

Expression and evolutionary features of the hexokinase gene family in *Arabidopsis*

Abhijit Karve · Bradley L. Rauh · Xiaoxia Xia ·
Muthugapatti Kandasamy · Richard B. Meagher ·
Jen Sheen · Brandon d. Moore

Received: 15 March 2008 / Accepted: 24 April 2008 / Published online: 15 May 2008
© Springer-Verlag 2008

Abstract *Arabidopsis* hexokinase1 (HXK1) is a moonlighting protein that has separable functions in glucose signaling and in glucose metabolism. In this study, we have characterized expression features and glucose phosphorylation activities of the six HXK gene family members in *Arabidopsis thaliana*. Three of the genes encode catalytically active proteins, including a stromal-localized HXK3 protein that is expressed mostly in sink organs. We also show that three of the genes encode hexokinase-like (HKL) proteins, which are about 50% identical to AtHXK1, but do not phosphorylate glucose or fructose. Expression studies indicate that both HKL1 and HKL2 transcripts occur in most, if not all, plant tissues and that both proteins are targeted within cells to mitochondria. The HKL1 and HKL2 proteins have 6–10 amino acid insertions/deletions (indels) at the adenosine binding domain. In contrast, HKL3 transcript was detected only in flowers, the protein lacks the noted indels, and the protein has many other amino acid changes that might compromise its ability even to bind glucose or ATP. Activity measurements of HXKs modified by

site-directed mutagenesis suggest that the lack of catalytic activities in the HKL proteins might be attributed to any of numerous existing changes. Sliding windows analyses of coding sequences in *A. thaliana* and *A. lyrata* ssp. *lyrata* revealed a differential accumulation of nonsynonymous changes within exon 8 of both HKL1 and HXK3 orthologs. We further discuss the possibility that the non-catalytic HKL proteins have regulatory functions instead of catalytic functions.

Keywords Glucose phosphorylation · Hexokinase · Hexokinase-like · Molecular evolution · Sliding windows

Abbreviations

GFP	Green fluorescent protein
HA	Hemagglutinin
HKL	Hexokinase-like
HXK	Hexokinase
Indels	Insertions/deletions
K_A	Number of nonsynonymous substitutions per nonsynonymous site
K_S	Number of synonymous substitutions per synonymous site

Electronic supplementary material The online version of this article (doi:10.1007/s00425-008-0746-9) contains supplementary material, which is available to authorized users.

A. Karve · B. L. Rauh · X. Xia · B. d. Moore (✉)
Department of Genetics and Biochemistry,
Clemson University, Clemson, SC 29634, USA
e-mail: moore8@clemson.edu

M. Kandasamy · R. B. Meagher
Genetics Department, University of Georgia, Athens,
GA 30602, USA

J. Sheen
Department of Molecular Biology,
Massachusetts General Hospital, Boston, MA 02114, USA

Introduction

Sugars are the primary currency in the metabolic economy of most cellular life. Contemporary research has revealed remarkable interconnections between the cellular and molecular processes that govern production or acquisition of sugars and their efficient utilization. In plants, sugars regulate plant growth and development by interacting with many different control processes, including ones with meristematic functions (Smeekens 2000; Francis and Halford

2006; Rolland et al. 2006). For example, in *Arabidopsis* both glucose and sucrose modulate the expression of nucleolin, a multi-function regulator of ribosome synthesis (Kojima et al. 2007). An increased amount of nucleolin was suggested to be a key component in the process by which sugars can enhance meristematic cell division activity. Interestingly, the targeted expression of cell wall invertase in apical meristems of *Arabidopsis* was shown to influence the developmental transition to flowering and ultimately to increase seed yield (Heyer et al. 2004).

Short-term treatments of *Arabidopsis* seedlings with glucose or sucrose have been shown to affect the expression of about 1,000–1,700 transcripts, depending on experimental conditions (Price et al. 2004; Osuna et al. 2007). Sugars can influence plant gene expression both through general metabolic effects and as signal molecules that can directly interact with sensor/transducer proteins (Sheen et al. 1999; Xiao et al. 2000). *Arabidopsis* hexokinase1 (HXK1) is perhaps the best characterized glucose signaling protein. The isolation and characterization of a null mutant of AtHXK1, *gin2-1*, revealed associated phenotypes including reduced shoot and root growth, reduced leaf expansion, increased apical dominance, delayed flowering and senescence, decreased auxin sensitivity, increased cytokinin sensitivity, and changes in transcript levels of several target genes (Moore et al. 2003). Furthermore, complementation of these phenotypes by transformation of *gin2-1* with a catalytically compromised HXK1 protein (S177A) demonstrated that the HXK metabolic function can be uncoupled from its signaling and related growth promoting functions. AtHXK1 is, therefore, a moonlighting protein (Moore 2004). The multi-function nature of AtHXK1 can be viewed as a novel cellular solution to integrate glucose metabolism with a separable glucose signal transduction process.

The biochemical basis for AtHXK1 function as a moonlighting protein has not been established. Analysis of crystal structures of other moonlighting proteins has shown that one mechanism for acquiring a moonlighting function is the development of distinct surface features that mediate protein-protein interactions (Jeffery 2004). While a crystal structure of AtHXK1 is not available, much might be learned by a close inspection of its evolutionary heritage and of existing structural homologs. Bork et al. (1992) suggested that a prokaryotic, dimeric ancestral ATPase has evolved through diverse processes into structurally related families of actin, hexokinase, and heat shock protein 70. The relatively sugar non-specific, but ATP-dependent hexokinases are characteristic of higher Eukarya and are thought to have arisen from a common ~50 kDa ancestral protein (Cárdenas et al. 1998). As far as is known, hexokinases are present in a given eukaryote as a multi-gene family. Surprisingly, within some of these gene families, individual members can have specialized non-catalytic,

regulatory functions. For example, there are two described hexokinase-like (HKL) proteins in *Aspergillus nidulans*. They both lack catalytic activity, but they are negative regulators for secretion of an extracellular protease in response to carbon starvation (Bernardo et al. 2007). In this case, sequence analysis suggested that specific amino acid changes relative to the canonical sequence are responsible for the lack of catalytic activity. It remains to be determined whether specialized regulatory HXKs occur within plant HXK gene families.

Phylogenetic analyses of a variety of plant HXKs indicate that these occur largely in two groups, ones with plastid signal peptides (Type A) and ones with N-terminal membrane anchors (Type B; Olsson et al. 2003). Direct experimental evidence for stromal-localized HXKs have been reported from moss (Olsson et al. 2003), tobacco (Giese et al. 2005), rice (Cho et al. 2006a), and tomato (Kandel-Kfir et al. 2006). Plastidic NtHXK2 is expressed mostly in certain starch-containing sink tissues, while plastidic LeHXK4 occurs in both source and sink organs, including non-starch containing fruits. Membrane-bound HXKs occur largely, but not exclusively, associated with mitochondria (Kandel-Kfir et al. 2006 and references therein; Damari-Weissler et al. 2006). AtHXK1 is predominantly associated with mitochondria, but also can occur in the nucleus (Cho et al. 2006b; Balasubramanian et al. 2007); both forms can modulate gene expression. Whether the nuclear form of AtHXK1 maintains its membrane anchor is not clear. Rice and maize have one or more cytosolic HXKs (da-Silva et al. 2001; Cho et al. 2006a), though these forms might occur only in monocots (Damari-Weissler et al. 2006). A number of different possible metabolic roles of HXKs have been recently described (Claeyssens and Rivoal 2007). Mitochondrial HXKs are thought to have preferred access to ATP produced in respiration for consumption by active metabolite fluxes through sucrose cycling, glycolysis, and sugar nucleotide syntheses (Rontein et al. 2002; Graham et al. 2007).

Among the better examined plant HXK families, rice has at least nine expressed HXKs (Cho et al. 2006a), tomato at least four HXKs (Kandel-Kfir et al. 2006), and *Arabidopsis* likely six HXKs (Rolland et al. 2002; Claeyssens and Rivoal 2007). We are interested in the function of *Arabidopsis* HXKs in organismal space. In this study, we describe their gene structures, their tissue and sub-cellular expressions, and we show that three of the six family members lack catalytic activity. We then did a detailed amino acid sequence analysis to identify key amino acid differences and to test a number of possible mechanisms by which catalysis might be compromised. The presence of non-catalytic HKL proteins in plants raises intriguing questions regarding their evolution and function. A comparison of HXK family coding sequences from *A. thaliana* and *Arabidopsis lyrata*

ssp. *lyrata* (hereafter *A. lyrata*) allowed us to identify regions of some gene orthologs that are undergoing possible differential selection.

Materials and methods

Plant material and growth conditions

Seeds of *Arabidopsis thaliana* (L.) Heyn. Ecotype Columbia (Columbia-0) were obtained from Arabidopsis Biological Resource Center (Ohio State University). Seeds of maize were purchased (line FR922 × FR967, Seed Genetics, Inc. Lafayette, IN, USA) and dark-grown for 9 days, followed by overnight greening (Jang and Sheen 1994). For most experiments, Arabidopsis was grown in soil in a growth chamber (Balasubramanian et al. 2007), except for collecting root tissue from plants grown by hydroponics (Tocquin et al. 2003). Leaf tissue from *A. lyrata* was kindly provided by Dr. Amy Lawton-Rauh.

