PROTEIN SEQUENCE MOTIFS

FAT: a novel domain in PIK-related kinases

Phosphatidylinositol kinases are found in all eukaryotes and serve important functions in phosphatidylinositol (PI) signaling pathways¹. In addition to the PIkinase domain, most of these proteins have a number of accessory domains, usually involved in protein–protein interactions, that specify the role in a given pathway. A few examples of such domain organizations are shown in Fig. 1a. Recently, a new subfamily of the PI-kinase superfamily has emerged, called PIK-related^{2,3}. Although these proteins are large (2000–4000 amino acids), they only share similarity to classical PI kinases in the ${\sim}300\text{-}\mathrm{amino}\text{-}\mathrm{acid}$ kinase domain.

Members of the PIK-related family appear functionally distinct, as none of them has been shown to phosphorylate lipids, such as PI; instead, many have Ser/Thr protein kinase activity^{4–7}. Despite this functional disparity, we will refer to this domain as the PI-kinase domain. Many PIK-related proteins are involved in cell-cycle checkpoint control [e.g. ATM, ATR, DNA-PK, ESR1 and Rad3 (reviewed in Ref. 8)]. Dysfunction can result in a range of diseases, including immunodeficiency, neurological disorder and cancer⁹.

It has previously been noted that members of the PIK-related family share a unique motif at the extreme C terminus¹⁰, which we call FATC. However, it has proved difficult to define shared domains in the large N-terminal portions. Although sequence similarity between various members extends upstream of the PIkinase domain, it usually tapers off in an irregular way. Several different N-terminal domain configurations have been hinted at in schematic diagrams, but no domain was supported by a multiple alignment^{5,11–14}. Tentative assignments of a leucine zipper and a DNA-polymerase-processivity-factorbinding site (P-site) have been made in the N-terminal region¹⁵. Although these regions were shown to be functionally important, the putative leucine-zipper motif is unlikely to be a true leucine zipper, given the atypical composition of the nonleucine residues¹⁶. The P-site motif is uncertain too, as five out of nine amino acids are never observed at the same position in other known cases¹⁷.

A group of proteins distantly related to PI kinases comprises the TRRAP



Figure 1

(a) Modular architecture and tree based on the PI-kinase domain (PI-KIN in red rectangle) of several representatives of the PI 3-/PI 4-kinases and all known PIK-related and TRRAP proteins. Within the PIK-related and TRRAP subfamilies, all members share the same domain architecture, as indicated by the tree sections colored blue and yellow. The FAT domain is only present in the FRAP, ATM and TRRAP subfamilies and always coexists with the FATC domain. Other domains in the PI 3-/PI 4-kinases are named according to Pfam (Ref. 21) but are shown without the 'PI3K' prefix. These are PI3K_C2 (PF00792), C2-like domain in PI kinases; PI3Ka (PF00613), accessory (PIK) domain; PI3K_rbd (PF00794), ras-binding domain. For more information on these domains, see http://www.cgr.ki.se/Pfam. The tree was generated by PHYLO_WIN (Ref. 22) using the neighbor-joining method with pairwise gap removal and PAM distances, from a multiple alignment of the PI-kinase domain in TRRAP is denoted (PI-KIN) and drawn with a hollow because it lacks the catalytic residues despite sequence homology. (b) Multiple alignment of the FAT domain in all known sequences. Residues were colored by Belvu based on average pairwise BLOSUM62 score exceeding 0.3 (gray shade) or 1.3 (blue). Sequence names are either the Swissprot ID or a TREMBL accession number followed by a dot and a name mimicking a Swissprot ID, except for TRRAP_HUMAN for which only an EMBL accession number is available. Because Q22258 could not be classified unambiguously we used a question mark.

PROTEIN SEQUENCE MOTIFS

TIBS 25 - MAY 2000

(b) Q62388.ATM_HUMAN TEL1 YEAST Q2258.? CAEEL ESRI YEAST Q24135.MBI41_DROME Q1358.ATR_HUMAN Q2391.RAD3 SCHPO O01436.FRAP_CAEEL FRAP HUMAN 014356.FRAP_SCHPO TOR1 YEAST P78527.DNAPK_HUMAN Q18627.TRAP_CAEEL Q48233.TRAP_ARATH AFO76974.TRAP_HUMAN YAMP_SCHPO YHP9_YEAST OC3288.ATM_HUMAN Ol4356.FRAP_SCHPO TOR1_YEAST P78527.DNAPK_HUMAN Q18667.TRRAP_CAEEL O48823.TRRAP_ARATH AF076974.TRRAP_HUMAN YAMB_SCHPO YHP9 YEAST Q62388.ATM HUMAN Q62388.ATM_HUMAN TELL YEAST Q22259.CAEL ESRL YEAST Q24135.ME141 DROME Q13535.ATR HUMAN Q92391.RAD3_SCHPO 001438.FRAP_CAEEL FRAP_HUMAN 014356.FRAP_SCHPO TOR1 YEAST Q18667.TRRAP_CAEEL Q48823.TRRAP_ARATH AF076974.TRRAP_HUMAN YAMB_SCHPO YHP9_YEAST YHP9_YEAST Q62388.ATM_HUMAN TEL1 YEAST Q22258.? CAEEL ESR1 YEAST Q24135.ME141 DROME Q13535.ATR HUMAN Q92391.RAD3_SCHPO 001438.FRAP_CAEEL G13535.ATR HUMAN O14356.FRAP_CAEEL O18527.DNAPK_HUMAN Q18667.TRRAP_CAEEL Q48823.TRRAP_ARATH AF076974.TRRAP_ARATH AF076974.TRRAP_ARATH NAME_SCHPO YHP9_YEAST YHP9_YEAST Q62388.ATM HUMAN TEL1 YEAST Q22258.? CAEEL ESR1 YEAST Q24135.MEI41 DROME Q13535.ATR HUMAN Q13535.ATR HUMAN Q1438.FRAP_CAEEL G14356.FRAP_SCHPO TOR1 YEAST P78527.DNAPK HUMAN Q18667.TRRAP_CAEEL Q48823.TRRAP_ARATH AP705974.TRRAP_ARATH AP705974.TRRAP_ARATH NAME_SCHPO VHD0_VEACT YAMB_SCHPO YHP9 YEAST Q62388.ATM HUMAN TEL1_YEAST Q22258.?_CAEEL 22258.? CAEEL ESRI YEAST Q24135.MEI41 DROME Q13535.ATR HUMAN Q92391.RAD3_SCHPO O01438.FRAP_CAEEL FRAP HUMAN O14356.FRAP_CAEEL O4852.TRAP_CAEEL O4823.TRRAP_CAEEL O4823.TRRAP_CAEEL O4823.TRRAP_ARATH AF076974.TRRAP_HUMAN YAME SCHPO YHP9_YEAST YHP9_YEAST Q62388.ATM_HUMAN TEL1 YEAST Q22258.? CAEEL ESR1 YEAST Q24135.ME141 DROME Q13535.ATR HUMAN Q92391.RAD3_SCHPO 001438.FRAP_CAEEL G13535.FRAP_CAEEL O14355.FRAP_CAEEL O14355.FRAP_CAEEL Q1867.TRRAP_CAEEL Q1867.TRRAP_CAEEL Q18623.TRRAP_ARATH AF076974.TRRAP_ARATH XAME_SCHPO YHP9_YEAST

