Theor Appl Genet (2001) 103:1046–1058

G. Vallad · M. Rivkin · C. Vallejos · P. McClean

Cloning and homology modelling of a Pto-like protein kinase family of common bean (*Phaseolus vulgaris* L.)

Received: 3 January 2001 / Accepted: 23 March 2001

Abstract Degenerate primers, based on conserved subdomains of several plant serine/threonine kinases (STK) similar to the tomato Pto protein kinase, were designed to amplify similar regions from the common bean genome. Sequence analysis of the products defined five distinct classes sharing from 56.9 to 63.9% amino-acid identity with Pto. Inter-class identity ranged from 61.2 to 81.4%. Each of the five classes contain the conserved residues found in subdomains II through IX of most STKs. Multiple sequence and neighbor-joining tree analysis suggest the Pto and the cloned common bean sequences define a unique class of plant protein kinases. Southern hybridization to common bean DNA determined that the sequence classes represent low to moderate copy number families. Using PCR amplification with class-specific primers followed by restriction enzyme digestion of the products, these five classes were found to be essentially monomorphic among 20 divergent common bean genotypes. Each class was determined to be expressed in a leaf mRNA population. Further analysis of the Sg5 class using 3′-RACE (rapid amplification of cDNA ends) identified seven unique family members. All Sg5 3′-RACE products share a high degree of identity, but contain numerous differentiating features that demonstrate the presence of microheterogeneity within the Sg5 class. Three-dimensional homology modelling demonstrated that *Pto* and Sg5–3e contain nearly all of the structural features found in type α cyclic AMP-dependent protein kinase (cAPKα) except α-helices within subdomains II and XI. Based on these homology models and models of ten other plant kinases, two subfamilies of plant protein kinase sequence could be differentiated based on subdomain XI structure. Database searches revealed that subdomains VIa, VIb, VIII and IX of the Pto-like class are unique to plant species, whereas for a second subfamily of plant protein kinases (containing the common bean kinase PvPKI) these subdomains are also similar to those found in non-plant eukaryotic species.

Keywords *Common bean · Phaseolus vulgaris* L. · Protein kinase · *Pto* · Kinase phylogeny

Communicated by M.A. Saghai Maroof

G. Vallad · M. Rivkin · P. McClean (✉)
Department of Plant Sciences, North Dakota State University, Fargo, ND 58105, USA
e-mail: phillip_mcclean@ndsu.nodak.edu

C. Vallejos
Horticultural Sciences Department, University of Florida, Gainesville, FL 32611. USA

Introduction

Protein kinases represent one of the largest family of related proteins in eukaryotic species. Recent genome sequencing efforts identified 118 protein kinase open reading frames (ORF) in yeast and 435 kinase ORFs in nematode (Chervitz et al. 1998). When these values are normalized based on genome size, each species contains an equivalent number of kinase genes. The *Arabidopsis* genome sequence project is also discovering many kinase-like sequences in that model species. Many of these were described and classified by Hardie (1999).

In plants, protein kinases play a crucial role in self-incompatibility systems (Nasrallah et al. 1994), plant hormone activation (Machida et al. 1997), and incompatible plant-pathogen interactions (Dröge-Laser et al. 1997). Of particular interest to the research reported here is the *Pto* gene of tomato (*Lycopersicon esculentum* L.). *Pto* encodes a cytoplasmic serine/threonine protein kinase (STK) (Martin et al. 1993) that interacts with the a*vrPto* gene product of the bacterial pathogen *Pseudomonas syringae* pv *tomato* (Scofield et al. 1996; Tang et al. 1996), an event that appears to launch a cascade of molecular events that triggers the hypersensitive disease-resistance response (Zhou et al. 1997). These experiments provided the first molecular confirmation of Flor's (1956) gene-for-gene hypothesis that predicted a host resistance (*R*) gene encodes a receptor that recognizes a ligand encoded or produced by the corresponding *Avr* gene.

Hybridization of a *Pto* clone to the DNA of a number of plant genotypes and species revealed the presence of numerous *Pto*-like sequences in susceptible and resistant tomato cultivars, as well as common bean and several other dicotyledonous species (Martin et al. 1993). This suggested that *Pto* characterizes a gene family with structural and perhaps functional similarities in various other plant species. PCR amplification is one technique that is appropriate to isolate homologous sequences from other species. For example, homology based cloning of fragments related to the NBS class of disease-resistance genes (Hammond-Kosack and Jones 1997) has been successful for a number of plant species (Kanazin et al. 1996; Leister et al. 1996; Yu et al. 1996; Leister et al. 1998), including common bean (Rivkin et al. 1999).

These results encouraged us to use the same approach to clone Pto homologs from common bean. This experimental approach would also allow us to test whether cloning with degenerative primers would preferentially select *Pto*-like sequences from among the many other protein kinase genes found in eukaryotic genomes (Hanks and Hunter 1995). Once the sequences were cloned we could determine the relatedness of these sequences to other eukaryotic protein kinases. Therefore, our specific objectives were to clone and classify *Pto*-like sequences from common bean, determine the distribution of these sequences among common bean genotypes, analyze the expression of the sequences, and compare the predicted protein structure of these *Pto*-like sequences with the structure of other protein kinases.

Materials and methods

DNA extraction

DNA was isolated from leaf tissue of common bean cultivars of Mesoamerican origin ('Aurora', 'Beryl', 'Domino', 'GN Harris', 'GN UI59', 'ICA-Bunsi', 'Mayflower', 'Olathe', 'Othello', 'Seafarer', 'Sierra', 'Sutter Pink', 'Topaz', 'UI114', and 'Viva') and Andean origin ['California Dark Red Kidney 82' (CDRK82), 'California Early Light Red Kidney' (CELRK), 'Isabella', 'Linden', and 'Montcalm'] by the method described by Doyle and Doyle (1990).

PCR amplification

Primers P3 and P2R (Table 1) were based on conserved regions of amino-acid identity between the STKs *Pto* (Martin et al. 1993), *Fen* (Loh and Martin 1995) and *Pti1* (Zhou et al. 1995) of tomato, and *MHK* (Moran and Walker 1993) and *APK1* (Hirayama and Oka 1992) of *Arabidopsis*. P3 is specific for STK subdomain II, and P2R is specific for subdomain IX. Reactions were carried out in 10 µl containing 10 µM of each primer, 30 ηg of genomic DNA (Sierra or Olathe), 1× PCR buffer (GibcoBRL), 1.5 mM $MgCl_2$, 0.25 mM of each dNTP, and 2.0 units of *Taq* polymerase. DNA was denatured at 94°C for 3 min and amplified for 35 cycles using the following parameters: 94°C for 1 min, 60°C for 1 min; and 72°C for 2 min.

Amplification was followed by a final extension of 72°C for 5 min. All amplification steps were carried out in a Perkin-Elmer/Cetus model-490 thermal cycler.

Cloning, sequencing, and analysis of PCR products

Amplified fragments were extracted from the agarose matrix using the GeneClean Kit (Bio 101, Inc., La Jolla, Calif.) and cloned into the pPCR cloning vector (Borovkov and Rivkin 1997). Clones were sequenced using the SequiTherm Cycle Sequencing Kit (Epicentre Technologies, Madison, Wis.), and the products were separated with a Model 4000 Automated DNA Sequencer (LI-COR, Inc., Lincoln, Neb.). DNA sequences and the derived amino-acid sequences were analyzed using the online Molecular Biology Workbench software package version 2.0 (http://bioweb.ncsa.uiuc.edu). Sequences were aligned using the ClustalX interface (Thompson et al. 1997) to the CLUSTAL W program (Thompson et al. 1994). ClustalX was also used to generate a rooted-bootstrapped gene tree using the neighbor-joining method (Saitou and Nei 1987). The bootstrap tree was based on 1000 replicates, and the *Drosophila* protein kinase Pelle was as the outgroup.