Cloning and plasmid constructs

AtHXK1 was previously cloned using *Bam*H1 and *Stu*I restriction sites into the HBT plant expression vector (Kovtun et al. 1998) followed either with a C-terminal double hemagglutinin (HA) tag (Moore et al. 2003) or with a C-terminal green fluorescent protein (GFP) fusion (Balasubramanian et al. 2007). Leaf or seedling cDNA libraries (see below) were used as template for PCR amplification of AtHXK2 (At2g19860, 5'-CGG GAT CCC GAT GGG TAA AGT GGC AGT TGC AAC G, 5'-AAA AGG CCT ACT TGT TTC AGA GTC ATC TTC), AtHXK3 (At1g47840, 5'-CGG GAT CCC G AT GTC ACT CAT GTT TTC TTC CCC TGT C, 5'-AAA AGG CCT GTA AAT GGA GTT AGT GGC CGC C), AtHKL1 (At1g50460, 5'-CGG GAT CCC GAT GGG GAA AGT GGC GGT TGC G, 5'-AAA AGG CCT TGA CTG TAA AGA GGC AAC GAG GAG), AtHKL2 (At3g20040, 5'-CGG GAT CCA TGG GGA AGG TTT TGG TGA TGT TG, 5'-AAA AGG CCT TAC GGA TGG TAT TGT TTG AAC AC), and AtHKL3 (At4g37840, 5'-TGC CAT GGC ATG ACC AGG AAA GAG GTG GTT C, 5'-GAA GGC CTC TTG CTT TCA GAA TCT TGA TGA). PCR products were then ligated into an HBT vector with the double HA tag, using the *Bam*HI/*Stu*I restriction sites for most constructs or available *Nco*I/*Stu*I sites for AtHKL3. All clones were validated by direct sequencing of plasmid DNA and by predicted sizes of the expressed proteins. The coding sequences were then sub-cloned into the same vector, but with a GFP tag.

Site-directed changes, insertions, and deletions of native sequences were all made by Quick Change (Stratagene).

For AtHXK1, the target amino acid changes and primers were as follows: N106Y (5'-GGA CCT AGG GGG GAC ATA CTT CCG TGT CAT GCG TG, 5'-CA CGC ATG ACA CGG AAG TAT GTC CCC CCT AGG TCC), G173A (5'-GGT AGA CAG AGG GAA TTA GCC TTC ACT TTC TCG TTT CC, 5'-GG AAA CGA GAA AGT GAA GGC TAA TTC CCT CTG TCT ACC), L251F (5'-G GAT GTT GTT GCT GTT ATT TTC GGC ACT GGG ACA AAC G, 5'-C GTT TGT CCC AGT GCC GAA AAT AAC AGC AAC AAC ATC), C159E (5'-G AAG TTT GTC GCT ACA GAA GAG GAA GAC TTT CAT CTT CC, 5'-GG AAG ATG AAA GTC TTC CTC TTC TGT AGC GAC AAA CTT C), and insert ⁴²⁸GITSGRSRSE⁴³⁷ (5'-CTG GGA AGA GAT ACT ACT AAA GGA ATC ACC AGC GGA AGA TCT AGA AGC GAG GAC GAG GAG GTG CAG AAA TCG G, 5'-C CGA TTT CTG CAC CTC CTC GTC CTC GCT TCT AGA TCT TCC GCT GGT GAT TCC TTT AGT AGT ATC TCT TCC CAG). For AtHKL1, amino acids ⁴²⁵GITSGRSRSE⁴³⁴ were deleted similarly (5'-GAT AGG CCG AGA TGG AAG CAG AAG TGA AAT CCA AAT G, 5'-CAT TTG GAT TTC ACT TCT GCT TCC ATC TCG GCC TAT C). All mutations were verified by DNA sequencing.

Analysis of mRNA expression by RT-PCR

Total RNA was prepared using the RNeasy kit (Qiagen) from 100 mg of corresponding plant tissue from *A. thaliana*. Root tissue was collected from plants grown in hydroponics, while all other tissues were from soil grown plants. One microgram of total RNA was converted to cDNA using the Protoscript II RT-PCR kit according to the manufacturer's instructions (New England BioLabs). PCR primer sequences were generated using the AtRTPrimer public database (Han and Kim 2006). Primers in all cases span one or more introns: AtHXK1 (5'-TGC TGC TTT CTT TGG CGA TAC AGT, 5'-AAA ATG GCG CTC TTT GGG TAG GTT; expected size = 505 bp), AtHXK2 (5'-ACA AAT GCA GCC TAT GTC GAA CGT G, 5'-TGT TCG GGG TCC TTA TGA TGA ATG G; expected size = 316 bp), AtHXK3 (5'-TCT CGA CCA CGC TCC AAT TAC ATC, 5'-AAT CAC ACC GAC CAT CAC ATC CTC; expected size = 702 bp), AtHKL1 (5'-GTT GGA GCC TTG TCG CTT GGA TAT T, 5'-CCT GCT CTT CGT GTA ACC ACA TCG; expected size = 521 bp), AtHKL2 (5'-CCC AGT CAA GCA GAC ATC CAT CTC A; 5'-TCG CCC AGA TAC ATC CCT CCT ATC A, expected size = 441 bp), and AtHKL3, (5'-TGG AAA CAC ACG GTC TGA AAA TTC G; 5'-TCA TCA CCA AGC ATT TCC CAA ACG, expected size = 736 bp). As a control for amount of tissue template, we routinely used AtUBQ5 (At3g62250, 5'-GTG GTG CTA AGA AGA GGA AGA, 5'-TCA AGC TTC AAC TCC TTC TTT; expected size =

254 bp). For PCR, we used 0.5–0.8 μ l of cDNA in 10 μ l reactions to first balance UBQ expression for the set of tissue cDNAs, with corresponding tissue template concentrations used thereafter for PCR reactions of varying cycle numbers for each product.

Protoplast transient expression and ^{35}S -labeling

Protoplasts were isolated from Arabidopsis leaves (Hwang and Sheen 2001) or greening maize leaves (Jang and Sheen 1994). These were transfected using the polyethylene glycol 4,000 (Fluka) protocol (Yoo et al. 2007) and 6–12 μ g of cesium chloride-purified plasmid DNA. In some experiments, newly synthesized proteins were labeled with [^{35}S]Met (Perkin Elmer) for 8 h, then collected from lysed protoplasts onto protein A agarose beads using anti-HA antibodies (Roche) as detailed previously (Balasubramanian et al. 2007). Proteins were solubilized in 2 \times SDS treatment buffer, electrophoresed on 10% SDS gels, and visualized by fluorography.

Enzyme activity assays

H XK was assayed largely as described by Doehlert (1989) in a medium containing 50 mM Bicine–KOH pH 8.5, 5 mM MgCl_2 , 2.5 mM ATP, 1 mM NAD, 15 mM KCl, 2 units of glucose 6-phosphate dehydrogenase (Sigma G8404), and either 2 mM glucose or 100 mM fructose. The increase in A_{340} was monitored over a 30 min interval and rates then calculated accordingly. Transfected protoplasts from greening maize leaves were used in either of two ways as a source of possible enzyme activity. First, frozen protoplast pellets were lysed by vortexing in a standard leaf extraction buffer containing 50 mM HEPES–KOH pH 7.5, 5 mM MgCl_2 , 1 mM EDTA, 15 mM KCl, 10% glycerol, 0.1% Triton X-100, and 1X protease inhibitor cocktail (Roche). This extract was then assayed directly. Second, HA-tagged proteins were isolated as described above. Thrice-washed beads were resuspended in enzyme activity assay buffer (minus sugar), transferred to cuvettes, and rates then measured after adding glucose or fructose. Immobilized protein had lower enzyme activity, but the recovery rates were very consistent ($20 \pm 2\%$). Included protoplast controls that received non-coding plasmid DNA had little or no background due to endogenous H XK activity that was initially present in the maize protoplasts.

Analysis of H XK family sequences in *A. thaliana* and *A. lyrata*

We first queried the NCBI database using BLAST and specifying the *A. lyrata* WGS first draft sequence database, using the coding sequence for each of the H XK family

member genes from *A. thaliana* as the query sequences. From this, we identified the homologous exons, introns, and splice sites for each of the six genes from the genome project of *A. lyrata*. To correct potential sequence errors we identified within exon 6 of *AtH XK3*, we directly sequenced the PCR product from first strand cDNA synthesis of the corresponding transcript. Otherwise, the available sequence information was robust with multiple reads and the splice sites were highly conserved.

All nucleotide and amino acid sequences were aligned manually using BioEdit (Hall 1999) and exported as Nexus or FASTA files. Phylogenetic analyses of the H XK and H KL loci were conducted using amino acid alignments in MEGA build 4.0.2 (Tamura et al. 2007). Phylogenetic trees were estimated using the neighbor-joining method, 1,000 bootstrap replicates, and the Dayhoff substitution model. Codon usage bias was examined using DnaSP v4.20.2 (Rozas et al. 2003), then calculated as effective number of codons (Wright 1990) and by the codon bias index (Morton 1993). Nucleic acid sequence diversity was estimated using DnaSP v4.20.2 (Rozas et al. 2003) to calculate ω ratios (number of nonsynonymous substitutions per nonsynonymous site/number of synonymous substitutions per synonymous site, K_A/K_S), following Nei and Gojobori (1986), and sliding window analysis of K_A/K_S ratios. All nucleotide sequence based analyses used paired alignments of coding sequences of *A. thaliana* and *A. lyrata* on a gene by gene basis.

Results

Gene structure of Arabidopsis H XK family

The *A. thaliana* genome potentially encodes six H XK related proteins (TAIR). By pair-wise Blast searches, these predicted proteins range from 45 to 85% identical to AtH XK1 (Table 1). As detailed below, we have designated these as H XK proteins or as hexokinase-like (H KL) proteins based, respectively, on whether they have apparent glucose phosphorylation activity or whether they lack catalytic activity. For reference, AtH XK1 is 36% identical to yeast H XK2 and about 70% identical to rice H XK2 and to tomato H XK1.

