DISTALGLK MEIAW ALGARGENG BLG TVSWRMGYA ALSERMGSS LCHQLFG. LARMLV.	320 YPFF A. KKESI DI. VRDDLGAIT. A. NRAHEN. A. NRAHEN. SOLVSPMH. SNPHCASTSSY LRS.	330 334
P76268 P77300 Q8EMM2 Q8EMM2 Q9K6K4 Q9K8K4 Q9K8K4 Q9K8K5 Q9K8V5 Q9KBV5 Q9KBV5 Q8KBV5 Q8KBV5 Q8KBV5 Q8K2V1 Q9KV13 Q88JF8 Q872778 Q8727778 Q8727778 Q8727778 Q87777778 Q877777778 Q877777778 Q877777777	230 -STERTITSTE TFTPATLASRE URTALITSOV PHTERTITORE PLTRHTSEPE KITANTITORS PTENTITNEN PLTRHTSEPE KITANTITER PLTRHTSEPE PLATENTONE OFADGTITVKK KSKVU OFTETTLVDRY OFGENARASTITE EPLYTERTCSSI ETPRLSADQLH TSRTHOPE: EKITDMDI TPYITDRAL KYITITNAU	240 ALLPVLDOVR ALLSVLDOVR ALLSVLDOVR ALSVLDOVR PRKALOKVL DEEGUOLFOU RLKKELKOVO VLKRELEKIR MFDELAHIR MFDELAHIR MFDELAHIR LLKELLAQTR TLKKDLQATR TLKDLQATR TLKDLQATR TLKELQATR SUKTEVARVR WLEEAAHVR WKTEVARVR VHTLLAGIR SUFALIQIR SUFALIQIE SUFALIQIE SUFALIQIE SUFALIQIE SUFALIQIE SUFALIQIE SUFALIQIE SUFALIQIE SUFALIQIE SUFALIQIE SUFALIQIE SUFALIQIE SUFALIQIE SUFALIQIE SUFALIQIE SUFALIQIE SUFALIQIE SUFALIQU	250 CGYGEDNEEQ CGYALDSEEN LGGYALDSEEN LGGYALDVEN LGYGYENEEN KOGYSIDDEEN TGYAPDEEII KGYAYDDEEII TGVAIDEEN LGYAYSELI TGVAISSEL LGYAYSEEN RGYAPDEEH LGYAYSTEEN LGYAYSTEEN LGYAYSTEEN CKGLHVHGLLI AGUCVVDEEI RGYAQUETS	260 DEGLECIAVP COVRCVAVP EGCRCIASP DEGITCIASP DEGITCIASP DEGITCIASP DEGITCIASP DEGITCIASP DEGITCIASP DEGITCIASP DIGITCIAST DIGITCIAST DIGITCIAST DIGITCIAST DIGITCIAST DEGUTASIAVP DEGUTASIAVP DEGUTASIAVP DEGUTASIAVP DEGUTASIAVP DIGITSIAVP	270 FDRFGVVI WINHESRVI TDYKGNI FDYQGAT FPNLGOPY FDSOROLL MKD.GEIA IFDINGFVY FDINGFVY FDINGFVY FDINGFVY FNHERQVA FNHERQVA FNHERQVA FNHERQVA KRPOSDEV RAGGEV RAGGES VI PAMORTVY LGRAGGVL LGRAGGLK VI JOSOSV VI JO	280 A.GLSISFFT A.ALSLSTLT A.SFSISGFT A.SFSISGFT A.VSISGFS A.SVSISGF A.GVSISGVA A.GVSISGVA Y.AVSISLST A.GVSISGVA A.GVSISGVA A.GISISGPA A.G	290 +LFSEERL.QE SRVDDAEL.AN NRVSIDRVNLE IRLSKEL.HE AMTDARL.OT DRMLKOKE.T YRLTEENK.TT YRLTEENK.TT SRLKQVGE.KN AFTEDOR.PY IRVTEERL.ST SRLTDRV.TE SRLTDRV.TE SRLTDRV.TE SRLTDREV.S RLAQSRLAAL TRLTDERVASU TRLTDERVASU RLAQSG.Q WTRSELQRF VASMEELVERY RLSIDEIRLKY RLSIDEIRLKY RLSIDEIRLKY ITYKEVLIOHI	300 	310 KISAQMGYH ALSRALGYP, LISSCPGYU KISQRLGYO, AISRNLGYR LISNCGMGYU AISRNLGYT GUSRK SISKCLGML RISKELGFL GUSRK SISKCLGML RISKELGFL MEIAM ALSARPAGEN MEIAM ALSARPAGEN MEIAM ALSARMGYA ALSENMGSYA ALSENMGSYA ALSENMGSYA LCHQLFG LRPULD LRPULD LRNIL.	320 YPFF A. XKESI DI. XRDDLGAIT. X. YRDDLGAIT. AT. A. NRAHEN. R. AHP. JFA. SGLVSPMH. SNPHCASTSSY JLRS.	330 334
P76268 P77300 Q&EMM2 Q&EMM2 QBYACA QB	230 	240 ALLPVLDOVR ALLSVLDOVR ALLSVLDOVR ALLSVLDOVR FMKALOKUL DEEGUOLFOU RLKKELKOVO VLKRELEKIR MYDELAHIR MYDELAHIR ALKELLAQTR TLKDLQATR TLKDLQATR TLKDLQATR TLKDLQATR MKTEVARVR WKTEVARVR WKTEVARVR WKTEVARVR WKTEVARVR DYHLLGIR SLFACIOVT ALLAEIDVTG DLAELOVTG	250 CGY4EDPEEQ CGYALDSEEN LGGYALDVEN LGYGYEMEEN TGYAEDNEEH KOGY5 DDEEN TGYAEDNEEH KOGY5 DDEEN TGYAEDDEE LI KGYAUDEEN LGYTVDKEEH KGY5FDDEEH LGYTVDKEEK RGY6FDDEEH KGYAFDEEN KRGYFDDEEH KGYAFDEEN KLGYAYSIEEN KGYGENCEEN KLGYAYSIEEN KGYGENCEEN KLGYAYSIEEN KGYAUCHTUDEY KGULTUDEY KGULTUDEY CGCCVCVEEH	260 DEGLECIANP DEGLECIANP DEGLECIANP DEGLECIANP DEGLECIANP DEGLECIGAP DEGLECIGAP DEGLECIGAP DEGLECIANP DIFEGSUS DIFEGSUS DIFEGSUS DIFEGSUS DIFEGSUS DIFEGSUS DIFEGSUS DIFESSION DIFE	270 FDRFGVVI WINHESRVI TDVKGNI FDVGGAT FPNLGOVP FDSOROLL MKD.GEIA FDNLGVV FDLGRVP FDLGRVP FDLGRVP FNLERQVA FNLERQVA FNLERQVA KINGGSVV KRPOSDEV FNLGRVP FNLGR	280 A.GLSISFPT A.ALSLSTLT A.SFSISGPT A.SFSISGPT A.VSISGPS A.SVSISGPT A.AVSISGPS A.GVSISGVA A.GVSISGVA A.GVSISGVA A.GVSISGVA A.GISIAGPE A.	290 +	300 	310 KISAQMGYH ALSRALGYP, LISSOFGYI KISQRLGYO, AISRNLGYR LISNCGMGYU AIEAKL EISRKLGYI EISRKLGYI EISRKLGYI EISRKLGYI EISRKLGYI EISRKLGYI EISRKLGYI EISRKLGYI EISRKLGYI EISRKLGYI AISASPAGES MEIAM AISASPAGES LCHQLFG LRYLLD LRNLL LRNLL	320 YPF A. KKESI DI. KKESI DI. KKESI DI. SI VRDDLGAIT. AT. A. SNRAHEN. R. AHP. LFA. SNPHCASTSSY LRS. SNPHCAS	330 334
P76268 P77300 QEEMM2 QEEMM2 Q9K6K4 Q9K6K4 Q3WX50 Q9K2 Q9K2 Q9K2 Q9K2 Q42968 Q42968 Q42968 Q42968 Q41755 Q9R201 Q88JP8 Q80755 Q9R201 Q88246 Q41755 Q9R201 Q82154 Q82154 Q82154 Q83046 Q83046 Q43992 Q8FUX8 Q8F	230 	240 ALLPVLDOVR ALLPVLDOVR ALLPVLDOVR ALLPVLDOVR ALLPVLDOVR PRKALOKUL DEEDLOLFOL DEEDLOLFOL DEEDLOLFOL DEELCEKIR MMFDELAHIR MMFDELAHIR MMFDELAHIR DLKVLDAVK DLKVLDAVK DLKVLDAVK DLKVLDAVK DLKVLDAVK DLKVLDAVK DLKVLDAVK DLKVLDAVK DLKVLDAVK DVHTLLAGIR DVHTLLAGIR DVHTLLAGIR DVHTLLAGIR DVHTLLAGIR DVHTLLAGIR DVHTLLAGIR DVHTLLAGIR DVHTLLAGIR DVHTLLAGIR DVHTLLAGIR DVHTLLAGIR DVHTLLAGIR DVHTLLAGIR DVHTLLAGIR DVHTLLAGIR DVHTLLAGIR DVHTLAGIR	250 CGYGEDBEEQ CGYALDSEEM LGYALDSEEM KOGYSIDDEN EGYADDEEN SGAADDEEN CGYAVDNEEM CGYAVDNEEM CGYAVDDEEN CGYAVDEEN CGYAVDEEN CGYAVDEEN CGYAVCE	260 DEGLECIAVP DEGLECIASP DEGLECIASP DEGLECIASP DEGLECIASP DEGLECIGAP DEGLECIGAP DEGLECIASP DEGLECIASP NYTAAIGAP NYTAAIGAP NYTAAIGAP NHTAAIAAP DEGLECIASC NHTAAIAAP DEGLECIASC NHTAAIAAP DEGLECIASC DEGLECI	270 FEDREGVUI WINHESRVI IVDYKGNI FDYQAT FPNHLGOPP FDSOROLL MKD.GELA FDINGYPV FDINGYPV FDINGYVV FNERQVA FNHERQVA F	280 A.GLSISFFT A.ALSLSTLT A.SFSISGFT A.AVSISGFS A.AVSISGFS A.AVSISGFS A.AVSISGFA A.GVSISGVA Y.AVSISLST A.GLSISGFA A.GLSISGFA A.GLSISGFA A.GLSISGFA A.AISISGFS A.AISISGFS A.AISISGFS A.ALSUSGFA A.ALSUSGFA A.ALSUSGFA A.ALVVGFS AALWVSTMG AALWVSTMG AALWVSTM AALMUSTM	290 +	300 	310 	320 YPF. A. XKESI DI. XRDDLGAIT. XRDDLGAIT. AA NRAHEN. R. AHP. JCR. SNPHCASTSSU JCRS A	330 334
