Plant, Cell & Environment

Plant, Cell and Environment (2012) **35**, 2143–2154 doi: 10.1111/j.1365-3040.2012.02544.x

Tapping natural variation at functional level reveals allele specific molecular characteristics of potato invertase *Pain-1*

ASTRID M. DRAFFEHN[1,*], PAWEL DUREK[2,†], ADRIANO NUNES-NESI[2,‡], BENJAMIN STICH[1], ALISDAIR R. FERNIE[2] & CHRISTIANE GEBHARDT[1]

[1]*Max Planck Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, 50829 Köln, Germany and* [2]*Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam, Germany*

ABSTRACT

Biochemical, molecular and genetic studies emphasize the role of the potato vacuolar invertase *Pain-1* in the accumulation of reducing sugars in potato tubers upon cold storage, and thereby its influence on the quality of potato chips and French fries. Previous studies showed that natural *Pain-1* cDNA alleles were associated with better chip quality and higher tuber starch content. In this study, we focused on the functional characterization of these alleles. A genotype-dependent transient increase of total *Pain-1* transcript levels in cold-stored tubers of six different genotypes as well as allele-specific expression patterns were detected. 3D modelling revealed putative structural differences between allelic *Pain-1* proteins at the molecule's surface and at the substrate binding site. Furthermore, the yeast *SUC2* mutant was complemented with *Pain-1* cDNA alleles and enzymatic parameters of the heterologous expressed proteins were measured at 30 and 4 °C. Significant differences between the alleles were detected. The observed functional differences between *Pain-1* alleles did not permit final conclusions on the mechanism of their association with tuber quality traits. Our results show that natural allelic variation at the functional level is present in potato, and that the heterozygous genetic background influences the manifestation of this variation.

Key-words: cold-sweetening; *Pain-1* invertase.

INTRODUCTION

The biosynthesis and degradation of starch and sugars, the major products of photosynthetic carbon fixation, is one of the best studied processes in plants at the molecular level. At low temperatures, plant metabolism undergoes a shift in the balance between starch degradation and glycolysis, which leads to sucrose accumulation. Sucrose is then enzymatically converted into the reducing sugars glucose and fructose. The accumulation of sugars as osmoprotective compounds in plant tissues in response to low temperatures is a widespread phenomenon, often referred to as cold-induced sweetening (Müller-Thurgau 1882).

In mature tubers of potato (*Solanum tuberosum*), invertases (EC 3.2.1.26) play a functional role in the accumulation of the reducing sugars glucose and fructose in response to cold storage. Several studies demonstrated increasing transcript accumulation (Zhou *et al*. 1994; Zrenner, Schuler & Sonnewald 1996; Bagnaresi *et al*. 2008) as well as increased enzymatic invertase activity upon exposure to low temperature (Rorem & Schwimmer 1963; Pressey & Shaw 1966). Down-regulation of invertase expression by antisense or siRNA constructs reduced the sugar accumulation in transgenic lines (Bhaskar *et al*. 2010). In plants, three different invertase isoforms can be distinguished based on solubility, subcellular localization, pH optimum and isoelectric point: (1) vacuolar; (2) cell wall-bound or apoplastic; and (3) neutral invertases (Tymowska-Lalanne, Kreis & Callow 1998; Roitsch & González 2004).

Potato invertase genes and cDNAs encoding apoplastic and vacuolar invertases have been cloned and characterized (Hedley *et al*. 1993, 1994; Zhou *et al*. 1994; Zrenner *et al*. 1996). Three invertase loci, *Pain-1*, Inv_{ap}-*a* and Inv_{ap}-*b*, have been mapped on potato chromosomes III, X and IX using invertase cDNA sequences as molecular markers (Chen, Salamini & Gebhardt 2001). The *Pain-1* locus consists of a single copy gene with a size of around 4 kb. The loci Inv_{ap}-*a* and Inv_{ap}-*b* each consist of two tandem duplicated genes located within 17 and 9 kbp genomic sequence, respectively (Draffehn *et al*. 2010).

Natural variation of starch and sugar content of potato tubers is controlled by multiple genetic and environmental factors and is therefore quantitative. Progress at the genetic level, together with the fact that carbohydrate metabolism is one of the best studied plant processes at the functional level, renders tuber starch and sugar content model traits for exploring the candidate gene approach in order to identify the molecular basis of quantitative trait loci (QTL) in

Correspondence: A. M. Draffehn. Fax: +49 (0) 221 506 2413; e-mail: draffehn@mpipz.mpg.de

Current addresses: *LIMES Institute, Carl-Troll-Straße 31, 53115 Bonn, Germany; †Institute of Pathology, Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10118 Berlin, Germany; ‡Universidade Federal de Viçosa, Departamento de Biologia Vegetal, 36570-000 Viçosa, MG, Brazil.

potato. Being a staple food, potato is not only grown for table use but is also processed into French fries and potato chips. Potato chip colour is an important tuber quality trait, which is correlated with the amount of reducing sugars in tubers and in this respect strongly influenced by carbohydrate metabolism. A high-reducing sugar content in raw tubers causes major problems during processing due to the non-enzymatic Maillard reaction (Shallenberger, Smith & Treadway 1959), which changes chips and French fries' colour from light yellow to dark brown, and causes a bitter taste (Roe, Faulks & Belsten 1990). Therefore, the development of processing varieties with low potential for reducing sugar accumulation, especially when stored at low temperatures, which is required to prevent tuber sprouting, is an ongoing activity in the breeding industry.

QTL mapping in experimental diploid mapping populations and association mapping in populations of tetraploid potato varieties and breeding clones revealed co-localization of all three invertase loci with QTL for tuber starch content and/or chip colour (Li *et al*. 2005b, 2008). The strongest associations with tuber quality traits were found with DNA polymorphisms at the *Pain-1* locus, which encodes a vacuolar invertase. This observation indicates that either the invertase locus itself or a physically linked locus is causal for the phenotypic variance. A causal relationship implies functional differences between natural invertase alleles. In the case of the tomato apoplastic invertase gene *Lin5,* which is causal for a fruit sugar content QTL, it has been shown that biochemical differences of *Lin5* natural alleles explain the QTL effect (Fridman *et al*. 2004). The cloning and characterization of full-length vacuolar invertase cDNA alleles from representative potato genotypes revealed large intraspecific molecular diversity. Eleven different cDNA alleles were identified, including two alleles associated with tuber quality traits (Draffehn *et al*. 2010).

As outlined previously, biochemical and molecular studies emphasize a role of invertase genes in the formation of reducing sugars in potato tubers. Genetic studies suggest that allelic variation at invertase loci contributes to the natural variation of tuber starch and sugar content, and thereby chip colour. In this study, we focused therefore on the functional analysis of natural *Pain-1* alleles, and addressed the questions (1) whether and how sequence polymorphisms of *Pain-1* alleles translate into structural and functional variation of the encoded proteins; and (2) whether we can identify functional and/or structural properties that distinguish associated from non-associated alleles. Therefore, we conducted comparative 3D protein modelling and determined enzymatic parameters of *Pain-1* cDNA alleles expressed in yeast. In addition, allele specific expression differences were analysed by quantifying the transcript levels of *Pain-1* alleles.

MATERIALS AND METHODS

Plant material, growing conditions and tuber cold storage

Potato cultivars 'Satina', 'Diana', and 'Theresa' were obtained from SaKa Pflanzenzucht GbR (breeding station Windeby, Germany) and Böhm-Nordkartoffel Agrarproduktion (breeding station Ebstorf, Germany). The diploid genotypes P18 and P40 were included in the analysis as parents of several experimental mapping populations developed at the MPI for Plant Breeding Research (Menendez *et al*. 2002). The analysed genotypes were the sources of 11 different *Pain-1* cDNA alleles (Draffehn *et al*. 2010). Names, accession numbers and origin of the *Pain-1* alleles are shown in Table 1. Plants were grown either in a greenhouse (day temperature 20–24 °C; night temperature 18 °C; additional light from 0600 to 2100 h) or in a Saran house under natural light and temperature conditions from May to September. Mature tubers were harvested after complete plant senescence. For cold storage experiments, the tubers were kept at 4 °C in the dark for 0 to 4 weeks.

