FAT: a novel domain in PIK-related kinases

Phosphatidylinositol kinases are found in all eukaryotes and serve important functions in phosphatidylinositol (PI) signaling pathways[1]. In addition to the PI-kinase 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 ~300-amino-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 cancer[9].

It has previously been noted that members of the PIK-related family share a unique motif at the extreme C terminus[10], 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 PI-kinase 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-factor-binding site (P-site) have been made in the N-terminal region[15]. 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 non-leucine residues[16]. The P-site motif is uncertain too, as five out of nine amino acids are never observed at the same position in other known cases[17].

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

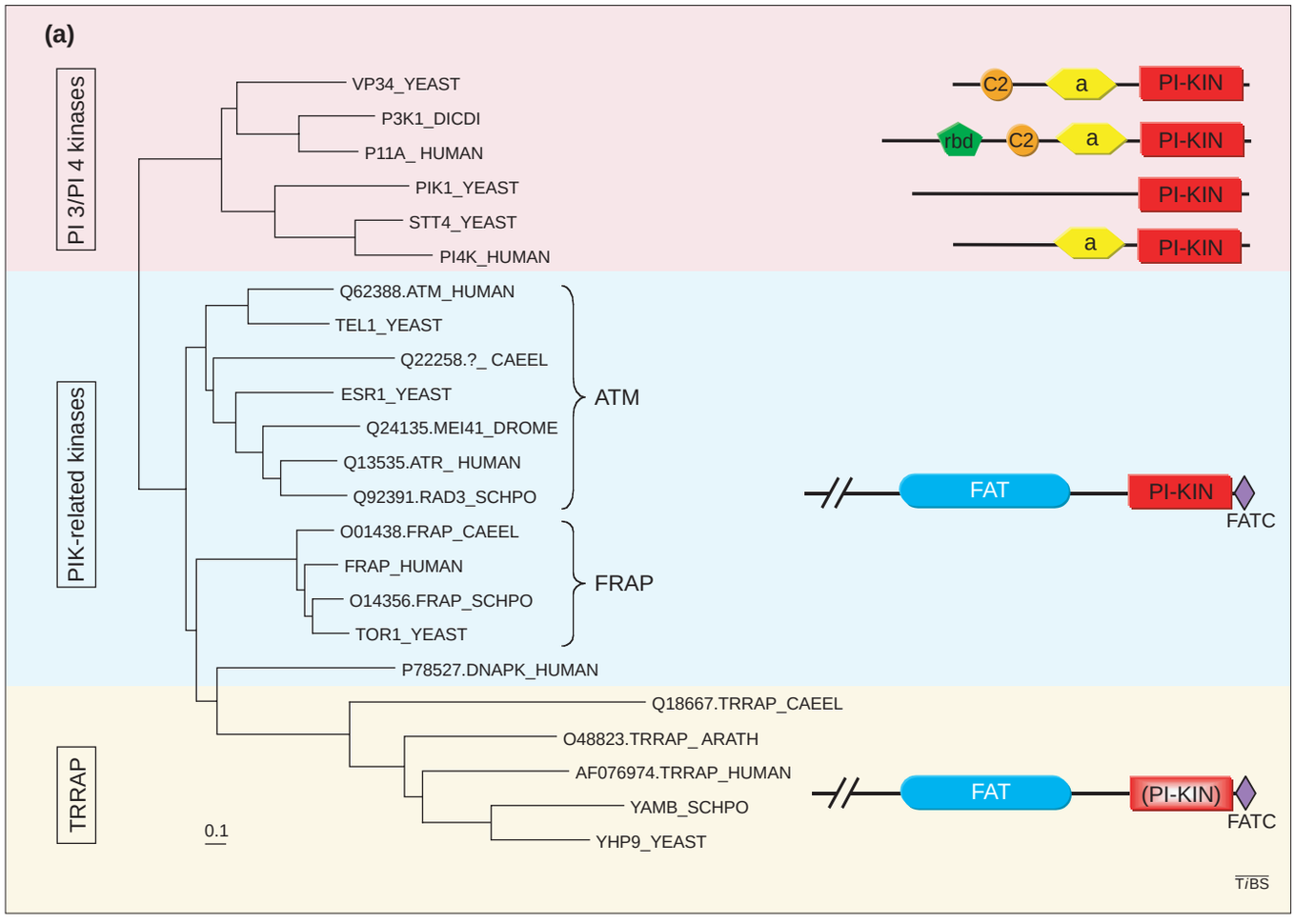
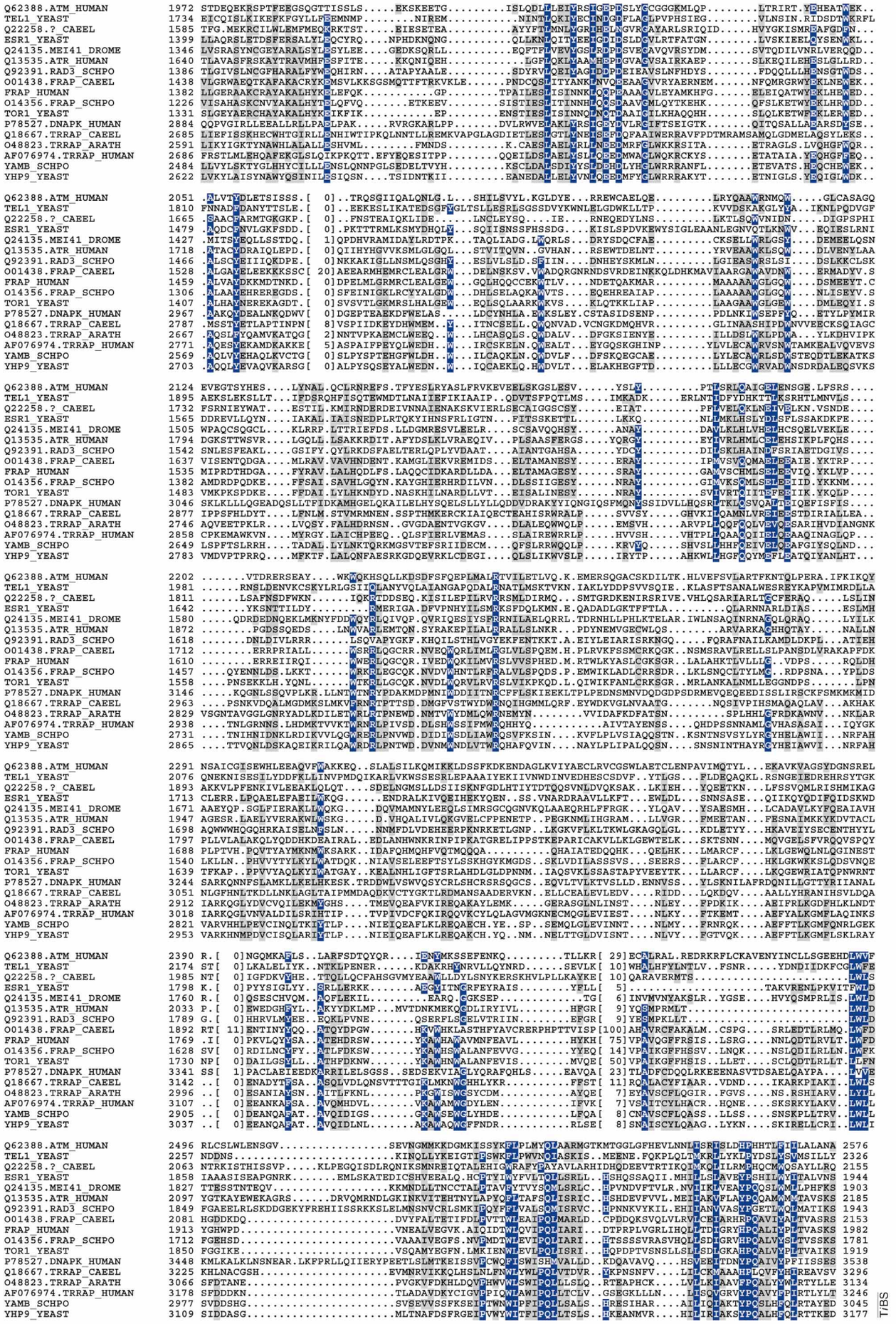