P76268 P77300 QEEMM2 QSEAM29 QSK6K4 QSWX50 P37671 P42968 QSWX50 QELJ4 P16528 P77734 QEELJ4 P16528 P77734 Q88JF8 Q80755 QSPU13 QSPU13 QS7278 QS2154 QS7278 QS2154 QS92154 QS920 QS304	230 RSTERTITSTE TFTPATLASRE LTRALITSOV PHTERTITORE PLTRHITSPE KITANTITRSY PLTRHITSEPE KITANTITRE PLTRHITSEPE LADGTITKL QFADGTITVKK KSKU QFTPTLVDKY QFTPTTVVKK QFGPNAFRSTDI QQAKAINTIE ELALQCTAS.I ETPRLSADQLHI TSRILHDPE E.KAITITDY FYITDEH EHAERLIR EHAERLIR EHAERLIR MC OULAPEN	240 ALLPVLDOVR ALLPVLDOVR ALLPVLDOVR ALLPVLDOVR ALLPVLDOVR PRKALOKUL DEFEUCIFO D	250 CGYGEDBEEQ CGYALDSEEN CGYALDSEEN CGYALDSEEN CGYALDSEEN CGYADDEEN SGAADDEEN SGAADDEEN CGYAVDEEN CGYAVDEEN CGYAVDEEN CGYAVDEEN CGYAVDEEN CGYAVDEEN CGYAVCEN CCAN CCAN CON CON CON CON CON CON CON CON CON CO	260 DEGLECIAVP 20GVRCVAVP 20GVRCVAVP EGLICIASP DEGITCIASP DEGITCIASP E	270 FEDREGVUI WINESRUI IVDXEGNI FDYGGAT FDHLGOVP FDSOROLL FDHLGOVP FDSOROLL FDHLGVP FDHERQVA FANDGTVU FNFCROSVU FNFCROSVU FNFCROSVU FNFCROSVU FNROSSVU FNAGGLASU FANGGTVI JGRAGGLASU JEGOSSVU JLGGOSSU JLGGOSGU JLLGSSFP RTPCGLAPL FRNDGESVE VLDVGGULAPL FRNDGESVE VLDVGGULAPL FRNDGESVE VLDVGGULAPL FRNDGESVE VLDVGGULAPL FRNDGESVE FRNDF FRNDF FRNDF FRNDF FRNDF FRNDF FRNDF FRNDF FRNDF FRNDF FRNDF FRNDF FRNDF FRNDF FRNDF FRNDF FRNDF FRNF FR	280 +A-GLSISFPT A.ALSLSTLT A.SFSISGPT A.SFSISGPT A.AVSISGPS A.SVSISGVA Y.AVSISGVA Y.AVSISLST A.GVSISGVA A.GVSISGVA A.GISIAGFE A.GISIAGFE A.GISIAGFE A.GISIAGFE A.AISISGPI A.AISISGPI A.AISISGPI A.ALSUSVGPA SQUTVVGPGS AALWVSIPVG AALWVSIPVG AALWVGTVS AALWVSIPVG AALWVGTVS AALWVSIPVG AALWVGTVS AALWVSIPVG AALWVGTVS AALWVSIPVG AALWVGTVS AALWVSIPVG AALWVGTVS AALWVSIPVG AALWVGTVS AALWVSIPVG AALWVGTVS AALWVSIPVG AALWVGTVS AALWVSIPVG AALWVGTVS AALWVSIPVG AALWVGTVS AALWVSIPVG AALWVGTVS AALWVSIPVG AALWCGGID VFN.CGGSFD VFN.CGGSFD AVS.CMGPAT	290 +	300 	310 	320 YPF. A. XKESI DI. X. VRDDLGAIT. X. XRAHEN RAHEN SNRAHEN SNRAHEN SNRAHEN SNRAHEN SNRAHEN SNRAHEN SNRAHEN H SNRAHEN S	330 334
P76268 P77300 QEEMM2 QSK6K4 QBY QSK6K4 QSWX50 P37671 P42968 QSWX50 QSUX50 QSUX50 QSUX50 QSUX50 QSV13 QSV13 QSV13 QSV15 QSV13 QSV15 QSV13 QSV15 QSV13 QSV15 Q	230 RSTERITISTE TFTPATLASRE LTALITISOV PHTERITICK KITANITISPE KITANITISPE KITANITISPE PLTRHISPE PLTRHISPE PLTRHISPE PLTRHISPE PLTRHISPE PLTRHISPE PLTRHISPE PLTRHISPE CONSTITUTE DALOGTAS.I ETPRLSADQLHI TSRILHDPE E.KITIMDU TP.YITDEH E.KATIMDU TP.YITDEH E.KATIMDU TP.YITDEH E.KATIMDU TP.YITDEH E.KATIMDU TP.YITDEH C.MWPSIKKR C.DWSIKKR C.DWSIKKR C.DWSIKKR C.DWSIKKR C.DWSIKKR	240 ALPVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR FFKALOKUL DEFEDIOLFOI DEFEDIOLFOI DEFEDIOLFOI DEFELIOLFOI DEFELIOLFOI DEFELIOLFOI DEFELIOLFOI DEFELIOLFOI DEFELIOLFOI DEFELIOLFOI DEFELION EVERHIKOIR EVERHIKOIR EVERHIKOIR EVERHIKOIR DIALEEFVTGI DIALECTOVR DIFILLSEIK FFLETLDAVR	250 CGYGEDBEEQ CGYALDSEEN CGYALDSEEN CGYALDSEEN CGYALDSEEN CGYADDEEN CGYADDEEN CGYAUDNEEN CGYAUDEEN CGYAUDEEN CGYAUDEEN CGYAUDEEN CGYAUDEEN CGYAUDEEN CGYAUDEEN CGYAUDEEN CGYAUDEEN CGYAUCEN CGY	260 DEGLECIAVP 20GVRCVAVP 20GVRCVAVP 20GVRCVAVP 20GVRCVAVP 20GVRCVAVP 20GVRCVAVP 20GVRCVAVP 20D1FCSGVS 2	270 FDRFGVVI WINESRVI IVDXFGNI FDVGGAT FDVGGAT FDHLGOVP FDSOROLL MED FDHLGOVP FDSOROLL FDHLGOVP FDHERQVA FANDGTVV FNFONEVV FNFONEVV FNFONEVV FNFONEVV FNFONEVV FNFONEVV FNANGGTV IVDASGVI LGRAGGLKI VIDASGUL VIDASGU	280 +A.GLSISFPT A.ALSLSTLT A.SFSISGFT A.SFSISGFT A.AVSISGFS A.AVSISGFS A.GVSISGVA Y.AVSISLST A.GVSISGVA A.GVSISGVA A.GISIAGFE A.GISIAGFE A.GISIAGFE A.GISIAGFS GVITUVGPGS GVITUVGPGS GVITUVGPGS AALMVSIPT AALGVSLPPP ASIGVOMSAG AUMTVVGPTS GVITUVGPGS AALMVSLPP AALMVSLPP ASIGVOMSAG AUMTVVGPS GVITUVGPGS AALMVSLPP AALMVSLPP AALMVSLPP AALMVSLPP AALMVSLPP AALMVSLPP AALMVSLPP AALMVSLPP AALMVSLPP AALMVSLPP AALMVSLPP AALMVSLAG AALMVSLA	290 +	300 	310 	320 DYPF. A. WRDDLGAIT. WRDDLGAIT. A. WRDDLGAIT. A. WRDDLGAIT. A. B. WRDDLGAIT. A. B. WRDDLGAIT. A. B. B. B. B. B. B. B. B. B. B	330 334