Table 1. Overview of the analysed *Pain-1* alleles and their genomic dosages as determined by pyrosequencing

Genotype	Ploidy[a]	*Pain-1* cDNA alleles	Accession number	SNP allele quantified by pyrosequencing[c]	Genomic dosage (%)
'Satina'	4n	*Pain1-Sa*[b]	JN661854	A_{1544}	25 (simplex)
		Pain1-Sb	JN661855	C_{1544}	75 (triplex)
'Diana'	4n	*Pain1-Da*[b]	JN661852	A_{1544}	25 (simplex)
		Pain1-Db	JN661856	C_{1596}	50 (duplex)
		Pain1-Dc	JN661859	T_{1574}	25 (simplex)
'Theresa'	4n	*Pain1-Tb*	JN661857	A_{612}	75 (triplex)
		Pain1-Tc	JN661860	G_{612}	25 (simplex)
P18	2n	*Pain1-P18a*	JN661853	A_{1544}	50 (heterozygous)
		Pain1-P18b	JN661858	C_{1544}	50 (heterozygous)
P40	2n	*Pain1-P40d1*	JN661861	A_{1267}	50 (heterozygous)
		Pain1-P40d2	JN661862	G_{1267}	50 (heterozygous)

[a]4n = tetraploid, 2n = diploid.
[b]Alleles *Sa* and *Da* were positively associated with chip colour and tuber starch content (Li *et al.* 2008; Draffehn *et al.* 2010).
[c]Additional information about pyrosequencing is listed in Supporting Information Table S1.

Yeast strains and growing conditions

The yeast strains were grown at 30 °C on solid (2% agar) or liquid (flasks shaking at 200 r.p.m.) YPD medium (Carl Roth GmbH, Karlsruhe, Germany) supplemented either with 2% glucose or sucrose as carbohydrate source depending on the genotype of the strain. The invertase deficient mutant strain *SUC2* (Gozalbo & Hohmann 1989) and the wild-type reference strain *FY1679* were obtained from EUROSCARF (European *Saccharomyces Cerevisiae* Archives for Functional Analysis, Frankfurt, Germany).

Invertase cDNA constructs and complementation of the yeast SUC2 mutant

For heterologous yeast expression, the *Pain-1* cDNA alleles were integrated in the 112 A1 NE yeast expression vector (Riesmeier, Willmitzer & Frommer 1992), which includes the *TRP1* gene as selectable marker. Pain-1/112 A1 NE constructs were produced by introducing *Not*I and *Bam*HI restriction sites using proof-reading high-fidelity *Taq*-Polymerase (Roche, Mannheim, Germany) and the primer combination forward: 5′-CCCCGCGGCCGCATGGCC ACGCAGTACC-3′ for *Not*I and reverse: 5′-CCCCG GATCCGATGAATTACAAGTCTTGCAAGGG-3′ for *Bam*HI. *Pain-1* PCR products and the yeast expression vector 112 A1 NE were digested with *Not*I and *Bam*HI and ligated using standard protocols. Pain-1/112 A1 NE constructs were verified by sequencing.

The invertase deficient yeast strain *SUC2* was transformed using a simplified method (Gietz & Schiestl 1995). Transformants growing on selective SD plates [6.7 g l^{-1} yeast nitrogen base without amino acids, 10x –Trp DO supplement (Clontech Laboratories, Inc., Mountain View, CA, USA), 2% agar, 2% sucrose] were transferred to fresh SD plates and checked for correct inserts by colony PCR using the primer pair 5′- CTCACCATATCCGCAA TGAC-3′ and 5′-CTTGAGTAACTCTTTCCTGTAGG TC-3′. Complemented strains were grown on selective SD media with sucrose as sole carbohydrate source.

Protein extraction from yeast cells

Protein extraction was carried out as described (Fridman *et al*. 2004) with slight modifications. Briefly, yeast cells were solubilized by adding acid-washed glass beads (425–600 micron, Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany). Cellular debris was subsequently removed by centrifugation (5000 *g*, 10 min, at 4 °C). The supernatant was loaded on a PD-10 column (GE Healthcare, Buckinhamshire, UK), equilibrated with protein extraction buffer (Fridman *et al*. 2004) without protease inhibitors, and eluted in 2 mL of the same buffer. Protein concentration was measured with Bradford dye reagent (Protein assay, Bio-Rad, Hercules, CA, USA).

Enzymatic invertase assay

Invertase enzymatic activity was measured using a modified protocol (Zrenner *et al*. 1995). Assay solution (100 μL) contained 20 μg total yeast protein, 20 mM NaOAc (pH 4.7) and 2.5–120 mM sucrose. The assay solution was incubated either for 1 h at 30 °C or for 1 h 30 min at 4 °C. The reaction was stopped by adding 10 μL 1 M $NaHPO_4$ (pH 7.2) and heating at 95 °C for 10 min. Controls containing the same reaction mixture were heat inactivated without incubation. Assay conditions were adjusted to be linear over time for at least 90 min and to depend linearly on the amount of yeast protein added up to 50 μg. The amount of glucose formed was measured using hexokinase (HK) and glucose-6-phosphate dehydrogenase (G6PDH) in a coupled enzymatic assay (Bondar & Mead 1974). The components of a 300 μL assay solution were: 100 mM Imidazol (pH 6.9 HCl), 5 mM $MgCl_2$, 2 mM NADP+, 1 mM ATP, 2U G6PDH (Roche Diagnostic GmbH, Mannheim, Germany) and 10 μL invertase assay solution. Absorbance was measured at 340 nm, and the mmol glucose h^{-1}*mg protein calculated. The apparent Michaelis constant K_m and the maximal velocity v_{max} were estimated from Lineweaver–Burk plots.

Three-dimensional protein modelling

Modelling of the molecular structure of vacuolar invertase alleles was based on the 3D and crystal structure of cell wall-bound invertase *AtcwINV1* from *Arabidopsis thaliana* (Verhaest *et al*. 2006) [Protein Data Bank code (PDB) 2AC1]. The models were comparative, superimposing two allelic sequences by Swiss-PDB Viewer (Guex & Peitsch 1997). Differences between structures were marked by colour code as judged by the root mean square between the structures and visualized by PyMol (Delano 2002). Models of the putative binding site included the substrate sucrose, and all amino acids in a 4 Å distance to sucrose were visualized. Binding site models were based on the crystal structure of *AtcwINV1* with sucrose (Lammens *et al*. 2008) (PDB 2QQU). In addition to the structural visualization of amino acid exchanges, also the electrostatic potential (EP) of the molecules was mapped at pH 4.7 mimicking vacuolar conditions. All models were predicted by homology modelling applying the HHpred interactive server for structure prediction (Söding, Biegert & Lupas 2005) and MODELLER (Sali & Blundell 1993) (Supporting Information Files S1–S6). Subsequently, the models were prepared for continuum electrostatics calculation utilizing the PDB2PQR package (Version: 1.3.0) (Dolinsky *et al*. 2004). For calculation, the AMBER force field and the protonation states at pH 4.7 were used (Li, Robertson & Jensen 2005a). The isoelectric surfaces were computed by the adaptive Poisson–Boltzmann Solver (APBS) (Baker *et al*. 2001) utilizing standard parameters at the temperature of 298 K. The comparison of the isoelectric surfaces revealed no significant changes upon lowering the temperature parameter. The isoelectric surfaces were visualized by PyMol.

