

Atypical Protein Kinases (aPKs)

Supplemental Online Material for Manning et al (Science 298:1912-1934)

Atypical protein kinases lack sequence similarity to the ePK domain HMM profile, but have been shown experimentally to have protein kinase activity, or are clear homologs of aPKs with demonstrated protein kinase activity. All aPK families are small, several having just one member in vertebrates, and none in invertebrates. Other aPK families may remain to be discovered by biochemical methods, but since most atypical families are small, and most biochemically-discovered kinases are ePKs, it is unlikely that many new atypical kinases will be discovered.. Conversely, some of these aPKs may be false positives, if the reports of kinase activity are not correct.

Alpha Kinases

The progenitors of this family are the myosin heavy chain kinases (MHCKs) of *Dictyostelium discoideum*. While these are evolutionarily restricted, they are similar to the eukaryotic elongation factor 2 kinase (eEF2K) found in most eukaryotes. Several other mammalian genes have been found to be homologous to these, including the channel kinases Chak1 and Chak2, which are multi-pass transmembrane proteins which act as kinases and as ion channels. Crystal structure of the CHAK1 gene (Yamaguchi et al, 2001) shows some structural similarity to the ePK domain fold, and catalytic activity has been demonstrated for many members of this family.

PIKK – Phosphatidyl inositol 3' kinase-related kinases

This family contains a phosphatidyl inositol 3,4, kinase domain (PI34K or PI3K), flanked by a N-terminal FAT domain and a C-terminal FATC domain (Bosotti et al, 2000). Five of the six human members of this family have experimentally verified protein kinase activity and probably do not function as phosphatidyl inositol kinases. Multiple sequence alignment shows that the PI34K domains of PIKKs form a distinct domain subfamily. The PI34K domain is structurally related to the ePK domain; a structural alignment of ePK and PI34K domains shows similar structure and conservation of most of the catalytically conserved residues, including the 'catalytic' K, the DFG motif and the HRD motif (modified to DRH) (see for instance Walker et al, 1999).

A6

This family consists of human A6 and A6r genes, along with homologs in fly, worm and yeast. A6 was first cloned as a phospho-protein by Beeler et al (1994), who demonstrated tyrosine kinase activity by bacterial expression A6 fusion protein in an *in vitro* kinase activity. A6r was discovered by Rohwer et al (1999) by two-hybrid interaction with PKC zeta; they show that both A6 and A6r bind ATP, however, they failed to see kinase activity by either protein. To the best of our knowledge, only one report supports the A6 family being a family of protein kinases.

ABC1/ADCK (ABC1 domain containing kinase)

This conserved family was identified as putative kinases by sequence alignment methods (Psi-Blast and HMMs) which show a domain that is weakly similar to the ePK domain, with particular conservation of the most conserved catalytic motifs. Their kinase similarity was first published by Leonard et al in 1998. Despite the lack of overall sequence conservation with the ePK domain, these kinases contain candidates for the most conserved kinase motifs, including the VAIK catalytic motif (VAVK, VAMK), the DFG motif, and a QTD motif that may take the place of the HRD motif.

PDK - Pyruvate Dehydrogenase Kinases

This family of mitochondrial kinases contains a domain which is similar to prokaryotic histidine kinases, but has been biochemically to phosphorylate serine rather than histidine. Crystal structures (Machius et al, 2001, Steussy et al, 2001) confirm that the PDK domain fold is similar to that of histidine kinases and Hsp90, and is distinct from the ePK domain.

RIO

This family has 3 clear subfamilies, with one member of each in fly, worm and human. Yeast has two members (in the RIO1 and RIO2 subfamilies) and the fungus *Aspergillus nidulans* has a member of the third subfamily, RIO3. Homologs are also present in several archeal genomes. Yeast RIO1 was recently published to have serine kinase activity by Angermayr et al (2002). The sequences do not align with the eukaryotic protein kinase domain, but many of the catalytic residues are strongly conserved in the RIO family, and overall structural similarity to ePKs has been predicted by Leonard et al (1998)

BRD - Bromodomain Kinases

This family consists of the BRD2 (RING3) kinase and homologs in human and model organisms. Dennis and Green (1996) first identified BRD2 as an autophosphorylating nuclear-specific protein in Hela extracts. Recombinant BRD2 expressed in *E. coli* showed kinase activity in an *in vitro* assay, which was abolished by mutation of the putative catalytic lysine (K578A). The kinase activity was only seen when the purified recombinant protein was first incubated with Hela cell extract and repurified, possibly due to the need for activation of BRD2 by phosphorylation by another kinase. Recombinant BRD2 purified from COS cells also showed *in vitro* kinase activity, without the need for an activation step. Alignment of BRD2 with its human homologs (BRD1, BRD3, BRD4), *Drosophila* fsh and an anonymous worm homolog shows the presence of two bromodomains, and a third conserved region, which contains some similarity to the ePK domain by secondary structure prediction (I. Grigoriev, unpublished).