Phylogenetic analysis of the AtH XKs reveals several interesting features (Fig. 1a). Bootstrap replicate values suggest that the gene pairs H XK1 and H XK2, and H KL1 and H KL2 are more closely related to each other than they are to other genes in the family. When analyzed with the 10 reported rice H XK family members, AtH KL1 and AtH KL2 form a related sub-group with OsH XK3 and OsH XK10 (Cho et al. 2006a). Within *A. thaliana*, H KL3 forms a distinct and not closely related sub-group. AtH KL3 also was reported not to form any closely related phylogenetic

Table 1 Selected properties predicted for HXK family proteins from *A. thaliana*

	% Identity	Transmembrane domain	Transit peptide
HXK1	100	8–24	No
HXK2	85	8–24	No
HXK3	53	No	1–34
HKL1	54	8–24	No
HKL2	49	4–20	No
HKL3	45	6–22	No

Amino acid sequences were subjected to pair-wise Blast analysis, and PSORT (<http://psort.nibb.ac.jp/>) and Predotar (<http://urgi.versailles.inra.fr/predotar/predotar.html>) targeting analyses

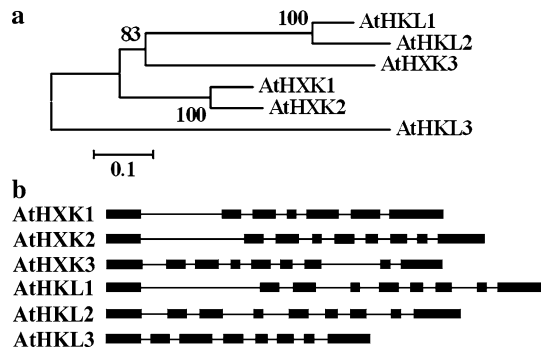


Fig. 1 a Phylogenetic relationships of predicted Arabidopsis HXK family proteins. As described below, proteins with detectable catalytic activity are designated as HXK proteins. Those without detectable catalytic activity are designated as HKL proteins. *HXK1* At4g29130, *HXK2* At2g19860, *HXK3* At1g47840, *HKL1* At1g50460, *HKL2* At3g20040, *HKL3* At4g37840. Phylogenetic reconstruction was done using MEGA4 (Tamura et al. 2007). The scale bar the frequency of amino acid substitutions per unit length. **b** Illustrations of gene structures were determined from genomic sequence data (NCBI, TAIR). Exons are indicated by rectangles, introns by lines. The lengths of both are proportional to the number of nucleotides that are present

sub-groups with rice HXKs (Cho et al. 2006a). The genome structure of the AtHXKs shows that most have nine exons, except for HXK1 which has seven and HKL3 which has eight (Fig. 1b). Most rice HXKs also have nine exons (Cho et al. 2006a). The intron structures of AtHXKs vary among the different family members. Intron1 in HXK1, HXK2, and HKL1 is relatively long, ranging from 625 to 804 nucleotides, while intron1 in HXK3, HKL2, and HKL3 is shorter, ranging from 72 to 183 nucleotides. HKL3 has short introns throughout the gene, averaging 81 nucleotides.

Organ expression of *AtHXKs* mRNAs and cellular expression of *AtHXKs* proteins

We examined by semi-quantitative RT-PCR the organ expression of transcripts for the *A. thaliana* HXK gene family (Fig. 2). HXK1 transcript was abundant in all organs

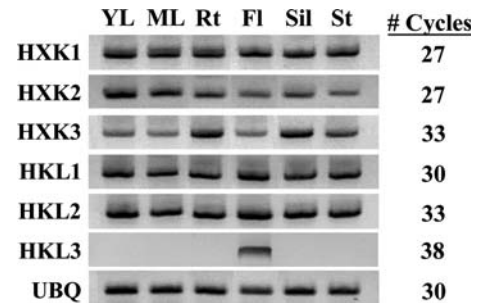
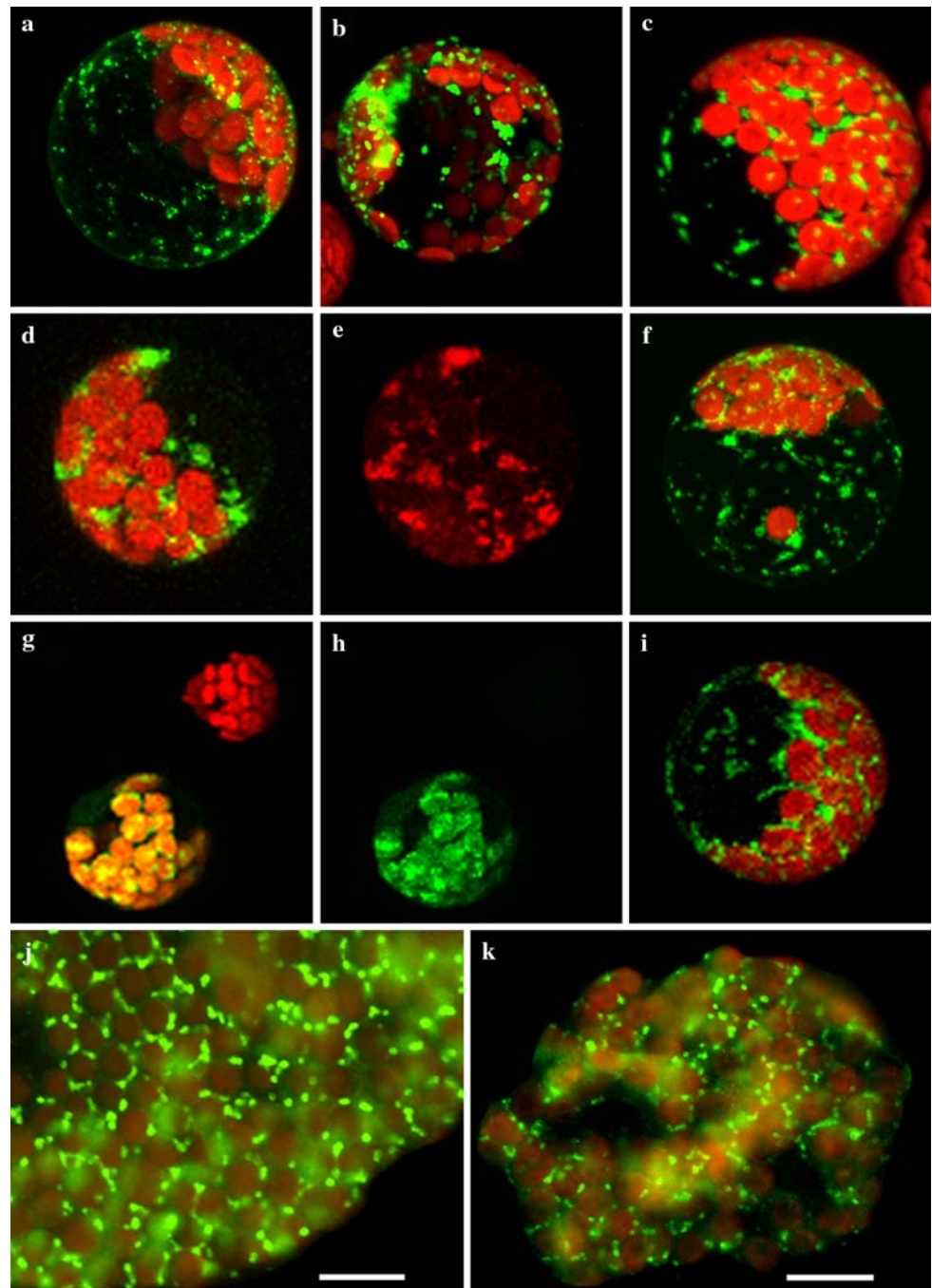


Fig. 2 Organ expression of Arabidopsis HXK family genes determined by RT-PCR. Ubiquitin5 (UBQ) was amplified as a control for amount of template. Product sizes ranged from 254 to 702 bp. Additional PCR reactions were run with all primer pairs at increased and decreased cycle numbers to establish reported conditions (data not shown). *YL* young leaves, *ML* mature leaves, *Rt* root, *Fl* flower, *Sil* silique, *St* stem

examined. HXK2 transcript was expressed to a relatively similar extent in leaves, but less so in sink tissues. HXK3 mRNA was relatively more abundant in sink tissues such as roots and siliques, but the increased number of PCR cycles used for its amplification indicates that it likely is not as abundant as is the HXK1 transcript. HKL1 and HKL2 transcripts were not expressed as highly as HXK1 mRNA, but it is noteworthy that all 3 of these transcripts were expressed at relatively similar levels in all organs examined. In contrast, HKL3 mRNA was detected only in flowers and at relatively much lower amounts.

Proteomic analyses have indicated that most of the HXK family members are associated with mitochondria (Heazlewood et al. 2004). Sequence analysis supports this evidence as well, indicating that HXK1, HXK2, HKL1, HKL2, and HKL3 all have predicted N-terminal transmembrane peptides that target to the mitochondria (Table 1). These possible targeting peptides are located upstream from the conserved large domain of the HXKs (Supplemental Table 1). However, HXK3 has a putative N-terminal transit peptide, which indicates that it might be expressed in plastids. We examined the subcellular localization of the AtHXK proteins by cloning these as C-terminal GFP fusions, followed by transient expression of their cDNAs in leaf protoplasts, and subsequent imaging of expressed fluorescence (Fig. 3a–i). As shown previously (Balasubramanian et al. 2007; Damari-Weissler et al. 2007), HXK1-GFP fluorescence occurred under these conditions only at the mitochondria. HXK2, HKL1, HKL2, and HKL3-GFP fluorescence also was observed only at the mitochondria. Expression of these proteins in protoplasts does apparently cause the mitochondria to aggregate as noted before with HXK1, but co-staining with MitoTracker dyes always showed their GFP fluorescence to be mitochondrial localized (e.g., compare fluorescence patterns in Fig. 3d, e). In contrast, HXK3-GFP fluorescence is expressed in chloroplasts

Fig. 3 Subcellular expression of Arabidopsis HXK family proteins. **a–i** C-terminal GFP fusion proteins were cloned and expressed in leaf protoplasts, then imaged with a scanning laser confocal microscope. GFP fluorescence is pseudocolored green. Chlorophyll autofluorescence from a long-pass filter is shown pseudocolored red. **a** Porin-GFP, a mitochondrial marker. **b** HXK1-GFP. **c** HXK2-GFP. **d** HKL1-GFP. **e** Protoplast in (d) showing selective staining with MitoTracker Red (pseudocolored red). **f** HKL2-GFP. **g** HXK3-GFP, showing merged GFP and chlorophyll fluorescence in lower protoplast (yellow-orange). The upper protoplast did not express GFP (red only). **h** Protoplasts in (g) showing only the GFP channel. **i** HKL3-GFP. **j, k** Cellular immunolocalization of HXK in leaf mesophyll cells of young seedlings. Following cryofixation and freeze substitution, cells were labeled with anti-HXK1 polyclonal antibody and then with secondary antibody conjugated to FITC. Leaf cells were observed by fluorescence microscopy. Images show a single optical section. Scale bars 10 μ m. **j** Leaf cells of Ler wildtype seedlings. **k** Leaf cells of *gin2-1* seedlings, an HXK1-null mutant



(Fig. 3g, h). In the presented image, much of the GFP fluorescence occurs inside the chloroplasts, but some also is located on the outer surface. The latter might be due to protein accumulation at sites of import. Internal accumulation of GFP was shown by using a specific band pass filter that excludes chlorophyll fluorescence. In contrast to these fluorescence patterns, transfected yeast HXK2-GFP is expressed exclusively in the cytosol of leaf protoplasts (data not shown). This localization is consistent with the absence of an N-terminal membrane anchor in SchHXK2.