Semi-quantitative RT-PCR

Total RNA was extracted from powdered tuber tissue using the Plant RNA Isolation Kit from Invitrogen (Karlsruhe,

Germany) following the manufacturer's protocol. Further RNA purification, removal of genomic DNA contamination and first-strand cDNA synthesis was performed as described in (Draffehn *et al*. 2010). PCR was performed using 2.5 μL cDNA solution diluted 1:100 and the Power SYBR® Green PCR Master Mix (Applied Biosystems, Warrington, UK) according to the supplier's protocol. *Pain-1* transcript levels were quantified using 10 pmol of the primer pair 5′-GGGACCATTTGGTGTCGTTG-3′and 5′-GCAAAGCTCTCCACAATTGAG-3′, and the following PCR conditions: initial denaturation at 95 °C for 10 min followed by 50 cycles of 95 °C for 15 s, 55 °C for 30 s and 72 °C for 45 s. At the end of the PCR, reaction were tested for undesired primer dimer formation by melting curve analysis (55 to 95 °C with a heating rate of 0.1 °Cs^{-1} and continuous fluorescence measurement). Average transcript levels were calculated from two to three biological and two technical replicates. Expression levels were normalized against *ef1-α* (Nicot *et al*. 2005).

Allele specific expression

The expression of specific *Pain-1* alleles was monitored by pyrosequencing (Ronaghi *et al*. 1996). PCR for biotin labelled products was performed in 25 μL containing 1× PCR buffer (100 mM Tris-HCl pH 8; 500 mM KCl, 1% Triton X-100), 3 mM $MgCl_2$, 25 mM dNTP, 10 pmol primers and 0.05U μL^{-1} *Taq*-DNA Polymerase [Fast Start High Fidelity PCR System (Roche)]. Cycling conditions were: 5 min initial denaturation at 94 °C, followed by 40 cycles of 30 s denaturation at 94 °C, 45 s annealing at 60 °C and 1 min extension at 72 °C. Reactions were finished by 10 min incubation at 72 °C. PCR products were examined on ethidium bromide stained agarose gels. The pyrosequencing procedure was performed as described (Royo, Hidalgo & Ruiz 2007) with slight modifications. The reaction was set up with 20 μL PCR product, 15 μL H_2O (LiChrosolv, Merck, Rahway, NJ, USA), 40 μL binding buffer and 5 μL Streptavidin-coated Super Paramagnetic beads (Invitrogen). Allele-specific sequencing primers are listed in Supporting Information Table S1. Alleles were quantified as the percentage of an allele specific nucleotide, where 100% was the sum of the peaks originating from all allele specific nucleotides per genotype. The distribution (in %) of these allele-specific nucleotides was automatically generated by the Biotage software package (Uppsala, Sweden). Allele specific expression analysis was performed with three biological replicates in three technical replicates.

RESULTS

Genotypic variation of total Pain-1 transcript levels in tubers during cold storage

Transcripts of vacuolar invertase accumulate in potato tubers during cold storage (Zhou *et al*. 1994; Zrenner *et al*. 1996; Bagnaresi *et al*. 2008). By monitoring *Pain-1* transcripts in three tetraploid and two diploid genotypes over a cold storage period of four weeks using semi-quantitative RT-PCR, we observed genotypic diversity with respect to transcript quantity, whereas the kinetics of transcript accumulation were similar (Fig. 1). With the exception of cv 'Diana', *Pain-1* transcript accumulation reached its maximum after 1 week of cold storage (1 wcs), and declined over the following 3 weeks. In 'Diana', *Pain-1* transcripts peaked later at 2 wcs and declined afterwards.

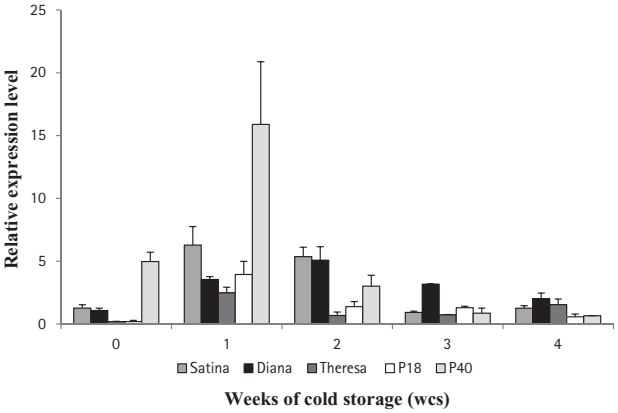

Figure 1. Genotypic variation of *Pain-1* total expression in potato tubers stored from 0 to 4 weeks at 4 °C. Standard deviations are derived from three technical replicates each of two biological replicates. The expression level was normalized against *ef1α*. The expression of the samples was calculated relative to the expression of cultivar 'Satina' at 0 weeks of cold storage (wcs), which was set as reference to the value '1'. 0 = no cold storage, 1 = one wcs, 2 = two wcs, 3 = three wcs, 4 = four wcs.

Before the onset of cold storage (0 wcs), *Pain-1* transcript levels showed already large variation between genotypes (Fig. 1). At 0 wcs, the transcript level in the diploid genotype P40 was 25-fold higher when compared with P18 and cv 'Theresa', and fivefold higher than in cvs 'Satina' and 'Diana'. The transcriptional up-regulation during cold storage compared with 0 wcs was around fourfold for 'Satina', 'Diana' and P40, while 'Theresa' transcripts increased 13-fold and P18 transcripts increased 21-fold. The overall total transcript abundance was lowest in 'Theresa', followed by P18, 'Diana', 'Satina' and finally P40, which possessed the largest amount of *Pain-1* transcripts.

Expression of specific Pain-1 alleles in tubers during cold storage

Semi-quantitative RT-PCR measured the sum of the transcriptional activity of all *Pain-1* alleles present in corresponding genotypes. In order to examine whether individual *Pain-1* alleles are preferentially expressed in cold stored tubers of the different genotypes, we developed a specific pyrosequencing assay for each of the 11 *Pain-1* alleles based on allele-specific single nucleotide polymorphisms (SNPs, Table 1). The *Pain-1* alleles *Sa* and *Da* in cvs 'Satina' and 'Diana' were of particular interest as they are positively associated with tuber quality traits, whereas *Sb,*

