

Genomic analysis of the eukaryotic protein kinase superfamily: a perspective

Steven K Hanks

Address: Department of Cell and Developmental Biology, Vanderbilt University School of Medicine, Nashville, TN 37232, USA.
E-mail: steve.hanks@vanderbilt.edu

Published: 29 April 2003

Genome Biology 2003, **4**:111

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2003/4/5/111>

© 2003 BioMed Central Ltd

Abstract

Protein kinases with a conserved catalytic domain make up one of the largest 'superfamilies' of eukaryotic proteins and play many key roles in biology and disease. Efforts to identify and classify all the members of the eukaryotic protein kinase superfamily have recently culminated in the mining of essentially complete human genome data.

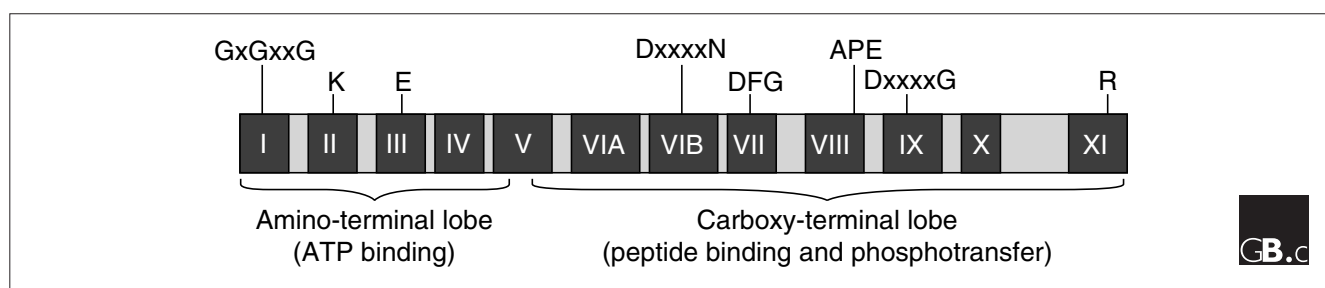
Phosphorylation by protein kinases is recognized as a major mechanism by which virtually every activity of eukaryotic cells is regulated, including proliferation, gene expression, metabolism, motility, membrane transport, and apoptosis. An ultimate goal of research into signal transduction is to reach a full understanding of the protein phosphorylation events that occur within individual cell types and how they eventually impact on cell behavior. A milestone *en route* to this ambitious goal is a determination of the number of protein kinases encoded by eukaryotic genomes and an assessment of their structures, functions, and evolutionary relationships. This article traces the progress made toward achieving these objectives in the pregenomic and genomic eras, which culminated recently with reports on the 'full complement' of human protein kinases.

The pregenomic era

About sixteen years ago, while working at the Salk Institute, my colleagues and I undertook a comparative analysis of all the available sequences of protein kinase catalytic domains [1]. This interest stemmed from my having identified several novel human protein kinases using a homology-based cDNA cloning strategy [2] and wanting to determine their relationships to other known protein kinases. In collaboration with the Salk's resident protein kinase guru Tony Hunter and bio-computing specialist Anne Marie Quinn, we aligned the homologous catalytic-domain amino-acid sequences of 65

distinct protein kinases from diverse eukaryotes (including 45 nonorthologous vertebrate enzymes) and constructed a phylogenetic tree to visualize their overall relationships [1]. The alignment (produced manually at the word-processor) defined the boundaries of the eukaryotic protein kinase (ePK) catalytic domain, revealed conserved subdomains that were never interrupted by amino-acid insertions, and identified highly conserved individual amino acids and motifs (Figure 1).

The phylogenetic tree revealed major clusters including the tyrosine kinases (the TK group), cyclic nucleotide- and calcium-phospholipid-dependent kinases (the AGC group; including the PKA, PKG, and PKC families) and calmodulin-dependent kinases (the CAMK group). These groupings indicated that ePK domain phylogeny reflects substrate specificity and/or mode of regulation and could therefore serve as a useful classification tool. Over the next 7 years I continued to add new sequences to the alignment as they became available and to construct phylogenetic trees as a means of classifying the burgeoning ePK superfamily. By early 1994, the ePK domain alignment had grown to contain 390 sequences including 205 non-orthologous vertebrate ePKs, and a fourth major ePK group (CMGC, comprising the CDK, MAPK, GSK, and CLK families) had been added through phylogenetic analysis [3]. The 390 ePK domain alignment was made publicly available through the Protein Kinase Resource website [4].