P76268 P77300 Q8EMM2 Q8EMM2 Q9K6X4 Q8EPA1 Q9WX50 Q9K6V5 Q9K8V5 Q8ELJ4 P16528 P77734 Q8ELJ4 P16528 Q87278 Q877732 Q8172732 Q817732 Q8172732 Q817732 Q817732 Q817732 Q817732 Q817732 Q817732 Q817732 Q81777732 Q8177777777732 Q81777777777777777777777777777777777777	230 RSTERITISTE TFTPATLASRE LTRALITSOV PHTERITISEE KITANTITORS PYTENITIELP SIGSGITIPEL PLANGITIPEL PLANGITIPEL PLANGITIPEL PLANGITIPEL PLANGITIPEL PLANGITIPEL PLANGITIPEL PLANGITIPEL SIGSGITIPEL PLANGITPEL SIGSGITIPEL PLANGITPEL SIGSGITPEL PLANGITPEL SIGSGITPEL PLANGITPEL SIGSGITPE	240 ALPVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR AMFDELASIN AMFDELASIN AMFDELASIN AMFDELASIN ALSVLDOVR AL	250 CGYGEDREEQ CGYALDSEEN CGYALDSEEN CGYALDSEEN CGYALDSEEN CGYALDSEEN CGYALDSEEN CGYALDSEEN CGYALDSEEN CGYALDSEEN CGYALDSEEN CGYALDSEEN CGYALDSEEN CGYALDSEEN CGWCYSCCOW CGWCYSSEEN CGWCYSSEEN CGWCYSCCOW CGWCYSSEEN CGWCYSCCOW CGWCYSSEEN CGWCYSCCOW CGWCYSSEEN CGWCYSCCOW CCWCYSCCOW CCWCYSCCOW CCWCYSCCOW CCWCYSCCOW CCWCYSCCOW CCWCYSCCOW CCWCYSCCOW CCWCYSCCOW CCWCYSCCOW CCWCYSCCOW CCWCYSCCOW CCWCYSCCOW CCWCYSCCOW CCWCYSCCOW CCWCYSCCOW CCWCYSC	260 EGGLRCIAVP EGGLRCIASP EGGLRCIASP EGGIRCIASP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIAAC EG	270 FEDREGUVI WINESRVI UVIXENTI FDYQGAT FDNHLGQVP FDSOROLL MKD, GELA FDHNGRVP FDINGRVV FDINGRVV FDINGRVV FNNCONSVV FNNCONSVV FNNCONSVV FNNCONSVV FNNCONSVV TANAGEVI FNAMGCI F	280 +	290 +	300 	310 	320 YPF A XKESI DI XKESI XKESI DI XKESI XKE	330 334
P76268 P77300 Q8EMM2 Q8EMM2 Q9K6X4 Q9WX50 Q9K6V5 Q9K05 Q9K05 Q9K05 Q8524 Q8524 Q8524 Q8524 Q8524 Q87278 Q87777777777	230 RSTERITISTE TFTPATLASRE LTRALITSOVI PHTERITICK KITANITISPE KITANITISPE KITANITISPE KITANITISPE SIGSGIIDEL PLANGIIDEL PLANGIIDEL PLANGIIDEL OFADGIIDEL PLANGIIDEL OFADGIIDEL PLANGIIDEL OFADGIIDEL SIGSGIIDEL PLANGIIDEL OFADGIIDEL OFADGIIDEL SIGSGIIDEL PLANGIIDEL SIGSGIIDEL PLANGIIDEL SIGSGIIDEL SIGSGIIDEL PLANGIIDEL SIGSGIIDEL SIGSGIIDEL PLANGIIDEL SIGSGI	240 ALPVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR AMFDELAKIN AMFDELAHIR AMFDELAHIR AMFDELAHIR ALSVLDOVR AL	250 CGYGEDDEEQ CGYALDSEEM CGYALDSEEM CGYALDSEEM CGYALDSEEM CGYALDSEEM CGYALDSEEM CGYALDSEEM CGYALDSEEM CGYALDSEEM CGYALDSEEM CGYALDSEEM CGYALDSEEM CGYALDSEEM CGYALDSEEM CGYALDSEEM CGWCYSDWCCD CGWCYSSEEM CGWCYSDWCCD CGWCYSDWCCD CGWCYSDWCCD CGWCYSDWCCD CGWCYSDWCCD CGWCYSDWCCD CGWCYSDWCCD CGWCYSDWCCD CGWCYSDWCCD CGWCYSDWCCD CGWCYSDWCCD CGWCYSDWCCD CGWCYSDWCCD CGWCYSDWCCD CGWCYSDWCCD CGWCYSDWCCD CGWCYSDWCCD CGWCYSDWCCD CGWCYSDWCCD CCCCCD CCCCCCC CCCCCCCCCCCCCCCCCC	260 EGGLRCIAVP CGVRCVAVP EGGLRCIASP EGIRCIASP EGIRCIGAP EGIR	270 FEDREGUVI WINESRUI YDYKGNI FDYGQAT FPUHLGQVP FDSOROLL MKD, GELA FPHHGQVP FDSOROLL FDHGQVF FDFRERVA FDHGRVF FDFRERVF FDHGRVF FDFRERVF FDHGRVF FNANGCI FDANGGVV YDASGQUI YDASGQUI YDASGQUI YDASGQUI YDASGQUI YDASGQUI YDASGQUI YDASGQUI LGRAGCH FPNMGQLA FRYDGTGL JLGSSFP CRPCGLAPL CRUGGUI YDASGQUI JLGSSFP CRUGGUI YDASGQUI JLGSSFP CRUGGUI YDASGQUI JLGSSFP CRUGGUI YDASGQUI JLGSSFP CRUGUI FNNMNNEN FNNMNEN FNNMNEN FNNMNEN FNNMNEN FNNMNEN FNNMNEN FNNMNEN FNNMNEN FNNMNEN FNNMNNEN FNNMNNEN FNNMNNEN FNNMNNEN FNNMNNEN FNNMNNEN FNNMNNEN FNNMNNEN FNNMNNEN FNNMNNEN FNNMNNEN FNNMNNEN FNNMNNEN FNNMNNEN FNNMNNEN FNNMNNNN FNNMNNNN FNNMNNNNN FNNMNNNNN FNNMNNNNN FNNMNNNNN FNNMNNNN FNNMNNNN FNNMNNNNN FNNMNNNNN FNNMNNNNN FNNMNNNNN FNNMNNNNN FNNMNNNNN FNNMNNNNN FNNMNNNNN FNNMNNNN FNNMNNNNNNNNNN	280 A.GLSISFPT A.ALSLSTLT A.SFSISGPT A.SFSISGPT A.SFSISGPT A.SFSISGPT A.SFSISGPT A.SFSISGPT A.SISISGPI A.SISISGPI A.GLSISGPI A	290 +	300 	310 	320 DYPF. A. WRDDLGAIT. WRDDLGAIT. A. WRDDLGAIT. A. WRDDLGAIT. A. B. WRDDLGAIT. A. B. WRDDLGAIT. A. B. B. B. B. B. B. B. B. B. B	330 334
P76268 P77300 Q8EMM2 Q8EMM2 Q9K6X4 Q9K6X4 Q9WX50 Q9K6V5 Q9K8V5 Q8ELJ4 P16528 P77734 Q8ELJ4 P16528 Q8V755 Q882F4 Q8	230 RSTERITISTE TFTPATLASRE LRTALITISOV PUTENTITSOV PUTENTITSOV PUTENTITNEL PLANDITUEL SIGSGIITOPLI PLANDITUEL SIGSGIITOPLI PLANDITUEL SIGSGIITOPLI PLANDITUEL SIGSGIITOPLI PLANDITUEL SIGSGIITOPLI PLANDITUEL SIGSGIITOPLI PLANDITUEL SIGSGIITOPLI PLANDITUEL SIGSGIITOPLI PLANDITUEL SIGSGIITOPLI PLANDITUEL SIGSGIITOPLI PLANDITUEL SIGSGIITOPLI PLANDITUEL SIGSGIITOPLI PLANDITUEL SIGSGIITOPLI SIGGII SIGSGIITOPLI SIGI	240 ALPVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR AMFDELASIN AMFDELASIN AMFDELASIN AMFDELASIN AMFDELASIN ALSVLDOVR A	250 CGYGEDDEEQ CGYALDSEEN C	260 EGGLRCIAVP CQGVRCVAVP EGGLRCIASP EGITCIASP EGITCIASP EGITCIASP EGITCIASP EGITCIASP EGITCIASP EGITCIASP EGITCIASP EGITCIASP EGITCIASP EGITSIAS EGITS	270 FEDREGUVI WINESRUI YDYKGNI FDYGQAT FPHLGQVP FDSOROLL MKD, GELA FDHNGRVP FDSOROLL MKD, GELA FDHRGVP FDHRGVP FDHRGVP FDHRGVP FDHRGVP FDHRGVP FNHCQVF CR. QGKDV KRPQSDV LGRAGGLA FAMGGTV LGRAGGLA FAMGGTV LGRAGGLA KTVRGQTV LGRAGGLA KTVRGQTV LGRAGGLA KTVRGQTV LGRAGGLA KTVRGQTV LGRAGGLA KTVRGQTV LGRAGGLA KTVRGQTV LGRAGGLA COLNGCCV FNNNNCH FNNNNCH FNNNNCH FNNNNCH COLNGCV FNNNNCH KNAGQCAL FNNNNCH KNAGQCAL FNNNNCH KNAGQCAL FNNNNCH KNAGQCAL FNNNNCH KNAGQCAL FNNNNCH KNAGQCAL FNNNNCH KNAGQCAL FNNNNCH KNAGQCAL FNNNNCH COLNGCV FNNNNCH COLNGCV FNNNNCH FNNNCH COLNGCV FNNNNCH FNNNNCH FNNNNCH FNNNNCH FNNNCH FNNNCH FNNNCH FNNNCH FNNNCH FNNNCH FNNNCH FNNNCH FNNNCH FNNNCH FNNNCH FNNNCH FNNNCH FNNNCH FNNNCH FNNNCH FNNNCH FNNNCH FNNCH FNNNCH FNNNCH FNNNCH FNNCH FNNCH FNNCH FNNNCH FNNC	280 +	290 +	300 	310 	320 DYPF. A. XKESI. DI. XK	330 334