	EELOYSENWKL
1346 LVSRASINGETARADSIDESIDESGEDASURLEQFIFICENCESIDSVEGARVQVSIDMSIQU 1640 TLAVASFRSKAVTRADVMERSIFIEKKONIGHKGPLOKLNAAMHEDEGVACVSAIKAEPSLKEOI	LVNRLVERQQD
1386 TLGIVSLNCGFHARALFYWEOHIRNATAPYAALESDYRVLOBIYAGIDDPDEIEAVSLNFHDYSFDOOL 1438 VLGEWAROTKAFAKACFYKEMSVLKKSGSMOTTFTRKVKLEDNDCOSUTTYANKINVOERAAGVVEYARENEMNFOMEG	LLH <mark>E</mark> NSGTWDS RWYEKLNEWEK
1382 LLGERAAKCRAYAKALHYKELEFOK	TWYEKLHEWED
1220 VISHARSCHVIARADAITELGEVQGFJKEV	TWFEKLERWED
2884 QQPYGIKLLEEALLKLIPAGLPAKQIIQSA 2685 LIEFISSKHECWHTGIRLLENHIWTIPKQLNNTLLREMKVAPGLAGDIETLESIGTLYNEISBFDQFAAIWERRAVFPDTMRAMSAMQLGDI	MELAQSYLEKS
2591 LIKYIGKTYNAWHLALALLESHVMLFMNDSK.CAESJAELWRLMDBDMRFCLWKKRSITAETRAGL 2686 FRSTLMLEHQAFEKGLSLQIKPKQTT.EFYEQESITPPQQEILDSDAELYSLHQBBDMMACLWQKRCKYSETATAI	S.LVQHGFWQR A.YEQHGFFEQ
2484 LLVYLSKTYGLHHYCILLLENSLQNNPGLSEDELTVYHKSCLDATSDIYYSLDEHDLYHCLWRRANFLETEVAT 2622 LVKYLAISYNAWYQSINILESIQSNSIDNTKIIEANEDATLELYVNLOEBDMFYCLWRRAKYTETNIGL	S.HEQCHEWEK S.YEQIGLWDK
2051 .NLVTYDLETSISSS.[0]TRQSGIIQALQNLG.LSHILSVYL.KGLDYERREWCAELQELRYQAAMRNMQM	GLCASAGQ
1810 FNNADEDANYTTSLE. [0]EEKESLIKATEDSGFUGLTSLLESRLSGSSDVYKWNLELGDWKLLTPKVVDSKAKGLYN 1665 .SAACEARMTGKGKP. [0]FNSTEAIQKLIDELNCLEYSQIERNEQEDYLNSLKTLSQUVNIDN	AIKNLPQDVGF DIGPSPHI
1479 .AQDCFNVLGKFSDD.[0].PKTTTRMLKSMYDHQLYSQIISNSSFHSSDGKISLSPDVKEWYSIGLEAANLEGNVQTLKNW 1427 .MITSWEOLLSSTDO.[1]OPDHVRAMIDAYLRDTPKTAOLIADG.LWORLSDRYSDOCFAECKSELLWRLGSY	VEQIESLRNI
1718 ATACYDRAIQLEPD.[0]QIIHYHGVVKSMLGLGQLSTVITQVNGVHANRSEWTDELNT	DLVENYLAA
1528 .ALGAVELEEKKKSSC[20] AEEARMHEMRCLEALGRODELNSKSV.VWADQRGNRNDSVRDEINKKQLDHKMAVIAARGAWAVDNW 1459 ALVAVDKKMDTNKD [0] DDELMLGDMPCLEALGRO GOLHOOCCERWTLV NDETOAKMAP MAAAAAVGLGOO	ERMADYV.S
1306 .ALAAMEHREREGDS.[0] SFEINIGKLRCYYALGDWDHLSELAQKAWTSEQEHREAIAPLAAAAAMGLGQW	NLISEYV.S
2967 .AAKONDEALNKOODU [0] DGEPTEAEKDFWELASLDCYNHLA.EMKSLEY.CSTASIDSENPDINKINSEPFY	QETYLPYMIR
2767 .MSSIMELLAFINFN[6]VSFIINKEIDAWBEM	AYLKDHVIPK
2771 .ROESYEKAMDKAKKE[5]ASPAIFPEYQUWEDH. MIRCSKELL.QUEALT. EYGQSKGHINPYUVLECARRVSNW 2569 .ROLVWEHAQLKVCTG[0]SLPYSPTEHGFWLDH. MILCAQKLN.QUEVLFDFSKQEGCAELYLECARRLSDW	TAMKEALVQVEVS STEQDTLEKATKS
2703 .MQQLWEVAQVKARSG[0]ALPYSQSEYALWEDNMIQCAEKLQ.HMDVLTELAKHEGFTDLLLECGMRVADW	NSDRDALEQSVKS
2124 EVEGTSYHESLYNAL.QCLRNREFS.TFYESLRYASLFRVKEVEELSKGSLESVYSLYPTFSRLQAIGSPEN 1895 AEKSLEKSLLTIFDSRQHFISQTEWMDTLNAIIEFIKIAAIPQDVTSFPQTLMSIMKADKERLNTHDFYDHKTT	SGE.LFSRS RHTLMNVLS
1732 FSRNIEYWATESTIL.KMIRNDERDEIVNNAIENAKSKVIERLSECAIGGSCSYEIATPFIVELOKLNSIVE 1565 DDREVLLQYNIAKAL.IAISNEDPLRTQKYIHNSFRLIGTNFITSSKETTLLKKQNLMKLHSLYDISF.	LKN.VSNDE LSSAKDKFE
1505 WPAQCSQGCLKLRRP.LTTRIEFDS.LLDGMRESVLEELRSCSAVQQHSYANAYDAVLKLHLVHGIHC: 1794 DGKSTTWSVRLGQLL.LSAKKRDIT.AFYDSLKLVRAEQIVPLSAASFERGSYQRGYEYLVRLHMLCBTEH:	SQELVEKLE SIKPLFQHS
1542 SNLESFEAKLGSIFY.QYLRKDSFAELTERLQPLYVDAATAIANTGAHSAYDCYDISKLHAINDFSR 1637 VISENTQDGAMLRAV.VAVHNDENT.KAMGLIEKVREMIDSELTAMANESYERAYIPMVSVCOMASIES	IAETDGIVS AIE.YKTRP
1535 MIPRDTHDGAFYRAV.LALHQDLFS.LAQQCIDKARDLLDABLTAMAGESYSRAV	VIQ.YKLVP IID.YKKNM
1483 VMKPKSPDKE FFDAI.LYLHKNDYD.NASKHILNARDLLVT BISALINESY NRAY SVIVRTOIITEFEE 3046 SKLKLLLOGRADOSLLTFIDKAMHGELOKAILELHYSORLSLLVLLODDVDRAKYYIONGIOSEMONYSSIDVLLHOSBUTKLOSVOATTE	IIK.YKQLP
2877 IPPSFHLDYTLFNLM.STVMRMNEN.SSPTHMKERCKIAIQECTEAHISRWRALPSVVSYGHVKILQAMNLVREIDE 2746 AQVETPKLELVOSY.FALHDENSN.GVGDAENTVCKGVDLALEOWNOLPEMSVHAEVPLUOOFCOLVEVOE	STDIRIALLEA
2858 CPKEMAWKVNMYRGY.LAICHPEEQ.QLSFIERLVEMASSLAIREWRRLPHVVSHVHTPLOAAQIIETOR 2649 LSPETSLEPH TADAL LVINKTOPKMGSUTERSPILECM OFSLEPWOOLD KEVKO SHISLUHHOFTVELOR	AAQINAGLQP
2783 VMDVPTPRRQMFKTF.LALQNFAESRKGDQEVRKLCDEGIQLSLKKWVSLPIRYTPAHKWLHGFQYMEFLE	ATQIYANLHT
2202VTDRERSEAYWKWQKHSQLLKDSDFSFQEPLMALETVILETLVQ.K.EMERSQGACSKDILTK.HLVEFSVLARTFKNTQ	LPERA.IFKIKQY
1811LSAFNSDFWKNIQKRTDDSEQ.KISILEPILRVRRSMLDIRMQSMTGRDKENIRSRIVE.VHLQSARIARLTCCFF	RAQLSLIN
1542QDRDEDNQEKLMKNYFDD QYRLQIVQP. QVRIQESIYSPRRNILGELQRRL TDRNHLPHLKTELAR. IWLNSAQINRNACQLQ	RAQLYILK
16/2 PDINEW GEDS IN WARDEN ON STRAKET INDEXAMPLEMENTS FOR WEDGES RAVARAGENO 1618 DNLDIVLRRR LSQVAPIGKF.KHQILSTHLVGYEKFINKKT.A.BIYLEIARISRKNGQFQRAFNAILKAMD	LDKPLATIEH
1712ERRPIALLBERDUGGER.NVEWUGLIHERGUVESPUER.H.PLEVESSMERKUGGENSMERAVERELSEPA 1610ERREIIRQIMWERLOGCQR.IVEDWQKILMVRSLVVSPHED.M.RTWLKYASLCGKSGR.LALAHKTLVLLLG.VI	DPSRQLDH
1457QTENNIDS.LKRIWKRRLEGCQK.NVDWHNILKFRALVISPQDS.F.EWWILLADLCKKSDK.LKLENQCHIILMG.K 1558PNSEKKLH.YQNLWTKRLLGCQK.NVDLWQRVLRVRSLVIKPKQD.L.QIWIKFANLCRKSGRMRLANKALMMLLEGGN	DPSNALPN
3146RQGRLSSQVPLKK.LLKNWINKYPDAKMDPMNIMDDITINKCFPLSKIEEKLTPDPEDNSMAVDQDGDPSDKMEVQEQEEDISSLI 2963PSNKVDQALMGDMKSLMKVFRNRTPTTSD.DMGFVSTWYDWRNQIHGMMLQRF.EYMDKVGLNVAATGNQSIVPIHSMAQ	
	RSCKFSMKMKMID AQLAVAKHAK
2829 VSGNTAVGGLGNRYADLKDILETUWRDKTFNEWD.NMTWYDMLQWRNEMYNVVIDAFKDFATSNSPLHHLEFKD 2938TNLGRNNS.LHDMKTVVKTWRNRLPIVSD.DLSHMSSIFMWRQHHYQAIVTAYENSSQHDPSSNNAMLGVHA	RSCKFSMKMKMID AQLAVAKHAK KAWNVNKLAR SASAIIQYGK
2829 VSGNTAVGGLGNRYADLKDILETWRLTTNEND.MTTWYDMLGWRSMYNVVIDAFKDFATSNSPLHLGFRD 2938TNLGRNS.LHDMKTVVKTWRNRLFIVSD.DLSHMSSIFMWRQHHYQAIVTAYENSSQHDPSSNNAMLGVHA 2731TNIHNIDNKLKDIKVVLQGWRELENVMD.DIDIWSDLAWFGVFKSINKVFLPLVSIAQQSTN.KSNTNSVSYLVRCYHE 2865TTVQNLDSKAQEIKRILQAWRDRLPNTWD.DVNMWNDLVTWRQHAFQVINNAYLPLIPALQQSNSNSNINTHAYRGYHE	RSCKFSMKMKMID AQLAVAKHAK KAWNVNKLAR SASAIIQYGK LAWIINRFAH IAWVINRFAH
2829 VSGNTAVGGLGARYADLKDILETWRLTTPAEND.AMTWYDMLGWRSMYNVVIDAFKDFATSASPLHELEFKDI 2938TNIGRNDS.LHDMKTVKTURNRLETVSD.DLSKSSIFMKGCHTQAIVTPATENSSOHDFSSNNAMLGVHA 2731TNIHNIDNKLRDIKVVLQGWRBRLENVWD.DIDIMSDLIAWRQSVFKSINKVFLPLVSIAQQSTNKSNTNSVSYLYRCYHE 2865TTVQNLDSKAQEIKRILQAWRDRLENTWD.DVNMMNDLVTWRQHAFQVINNAYLPLIPALQQSNSNSNINTHAYRCYHE 2291 NSAICGISEWHLEEAQVFWAKKEQSLALSILKQMIKKLDSSFKDKENDAGLKVIYAECLRVCGSWLAETCLENPAVIMQTYLEKAVI	RSCRFSMKMKMID AQLAVAKHAK KAWNVNKLAR SASAIIQYGK LAWIINRFAH IAWVINRFAH KVAGSYDGNSREL
2929 VSGNTAVGGLGARYADL&DILETWRLHTPAEND.MMTWTYDHUWRSMYINVVIDAFKDFATSNSPLHELEFKDI 2938TNIGRNDS.LHDMKTVKTYGRRELEFVSD.DLSTGSSIFMRGCHYQAIVTFATSNSQHDFSSNNAMLCVHA. 2731TNIHNIDNKLKDIKVVLQGKRELEFVSD.DLDITGSDLARGQSVFKSINKVFLPLVSIAQQSTNKSNTNSVSVLYRCYHE 2865TTVQNLDSKAQEIKRILQAWRDRLENTWD.DDVNMMDLVTWRQHAFQVINNAYLPLIPALQQSNSNSNINTHAYRGYHE 2291 NSAICGISEWHLEEAQVFWAKKEQSLALSILKQMIKKLDSSFKDKENDAGLKVIYAECLRVCGSWLAETCLENPAVIMQTYLEKAVV 2076 QNEKNISESILYDDFKLLINVPMDQIKARLVKWSSESRLEPAAAIYEKIVNWDINVEDHESCSDVFYTLGSFLDEQAQKL.RSNG 1893 AKKVLPFENKIVLEEAKLQLQTSDELNGMSLLDSIISKNFGDLHTIYDTQQSVMLDVQKSAKLKIEHYQEETKNLEFSY	RSCKFSMKMKMID AQLAVNKLAR KAWNVNKLAR SASAIIQYGK LAWIINRFAH IAWVINRFAH KVAGSYDGNSREL EIRDREHRSYTGK VQMLRISHMIKAG