**Figure 1**

The ePK catalytic domain. The 12 conserved subdomains are indicated by Roman numerals. The positions of amino-acid residues and motifs highly conserved throughout the ePK superfamily are indicated above the subdomains, using the single-letter amino-acid code with x as any amino acid. Crystal structures show that ePK domains adopt a common fold consisting of amino-terminal and carboxy-terminal lobes connected by a hinge region. Binding of Mg-ATP is largely the function of the amino-terminal lobe and hinge region, while peptide-substrate binding is mediated by the carboxy-terminal lobe. Particularly important for catalytic function are the invariant lysine in subdomain II and the invariant aspartate in subdomain VII that function to anchor and orient ATP, and the invariant aspartate in subdomain VIB which is the likely catalytic base in the phosphotransfer reaction. More detailed discussions of ePK subdomains and conserved residues in relation to crystal structures and catalytic function can be found in [3,4,12,13].

The genomic era

By 1995, with the advent of genome-sequencing projects, the task of cataloging and classifying the members of the ePK superfamily had grown to become too distracting from my funded research and I discontinued my efforts in this area. Tony Hunter continued to work with bioinformaticians at SUGEN, Inc. (including Greg Plowman, Gerard Manning, and Sucha Sudarsanam) to characterize the full ePK complements of model eukaryotes from genomic sequence data [5,6]. By the time of a recent report [7], their efforts had resulted in the identification and classification of 115 distinct ePKs from budding yeast (around 2% of all genes), 434 from *Caenorhabditis elegans* (about 2.5% of all genes), and 223 from *Drosophila*. In addition they described the complement of 'atypical protein kinases' (aPKs) from these species: 15 from yeast, 20 from *C. elegans*, and 16 from *Drosophila*. (The aPKs are a variety of protein kinases that lack strong sequence similarity to the classical ePK domain but have been shown experimentally to have protein kinase activity; well-known examples are the 'lipid kinases' of the phosphatidylinositol 3'-kinase (PI3K) family, some of which have been shown experimentally to have protein kinase activity.)

As a result of their comprehensive analyses of 'kinomes', the SUGEN investigators were able to define three new major groups within the broad ePK classification scheme: first, the STE group, which includes ePKs that function in the MAPK kinase cascades that were first described through characterization of yeast *sterile* mutants; second, the CK1 group, including the casein kinase 1 family and related enzymes, which is greatly expanded in the worm; and third, the TKL ('tyrosine-kinase like') group that includes the STKR family of TGFbeta serine/threonine kinase receptors and is phylogenetically close to the tyrosine kinases (TKs). Many distinct kinase families within the AGC, CAMK, CMGC, STE, and CK1 groups have representatives from all three species, supporting the idea of an early evolutionary origin and critical

function in basic cellular processes. Members of the TK and TKL groups are notably absent from yeast, consistent with the known functions of these ePKs in intercellular signaling events associated with metazoan complexity. More discussion of the evolutionary relationships among the ePKs identified through the SUGEN genome-mining efforts has been published elsewhere [7]. The SUGEN kinase.com website [8] includes links to all their published work on protein kinase analysis as well as 'KinBase', a very useful searchable database that holds information on all the protein kinase genes found in the yeast, worm, fly, and human (see below) genomes.

Human protein kinases

The completion of the first draft of the human genome sequence presented an opportunity to determine the full complement of human protein kinases. The first analysis came from a group led by Mitch Kostich at Schering-Plough Research Institute (SPRI) [9]. This group mined public GenBank records (available before December, 2001) for ePK sequences by performing BLAST searches using known ePK domains as queries. The resulting hits were consolidated, and efforts were made to remove non-human sequences, pseudogenes, and poor-quality sequences that could represent duplicate hits. The SPRI investigators chose to err on the side of inclusion rather than exclusion, however, and many cases of 'single hit' sequences were retained. Their effort culminated in a collection of 510 potentially unique human ePKs. A color-coded alignment that accompanied their article [9] nicely illustrates the ePK domain sequence conservation.