P76268 P77300 Q8EMM2 Q8EMM2 Q9K6X4 Q9K6X4 Q9K6X4 Q9K6X5 Q9K6V5 Q9K8V5 Q8ELJ4 P16528 P77734 Q8ELJ4 P16528 Q8V755 Q882F4 Q80755 Q882F4 Q897 Q897 Q897 Q897 Q897 Q897 Q877 Q877	230 RSTERTITSTE TFTPATLASRE LTTALTITSOVI PHTRATLASRE KITANTITONO PLTRATISPE KITANTITONO PLTRATISPE KITANTITPL SIGSGTITPEL ILADGTITVKL QFADGTIVKL QFADGTIVKL QFADGTIVKL QFADGTIVKL DGTPTTLVDMP OFGFNARSTOP LLALQTAS. K. YTYTNNI TP. YTTDEH E. HARFLIE E. HARFLIE E. HARFLIE SPIGRAEINE SPI	240 ALPVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR ALSVLDOVR AMFDELAHIR AMFDELAHIR AMFDELAHIR AMFDELAHIR AMFDELAHIR ALSVLDOVR A	250 CGYGEDDEEQ CGYALDSEEN C	260 EGLRCIAVE EGGLRCIASP EGGIRCIASP EGGIRCIASP EGGIRCIGAP EGGIRCIGAP EGGIRCIGAP EGLVCLASP IGLWCVCP IGLWCVCP IGLWCVCP IGLWCVCP EGVGALAAAAAP IGLSSLAVE IGLWCVCASAA IGTSSLAVE IGLWCVCASAA IGTSSLAVE IGLWCVCASAA IGTSSLAVE IGLWCVCASAA IGLSSLAVE IGLWCVCASAA IGLSSLAVE IGLWCVCASAA IGLSSLAVE IGLWCVCASAA IGLSSLAVE IGLWCVCASAA IGLSSLAVE IGLWCVCAAA IGLSAAAAA IGLSAAAAA IGLSAAAAA IGLSAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	270 FFDRFGVVI WINESRVI FDVKGNI FDVKGNI FDVGAT FDNLGQV FDSOROLL MKD, GELA FDHNGRV FDNERQVA FDIGRV FDIGRV FDIGRV FDNGRV FNRCQVA FDNGRV FNRCQVA FNRCQV	280 +	290 +	300 	310 	320 DYPF. A. XKESI DI. X. WRDDLGAIT. A. WRDDLGAIT. A. MR. A. SINFAHEN. R. A. SINFAHEN. SINFACASTSS: DLRS. SPFDLQAS. REAP. A. SPFDLQAS. REAP. A. SPFDLQAS. REAP. A. SPFDLQAS. REAP. A. SPFDLQAS. REAP. A. SPFDLQAS	330 334
P76268 P77300 Q8EMM2 Q8EMM2 Q9K6X4 Q9F6K4 Q9K6V5 Q9K6V5 Q9K8V5 Q8ELJ4 P16528 P77734 Q8ELJ4 P16528 Q8V755 Q882F8 Q80755 Q882F8 Q882F4 Q882F8 Q882F4 Q83046 Q83046 Q43992 Q88046 Q43992 Q8F053 Q88046 Q43992 Q8F053 Q88046 Q43992 Q8F053 Q88046 Q817278 Q88046 Q43992 Q88046 Q43992 Q88046 Q817732 Q88046 Q87732 Q88047 Q88047 Q80732 Q80047 Q80732 Q80047 Q80732 Q80047 Q80732 Q80047 Q80732 Q80047 Q80732 Q807732 Q80732 Q80732 Q80732 Q80732 Q80732 Q80732 Q807732 Q80742 Q8074	230 RSTERTITSTE TFTPATLASRE LTTALTITSOVI PHTRATLASRE KITANTITOSV PLTRATISPE KITANTITOSV PLTRATISPE KITANTITPL PLTRATITUSV IGGORTUVKK QFADGTUVKK QFADGTUVKK QFADGTUVKK COQARATNTIE ELADGTIVKL QFADGTUVKK COQARATNTE ELALQOTAS. COMPAFIC PLANGTIVK COMPAFIC PLANGTUVK COMPAFIC PLANGTUVK RSV QFTPTIVOMP OFGPTALANDA COMPAFIC PLANGTUKK RSV MUSIKK COMPAFIC PLANGTUKK RSV MUSIKK SPIGRAEINFE SPIGR	240 ALLPVLOOVR ALMSALQTR PKRAALQKU' JEECULOFOU JE	250 CGYCEDUECQ CGYALDEED CGYALDEED CGYALDEED CGYALDEEN CGYSIDDEEN CGYSIDDEEN CGYSIDDEEN CGYSIDDEEN CGYSIDDEEN CGYSIDDEEN CGYSIDS CGYSIDS CGYSISSEL CGYSISSEL CGYSISSEL CGYSISSE CGYSISSE CGS CGS CGS CGS CGS CGS CGS CGS	260 EGLRCIAVE EGGLRCIASP EGGIRCIASP EGGIRCIASP EGGIRCIASP EGGIRCIGAP EGGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP IGLRSIAVE	270 FDRFGVVI WNHESKVI FDRFGVVI FDRJGQAIT FDNLGQVP FDSQRQLL MKD, GELA FDINLGVV FDDLGKVV FDDLGKVV FDDLGKVV FDL	280 A.GLSISFFT A.ALSLSTIT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSISGFT A.JSISGFT A.GSISISGFA A.GSISISGFA A.GSISISGFA A.GSISISGFA A.JSISGFS IGVISVGFA A.JSISGFS GVITVVGFCS ALMVSTHVG ALMVSTHVG ALMVSTHVG ALMVSTHVG ALMVSTHVG ALMVSTHVG ALMVSTHVG ALMVSTHVG ALMVSTHVG ALMVSTHVG ALMVSTHVG ALMVSTHVG ALMVSTHVG ALMVSTHVG ALMVSTHVG ALMVSTHVG ALMVSTHVG ALMVSTHVG ALMVSTHVG ALSVGFAL ALSVGFA	290 +LRFSEERL.QE SRVDDAEL.AN NRVSIDRVNLE IRLSKEL.HE ARMTDARL.QT DRMLKQKE.T RKFTERKI.EX SRLSKUCKE.N RKFTERKI.EX SRLKPRCKE.ST IEVSESNI.FY SRLKPRCKE.ST IEVSESNI.FY SRLKPRCKE.ST IEVSESNI.FY SRLKPRCKE.ST RITEREN.FY SRLKPRCKE.ST RITEREN.FY SRLKPRCKEN.FY SRLTEDRY.YE SRLTEDRY.YE SRLTEDRY.YE SRLKPRCKEN.FY SRLKPRCKEN.F	300 	310 	320 DYPF. A A XKESI DI X VRDDLGAIT MT. A NRAHEN R A B SIVSPMH. SNPHCASTSSY DLRS REAP I SPFDLQAS. REAP I SPFDLQAS. REAP I SPFDLQAS. REAP I SPFDLQAS.	330 334
P76268 P77300 Q8EMM2 Q8EMM2 Q9K6X4 Q956K4 Q956K4 Q956K4 Q956K4 Q9505 Q956K4 Q9505 Q9505 Q9505 Q9505 Q9505 Q9505 Q9705 Q85246 Q85246 Q85246 Q85246 Q85246 Q85246 Q85246 Q85245 Q98524 Q85246 Q85245 Q9992 Q85246 Q85245 Q9992 Q85246 Q85245 Q9992 Q85245 Q9992 Q85245 Q7WB55 Q7WB55 Q7WB55 Q7WB51 Q85245 Q7WB55 Q7WB51 Q85245 Q7055 Q	230 RSTERTITSTE TFTPATLASRE LTTALTITSOVI PHTRHTISPE KITANTITORX PTTRNTITREL PKTPNTITHEL PKTPNTITHEL PKTPNTITHEL PLADGTIVTKL QFADGTIVTKL QFADGTIVTKL QFADGTIVTKL QFADGTIVTKL QFATTATLVSPU DGFTPTLAND COMMAND TFTLAND TFTLAND COMMAND TFTLAND TFTLAND TFTLAND TFTLAND TFTLAND TFTLAND TFTLAND TFTLAND TFTLAND TFTLAND TFTLAND TFTLAND TTTDEH C. SELLE SPIGRAEINE SPI	240 ALLPVLOOVR ALMSALQTR PKRAALQKU' JEECULOFOU JE	250 CGYCEDUREQ CGYALDRED CGYAL	260 EGLRCIAVE EGGLRCIASP EGGIRCIASP EGGIRCIASP EGGIRCIASP EGGIRCIGAP EGGIRCIGAP EGLYCLAAP EGLYCLASP IGLWSCIAVE INNTTALAP INNTTALAP INNTTALAP INNTTALAP INNTTALAP ISGTSANAAP IGLSIAVE IGLSIAV	270 FDRFGVVI WINESRVI FDRFGVVI FDRFGVVI FDNLGQVP FDNLGQVP FDNLGQVP FDNLGVV FDNLGVV FDNDRGVV FDLFGVV FDLFGVV FDLFGVV FDNDRGVV FDNDRGVV FDNGVC FDNDVC	280 A.GLSISFFT A.ALSLSTIT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSISGFT A.JSISGFT A.GISISGLA A.GISISGLA A.GISISGLA A.GISISGLA A.JSISGFT GUITVUGFOS ALIVISTM ALMUSTM ALMUSTM ALMUSTM ALMUSTM ALMUSTM ALMUSTM ALMUSTM ALMUSTM ALMUSTM ALMUSTM ALMUSTM ALMUSTM ALMUSTM ALMUSTM ALMUSTM ALMUSTM ALMUSTM ALMUSTM ALMUSTM ALSIGPTG ALSISGPTG ALSISAPAE ALSISGFTG ALSISAPAE	290 +LHFSEERL.QE SRVDDAEL.AN NRVSIDRVNLE IRLSKEL.HE ARMTDARL.QT DRMLKQKE.T RKFTERKI.ES SRUDRVCE SRUDRV	300 	310 	320 DYPF. A A WRDDLGAIT. MT. A WRDDLGAIT. MT. A MP. FPA. SOLVSPMH. SNPHCASTSSY DLRS. REAP. A NPCASTSSY DLRS. REAP. A SPFDLQAS. REAP. A SPFDLQAS. REAP. A SPFDLQAS. SPFDLQA	330 334