As a complementary approach to demonstrate the subcellular localization of AtHXKs, we have immunostained seedling leaves after their cryofixation and acetone freeze substitution. First, samples were prepared from wild type Ler. These were incubated with a polyclonal antibody to AtHXK1, then a FITC-conjugated secondary antibody, and were subsequently viewed by epifluorescence (Fig. 3j). As noted previously with wild type cells (Balasubramanian et al. 2007), we observed HXK1 antigen in numerous elongated shapes which appear to be mitochondria and that

were often located close to chloroplasts. This staining likely is due to some combination of HXK1, HXK2, HKL1, and/or HKL2 proteins. As a comparison, we next observed HXK antigen in leaves of *gin2-1* (Fig. 3 k), a HXK1-null mutant (Moore et al. 2003). We observed a readily detectable, but much reduced level of fluorescence, again associated with elongated, apparent mitochondrial structures. FITC fluorescence was not observed in the cytoplasm or elsewhere in either *gin2-1* or Ler mesophyll cells. Chloroplasts in these cells, however, were not labeled by antibody, perhaps due to the HXK3 protein being present in this tissue only at very low levels, if at all.

Protein sequence analyses of *AtHXKs*

The HXK tertiary structure is one of the better studied proteins from mammals and yeast. HXK proteins contain a large and a small domain. The sugar binding site is largely located in the small domain, with four additional peptide segments (Loops 1–4) that are induced to move upon binding the sugar ligand. These loops both complete the sugar binding sites as well as “pre-form” the nucleotide binding site (Kuser et al. 2000). In fact, analysis of HXK sequences from mammalian, yeast, and Arabidopsis (*AtHXK1*) using crystal structures, sequence elements, and/or space filling models indicate that most of the conserved amino acid residues occur at the cleft of the two domains and form the glucose and ATP binding sites (Kuser et al. 2000).

In order to better associate structural differences of *AtHXKs* with their possible functional differences, we have aligned and analyzed their 1° amino acid sequences (Fig. 4). Homologous regions and residues in *AtHXKs* were assigned based on reported detailed analyses of HXK2 from *Saccharomyces cerevisiae* (Bork et al. 1992; Kuser et al. 2000). By inspection, HXK1, HXK2, and HXK3 were much more similar to each other in the longer motifs designated phosphate 1, connect 1, phosphate 2, connect 2, and adenosine, than were HKL1, HKL2, and HKL3. The relative divergence of HKL3 is particularly pronounced in the phosphate 1, sugar, and connect 1 motifs. Most noticeably, both HKL1 and HKL2 have an indel of 10 and 6 amino acids, respectively, at the adenosine binding site. These both also have relatively divergent phosphate 1 and connect 1 motifs. Within the core sugar binding motif (LGFTFSFP-Q-L/I), the sequence is best conserved among HXK1, HXK2, and HXK3, with a limited divergence in HKL1 and HKL2, but extensive divergence in HKL3 (see also Supplemental Table 1). However, with reference to the HXK1 sequence, the key glucose contact residues S177, N256, E284, and E315 are conserved in all of the HXKs, with the single exception of T175 for S177 in HKL3. Among the 4 key noted loops, loop 2 is not conserved at the level of amino acid identity between *SchHXK2*

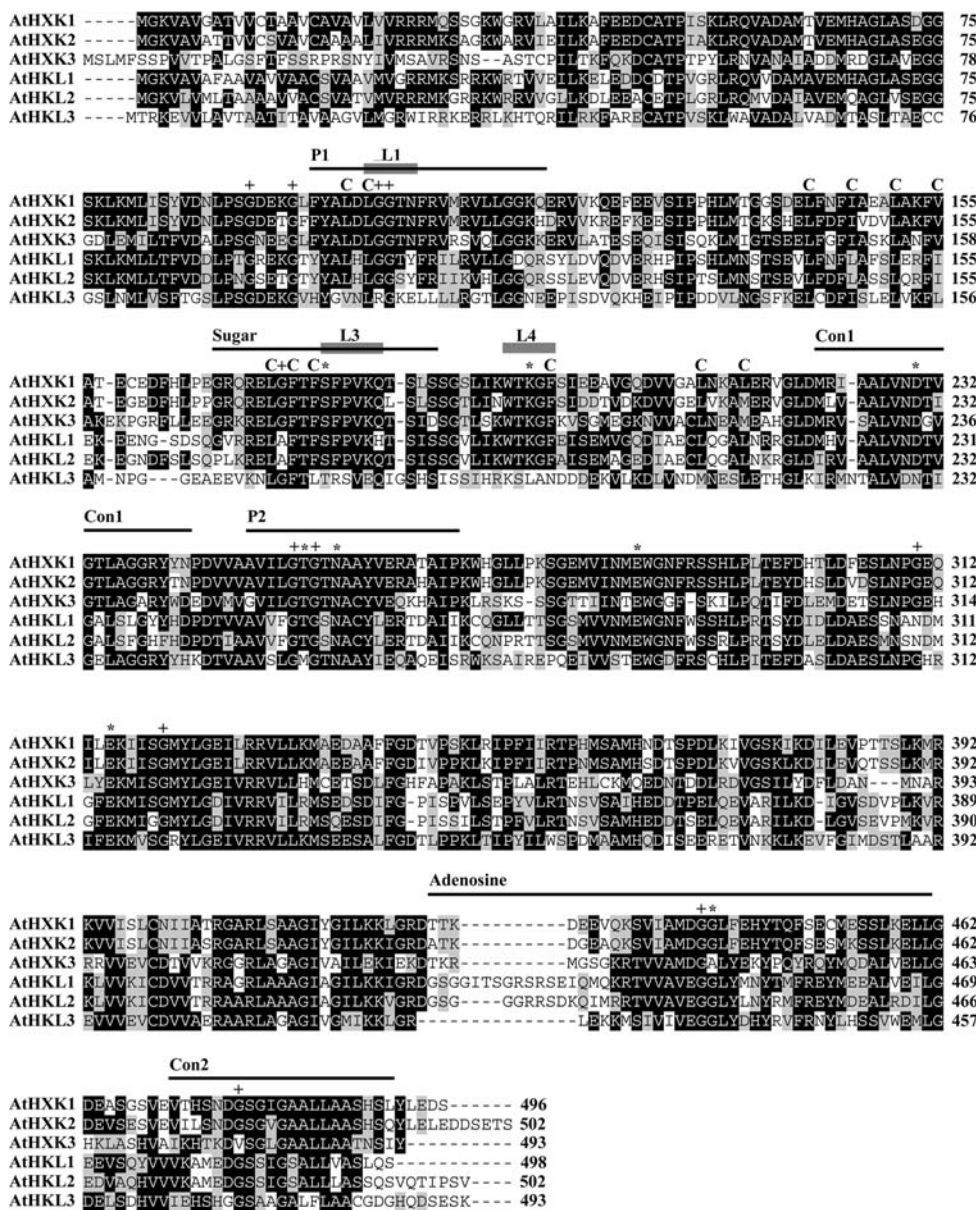
and any of the *AtHXKs*, but the other 3 loops have varying levels of identity. Loops 1, 3, and 4 diverge substantially in HKL3, but are identical or very similar in all other *AtHXKs*, with respect to *SchHXK2* (see also Supplemental Table 1). Among the 12 hydrophobic residues previously assigned to a channel in the small domain, these are either conserved or identical in all of the *AtHXKs*, with the exception of 2 different ones in HKL3 at F197 and L211 (Fig. 4). Also, among eight key glycine residues thought to be located at the end of α -helices or β -sheets, these are all conserved identically among HXK1, HXK2, and HXK3 proteins. The HKL predicted proteins though do show some divergence in these features. This includes substitutions for G103 in HKL3, for G173 and G310 in HKL1 and HKL2, and for G479 in all 3 HKL proteins. Finally, there are 2 recognized catalytic residues in HXK1, K195 and D230. Both are conserved in all of the *AtHXKs* except for HKL3 (L194 and N230). In summary, the amino acid sequence analysis shows a broad pattern of conserved key motifs and residues among HXK1, HXK2, and HXK3 predicted proteins. HKL3 protein lacks many recognized residues important for sugar and adenylate binding and for enzyme catalysis. In contrast, HKL1 and HKL2 proteins have most of the residues known to be important for sugar binding, but they do have a few noted key residue changes and also have an indel at the adenosine binding domain. Additionally though, HKL1 and HKL2 proteins do have changes in many residues relative to HXK1 protein, other than those specifically noted above.