TAF – TATA binding factor associated factors

TAF1 (TAF II-250) is a component of the basal transcriptional machinery, and exists as a single copy gene in all fully-sequenced eukaryotes. It has no close homologs. Dikstein et al (1996) report that TAF1 is a protein kinase and contains two regions which can independently phosphorylate the basal transcription factor RAP74. *In vitro* kinase assays

were carried out with immunopurified TFIID from HeLa cells or with cloned TAF1 transfected into insect Sf9 cells or E. coli. Deletion mapping showed that two independent regions, each less than 470 AA long had kinase activity, though neither had significant sequence similarity to each other, to protein kinases, or to any other proteins. Later studies (Solow et al, 2001, O'Brien and Tijan, 1998) confirm the result and perform a finer mapping of the N-terminal kinase region. In 2002, Wang and Page revealed the presence of TAF1L, a retrotransposed copy of TAF1 present in human and old-world primates, which is expressed during spermatogenesis and substitutes for TAF1 in a cellular assay.

BCR

Best known as the fusion partner of the Abl kinase in chronic myelogenous leukemia, the BCR gene itself also has protein kinase activity. Maru and Witte (1991) showed that highly-purified BCR has auto- and trans-phosphorylation activities, and later mapped cysteines and tyrosines that were critical for kinase activity. The kinase domain appears to be a recent addition to the protein: the human ABR gene is about 70% identical in AA sequence, but lacks the N-terminal putative kinase domain. The Drosophila gene EG:23E12.2 (gi|7289304) also lacks the N-terminal putative kinase domain, but is 36% identical over 750 AA in the C-terminus. The region conserved between these three proteins includes a GTPase activator domain.

FASTK

The human Fas-activated s/t kinase (FASTK) was characterised by Tian et al (1995) as a kinase which was dephosphorylated and activated by Fas-mediated apoptosis. The nuclear TIA-1 RNA-binding protein, a putative apoptosis effector, was identified as a substrate. Differential expression of FASTK in apoptotic cells has also been reported by Brutsche et al (2001). A close mouse ortholog is known (NP_148936.1), with 89% AA identity. Two other very distant putative mammalian homologs have been sequenced, but with ~27% identity between these genes, are not close enough to confidently assign a kinase function to them.

G11

This family consists of a single gene called G11 or STK19, which was shown by Gomez-Escobar et al (1998). to have serine/threonine kinase activity against alpha casein and histone, showed ATP-binding function and identified a putative catalytic lysine required for function (unlike ePKs, this lysine is near the C-terminal end of the protein). The G11 protein was made in transfected insect or COS-7 cells and immunoprecipitated, so it remains possible that the kinase activity was due to a tightly bound protein. Clear homologs are found in rat and mouse (~85% AA identity throughout), and a divergent putative homolog fragment has been sequenced in zebrafish (45% identity over 57 AA), but no obvious homolog has been seen in any other organism.

TIF1 – Transcriptional Intermediary Factor 1 family

A family of three Transcriptional Intermediary Factor 1 genes (TIF1a,b,g), of which TIF1a has been shown to be a protein kinase (Fraser et al, 1998); the two other genes are similar across their full length and so also likely to be kinases. A single TIF1 exists in the *Drosophila* and mosquito genomes but not in *C. elegans*. Autophosphorylation activity was detected in immunopurified, baculovirus-produced protein. Like the TAFs, TIFs are involved in the transcriptional machinery, and are thought to phosphorylate several other TATA-associated factors. Also similar to the TAFs and to BRD kinases, the TIFs contain bromodomains, suggesting that they may be in some way involved in the kinase functions of these proteins. A fourth putative member of this family exists in human, known as KIAA0298 (Genbank sequence gi|13509324). However, it is divergent in regions that are conserved between the other TIF1 family members. It may therefore not have conserved the kinase sequence or activity and has not been included in our kinase catalog.