Cleaved amplified polymorphic sequence (CAPS) analysis

DNA (20 ηg) from common bean genotypes was amplified using the class-specific cDNA amplification conditions described above. Amplified products were then digested with restriction endonucleases, whose recognition sequences were previously determined by sequence analysis to be present in a specific *Pto*-like class. These enzymes were: *Hinf*I, *Pst*I and *Rsa*I (Og9); *Msp*I, *Pst*I and *Rsa*I (Sg1 and Og8); *Bgl*II, *Hae*III and *Rsa*I (Sg2 and Og5); and *Hin*fI, *Pvu*II, and *Xcm*I (Sg5).

Table 1 Primers used in DNA and cDNA amplification of Pto-like fragments of common bean. Primers are listed in a in 5′→3′ direction from left to right. Class-specific primers which include a 5′ or 3′ suffix indicated the region to which the primer is complementary

Primer	Sequence[a]
Pto-based primers	
P3 (Pto residues 47–53)	TNGGNSANGGNGKNTTYGG
P2R (Pto residues 224–230)	ACNCCRAANGARTANACRTC
Class-specific primers	
Og9–5′	GCCGTGGAAGCAAAGGGTCG
Og9–3′	CACCTGAGTGTTCACATGACTCATCG
Sg1–5′	CGGTCGCGAGTTCCTGAACG
Sg1–3′	GGGCCTTCGAAGTTCCAATTGG
Og8–5′	GTGCGCACGAGTTCATGAACG
Og8–3′	GGTACTCACGTGGGACTTGTCCATG
Sg5–5′	CACCAGTCGTGACATACGATTCAGTACC
Sg5–3′	CCAGGGTCAGGTCAAGGCCTTC
Og5–5′	CGCAATCAGAACAAGGAGTGAACG
Og5–3′	GCTACCCTTCACCACAGTACTAACATGACC

[a] R=A or G; W=A or T; Y=C or T; S=C or G; K=T or G; N=A,T,C,G

DNA hybridizations

Two micrograms of genomic DNA were digested for 4 to 6 h at 37°C with 20 units of a restriction enzyme. Electrophoresis, Southern transfer and subsequent hybridization followed the conditions described for Sierra and Olathe DNA in Rivkin et al. (1999). The final wash was 0.1× SSC, 0.005% SDS, 65°C. Restriction digestion, electrophoresis, Southern transfer and subsequent hybridization of Calima and Jamapa were as described by Vallejos et al. (1992). The final wash was with 0.5× SSPE, 0.1% SDS, 65°C.

cDNA preparation and amplification

cDNA was prepared from 2-weeks old, growth-chamber grown leaves, as described by Rivkin et al. (1999). Double-stranded cDNA was prepared using the Marathon cDNA Amplification Kit according to the manufacture's specifications (CLONTECH Laboratories). cDNA amplifications were carried out using a set of class-specific primers (CSP) (Table 1) derived from the sequences conserved within all members of a specific *Pto*-like class. The class-specific amplification conditions were identical to the initial DNA amplification except that 2 µM of each primer was used, annealing was carried out at 65°C, and 1 ml of a 250-fold diluted cDNA solution was substituted for the DNA. RACE (rapid amplification of cDNA ends) reactions were carried out in 50-µl volumes using protocols outlined in the Advantage cDNA PCR Kit and Marathon cDNA Amplification Kit (CLONTECH Laboratories). RACE amplification conditions consisted of an initial denaturation at 94°C for 1 min, followed by five cycles of 94°C for 30 s and 72°C for 10 min; five cycles of 94°C for 30 s and 70°C for 10 min, and 35 cycles of 94°C for 20 s and 68°C for 10 min.

Homology modelling

Secondary and tertiary structures were determined using the SWISS-MODEL program (Guex and Peitsch 1997; http://www.expasy.ch/swissmod/SWISS-MODEL.html). The SWISS-MODEL program first utilizes a homology based BLAST search (Altschul et al. 1997) of the Protein Data Bank (http://www.rcsb.org) to identify proteins with sequence similarity to the sequence to be modelled. The five sequences with identity greater than 25% to the query sequence and having a projected model size larger than 20 residues are compared against the query sequence by use of the SIM program (Huang and Miller 1991) to create an alignment that serves as a guide during model building. The ProModII program (Peitsch 1996) creates the consensus three-dimensional structure. The model is then processed by the Gromos96 program (http://igc.ethz.ch/gromos/) which minimizes the energy required to fold the protein into a final conformation. We used the fully automated First Approach Mode option. The quality of the model was assessed by calculating the root mean square deviation (RMSd) between the target sequence and each protein used in the modelling of that sequence. The RMSd value was calculated using the following steps in the Swiss-PdbViewer software (Guex and Peitsch 1997): (1) the appropriate .pdb files containing the three-dimensional modelling information are loaded, and the target sequence is made the active layer; (2) the two models are super-imposed using the Magic Fit and Improve Fit tools; and (3) the RMSd value for the carbon-α backbone is calculated using the Calculate RMS tool.

Results

Isolation and characterization of common bean Pto-like sequences

Amplification of Sierra and Olathe genomic DNA with degenerate primers P3 and P2R generated fragments estimated as 550-bp, 900 bp and 950 bp. Because the 550 bp fragment corresponded to the expected size of the amplification product (assuming the absence of introns), it was isolated for further analysis. This fragment was cloned, and 18 colonies with identical insert size were sequenced. Clones were labeled as "Sg" or "Og" to designate products amplified from *S*ierra or *O*lathe *g*enomic DNA, respectively.

Percent amino-acid identity among each clone pair was determined, and clones with greater than 85% identity were grouped into a single class. The five classes, Og5, Og8, Og9, Sg1 and Sg5 consisted of one, five, four, six and two clones, respectively. Og5 is the lone member of its class and shares less than 70% identity to any other sequence isolated (Table 2). Amino-acid identity among the classes ranged from 61.1% (Sg5 vs Sg1) to 81.4% (Og8 vs Sg1) (Table 2).

Alignment and identification of a *Pto*-like class of protein kinases

A sequence representing each of the five common bean sequence classes was aligned with kinase subdomains II–IX of the tomato genes *Pto*, *Fen* and *Pti1*, and the kinase region of the rice resistance gene *Xa21* (Song et al. 1995). The alignment identified specific residues in the common bean sequences that correspond to the conserved residues found in all STKs (Fig. 1) (Hanks and Quinn 1991). Additionally, all seven subdomains expected for a STK fragment derived from amplification with the selected primers were present in each common bean class.

A BLAST search (Altschul et al. 1997) of the NCBI non-redundant nucleotide and peptide sequence databases (http://www.ncbi.nlm.nih.gov) for sequences with significant homology to each common bean class was performed. Both Pto and Fen gave high comparison scores (Pto: expect value $<9^{-48}$; Fen: expect value $<5^{-46}$). Among all sequences, classes Og5 and Sg5 were most similar to Pto (expect values=7^{-60} and 2^{-52}, respectively).

High amino-acid identity was observed between Pto and Og5 (63.9%) and Fen and Sg1 (55.4%) (Table 2). By comparison, the identities between the common bean STK sequences and 17 other STKs were lower with values ranging from 49.7% (TMK1 vs Sg5) to 19.4% (ASK2 vs Og5). Pto and Fen showed a similar low level of identity to these other STKs as well: 45.7% (Pro25 vs Pto) to 22.2% (cAPKα vs Pto).