2929 VSGNTAVGGLGARYADLED LEVENUKTURNELTENENDI.NMTWYDMLGWRSMYNVVIDAFKDFATSMSPLHELEFRU 2938TNLGRNDS.LHDMKTUVKTURNELTIVSD.DLSLSSIFMMEQHYQAIVTATENSSQHDFSSNNAMLCVHA. 2731TNIHNIDNKLADIKVULOGKRBLENVND.DLDIGSSITAMRQSVFKSINKVFLEUVSIAQQSTNKSNTNSVSYLYRCYHE 2865TTVQNLDSKAQEIKRILQAMRDRLENTWD.DUNMANDLVTWRQHAFQVINNAYLPLIPALQQSNSNSNINTHAYRCYHE 2291 NSAICGISEWHLEEAQVFWAKKEQSLALSILKQMIKKLDSSFKDKENDAGLKVIYAECLRVCGSWLAETCLENPAVIMQTYLEKAV 2076 QNEKNISSESILDDFKLINDYPMDIKAKLVKSSSESKLEPAAITKKIIVNDINVEDHESCSDVF.YTLGSFLDEQAQKL.RSNG 1893 AKKVLPFENKIVLEEAKUQLJTSDELNGMSLLDSISKNFGDLHTIYTDQSVNLDVQKSAK.LKTEHYQEETKNLFSS 1713 CLERR.LEQAELEFAELLWQKGENDRALKIVQBIHEKYQENSS.VNARDRAAVLLKFTEWLDLSNNSASEQIIKK 1671 AAEYQF.SGITERAKLLWQKGQVMAMNYLEGQISMKSQCGQWVKQLAAEQRLFFRGKYLQAVYSAESMHLCAD.	RECKFSMKMKMID AQLAVAKHAK KAWNYNKLAR SASAIIQYGK LAWIINRFAH IAWVINRFAH KVAGSYDGNSREL EIRDREHRSYTGK VQMLRISHMIKAG QYQDIFQIDSKND AVLKYFGKAIAVH
2929 VSGNTAVGGLGNRYADLKDILETNREHDINNETWYDMILGNRSMYNVVIDAFKDFATSNSPLHELEFKDL 2938TNIGRNNS.LHDMKTVNYNRRALEIVSD.DIDITSSITAMRGSWFKSINKVFLPLVSIAQQSTNKSNTNSVSVLYRCYHE 2865TTVQNLDSKAQEIKRILQANRDELENTWD.DIDITSSITAMRGSWFKSINKVFLPLVSIAQQSTNKSNTNSVSVLYRCYHE 2291 NSAICGISEWHLEEAQVFMAKKEQSLALSILKQMIKKLDSSFKDKENDAGLKVIYAECLRVCGSWLAETCLENPAVIMQTYLEKAVI 2076 QNEKNISSESILJDDFKLINVPMDIKARLVKNSSESRLEPAAIYEKIVNNDINVEDHESCSDVF.YTLGSFLDEQAQKL.RSNG 1893 AKKVLPFENKIVLEEAKLQLGTSDELNGNSLLDSISFKDKENDAGLKVIYAECLRVCGSWLAETCLENPAVIMQTYLEKAVI 2076 QNEKNISSESILJDDFKLINVPMDIKARLVKNSSESRLEPAAIYEKIVNNDINVEDHESCSDVF.YTLGSFLDEQAQKL.RSNG 1893 AKKVLPFENKIVLEEAKLQLGTSDELNGNSLDSISISFKDKENDAGLKVIYAECLRVCGSWLAETCLENPAVIMQTYLEKAVI 2076 QNEKNISSESILJOPKLINVPMDIKARLVKNSSESRLEPAAIYEKIVNNDINVEDHESCSDVF.YTLGSSDESTKNLFSS 1713 CLERR.LEQAELEFAELIKGGENDRALKIVQSIHEKYGEN.SS.VNARDRAVLLKFTEMLDLSNNSASE.QIIK 1671 AAEYQP.SGLFIERAKLUQKGDVMAMNYLEEQLSIMRSGCQGNVKQLAAEQRHLFFRGKYLQAV.YSAESMHLCADL 1947 AGESR.LAELYVERAKNUKSGD.VHAALIVLQKGVELCFPENETPEGKNMLIHGRAM.LLVGGFMEETAN.FESN. 1698 AQWWHEQGGHKRAISELN.N.NNMFDLVDEHERFKDRKETLGMP.LKGKVFKLKIKMLGKAQCLG.K.KDLETYY.HKAV.	RECKFSMKMKMID AQLAVAKHAK KAMNVNKLAR SASAIIQYGK LAWIINRFAH IAWVINRFAH KVAGSYDGNSREL EIDDREHRSYTGK VQMLRISHMIKAG QYQDIFQIDSKWD AVLKYFQEAIAVH AIMKKYRDVTACL EIYSECENTHYYL
 2930 VSGNTAVGGLGARYADLØD LEVRALUTNENDINGT VENDLGARSENTA	RECKFSMKMKMID AQLAVAKHAK KAMNVNKLAR SASAIIQYGK LAWIINRFAH IAWVINRFAH IAWVINRFAH EIEDREHRSYTGK VOMLRISHMIKAG QYQDIFQIDSKND AVLKYFQEAIAVH AIMKYKDVTACL EISECENTHYYL EISFVRQQVSPQY EWQLNLQGINEST
 2930 VSGNTAVGGLGNRYADLKD1EFINGHULMTNNENDIN.NNTWYDMLGWRSMYN	RECKFSMKMKMID AQLAVAKHAK KAMNVNKLAR SASAIIQYGK LAWIINRFAH IAWVINRFAH IAWVINRFAH EIDDREHRSYTCK VQMLRISHMIKAG QYDIFQIDSKND AVLKYFQEAIAVH AIMKYKNOVTACL EIYSECENTHYYL EISFVRQUYSQY EWQLNLGUINEST KWKKSLQDSVNQE EWRIATQPWRNT
 2930 VSGNTAVGGLGNRYADLKD1ETVNENDI.NMTWTYDHLGNRSMYN	RECKFSMKMKMID AQLAVAKHAK KAMNVNKLAR SASAIIQYGK LAWIINRFAH IAWVINRFAH IAWVINRFAH EIDREHRSYTCK VQMLRISHMIKAG QYQDIFQIDSKMD AVLKYFQEAIAVH AIMKKKNOVTACL EISFECENTHYLL ELSFVRQVVSPQY EWQLNLGGINEST KWKKSLQDSVNQE EWRIATQPSVRQN
 2930 VSGNTAVGGLGNRYADLEDILETINGLGTINGENDINNETWYIDHLGNRENETNY	RECKFSMKMKMID AQLAVAKHAK KAMNVNKLAR SASAIIQYGK LAWIINRFAH IAWVINRFAH IAWVINRFAH EIEDREHRSYTGK VQMLRISHMIKAG QYQDIFQIDSKMD AVLKYFQEAIAVH AIMKKKDJDSKND AVLKYFQEAIAVH EISFVRQVSPQY EWQLNLQGINEST KWKKSLQDSVNQE EWRIATQFNWRMT LLGTTYRIIANAL YHRANIHSVLDQA RLKGDFHLKLNDT ALKGMFLAVINS
 293 VSGNTAVGGLGNRYADLÆDILETINGENDINNETVY TUDIUGNENEMIN	RECKFSMKMKMID AQLAV AKHAK KAMNV NKLAR SASAI IQYGK LAWII NRFAH IAWVI NRFAH EIEDREHRSYTGK VOMLRISHMIKAG QYQDIFQLOSKMD AVLKYFQEAIAVH AUKKYKDOXSAU ELSFVRQVSPQY EWGLNLQGINEST KWKSLQOVSPQY EWGLNLQGINEST KWKSLQOVSPQY EWGLATQPNWRWT LLGTTYRIIANAL YHRANIHSVLDQA RLEGDFHLKLNDT ALEGMFLAQINKS
 293 VSGNTAVGGLGNRYADLÆDILETINENDI.NMTUNTIDELUNENENTN	RECKFSMKMKMID AQLAV AKHAK KAWNV NKLAR SASAI IQYGK LAWII NRFAH IAWVI NRFAH EIEDREHRSYTGK VOMLRISHMIKAG QYQDIFQIDSKMD AVLKYFQEAIAVH AIMKKYKDYTACL EISSECENTHYLL ELSFVRQVSPQY EWGLNLQGINEST KWKKSLQOVSPQY EWGLNLQGINEST KWKSLQDFHLKLNDT ALKGMFLALQINKS TLKGMFQNRLGEK TLKGMFQNRLGEK
2939 VSGNTAVGGLGNRYADLÆDILETURENDINNT VYDIDUGNRSMYN	RECKFSMKMKMID AQLAV AKHAK KAWNV NKLAR SASAI IQYGK LAWII NRFAH IAWVI NRFAH EIEDREHRSYTGK VOMLRISHMIKAG QYQDIFQIDSKMD AVLKYFQEAIAVH AIMKKYKDYTACL EISSPRQCYSPQY EWGLNLQGINEST KWKKSLQDFULKUNT LLGGTYRIIANAL YHRANIHSVLDQA RLKGMFLALQINKS TLKGMFQNRLGKK TLKGMFQNRLGKK TLKGMFQNRLGKK TLKGMFQNRLGKK TLKGMFQNRLGKK TLKGMFQNRLGKK TLKGMFQNRLGKK TLKGMFQNLGKKFE
2939 VSGNTAVGGLGARYADLÆDILETURENDINNT VYDIDUGURENEMIN	RECKFSMKMKMID AQLAV AKHAK KAWNV NKLAR SASAI IQYGK LAWII NRFAH IAWVI NRFAH EIEDREHRSYTGK VQMLRISHMIKAG QYQDIFQIDSKMD AVLKYFQEAIAVH AIMKKYKDYTACL EISSVRQCYSPQY EWGLNLQGINEST KWKKSLQOVSPQY EWGLNLQGINEST KWKSLQOSVNOE EWRIATQPNWRNT LGGTYRTIANAL YHRANIHSVLDQA RLKGDFHLAQINKS TLKGMFQNRLGEK TLKGMFQNRLGEK TLKGMFQNRLGEK TLKGMFQNRLGEK TLKGMFQNRLGEK TLKGMFQNRLGEK TLKGMFQNRLGEK TLKGMFQNRLGEK TLKGMFQNRLGEK TLKGMFQNLGEK TLKGMFQNLGEK TLKGMFQNLGEK TLKGMFQNLGEK TLKGMFQNLGEK TLKGMFQNLGEK TLKGMFQNLGEK TLKGMFQNLGEK TLKGMFQNLGEK TLKGMFQNLGEK TLKGMFQNLGEK TLKGMFQNLGEK TLKGMFQNLGEK TLKGMFQNLGEK TLKGMFQNLGEK TLKGMFQNLGEK TLKGMFQNLGEK TLKGNFQNLGEK TLKKK TLKKYK TLKKK TLKKK TLKKK TLKKK TLKKK TLKKK T