H11

The H11 gene is a homolog of the ICP10 gene of Herpes simplex virus, both of which have been indicated to have kinase activity (Nelson et al, 1996, Smith et al, 2000). H11 kinase activity was demonstrated in protein immunopurified from bacterial and eukaryotic expression systems. The kinase activity was lost when a putative catalytic lysine was mutated. Smith et al noted weak similarities between H11 and the FASTK atypical kinase, but it is not known if these are significant. H11 belongs to the HSP20/Alpha Crystallin family of proteins, and has a number of close and moderate homologs. Since the kinase region hasn't been mapped in this gene, it was not possible to confidently say if any of these homologs also have kinase activity, so only H11 has been included in the kinome catalog.

References

- Angermayr, M., A. Roidl, et al. (2002). "Yeast Rio1p is the founding member of a novel subfamily of protein serine kinases involved in the control of cell cycle progression." *Mol Microbiol* **44**(2): 309-24.
- Beeler, J. F., W. J. LaRochelle, et al. (1994). "Prokaryotic expression cloning of a novel human tyrosine kinase." *Mol Cell Biol* **14**(2): 982-8.
- Bosotti, R., A. Isacchi, et al. (2000). "FAT: a novel domain in PIK-related kinases." *Trends Biochem Sci* **25**(5): 225-7.
- Brutsche, M. H., I. C. Brutsche, et al. (2001). "Apoptosis signals in atopy and asthma measured with cDNA arrays." *Clin Exp Immunol* **123**(2): 181-7.
- Denis, G. V. and M. R. Green (1996). "A novel, mitogen-activated nuclear kinase is related to a *Drosophila* developmental regulator." *Genes Dev* **10**(3): 261-71.
- Dikstein, R., S. Ruppert, et al. (1996). "TAFII250 is a bipartite protein kinase that phosphorylates the basal transcription factor RAP74." *Cell* **84**(5): 781-90.

- Gomez-Escobar, N., C. F. Chou, et al. (1998). "The G11 gene located in the major histocompatibility complex encodes a novel nuclear serine/threonine protein kinase." J Biol Chem **273**(47): 30954-60.
- Leonard, C. J., L. Aravind, et al. (1998). "Novel families of putative protein kinases in bacteria and archaea: evolution of the "eukaryotic" protein kinase superfamily." Genome Res **8**(10): 1038-47.
- Machius, M., J. L. Chuang, et al. (2001). "Structure of rat BCKD kinase: nucleotide-induced domain communication in a mitochondrial protein kinase." Proc Natl Acad Sci U S A **98**(20): 11218-23.
- Maru, Y. and O. N. Witte (1991). "The BCR gene encodes a novel serine/threonine kinase activity within a single exon." Cell **67**(3): 459-68.
- Nelson, J. W., J. Zhu, et al. (1996). "ATP and SH3 binding sites in the protein kinase of the large subunit of herpes simplex virus type 2 of ribonucleotide reductase (ICP10)." J Biol Chem **271**(29): 17021-7.
- O'Brien, T. and R. Tijan (1998). "Functional Analysis of the Human TAFII-250 N-terminal Kinase Domain." Molecular Cell **1**: 905-911.
- Rohwer, A., W. Kittstein, et al. (1999). "Cloning, expression and characterization of an A6-related protein." Eur J Biochem **263**(2): 518-25.
- Smith, C. C., Y. X. Yu, et al. (2000). "A novel human gene similar to the protein kinase (PK) coding domain of the large subunit of herpes simplex virus type 2 ribonucleotide reductase (ICP10) codes for a serine-threonine PK and is expressed in melanoma cells." J Biol Chem **275**(33): 25690-9.
- Solow, S., M. Salunek, et al. (2001). "Taf(II) 250 phosphorylates human transcription factor IIA on serine residues important for TBP binding and transcription activity." J Biol Chem **276**(19): 15886-92.
- Steussy, C. N., K. M. Popov, et al. (2001). "Structure of pyruvate dehydrogenase kinase. Novel folding pattern for a serine protein kinase." J Biol Chem **276**(40): 37443-50.
- Tian, Q., J. Taupin, et al. (1995). "Fas-activated serine/threonine kinase (FAST) phosphorylates TIA-1 during Fas-mediated apoptosis." J Exp Med **182**(3): 865-74.
- Walker, E. H., O. Perisic, et al. (1999). "Structural insights into phosphoinositide 3-kinase catalysis and signalling." Nature **402**(6759): 313-20.
- Wang, P. J. and D. C. Page (2002). "Functional substitution for TAF(II)250 by a retroposed homolog that is expressed in human spermatogenesis." Hum Mol Genet **11**(19): 2341-6.
- Yamaguchi, H., M. Matsushita, et al. (2001). "Crystal structure of the atypical protein kinase domain of a TRP channel with phosphotransferase activity." Mol Cell **7**(5): 1047-57.