P76268 P77300 Q8EMM2 Q9K6K4 Q9K6K4 Q9K6K4 Q9K8C5 Q9K8C5 Q9K8C5 Q9K8U5 Q9K8U5 Q9K8U5 Q9K8U5 Q9K8U5 Q9K8U5 Q9K8U5 Q9K2U1 Q9K2U5 Q9K2U1 Q8U7A8 Q8U7A8 Q8U7A8 Q8U7A8 Q8U7A8 Q8U7A8 Q8U7A8 Q8U7A8 Q817V5 Q9K2U1 Q9FXC6 Q9	230 RSTERTITSTE FTFTATLASRE LTTALTITSOVI PHTERTITCK PHTERTITCK PHTERTITCK PHTERTITCK PHTERTITCK PLANGTACK PLANGT	240 ALLPVLOOVR ALMSALQTR PKRAALQKU' JEECUCLFQU JEECUCLFQU JEECULFQU JEECULFQU JEECULFQU JEECULFU JEECULFU JEECULFU JEECULFU JEECULFU JEECULFU JEECULFU JEECULFU JEECULFU JEECULFU JEECULFU JEECULFU JEECULFU JEECULFU JEELAIVE	250 CGYCEDUPEO CGYALDVENI LGYCYCHEENI CGYSLDVENI LGYCYCHEENI CGYSLDVENI CGYSLDVENI CGYSLDDEEII CGYSLDDEEII CGYSLDDEEII CGYSLDVENI CGYSLDVE	260 EGLRCIAVE EGGLRCIASP EGGIRCIASP EGGIRCIASP EGGIRCIASP EGGIRCIGAP EGGIRCIGAP EGLYCLAAP EGLYCLASP IGLWSCIAVE INNTTALAP INNTTALAP INNTTALAP INNTTALAP INNTTALAP INNTTALAP INNTTALAP IGLSSIAVE	270 FDRFGVVI WINESRVI FDVKGNI FDVGAIT FDVLGQIT FDVLGQIT FDVLGQIT FDVLGQIT FDVLGQIT FDVLGQIT FDVLGQIT FDVLGQIT FDVLGQIT FNTONEVV FDVLGQIT FNTONEVV FDVLGQIT FNTONEVV FDVLGQIT FNTONEVV FDVLGQUT FNTONEVV FDVLGQUT FNTONEVV FDVLGQUT FNTONEVV FDVLGQUT FNTONEVV FDVLGQUT FNTONEVV FDVLGQUT FNNNE	280 A.GLSISFFT A.ALSLSTIT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSISGFT A.GISISGLA A.GISISGLA A.GISISGLA A.GISISGLA A.GISISGLA A.JSISGFT GUITVUGPCS ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALSISGFT GVITVUGQSISA ALSISGFTG ALSVGFAL ALSUSGFT ALSVGFAL ALSV	290 +LPFSEERL.QE SRVDDAEL.AN NRVSIDRVNLE IRLSKEL.HE ANTDARL.QT DRMLKQKE.T DRMLKQKE.T RKFTERKI.ES SRLSQUGE.KN ARFTEDRL.PY IRVTERKI.ST IEVSESNI.PY SRLKPREW.PE SRLTDDRV.TE SRLKPREW.PE SRLTDDRV.TE IFADAGSLAGG.Q WITRSLEQRE RURDERQUGE IFADAGG.Q ULDRKAGG.Q ULDRKAGG.Q ULDRKAGG.Q ULDRKAGG.Q ULDRKAGG.Q ULDRKAGG.Q ULDRKAGG.Q ULDRKAGG.Q ULDRKAGG.Q ULDRKAGG.Q ULDRKAGG.Q ULDRKAGG.Q ULDRKAGG.Q ULDRKAGG.Q IAPRGVCIGSI IMPROVIDION SADAGSTRELEVEN KULADERLEVEN KULADERLEVEN SADAGSTRELEV	300 	310 	320 YPFF A A YRDDLGAIT YRDDLGAIT A YRDDLGAIT A A NRAHEN R A P C PA SOLVSPMH SOLVSPMH SOLVSPMH SOLVSPMH SOLVSPMH A SOLVSPMH	330 334
P76268 P77300 Q8EMM2 Q9K6K4 Q9K6K4 Q9K6K4 Q9K805 Q9K805 Q9K805 Q9K805 Q9K805 Q9K805 Q9K805 Q9K805 Q9K805 Q9K805 Q9K805 Q9K201 Q9K205 Q9K201 Q9K205 Q9K201 Q9K205 Q9	230 RSTERTITSTE FTFTATLASRE LTTALTITSOVI PHTERTITCK PHTERTITCK PITENTITSPE KITANTITORS PITENTITREL PKTENTITREL PLADGTTVTKL QFADGTTVCKL QFADGTTVCKL QFADGTTVCKL QFADGTTVCKL COGASTITED ELAALQOTAS. COGASTITED ELAALQOTAS. COGASTITED ELAALQOTAS. TETPLISADOLH TS. RTHDPS TETPLISADOLH TS. RTHDPS TETPLISADOLH TS. RTHDPS TC. MSRICH COGASTITED ELABELLES SPIGRAE IMPS CG. SELLLES SPIGRAE IMPS CG. SELLES SPIGRAE IMPS CG. SELLES SPIGRAE SP	240 ALLPVIDOVR ALMSALQTR PRKAALQKU' JEECUCLFQU JEECUCLFQU JEECULFQU JEECULFQU JEECULFQU JEECULFU JEECULFU JEECULFU JEECULFU JEECULFU JEEULFU	250 CGYCEDUREQU CGYALDNEEN CGYCEDUREQU CGYALDNEEN CGYSLDNEEN CGYSLDNEEN CGYSLDNEEN CGYSLDDEEN CGYSLDDEEN CGYSLDDEEN CGYSLDS CGYSLDS CGYSLDS CGYSLDS CGYSLDS CGYSLS CG	260 EGLRCIAVE EGGLRCIASP EGGIRCIASP EGGIRCIASP EGGIRCIASP EGGIRCIASP EGIRCIGAP EGIRCIGAP EGLYCLASP IGLOSCIAVE INTTALAR INTTA	270 FPDRFGVVI WINESRVI IFDVGGAIT FPINLGQVP FPDSQRQLL MKD.GELA FPINLGVP FPDSRGVVI FPDLFGVP FPDLFG	280 A.GLSISFFT A.ALSLSTIT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSISGFT A.GSISGFA A.GSISGFA A.GSISGFA A.GSISGFA A.GSISGFA A.GSISGFA A.GSISGFA A.JSISGFF GUITVUGPGS ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALMVSTWG ALSIGFTG ALSVGFAL ASSIGFTG ALSVGFAL ASSIGFTG ALSVGFAL ASSIGFTG ALSVGFAL ASSIGFTG ALSVGFAL ASSIGFTG ALSVGFAL ASSIGFTG ALSVGFAL ASSIGFTG ALSVGFAL ASSIGFTG ALSVGFAL ASSIGFTG ALSVGFAL ASSIGFTG ALSVGFAL ASSIGFTG ALSVGFAL ASSIGFTG ALSVGFAL ASSIGFTG ALSVGFAL ASSIGFTG ALSVGFAL ASSIGFTG ALSVASTWF ASSIGFTG ALSVASTWF ASSIGFTG ALSVASTWF ASSIGFTG ALSVASTWF ASSIGFTG ALSVASTWF ASSIGFTG ALSVASTWF ASSIGFTG ALSVASTWF ASSIGFTG ALSVASTWF ASSIGFTG ALSVASTWF ASSIGFTG ALSVASTWF ASSIGFTG ALSVASTWF ASSIGFTG ALSVASTWF ASSIGFTG ALSVASTWF ASSIGFTG ALSVASTWF ASSIGFTG	290 +LFSEERL.QE SRVDDAEL.AN NRVSIDXVLE IRLSKEL.HE ANTDARL.QT DRMLKQKE.T DRMLKQKE.T RKFTERKI.ES SRLSQUGE.KN ARFTEDRL.FY IEVSESNI.FY IEVSESNI.FY SRLKPREW.FE SRLTDDRV.TE SRLTEDRLF.VS ARLAQSRLAL IFADAGSLAU (UNDKAGG.Q RVTRSLEQEF RVTRESLEQEF RVTRESLEQEF RVTRESLEVENY RTKELVENY RTKELVLIQHI RTKELVING RUTRESLEVENY RTKEVLIQHI RTKEVLIQHI MYGPQ'LIQQY IAPRGVC'GEL RDGRINHDWY VISDRAUCY UNDKAGG.Q QUPRELVENY RTKEVLIQHI MYGPQ'LIQQY IAPRGVC'GEL RDGRINHDWY VISDRAUCY SCHENER RUDERLS.F RUDAERLENDI YNTKEDLTGW RUDAERLS.F RUDAERLS.F RUDAERLS.F RUDAECH.RA	300 YVAMLHTAAR FRECLQQAGL LIPPLKQAAA VITNLRDTAK LIPPLKQAAA VITNLRDTAK LIPPLKQAAA VITNLRDTAK LIPPLKQAAA VITNLRDTAK LIPPLKQAAA VITNLRDTAK LIKVTLKNAAK FIAKVKETAH LIKVTLKNAAK FIAKVKETAH LIKVTLKNAAK