2929 VSGNIAVGGLGARYADLÆDILETURENDINNETVYTDELGARSSIFAMAGGHYG. AIVTAFKDATENSGHDFSSNAAHLGYRA 2731TNIGNNS.LHDMKTVKTURGNETIVSD.DIZDISSIFAMAGGYFKSINKVFFLVSIAQQSTN.KSNTNSVSYLYRCYHE 2865TYVQNLDSKAQEIKILQAARDLEPNVMD.DIDISSIFAMAGGYFKSINKVFFLVSIAQQSTN.KSNTNSVSYLYRCYHE 2291 NSAICGISEWHLEERAQVFLAKKEG.SLALSILKQNIKKLDSSFKDKENDAGLKVIYAECLRVCGWLAETCLENPAVIMGTVIEKAV 2076 QNEKNISESILVDDFKLLINVPMODIKARLVKNSSSSILEPAAAIYEKIVNMDINVEDHESCSDVF.YTLGSFLDEQAQKL.RSNG 201693 AKKVLPFENKUVLEERAVGLOTSELLINGKLLDSIISKNFGDLHTIYTDTQGVNLDVKSGK.LKIEHVGEETKN.LFSS 21713 CLERR LEQAELEFAFILKKGGENDRALKIVQSIHEKYOEN.SS.VMARDRAAVLKFTEMDL.SNNSASE.QIIK 2017 AGESR.LEELYVERAKULSKGDQVMAMNYLEGUSIMRSGCQGNVKQLAAEQRHLFFRGK.YLQAV.YSAESMH.LCAD. 2018 AKVUPFENKUVJQOLKALSSDVHQALVVLQKGVELCFPENETP.PEKKNMLINGRGAGLGG.KDLETYY.HKAV. 2017 AGESR.LEELYVERAKULSKGD.VHQALVVLQKGVELCFPENETP.PEKKNMLINGRGAUGGG.KDLETYY.HKAV. 2018 AQWWHQGQHRKAISELMESLN.NNNFDLVDEHEERFKNRKETLGPP.LKGKVFLKLIKMLGKAGQLGG.KDLETYY.HKAV. 2019 AGESR.LEELYVERAKULSKGD.VHQALVVLQKGVELCFPENETP.PEKKNMLINGRAVULFFT.LVGK.YLLAKKVLYQDHKHEARAN.LDVGK.MQVG 2016 AGESR.LEELYVERAKULSKGD.VHQALVVLQKGVELCFPENETP.PEKCHNMLINGRAGQLGG.KDLETY.HKAV. 2019 FXAD.PUVYALKINGKG.D.VHQALVVLQKGVELCFPENETP.PEKCHARICAXVLLKGGGUGLK.KLLAKFT.LKG 2019 AGWWHQGQHRKAISELMESLN.NNNFDLVDEHEERFKNRKETLGPP.LEKGVVFLKLIKMLGKAGQLGG.KDLETY.HKAV. 2019 FXAD.PVVVALKINGKGARAL.DAHNNKNINFFFXATGRAFLIPPSTKEPARICAXVLLKGGWTELK.SKTSNN.MQVG 2016 AKKLN.PPVVYALKINGKSARK.IDAFQMUHVQTMQQA. 2014 SARKQNNFSLAKKLIKGKKS.KTRDWLVSVQSVCISULSKGSNGGGC.SUVLIVKTVSLLDLSNSSS.SERS.FLARCF.HKLG 2019 KKLD.PVVVALKINGKSS.KTRDVKSVQSVCISULSKGNGGGC.SUVLISKSSTAPVVEYTK.LLARCF.LKLQ 2021 VARVHLEPVCINQLKKINGKSNEVYCNGKS.MKRQARVLKKGERASGINLINST.NLEY.PVERT.A.AFFY. 2016 LARKQCUVVALDLINSRIFTP.NIETQEAFLKLREQAKCH.YQ.NN.NELCEALEVLDVRLD.LQKDQV.AALU 2022 VARVHLEPVCINQLKKYLYKSVCASVLERQAKCH.YQ.NN.NELCEALEVLDV.NLVKY.SLLCHAFSK.LLAKPKN. 2034 SARKONNFSLAKKLESS.KTRDVKSEFENKQTLLKK[29]ECALRAL.RECRKRFLCKAVENYIN 2144 STU.JQAKAKLS.LARFSDTQYQRIENKKSEFENKQTLLKK[29]ECALRAL.	RECKFSMKMKMID AQLAV AKHAK KAWNV NKLAR SASAI IQYGK LAWII NRFAH IAWVI NRFAH IAWVI NRFAH EIEDREHRSYTGK VQMLRISHMIKAG QYQDIFQIDSKMD AVLKYFQEAIAVH AINKKYKDYTACL EISSFVRQVSPQY EWGLNLQGINEST KWKSLQOVSPQY EWGLNLQGINEST KWKSLQOVSPQY EWGLNLQGINEST KWKSLQONNC EUSFVRQVSPQY EWGLNLQGINEST TLKGMFLAQINKS TLKGMFLAX TLKGMFLAQINKS TLKGMFLAX TLKG
2939 VSGNIAVGGLGARYADL&DIGLTUNGLGARLETUNG. NATUWIDALGARGAN,	RECKFSMKMKMID AQLAV AKHAK KAWNV NKLAR SASAI IQYGK LAWII NRFAH IAWVI NRFAH IAWVI NRFAH EIEDREHRSYTGK VQMLRISHMIKAG QYQDIFQIDSKND AVLKYFQEAIAVH AINKKYKDYTACL EISFVRQVSPQY EWGLNLQGINEST KWKSLQOVSPQY EWGLNLQGINEST KWKSLQOVSPQY EWGLNLQGINEST KWKSLQONNC EUSFVRQVSPQY EWGLNLQGINEST TLKGMFLAQINKS TLKGNKS TLKGMFLAQINKS TLKGNKS T
2929 VSGNIAVGGLGARYADLÆDILETURENDINNETVYTDELGARSSIFMAGGHYQ. AIVTAFKDATENSDEDESSINAMLCYHAL 2731TNIGNNS.LHDMKTVKTURGNETIVSD.DISNSSIFMAGGEVFXSINKVFFLEVSIAQQSTN. KSNTNSVSYLYRCYHE 2865TYVQNLDSKAQEIKILQAMERDLEPNVMD.DIDINSDLTAMEQSVFKSINKVFFLEVSIAQQSTN. KSNTNSVSYLYRCYHE 2291 NSAICGISEWHLEERAQVFLAKKEQ.SLALSILKQMIKKLDSSFKDKENDAGLKVIYAECLRVCGWLAETCLENPAVIMQTYLEKAV 2076 QNEKNISESILVDPKLLINVPMODIXARLVKNSSSSILEPAAAIYEKITVNMDINVEDHESCSDVF.YTLGSFLDEQAQKL.RSNG 1893 AKVLPFENKIVLEERALVGLTSELINGSLLDSIISKNFGDLHTIYTDTQQSVNLDVKSSK.LKIEHYQEETKN.LFSS 1713 CLERR.LEQALEFAELLVQKGENDRALKIVQSIHEKYQEN.SS.VMARDRAAVLKFTEMDL. SNNSASE.QIIK 1671 AAEYOP.SGLFIERAKLUQKG. DQVMANNYLEGLSIMRSGCQGNVKQLAAEQRHLFFGK. YLQAV.YSABSMH. LCAD 1947 AGESR.LAELYVERAKULSGG. D.VHQALVVLQKGVELGFPENETP.PECKNNLINGRAM.LLVG.FMEETAN.FESN 1698 AQWWHQGHKRAISELMESINNNNFDLVDEHERFKNRKETLGAP.LKGKVFLKLIKNIGKAGQLGG.KDLETY.HKAV 1797 PLUVLALKACUQDHKDEIRAL.BLANHNNKINFPIFKATGRIGDS.SKLUVILASSSYS SEERS.FLARCF.HKLG 1640 LKLN.PEVYTAKKNIGSG.J.VHQALVUQKGVELGFPENETP.LGKVKVEKLIKNIGKAGQLGG.KDLETY.HKAV 1797 PLUVLALKGVUQDHKDEIRAL.BLANHNKINKINFFFKATGRIGDS.SKLUVILASSSYS SEERS.FLARCF.HKLG 1643 AKKUNFFSLUKINGKAK.IDAFQHNQHVQTMQQAGHAIATEDQQHGLHK.LMARCF.LKLG 1644 SARKONNFSLUKKIKKINGSK.KIDAFQGRAVKCILGKVEKKGSGCG.EQUIVKIKTSLLDENNYSS.YLSKNILAFRDNII 3051 NLGFHNITKDLNKLAGITATPMNDAQDKVCTYGKTLERMANSAADEVKN.GERASGINLINST.NLEY.FPKIK.AETF 3161 LARKQCIVVVADLINKINGKSS.KIDAFQGRAVKCILGKVEKKGNECMQCLEVINKT.NLEY.FPKKK.AETF 3244 SARKONNFSLUKKIKSK.TRDDVISVOYGYCICHSCICHSQCKSCILLGVINNT.NLEY.FFKEMT.AEFF 3253 VARKHNMPDVCISQLARITIP.NIEIQEAFLKLREQAKCH.YQ.NM.NELTIGLDVISNT.NLEY.FFKEMT.AEFF 3249 R.[0]NGQMKALSLS.LARFSDTQYQR.IENMKSEFENKQTLLKR[29]ECHRAL.REDKKRFLCKAVENYIM 2174 ST [0]LKALELIYK.NTKLPENER.KSK.KDAKAYLMENKLQVINNS.SELUCALEVINNT.NLEY.FTKEMT.AEFF 3253 VARKHNMPDVCISQLARITIP.NIEIQEAFLKLREQAKCH.YQ.NM.NELTIGLDVISNT.NLY.FTKEMT.AEFF 3254 VARKHNMPDVCISQLARITIP.NIEIQEAFLKLREQAKCH.YQ.NM.NELTIGLDVISNT.NLY.FTKEMT.AEFF 3254 VARKHNMPDVCISQLARITIP.NIEIQEAFLKLREQAKCH.YQ.NM.NELTIGLDVISNT.NLY.F	RECKFSMKMKMID AQLAVAKHAK KAMNVNKLAR SASAIIQYGK LAWIINRFAH IAWVINRFAH IAWVINRFAH EIDDREHRSYTCK VQMLRISHMIKAG QYDIFQIDSKND AVLKYFQEAIAVH AIMKYKNOVTACL EIYSECENTHYYL ELSFVRQUVSPQY EWQLNLGINEST KWKKSLQDSVNQE EWQLNLGINEST KWKKSLQDSVNQE EWRIATQPHUKLNDT ALKGMFLAQINKS TLKGMFDAWRLGEK TLKGMFDKRLGEK TLKGMFDKRLGEK TLKGMFDKRLGEK TLKGMFDKLAY CLLSGEEHDLWVF NDIIDKFCIWF NDIIDKFCIWF NDIIDKFCIWF DONTHENLD CUTURUT.LWFD QDITLRLN.LWFK
 293 VSGNTAVGGLGNRYADLÆDILETURENDINGNENTING VDELGUNGNENTING VVIJDAVKDEATENS	RECKFSMKMKMID AQLAVAKHAK KAMNVNKLAR SASAIIQYGK LAWIINRFAH IAWVINRFAH IAWVINRFAH EIDDREHRSYTCK VQMLRISHMIKAG QYDIFQIDSKND AVLKYFQEAIAVH AIMKKINOVTACL EIYSECENTHYYL ELSFVRQOVSPQY EWQLNLGGINEST KWKKSLQDSVNQE EWGLNLGGINEST KWKKSLQDSVNQE EWGLNLGGINEST KWKKSLQDSVNQE EWGLNLGGINEST LLGTTYRIIANAL YHRANIHSVLDQA RLKGDFHLKLNDT ALKGMFLAQINKS TLKGMFDAVING TLKGMFDSKLRAY CLLSGEEHDLWVF NJIIDKFGLWFE LWWS RENLFXUITFWLD QSMPRLIS.LWLD DLWD EDTLRIM.LWFN QDTLRLIT.LWFN QAYPALVWE
2939 VSGNIAVGGLGARYADLÆDILETURENTIVSD.DLSNSSIFMARGEHYQ. ALVTAFKDATENS	RECKPSMKMKMID AQLAV AKHAK KAMNV NKLAR SASAI IQYGK LAWII NRFAH IAWVI NRFAH IAWVI NRFAH EIDDREHRSYTCK VQMLRISHMIKAG QYQDIFQIDSKND AVLKYFQEAIAVH AIMKKINDVTACL EISPVRQVSPQY EWQLNLGGINEST KWKKSLQDSVNQE EWGLNLGGINEST KWKKSLQDSVNQE EWGLALGGINEST KWKKSLQDSVNQE EWGLALGGINEST KWKKSLQDSVNQE EWGLALGGINEST KWKKSLQDSVNQE EWGLALGGINEST KWKKSLQDSVNQE EWGLALGGINEST KWKKSLQDSVNQE EWGLALGGINEST KWKKSLQDSVNQE EWGLALGGINEST KWKKSLQDSVNQE EWGLALGGINEST KWKKSLQDSVNQE EWGLALGGINEST KWKKSLQDSVNQE EWGLALGGINEST KWKSLQDSVNQE EWGLAUG ENGLAUN ILGGFUN ILGGFUN STRENEVITFWID QSMPRLIS.LWID QDTLRLIN.LWFR QATPA LWUS RSHIARY.LYL