The SUGEN group, led by Gerard Manning and Sucha Sudarsanam, carried out a more comprehensive effort to describe and classify all human ePKs [10]. They employed a dataset that included, in addition to the public databases, genomic reads from Celera that are not publicly available, non-public expressed sequence tags (ESTs) from Incyte and SUGEN, and they searched using a hidden Markov model of

the ePK domain that allowed detection of very divergent family members. The sequence data were further searched for members of the various known aPK families. Using stringent criteria to eliminate false positives (including verification of novel sequences by cDNA cloning) they compiled a list of 478 human members of the ePK superfamily and another 40 aPKs, bringing their human kinome total to 518 (approximately 1.7% of all predicted human genes). They also identified 106 ePK or aPK pseudogenes.

A comparison of the SPRI-510 and SUGEN-518 lists reveals 474 protein kinases in common (see the additional file available with this article online). Of the 44 SUGEN-specific kinases, 32 are aPKs; the other 8 aPKs identified by SUGEN, from the ABC1 and RIO families, were included in the SPRI list as a result of their having weak ePK domain similarity. Of the remaining 12 SUGEN-specific ePKs, five (TAK1, MLKL, NEK5, SgK307, and TBCK) were not available in the public data used in the SPRI analysis; another five (SgK196, SgK223, SgK424, SgK493, and Slob) have rather divergent ePK domains that lack many of the highly conserved residues and are unlikely to have catalytic activity, so it is easy to see how these might have been excluded by visual inspection; and the final two are SgK110 and NEK10. SgK110 was actually detected by the SPRI search, but it was erroneously merged with a related sequence AC008735_EPK1 (SgK069) on the same genomic contig; and it is unclear why the SPRI group missed NEK10. Most, if not all, of the 36 SPRI-specific ePKs represent over-inclusion errors (Table 1): 14 correspond to sequences determined to be pseudogenes by the SUGEN group; 19 are based on single sequences that are (or appear to be) either poor-quality duplicates of other ePKs or interspecies contaminants; and the remaining three are duplicates arising by virtue of non-overlapping partial sequences.

Thus the SUGEN compilation of 478 human ePK superfamily genes represents the accurate count based on current sequence data. If one subtracts those that lack key conserved residues, we are left with 428 human ePKs with known or likely kinase function (Table 2), 99% of which were included in the SPRI list; 365 of these fall within the seven major ePK groups: TK, 84 in total; CAMK, 66; AGC, 61; CMGC, 61; STE, 45; TKL, 37; and CK1, 11. The remaining 63 are in the 'Other' category, falling outside the main ePK group branches. Krupa and Srinivasan [11] have also recently searched the public human genome data with a focus on identifying functional protein kinases; their efforts resulted in a list of 448 distinct human ePK sequences, but around 90 of these appear to represent duplicate entries, and no novel protein kinases were identified that were not present in the SUGEN compilation.

Usefulness of the kinome data

Knowing the full complement of ePK family members and functional ePKs encoded by eukaryotic genomes will have great impact upon many areas of scientific investigation. As mentioned above, an obvious benefit relates to understanding