Catalytic activities of native and modified *AtHXK* family proteins

We assessed the catalytic competency of HXK family proteins after transient expression of corresponding cDNAs in maize leaf protoplasts. In these experiments, HXK family genes were cloned as C-terminal fusions to double HA tags. Each HA tag is ten amino acids and the double tag does not appear to interfere with protein catalytic activity as measured with different C-terminal tags (double HA, Flag, GFP; unpublished data). Protein expression first was monitored by direct labeling using ^{35}S -Met (Fig. 5a) to adjust the amounts of transfected cDNAs to yield comparable amounts of expressed proteins for activity assays. In one approach to assess possible protein catalytic activities, we measured enzyme activities from protoplasts following their transient transformation with selected amounts of target cDNAs and an incubation period to allow the proteins to accumulate (Fig. 5b). This approach was rapid, but does include as background the endogenous HXK activity that is present in the protoplasts. Nonetheless, we found that three of the six HXK genes encode proteins with glucokinase activity. HXK1 and HXK2 were both shown previously to

Fig. 4 Amino acid sequence alignment for predicted Arabidopsis HXK family proteins. Sequences were aligned using ClustalW in BioEdit (Hall 1999).

Annotations are based on regions homologous to yeast hexokinase II (Kuser et al. 2000). *P1* phosphate 1, *Con 1* connect 1, *P2* phosphate 2, *Con 2* connect 2, *L1* loop1, *L3* loop3, *L4* loop4, *C* predicted hydrophobic channel amino acid; + predicted conserved glycine residue; *other predicted conserved residues (following AtHXK1 sequence, S177, K195, D230, T253, N256, E284, E315, and G441)



have catalytic activity (Jang et al. 1997), while the finding that HXK3 can phosphorylate glucose is novel. The three expressed proteins that apparently lack catalytic activity are designated, as mentioned earlier, as HKL proteins. Comparable results were obtained with 0.1 M fructose or increased glucose concentrations in the assay medium. That is, HXK3 can phosphorylate fructose also, while HKL1, HKL2, and HKL3 proteins cannot phosphorylate fructose either (data not shown).

To minimize the appreciable background HXK activity that is present in the protoplasts, we also assayed the expressed proteins after their immunoprecipitation from lysed protoplasts, captured using anti-HA antibody and Protein A agarose beads (Fig. 5c). Using the washed beads allowed us to exclude most, if not all, of the endogenous

activity and thereby better determine whether a given construct has any measurable catalytic activity. As shown, HXK1, HXK2 and HXK3 again all have easily measured glucose phosphorylation activity, while the three HKL proteins lack any such activity. As before, substitution of 0.1 M fructose for 2 mM glucose in the assay gave the same qualitative results (data not shown). While these protein expression assays are very useful for establishing whether a protein has catalytic activity, these preparations are not well suited for rigorous kinetic analyses.

The lack of glucose phosphorylation activity in the HKL proteins is not surprising in view of the key amino acid changes that were noted above, especially for HKL3. However, we asked whether the absence of catalytic activity in the HKL1 and HKL2 proteins might require a suite of

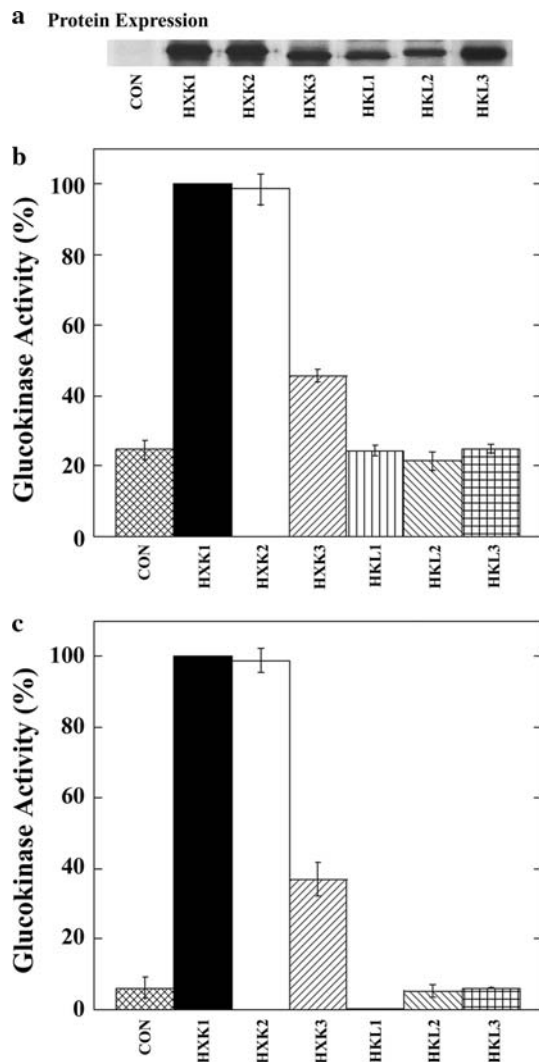


Fig. 5 Glucokinase activity measured after transient expression of different cDNAs in protoplasts from greening maize leaves. **a** The fluorogram shows protein expression levels from all of the HXK cDNAs, following their labeling with ³⁵S-Met. In this gel image, band intensities ranged from 60% (HKL2) to 100% (HKL3), as measured using ImageJ software (<http://rsb.info.nih.gov/ij/>). Target cDNA amounts were slightly adjusted for better balanced expression in subsequent transfections. **b** Full length HXK-HA family proteins were expressed and assayed directly after protoplast lysis. **c** Full length HXK-HA family proteins were expressed, then immunoprecipitated after protoplast lysis, and collected onto Protein A agarose beads prior to assay. Activity values are means ± SD of measured activities from 3 to 4 separate protoplast transfections with each construct. Activities are expressed as percent of HXK1. Notably, control values in **c** represent assay detection limits that are inherent when using enzyme-bound beads in an optical assay. CON control, without transfected HXKs

changes relative to the active enzymes or whether there might be one or just a few key changes that are sufficient to make the enzyme inactive. A further examination of the amino acid sequences revealed that the HKL1 and HKL2 proteins have equivalent changes at 39 positions relative to

conserved residues in the three proteins with catalytic activity. For example, in the core sugar binding motif, Gly173 is substituted with Ala in both HKL1 and HKL2. We therefore carried out site directed mutagenesis of both HXK1 and HKL1 in order to test whether key amino acid changes might compromise catalytic activity of HXK1 or might restore activity of HKL1. The target amino acids included Asn106 (located next to Loop 1), Gly173 (located within the sugar binding domain), Leu 251 (located within phosphate 2), insertion into HXK1 (at the corresponding position, K427) of the additional 10 amino acids present at the adenosine domain of HKL1, and removal of the 10 amino acid adenosine indel from HKL1. As a possible negative control, Cys159 of HXK1 was changed to Glu. Cys159 is one of only three amino acids among all six HXKs in which there is variation among either five or all six of the amino acids. We also included the previously altered construct S177A, which has no catalytic activity (Moore et al. 2003).

In this experiment, expressed proteins were again collected onto agarose beads in order to minimize the background glucokinase activity endogenous to the protoplast. Most of the amino acid changes did substantially impact the glucokinase activity (Fig. 6). Changing N106Y, G173A, and L251F reduced enzyme activity by 75, 90, and 95% respectively. The ten amino acid insertion into HXK1 reduced activity about 55%, and removal of the indel from HKL1 did not restore any detectable activity. Surprisingly, the C159E mutation stimulated activity twofold. We observed the same results with all constructs expressed in parallel transfections in which enzyme activities were measured directly from the lysed protoplasts (data not shown). From these data, we suggest that there likely are many non-conserved amino acids in the HKL proteins that could each compromise catalytic activity. However, we cannot exclude the possibility that particular combinatorial changes among all of those present might have compensatory effects that otherwise would result with changes to single key amino acids.

Comparison of HXK family sequences from *A. thaliana* and *A. lyrata*

The *A. lyrata* genome is currently being sequenced through the Joint Genome Institute of the Department of Energy (<http://www.jgi.doe.gov/index.html>). The North American species, *A. lyrata*, is thought to have diverged from *A. thaliana* about 5 million years ago (Koch and Matschinger 2007). In contrast to *A. thaliana*, *A. lyrata* is self-incompatible and out-crossing. We have compared the HXK family member gene sequences between *A. thaliana* and *A. lyrata* in order to identify possible regions of gene orthologs that might be undergoing differential selection. For all of the AIHXX family members, the available sequence data were

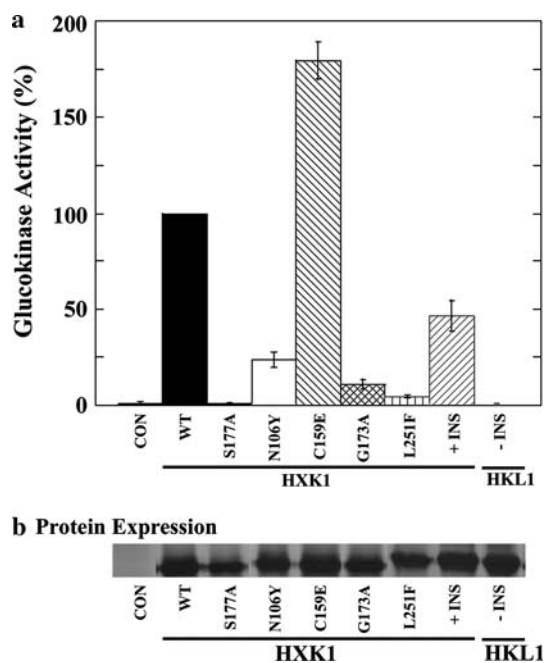


Fig. 6 Glucokinase activity of HXXK1 and HKL1 proteins modified by site directed mutagenesis. **a** Modified proteins were expressed in protoplasts from greening maize leaves, then immunoprecipitated after protoplast lysis, and collected onto Protein A agarose beads prior to assay. Activity values are means \pm SD of measured activities from 3 to 4 separate protoplast transfections with each construct, expressed as percent of WT HXXK1. **b** The fluorogram shows protein expression levels following labeling with ^{35}S -Met. Protein bands correspond precisely as labeled in **a**. In this gel image, band intensities ranged from 75% (S177A) to 100% (C159E), as measured using ImageJ software. HXXK1 + INS = HXXK1 with a 10 amino acid insertion at the adenosine domain; HKL1 – INS = HKL1 with deleted ten amino acid indel at the adenosine domain; see text for further details

sufficient to enable us to identify the homologous exons, introns, and splice sites for each gene except *AtHXXK3*. In the case of *AtHXXK3* direct sequencing resolved a discrepancy in the first draft sequence within exon 6. The HXXK gene structures are very similar in *A. lyrata* as compared with those in *A. thaliana* (Fig. 1, Supplemental Fig. 1). For example, *AtHKL3* has many short introns, as does *AtHKL3*. The most notable difference in overall gene structure is that intron 1 of *AtHKL1* is about 200 nucleotides shorter than intron 1 of *AtHKL1*.