FIAKVKETAH LIKNYMEASU GELVRUTAASD LHDAAAEVAW QGELVRTAASD LHDAAAEVAW GPALLAWREA AARRILETAT LPILLAAREA AARRILETAT LPILLAAREA AARRILETAT LPILLAAREA QUKVSAENIE GPALTEMVHM GPALIAWVEA GPALICAWVEA GPALICAWVEA LOSAKACAA VKBILKEAAAB LOSAKACAA VKBILKEAAA VKBILKEAAA VKBILKEAAA VKBILKEAAA VKBILKEAAA VKBILKEAAA VKBILKEAAA VKBILKEAAA VKBILKEAAA VKBILKEAAA VKBILKEAAA VKBILKEAAAA VKBILKEAAA VKBILKEAAAA VKBILKEAAAA VKBILKEAAAA VKBILKEAAAA VKBILKEAAAA VKBILKEAAAA VKBILKEAAAA VKBILKEAAAA VKBILKEAAAA VKBILKEAAAAA VKBILKEAAAAA VKBILKEAAAAA VKBILKEAAAAA VKBILKEAAAAA VKBILKEAAAAA VKBILKEAAAAA VKBILKEAAAAA VKBILKEAAAAAA VKBILKEAAAAAAA VKBILKEAAAAAAAA VKBILKEAAAAAA VKBILKEAAAAAAAA VKBILKEAAAAAAAAAA VKBILKEAAAAAAAAAAA VKBILKEAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	310 	320 YPFF A YRDDLGAIT YRDDLGAIT A YRDDLGAIT A YRDDLGAIT A A NRAHEN R A P CFA SOLVSPMH SOLVSPMH SOLVSPMH SOLVSPMH SOLVSPMH SOLVSPMH SOLVSPMH SOLVSPMH SOLVSPMH SOLVSPMH SOLVSPMH SOLVSPMH SOLVSPMH SOLVSPMH SOLVSPMH SOLVSPMH SOLVSPMH SOLVSPMH SOLVSPMH SOLVSPM	330 334
P76268 P77300 Q8EMM2 Q9K6K4 Q9K6K4 Q9K6K4 Q9K8K5 Q110 P37671 P42968 Q9K8V50 Q9K8V50 Q9K8V50 Q9K8V50 Q9K8V50 Q9K8V50 Q9K201 Q9K200 Q9K201 Q9K200 Q9K20	230 RSTERTITSTE TFTPATLASRE LTALITSOVI PHTRATLASRE KITANITISVE KITANITISVE KITANITISVE KITANITISVE KITANITISVE SIGSGTIPE LARGTIVKL QFADGTIVKL QFADGTIVKL QFADGTIVKL QFADGTIVKL MYHATLVSVU OFGFPTIVVDMP OFGFPTIVVDMP GGARATNITE ELALQCTAS. TRIHDES TETPLISADOLH TS. RTHDES TETPLISADOLH TS. RTHDES C. MURSIRE C. MUSIC MARGEALES SPIGRAE INFE SPIGRAE INFE SPIGRAELEE SPIGRAE	240 ALLPVIDOVRI ALMSALQTR PRKAALQKU' JEECUCLFQU JEECUCLFQU JEECUCLFQU JEECULFQU JEECULFQU JEECULFU JEECULFU JEECULF J	250 CGYCEDUREQU CGYALDRED CGYA	260 EGLRCIAVE EGGLRCIASP EGGIRCIASP EGGIRCIASP EGGIRCIASP EGGIRCIGAP EGGIRCIGAP EGGIRCIGAP EGUDIFCSGVS IGLGVSCIAVP IGLGVSCIAVP INTTAIAP IGLGVSCIAVP INTTAIAP IGLSCIAVP IGLGVSCIAVP IGLGVSCIAVP IGLGVSCIAVP IGLSCIASP IGLSSIAVP IGLSSIAVP IGLSCIASP IGLSCIAVP	270 FPDRFGVVI WINESRVI IFDVGAIT FPDLGQP FPDSQQLL MKD.GELA FPDHLGQVP FPDSRQVL FPDLGRVP FPDLFRVP	280 A.GLSISFFT A.ALSLSTIT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSSISGFT A.JSISGFT A.GSISGLA A.GSISGLA A.GSISGLA A.GSISGLA A.GSISGLA A.JSISGFT GUITVUGPGS ALJVUSTA ALSUSGFA ALSUSFA ALSUSFA ALSUSGFA ALSUSFA AL	290 +LHFSEERL.QE SRVDDAEL.AN NRVSIDXVLE IRLSKEL.HE ANTDARL.QT DRMLKQKE.T URLTEENK.TT RKFTERKI.ES SRLKQVGE.KN ARFTEDRL.FY IEVSESNI.FY SRLKPREW.FE SRITDDRV.TE SRLKPREW.FE SRITDDRV.TE RUTEREN.ST IEVSESNI.FY VINDKAGG.Q VITRSLEQRE VINDKAGG.Q VITRSLEQRE VINDKAGG.Q URDKAGG.Q URDKAGG.Q URDKAGG.Q URDKAGG.Q URDKAGG.Q URDKAGG.Q URDKAGG.Q IAPROVISION RUTRSLEQRE RUDEIVENY RLSIDEIRLNY VINDKAGG.Q UPREKLEDDI TMTRARVAREL RUGPQILIOQV VINDKAGG.Q UPREKLEDDI TMTRARVAREL RUGPGINHOW VINDKADCHING VINDERLS.F RUDAERSTREL RUDEINNIN VINDERLS.F RUDAERSTREL SEDLQAERENI NMCAAFGG.R RUDDELS.F RUDAEQELA.M	300 	310 	320 YPFF. A A YRDDLGAIT. YRDDLGAIT. A NRAHEN. R A DI SNRAHEN. R A A SNRAHEN. SNRAHEN. SNRAHEN. SNRAHEN. SNRAHEN. SNRAHEN. SNRAHEN. SNRAHEN. SNRAHEN. SNRAHEN. SNRAHEN. SNRAHEN. SNRAH. SSRAH. SEL. SSPVADD. JARSNOSKRNU. AGRRSGML. UT.	330 334
P76268 P77300 Q8EMM2 Q9K6K4 Q2EXA29 Q9K6K4 Q2858K2 Q9WR50 P3767724 P162968 Q9K8U50 Q9K8U50 Q9K8U50 Q9K8U50 Q9K8U50 Q9K2U1 Q9K2U1 Q80255 Q9K2U1 Q881V2 Q882V3	230 RSTERTITSTE FTFTATLASRE KITANITSOVI PHTRHTISPE KITANITISVE PUTRHTISPE KITANITIRSL PUTRHTISPE KITANITIRSL PUTRHTISPE SIGSGTIPEL PLANGTTIKLL PADGTVUKK KSKVI AYHATLVSVV OFGPTITLVDMP OFGPTNARST CQAAKAINTIE ELANGTTVKL QAAKAINTIE ELANGTTVKL OFAGNARST GOARAINTE ENTRISPE ETPRISADOLH TS. RTHDDES C. MARGEN RTTIDEN F. YITTDEH ENMPSIRK RMMSIRK MSCMARE PHSJMNPHRK DBG. ELFLE SPTGRABINE SPTGRABINE SPTGRABINE CDMSRLR MACDEGAALER VRHTPHTKTPES LESLEPCKQ RFTBATLITPE LARDELPLE STRAFLICE CLAREPLKLBA RFSARDISRD CLAREPLKLBA RFSARDISRD MAGNARST RFSARDISRD CLAREPLKLBA RFSARDISRD MAGNARST CLAREPLKLBA RFSARDISRD MAGNA STRAFLST	240 ALPVIDOVRI ALMSALACTR ALFVIDOVRI ALMSALACTR PRKAALOKU' DEECICLFOU D	250 CGYCEDUREQU CGYALDNEEN CGYCEDUREQU CGYALDNEEN LGYCYCHEEN CGYALDNEEN	260 EGLRCIAVP EGLRCIASP EGLRCIASP EGIRCIASP EGIRCIASP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIGAP EGIRCIAVP IGLRCIASD IGLRSIAVP EGIRCIASA IGTSALAVV IGLRCIASA IGTSALAVV EGIRCIASS EGIRCIAVP EGIRCIASP EGIRCIAVP EGIRCIASP EGIRCIAVP EGIRCIASP EGIRCIAVP EGIRCIAVP EGIRCIASP EGIRCIAVP EGIRCI	270 FPDRFGVVI WINESRVI IPDVGAIT FPDVGAIT	280 A.GLSISFFT A.ALSLSTIT A.SFSISGFT A.SFSISGFT A.VSISGFS A.SFSISGFT A.VSISGFS A.SFSISGFT A.SFSISGFT A.SISGFT A.SISGFS A.SISISGFS A.GLSISGFS A.GLSISGFS A.GLSISGFS A.GLSISGFS A.SISVGFA A.SISVG	290 +LIFSEERL QE SRVDDAEL AN NRVSIDXELSE RUDDAEL AN NRVSIDXELSE SRVDDAEL AN NRVSIDXE SRVDDAEL AN NRVSIDXE SRVDAEL SRVDAEL SRVDAEL SRVDAEL SRVDAEL SRVDAEL SRVDAEL SRUDX SRLXER SRITEDRL FY SRITEREN F SRITEDRU SRUDXES SRITEDRU SRUDXES SRITER S	300 	310 	320 YPFF. A. YRDDLGAIT. YRDDLGAIT. YRDDLGAIT. A. NRAHEN. R. A. SNRAHEN. R. A. SNRAHEN. SNRAHEN. SNPHCASTSSY JLRS. SPFDLQAS. S	330 334