Figure 1

(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 generated by CLUSTALW (Ref. 23). 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.

0968–0004/00/$ – See front matter PII: S0968-0004(00)01563-2

(b)


Q62388.ATM_HUMAN 1972 STDEQEKRSPTFEEGSQGTTISSLS.....EKSKEETG.....ISLQDLVLEIVRSIGEPLSLYCGGGKMLQP.....LTRIRT.YEHEATWEK..
TEL1_YEAST 1734 EICQISLKIEKFKFGYLLPEEMNMP.....NIREM.....NINTLQKIVECINDGLFLACLPVPHSIEG.....VLNSIN.RIDSDTWKRFL
Q22258.?_CAEEL 1585 TFG.MEKRCILWLEMFMEQKRKTST.....EIESETEA.....AYYFTLMNLVGRIHELNGVRCAYARLSRIQID.....HVYGKISMREAFGDEN...
ESR1_YEAST 1399 LLAQRSLETDSFERSALYLEQCYRQ....NPHDKNQNG.....QLLKNLQITVEEICDISLDGVLRTFATGN.....LVSKIEELQYSENWKL..
Q24135.MEI41_DROME 1346 LVSRASYNCGEYARALSYLESYLEE....GEDKSQRLL.....EQFTFLVEVVGSLRDPLSVECAVQVRSYDM.....SVTQDILVNRLVERQQD..
Q13535.ATR_HUMAN 1640 TLAVASFRSKAYTRAVMHPESFITE....KKQNIQEH.....LGPLQKLYAAMHEPIGVACVSAIRKAEP.....SLKEQILEHESLGLLRD..
Q92391.RAD3_SCHPO 1386 TLGIVSLNCGFHARALFYWEQHIRN....ATAPYAALE.....SDYRVLQEIYAGIDDPLEIEAVSLNFHDYS.....FDQQLLLHENSGTWDS..
O01438.FRAP_CAEEL 1438 VLGRWAEQTKAFAKACRYKEMSVLKKSGSMQTTFTRKVKLE.....PNDCQSLITYANKLNVQEEAAGVVRYAERNEM.....NFQMRGRWYEKLNEWEK..
FRAP_HUMAN 1382 LLGERAAKCRAYAKALHYKELEFQK.....GP.....TPAILESLISINNKLQQPEAAAGVLEYAMKHFG.....ELEIQATWYEKLHEWED..
O14356.FRAP_SCHPO 1226 VISAHASKCNVYAKALHYTELQFVQ.....ETKEEV.....SISTIESLITINNHLQQSPAAVGMLQYTKEHK.....QFSLKETWYEKLHRWDD..
TOR1_YEAST 1331 SLGEYAERCHAYAKALHYKEIKFIK.....EP.....ENSTIESLITINNQLNQTEAAIGILKHAQQHH.....SLQLKETWFEKLERWED..
P78527.DNAPK_HUMAN 2884 QQPVGIRLLEEALLRLLPAELPAK.....RVRGKARLPP.....DVLRWVELAKLYRSIGEYEVLRGIFTSEIGTK.....QITQSALLAERASDYSE..
Q18667.TRRAP_CAEEL 2685 LIEFISSKHECWHTGIRLLENHIWTIPKQLNNTLLREMKVAPGLAGDIETLESLGTLVNEISEFLQFAAIWERRAVFPDTMRAMSAMQLGDMELAQSYLEKS..
O48823.TRRAP_ARATH 2591 LIKYIGKTYNAWHLALALLESHVML.....FMNDS.....K.CAESLAELYRLLNEELMRFGLWKKRSITA.....ETRAGLS.LVQHGFWQR..
AF076974.TRRAP_HUMAN 2686 FRSTLMLEHQAFEKGLSLQIKPKQTT.EFYEQESITPP.....QQEILDSLAELYSLIQEELMWAGLWQKRCKYS.....ETATAIA.YEQHGFEQ..
YAMB_SCHPO 2484 LLVYLSKTYGLHHYCILLLENSLQNNPGLSEDELTVYH.....KSCLDAISDIVYSLIDEHILYHGLWRRRANFL.....ETEVATS.HEQCHEWEK..
YHP9_YEAST 2622 LVKYLAISYNAWYQSINILESIQSN....TSIDNTKII.....EANEDALLELVNLQEELMFYGLWRRRAKYT.....ETNIGLS.YEQIGLWDK..