The five common bean sequences, the 87 *Arabidopsis* kinase sequences characterized by Hardie (1999), and 13 other plant kinase sequences found in GenBank were

Table 2 Percent amino acid identity between common bean *Pto*-like sequences and the corresponding kinase subdomains II–IX of 19 other plant serine/threonine kinases[a]

Kinase	Kinase																						
	Sg1	Og9	Og5	Sg5	Pto	Fen	APK1	PRO25	TMK1	Pti1	AtPK1	ZmPPK	AtPK5	OsG11A	PVPK1	PsPK5	ASK1	ASK2	cAPKα	ZMPK1	SRK6	Xa21	TMKL1
Og8	81.4	79.7	68.8	64.7	61.7	59.0	47.6	44.3	45.0	41.4	23.3	24.0	24.0	22.8	25.1	24.7	21.8	22.4	21.4	34.1	38.5	32.0	28.0
Sg1		75.6	63.5	61.2	56.9	55.4	46.5	47.3	45.6	45.0	25.2	24.0	24.0	22.8	25.1	24.1	22.4	22.4	23.9	34.7	37.9	32.5	32.1
Og9			67.6	63.5	59.9	56.6	46.5	46.1	46.7	42.6	23.3	25.1	23.4	23.4	25.1	23.5	20.5	21.2	21.4	36.5	42.6	37.9	31.0
Og5				66.5	63.9	62.3	44.7	42.5	45.6	40.9	24.7	25.1	24.6	23.4	25.7	24.1	21.9	19.4	21.3	36.5	38.5	30.2	29.6
Sg5					61.1	56.6	43.5	47.3	49.7	42.6	24.8	24.8	24.8	23.0	24.8	23.8	22.7	22.7	24.1	40.6	41.9	34.7	28.7
Pto						85.0	42.5	45.7	42.9	41.7	23.7	23.6	23.6	22.4	23.6	26.8	24.2	24.2	22.2	43.3	44.0	30.1	31.9
Fen							43.0	43.8	42.2	43.4	24.4	23.9	23.9	22.7	24.5	27.8	22.9	23.5	24.4	44.4	43.3	31.1	31.1
APK1								45.6	45.0	42.0	25.2	23.8	23.8	23.8	25.0	25.1	19.1	21.7	27.3	42.5	36.7	35.5	28.6
PRO25									43.5	42.8	25.5	22.9	22.3	21.7	22.3	23.6	21.3	23.9	25.0	40.6	41.1	33.9	30.5
TMK1										42.1	29.6	28.0	26.8	27.4	28.6	29.3	24.4	25.0	27.2	42.2	41.1	35.3	30.2
Pti1											23.6	26.5	26.5	25.3	26.5	27.3	22.7	23.4	23.9	38.9	41.7	32.0	33.1
AtPK1												41.6	41.6	41.6	41.0	45.2	31.7	31.7	40.8	25.3	23.3	23.6	22.2
ZmPPk													85.3	85.3	79.4	57.5	25.3	26.5	39.6	21.7	23.4	25.1	19.9
AtPK5														84.8	82.1	56.0	26.5	27.8	39.6	20.5	22.2	26.3	19.3
OsG11 A															78.5	55.4	25.3	26.5	40.9	21.1	21.6	24.6	20.5
PVPK1																57.5	25.9	27.2	40.2	22.3	22.8	25.7	20.5
PsPK5																	28.4	27.8	38.0	25.5	22.9	24.7	23.6
ASK1																		86.7	28.6	26.5	22.4	22.2	16.1
ASK2																			31.2	27.1	23.1	23.4	19.8
APK-Cα																				27.7	21.9	21.3	21.3
ZMPK1																					42.9	36.7	31.5
SRK6																						32.9	30.8
Xa21																							31.4

[a] For the common bean Pto-like sequences the GenBank nucleotide accession numbers are in parenthesis: Og8 (AF363818), Sg1 (AF363810), Og9 (AF363815), Og5 (AF363819) and Sg5 (AF363807). GenBank protein accession numbers are in parenthesis following the names of the kinase proteins: Pto (626010), Fen (557882), APK1 (1168470), PRO25 (423988), TMK1 (1174718), Pti1 (1155217), AtPK1 (1170689), ZmPPK (320662), AtPK5 (421843), OsG11 A (1346057), PVPK1 (125568), PsPK5 (556347), ASK1 (1168529), ASK2 (1168530), APK-Cα (125205), ZMPK1 (1346396), SRK6 (730831), Xa21 (2130082) and TMKL1 (313190)

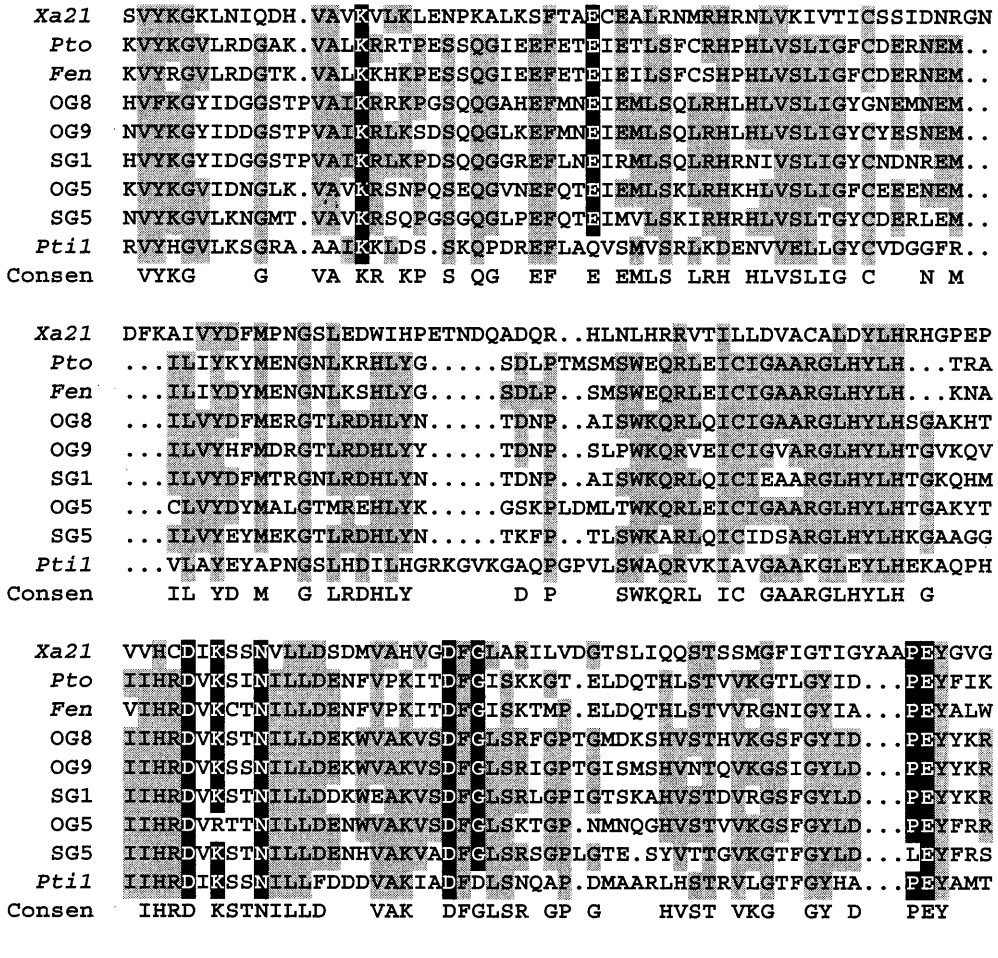

Fig. 1 Amino-acid alignment between common bean Pto-like sequences and Pto, Fen and Pti1 of tomato, and Xa21 of rice. "Consen" is the consensus sequence. Those invariant residues found in all protein kinases (Hanks and Quinn 1991) are highlighted in *black*. Residues found in >50% of sequences analyzed are highlighted in *gray*. *Dots* represent gaps introduced in a sequence for alignment purposes. Only those regions homologous to the common bean sequences are shown for the other genes

analyzed. Prior to detailed analysis, the 105 amino-acid sequences were aligned, and the non-common bean sequences were all trimmed to include only kinase subdomains II through IX. In addition, the insert between subdomains VII and VIII common to sequences similar to the *P. vulgaris* kinase PvPK1 class (Hardie 1999; equal to class AGC-VIII, Hanks and Hunter 1995) were also trimmed. The trimmed sequences were aligned, and a rooted, bootstrapped neighbor-joining gene tree was constructed (Fig. 2).