Table 1
Putative ePKs identified by SPRI but not SUGEN

Category	SPRI name	Comment
14 pseudogenes	LOC95530	Corresponds to SUGEN MAP2K2ps
	AC008014_EPK1	Corresponds to SUGEN MARKps22
	AC023095_EPK1	Corresponds to SUGEN PLK1ps1
	AC024933_EPK1	Corresponds to SUGEN MARKps30
	AC091554_EPK1	Corresponds to SUGEN SRPK2ps
	AI621045	Corresponds to SUGEN SgK384ps
	AI991174	Corresponds to SUGEN TSSKps2
	AL138964_EPK1	Corresponds to SUGEN MARKps23
	AL161450_EPK1	Corresponds to SUGEN CK1g2ps
	AL390061_EPK1	Corresponds to SUGEN MNK1ps
	BG742738	Corresponds to SUGEN TLK2ps2
	LOC65348	Corresponds to SUGEN CDK7ps
	LOC65743	Corresponds to SUGEN PAK2ps
	LOC85643	Corresponds to SUGEN CLK2ps
11 duplicates from poor-quality single-EST sequences	AI342911	Corresponds to SUGEN CLK3
	AA565761	Corresponds to SUGEN CLK3
	AI826899	Corresponds to SUGEN CDK4
	AW148845	Corresponds to SUGEN ILK
	AW674733	Corresponds to SUGEN ILK
	AA730234	Corresponds to SUGEN MAP2K7
	AI075923	Corresponds to SUGEN TSSK4
	AI206530	Corresponds to SUGEN TLK1
	AW518044	Corresponds to SUGEN MPSK1
Four non-human sequences	AL538014	Corresponds to SUGEN BRSK2
	BG986727	Corresponds to SUGEN Obscn
	AC025598	<i>Mus musculus</i> MAP2K6
	AC026940_EPK1	<i>Mus musculus</i> CDK7
	PKCETB	<i>Rattus norvegicus</i> PKCeta
	CAMKII	<i>Sus scrofa</i> sequence, subsequently recalled
Three duplicates from non-overlapping partial sequences	AC023837_EPK1	Part of EphA6
	AA263006	Part of HIPK1 (along with BE180036)
	BE567816	Part of YANK1 (along with BG036777)
Four others	AW499744	Single EST; does not seem to encode an ePK
	AJ336398_EPK1	Single genomic sequence, possibly prokaryotic
	AL512402_EPK1	Single genomic sequence, possibly nonhuman
	ILK2	Possibly murine ILK; published as human [16]


Table 2**The 428 human ePKs with known or likely kinase catalytic function**

Group	Number within group	Family	Number within family	Family members (SUGEN nomenclature)
TK	84	Ack	2	ACK, TNK1
		Abl	2	ABL, ARG
		Csk	2	CSK, CTK
		FAK	2	FAK, PYK2
		Fer	2	FER, FES
		JAK	4	JAK1, JAK2, JAK3, TYK2
		Src	11	BLK, BRK, FGR, FRK, FYN, HCK, LCK, LYN, SRC, SRM, YES
		Syk	2	SYK, ZAP70
		Tec	5	BMX, BTK, ITK, TEC, TXK
		Alk	2	ALK, LTK
		Axl	3	AXL, MER, TYRO3
		DDR	2	DDR1, DDR2
		EGFR	3	EGFR, HER2/ErbB2, HER4/ErbB4
		Eph	12	EphA1, EphA2, EphA3, EphA4, EphA5, EphA6, EphA7, EphA8, EphB1, EphB2, EphB3, EphB4
		FGFR	4	FGFR1, FGFR2, FGFR3, FGFR4
		InsR	3	IGF1R, INSR, IRR
		Lmr	3	LMR1, LMR2, LMR3
		Met	2	MET, RON
		Musk	1	MUSK
		PDGFR/VEGFR	8	FLT3, FMS, KIT, PDGFRa, PDGFRb, FLT1, FLT4, KDR
		Ret	1	RET
		Ror	2	ROR1, ROR2
		Sev	1	ROS
		Tie	2	TIE1, TIE2
		Trk	3	TRKA, TRKB, TRKC
AGC	61	PKA	5	PKACa, PKACb, PKACg, PRKX, PRKY
		PKG	2	PKG1, PKG2
		PKC	9	PKCa, PKCb, PKCd, PKCe, PKCg, PKCh, PKCi, PKCt, PKCz
		AKT	3	AKT1, AKT2, AKT3
		DMPK	7	CRIK, DMPK1, DMPK2, MRCKa, MRCKb, ROCK1, ROCK2
		GRK	7	BARK1, BARK2, GPRK4, GPRK5, GPRK6, GPRK7, RHOK
		MAST	5	MAST1, MAST2, MAST3, MAST4, MASTL
		NDR	4	LATS1, LATS2, NDR1, NDR2
		PKB	1	PDK1
		PKN	3	PKN1, PKN2, PKN3
		RSK	9	MSK1, MSK2, RSK1, RSK2, RSK3, RSK4, SgK494, p70S6K, p70S6Kb
		SGK	3	SGK, SGK2, SGK3
		YANK	3	YANK1, YANK2, YANK3
CAMK	66	CAMK1	5	CaMK1a, CaMK1b, CaMK1d, CaMK1g, CaMK4
		CAMK2	4	CaMK2a, CaMK2b, CaMK2d, CaMK2g
		CAMKL	20	AMPKa1, AMPKa2, BRSK1, BRSK2, CHK1, HUNK, LKB1, MARK1, MARK2, MARK3, MARK4, MEL, NIM1, NuaK1, NuaK2, PASK, QIK, QSK, SIK, SNRK
		DAPK	5	DAPK1, DAPK2, DAPK3, DRAK1, DRAK2
		DCAMKL	3	DCAMKL1, DCAMKL2, DCAMKL3
		MAPKAPK	5	MAPKAPK2, MAPKAPK3, MAPKAPK5, MNK1, MNK2
		MLCK	4	caMLCK, skMLCK, smMLCK, SgK085
		PHK	2	PHKg1, PHKg2
		PIM	3	PIM1, PIM2, PIM3
		PKD	3	PKD1, PKD2, PKD3
		PSK	1	PSKH1
		RAD53	1	CHK2
		Trio	4	Obscn, SPEG, Trad, Trio