Q62388.ATM_HUMAN 2051 ALVTYDLETSISSS.[0]..TRQSGIIQALQNLG.L...SHILSVYL.KGLDYE...RREWCAELQE.....LRYQAAWRNMQW.....GLCASAGQ
TEL1_YEAST 1810 FNNADFDANYTTSLE.[0]..EEKESLIKATEDSGFYGLTSLLESRLSGSSDVYKWNLELGDWKLLTP.....KVVDSKAKGLYLA..IKNLPQDVGF
Q22258.?_CAEEL 1665 SAACFARMTGKGKP.[0]..FNSTEAIQKLIDE.....LNCLEYSQ....IE....RNEQEDYLNS.....LKTLSQWVNIDN.....DIGPSPHI
ESR1_YEAST 1479 AQDCFNVLGKFSDD.[0]..PKTTTRMLKSMYDHQLYI...SQIISNSSFHSSDG....KISLSPDVKEWYSIGLEAANLEGNVQTLKNWV...EQIESLRNI
Q24135.MEI41_DROME 1427 .MITSYEQLLSSTDQ.[1]QPDHVRAMIDAYLRDTPK...TAQLIADG.LWQRLS...DRYSDQCFAE.....CKSELLWRLGSY....DEMEELQSN
Q13535.ATR_HUMAN 1718 ATACYDRAIQLEPD.[0]QIIHYHGVVKSMLGLGQL...STVITQVN..GVHAN...RSEWTDELNT.....YRVEAAWKLSQN....DLVENYLAA
Q92391.RAD3_SCHPO 1466 ALSCYEIIIQKDPE.[0]NKKAKIGLLNSMLQSGHY...ESLVLSLD.SEIIN....DNHEYSKMLN.....LGIEASWRSLSI....DSLKKCLSK
O01438.FRAP_CAEEL 1528 ALGAYELEEKKKSSC[20]AEEARMHEMRCLEALGRW...DELNSKSV.VWADQRGNRNDSVRDEINKKQLDHKMAVIAARGAWAVDNW....ERMADYV.S
FRAP_HUMAN 1459 ALVAYDKKMDTNKD.[0]DPELMLGRMRCLEALGEW...GQLHQQCCEKWTLV....NDETQAKMAR.....MAAAAAWGLGQW....DSMEEYT.C
O14356.FRAP_SCHPO 1306 ALAAYEHRERGDS.[0]SFEININGKLRCYYALGDW...DHLSELAQKAWVTS....EQEHREAIAP.....LAAAAAWGLGQW....NLISEYV.S
TOR1_YEAST 1407 ALHAYNEREKAGDT.[0]SVSVTLGKMRSLHALGEW...EQLSQLAARKWKVS....KLQTKKLIAP.....LAAGAAWGLGEW....DMLEQYI.S
P78527.DNAPK_HUMAN 2967 AAKQYDEALNKQDWV[0]DGEPTEAEKDFWELAS.....LDCYNHLA.EWKSLEY.CSTASIDSENP.....PDLNKIWSEPFYQ...ETYLPYMIR
Q18667.TRRAP_CAEEL 2787 .MSSTYETLAPTINPN[8]VSPIIDKEYDHWMEM..YI...ITNCSELL.QWNVADVCNGKDMQHVR.....GLINAASHIPDWNVVEECKSQIAGC
O48823.TRRAP_ARATH 2667 AQSLFYQAMVKATQG[2]NNTVPKAEMCLWEEQ..WI...LHCASQLS.QWDALV...DFGKSIENYE.....ILLDSLWKLPDWA...YLKDHVIPK
AF076974.TRRAP_HUMAN 2771 AQESYEKAMDKAKKE[5]ASPAIFPEYQLWEDH..WI...IRCSKELN.QWEALT...EYGQSKGHINP.....YLVLECAWRVSNWTAMKEALVQVEVS
YAMB_SCHPO 2569 AQLVYEHAQLKVCTG[0]SLPYSPTEHGFWLDH..WI...ILCAQKLN.QWDVLF...DFSKQEGCAE.....LYLECAWRLSDWSTEQDTLEKATKS
YHP9_YEAST 2703 AQQLYEVAQVKARSG[0]ALPYSQSEYALWEDN..WI...IQCAEKLQ.HWDVLT...ELAKHEGFTD.....LLLECGWRVADWNSDRDALEQSVKS