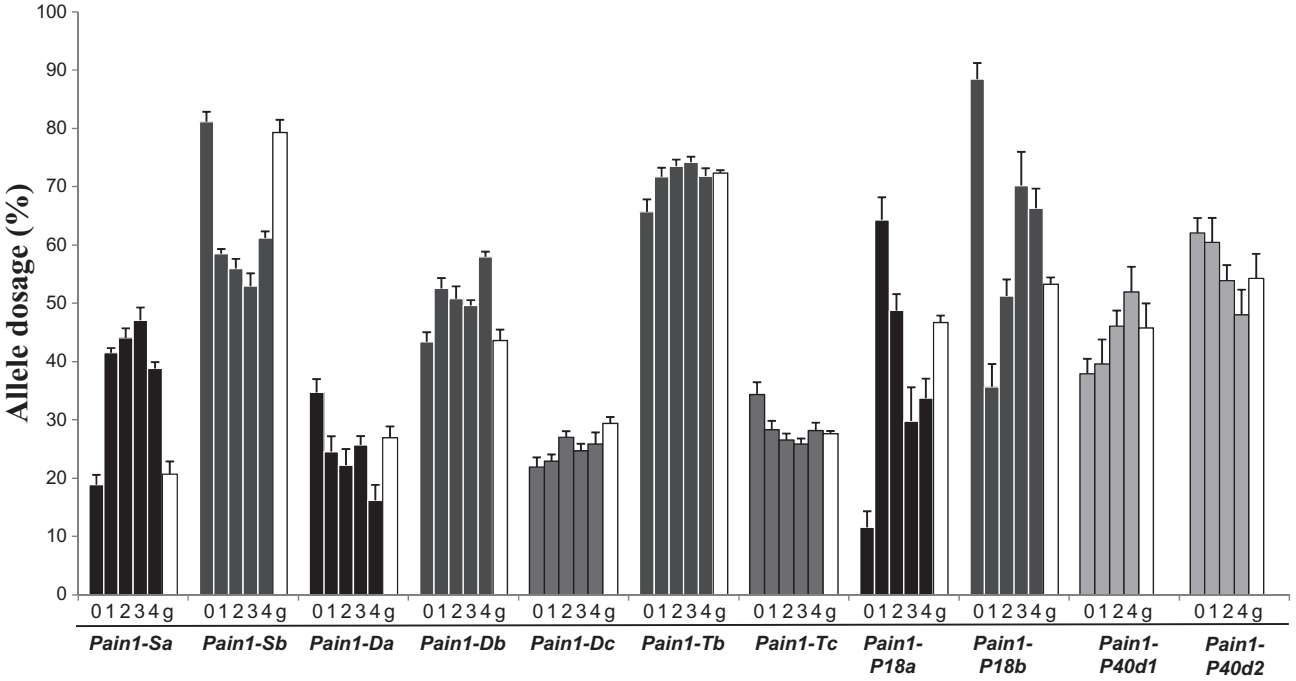

Figure 2. *Pain-1* allele-specific expression in tubers during 4 weeks of cold storage (wcs) at 4 °C. The expression level of each allele is shown as the percentage of allele-specific nucleotides (Table 1). One hundred percent is defined as the sum of the peaks originating from all allele-specific nucleotides per genotype. The genomic allele dosage is shown as white bar to the right of each allele. Alleles in the same phylogenetic group *a*, *b*, *c* and *d* are coloured the same. Standard deviations were derived from three technical replicates each of three biological replicates. 0 = no cold storage, 1 = one wcs, 2 = two wcs, 3 = three wcs, 4 = four wcs; g = genomic allele dosage. Due to technical reasons, for the diploid genotype P40, no allele-specific expression could be determined at 3 wcs.

Db, Tb, Dc and *Tc* alleles showed no significant associations (Draffehn *et al*. 2010). The status of the four alleles isolated from P18 and P40 regarding association with tuber quality traits is unknown. The pyrosequencing assays were first performed with genomic DNA as template to determine the allele dosage in the tetraploid cultivars. The associated alleles *Sa* and *Da* were present in single dose in cvs 'Satina' and 'Diana', respectively. The dosage of the remaining five alleles of the tetraploid cultivars varied between simplex and triplex (Table 1). Using tuber cDNA as template, we then quantified the allele specific expression in cold stored tubers in comparison with the genomic allele dosage (25, 50 or 75%). This revealed that depending on the genotype and cold storage condition, expression of several *Pain-1* alleles deviated from the genomic allele dosage, indicating allele-specific transcriptional regulation (Fig. 2). The 'Satina' allele *Sa* was overrepresented up to 46% (3 wcs) during cold storage compared with its simplex allele dosage. Accordingly, the allele *Sb* was underrepresented, accounting for up to 20% less transcripts than expected from its triplex allele dosage. The three 'Diana' alleles *Da*, *Db*, and *Dc* displayed subtle but diverse allele specific regulation during cold storage. The simplex allele *Da* displayed a 10% higher transcript level at 0 wcs but then decreased continuously over time. The transcript abundance of the 'Theresa' alleles *Tb* and *Tc* hardly deviated from the proportion expected from a triplex and simplex allele dosage, respectively. Without differential allelic expression, each allele of a heterozygous diploid represents 50% of the transcript. The alleles of the diploid genotypes P18 and P40 deviated from this expectation. Transcripts of the allele *P18a* were strongly underrepresented at 0 wcs and increased transiently during cold storage. The allele *P18b* behaved complementarily. The allele *P40d1* was less expressed than *P40d2* at 0 wcs but increased over the cold storage period until it was overrepresented at 4 wcs (Fig. 2). Due to technical reasons, no allelic expression pattern could be detected in P40 at 3 wcs.

Comparison of Pain-1 allelic proteins by 3D modelling

Pain-1 alleles separate in four phylogenetic subgroups *a*, *b*, *c* and *d* (Draffehn *et al*. 2010). Alleles within subgroups *b* and *c* are identical at the amino acid level, whereas alleles within subgroups *a* and *d* have two and five amino acid differences, respectively (Supporting Information Figs S1 & S2). To detect conformational changes, which might cause differential protein performance, we conducted comparative protein modelling between pairs of *Pain-1* alleles based on the crystal structure of the *A. thaliana* cell wall-bound invertase *AtcwINV1* (Verhaest *et al*. 2006). The associated *Pain-1* alleles *Sa* and *Da* (group a), which differ by two amino acids, and the alleles *Pain1-Db*, *-Dc* and *-P40d1* as representatives of subgroups *b*, *c* and *d*, respectively, were selected for 3D protein modelling (Supporting Information Fig. S2). Additionally, we generated a virtual allele,

Pain1-DaT, by replacing lysine in the deduced polypeptide sequence of *Pain1-Da* at position 515 (Lys515) by threonine (Thr515), which is present in alleles of subgroups *b*, *c* and *d* (Draffehn *et al.* 2010). The non-conservative amino acid change Lys515Thr results from the SNP at position 1544 of the *Pain-1* cDNA. The SNP allele A_{1544} translating into Lys515 is specific for *Pain-1* alleles in subgroup *a*, and was associated with better potato chip quality and higher tuber starch content (Draffehn *et al.* 2010). The comparison between *Pain-1* alleles *Da* and *DaT* might therefore hint at a structural difference between associated and non-associated invertase alleles.

Superimposition of nine pairs of modelled *Pain-1* alleles visualized putative structural differences at the molecule's surface, which affected 28 amino acid residues. Four domains were identified that showed differences in surface topology of variable intensity (Fig. 3a, Supporting Information Fig. S3). Domain A comprised the seven C-terminal amino acid positions Ser633, Phe634, Pro635, Leu636, Gln637, Asp638 and Leu639. Differences in the A domain were detected in all superimpositions most prominently in the models *Pain1-Da* on *Pain1-Db*, *Pain1-Da* on *Pain1-P40d1* and *Pain1-Dc* on *Pain1-P40d1*. Structural changes in domain B consisting of a single asparagine at position 147 were visible in five of the nine models. Domain C, which is part of the putative sucrose binding site, included the amino acid residues Leu332, Asp333, Asp334, Asn335 and Lys336,