<pre>299 VSGNTAVGGLGARYADL&DIGLTVAKUMERLETVSD.DISMSSIFMAGGHYQAIVTAFNSSGHDFSSNAAHLCYHA 2731TNIGNNS.LHDMKTVKTUKTUKTUSD.SIFMAGGHYQAVTFLEVENSGHDFSSNAAHLCYHA 2751TYUGNLDSKAGEIKFILGAMERDELENVUD.DIDLSSDIAMEGSVFKSINKVFFLUSIAQQSTN.KSNTNVSVIYECYHE 2291 NSAICGISEWHLEEAQVFLAKKEGSLALSILKQMIKKLDSSFKDKENDAGLKVIYAECLRVCGSWLAETCLENPAVIMQTVIEKAV 2076 QNEKNISESILVDOFKLINVPHODIKARLVKNSSESRLEPAAAIYEKITVNDINVEDHESCSDVF.YTLGSFLDEQAQKL.RSNG 1893 AKKVLPFENKIVLEEAAUGLGTSELEINGSLLDSIISKNFGDLHTIYTDTQGSVNLDVKSSK.LKIEHVGEETKN.LFSS 1713 CLERR.LEQAELEFAEILKGGENDGALLKIVQEIHEKYGEN.SS.VWARDRAAVLLKFTEWLDL.SNNSASEQIIK 1671 AAEVOP.SGLFEERAKLUSGGENDGALLKIVQEIHEKYGEN.SS.VWARDRAAVLKFTEWLDL.SNNSASEQIIK 1673 AGWUHGGGHKAAISELMSSLNNNNFDLVDEHEGISIMSGCQGNVKQLAAEORHFFEKK.YLGAVSSKSML.CAD 1947 AGESR.LAELYVERAKWLSKG. D.VHQALIVLQKGVELCFPENETP.PEGKNMLIHGRAMLLUGR.FMEETAN.FESN 1698 AQWWHGGGHKAAISELMSSLNNNNFDLVDEHEEREKNRKETIGAP.LKGKVELKLKKWIGKAGQLGG.KDLETY.HKAV 1797 PLUVLALARGUYQODHKDEAIRAL.EDLANHWNKINFFEXATGRALIPSTKEAARICAVLLKLEGWTELK.SKISNN.MGVG 1688 PLPTVH.PQVTYAYKKNNKSAKK.IDAFQHQHVQTMQQA. QHAIATEDQQH.QELHK.LMARCF.LKLG 1630 IKKLN.PEVVYAQKKINGAGK.IDAFQHQHVQTMQQA. 0HAIATEDQUHK.QELKKINGAGK.IDAFQHQHVQTMQQA. 0HAIATEDQUHK.QELKKINGAGK.IDAFQHQHVQTMQQA. 10451 NLGFHNITKDLNKLAGGTAIPMMDAQDKVCTYGKTLENDANSAADERVKN.ELCEALEVLEVVRIDD.LQKDQV.AALU 1912 IARKQGUVVALDUIKSKINGS.SKLVGLESKCKSGGCS.EQULTVKTSLLENNYSS.YLSKNILAFRDQNI. 3051 NLGFHNITKDLNKLAGGTAIPMMDAQDKVCTYGKTLENDANSAADERVKN.ELCEALEVLEVVRIDD.LQKDQV.AALU 1914 STKQUVUVUCULEKKYGSSTHEYQEAFVKIRGQAKCT.YQ.NM.NELTTGLDVISNTNLEY.FTKEMT.AEFF 2953 VARKHNMPDVCISQLARTYTLPNIETQEAFLKLREQAKCH.YQ.NM.NELTTGLDVISNTNLYY.FFVDKK.AEFF 2954 VARKHNMPDVCISQLARTYTLP.NIETQEAFLKLREQAKCH.YQ.NM.NELTTGLDVISNTNLYY.FGTVQK.AEFF 2955 VARKHNMPDVCISQLARTYTLP.NIETQEAFLKLREQAKCH.YQ.NM.NELTTGLDVISNTNLYY.FGTVQK.AEFF 2956 VARKHNMPDVCISQLARTYTLP.NIETQEAFLKLREQAKCH.YQ.NM.NELTTGLDVISNTNLYY.FGTVQK.AEFF 2957 VARKHNMPDVCISQLARTYTLP.NIETQEAFLKLREQAKCH.YQ.NM.NELTTGLDVISNTNLYY.FTNETKA.AEFF 2958 VARKHNMPD</pre>	RECKFSMKMKMID AQLAV AKHAK KAMNV NKLAR SASAI IQYGK LAWII NRFAH IAWVI NRFAH IAWVI NRFAH IAWVI NRFAH EIEDREHRSYTGK VQMLRISHMIKAG QYQDIFQIDSKND AVLKYFQEAIAVH AIMKKIKDJDSKND AVLKYFQEAIAVH AIMKKIKDYTACL EISFVRQVSPQY EWGLNLGGINEST EWGLNLGGINEST EWGLNLGGINEST EWGLNLGGINEST EWGLNLGGINEST EKMEKSLDDSVNGE EWGLAUGGINEST EKMEATEN LGGTYRIIANAL YHRANIHSVLDQA RLKGMFLAQINKS TLKGMFQNRLGEK TLKGMFQNRLGEK TLKGMFQNRLGEK TLKGMFQNRLGEK TLKGMFQNRLGEK TLKGMFLSKLRAY CLLSGEEHDLWVP NDIIDKFCGLWFE LWDS RENLFYVITFWID QSMPRLIS.LWVD QDTLRLT.LWPD QDTLRLT.LWPD QDTLRLT.LWPD QAYPALVVE RKFLAKY.LWVL
<pre>299 VSGNIAVGGLENKYADLKDILETURLETURINED.NMT.WTUNIDURUNENKI. VVIDARKDFATSSPLHHLEFKD 2936TTVGNIDSKLEDKYTVGTENKELEFKYVES DISLESSFFWR/OHTQAVTLPLUFALQGSISSININTKAYEYLE 2965TTVGNIDSKAGEIKHILGAARDELFNYWE DUNMENDLYTHEGRAFQUINNAVLPLIFALQGSIKSNINSVSULYECTHE 2991 NSAICGISEWHLEEAQVFWAKKEQ.SLALSILKQMIKKLDSSFKDEENDAGLKVIYAECLRVCGSWLAETCLENFAVINQTYLEKAV 2076 QMEKNISESILYDDFKLLINVEMGLURINED.DUNMENDLYTHEGRAFQUINNAVLPLIFALQGSNSSNINTHAYE'THE 2911 NSAICGISEWHLEEAQVFWAKKEQ.SLALSILKQMIKKLDSSFKDEENDAGLKVIYAECLRVCGSWLAETCLENFAVINQTYLEKAV 2076 QMEKNISESILYDDFKLLINVEMGLURINVEDHESCSDVF. YTLGSFLDEQAQKL.RSNG 1893 AKVLPFFENKIVLEEAKLGUG ENDRALKIVQBIHEKYQEN.SS.VNABRAAVLLKFTEWDL SNISASE.QIEK 1617 AAEYOP.SGETHERAKLLGGG ENDRALKIVQBIHEKYQEN.SS.VNABRAAVLLKFTEWDL SNISASE.QIEK 1619 AGWWHQQGHKAISELMSGGD.VHQALUVLGGVELCFPENETP.PECKNELHFFRGKLVQAV. YSABSMHLCAD 1947 AGESE.LAELVVERAKMI GKGD.VHQALUVLGGVELCFPENETP.PECKNELHFFRGKLVQAVSABSMHLCAD 1948 AGWWHQQGHKAISELMG SLNNMFDLVDEHEERPKNRKETLGNP.LKGKVFLKLIKWLGKAQQLGL. KDLETYY.HKAV 1797 PLLVLALAKGLYQDDEKDEATAAL.EDLANHNNKINPIFKATGELPFSTKEPARICAKVLLKLGEATELKSKISNN. MQVG. 1688 ELPTYH.PQVTYAYMKMM KSARKIDAFQHQHFVQTMQQQAOHAIATEDQHKGELKK.LMARCF.LKLG 1639 TFKAPPPVVYAQLKYTMATGAYKEADNHLIGFTSKLAHDLGLPFNNM.IAGSVKJ.SEESS.ELEACF.HKLG 1639 TFKAPPPVVYAQLKYTMATGAYKEADNHLIGFTSKLAHDLGLPFNNMIAGSVKJ.SEESS.ELAACFLKLG 1639 TFKAPPPVVYAQLKYTMATGAYKEADNHLIGFTSKLAHDLGLPNNM.IAGSVKJ.SEESS.NINSS.Y.VILASKNILAFEDQNT. 1051 NLGFHNITKVLLNKLGETATPMMDADKVCTYGKTLENGARACHEKKGERASGUNINST. NLEY.FPDKIK. AEFF 2953 VARKHNMEDVCIJEKKMESK.TDDWVCAFVKLENGARGGCS.EQULVITYVSLL.HENNYSS.Y.VILASKNILAFEDQNT. 1915 NTG 01 GGPKVTHE.TTPL.NTETQEAFLKRGARACHEKKGERASGUNINST. NLEY.FFDKK. AEFF 2953 VARKHNMEDVCIJEKKMESK.TDDVGARYKIKENGARCH.YQ. NM.NELTTGLDVISNT. NLEY.FTKEMT.AEFF 2953 VARKHNMEDVCIJEKKMESK.TDDVGARYKIKENGARCH.YQ. NM.NELTTGLDVISNT. NLEY.FTKEMT.AEFF 2953 VARKHNMEDVCIJEKKMESK.TDDVGARKIKENGARCH.YKFSS.MULLAKAKSENKULVYLKAKKENKIKANGKAL.YFKEMS</pre>	RECKFSMKMKMID AQLAV AKHAK KAMNV NKLAR SASAI IQYGK LAWII NRFAH IAWVI NRFAH IAWVI NRFAH EIEDREHRSYTGK VQMLRISHMIKAG QYQDIFQIDSKND AVLKYFQEAIAVH AUKKYFQEAIAVH AUKKYFQEAIAVH AUKKYFQUSQUSPQY EWQLNLQGINEST KWKKSLQDSVNQE EWQLNLQGINEST KWKKSLQDSVNQE EWQLNLQGINEST KWKKSLQDSVNQE EWQLNLQGINEST KWKKSLQDSVNQE EWQLNLQGINEST KWKKSLQDSVNQE EWQLNLQGINEST KWKKSLQDSVNQE EWQLNLQGINEST KWKKSLQDSVNQE EWQLNLQGINEST KWKKSLQDSVNQE EWQLNLQGINEST KWKKSLQDSVNQE EWQLNLQGINEST KWKKSLQDSVNQE EWQLNLQGINEST KWKKSLQDSVNQE EWQLNLQGINEST KWKLKSLQDSVNUE ENFLYNUE ILGTYFNUD QSTRLLT. WFD QTLRLLT. WFD QTLRLLT. UWF QTLRLLT. WFN AKYLAKY.LWFL RELLCRI.LWFI EBLLCRI.LWFI
293 VSGRIAVGGLGNNYADLED LEYWKHETPNEWD. NHIVYTYDELUMENENYN. VVIDAFKDATSN. SPLHHLCFKD 2936TNUGRNNS.LDWKTVVUKWWN.RDKDI'SD. DLSG SSJFWMCHIMYQAITTAYENSS. OHDPSNNAHLCFKA. 2731TNIHNIDMKLEDIKVVIQGWERDEPNWD. DLDI SDLIAWGSVFKSIN. KYFLEVSLAQGSTN. KSNINSYSIYKTYT 2865TTYQNLDEKAGEIKRILGWARDELPNWD. DLNIAWLGYFKSIN. KYFLEVLYSLAQGSTN. KSNINSYSIYKTYT 2931 NSAICGISEWHLEEAQVFWARKEQ. SLALSILKQMIKKLDSSFKDKENDAGLKVIYAECLRVCGSWLAETCLENPAVIMQTYLEKAV 2036 (DENNISESILYDDFKLINVPHDQIKARLVKWSSESRLEPAAIYEKIIVMDINVDHESCSDVF. YTLGS. FLDEQACKL.RSNG 2039 AKKVLPFENKIVLEEAKLQUGT. SPELMOMSLLDSISISNFODLHTIYDTOGSVNLDVQKSKAL.LKIE. VQEETKN. LFSS 2031 (DENKISESILYDDFKLINVPHDQIKARLVKWSSESRLEPAAIYEKIIVMDINVDHESCSDVF. YTLGS. FLDEQACKL.RSNG 2040 (DENKISESILYDDFKLINVPHDQIKARLVKWSSESRLEPAAIYEKIIVMDINVDHESCSDVF. YTLGS. FLDEQACKL.RSNG 2041 (DENKISESILYDDFKLINVPHDQIKARLVKWSSESRLEPAAIYEKIIVMDINVDHESCSDVF. YTLGS. FLDEQACKL.RSNG 2041 (DENKISESILYDDFKLINVPHDQIKARLVKWSSESRLEPAAIYEKIIVNDGINVQLASCHLFT. EMLDL. SNNSASE. QIIK 2041 (DENKISESILYDPKLEFAKLLQKG. DQVMANNYLEGLSINMSGCQNKVQLASCRILKFT. EMLDL.SNNSASE. QIIK 2042 (DENKISSING) 2044 (DENKISSING) 2044 (DENKISSING) 2044 (DENKISSING) 2044 (DENKISSING) 2044 (DENKISSING) 2047 (DENKISSING) 2049 (DENKISSING) 2049 (DENKISSING) 2049 (DENKISSING) 2040 (DENKISSING) 2040 (DENKISSING) 2040 (DENKISSING) 2041	RECKFSMKMKMID AQLAV AKHAK KAMNV NKLAR SASAI IQYGK LAWII NRFAH IAWVI NRFAH IAWVI NRFAH EIEDREHRSYTGK VQMLRSHMIKAG QYQDIFQIDSKND AVLKYFQEAIAVH AUKKYFQEAIAVH AUKKYFQUSAY EUSPYQOVSPQY EWGLNLQGINEST KWKKSLQDSVNQE EUSPYQOVSPQY EWGLNLQGINEST KWKKSLQDSVNQE EWRIATQPNWRNT LLGTTYRIIANAL YHRANIHSVLDQA RLEGDFHLKLNDT ALKGMFLSKLRAY CNLISGEEHDLWUF NJIIDKFGLWF SENLPKVITFWD QSMFLIS.LWD DIILKCGLWF DETTERLMQ.LWFD QDILRLN.LWF QAYPALWD REKJAKY.LWTL RELLCRI.LWF