Q62388.ATM_HUMAN 2124 EVEGTSYHES....LYNAL.QCLRNREFS.TFYESLRYASLFRVKEVEELSKGSLESV.....YSLY.....PTLFSRLQAIGELENSGE.LFSRS....
TEL1_YEAST 1895 AEKSLEKSLLT...IFDSRQHFISQTEWMDTLNAIIEFIKIAAIP..QDVTSFPQTLMS....IMKADK.....ERLNTIIDFYDHKTTLKSRHTLMNVLS....
Q22258.?_CAEEL 1732 FSRNIEYWAT....ESTIL.KMIRNDERDEIVNNAIENAKSKVIERLSECAIGGSCSY.....EIAT.....PPLVELQKLNEIVELKN.VSNDE....
ESR1_YEAST 1565 DDREVLLQYN....IAKAL.IAISNEDPLRTQKYIHNSFRLIGTN...FITSSKETTL.....LKKQ.....NLLMKLHSLYDLSFLSSAKDKFE....
Q24135.MEI41_DROME 1505 WPAQCSQGCL....KLRRP.LTTRIEFDS.LLDGMRESVLEELR....SCSAVQQHSY.....ANAY.....DAVLKLHLVHELHCSQELVEKELE....
Q13535.ATR_HUMAN 1794 DGKSTTWSVR....LGQLL.LSAKKRDIT.AFYDSLKLVRAEQIV...PLSAASFERGS....YQRGV.....EYLVRLHMLCELEHSIKPLFQHS....
Q92391.RAD3_SCHPO 1542 SNLESFEAKL....GSIFY.QYLRKDSFAELTERLQPLYVDAAT....AIANTGAHSA.....YDCV.....DILSKLHAINFFSRIAETDGIVS....
O01438.FRAP_CAEEL 1637 VISENTQDGA....MLRAV.VAVHNDENT.KAMGLIEKVREMIDS...ELTAMANESY.....ERAV.....IPWVSVQQMASLEEAIE.YKTRP....
FRAP_HUMAN 1535 MIPRDTHDGA....FYRAV.LALHQDLFS.LAQQCIDKARDLLDA...ELTAMAGESY.....SRAV.....GAWVSCHMLSSLEEVIQ.YKLVP....
O14356.FRAP_SCHPO 1382 AMDRDPQDKE....FFSAI.SAVHLGQYN.KAYGHIERHRDILVN...DLSSIIGESY.....NRAV.....GIWVKSQMLSSLEEIID.YKKNM....
TOR1_YEAST 1483 VMKPKSPDKE....FFDAI.LYLHKNDYD.NASKHILNARDLLVT...EISALINESY.....NRAV.....SVIVRTQITTFEEIIK.YKQLP....
P78527.DNAPK_HUMAN 3046 SKLKLLLQGEADQSLLTFIDKAMHGELQKAILELHYSQELSLLYLLQDDVDRAKYYIQNGIQSFMQNVSSIDVLLHQSRITKLQSVQALTEIQEFISFIS....
Q18667.TRRAP_CAEEL 2877 IPPSFHLDYT....LFNLM.STVMRMNEN.SSPTHMKERCKIAIQECTEAHISRWRALP.....SVVSY.....GHVKILQAMNLVRELESSTDIRIALLEA..
O48823.TRRAP_ARATH 2746 AQVEETPKLR....LVQSY.FALHDRNSN..GVGDAENTVGKGV....DLALEQWWQLP.....EMSVH.....ARVPLLQQFQLVEVQESARIHVDIANGNK
AF076974.TRRAP_HUMAN 2858 CPKEMAWKVN....MYRGY.LAICHPEEQ..QLSFIERLVEMAS....SLAIREWRRLP.....HVVSH.....VHTPLLQAAQQIILQEAAQINAGLQP...
YAMB_SCHPO 2649 LSPFTSLRRH....TADAL.LYLNKTQRKMGSVTEFSRIIDECM....QFSLRRWQQLP.....KRVVQ.....SHVSLLHHFQEIVLQEAFGIYSQLND...
YHP9_YEAST 2783 VMDVPTPRRQ....MFKTF.LALQNFAESRKGDQEVRKLCDEGI....QLSLIKWVSLP.....IRYTP.....AHKWLLHGFQQYMFLEATQIYANLHT...