Alignment and cluster analysis with complete sequences are preferred to comparisons with partial sequence data. Therefore, we were concerned that the trimmed sequences may not cluster in the same manner as the full sequences. Our results suggest this concern was unfounded. The gene tree we developed using the trimmed sequences had a one-to-one correspondence with 15 of the 16 clusters discovered by Hardie (1999). Each of the these clusters was supported by high bootstrap values (>804). Only the receptor-like kinase cluster described by Hardie (1999) was not supported by our analysis. The results suggests that subdomains II through IX contain sufficient sequence data to represent clusters defined by analyses with full sequence data.

Only four of the non-*Arabidopsis* sequences included in this analysis clustered with *Arabidopsis* kinases. The pea kinase PsPK5 formed a highly supported two-member cluster with NPH1. Additionally, common bean kinase PvPK1, rice kinase OsG11, and maize kinase ZMPKK formed a highly supported cluster with *Arabidopsis* kinases PK64 and PK5. These three non-*Arabdidopsis* kinases are unique because they contain an insert between subdomains VII and VIII. The two *Arabidopsis* sequences also contain inserts (68 and 67 residues, respectively) at the same location. The fact that these insert sequences were removed prior to alignment and tree building suggest that sequences other than those define the kinase members of this cluster.

A significant result of our research is the observation that the five cloned common bean sequences do not cluster with any of the 16 groups defined by Hardie. Rather these sequences along with tomato kinases Pto and Fen,

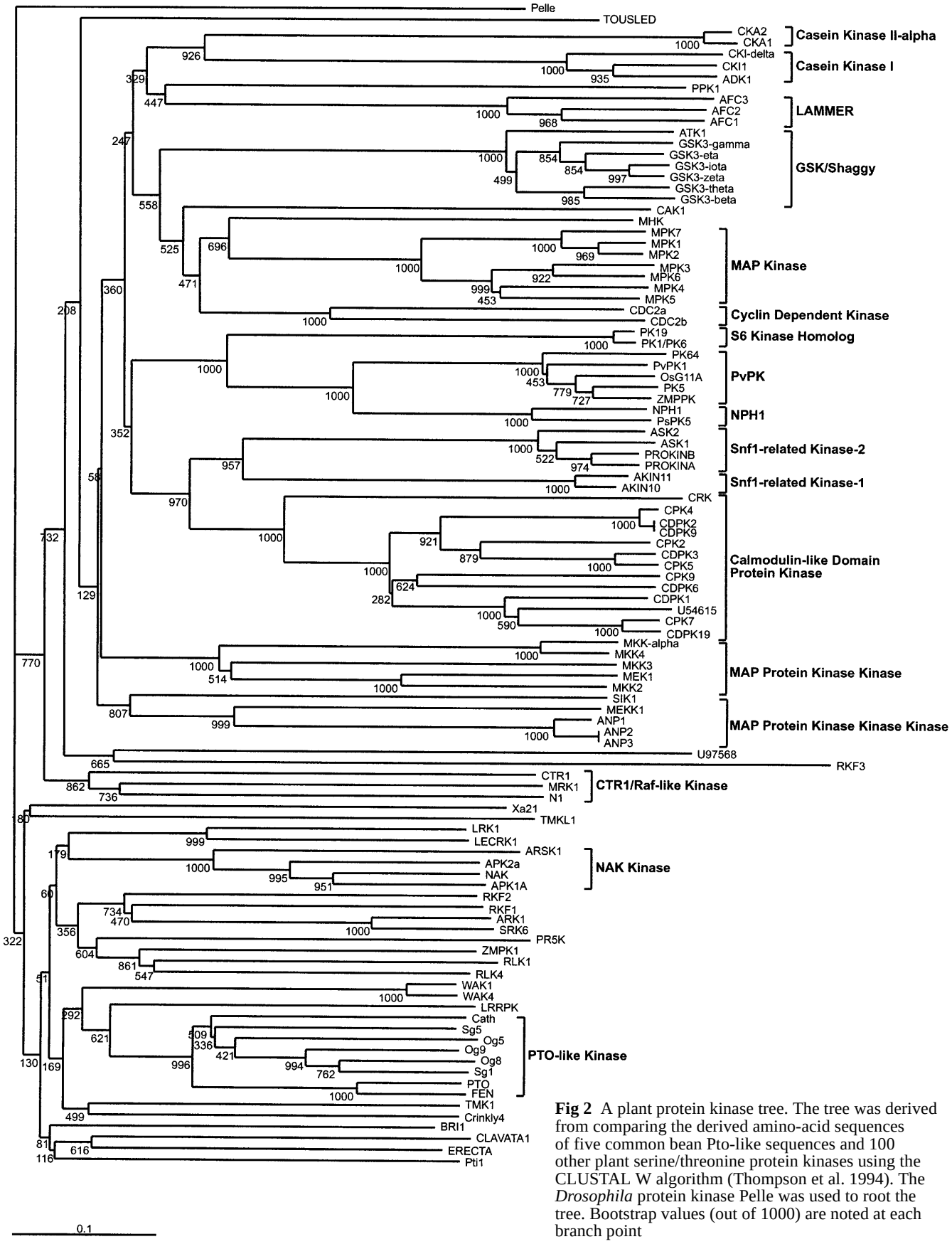

Fig 2 A plant protein kinase tree. The tree was derived from comparing the derived amino-acid sequences of five common bean Pto-like sequences and 100 other plant serine/threonine protein kinases using the CLUSTAL W algorithm (Thompson et al. 1994). The *Drosophila* protein kinase Pelle was used to root the tree. Bootstrap values (out of 1000) are noted at each branch point

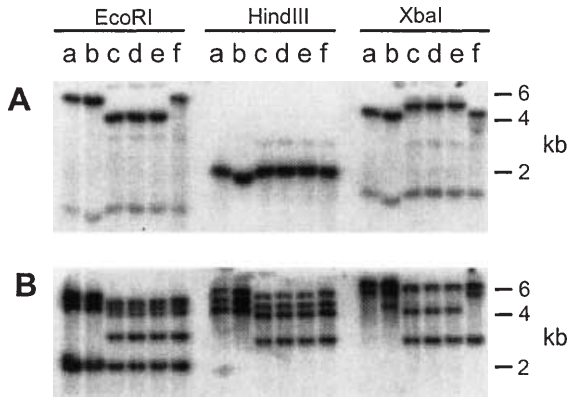

Fig. 3 RFLP hybridization patterns for *Pto*-like class probes Og5 (*A*) and Sg5 (*B*). Genotypes are Sierra (*a*), Olathe (*b*), Montcalm (*c*), California Early Light Red Kidney (*d*), California Dark Red Kidney 82 (*e*) and ICA-Bunsi (*f*)

and periwinkle kinase CrRLK1 form a highly supported cluster. We have termed this group of kinase sequences the Pto-like kinase cluster.

Identification of polymorphic Pto-like sequences in common bean

CAPS analysis was utilized to estimate the diversity of Pto-like sequences among 20 common bean genotypes. These genotypes were selected because they represent diverse genotypes based on pedigree analysis (McClean et al. 1993). For all primer pairs except Og9, the restriction pattern for a given amplification product was monomorphic among all genotypes. Digestion of Og9 products with *Hinf*I, *Pst*I and *Rsa*I generated polymorphic products unique to the kidney cultivars CELRK, CDRK82 and Isabella.

Southern hybridization, with probes derived by amplifying cloned fragments with class-specific primers, identified numerous polymorphisms among the genotypes Sierra, Olathe, Montcalm, CELRK, CDRK and ICA-Bunsi. Og5 was the only probe that hybridized strongly to a single fragment (Fig. 3A). Sg1, Sg5 (Fig. 3B), Og8 and Og9 probes exhibited unique hybridization patterns indicative of moderately sized multigene families. The hybridization of the Sg1 probe to the genotypes Calima and Jamapa produced a complex pattern indicative of a relatively high copy number, multigene family (data not shown).