293 VSGRTAVGGLGNNYADLED LETWIKHTPNEWD. NUTVUTUNLUMRENYN. VVIDAFKDATSN. SPLHHLCFKD 2936TNUGRNNS.LDWKTVVUKWENKEPTSD. DISES SIFWMCHING	RECKPSMKMKMID AQLAV AKHAK KAMNV NKLAR SASAI IQYGK LAWII NRFAH IAWVI NRFAH IAWVI NRFAH ELEDREHRSYTGK VQMLISHMIKAG QYQDIFQIDSKMD AVLKYFQEAIAVH AUKKYKDYACL EISPYRQVGYSQY EWGLNLQGINEST KWKKSLQDSVNQE EUSPYRQVGYNQ EWGLNLQGINEST KWKKSLQDSVNQE EUSPYRQVGYNQ EWGLNLQGINEST KWKKSLQDSVNQE EUSPYRQVGYNQ EWGLNLQGINEST KWKKSLQDSVNQE EUSPYRQVGYNQ EWGLNLQGINEST KWKKSLQDSVNQE EUSPYRQVGYNQ EWGLNLQGINEST KWKKSLQDSVNQE EUSPYRQVGYNQ EWGLNLQGINEST KWKKSLQDSVNQE EUSPYRQVGYNQ EUSPYRQVGYNQ EUSPYRQVGYNQ EUSPYRQVGYNQ ENTELLSU QDIFLELM QNTALLTVL WPD DITLELLALWP RELLCRI.LWYL MSLLYQ 2326 SAYLLRQ 1944
2899 VSGNTAVGGLGNAYADLKDILETURENTENLEND. NUTVUVDEDUMENENNVVIDAFKDFATBASPLEHLEFPAL. 2938THLGRNS.LEMDKTVVKT RANDE VSD. DEGE SST PHWCHHVY	RECKFSMKMKMID AQLAV AKHAK KAMNV NKLAR SASAI IQYGK LAWII NRFAH IAWVI NRFAH IAWVI NRFAH EIEDREHRSYTGK VQMLRISHMIKAG QYQDIFQIDSKMD AVLKYFQEAIAVH AUKKYKDYACL EISFVRQVSPQY EWGLNLQGINEST KWKKSLQDSVNQE EUSFVRQVSPQY EWGLNLQGINEST KWKKSLQDSVNQT EUSFVRQVSPQY EWGLNLQGINEST KWKKSLQDSVNQT EUSFVRQVSPQY EWGLNLQGINEST KWKKSLQDSVNQT EUSFVRQVSPQY EWGLNLQGINEST KWKKSLQDSVNQT EUSFVRQVSPQY EWGLNLQGINEST KWKKSLQDSVNQT EUSFVRQVSPQY EWGLNLQGINEST KWKKSLQDSVNQT EUSFVRQVSPQY EWGLNLQGINEST KWKKSLQDSVNQT EUSFVRQVSPQY EWGLNLQGINEST KWKSLQDSVNZ EUSFVRQVSPQY EWGLNLQVSPQY UNANINSVLQV REKGFQNLGK SCHUCKSLQVSPQY CLLSGEEHDLWUF NDIIDKFCGLWF QDTRLLS.LWF QDTLRLLM.LWF QDTLRLLMVF QDTLRLLT.LWF QDTLRLLT.LWF QDTLRLLT.LWF RELLCRI.LWF MSLLYY 2326 SAYLLRQ 2155 TALVNS 1944 UPHFKS 1903
289 VSONTAVGGUNNTADLEDLETHELTENENDINGEN NMILVUDGLUMMENTNVIDAFKDAATSM	RECKFSMKMKMID AQLAV AKHAK KAMNY NKLAR SASAI IQYGK LAWII NRFAH IAWVI NRFAH IAWVI NRFAH EIEDREHRSYTGK EIEDREHRSYTGK EVAGSYDGNSREL EIEDREHRSYTGK VOMLRISHMIKAG QYQDIFQIDSKMD AVLKYFQEAIAVH AUKKYKDYGEAIAVH AUKKYKDYGEAIAVH AUKKYKDYGUSY EWGLNLQGINEST KWKSLQOVNCE EWRIATQPNWRNT LGGTTWRIJANAL YHRANIHSVDQA RLEGGFHLSKLRAY CLLSGEEHDLWYF NDIIDKFCGLWFE LWFS ENLPKVITFNLD QSMPRLS.LWF QDTLRLM.LWF QDTLRLT.LWF QMYALWF QMYALWF QMYALWF QMYALWF SCHLARY.ST CLLSGE ENDLWYF NDIIDKFCGLWFE CLLSGEENDLWYF QDTLRLT.LWF QDTLRLT.LWF QMYALWF QMYALWF SCHLARY.ST S
289 VSGNTAVGGGUNTADLEDLETTELETTELETTELETTELETTELETTELETT	RECKFSMKMKMID AQLAV AKHAK KAMNY NKLAR SASAI IQYGK LAWII NRFAH IAWVI NRFAH IAWVI NRFAH EIEDREHRSYTGK VOMLRISHMIKAG QYQDIFQIDSKMD AVLKYFQEAIAVH AUKKYKDYTACL EISSFXQCVSPQY EWQLINQGINEST KWKSLQOVSPQY EWQLINQGINEST KWKSLQOSVNCE EUSFYRQCVSPQY EWQLINQGINEST KWKSLQOSVNCE EWRIATQPNWRWT LGGTTKHIANAL YHRANIHSVLQA ALKGMFLAQINKS TLKGMFQNRLGEK TLKGMFQNRLGEK TLKGMFQNRLGEK TLKGMFLSKLAY CLLSGEEHDLWY NDIIDKFCGLWFE NDIIDKFCGLWFE NDIIDKFCGLWFE NDIIDKFCGLWFE NDIIDKFCGLWFE RENLFXVITFNLD QSMFLIS.LWLD QTLRLLN.LWFE RSHIARVLVLL RKFIARY
2839 VSUNIAVGELGENTADLEN LELEN HUR PEND. NET VEUDEL WARNENS	RECKPSMKMKMID AQLAV AKHAK KAMNV NKLAR SASAI IQYGK LAWII NRFAH IAWVI NRFAH IAWVI NRFAH EIEDREHRSYTGK VOMLRISHMIKAG QYQDIFQIDSKMD AVLKYFQEAIAVH AUKKYKDYGEAIAVH AUKKYKDYGEAIAVH AUKKYKDYGUSYY EWGLNLQGINEST KWKSLQOVSPQY EWGLNLQGINEST KWKSLQOVSPQY EWGLNLQGINEST KWKSLQDSVNCE EWFLATQPNWRWT LLGGTYKIIANAL YHRANIHSVLDQA RLEGGFHLKLNDT ALKGMFQNRLGEK TLEGMFGNRLGEK TLEGMFGNRLGEK TLEGMFLSKLRAY CLLSGEEHDLWYF NDIIDKFGGLWFE NDIIDKFGGLWFE NDIIDKFGGLWFE NDIIDKFGLWFE NDIIDKFGLWFE SENLPKVITFWLD QSNPLIS.LWD QDILRLIN.LWFK SCHIAKI.LWIS SCHIAKI.LWIS SCHIAKI.LWIS SCHIAKI.LWIS SCHIAKI.LWIS SCHIAKI.LWIS CLLSGEENDLWJF CUTLRLT.LUFN QTILRLT.LUFN CAYAS.LWIS SCHIAKI.LWIS SCHIAKI.LWIS SCHIAKI.LWIS SCHIAKI.LWIS SCHIAKI.LWIS CAYASS 2165 DAATIKS 1943 LTVASKS 2153 TVASKS 2153 TVASKS 2153 TVASKS 1943 LTVASKS 1943 LTVASKS 1943
2839 VSGNIAVGGLGMATADLEN LEN KLUP VERD LNEV VUDLUM ALBEYNVIDAKDFATEN	RECKPSMKMKMID AQLAV AKHAK KAMNV NKLAR SASAI IQYGK LAWII NRFAH IAWVI NRFAH IAWVI NRFAH EIEDREHRSYTGK VOMLRISHMIKAG QYQDIFQIDSKMD AVLKYFQEAIAVH AUKKYKDOXSCU EISFVRQVSPQY EWGLNLQGINEST KWKSLQOVSPQY EWGLNLQGINEST KWKSLQDSVNQE EWGLATQPNWRWT LLGTTYRIIANAL YHRANIHSVLDQA RLEGDFHLKLNDT ALKGMFQNRLGEK TLEGMFGNRLGEK TLEGMFGNRLGEK TLEGMFGNRLGEK TLEGMFLSKLRAY CLLSGEEHDLWYF NDIIDKFGGLWFE LWB SHIARV.LWD QUTLRLIN.LWFK QDILRLIN.LWFK QDILRLIN.LWFK RKHIAKI.LWU SRHIARV.LWU LKYLAW.S SHIARV.LWU RKYLAV.LWU LKYLAV.S SHIARV.LWU LALANA 2576 YMSILLY 2326 SAVLRQ 2155 TALVNS 1943 LTVASS 2153 LTVASS 1943 LTVASS 1943 L
2239 USGNIAVGBLORKTADLEDILETRELIEVENDI. KNIVUESSERVENTO. AIVIALESS. GLOPATSK. SLUHELEYED 2330TUGENKS LUNKUVYE GETVED DETVED DETVED SOFTME GENET. AIVIALESS. GLOPATSK. SLUHELEYED 2340TUGENKS LUNKUVESSERVENTO. AVIA 2350TUGENKS LUNKUVESSERVENTO. AVIALESSERVENTO. 2365TUGENKS LUNKUVESSERVENTED STEVENTO. 2365TUGENKS LUNKUVESSERVENTED STEVENTO. 2366TUGENKS LUNKUVESSERVENTED STEVENTO. 2366TUGENKS LUNKUVESSERVENTED STEVENTO. 2367SENTISSEILUDGEN LUNKUVESSERVENTED SERVENTO. 2368SENTISSEILUDGEN LUNKUVESSERVENTED SERVENTO. 2368SENTISSEILUDGEN LUNKUVESSERVELINVON UNDUKUVESSERVENTED SEGUNALETCLENPAVINGTVIRAVU 2369SENTISSEILUDGEN LUNKUVESSERVELINVON UNDUKUVESSERVENTESSEGUVESSERVENTED SEGUNALETCISSE 2360SENTISSEILUDGEN LUNKUVESSERVELINVON UNDUKUSSESSERVELINGEN LUNKUVESSERVELINGEN 2361SENTISSEILUDGEN LUNKUVESSERVELINSGEGENVELINGEN LUNKUSSERVELINGEN LUNKUVESSERVELINGEN 2364SENTISSEILUDGEN LUNKUVESSERVELINGEN SEGUNALISEN SEGUNALISEN LUNKUVESSERVELINGEN 2364SENTISSEILUDGEN LUNKUVESSERVELINGEN SEGUNALISEN	RECKPSMKMKMID AQLAV AKHAK KAMNV NKLAR SASAI IQYGK LAWII NRFAH IAWVI NRFAH IAWVI NRFAH EIEDREHRSYTGK VOMLRISHMIKAG QYQDIFQIDSKMD AVLKYFQEAIAVH AUKKYKDOXSAU EISFVRQVSPQY EWGLNLQGINEST KWKSLQOVSPQY EWGLNLQGINEST KWKSLQDSVNQE EWGLATQPNWRWT LLGGTHRINAL VHRANISVLDQA RLEGFFLQVIFQWWF TLEGMFLSKLAZY CLLSGEEHDLWYF NDIIDKFCGLWFE NDIIDKFCGLWFE NDIIDKFCGLWFE NDIIDKFCGLWFE NDIIDKFCGLWFE NDIIDKFCGLWFE NDIIDKFCGLWFE NDIIDKFCGLWFE NDIIDKFCGLWFE NDIIDKFCGLWFE NDIIDKFCGLWFE SHARV. LWU CLLSGEEHDLWYF NDIILKLS LWU QUTLRLIN. LWFK SHIAKI.LWI SSHIARV.LYU RKFIAKI.LWI SSHIARV.LYU LATANS 1943 UTVASKS 2155 TALVNS 1943 UTVASKS 1944 UTVASKS 1943 UTVASKS 1943 UTVASKS 1943 UTVASKS 1944 UTVASKS 1943 UTVASKS 1943 UTVASKS 1943 UTVASKS 1943 UTVASKS 1944 UTVASKS 1943 UTVASKS 1943 UTVASKS 1944 UTVASKS 1943 UTVASKS 1943 UTVASKS 1943 UTVASKS 1944 UTVASKS 1943 UTVASKS 1943 UTVASKS 1944 UTVASKS 1943 UTVASKS 1943 UTVASKS 1943 UTVASKS 1944 UTVASKS 1943 UTVASKS 1943 UTVASKS 1944 UTVASKS 1943 UTVASKS 1943 UTVASKS 1944 UTVASKS 1943 UTVASKS 1943 UT