The shared predicted amino acid identities were generally very high between all of the orthologues (e.g., >97%) except for HXXK3 and HKL3 (Table 2, Supplemental Fig. 2). For example, the predicted HKL1 and HKL2 proteins from *A. lyrata* have exactly the same indels at the adenosine binding domain as the orthologous proteins from *A. thaliana*. In contrast, the HXXK3 and HKL3 gene pairs have lower percent identity values of 89 and 93, respectively, and have some gaps in their aligned sequences. A phylogenetic tree based on all 12 genes indicates that the orthologous genes are much more related to each other,

Table 2 Comparisons of HXXK amino acid sequences from *A. thaliana* and *A. lyrata*

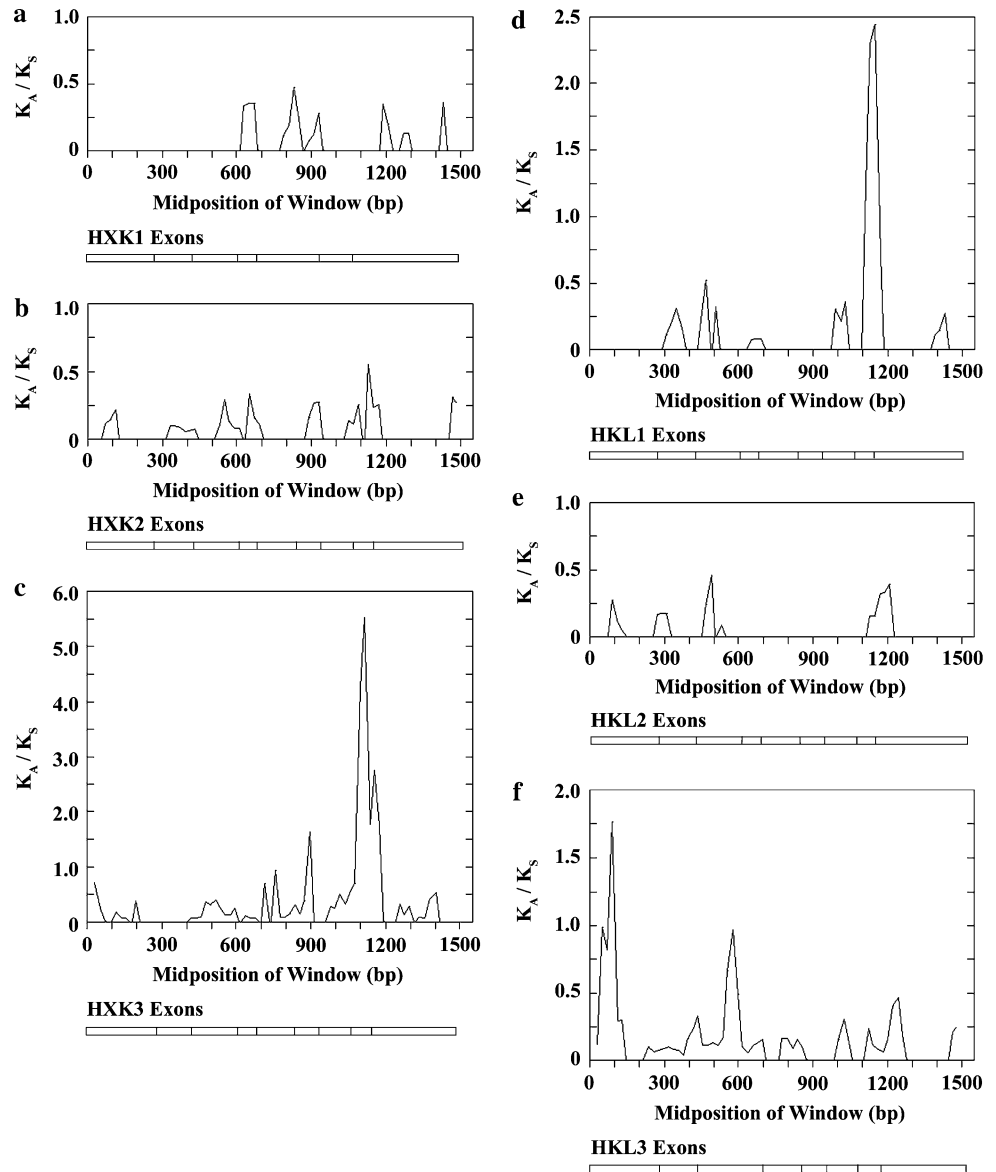
	% Identity	% Similarity	# of Gaps
HXXK1	98	99	0
HXXK2	97	98	0
HXXK3	89	94	4
HKL1	97	99	0
HKL2	98	99	0
HKL3	93	95	10

than to other family members (Supplemental Fig. 1). These relationships also support a previous more global phylogenetic analysis that includes additional plant HXXK sequences (Cho et al. 2006a). Notably, the HXXK1 and HXXK2 protein homologs have accumulated fewer amino acid changes from the nearest common ancestral sequence than have the HKL1 and HKL2 protein homologs.

One approach to test for possible differential rates of sequence evolution within a gene pair is to compare rates of synonymous and nonsynonymous nucleotide substitutions. Because of multiple indels particularly in the 3' end of the *HKL1* and *HKL2* genes, analyses of genomic DNA sequences were only possible for comparing specific *A. thaliana* and *A. lyrata* orthologs (for example, *AtHKL1* versus *AtHKL1*). Although it would be informative to consider divergence of the sequences coding for exons plus introns between the two species, the presence of non-triplet indels made this not feasible using the polymorphism and divergence analyses of DnaSP since any non-triplet indels lead to false reading frames in the sequence containing the possible sequence deletion. For example, the *HKL1* and *HKL2* genes have 13 indels total, of which only one is a triplet, though all of the indels do occur in groups that add up to a triplet. After analyzing the coding sequences for possible codon usage bias (none detected, data not shown), exon sequences for gene orthologs were then examined.

To test for the potential contribution of selection on sequence divergence between species and genes, the rates and patterns of sequence divergence of each member of the HXXK family of genes were tested between *A. thaliana* and *A. lyrata* (Fig. 7). K_A/K_S ratios of less than 1 indicate coding regions that are selectively constrained (purifying selection with higher rates of synonymous versus nonsynonymous mutation). Regions with values similar to 1 indicate neutral evolution. Regions having ratios greater than 1 indicate a higher rate of nucleotide substitutions that change the amino acid sequence versus synonymous mutations (adaptive selection; for further descriptions see Parmley and Hurst 2007). All HXXK family genes have consistently low K_A and K_S values and most have K_A/K_S values much lower than 1: HKL1 and HKL2 have overall

Fig. 7 Sliding window analyses of K_A/K_S of selected HXK and HKL loci. Analyses were done using DnaSP (Rozas et al. 2003) as described further in “Materials and methods”. **a** HXK1. **b** HXK2. **c** HXK3. **d** HKL1. **e** HKL2. **f** HKL3. Inserted exon bars correspond in length to predicted numbers of amino acids. *bp* Base pairs. Undefined estimates of K_A/K_S (because $K_S = 0$ at these sites) occurred in the following sliding windows: HXK1 (at midpoints 710, 730, 750, 1170, 1310, 1330, 1350, 1370, 1390, and 1410 bp), HXK2 (at midpoint 630 bp), HXK3 (at midpoints 216, 236, 256, 276, 336, 356, 376, 396, 736, 916, 936, 1196, and 1422 bp), HKL1 (at midpoints 490, 1070, and 1090 bp), HKL2 (at midpoints 510 and 1230 bp), and HKL3 (at midpoint 757 bp)



ratios of 0.08 and 0.05, respectively; HXK1 and HXK2 have ratios of 0.09 and 0.06, respectively; HKL3 has a ratio of 0.12; and, HXK3 has a higher ratio of 0.44. These overall low values indicate that all members of the HXK gene family are active in these two sister Arabidopsis species and are being selectively constrained to their current amino acid sequences. To further extend this analysis, sliding windows of K_A/K_S were calculated using window sizes of 60 nucleotides and a step size of 20 nucleotides (Fig. 7). The window size roughly correlates with the size of some structural elements of the proteins. Most regions of the presented genes are constrained to 0 across much of the gene and only short regions have increased K_A/K_S ratios. The primary exception is a large peak in K_A/K_S value that corresponds to exon 8 in both *HKL1* and *HXK3*. This region is somewhat upstream of the noted large indel of *HKL1*, but

in exon 8 there are no apparent functional protein motifs or key amino acids that have been described. K_A/K_S ratios of the *HXK2* gene pair were similar in profile to *HXK1*, while analysis of *HKL3* showed a prominent peak in K_A/K_S in exon 1.