Figure 1. Multialignment of the 53 sequences used as the seed to construct the IcIR family profile. Shaded in green is the conserved segment that best defined the IcIR family. Bars in blue *above* the sequence indicate the HTH binding motif. Highlighted in light brown are the residues that are conserved in $\geq 60\%$ of the aligned sequences, and in purple are shown the amino acids with $\geq 80\%$ conservation.

No.	Accession no.	Description	Microorganism	Profile N-score	No. of amino acids
1	Q9AH06	Putative transcriptional regulator	Rhodococcus erythropolis	8.465	274
2	Q62IS0	Putative transcriptional regulator	Burkholderia mallei	8.337	180
3	Q9EWL2	Putative transcriptional regulator	Streptomyces coelicolor	8.189	255
4	Q6FBA6	Putative transcriptional regulator	Acinetobacter sp.	7.932	259
5	Q82D43	Putative differentiation regulon	Streptomyces avermitilis	7.895	213
6	Q9HW60	Probable transcriptional regulator	Pseudomonas aeruginosa	7.877	256
7	Q93T33	Differentiation regulon SamR	Streptomyces ansochromogenes	7.711	213
8	Q5YU96	Putative transcriptional regulator	Nocardia farcinica	6.441	299
9	Q5Z0G4	Putative transcriptional regulator	Nocardia farcinica	6.386	221
10	Q7W1I1	Putative transcriptional regulator	Bordetella parapertussis	5.558	236
11	Q7WPH1	Putative transcriptional regulator	Bordetella bronchiseptica	5.558	236
12	Q762I2	Hypothetical protein orf12	Rhodococcus rhodochrous	5.190	202
13	Q5PJJ6	Probable global regulatory protein homolog	Salmonella paratyphi	5.117	228
14	Q8ZM49	Putative transcriptional regulator	Salmonella typhimurium	4.914	228
15	Q7W4F4	Putative DNA-binding protein	Bordetella parapertussis	4.730	250
16	Q57K18	Putative transcriptional regulator	Salmonella cholerae-suis	4.712	238
17	Q7WFW4	Putative DNA-binding protein	Bordetella bronchiseptica	4.620	250
18	Q7WBG1	Putative transcriptional regulator	Bordetella parapertussis	4.546	235
19	Q7WMY2	Putative transcriptional regulator	Bordetella bronchiseptica	4.546	235
20	Q5YWY8	Hypothetical protein	Nocardia farcinica	4.197	295
21	Q938D9	Putative transcription regulator	Mycobacterium smegmatis	4.068	229
22	Q7VV60	Putative DNA-binding protein	Bordetella pertussis	3.957	250
23	Q7NW36	Hypothetical protein	Chromobacterium violaceum	3.810	100
24	Q5YWX8	Putative transcriptional regulator	Nocardia farcinica	3.608	294
25	Q73W85	Hypothetical protein	Mycobacterium paratuberculosis	3.479	295
26	Q4NBL4	Regulatory protein, IclR	Arthrobacter sp. FB24	3.295	128
27	Q6UP88	Putative enoyl-CoA hydratase	Alcaligenes eutrophus	3.258	448
28	Q8Z3V0	Probable global regulatory protein	Salmonella typhi	3.258	228
29	Q89SX5	Transcriptional regulatory protein	Bradyrhizobium japonicum	3.166	250
30	Q93RL8	IclR-like protein (fragment)	Acinetobacter sp. NCIMB9871	3.111	127
31	Q60G69	Putative regulatory protein (fragment)	Rhodococcus sp. DFA3	3.019	129
32	Q9AGJ8	Probable transcriptional regulator	Corynebacterium glutamicum	2.872	137
33	Q5YTZ3	Putative transcriptional regulator	Nocardia farcinica	2.798	288
34	Q6QID7	Gp12	Burkholderia cenocepacia	2.761	100
35	Q9S4Y5	Glyoxylate regulatory protein (fragment)	Salmonella enteritidis	2.725	148
36	Q65WD5	Hypothetical protein	Mannheimia succiniciproducens	2.504	108
37	Q5GRB7	Probable transcriptional regulator	Alcaligenes xylosoxydans	2.375	71
38	Q7W593	Putative transcriptional regulator	Bordetella parapertussis	2.338	215
39	Q7WCS4	Putative transcriptional regulator	Bordetella bronchiseptica	2.338	215
40	Q7VZR7	Putative transcriptional regulator	Bordetella pertussis	2.320	215
41	O67479	Hypothetical protein aq_1510	Aquifex aeolicus	2.136	219