proteins¹⁸. Their PI-kinase domain lacks the catalytic residues and indeed, none of them has been shown to possess kinase activity. The TRRAP proteins contain regions of similarity to PIK-related proteins in the neighboring regions of this aberrant PI-kinase domain.

Here we describe a novel homology domain spanning ~500 amino acids N-terminal to the PI-kinase domain in the PIK-related and TRRAP subfamilies (Fig. 1b). The multiple alignment was constructed by iterative hidden Markov model searching using HMMER2 (see http://hmmer.wustl.edu) and manual alignment refinement. Because the middle portion in the multiple alignment is poorly conserved (see Fig. 1b), we were tempted to propose two domains. However, because all members contain conserved motifs over the entire region, we defined a single domain, called FAT, after representatives of the three main groups sharing the domain (FRAP, ATM, and TRRAP; see Fig. 1a). This domain is not found outside these subfamilies. Because the previously mentioned extreme C-terminal domain is also only found in these subfamilies, we call it FATC. The FAT and FATC domains only occur in combination, suggesting that they interact with each other. It is possible that they fold together in a configuration that ensures proper function of the PIkinase domain, which is wedged in between the FAT and FATC domains. The FATC domain is probably too small (~35 amino acids) to fold independently, but because it is more conserved than the FAT domain (34% versus 16% average identity), it could be more important for catalytic activity than the FAT domain.