Discussion

In this study, we have reported expression characteristics of the six HXK family members of *A. thaliana*. Among the HXK proteins that have hexose phosphorylation activity, AtHXK3 has not been previously recognized. We have shown that AtHXK3 is a stromal-localized protein (Fig. 3), likely expressed at low abundance primarily in roots, stems, and siliques (Fig. 2). As noted with other plastid HXKs

(Cho et al. 2006a; Kandel-Kfir et al. 2006), the presence of the transit peptide in AtHXX3 might have led to this protein not being identified in a previous complementation study of yeast cells deficient in glucose phosphorylation activity, using Arabidopsis cDNAs (Jang et al. 1997). The organ distribution of AtHXX3 is similar to that observed for tobacco plastid NtHXX2 from promoter-GUS studies (Giese et al. 2005). NtHXX2 expression was further shown to be localized in specialized tissues such as guard cells, root tips, xylem parenchyma and the vascular starch sheath. Accordingly, NtHXX2 was suggested to function primarily in starch degradation (Giese et al. 2005). On the other hand, LeHXX4 expression in plastids is perhaps relatively more wide-spread among different tomato tissues and its expression in fruits is not associated with starch degradation (Kandel-Kfir et al. 2006). As noted in both of these studies and elsewhere (Olsson et al. 2003), plastid HXXs can have an important function also in metabolizing imported glucose for production of erythrose 4-phosphate to supply the shikimic acid pathway for synthesis of some 2° metabolites. Since AtHXX3 mRNA is expressed mostly in sink tissues (Fig. 2), we suggest that this protein might have a more pronounced role in phosphorylating glucose that is imported into the plastid for biosynthetic processes.

Three of the six HXX family genes in *A. thaliana* encode proteins that do not phosphorylate glucose or fructose (Figs. 2, 5). Therefore we suggest that non-catalytic HXXs most likely do exist in plants and that particular HKL proteins might occur in most tissues. Their lack of catalytic activity would account for these also not being identified in the previous yeast complementation study by Jang et al. (1997). All three encoded HKL proteins are about 50% identical to HXX1 and they do cross-react well with a polyclonal anti-HXX1 antibody (Karve and Moore, unpublished data). The observed presence of HXX1-related antigen in leaves of *gin2-1* (Fig. 4) supports our suggestion that the HKL proteins might be expressed in leaves and elsewhere in Arabidopsis.

Several of the reported expression characteristics of HKL1 and HKL2 proteins indicate they might have a regulatory function in *A. thaliana*. First, transcripts for both genes were present at similar levels in all organs examined (Fig. 2). Function in a broad tissue context could be important if these proteins somehow affect HXX-dependent glucose signaling or possibly other wide-spread regulatory processes. Second, both HKL proteins are targeted to mitochondria, as predominantly is the case for AtHXX1 (Fig. 3). We have previously shown that mitochondrial targeted HXX1 can bind to porin in the outer membrane and can mediate at least some aspects of glucose signaling (Balasubramanian et al. 2007). In this regard, it will be important to establish whether these proteins might interact with AtHXX1. A third important finding is that AtHXX1

catalytic activity is readily compromised by any of a number of single amino acid changes (Fig. 6, Moore et al. 2003). Since there are possibly many such changes in the 1° amino acid sequences of the HKL proteins (Fig. 4, Supplemental Table 1), we surmise that the presence of catalytically defective HXXs is not simply the result of a chance mutation. Yet to be established is whether these proteins can bind glucose and with what affinity. On the one hand, the G173A change in the sugar binding domain of AtHKL1 and AtHKL2 could impact their ability to bind glucose, since the corresponding change in AtHXX1 did substantially reduce catalytic activity (Fig. 6). However, in the absence of a detailed structure analysis or ligand binding assays, one cannot rule out that glucose or even other sugars might bind with sufficient affinity in the HKL proteins as to be biologically relevant. From their primary sequence analysis, we predict that both HKL1 and HKL2 proteins do have extensive conformational flexibility as inferred by the presence of conserved key loop motifs and most of the important glycine residues at the ends of structural elements (see Supplementary Table 1). That is, the primary recognized elements required for a glucose-dependent conformational change in protein structure are largely conserved and could be exploited by cell regulatory mechanisms. The described expression and sequence characteristics support the hypothesis that the HKL1 and HKL2 proteins might affect glucose signaling or related processes.

The AtHKL3 protein has quite different expression and sequence characteristics from the other two non-catalytic AtHXXs. Among plant organs that we examined, HKL3 mRNA expression was restricted only to flowers (Fig. 2). This finding largely supports the conclusion by Claeysen and Rivoal (2007) from a survey of transcriptional profiling experiments, that *AtHKL3* is expressed most abundantly in male reproductive parts of the flower. Interestingly, among rice HXXs, OsHXX10 is expressed in pollen, but only there (Cho et al. 2006a). Our analysis of the predicted amino acid sequence of AtHKL3 (Fig. 4, Supplemental Table 1) indicated that it would not likely bind glucose or ATP as substrates, in contrast to other HXX family proteins, and that it does not have the conserved mobile loops or elements required for structural flexibility that are characteristic for this family of proteins. We suggest that this protein might have been recruited by evolutionary processes to have a much different function than those of other family members. Notably, cluster analysis of amino acid sequences of rice and Arabidopsis HXXs, indicate that the AtHKL3 protein occurs as an isolated group established prior to the separation of monocots and dicots (Cho et al. 2006a).

Non-catalytic HXXs have been identified in a variety of fungi including *Saccharomyces cerevisiae* and *A. nidulans* (Katz et al. 2000; Bernardo et al. 2007), in *Drosophila melanogaster* (Kulkarni et al. 2002), and now in *A. thaliana*.

Previous phylogenetic analysis suggests that the non-catalytic HXKs evolved independently in different lineages (Bernardo et al. 2007). Nonetheless, there are some intriguing comparisons in the primary sequences of these proteins. The fungal and fly HKL proteins were noted as having a number of altered residues in the sugar binding domain, as well as many of the fungal proteins having an indel of about 20–25 amino acids at the adenosine domain (Bernardo et al. 2007). Arabidopsis HKL1 and HKL2 proteins have much better conserved sugar binding domains (one or two substitutions) and have a similarly positioned indel of 6–10 amino acids in the adenosine domain. The predicted adenosine domain in the Arabidopsis catalytic HXKs extends eight amino acids towards the N-terminus relative to the adenosine domain in yeast HXK2 (Kuser et al. 2000). The prominent indels in the AtHKL1 and AtHKL2 proteins occur within this extension and are located four amino acids to the N-terminal side of the corresponding indel in *A. nidulans* HXKC and HXKD proteins. Whether these indel sequences in AtHKL1 and AtHKL2 actually function in adenosine binding is uncertain. Alternatively, these indels might be important for possible regulatory functions of the non-catalytic HXKs. Among the rice HXK family, all nine expressed members are thought to have catalytic activity, based on their ability to complement the HXK-deficient, triple mutant of yeast (Cho et al. 2006a). However, OsHXK3 and OsHXK10 do form a distinct phylogenetic group with AtHKL1 and AtHKL2 proteins (Cho et al. 2006a). Upon sequence inspection, both rice proteins also contain a similar indel of nine amino acids at the same aligned position near the beginning residue of the predicted adenosine domain (data not shown). The sugar binding domains of OsHXK3 and OsHXK10 do not have the homologous substitution of A for G as do AtHKL1 and AtHKL2 (LAFTF-SFP-Q), though these domains do have 1 or 2 changes elsewhere.

The different expressed features within the HXK family of a given plant or fungal species reflect some remarkable apparent evolutionary trends. In both *A. thaliana* and *S. cerevisiae*, there occurs at least one moonlighting HXK which has separable functions as both a metabolic catalyst and a glucose sensor/transducer (Moreno and Herrero 2002; Moore et al. 2003). Additionally, there occur proteins within the families that have apparent specialized metabolic roles or specialized regulatory roles. The metabolic only catalysts include yeast HXK1 and glucokinase1 (Santangelo 2006), and we suggest might also include plant plastid HXKs. Among non-catalytic HKL proteins, there are two HKL proteins in *S. cerevisiae* (Bernardo et al. 2007), one of which, EMI2, is required for induction of a meiosis-specific transcription factor (Daniel 2005). In *A. nidulans*, the non-catalytic proteins AnHXKC and AnHXKD both are thought not to have a role in glucose signaling, but genetic

evidence indicates that instead both are negative regulators for a secreted extracellular protease during carbon starvation (Bernardo et al. 2007). AnHXKC is associated with mitochondria, while AnHXKD is present in the nucleus. As noted above, it is yet to be demonstrated whether the Arabidopsis HKL proteins do have specialized regulatory functions. Notably, AnHXKD is transcriptionally induced by carbon starvation, yet neither AtHKL1 nor AtHKL2 are apparently induced by starvation conditions (see Supplementary Tables in Baena-González et al. 2007).

While non-catalytic HXKs are present in the relatively distant lineages of fungi and at least Arabidopsis, we also interrogated the HXK nucleotide sequences of *A. thaliana* and *A. lyrata* to test for possible different relative rates of evolution among the respective orthologs. HXK3 has a higher K_A/K_S value than do all the other loci examined. This indicates those amino acid sequences are less constrained overall for accumulating changes. Notably, the divergence in amino acid sequences between AtHXK3 and AtHXK3 orthologs also is greater than for the other catalytic HXKs (Table 2). We suggest that the *HXK3* gene is evolving at an increased rate relative to the other genes for the catalytic HXK proteins and for the HKL proteins. In contrast, among the orthologs for HXK1, HXK2, HKL1, and HKL2, the genes are evolving overall at similar rates. However, both HKL1 and HXK3 orthologs have a pronounced peak of K_A/K_S ratio within exon 8 (Fig. 7). This indicates a possible adaptive selection process for these sequences. The significance of this observation is not yet clear, but does warrant investigation.

For the global plant HXK gene family, it will be interesting to establish whether the noted indel that occurs in Arabidopsis HKL1 and HKL2 might be useful as a molecular marker for phylogeny studies. To that end, it is important to verify if OsHXK3 and OsHXK10 actually do have catalytic activity as suggested (Cho et al. 2006a), since their sequences also contain a similar indel at the homologous position as occurs in the reported non-catalytic HKL proteins.