Table 1. Proteins listed as IclR family members in Interpro (Zdobnov and Apweiler 2001) but detected as non-IclR family members by the new profile

their identification as members of the Ic1R family, but at present, such information is not available.

The IclR profile with an N-score threshold of 8.5 unequivocally identified proteins as members of the IclR family, and no false positives were found among all prokaryotic proteins that were analyzed. These results indicate that the new profile is highly effective in detecting members of the IclR family.

Using the profile defined above for the IcIR family, we searched for members of this family in 228 complete microbial genomes available in NCBI (release 13-8-05). This resulted in the detection of 477 IcIR members in 91 microbial genomes belonging to 60 genera of Grampositive, α , β , and γ -proteobacteria and archaea, indicat-

ing a wide taxonomic distribution. This information can be accessed at http://www.bactregulators.org/.

The database of bacterial transcriptional regulators: BacTregulators

The profile that best defines the IcIR family members, the sequences of all members of the family, their sequence alignment, as well as the available structural information together with a number of references on IcIR proteins have been gathered in the BacTregulators database (http://www.bactregulators.org). This database, which can be searched with a number of different parameters such as organism, name of the regulator, accession code, or simple

text information as input information, is, in our view, a convenient tool to identify and study IclR family members.

The structural information available for IclR family proteins supports the profile as a useful tool for assigning proteins to this family

Currently, five PDB entries are available that contain structural information on IclR family members. The only full-length 3D structure of an IclR-family member is that of *Thermotoga maritima* TM0065 (PDB: 1MKM) (Zhang et al. 2002). The other four structures correspond to the effector binding domains of IclR, the glyoxylate shunt regulatory protein, YaiJ and KdgR from *E. coli* (PDB: 1TF5, 1TF1, 1YSQ, and 1YSP, respectively). All structures have in common that they were obtained in the absence of target promoter DNA or effector molecules. Structural alignments with the DALI algorithm have shown that these proteins share a similar structure, as witnessed by Z-scores >22 (see http://www.bactregulators.org/structure.php).

The TM0065 IcIR protein was shown by X-ray crystallography (Zhang et al. 2002) to consist of two α/β domains: a small N-terminal DNA-binding domain with the HTH motif and a larger C-terminal effector-binding domain (Fig. 2). The latter domain consists of a fivestranded, curved β -sheet, which is flanked on both sides by several α -helices. The 79–amino acid fragment that contains the IcIR profile is highlighted in yellow in Figure 2. The profile sequence forms a long loop starting at Gly151, followed by a sequence of three helices (H6–H8, of which H6 is buried and H7 and H8 are surfaceexposed), and terminates with strands S5 and S6, which form the flanking part of the sheet (Fig. 2). The amino acids with the highest score (indicating that little variation is tolerated) in this new profile are shown in ball-andstick mode. Gly151, which is labeled in Figure 2, has been proposed to play a key role in tetramerization of the protein, which is likely to occur when the protein is bound to DNA (Zhang et al. 2002). This role in tetramerization is thus likely to be responsible for the high score of Gly151 in the IclR profile. The remaining high-scoring amino acids are all located on the loop, the short buried helix, and the two strands. None of the important amino acids is located on the two long surface-exposed H7 and H8 helices. All the important amino acids are buried to a large degree and maintain multiple interactions with neighboring residues. These residues thus fulfill an important structural role, which accounts for their weight in the IclR profile.

Acknowledgments

Work at the authors' laboratory was supported by grants CICYT (2003-0515), Ministerio de Medio Ambiente (059/2004/3), and Junta de Andalucía to group CIV191. We thank M. Mar Fandila and Carmen Lorente for secretarial assistance, and Karen Shashok for improving the use of English in the manuscript.



Figure 2. Schematic representation of the three-dimensional structure of the IcIR dimer of *Thermotoga maritima*. Secondary structure elements are annotated, and the helix-turn-helix DNA binding domain (HTH) is shown in purple. The 79–amino acid fragment comprising the new IcIR profile is highlighted in yellow. The nine amino acids with the highest score in the IcIR profile are shown in ball-and stick-mode. Gly151 proposed to be involved in tetramerization is annotated.

References

- Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etweiler, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., and Sonnhammer, E.L. 2002. The Pfam protein families database. *Nucleic Acids Res.* 30: 276–280.
- Bucher, P., Karplus, K., Moeri, N., and Hoffman, K. 1996. A flexible motif search technique based on generalized profiles. *Comput. Chem.* 20: 3–24.
- Busenlehner, L.S., Pennella, M.A., and Giedroc, D.P. 2003. The SmtB/ArsR family of metalloregulatory transcriptional repressors: Structural insights into prokaryotic metal resistance. *FEMS Microbiol. Rev.* 27: 131–143.
- Gallegos, M.T., Schleif, R., Bairoch, A., Hofmann, K., and Ramos, J.L. 1997. The AraC/XylS family of transcriptional regulators. *Microbiol. Mol. Biol. Rev.* 61: 393–410.
- Gerischer, U., Segura, A., and Ornston, N.L. 1998. PcaU, a transcriptional activator of genes for protocatechuate utilization in. Acinetobacter. J. Bacteriol. 180: 1512–1524.
- Guazzaroni, M.-E., Terán, W., Zhang, X., Gallegos, M.T., and Ramos, J.L. 2004. TtgV bound to a complex operator site represses transcription of the promoter for the multidrug and solvent extrusion TtgGHI pump. *J. Bacteriol.* **186**: 2921–2927.
- Hofmann, K., Bucher, P., Falquet, L., and Bairoch, A. 1999. The Prosite database, its status in 1999. Nucleic Acids Res. 27: 215–219.
- Jiang, H. and Kendrick, K.E. 2000. Characterization of ssfR and ssgA, two genes involved in sporulation of Streptomyces griseus. J. Bacteriol. 182: 5521–5529.
- Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Barrell, D., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P., et al. 2003. The InterPro Database, 2003 brings increased coverage and new features. *Nucleic Acids Res.* 31: 315–318.
- Nègre, D., Cortay, J.C., Galinier, A., Sauve, P., and Cozzone, A.J. 1992. Specific interactions between the IcIR repressor of the acetate operon of *Escherichia coli* and its operator. J. Mol. Biol. 228: 23–29.
- Nguyen, C.C. and Saier Jr., M.H. 1995. Phylogenetic, structural and functional analyses of the LacI-GalR family of bacterial transcription factors. *FEBS Lett.* 377: 98–102.
- Orth, P., Schnappinger, D., Hillen, W., Saenger, W., and Hinrichs, W. 2000. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat. Struct. Biol.* 7: 215–219.

- Pabo, C.O. and Sauer, R.T. 1992. Transcription factors: Structural families and principles of DNA recognition. Annu. Rev. Biochem. 61: 1053–1095.
- Ramos, J.L., Gallegos, M.T., Marqués, S., Ramos-González, M.I., Espinosa-Urgel, M., and Segura, A. 2001. Response of gram-negative bacteria to certain environmental stresses. *Curr. Opin. Microbiol.* 4: 166–171.
- Ramos, J.L., Martínez-Bueno, M., Molina-Henares, A.J., Terán, W., Watanabe, K., Zhang, X., Gallegos, M.T., Brennan, R., and Tobes, R. 2005. The TetR family of transcriptional repressors. *Microbiol. Mol. Biol. Rev.* 69: 326–356.
- Reverchon, S., Nasser, W., and Robert-Baudouy, J. 1991. Characterisation of kdgR, a gene of Erwinia chrysanthemi that regulates pectin degradation. Mol. Microbiol. 5: 2203–2216.
- Rigali, S., Derouaux, A., Giannotta, F., and Dusart, J. 2002. Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. J. Biol. Chem. 277: 12507–12515.
- Schell, M.A. 1993. Molecular biology of the LysR family of transcriptional regulators. Annu. Rev. Microbiol. 47: 597–626.
- Schultz, J., Copley, R.R., Doerks, T., Ponting, C.P., and Bork, P. 2000. SMART: A Web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* 28: 231–234.
- Schumacher, M.A., Miller, M.C., Grkovic, S., Brown, M.H., Skurray, R.A., and Brennan, R.G. 2002. Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR. *EMBO J.* 21: 1210– 1218.
- Yamamoto, K. and Ishihama, A. 2003. Two different modes of transcription repression of the *Escherichia coli* acetate operon by IclR. *Mol. Microbiol.* 47: 183–194.
- Zdobnov, E.M. and Apweiler, R. 2001. InterProScan—An integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17:** 847– 848.
- Zhang, R.G., Kim, Y., Skarina, T., Beasley, S., Laskowski, R., Arrowsmith, C., Edwards, A., Joachimiak, A., and Savchenko, A. 2002. Crystal structure of *Thermotoga maritima* 0065, a member of the IclR transcriptional factor family. J. Biol. Chem. 277: 19183–19190.
- Zhang, H.-B., Wang, C., and Zhang, L.H. 2004. The quormone degradation system of *Agrobacterium tumefaciens* is regulated by starvation signal and stress alarmone (p)ppGpp. *Mol. Microbiol.* 52: 1389–1401.