Table 2 (continued)

Group	Number within group	Family	Number within family	Family members
CMGC	61	TSSK	5	SSTK, TSSK1, TSSK2, TSSK3, TSSK
		CAMK-Unique	1	STK33
		CDK	20	CCRK, CDC2, CDK2, CDK3, CDK4, CDK5, CDK6, CDK7, CDK8, CDK9, CDK10, CDK11, CHED, CRK7, PCTAIRE1, PCTAIRE2, PCTAIRE3, PFTAIRE1, PFTAIRE2, PITSLRE
		MAPK	14	Erk1, Erk2, Erk3, Erk4, Erk5, Erk7, JNK1, JNK2, JNK3, NLK, p38a, p38b, p38d, p38g
		GSK	2	GSK3A, GSK3B
		CLK	4	CLK1, CLK2, CLK3, CLK4
		CDKL	5	CDKL1, CDKL2, CDKL3, CDKL4, CDKL5
		DYRK	10	DYRK1A, DYRK1B, DYRK2, DYRK3, DYRK4, HIPK1, HIPK2, HIPK3, HIPK4, PRP4
		RCK	3	ICK, MAK, MOK
STE	45	SRPK	3	MSSK1, SRPK1, SRPK2
		STE7	7	MAP2K1, MAP2K2, MAP2K3, MAP2K4, MAP2K5, MAP2K6, MAP2K7
		STE20	28	GCK, HPK1, KHS1, KHS2, LOK, MST1, MST2, MST3, MST4, MYO3A, MYO3B, OSR1, PAK1, PAK2, PAK3, PAK4, PAK5, PAK6, SLK, STLK3, TAO1, TAO2, TAO3, YSK1, ZC1/HGK, ZC2/TNIK, ZC3/MINK, ZC4/NRK
		STE11	8	MAP3K1, MAP3K2, MAP3K3, MAP3K4, MAP3K5, MAP3K6, MAP3K7, MAP3K8
		STE-Unique	2	COT, NIK
TKL	37	IRAK	2	IRAK1, IRAK4
		LISK	4	LIMK1, LIMK2, TESK1, TESK2
		LRRK	2	LRRK1, LRRK2
		MLK	9	DLK, HH498, LZK, MLK1, MLK2, MLK3, MLK4, TAK1, ZAK
		RAF	3	ARAF, BRAF, RAF1
		RIPK	5	ANKRD3, RIPK1, RIPK2, RIPK3, SgK288
		STKR	12	ACTR2, ACTR2B, ALK1, ALK2, ALK4, ALK7, BMPR1A, BMPR1B, BMPR2, MISR2, TGFbR1, TGFbR2
CKI	11	CKI	7	CK1a, CK1a2, CK1d, CK1e, CK1g1, CK1g2, CK1g3
		TTBK	2	TTBK1, TTBK2
		VRK	2	VRK1, VRK2
Other	63	Aur	3	AurA, AurB, AurC
		BUB	1	BUB1
		Bud32	1	PRPK
		CAMKK	2	CaMKK1, CaMKK2
		CDC7	1	CDC7
		CK2	2	CK2a1, CK2a2
		IKK	4	IKKa, IKKb, IKKe, TBK1
		IRE	2	IRE1, IRE2
		MOS	1	MOS
		NAK	4	AAK1, BIKE, GAK, MPSK1
		NEK	11	NEK1, NEK2, NEK3, NEK4, NEK5, NEK6, NEK7, NEK8, NEK9, NEK10, NEK11
		NKF1	3	SBK, SgK069, SgK110
		NKF2	1	PINK1
		NKF4	2	CLIK1, CLIK1L
		PEK	4	GCN2, HRI, PEK, PKR
		PLK	4	PLK1, PLK2, PLK3, PLK4
		TLK	2	TLK1, TLK2
		TOPK	1	PBK
		TTK	1	TTK
		ULK	4	Fused, ULK1, ULK2, ULK3
		VPS15	1	PIK3R4
		WEE	3	MYT1, Wee1, Wee1B
		Wnk	4	Wnk1, Wnk2, Wnk3, Wnk4
		Other-Unique	2	KIS, SgK496