Q62388.ATM_HUMAN 2202 .......VTDRERSEAY....WKWQKHSQLLKDSDFSFQEPLMALRTVILETLVQ.K.EMERSQGACSKDILTK.HLVEFSVLARTFKNTQLPERA.IFKIKQY
TEL1_YEAST 1981 ......RNSLDENVKCSKYLRLGSIIQLANYVQLAIANGAPQDALRNATLMSKTVKN.IAKLYDDPSVVSQIE..KLASFTSANALWESREYKAPVMIMRDLLA
Q22258.?_CAEEL 1811 ......LSAFNSDFWKN......IQKRTDDSEQ.KISILEPILRVVRSMLDIRMQ....SMTGRDKENIRSRIVE.VHLQSARIARLTGCFERAQ.....LSLIN
ESR1_YEAST 1642 .......YKSNTTILDY.........RMERIGA.DFVPNHYILSMRKSFDQLKMN.E.QADADLGKTFFTLA.........QLARNNARLDIAS.....ESLMH
Q24135.MEI41_DROME 1580 .....QDRDEDNQEKLMKNYFDDWQYRLQIVQP.QVRIQESIYSFRNILAELQRRL.TDRNHLLPHLKTELAR.IWLNSAQINRNAGQLQRAQ.....LYILK
Q13535.ATR_HUMAN 1872 ........PGDSSQEDS....LNWVARLEMTQN.SYRAKEPILAIRRALLSLNKR...PDYNEMVGECWLQS.....ARVARKACHHQTAY.....NALLN
Q92391.RAD3_SCHPO 1618 .......DNLDIVLRRR......LSQVAPYGKF.KHQILSTHLVGYEKFENTKKT.A.EIYLEIARISRKNGQ....FQRAFNAILKAMDLDKPL...ATIEH
O01438.FRAP_CAEEL 1712 ........ERRPRIALL......WSRRLQGCRR.NVEQWQRLIMLRGLVLSPQEM.H.PLRVKFSSMCRKQGK..NSMSRAVLRELLSLPANSDLVRAKAPFDK
FRAP_HUMAN 1610 ........ERREIRQI......WWERLQGCQR.IVEDWQKILMVRSLVVSPHED.M.RTWLKYASLCGKSGR..LALAHKTLVLLLCE..VDPS.....RQLDH
O14356.FRAP_SCHPO 1457 ....QYENNLDS.LKKT......WRKRLEGCQK.NVDVWHNTLRFRALVLSPQDS.P.EMWIKLADLCRRSDR..LKLSNQCLTYLMCE..RDPSNA....YPLDS
TOR1_YEAST 1558 ....PNSEKKLH.YQNL......WTKRLLGCQK.NVDLWQRVLRVRSLVIKPKQD.L.QIWIKFANLCRKSGR..MRLANKALNMLLEGGNDPS.....LPN
P78527.DNAPK_HUMAN 3146 .....KQGNLSSQVPLKR.LLNTWTNRYPDAKMDPMNIWDDIITNRCFFLSKIEEKLTPLPEDNSMNVDQDGDPSDRMEVQEQEEDISSLIRSCKFSMKMKMID
Q18667.TRRAP_CAEEL 2963 ....PSNKVDQALMGDMKSLMKVERNRTPTTSD.DMGFVSTWYDWRNQIHGMMLQRF.EYWDKVGLNVAATG.....NQSIVPIHSMAQAQLAV...AKHAK
O48823.TRRAP_ARATH 2829 VSGNTAVGGLGNRYADLKDILETWRLRTPNEWD.NMTVWYDMLQWRNEMYN.....VVIDAFKDFATSN.....SPLHHLCFRDKAWNV...NKLAR
AF076974.TRRAP_HUMAN 2938 ....TNLGRNNS.LHDMKTVVKTWRNRLPIVSD.DLSHWSSIFMWRQHHYQ.....AIVTAYENSS....QHDPSSNNAMLCVHASASAI...IQYGK
YAMB_SCHPO 2731 ....TNIHNIDNKLRDIKVVLQGWRERLPNVWD.DIDIWSDLIAWRQSVFKSIN....KVFLPLVSIAQQSTN..KSNTNSVSYLYRCYHELAWII...NRFAH
YHP9_YEAST 2865 ....TTVQNLDSKAQEIKRILQAWRDRLPNTWD.DVNMWNDLVTWRQHAFQVIN....NAYLPLIPALQQSN....SNSNINTHAYRCYHEIAWVI...NRFAH