Expression of Pto-like sequences in common bean

To characterize the expression of these sequences, cDNA from 2-week old bean leaves was used as a substrate for PCR amplification with class-specific primers (Table 1). This cDNA sample was previously shown to be free of contaminating DNA (Rivkin et al. 1999). The fact that each reaction produced a product of the appropriate size demonstrates that each common bean Pto-like class is represented in the transcript population of the source tissue (Fig. 4).

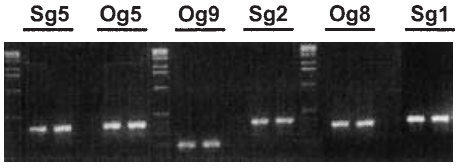

Fig. 4 Amplification of common bean DNA and cDNA with primers specific to each common bean Pto-like class. Each pair of lanes consists of DNA (left) and cDNA (right) amplification with primers to the class noted above the lanes

To further define the nature of these *Pto*-like transcripts, we chose to study the Sg5 class in more detail because it has a high degree of identity with Pto (Table 2; Fig. 2) and is the only class with the specific threonine residue found to be essential for the Pto/avrPto interaction (Frederick et al. 1998). 3′-RACE products of Sg5-class transcripts were amplified using the Sg5–5′ class-specific primer (Table 1) in conjunction with the linker-specific primer, Ap1. This procedure yielded a 1.3-kb fragment that was cloned, and sequence analysis revealed several unique products. Southern analysis with clone Sg5–3e gave the same hybridization pattern as the Sg5 clone (data not shown).

Seven Sg5 3′-RACE clones were placed in one of two main groups based on insertions-deletions in the 3′-untranslated region (3′-UTR). Members of one group (Sg5–3c, Sg5–3f, Sg5–3g and Sg5–3h) contain identical deletions of 1, 3, 6, 15, 32 and 61 nucleotides within the 3′-UTR. Sg5–3g is further distinguished by one 5-nucleotide deletion that introduces a stop codon that would result in a truncated protein 80 amino acids shorter than the other products. An in-frame six-nucleotide deletion in Sg5–3h results in a product two amino acids shorter than the 312 amino-acid product predicted for the other Sg5 3′-RACE clones. The second group of sequences (Sg5–3b, Sg5–3d and Sg5–3e) lack these large 3′-UTR deletions and contain two unique UTR deletions of 1 and 3 nucleotides.

The derived protein sequence for each Sg5 3′-RACE clone (Fig. 5) shows the presence of specific residues conserved in each of the 12 subdomains of typical STKs (Hanks and Hunter 1995). The sequence for all Sg5 class members except Sg5–3g extend 62 amino acids past the carboxy terminal of Pto. Due to the premature stop codon described above, Sg5–3g terminates 19 amino acids before the carboxy terminal of Pto.

A comparison of the derived amino-acid sequences of the Sg5 3′-RACE sequences identified 42 amino-acid substitutions, excluding the truncated positions in Sg5–3g. Ten of these are conservative. Of these 42 substitutions, 14 are shared among two or more sequences, and 28 are unique to a single sequence. Seven of the unique amino-acid substitutions are related to the nucleotide deletions that occurred in Sg5–3g and Sg5–3h and are a result of two and three nucleotide changes within specific codon positions. Six other unique amino-acid substitutions are also the result of two and three nucleotide changes. The remaining 15 unique substitutions are

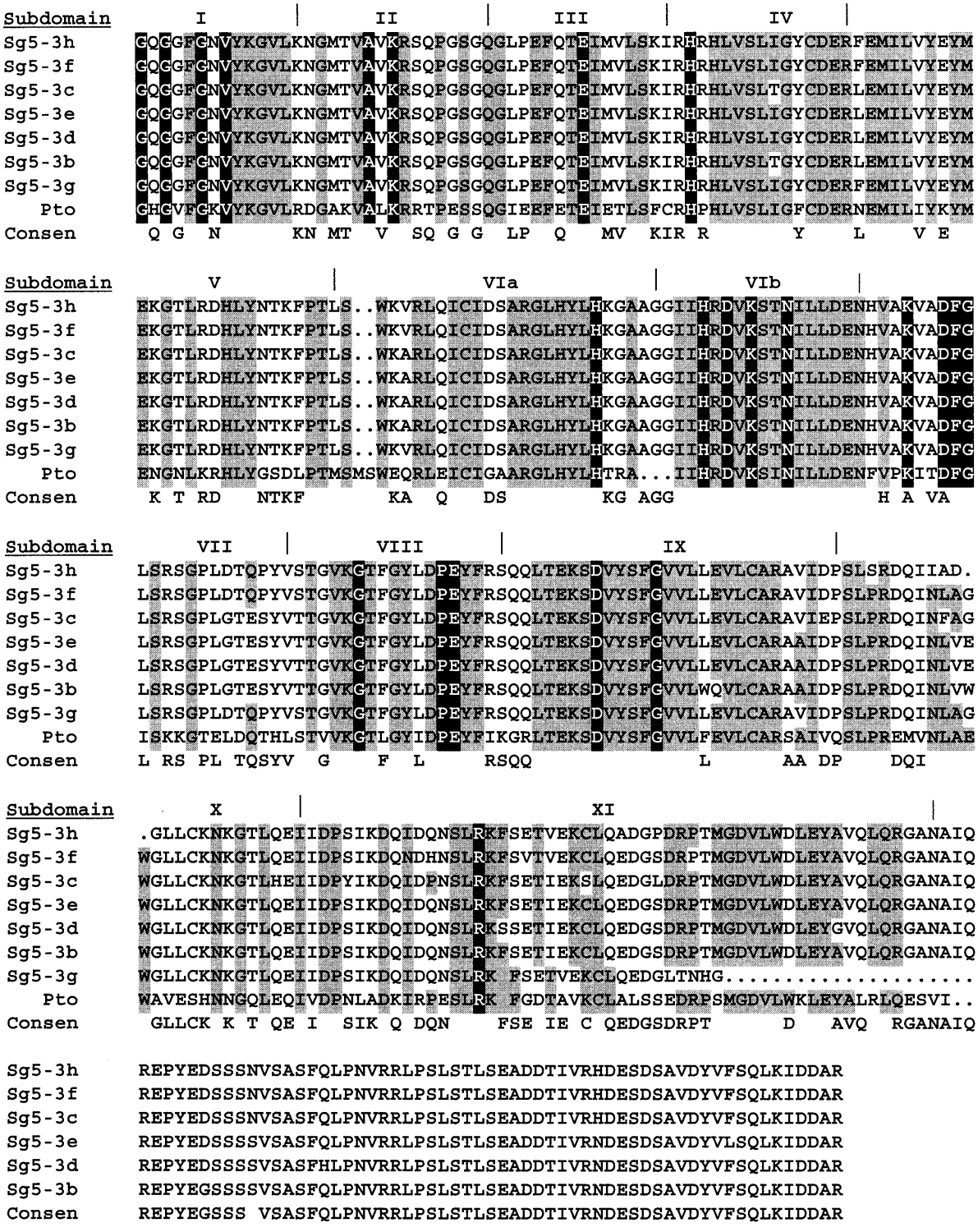

Fig. 5 Amino-acid alignment of Sg5 class members and Pto of tomato. *Black highlight* denotes residues common to most STKs (Hanks and Hunter, 1995) and *gray highlight* denotes amino acids identical between Pto and Sg5 clones. *Dots* represent gaps introduced in a sequence for alignment purposes. "Consen" is the consenesus sequence. Roman numerals identify the STK subdomains. The GenBank nucleotide accessions numbers for each member of the Sg5 class are in parenthesis: Sg5–3b (AF363828), Sg5–3c (AF363829), Sg5–3d (AF363827), Sg5–3e (AF363826), Sg5–3f (AF363825), Sg5–3g (AG363824) and Sg5–3h (AF363823)

due to single nucleotide changes within a codon position. All 14 of the shared amino-acid substitutions are due to identical, single-nucleotide changes within a codon position. Of the 139 nucleotide substitutions that occurred over all analyzed sequences (excluding gaps), 73.4% were shared across three or more sequences, and 23.0% were unique to any single sequence. Taken together, these results suggest that each Sg5 3′-RACE clone is unique and not an amplification artifact.