	VIB	VIII
TK	H r D l A a R N R t A	l P i k W m a p E f v F t p l
AGC	Y R D l K p e N H i l d	G T p e y i a p E d m
CAMK	H r D l k p e N l i a	g t p x y a a p e s f v
CMGC	H r D l K p e N t i s	V T r w Y R a p e S l y p
STE	H r D i k g x N v a	G t p y w a p E f y
TKL	h r D l k s x N	g t x r y m a p E s a w
CK1	h r D V K p x N I	G T a r y a S i a v m


Figure 2

Conserved residues implicated in peptide-substrate recognition. Consensus motifs for the catalytic loop region in subdomain VIB and activation loop region in subdomain VIII were determined for the members of each of the seven major ePK groups with known or likely kinase activity. Invariant residues at a given position are indicated by single upper-case letters. Two upper-case letters at a single position indicate that either of two residues are strictly conserved, the most frequent shown in the top row. Positions in which more than two amino acids are present are indicated with lower-case letters; a single letter indicates that only one residue is highly conserved, two letters indicate that either of two residues are frequently conserved (most frequent on the top row), and 'x' indicating poor positional conservation. Residues highlighted in outline are notably conserved within an ePK group and are thought to function in the recognition of peptide substrates specifically targeted by the members of the group.

of how signal transduction pathways evolved during the course of eukaryotic evolution. Both SUGEN [10] and Krupa and Srinivasan [11] extended their analyses to describe other domains present in the various human ePKs which are likely to function in directing the enzymes to relevant substrates or modulating kinase activities. Further analysis of the ePK domain sequences uniquely conserved within the major groups and families, together with comparisons of ePK domain crystal structures, should ultimately allow a full understanding of how different classes of peptide substrate are recognized. For example, Figure 2 shows consensus sequences for the catalytic loop region in subdomain VIB (which includes the invariant aspartate thought to function as the catalytic base) and the activation loop region in subdomain VIII (which includes the highly conserved