Q62388.ATM_HUMAN 2291 NSAICGISEWHLEEAQVFWAKKEQ..SLALSILKQMIKKLDSSFKDKENDAGLKVIYAECLRVCGSWLAETCLENPAVIMQTYL...EKAVKVAGSYDGNSREL
TEL1_YEAST 2076 QNEKNISESILYDDFKLLINVPMDQIKARLVKWSSESRLEPAAAIYEKIIVNWDINVEDHESCSDVF..YTLGS...FLDEQAQKL.RSNGEIEDREHRSYTGK
Q22258.?_CAEEL 1893 AKKVLPFENKIVLEEAKLQLQT....SDELNGMSLLDSIISKNFGDLHTIYTDTQQSVNLDVQKSAK..LKIEH....YQEETKN...LFSSVQMLRISHMIKAG
ESR1_YEAST 1713 CLERR.LPQAELEFAEILVKQG......ENDRALKIVQEIHEKYQEN..SS.VNARDRAAVLLKFT...EWLDL...SNNSASE...QIIKQYQDIFQIDSKWD
Q24135.MEI41_DROME 1671 AAEYQP.SGLFIERAKLLVQKG.....DQVMAMNYLEEQLSIMRSGCQGNVKQLAAEQRHLFFRGK...YLQAV...YSAESMH...LCADAVLKYFQEAIAVH
Q13535.ATR_HUMAN 1947 AGESR.LAELVVERAKWLVSKG.....D.VHQALIVLQKGVELCFPENETPP..PEGKNMLIHGRAM....LLVGR...FMEETAN...FESNAIMKKYKDVTACL
Q92391.RAD3_SCHPO 1698 AQWWWHQSQHRKAISELNESLN.....NNMFDLVDEHEERPKNRKETLGNP..LKGKVFLKLTKWLGKAGQLGL...KDLETYY...HKAVEIYSECENTHYYL
O01438.FRAP_CAEEL 1797 PLLVLALAKQLYQDDHKDEAIRAL..EDLANHWNKRINPIPKATGRELIPPSTKEPARICAKVLLKLGEWTELK...SKTSNN....MQVGELSFVRQQVSPQY
FRAP_HUMAN 1688 PLPTVH.PQVTYAYMKNMVKSARK..IDAFQHMQHFVQTMQQQA.....QHAIATEDQQHK...QELHK...LMARCF....LKLGEWQLNLQGINEST
O14356.FRAP_SCHPO 1540 LKLLN..PHVVYTYLKYLWATDQK..NIAVSELEEFTSYLSSKHGYKMGDS..SKLVDILASSSVS...SEERS...FLARCF....HKLGKWKKSLQDSVNQE
TOR1_YEAST 1639 TFKAP..PPVVYAQLKYIWATGAY..KEALNHLIGFTSRLAHDLGLDPNNM..IAQSVKLSSASTAPYVEEYTK...LLARCF....LKQGEWRIATQPNWRNT
P78527.DNAPK_HUMAN 3244 SARKQNNFSLAMKLLKELHKESK.TRDDWLVSWVQSYCRLSHCRSRSQGCS..EQVLTVLKTVSLLD.ENNVSS...YLSKNILAFRDQNILLGTTYRIIANAL
Q18667.TRRAP_CAEEL 3051 NLGFHNLTKDLLNKLAGLTAIPMMDAQDKVCTYGKTLRDMANSAADERVKN..ELLCEALEVLEDV....RIDD...LQKDQV....AALLYHRANIHSVLDQA
O48823.TRRAP_ARATH 2912 IARKQGLYDVCVQILEKMYGHS....TMEVQEAFVKIREQAKAYLEMK.....GERASGLNLINST....NLEY...FPDKIK....AEIFRLKGDFHLKLNDT
AF076974.TRRAP_HUMAN 3018 IARKQGLVNVALDILSRIHTIP....TVPIVDCFQKIRQQVKCYLQLAGVMGKNECMQGLEVIEST....NLKY...FTKEMT....AEFYALKGMFLAQINKS
YAMB_SCHPO 2821 VARVHHLPEVCINQLTKIYTLP....NIEIQEAFLKLREQAECH.YE..SP..SEMQLGLEVINNT....NLMY...FRNRQK....AEFFTLKGMFQNRLGEK
YHP9_YEAST 2953 VARKHNMPDVCISQLARIYTLP....NIEIQEAFLKLREQAKCH.YQ..NM..NELTTGLDVISNT....NLVY...FGTVQK....AEFFTLKGMFLSKLRAY