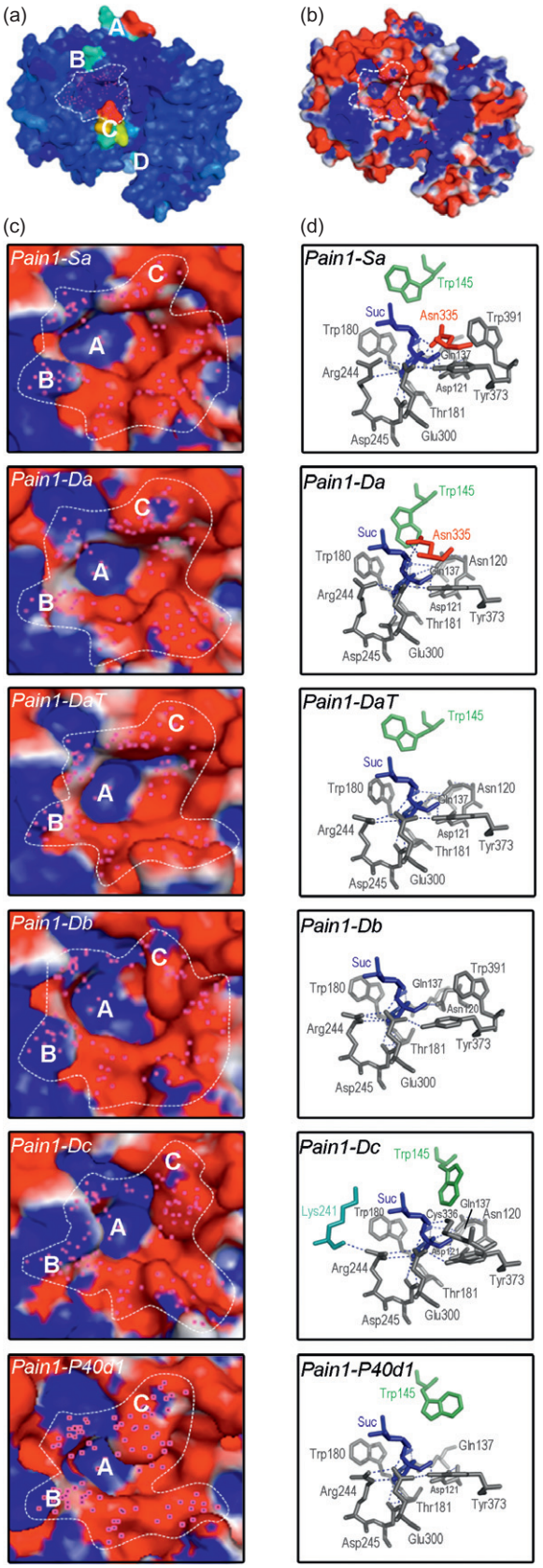

Figure 3. Steric and electrostatic differences between *Pain-1* alleles. (a) As representative model for structural differences between *Pain-1* alleles, the comparison of *Pain1-Da* and *Pain1-Db* is shown. The whole set of superimposed invertase alleles is shown in Supporting Information Fig. S3. The modelling was based on the superimposition of two allelic sequences. Structural differences manifested in four major domains with variable intensities, indicated by A, B, C and D. The intensity of structural differences is colour coded: blue = no structural difference; red = strong structural difference; colours in between represent transitions of different structural intensities from red (strong) to blue (none). The white dotted line depicts the putative sucrose binding site. The amino acids, which form the putative sucrose binding site are indicated by red dots. (b) As representative model for the electrostatic potential (EP) of all invertase alleles, the allele *Pain1-Da* is shown. The EP models of all invertase alleles are shown in Supporting Information Fig. S4. EP models were generated at pH 4.7, mimicking vacuolar conditions. Blue = positively charged, red = negatively charged, white = neutral. The white dotted line depicts the putative sucrose binding site. (c) Zoom-in view on the allele-specific EP models of the putative sucrose binding site. Three variable domains are indicated by A, B and C. White dotted line and red dots as in (a). (d) Allele-specific models of the amino acids placed within 4 Å distance of the catalytic site. Putative H-bonds to the substrate sucrose are shown as dotted lines. The substrate sucrose is highlighted in blue, Trp145 in green, Lys241in light blue, Asn335 in red, and amino acid residues without positional differences between the *Pain-1* alleles are shown in grey. The amino acids are numbered according to their position in the protein counting the start methionine as '1'. Arg, arginine; Asn, asparagine; Asp, aspartic acid; Gln, glutamine; Glu, glutamic acid; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; Suc, sucrose.

Table 2. Amino acid residues forming the putative catalytic site of *Pain-1* alleles

Pain-1 amino acid position	Equivalent *AtcwINV1* amino acid position	*Pain1-Sa*	*Pain1-Da*	*Pain1-DaT*	*Pain1-Db*	*Pain1-Dc*	*Pain1-P40d1*
Asn120	Asn22	No	Yes[a]	Yes[a]	Yes[a]	Yes[a]	No
Asp121	Asp23	Yes[a]	Yes[a]	Yes[a]	No	Yes[a]	Yes[a]
Gln137		Yes[a]	Yes[a]	Yes[a]	Yes[a]	Yes[a]	Yes[a]
Trp145		Yes	Yes	Yes	No	Yes	Yes
Trp180	Trp82	Yes	Yes	Yes	Yes	Yes	Yes
Thr181	Ser181	Yes[a]	Yes[a]	Yes[a]	Yes[a]	Yes[a]	Yes[a]
Lys241		No	No	No	No	Yes	No
Arg244	Arg148	Yes[a]	Yes[a]	Yes[a]	Yes[a]	Yes[a]	Yes[a]
Asp245	Asp149	Yes[a]	Yes[a]	Yes[a]	Yes[a]	Yes[a]	Yes[a]
Glu300	Glu203	Yes[a]	Yes[a]	Yes[a]	Yes[a]	Yes[a]	Yes[a]
Asn335		Yes[a]	Yes[a]	No	No	No	No
Lys336		No	No	No	No	Yes[a]	No
Tyr373		Yes	Yes	Yes	Yes	Yes	Yes
Trp391		Yes	No	No	Yes	No	No

[a]Amino acids, which are involved in H-bond formation to the substrate sucrose.
The numbers of the amino acids of the catalytic site refer to their position in the protein counting the start methionine as '1'.
AtcwINV1, Arabidopsis thaliana cell wall-bound invertase 1; Arg, arginine; Asn, asparagine; Asp, aspartic acid; Gln, glutamine; Glu, glutamic acid; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine.

and was detectable in all superimposed models. The strongest structural differences in this domain were present in models superimposing *Pain1-Da* on *Pain1-Db*, *Pain1-Da* on *Pain1-P40d1*, *Pain1-Db* on *Pain1-P40d1*, and *Pain1-Dc* on *Pain1-P40d1*. The fourth domain D was formed by the amino acids Lys420, Lys421, Thr422, Gly423 and Thr424. Conformational changes in domain D manifested in the three superimposed models *Pain1-Da* on *Pain1-P40d1*, *Pain1-Db* on *Pain1-P40d1* and *Pain1-Dc* on *Pain1-P40d1* (Supporting Information Fig. S3). The superimposition of *Pain1-Da* on the virtual allele *Pain1-DaT* revealed structural changes in domains A, B and C. With one exception (Gly423Arg in allele *Pain1-P40d1*) (Draffehn *et al*. 2010), none of the amino acid positions in domains A, B, C and D was polymorphic in the analysed *Pain-1* alleles (Supporting Information Fig. S2). The putative structural changes at the enzyme's surface are therefore indirectly caused by polymorphisms in other regions of the molecule, which seem to influence its steric properties.

The putative catalytic sucrose binding site of the cell wall-bound invertase of *A. thaliana* is composed of 13 amino acids, notably Asn22, Asp23, Trp82, Ser83, Arg148, Asp149, Gly166, Met201, Glu203, Cys204, Asp240, Tyr279 and Ala280 (Verhaest *et al*. 2006). Equivalent amino acids were found in the putative catalytic sites of the *Pain-1* alleles, which were modelled together with the bound substrate sucrose. There were 9 to 12 of 14 amino acid positions, none of them variable among *Pain-1* alleles, placed within 4 Å distance from the sucrose molecule (Table 2), the maximum distance where hydrogen bonds can still be built (Jeffrey 1997). Seven amino acids (Gln137, Trp180, Thr181, Arg244, Asp245, Glu300, Tyr373) were common to the catalytic sites of all six *Pain-1* allelic models. Additionally, the amino acids Asn120, Asp121 and Trp145 formed part of the catalytic site of four to five alleles (Table 2). Asn120 and Asp121 are members of the conserved region NDPNG whereas Arg244 and Asp245 belong to the conserved motif FRDP, and Glu300 to the conserved motif $WEC^{V}/_{P}D$. All three conserved regions play a crucial role in the catalytic mechanism of hydrolysing the glycosidic bond (Reddy & Maley 1990). Notably, amino acid Asn335 was unique for the putative catalytic sites of the associated alleles *Pain1-Sa* and *-Da*. Intriguingly, the mutation of amino acid Lys515 to Thr515 in the virtual allele *DaT* eliminated Asn335 from its putative catalytic site (Fig. 3d, Table 2). When focusing on the putative catalytic site, the seven conserved amino acids present in all six *Pain-1* alleles showed little spatial variation. Structural differences were introduced by Asn335 specific for alleles *Pain1-Sa* and *Pain1-Da*, Lys336 and Lys241 specific for *Pain1-Dc* and Trp145, which was part of the catalytic site of all alleles except *Pain1-Db*. Trp145 is not involved in H-bond formation with the substrate sucrose (Table 2) but its indol ring displayed different positional angels in five alleles (Fig. 3d). It is noteworthy that the sequences of *Da* and *DaT* not only differ in the elimination of Asn335 from the putative catalytic site of *DaT* but also display the steric difference of Trp145. Modelling of the electric potential of the *Pain-1* alleles also uncovered differences all over the molecule's surface and at the putative catalytic site (Fig. 3b,c, Supporting Information Fig. S4). In accordance with the partially negative character of sucrose (-OH groups), all alleles are positively charged in the depth of their binding site (Fig. 3c, domain A). Nevertheless, charge differences occur between the putative allelic sucrose binding sites. Domain B composed of Arg244 and Asp245 was positively charged and most prominent in the alleles *Sa* and *Db*. Domain C, which is defined by Gln137