glutamine in the 'APE' motif) - two regions that have been recognized as being primarily involved in peptide-substrate recognition [12,13]. A number of group-specific differences are apparent (highlighted in Figure 2) that correlate with unique peptide-recognition tendencies for the ePKs that fall within a given group [14]. Beyond sequence analysis, the kinome data will allow for the development of comprehensive tools (such as full-length cDNAs, microarrays, antibodies, and fusion protein and RNAi constructs) that will greatly aid laboratory investigations aimed at understanding cell signaling through analysis of kinase function. As an example of such proteomic approaches to the study of protein kinases, nearly all yeast protein kinases have been expressed in bacteria and analyzed for their ability to phosphorylate an array of protein or peptide substrates using protein-chip technology [15]. Finally, the human kinome data will have benefits in the understanding and treatment of human diseases. The ePK genes that map within disease loci are attractive etiological candidates, and knowledge of the full repertoire of human protein kinases will greatly aid in the development of drugs that target specific protein kinases or protein kinase families whose function contributes to disease-associated cellular defects.

Acknowledgements

I am indebted to Gerard Manning of SUGEN, Mitch Kostich of Schering-Plough Research Institute, and N. Srinivasan of the Indian Institute of Science, for their contributions and comments regarding comparative analysis of their respective human protein kinase compilations.

Additional data file

An additional data file with the 474 protein kinases in common between the SPRI-510 and the SUGEN-518 lists is available with the online version of this article.

References

- Hanks SK, Quinn AM, Hunter T: **The protein kinase family: conserved features and deduced phylogeny of the catalytic domains.** *Science* 1988, **241**:42-52.
- Hanks SK: **Homology probing: identification of cDNA clones encoding members of the protein-serine kinase family.** *Proc Natl Acad Sci USA* 1987, **84**:388-392.
- Hanks SK, Hunter T: **The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification.** *FASEB J* 1995, **9**:576-596.
- Protein Kinase Resource** [<http://kinases.sdsc.edu/html/index.shtml>]
- Hunter T, Plowman GD: **The protein kinases of budding yeast: six score and more.** *Trends Biochem Sci* 1997, **22**:18-22.
- Plowman GD, Sudarsanam S, Bingham J, Whyte D, Hunter T: **The protein kinases of *Caenorhabditis elegans*: a model for signal transduction in multicellular organisms.** *Proc Natl Acad Sci USA* 1999, **96**:13603-13610.
- Manning G, Plowman GD, Hunter T, Sudarsanam S: **Evolution of protein kinase signaling from yeast to man.** *Trends Biochem Sci* 2002, **27**:514-520.
- Kinase.com** [<http://kinase.com/>]
- Kostich M, English J, Madison V, Gheys F, Wang L, Qiu P, Greene J, Laz TM: **Human members of the eukaryotic protein kinase family.** *Genome Biol* 2002, **3**:research0043.1-12.

10. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S: **The protein kinase complement of the human genome.** *Science* 2002, **298**:1912-1934.
11. Krupa A, Srinivasan, N: **The repertoire of protein kinases encoded in the draft version of the human genome: atypical variations and uncommon domain combinations.** *Genome Biol* 2002, **3**:research0066.1-14.
12. Taylor SS, Radzio-Andzelm E, Hunter T: **How do protein kinases discriminate between serine/threonine and tyrosine? Structural insights from the insulin receptor protein-tyrosine kinase.** *FASEB J* 1995, **9**:1255-1266.
13. Johnson LN, Lowe ED, Noble MEM, Owen DJ: **The structural basis for substrate recognition and control by protein kinases.** *FEBS Lett* 1998, **430**:1-11.
14. Kreegipuu A, Blom N, Brunak S, Järvi J: **Statistical analysis of protein kinase specificity determinants.** *FEBS Lett* 1998, **430**:45-50.
15. Zhu H, Klemic JF, Chang S, Bertone P, Casamayor A, Klemic KG, Smith D, Gerstein M, Reed MA, Snyder M: **Analysis of yeast protein kinases using protein chips.** *Nat Genet* 2000, **26**:283-289.
16. Janji B, Melchior C, Vallar L, Kieffer NL: **Cloning of an isoform of integrin-linked kinase (ILK) that is upregulated in HT-144 melanoma cells following TGF-beta1 stimulation.** *Oncogene* 2000, **19**:3069-3077.

Editor's note: The author has declared that he has no affiliation with SUGEN or Schering-Plough.