Acknowledgments We appreciate technical assistance from Ms. Betsy Metters. We thank also Dr. Amy Lawton-Rauh for consultation on nucleotide sequence analyses and for supplying *A. lyrata* tissue and DNA. This work was supported by the U.S. Department of Agriculture-National Research Initiative Plant Biochemistry Program (grant no. 2001-035318) and the South Carolina Agricultural Experiment Station (technical contribution No. 5389 of the Clemson University Experiment Station). This material is based upon work supported by CSREES/USDA, under project number SC-170090.

References

- Balasubramanian R, Karve A, Kandasamy M, Meagher RB, Moore BD (2007) A role for F-actin in hexokinase-mediated glucose signaling. *Plant Physiol* 145:1423–1434

- Baena-González E, Rolland F, Thevelein JM, Sheen J (2007) A central integrator of transcription networks in plant stress and energy signaling. *Nature* 448:938–943
- Bernardo SMH, Gray K-A, Todd RB, Cheetham BF, Katz ME (2007) Characterization of regulatory non-catalytic hexokinases in *Aspergillus nidulans*. *Mol Genet Genomics* 277:519–532
- Bork P, Sander C, Valencia A (1992) An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. *Proc Natl Acad Sci USA* 89:7290–7294
- Cárdenas ML, Cornish-Bowden A, Ureta T (1998) Evolution and regulatory role of the hexokinases. *Biochim Biophys Acta* 1401:242–264
- Cho J-I, Ryoo N, Ko S, Lee S-K, Lee J, Jung K-H, Lee Y-H, Bhoo SH, Winderickx J, An G, Hahn T-R, Jeon J-S (2006a) Structure, expression, and functional analysis of the hexokinase gene family in rice (*Oryza sativa* L.). *Planta* 224:598–611
- Cho Y-H, Yoo S-D, Sheen J (2006b) Regulatory functions of nuclear hexokinase1 complex in glucose signaling. *Cell* 127:579–589
- Claeysen E, Rivoal J (2007) Isozymes of plant hexokinase: occurrence, properties and functions. *Phytochem* 68:709–731
- Damari-Weissler H, Kandel-Kfir M, Gidoni D, Mett A, Belausov E, Granot D (2006) Evidence for intracellular spatial separation of hexokinases and fructokinases in tomato plants. *Planta* 224:1495–1502
- Damari-Weissler H, Ginzburg A, Gidoni D, Mett A, Krassovskaya I, Weber APM, Belausov E, Granot D (2007) Spinach SoHKK1 is a mitochondria-associated hexokinase. *Planta* 226:1053–1058
- Daniel J (2005) Sir-dependent downregulation of various aging processes. *Mol Genet Genomics* 274:539–547
- da-Silva WS, Rezende GL, Galina A (2001) Subcellular distribution and kinetic properties of cytosolic and non-cytosolic hexokinases in maize seedling roots: implications for hexose phosphorylation. *J Exp Bot* 52:1191–1201
- Doehlert DC (1989) Separation and characterization of four hexose kinases from developing maize kernels. *Plant Physiol* 89:1042–1048
- Francis D, Halford NG (2006) Nutrient sensing in plant meristems. *Plant Mol Biol* 60:981–993
- Giese J-O, Herbers K, Hoffmann M, Klösgen RB, Sonnewald U (2005) Isolation and functional characterization of a novel plastidic hexokinase from *Nicotiana tabacum*. *FEBS Lett* 579:827–831
- Graham JWA, Williams TCR, Morgan M, Fernie AR, Ratcliffe RG, Sweetlove LJ (2007) Glycolytic enzymes associate dynamically with mitochondria in response to respiratory demand and support substrate channeling. *Plant Cell* 19:3723–3738
- Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/NT. *Nucleic Acids Symp Ser* 41:95–98
- Han S, Kim D (2006) AtRTPrimer: database for Arabidopsis genome-wide homogeneous and specific RT-PCR primer-pairs. *BMC Bioinform* 7:179
- Heazlewood JL, Tonti-Fillippini JS, Gout AM, Day DA, Whelan J, Millar AH (2004) Experimental analysis of the Arabidopsis mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. *Plant Cell* 16:241–256
- Heyer AG, Raap M, Schroeder B, Marty B, Willmitzer L (2004) Cell wall invertase expression at the apical meristem alters floral, architectural, and reproductive traits in *Arabidopsis thaliana*. *Plant J* 39:161–169
- Hwang I, Sheen J (2001) Two-component circuitry in Arabidopsis cytokinin signal transduction. *Nature* 413:383–389
- Jang J-C, León P, Zhou L, Sheen J (1997) Hexokinase as a sugar sensor in higher plants. *Plant Cell* 9:5–19
- Jang J-C, Sheen J (1994) Sugar sensing in higher plants. *Plant Cell* 6:1665–1679
- Jeffery CJ (2004) Molecular mechanisms for multitasking: recent crystal structures of moonlighting proteins. *Curr Opin Struct Biol* 14:663–668
- Kandel-Kfir M, Damari-Weissler H, German MA, Gidoni D, Mett A, Belausov E, Petreikov M, Adir N, Granot D (2006) Two newly identified membrane-associated and plastidic tomato HXKs: characteristics, predicted structure and intracellular localization. *Planta* 224:1341–1352
- Katz ME, Masoumi A, Burrows SR, Shirliff CG, Cheetham BF (2000) The *Aspergillus nidulans xprF* gene encodes a hexokinase-like protein involved in the regulation of extracellular proteases. *Genetics* 156:1559–1571
- Koch MA, Matschinger M (2007) Evolution and genetic differentiation among relatives of *Arabidopsis thaliana*. *Proc Natl Acad Sci (USA)* 104:6272–6277
- Kojima H, Suzuki T, Kato T, Enomoto K-I, Sato S, Kato T, Tabata S, Sáez-Vasquez J, Echeverría M, Nakagawa T, Ishiguro S, Nakamura K (2007) Sugar-inducible expression of the nucleolin-1 gene of *Arabidopsis thaliana* and its role in ribosome synthesis, growth and development. *Plant J* 49:1053–1063
- Kovtun Y, Chiu W-L, Zeng W, Sheen J (1998) Suppression of auxin signal transduction by a MAPK cascade in higher plants. *Nature* 395:716–720
- Kulkarni GV, Deobagkar DN, Deobagkar DD (2002) Hexokinase like protein from *Drosophila melanogaster*. *Dros Inf Serv* 85:65–68
- Kuser PR, Krauchenco S, Antunes OAC, Polikarpov I (2000) The high resolution crystal structure of yeast hexokinase PII with the correct primary sequence provides new insights into its mechanism of action. *J Biol Chem* 275:20814–20821
- Moore B, Zhou L, Rolland F, Hall Q, Cheng W-H, Liu Y-X, Hwang I, Jones T, Sheen J (2003) Role of the Arabidopsis glucose sensor HXK1 in nutrient, light, and hormonal signaling. *Science* 300:332–336
- Moore B (2004) Bifunctional and moonlighting enzymes: lighting the way to regulatory control. *Trends Plant Sci* 9:221–228
- Moreno F, Herrero P (2002) The hexokinase 2-dependent glucose signal transduction pathway of *Saccharomyces cerevisiae*. *FEMS Microbiol Rev* 26:83–90
- Morton BR (1993) Chloroplast DNA codon use: evidence for selection at the *psbA* locus based on tRNA availability. *J Mol Evol* 37:273–280
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426
- Olsson T, Thelander M, Ronne H (2003) A novel type of chloroplast stromal hexokinase is the major glucose-phosphorylating enzyme in the moss *Physcomitrella patens*. *J Biol Chem* 278:44439–44447
- Osuna D, Usadel B, Morcuende R, Gibon Y, Bläsing OE, Höhne M, Günter M, Kamlage B, Trethewey R, Scheible W-R, Stitt M (2007) Temporal responses of transcripts, enzyme activities and metabolites after adding sucrose to carbon-deprived Arabidopsis seedlings. *Plant J* 49:463–491
- Parnley JL, Hurst LD (2007) How common are intragene windows with KA > KS owing to purifying selection on synonymous mutations? *J Mol Evol* 64:646–655
- Price J, Laxmi A, St Martin SK, Jang J-C (2004) Global transcription profiling reveals multiple sugar signal transduction mechanisms in Arabidopsis. *Plant Cell* 16:2128–2150
- Rolland F, Baena-Gonzalez E, Sheen J (2006) Sugar sensing and signaling in plants: conserved and novel mechanisms. *Annu Rev Plant Biol* 57:675–709
- Rolland F, Moore B, Sheen J (2002) Sugar sensing and signaling in plants. *Plant Cell* 14(Suppl):S185–S205
- Rontein D, Dieuaide-Noubhani M, Dufourc EJ, Raymond P, Rolin D (2002) The metabolic architecture of plant cells. *J Biol Chem* 277:43948–43960

- Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497
- Santangelo GM (2006) Glucose signaling in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* 70:253–282
- Sheen J, Zhou L, Jang J-C (1999) Sugars as signaling molecules. *Curr Opin Plant Biol* 2:410–418
- Smeeckens S (2000) Sugar-induced signal transduction in plants. *Annu Rev Plant Physiol Plant Mol Biol* 51:49–81
- Tamura K, Dudley J, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 10:1093
- Tocquin P, Corbesier L, Havelange A, Pieltain A, Kurtem E, Bernier G, Périlleux C (2003) A novel high efficiency, low maintenance, hydroponic system for synchronous growth and flowering of *Arabidopsis thaliana*. *BMC Plant Biol* 3:2
- Wright F (1990) The “effective number of codons” used in a gene. *Gene* 87:23–29
- Xiao W, Sheen J, Jang J-C (2000) The role of hexokinase in plant sugar signal transduction and growth and development. *Plant Mol Biol* 44:451–461
- Yoo S-D, Cho Y-H, Sheen J (2007) *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat Prot* 2:1565–1572