Table 3. K_m (mM) and v_{max} (mmol h^{-1}*mg protein) of *Pain-1* invertase alleles at 30 and 4 °C

Allele	K_m (30 °C)	K_m (4 °C)	v_{max} (30 °C)	v_{max} (4 °C)
Pain1-Sa	22.1 ± 1.4	2.3 ± 0.3	4.7 ± 1.0	1.7 ± 0.1
Pain1-Da	19.9 ± 1.4	4.0 ± 0.3	12.4 ± 0.8	2.8 ± 0.1
Pain1-Db	19.6 ± 1.4	2.7 ± 0.3	11.2 ± 0.8	2.7 ± 0.1
Pain1-Dc	15.6 ± 1.4	3.7 ± 0.3	6.2 ± 1.0	2.2 ± 0.1
Pain1-P40d1	17.2 ± 1.4	3.4 ± 0.3	5.8 ± 1.0	2.1 ± 0.1
FY 1679	21.0 ± 1.6	21.1 ± 2.5	23.9 ± 4.5	7.7 ± 0.9

Standard errors derived from three biological replicates for the associated alleles *Pain1-Sa* and *Pain1-Da*, and the wild-type reference strain *FY 1679*, and from two biological replicates for the other alleles done in technical replicates to obtain six measurements. To make assays of different invertase isoforms comparable, the yeast reference strain *FY 1679* was used as positive control.

and Trp145, showed a small positively charged area surrounded by a negatively charged region, which appears in *Pain1-Da*, *Dc* and *P40d1*.

Variation of biochemical properties of Pain-1 alleles

Plant invertases have been functionally characterized in heterologous systems such as yeast (*Saccharomyces cerevisae*) (Fridman *et al*. 2004). The yeast mutant *SUC2* (Gozalbo & Hohmann 1989) lacks invertase activity, and is unable to use sucrose as sole carbon source. Complementation of the *SUC2* mutant phenotype was achieved by all *Pain-1* cDNA alleles (Supporting Information Fig. S5). The recombinant *Pain-1* proteins were extracted from yeast cultures and partially purified. Western blot analysis with an antibody raised against potato vacuolar invertase (Burch *et al*. 1992) showed similar protein levels of recombinant potato invertase (Supporting Information Fig. S6). Protein extracts were used to determine the apparent kinetic constants substrate affinity [$K_{m\ (sucrose)}$ (hereafter called K_m)] and maximal velocity (v_{max}) at 30 and at 4 °C. Assays at 4 °C mimicked cold storage conditions, and should elucidate whether the allelic enzymes performed differently under cold treatment. The results of the biochemical analysis at 30 and 4 °C are shown in Table 3. The *P*-values for significant differences between all pairs of alleles are shown in Supporting Information Tables S2–S5. The wild-type strain *FY 1679* displayed a similar K_m at both temperatures, while v_{max} was three times lower at 4 °C compared with 30 °C. All *Pain-1* alleles tested showed significant variation of K_m and v_{max} values, depending on genotype and temperature. Most similar to each other were *Pain-1* alleles *Dc* and *P40d1*, exhibiting similar biochemical characteristics at both temperatures but they differed from *Sa*, *Da* and *Db*, particularly for v_{max} at 4 °C. The K_m value of the *Pain-1* alleles at 30 °C was in the same order of magnitude as the wild type yeast allele, ranging between 15.6 mM for *Pain1-Dc* and 22.1 mM for *Pain1-Sa*. Maximal velocities were lower, varying from 4.7 mmol*h^{-1}*mg $protein^{-1}$ in the allele *Pain1-Sa* to 12.4 mmol*h^{-1}*mg $protein^{-1}$ in *Pain1-Da*. At 30 °C, the K_m values of the *Pain-1* alleles *Sa*, *Da* and *Db* were similar to each other, whereas v_{max} values differed significantly, with *Da* and *Db* having approximately three times higher v_{max} values than *Pain1-Sa* (Table 3). At 4 °C, K_m and v_{max} of all *Pain-1* alleles decreased four- to ninefold and two- to fourfold, respectively. K_m values varied between 2.3 mM for *Pain1-Sa* and 4.0 mM for *Pain1-Da,* and v_{max} values ranged from 1.7 to 2.8. mmol*h^{-1}*mg $protein^{-1}$ for *Pain1-Sa* and *Pain1-Da,* respectively. The v_{max} value of *Pain1-Sa* differed significantly from the other four alleles.

DISCUSSION

In a previous paper (Draffehn *et al*. 2010), we described the isolation, sequence diversity and marker-trait associations of 11 full-length cDNA alleles of *Pain-1* from six heterozygous potato genotypes. The 11 alleles can be assigned to four phylogenetic groups *a*, *b*, *c* and *d* (Supporting Information Fig. S1). The cDNA alleles *Sa* and *Da* in group *a* share three group-specific SNP alleles (C_{552}, A_{718}, A_{1544}), which are strongly associated with better chip quality and higher tuber starch content (Li *et al*. 2008; Draffehn *et al*. 2010). In the case that DNA variation at the *Pain-1* locus is causal for a proportion of the observed phenotypic variation, functional differences are expected between associated and non-associated alleles. In the present study, we addressed the questions (1) whether and how the sequence variation in *Pain-1* translates into structural and functional variation of the encoded proteins; and (2) whether we can identify functional and/or structural properties that distinguish the associated alleles *Sa* and *Da* from non-associated alleles. Therefore, we conducted comparative 3D protein modelling and determined K_m and v_{max} values of heterologously expressed *Pain-1* cDNA alleles. In addition, differences in allelic expression were analysed by quantifying the transcript levels of *Pain-1* alleles *Da* and *Sa* and of alleles representative for groups *b, c* and *d* in tubers of the tetraploid genotypes 'Satina', 'Diana' and 'Theresa' and the diploid genotypes P18 and P40.

Consistent with earlier studies (Zhou *et al*. 1994; Zrenner *et al*. 1996; Bagnaresi *et al*. 2008), we observed strong up-regulation of potato *Pain-1* transcripts in tubers after 1 to 2 weeks of cold storage, followed by a decline in the 3rd and 4th week (Fig. 1). This result emphasizes the impact of *Pain-1* on cold sweetening. However, transcript levels were

highly variable between the genotypes. Particularly, P40 showed much higher *Pain-1* transcript levels than all other genotypes. Whether this is due to the 50% DNA of the wild species *Solanum spegazzinii* present in P40 (Barone *et al*. 1990), or to combined effects of transcriptional modifiers in this genetic background, is unclear.