Q62388.ATM_HUMAN 2390 R.[0]NGQMKAFLS..LARFSDTQYQR....IENYMKSSEFENKQ.....TLLKR[29]ECALRAL.REDRKRFLCKAVENYINCLLSGEEHDLWVF
TEL1_YEAST 2174 ST[0]LKALELIYK..NTKLPENER......KDAKRHYNRVLLQYNRD.....SEVLK[10]WHALHFYLNTLV..FSNR.....YDNDIIDKFCGLWFE
Q22258.?_CAEEL 1985 NT[0]IGFDKVYHE..TTQLLQCFAHSGVMYEAAWLLDYLSNYKERSKHVLPLLKAYKE[10]QARAVERMTS.....LWFS
ESR1_YEAST 1798 K.[0]PYYSIGLYY..SRLLERKK.....AEGYITNGRFEYRAIS.....YFLL[5].....TAKVRENLPKVITFWLD
Q24135.MEI41_DROME 1760 R.[0]QSESCHVQM..AQFLEKIL.....EARQ.CGKSEP.....TG[6]INVMVNYAKSLR..YGSE....HVYQSMPRLIS.LWLD
Q13535.ATR_HUMAN 2033 P.[0]EWEDGHEYL..AKYYDKLMP...MVTDNKMEKQCDLIRYIVL.....HFGR[9]YQSMPRMLT.....LWLD
Q92391.RAD3_SCHPO 1789 G.[0]HHRVLMMEE..EQKLPVNE.....QSERFLSCELVTRIIN.....EFGR[9]YESMPKLLT.....LWLD
O01438.FRAP_CAEEL 1892 RT[11]ENTINYYQQ..ATQYDPGW.....HKVWHKLASTHFYAVCRERPHPTTVISP[100]AHAVRCFAKALM..CSPG....SRLEDTLRLMQ.LWFD
FRAP_HUMAN 1769 .I[0]PKVLQYYSA..ATEHDRSW.....YKAWHAWAVMNFEAVL.....HYKH[75]VPAVQGFFRSIS..LSRG....NNLQDTLRVLT.LWFD
O14356.FRAP_SCHPO 1628 SV[0]RDILNCYFY..ATLFDKSW.....YKAWHSWALANFEVVG.....YYEQ[14]VPAIKGFFHSSV..LNQK....NSLQDILRLLN.LWFK
TOR1_YEAST 1730 NP[0]DAILGSYLL..ATHFDKNW.....YKAWHNWALANFEVIS.....MVQE[50]VPAIKGFFHSIS..LLET....SCLQDTLRLLT.LFN
P78527.DNAPK_HUMAN 3341 SS[1]PACLAEIEEDKRRILELSGSS..SEDSEKVIAGLYQRAFQHLS.....EAVQA[23]TLADFCDQQLRKEEENASVTDSAELQAYPA....LVVE
Q18667.TRRAP_CAEEL 3142 ..[0]ENADYTFSA..ASQLVDLQNSVTTTGIKLMKNWGHHLYKR.....FFST[11]RQALACYFIAAR..VDND....IKARKPIAKI..LWLS
O48823.TRRAP_ARATH 2996 ..[0]ESANIAMSN..AITLFKNL.....PKGWISWGSYCDM.....AYQE[7]EYAVSCFLQGIR..FGVS....NS.RSHIARV..LYLL
AF076974.TRRAP_HUMAN 3107 ..[0]EEANKAFSA..AVQMHDVL.....VKAWAMWGDYLEN.....IFVK[7]VSAITCYLHACR..HQNE....SKSRKYLAKV..LWLL
YAMB_SCHPO 2905 ..[0]DEANQAFAT..AVQIDIGS.....GKAWSEWGLYHDE.....LFQA[8]CNAVSCFLQASS..LLSS....SNSKPLLTRV..LWLL
YHP9_YEAST 3037 ..[0]EEANQAFAT..AVQIDLNL.....AKAWAQWGFFNDR.....RLSE[8]SNAISCYLQAAG..LYKN....SKIRELLCRI..LWLI