Table 3 Sg5–3e and Pto homology modelling statistics

Comparison	Statistic			
	P(N)[a]	%ID[b]	# Insertions[c]	RMSd[d]
SG5–3e vs				
Tyrosine kinase domain (chain A) of fibroblast growth factor receptor complexed with SU5402 (1FGI)[e]	2.4×10^{-22}	29.6	5 (11)	0.49
Tyrosine kinase domain (chain A) of fibroblast growth factor receptor complexed with SU4984 (1AGW)	5.6×10^{-21}	29.6	5 (11)	0.55
Tyrosine kinase domain (chain B) of fibroblast growth factor receptor complexed with SU4984 (1AGW)	1.2×10^{-20}	32.1	8 (24)	0.55
Human tyrosine kinase c-src (1FMK)	1.1×10^{-19}	40.6	11 (42)	1.17
Human lymphocyte kinase (activated-form) (3LCK)	1.8×10^{-19}	33.2	11 (28)	1.16
Pto vs				
Tyrosine kinase domain (chain B) of fibroblast growth factor receptor complexed with SU4984 (1AGW)	3.4×10^{-21}	23.0	9 (26)	1.47
Human lymphocyte kinase (activated-form) (3LCK)	1.5×10^{-18}	33.4	6 (9)	0.91
Human tyrosine kinase c-src (1FMK)	7.2×10^{-18}	30.5	8 (28)	1.37
Human insulin receptor tyrosine kinase domain (1IRK)	5.5×10^{-17}	28.5	9 (17)	1.08
Activated insulin tyrosine kinase in complex with peptide substrate and ATP analog (1IR3)	7.1×10^{-16}	28.1	5 (17)	0.87

[a] Blast-search-generated Poisson unlikelihood probability
[b] Percent identity between the two proteins
[c] Number of insertions created during alignment; the number of amino acids represented by the insertions are noted in parentheses
[d] Root mean square deviation
[e] The Protein Data Base accession number is given in parenthesis

Homology modelling of Sg5–3e, Pto, and other plant protein kinases

Homology models of the three-dimensional structure of Pto and Sg5–3e were developed using the SWISS-MODEL program. Three of the five sequences used to model the two sequences were in common (Table 3). Sequence identity between the target sequence and the sequences upon which the models were built ranged from 29.6 to 40.6% for Sg5–3e and 23.0 to 33.4% for Pto. The observation that the root mean square deviations (RMSd) were low (Table 3) given this level of sequence identity (Chothia and Lesk 1986) suggests these are reliable models.

The two structural models were compared to the catalytic subunit of the type α cyclic AMP-dependent protein kinase (cAPKα), the first protein kinase for which a crystal structure was determined (Knighton et al. 1991). The structure of this protein is typically used when comparing protein kinase structures (Hanks and Hunter 1995). Thirteen of the seventeen cAPKα secondary structures were identified in both Pto and Sg5e (Fig. 6). The subdomain I β-strand 1 was present in Pto, but missing in Sg5–3e. This is a reflection of the fact that Sg5–3e begins seven amino acids downstream from beginning of this domain. Sg5–3e is also missing β-strand 9 located in subdomain VII. Additionally, Sg5–3e contains an α-helix in subdomain VIII not found in either the Pto or cAPKα structures. The major differences between Sg5–3e and Pto and cAPKα are the absence of the cAPKα-specific, α-helix B in subdomain II and α-helices H and I in subdomain XI.

Hanks and Hunter (1995) classified all but two of the plant protein kinases known at the time into two subfamilies, OPK-X and AGC-VIII. Pto was placed in the OPK-X group. To compare the models of Sg5–3e and Pto with other members of this subfamily, homology models were built for APK1, PRO25 and TMKl1 of *Arabidopsis thailana*, SRK6 of *Brassica oleracea*, and ZMPK1 of *Zea mays*. In each case, the RMSd values (range: 0.63–1.47) comparing the target structure and the structure of the five proteins suggested these were reliable homology models. As with Sg5–3e and Pto, nearly all of the major structures found in cAPKα were present in these models (data not shown). As with Sg5–3e and Pto, the notable differences were the absence of the α-helix B in subdomain II and α-helices H and I in subdomain XI.

PvPK1 (Lawton et al. 1989), a common bean protein kinase, is a member of the second plant kinase subfamily, AGC-VIII. Other members include OsG11A (*Oryzae sativa*), ZmPKK (*Zea mays*), AtPK5 (*A. thaliana*), and PsPK5 (*Pisum sativum*). The identity between these kinases and the five classes of common bean Pto-like sequences described here is low (range: 23.5–25.7%; Table 2). In contrast, the identity among these five kinases is high (range: 55.4–85.3%; Table 2). As discussed above, our analysis showed these kinases, along with AtPk64, also form a well-supported cluster (Fig. 3).

Because of their divergence with the Pto-like sequences, we built homology models of these proteins to compare them with the models of Sg5–3e, Pto and cAPKα. For all five models, the range of RMSd values (0.46–1.14) suggest these models are reliable. The five