Besides monitoring total *Pain-1* transcript accumulation, we quantified the expression of individual *Pain-1* alleles in comparison with their genomic dosage before and after cold storage (Table 1, Fig. 2). Interestingly, transcript levels of all alleles varied during cold storage and, except for *Tb,* deviated from the expected genomic proportions. For none of the alleles in the same homology group, we observed a similar expression pattern. All group *a* alleles (*Sa, Da* and *P18a*) behaved differently, not only with respect to the pattern expected from their genomic dosage but also based on their phylogenetic relationship. *Sa* and *Da* showed opposite expression patterns during cold storage, *Sa* being strongly overrepresented and *Da* being underrepresented relative to the other alleles. Silencing of vacuolar invertase in transgenic lines resulted in up to 97% reduction of invertase activity, significantly lower glucose and fructose levels, and improved chip quality (Bhaskar *et al*. 2010). Compared with studies using *RNAi* transgenic lines, the differences in allelic expression observed in this study appear small. However, when dealing with natural variation, quantitative rather than qualitative differences are expected. Small expression differences between invertase alleles might translate in quantitative effects on sugar content and chip quality. *RNAi* lines with only 38% silencing of invertase also displayed a more subtle quantitative phenotype. Reducing sugar accumulation was much slower but reached the same absolute concentration after cold storage than wild-type plants (Bhaskar *et al*. 2010). Sequence polymorphisms in cis-acting regulatory elements or in trans-acting regulators of *Pain-1*, present in the corresponding heterozygous genetic backgrounds, might also cause these expression differences. This is further substantiated as invertase alleles interact with other loci (Li, Paulo, van Eeuwijk *et al*., 2010). Another explanation for differential allelic transcript abundance is transcript stability. Several studies in humans reported that synonymous SNPs affect mRNA secondary structure, thereby influencing mRNA degradation and modification (e.g. splicing), and consequently protein abundance (Chamary, Parmley & Hurst 2006; Nackley *et al*. 2006). Furthermore, such SNPs can also affect enzymatic substrate specificities (Kimchi-Sarfaty *et al*. 2007) by delaying co-translational folding due to 'silent' SNPs introduced in rare codons. Whether synonymous SNPs at the *Pain-1* locus (Draffehn *et al*. 2010) have similar effects needs further investigation. In conclusion, our results demonstrate that transcriptional regulation of *Pain-1* invertase alleles is highly variable and genotype dependent.

Having established variability at the transcript level, we next turned our attention to the protein level. Comparative 3D modelling of pairs of allelic *Pain-1* proteins indicated steric differences of variable intensity in four regions on the protein's surface (Fig. 3a, Supporting Information Fig. S3). Differences in regions B and D did not show specificity for associated compared with non-associated alleles. The strongest steric changes resulted from comparisons with allele *P40d1*, which differed by the highest number of amino acids from all other alleles (Supporting Information Fig. S2). Intriguingly, the single non-conservative amino acid change (Lys515Thr) by which the virtual protein *DaT* differs from *Da*, was sufficient to induce structural changes in regions A, B and C. The presence or absence of variation in regions A, B, C and D did not seem to be responsible though for functional differences between associated and non-associated alleles. These protein regions might be under weaker natural selection allowing mutations to persist during evolution. Nevertheless, there are examples that structural differences on the protein's surface lead to altered structure and therefore performance of enzyme complexes (Veitia, Bottani & Birchler 2008). Invertases do form homomeric protein complexes (Ross, McRae & Davies 1996). Performance of the complex might depend on the allelic composition of the subunits.

Modelling the catalytic sucrose binding site of *Pain-1* revealed further differences between six *Pain-1* alleles (Table 2, Fig. 3c,d). These steric changes were the result of combinations of amino acid substitutions at positions remote from the catalytic site. Most conspicuously, the indol ring of Trp145, in close proximity to the sucrose binding site, occupied a different position in each of the five allelic models. Interestingly, Asn335, which is included in the variable region C, was the only amino acid that was placed specifically in the putative catalytic site of the associated alleles *Sa* and *Da*, close to the functionally important Asn120 (Reddy & Maley 1990). *In silico* introduction of the mutation Lys515Thr in *DaT* removed Asn335 from the catalytic site (Fig. 3d). This difference might be the structural basis for the association of *Sa* and *Da* with tuber quality traits. Furthermore, the allelic structural variation resulted in a different size and shape of the substrate binding pocket and in different electrostatic potentials around the catalytic site (Fig. 3c), which might have an effect on the enzyme's kinetic properties.

Kinetic parameters were analysed after successful complementation of the yeast *SUC2* mutant with five *Pain-1* alleles, by determining apparent K_m and v_{max} at 30 and 4 °C. Invertase enzyme kinetics at 4 °C have not been described so far. Significant differences were observed between alleles for both biochemical parameters at both temperatures (Table 3, Supporting Information Tables S2–S5). The fact that differences were less dramatic than reported for two invertase alleles of *Solanum lycopersicum* and *Solanum pennellii*, underlying a fruit sugar content QTL in tomato (Fridman *et al*. 2004) can be a result of analysing intraspecific rather than interspecific natural variation. However, it is important to note that *S. lycopersicum* has an unusually high K_m for sucrose. The alleles *Dc* and *P40d1* showed similar K_m and v_{max} values at both temperatures, despite the structural differences highlighted by 3D protein modelling. The only structural feature that might be correlated with the similar catalytic properties of

Dc and *P40d1* was the similar orientation of Trp145 in the models of the sucrose binding site (Fig. 3d). The most striking difference at 30 °C was the approximately twofold higher v_{max} value of alleles *Da* and *Db* compared with *Sa*, *Dc* and *P40d1*. As the relative amount of potato invertase protein in total yeast protein was similar in all strains (Supporting Information Fig. S6), the large difference observed for v_{max} is not due to variable protein abundance. The protein structures modelled at 25 °C (298 K) also did not indicate a possible reason for the highly significant differences between v_{max} values observed at 30 °C. The physiological sucrose concentration in tubers without cold storage is approximately 5 mM and increases rapidly up to 20 to 30 mM after 4 weeks' cold treatment (Junker *et al*. 2006; Bhaskar *et al*. 2010). As tuber sugar content is a complex trait, sucrose concentration is highly variable (Menendez *et al*. 2002). At ambient temperatures, the analysed *Pain-1* alleles face sucrose concentrations well below their apparent K_m values and therefore do not operate at v_{max}. Conversely, at low temperatures, *Pain-1* alleles operate at v_{max} when sucrose concentrations are high (20–30 mM). However, in genotypes with low sucrose levels even under cold storage (e.g. below 5 mM), small differences in K_m as observed in our study (Table 3) could have an impact on reducing sugar accumulation depending on the allele combination present in a given genotype. Neither K_m nor v_{max} values on their own pointed to a mechanism that could explain the difference between the associated and non-associated alleles. Altered glycosylation mechanisms in yeast (Sturm *et al*. 1995) potentially influence the kinetic properties of the heterologously expressed alleles compared with the *in vivo* situation. However, given the large number of invertase isoforms in potato and the difficulty to purify allelic enzymes, yeast represented the best available system to test their biochemical properties.

The decreased sucrose turnover of *Pain-1* at low temperature is compensated *in vivo* in a genotype-dependent manner by increased substrate affinity, increased protein abundance and enzyme activity (Rorem & Schwimmer 1963; Pressey & Shaw 1966), as implicated by the increased expression during cold storage (Zhou *et al*. 1994; Zrenner *et al*. 1996; Bagnaresi *et al*. 2008). As indicated previously, the actual *Pain-1* activity in tissues of an individual genotype depends therefore on several factors. Subtle biochemical differences caused by a combination of factors rather than a specific feature of the protein might be responsible for the phenotypic outcomes associated with the alleles *Sa* and *Da*. Furthermore, the associated SNPs in the *Pain-1* coding sequence could be included in a larger haplotype block comprising promoter regions, which affect transcriptional regulation. As such sequence differences can modify the binding of regulatory proteins, which may also vary in different genetic backgrounds, a multitude of transcriptional readouts is possible. Expression analysis on a population scale is required to test whether quantitative *Pain-1* expression (eQTL) contributes to the natural variation of tuber cold sweetening and which polymorphism underlies this phenomenon.