Q62388.ATM_HUMAN 2496 RLCSLWLENSGV.....SEVNGMMKKDGMKISSYKPLPLMYQLAARMGTKMTGGLGFHEVLNNLISRISLDHEHHTLSIILALANA 2576
TEL1_YEAST 2257 NDDNS.....KINQLLYKEIGTIPSWKQLPWVNQLASKIS..MEENE.FQKPLQLTWKRLLYKLEYDSLYSVMSILLY 2326
Q22258.?_CAEEL 2063 NTRKISTHISSVP.....KLPEGQISDLRQNIKSMNREIQTALEHIGWRAFYPAYAVLARHIDHQDEEVTRTIKQIMKQLILRMPHQCMWQSAYLLRQ 2155
ESR1_YEAST 1858 IAAASISEAPGNRK.....EMLSKATEDICSHVEEALQ.HCPTYIQYFVLTQLSRLL..HSHQSSAQII.MHILLSLAVEYFSHILWYITALVNS 1944
Q24135.MEI41_DROME 1827 TTESSTNTEQV.....KKMNDLLTNCCTALPTAVEYTVYSQMLSRLC..HPVNDVFTVLR.NVLIKLVEAYPQQSLWMLLPHFKS 1903
Q13535.ATR_HUMAN 2097 YGTKAYEWEKAGRS.....DRVQMRNDLGKINKVITEHTNYLAPYQPLTAFSQLISRIC..HSHDEVFVVL.MELIAKVFLAYPQQAMWMWTAVSKS 2185
Q92391.RAD3_SCHPO 1849 FGAEELRLSKDDGEKYFREHIISSRKKSLELMNSNVCRLSMKIPQYFPLVALSQMISRVC..HPNNKVYKIL.EHILANVVASYEGETLWQLMATIKS 1943
O01438.FRAP_CAEEL 2081 HGDDKDQ.....DVYFALTETIFDLPVTTWLEAIPQLMARLD..CPDDQKSVQLVLRVLCELARHRPQAVIWALTVASRS 2153
FRAP_HUMAN 1913 YGHWPD.....VNEALVEGVK.AIQIDTWLQVIPQLIARI...DTPRPLVGRLIHQLTDIGRYHPQALIWPLTVASKS 1982
O14356.FRAP_SCHPO 1712 FGEHSD.....VAAAIVEGFS.NVPMDTWLEVIPQLIARI...HTSSSSVRASVHQLLSDIGRVHPQALVWSLTVSSKS 1781
TOR1_YEAST 1850 FGGIKE.....VSQAMYEGFN.LMKIENWLEVLPQLISRI...HQPDPTVSNSLLSLLSDLGKAHPQALVWPLTVAIKS 1919
P78527.DNAPK_HUMAN 3448 KMLKALKLNSNEAR.LKFPRLLQIIERYPEETLSLMTKEISSVPCWQPISWISHMVALLD..KDQAVAVQ....HSVEEITDNYPQAIVWPFIISSES 3538
Q18667.TRRAP_CAEEL 3225 KHLNACGSH.....EVMNRVIKKQLHSLNLFNWLYWLPQLVTDVR..YKPNSNFV....LILCKAAAHFLQVFYHREAVSV 3296
O48823.TRRAP_ARATH 3066 SFDTANE.....PVGKVFDKHLDQVPHWVWLSWIPQLLSLQ..RTEAPHCK....LVLLKIAAVFPQALYWWLRTYLLE 3134
AF076974.TRRAP_HUMAN 3178 SFDDDKN.....TLADAVDKYCIGVPPIQWLAWIPQLLTCLV..GSEGKLLLN....LISQVGRVYPQAVYGPIRTLYLT 3246
YAMB_SCHPO 2977 SVDDSHG.....SVSEVVSSFKSEIPTWNWIPFIPQLLSALS..HRESIHAR....AILIQIAKTYPQSLHGQLRTAYED 3045
YHP9_YEAST 3109 SIDDASG.....MLTNAFDSFRGEIPVWYWIPFIPQLLTSLS..HKEANMVR....HILIRIAKSYPQALHGQLRTTKED 3177

TIBS

proteins[18]. 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 PI-kinase 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 complexes[24]. We therefore speculate that the FAT domain could be of importance as a structural scaffold or as a protein-binding domain, or both.

References

1 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 functions of the mammalian target of rapamycin by the phosphoinositide 3-kinase 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
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
19 Baskaran, R. *et al.* (1997) Ataxia telangiectasia mutant protein activates c-Abl 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 ligand-modulated 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 functions[3]. 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 organisms[4]. 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 cerevisiae*[5,6] 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

 PII: S0968-0004(00)01579-6