In the FRAP and TRRAP groups it is quite clear from the tree in Fig. 1a which *Caenorhabditis elegans* proteins (Q18667 and O01438) are the likely orthologs to human counterparts. For the ATM group, however, it is unclear whether the *C. elegans* protein Q22258 is orthologous to ATM or ATR; perhaps it is orthologous to both.

The functions of the FAT and FATC domains still need to be elucidated experimentally. Data from deletion experiments regarding the functional importance of the N-terminal region in ATM (Refs 19,20) and RAD3 (Ref. 15) have proved contradictory. The region upstream of the PI-kinase domain, including the FAT domain, contains numerous regions of low sequence complexity. These are not of a standard type, such as coiled-coil, but many are enriched in leucine and glutamate. The FAT domain has diverged much faster than the catalytic PI-kinase domain (16% versus 28% average identity). Such sequence properties are typically found in proteins with an extended, non-globular structure, or proteins that form multimeric protein complexes²⁴. We therefore speculate that the FAT domain could be of importance as a structural scaffold or as a protein-binding domain, or both.

References

- Majerus, P.W. et al. (1990) Recent insights in phosphatidylinositol signaling. Cell 63, 459–465
- 2 Carr, A.M. (1997) Control of cell cycle arrest by the Mec1sc/Rad3sp DNA structure checkpoint pathway. *Curr. Opin. Genet. Dev.* 7, 93–98
- 3 Hoekstra, M.F. (1997) Responses to DNA damage and regulation of cell cycle checkpoints by the ATM protein kinase family. *Curr. Opin. Genet. Dev.* 7, 170–175
- 4 Hartley, K.O. *et al.* (1995) DNA-dependent protein kinase catalytic subunit: a relative of phosphatidylinositol 3-kinase and the ataxia telangiectasia gene product. *Cell* 82, 849–856
 5 Bentley, N.J. *et al.* (1996) The *Schizosaccharomyces*
- pombe rad3 checkpoint gene. *EMBO J.* 15, 6641–6651 6 Brunn, G.J. *et al.* (1996) Direct inhibition of the signaling
- billini, G.J. et al. (1990) Inter immotion of the signaling functions of the mammalian target of rapamycin by the phosphoinositide 3kinase inhibitors, wortmannin and LY294002. EMBO J. 15, 5256–5267
- 7 Rotman, G. and Shiloh, Y. (1998) ATM: from gene to function. *Hum. Mol. Genet.* 7, 1555–1563
- 8 Zakian, V.A. (1995) ATM-related genes: what do they tell us about functions of the human gene? *Cell* 82, 685–687
- **9** Ziv, Y. *et al.* (1997) Recombinant ATM protein complements the cellular A–T phenotype. *Oncogene* 15, 159–167
- 10 Keith, C.T. and Schreiber, S.L. (1995) PIK-related kinases: DNA repair, recombination, and cell cycle checkpoints. *Science* 270, 50–51

PROTEIN SEQUENCE MOTIFS

- 11 Kato, R. and Ogawa, H. (1994) An essential gene, ESR1, is required for mitotic cell growth, DNA repair and meiotic recombination in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 22, 3104–3112
- 12 Meyn, M.S. (1995) Ataxia-telangiectasia and cellular responses to DNA damage. *Cancer Res.* 55, 5991–6001.
- 13 Savitsky, K. *et al.* (1995) A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* 268, 1749–1753
- 14 Carpenter, C.L. and Cantley, L.C. (1996) Phosphoinositide kinases. *Curr. Opin. Cell Biol.* 8, 153–158
- 15 Chapman, C.R. *et al.* (1999) Requirement of sequences outside the conserved kinase domain of fission yeast Rad3p for checkpoint control. *Mol. Biol. Cell* 10, 3223–3238
- 16 Bornberg-Bauer, E. et al. (1998) Computational approaches to identify leucine zippers. Nucleic Acids Res. 26, 2740–2746
- 17 Montecucco, A. et al. (1998) DNA ligase I is recruited to sites of DNA replication by an interaction with proliferating cell nuclear antigen: identification of a common targeting mechanism for the assembly of replication factories. *EMBO J.* 17, 3786–3795
- 18 McMahon, S.B. et al. The novel ATM-related protein TRRAP is an essential cofactor for the c-Myc and E2F oncoproteins. Cell (1998) 94, 363–374
- Baskaran, R. et al. (1997) Ataxia telangiectasia mutant protein activates cAbl tyrosine kinase in response to ionizing radiation. Nature 387, 516–519
- 20 Morgan, S.E. and Kastan, M.B. (1997) p53 and ATM: cell cycle, cell death, and cancer. *Adv. Cancer Res.* 71, 1–25
- 21 Bateman, A. et al. (1999) Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. Nucleic Acids Res. 27, 260–262
- 22 Galtier, N. et al. (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. 12, 543–548
- 23 Thompson, J.D. et al. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680
- 24 Wright, P.E. and Dyson, H.J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm. J. Mol. Biol. 293, 321–331

ROBERTA BOSOTTI AND ANTONELLA ISACCHI

Dept of Biology, Pharmacia & Upjohn, Viale Pasteur 10, 20014, Nerviano (MI), Italy.

ERIK L.L. SONNHAMMER

Center for Genomics Research, Karolinska Institutet, 171 77 Stockholm, Sweden. Email: Erik.Sonnhammer@cgr.ki.se

Nuclear receptors arose from pre-existing protein modules during evolution

Nuclear receptors (NRs) are ligandmodulated transcription factors comprising multiple domains that include a highly conserved DNA-binding domain (DBD) and a ligand-binding domain (LBD). Here, we support the hypothesis that nuclear receptors arose in evolution from a fusion event linking pre-existing and independent protein modules^{1,2} (see also: http://www.biomednet.com/hmsbeagle/03 /cutedge/day1.htm).

One view held for the origin of multidomain proteins is that they arose by DNA shuffling and rearrangement, bringing together pre-existing protein modules in new chimeric combinations to

create proteins with potentially new functions³. The origin of NRs might comply with this view. For example, the zinc-finger DNA-binding elements seen in the NR show a high degree of structural similarity to LIM (Lin-11, Isl-1 and Mec-3) and GATA (GATA-DNA-binding transcription factors) domains, which are present in both unicellular and multicellular organisms⁴. Therefore, it is believed that such zinc-finger DBDs might well have evolved from LIM and GATA proteins. However, a candidate LBD precursor has remained elusive. Our analyses provide evidence that Pex11p, a protein found in all major eukaryotic

Kingdoms, could be an ancient relative of the LBD.

Pex11p is a peroxisomal membrane protein of 235 amino acid residues implicated in peroxisome proliferation in the yeast Saccharomyces cerevisiae5,6 and in human cells^{7,8}. Although its precise function is unknown (for a critical discussion, see Ref. 9), Pex11p-deficient S. cerevisiae also displays a marked inability to degrade fatty acids (E.H. Hettema and C.W.T. van Roermund, unpublished). To obtain additional clues as to the possible function of Pex11p, we searched for similarities to other proteins. By combining both BLAST2 searches and CLUSTALX analyses, we observed a highly significant amino acid sequence similarity (30% identity, 50% similarity) between amino acids 2–187 of Pex11p and the LBD of NRs, in particular the peroxisome