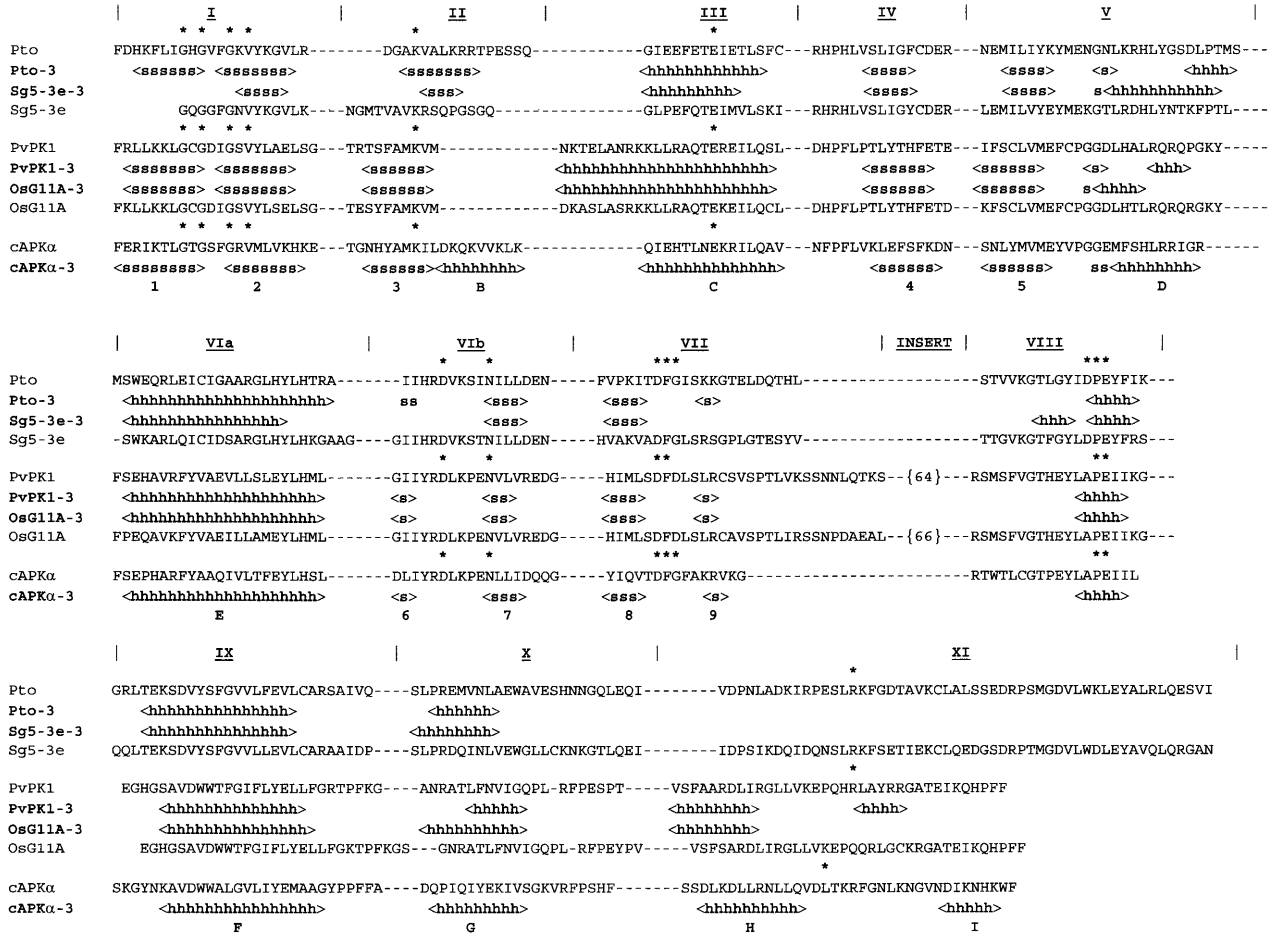

Fig. 6 Homology based structural models of Sg5–3e, and protein kinases Pto, PvPKI and OsG11a. "*" represent the conserved amino-acid residues found in STKs. "INSERT" refers to the 64 and 66 amino-acid inserts between subdomains VII and VIII of PvPK1 and OsG11A, respectively. The structure symbols are: *s*=β-sheet; *h*=α-helix. A *line* abbreviated by the protein designation followed by "–3" (Pto-3, for example) delineates the structural features identified by homology modelling. The *letters and numbers* below cAPKα are the designations for each structure

models were quite similar, and two of these (PvPK1 and OsG11 A) are shown in Fig. 6. These models are also very similar to cAPKα. The only structure missing is cAPKα-specific, α-helix B in subdomain II. Unlike Sg5–3e and Pto, subdomain XI α-helices H and I are present in these models.

Discussion

Genomic sequencing efforts demonstrate that protein kinases are one of the largest gene families found within many model species (Chervitz et al. 1998). To-date, the NCBI nr database contains 5271 kinase sequences (http://www.sdsc.edu/kinases/pkr/pk_catalytic/pk_cat_list.html). A system was designed to uncover structural and functional features of kinases from unrelated species and to order these into subfamilies (Hanks and Hunter 1995). This system groups kinases based on amino-acid similarity in the catalytic domains of the protein and has proven useful because it identifies protein kinases with similar three-dimensional structure, substrate specificity and cellular regulation. Fifty seven subfamilies were identified, and plant protein kinases were placed in two subfamilies. One subfamily (AGC-VIII) contains the common bean kinase PvPKI (Lawton et al. 1989) and are related to cyclic nucelotide-regulated protein kinases. The second subfamily (OPK-X), containing the tomato Pto kinase, are not structurally related to other kinases. More recently, Hardie (1999) classified *Arabidopsis* kinase sequences into 16 clusters.

The ubiquitous presence of protein kinases in eukaryotes make them good candidates for PCR-based cloning. The shared subdomain structure of protein kinases, though, presents the significant challenge of selectively cloning members for a specific kinase group. Therefore, it is important to carefully select primers that are unique to one, but not all, kinase groups. To determine its feasibility, we decided to concentrate on cloning common bean orthologs to the tomato Pto kinase because it represents the only disease resistance gene that encodes a sole kinase protein product. Several lines of evidence suggest we were successful. First, we identified five classes of related sequences that exhibit a higher degree of amino-

acid identity to Pto. These sequences are also similar to Fen, a tomato kinase, closely related to Pto. This identity is much greater than for other plant kinases and in some cases the identity with other plant protein kinaes is very low.

Secondly, cluster analysis defined a well-supported gene family that consists of the common bean sequences along with Pto and Fen. This analysis included a large array of *Arabidopsis* seqeunces, along with several other plant kinases, and supported those kinase groups defined by Hardie (1999). Given that those other groups represent kinases with many different cellular and regulatory functions, and the sequences we cloned, along with Pto and Fen, were not similar to any of those groups, the kinase domains within this cluster may be related by function. Clearly, the cloning and sequence analysis of full-length common bean homologs would provide support for this hypothesis.

Homology modelling provided further support for our cloning approach. The models of Sg5–3e and Pto are very similar, but lack several key structures found in cAPKα. These two models are also similar to those kinases determined by Hanks and Hunter (1995) to be related to Pto. In contrast, these models differ from those of the AGC-VIII subfamily [PvPK gene family of Hardie (1999)] of plant kinases by the absence of subdomain XI α-helices.

At the individual class level, PCR/RFLP analysis detected polymorphisms for only the Og9 class among 3 of 20 common bean genotypes surveyed. Because this analysis consisted of genotypes representing Andean and Mesoamerican origin, it appears that each Pto-like class is conserved in the common bean gene pool. The isolation of seven unique members of the Sg5 class, though, demonstrates that microheterogeneity within a class does exist. This is consistent with the hybridization data showing Sg5 to be a moderately sized multigene family.

The complex hybridization patterns did not allow us to determine the map location for Sg5 and the other kinase-like sequences. The fact that the number of Sg5-hybridizing fragments is less than the number of unique family members suggests that some class members may be clustered in the common bean genome. Our database searches discovered that nine sequences with 60–68% identity (73–81% similar) to the Sg5 class member Sg5–3e are found on four of the five *Arabidopsis* chromosomes, and four of the copies are located at genetically distinct locations of chromosome 5. Further experiments are necessary to determine if a similar genomic distribution of this class is observed in common bean.

The observation that highly similar sequences are found in other plant species prompted us to perform a sequence relatedness analysis of these Pto-like sequences from common bean and other eukaryotic kinases. In particular we were interested in determining if any particular subdomain(s) was (were) highly conserved, and if so what patterns could be recognized. Our modelling experiments provided important structural data necessary for the analysis. We initially used each Sg5–3e and PvPKI subdomain sequence as a query against the GenBank database using the Basic Blast search (tblastn) at http://www.ncbi.nlm.nih.gov/BLAST/ (search criteria: expect value <10; hits =500.) Sequences with >90% identity were found for each subdomain except Sg5–3e I, II and III and PvPKI II. For these subdomains no matches were reported.

For subdomains Sg5–3e VIb and IX, and PvPKI VIa, VIb and IX, greater than 100 matches were detected. Because these domains participate in such important kinase functions as support (VIa), catalysis (VIb) and peptide recognition (IX) (Hanks and Hunter 1995), we next searched the database using the combined SG5–3e and PvPKI sequences for subdomains VIa and VIb, and VIII and IX. Subdomain VIII was included because it plays a major role in the recognition of peptide substrates (Hanks and Hunter 1995). As an example, Thr-204, found within this subdomain, was recently determined to be the essential amino acid enabling the interaction between resistance gene Pto and avrPto, the corresponding pathogen avirulence factor (Frederick et al. 1998). Sequences within this subdomain have also been shown to be required for autophosphorylation (Steinberg et al. 1993).

For each of these combined subdomains, greater than 100 sequences with expect values less than 3×10^{-5} and similarity greater than 70% were located in the database. A most-important observation was that all of the sequence matches for each Sg5–3e query were from plant species such as *Arabidopsis*, rice, maize, potato, tomato, carrot and rose. This is somewhat analogous to yeast where protein kinases unique to that species were discovered (Hunter and Plowman 1997). Whereas some protein kinases homologous to Sg5–3e are receptor-like protein kinases involved in diverse cellular functions such as self-incompatibility (BNASRECKIN; Goring et al. 1993) and leaf and endosperm differentiation (CRINKLY4; Becraft et al. 1996), and Pto is a cytoplasmic protein kinase without a receptor domain, it could be suggested that the conserved structure of the kinase domain provides such a unique regulatory function that it was conserved during evolution among different plant species.

In contrast, the PvPKI-searches identified sequences from several plant species, but also from such diverse eukaryotic species as human, rat, mouse, chicken, *Drosophila*, *Ustilago*, *Dictyostelium*, *Leishmania*, *Ascaris*, *Euglena* and *Plasmodium*. In each case, the similarity between the query and subject sequence was >70%. As an example, ribsomal S6 protein kinase and cAMP dependent protein kinase were two proteins from several non-plant species that were found to be similar to subdomains VIa, VIb, VIII and IX of PvPKI. These similarities were as high as 87% (63% identity, no gaps) for human *RPS6KA2* (subdomains VIa and VIb) and 79% (64% identity; one gap) for human *RPS6KA4* (subdomains VIII and IX).

Together these results suggest that common bean protein kinase sequences similar to Sg5–3e evolved after the divergence of plants from other eukaryotes whereas those sequences related to PvPKI are more ancestral and predate the divergence. Furthermore, the fact that our previous PvPK1 searches using individual subdomain I–V and X–XI sequences as a query did not return any non-plant species sequences suggests that VIa, VIb, VIII and IX are the conserved ancestral subdomains.

Acknowledgments This research was supported by USDA/NRI Plant Genome grant 95–37300–1592. The authors would also like to thank Drs. Ted Helms and Richard Novy for critically reviewing this manuscript.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Becraft PW, Stinard PS, McCarty DR (1996) CRINKLY4: A TNFR-like receptor kinase involved in maize epidermal differentiation. Science 273:1406–1409

Borovkov AY, Rivkin MI (1997) *Xcm*I-containing vector for direct cloning of PCR products. BioTechniques 22:812–814

Chervitz SA, Aravind L, Sherlock G, Ball CA, Koonin EV, Dwight SS, Harris MA, Dolinski K, Mohr S, Smith T, Weng S, Cherry JM, Botstein D (1998) Comparison of the complete protein sets of worm and yeast: orthology and divergence. Science 282:2022–2028

Chothia C. Lesk AM (1986) The relation between the divergence of sequence and structure in proteins. EMBO J 5:823–826

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Dröge-Laser W, Kaiser A, Lindsay WP, Halkier BA, Loake GJ, Doerner P, Dixon RA, Lamb C (1997) Rapid stimulation of a soybean protein-serine kinase that phosphorylates a novel bZIP DNA-binding protein, G/HBF-1, during the induction of early transcription-dependent defenses. EMBO J 16:726–738

Flor HH (1956) The complementary genic systems in flax and flax rust. Adv Genet 8:29–54

Frederick RD, Thilmony RL, Sessa G, Martin GB (1998) Recognition specificity for the bacterial avirulence protein avrPto is determined by Thr-204 in the activation loop of the tomato Pto kinase. Mol Cell 2:241–245

Goring DR, Glavin TL, Schafer U, Rothstein SJ (1993) An S receptor kinase gene in self-compatible *Brassica napus* has a 1-bp deletion. Plant Cell 5:531–539

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative modelling. Electrophoresis 18:2714–2723

Hammond-Kosack KE, Jones JDG (1997) Plant disease resistance genes. Annu Rev Plant Physiol Mol Biol 48:575–607

Hanks SK, Hunter T (1995) The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J 9:576–596

Hanks SK, Quinn AM (1991) Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. Methods Enzymol 200:38–81

Hardie DG (1999) Plant protein serine/threonine kinases: classification and functions. Annu Rev Plant Physiol Mol Biol 50: 97–131

Hirayama T, Oka A (1992) Novel protein kinase of *Arabidopsis thaliana* (APK1) that phosphorylates tyrosine, serine and threonine. Plant Mol Biol 20:653–662

Huang X, Miller M (1991) A time-efficient, linear-space local similarity algorithm. Adv Appl Math 12:337–367

Hunter T, Plowman GD (1997) The protein kinases of budding yeast: six score and more. Trends Biochem Sci 22:18–22

Kanazin V, Marek LF, Shoemaker RC (1996) Resistance gene analogs are conserved and clustered in soybean. Proc Natl Acad Sci USA 93:11746–11750

Knighton DR, Zheng J, Ten Eyck LF, Ashford VA, Xuong N-H, Taylor SS, Sowadski JM (1991) Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. Science 253:407–420

Lawton MA, Yamamoto RT, Hanks SK, Lamb CJ (1989) Molecular cloning of plant transcripts encoding protein kinase homologs. Proc Natl Acad Sci USA 86:3140–3144

Leister D, Ballvora A, Salamini F, Gebherdt C (1996) A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plant. Nature Genet 14:421–429

Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Devos K, Graner A, Schulze-Lefert P (1998) Rapid reorganization of resistance gene homologues in cereal genomes. Proc Natl Acad Sci USA 95:370–375

Loh Y-T, Martin GB (1995) The disease-resistance gene *Pto* and the fenthion-sensitivity gene *Fen* encode closely related functional protein kinases. Proc Natl Acad Sci USA 92:4181–4184

Machida Y, Nishihama R, Kitakura S (1997) Progress in studies of plant homologs of mitogen-activated protein (MAP) kinase and potential upstream components in kinase cascades. Crit Rev Plant Sci 16:481–496

Martin GB, Brommonschenkel JC, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262:1432–1436

McClean PE, Myers JR, Hammond JJ (1993) Coefficient of parentage and cluster analysis of North American dry bean cultivars. Crop Sci 33:190–197

MoranTV, Walker JC (1993) Molecular cloning of two novel protein kinase genes from *Arabidopsis thaliana*. Biochim Biophysica Acta 1216:9–14

Nasrallah JB, Stein JC, Kandasamy MK, Nasrallah ME (1994) Signalling the arrest of pollen tube development in self-incompatible plants. Science 266:1505–1508

Peitsch MC (1996) ProMod and Swiss-Model: internet-based tools for automated comparative protein modelling. Biochem Soc Trans 24:274–279

Rivkin MI, Vallejos CE, McClean PE (1999) Disease resistance related sequences in common bean. Genome 42:41–47

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Scofield SR, Tobias CM, Rathjen JP, Chang JH, Lavelle DT, Michelmore RW, Staskawicz BJ (1996) Molecular basis of gene-for-gene specificity in bacterial speck disease of tomato. Science 274:2063–2065

Song W, Wang G, Chen L, Kim H, Pi L, Holsten T, Gardner J, Wang B, Zhai W, Zhu L, Fauquet C, Ronald P (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 270:1804–1806

Steinberg RA, Cauthron RD, Symcox MM, Shuntoh H (1993) Autoactivation of catalytic subunit (C) of cyclic AMP-dependent protein kinase by phosphorylation at threonine 197. Mol Cell Biol 13:2332–2341

Tang X, Frederick RD, Zhou J, Halterman,DA, Jia Y, Martin GB (1996) Initiation of plant disease resistance by physical interaction of AvrPto and Pto kinase. Science 274:2060–2062

Thompson JD, Higgins DG, Gibson TJ (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

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Vallejos CE, Sakiyama NS, Chase CD (1992) A molecular marker-based linkage map of *Phaseolus vulgaris* L. Genetics 131: 733–740

Yu YG, Buss GR, Saghai Maroof MA (1996) Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. Proc Natl Acad Sci USA 93:11 751–11756

Zhou J, Loh Y-T, Bressan RA, Martin GB (1995) The tomato gene *Pti1* encodes a serine/threonine kinase that is phosphorylated by Pto and is involved in the hypersensitivity response. Cell 83:925–935

Zhou J, Tang X, Martin GB (1997) The Pto kinase conferring resistance to tomato bacterial speck disease interacts with proteins that bind a *cis*-element of pathogenesis-related genes. EMBO J 16:3207–3218