Taken together, the results of this and the previous paper (Draffehn *et al*. 2010) clearly demonstrate that the potato invertase gene *Pain-1* shows natural variation of structure and function. Whether natural variation of *Pain-1* is directly responsible for part of the phenotypic variation of tuber chip quality is still unclear. We cannot exclude the possibility that a physically linked gene, which is in linkage disequilibrium with *Pain-1,* is causing the phenotypic effects. It is also possible that more than one gene control the sugar QTL overlapping with the *Pain-1* locus. Inspecting the superscaffold PGSC0003DMB000000605 of 268.6 kbp showed that it contains 19 annotated genes (PGSC 2011), one of which (PGSC0003DMG400013856) corresponds to *Pain-1*. Interestingly, a putative trehalose 6-phosphate phosphatase (PGSC0003DMG400013857; TPP) is located at 30 kbp distance from *Pain-1*. Evidence accumulated recently that trehalose and its precursor trehalose 6-phosphate play a role in abiotic stress response, particularly during osmotic stress as consequence of cold and salt treatment in *Arabidopsis* (Iordachescu & Imai 2008).

In conclusion, the present study characterizes transcriptional, structural and biochemical diversity of natural *Pain-1* invertase alleles, and highlights the impact of the genetic background on allele-specific transcript patterns. Although the observed variation did not reveal a specific mechanism explaining the differences between associated and non-associated *Pain-1* alleles, our data demonstrate the importance of considering natural variation when studying structure and function of genes and their encoded proteins, especially, when dealing with a highly polymorphic, polyploid species such as potato.

ACKNOWLEDGMENTS

Part of this work was carried out in the Department of Plant Breeding and Genetics headed by Maarten Koornneef, and was funded by the Max Planck Society. The authors thank Dirk Walther of the Max Planck Institute of Molecular Plant Physiology in whose group Pawel Durek carried out the 3D modelling. The authors also thank Claude Urbany for general discussion and support.

REFERENCES

Bagnaresi P., Moschella A., Beretta O., Vitulli F., Ranalli P. & Perata P. (2008) Heterologous microarray experiments allow the identification of the early events associated with potato tuber cold sweetening. *BMC Genomics* **9,** 176.

Baker N.A., Sept D., Joseph S., Holst M.J. & McCammon J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* **98,** 10037–10041.

Barone A., Ritter E., Schachtschabel U., Debener T., Salamini F. & Gebhardt C. (1990) Localization by restriction fragment length polymorphism mapping in potato of a major dominant gene conferring resistance to the potato cyst nematode *Globodera rostochiensis*. *Molecular and General Genetics* **224,** 177–182.

Bhaskar P.B., Wu L., Busse J.S., Whitty B.R., Hamernik A.J., Jansky S.H., Buell C.R., Bethke P.C. & Jiang J. (2010) Suppression of the

vacuolar invertase gene prevents cold-induced sweetening in potato. *Plant Physiology* **154,** 939–948.

Bondar R.J.L. & Mead D.C. (1974) Evaluation of glucose-6-phosphate dehydrogenase from leuconostoc mesenteroides in the hexokinase method for determining glucose in serum. *Clinical Chemistry* **20,** 586–590.

Burch L.R., Davies H.V., Cuthbert E.M., Machray G.C., Hedley P. & Waugh R. (1992) Purification of soluble invertase from potato. *Phytochemistry* **31,** 1901–1904.

Chamary J.V., Parmley J.L. & Hurst L.D. (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nature Reviews Genetics* **7,** 98–108.

Chen X., Salamini F. & Gebhardt C. (2001) A potato molecular-function map for carbohydrate metabolism and transport. *Theoretical and Applied Science* **102,** 284–295.

Delano W.L. (2002) *The PyMOL Molecular Graphics System*. DeLano Scientific, San Carlos, CA, USA.

Dolinsky T.J., Nielsen J.E., McCammon J.A. & Baker N.A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research* **32,** W665–W667.

Draffehn A.M., Meller S., Li L. & Gebhardt C. (2010) Natural diversity of potato (*Solanum tuberosum*) invertases. *BMC Plant Biology* **10,** 271.

Fridman E., Carrari F., Liu Y.-S., Fernie A.R. & Zamir D. (2004) Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science* **305,** 1786–1789.

Gietz R.D. & Schiestl R.H. (1995) Transforming yeast with DNA. *Methods in Molecular and Cellular Biology* **5,** 255–269.

Gozalbo D. & Hohmann S. (1989) The naturally occurring silent invertase structural gene suc2 zero contains an amber stop codon that is occasionally read through. *Molecular and General Genetics* **216,** 511–516.

Guex N. & Peitsch M.C. (1997) SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis* **18,** 2714–2723.

Hedley P.E., Machray G.C., Davies H.V., Burch L. & Waugh R. (1993) cDNA cloning and expression of a potato (*Solanum tuberosum*) invertase. *Plant Molecular Biology* **22,** 917–922.

Hedley P.E., Machray G.C., Davies H.V., Burch L. & Waugh R. (1994) Potato (*Solanum tuberosum*) invertase-encoding cDNAs and their differential expression. *Gene* **145,** 211–214.

Iordachescu M. & Imai R. (2008) Trehalose biosynthesis in response to abiotic stresses. *Journal of Integrative Plant Biology* **50,** 1223–1229.

Jeffrey G.A. (1997) *An Introduction to Hydrogen Bonding*. Oxford University Press, Oxford.

Junker B.H., Wuttke R., Nunes-Nesi A., Steinhauser D., Schauer N., Bussis D., Willmitzer L. & Fernie A.R. (2006) Enhancing vacuolar sucrose cleavage within the developing potato tuber has only minor effects on metabolism. *Plant & Cell Physiology* **47,** 277–289.

Kimchi-Sarfaty C., Oh J.M., Kim I.W., Sauna Z.E., Calcagno A.M., Ambudkar S.V. & Gottesman M.M. (2007) A 'silent' polymorphism in the MDR1 gene changes substrate specificity. *Science* **315,** 525–528.

Lammens W., Le Roy K., Van Laere A., Rabijns A. & Van den Ende W. (2008) Crystal structures of arabidopsis thaliana cell-wall invertase mutants in complex with sucrose. *Journal of Molecular Biology* **377,** 378–385.

Li H., Robertson A.D. & Jensen J.H. (2005a) Very fast empirical prediction and rationalization of protein pKa values. *Proteins* **61,** 704–721.

Li L., Strahwald J., Hofferbert H.R., Lubeck J., Tacke E., Junghans H., Wunder J. & Gebhardt C. (2005b) DNA variation at the invertase locus invGE/GF is associated with tuber quality traits in populations of potato breeding clones. *Genetics* **170,** 813–821.

Li L., Paulo M.J., Strahwald J., Lübeck J., Hofferbert H.R., Tacke E., Junghans H., Wunder J., Draffehn A. & van Eeuwijk F. (2008) Natural DNA variation at candidate loci is associated with potato chip color, tuber starch content, yield and starch yield. *Theoretical and Applied Genetics* **116,** 1167–1181.

Li L., Paulo M.-J., van Eeuwijk F. & Gebhardt C. (2010) Statistical epistasis between candidate gene alleles for complex tuber traits in an association mapping population of tetraploid potato. *Theoretical and Applied Genetics* **121,** 1303–1310.

Menendez C.M., Ritter E., Schafer-Pregl R., Walkemeier B., Kalde A., Salamini F. & Gebhardt C. (2002) Cold sweetening in diploid potato: mapping quantitative trait loci and candidate genes. *Genetics* **162,** 1423–1434.

Müller-Thurgau H. (1882) Über Zuckeranhäufung in Pflanzentheilen in Folge niederer Temperatur. *Landwirtschaftliches Jahrbuch* **11,** 751–828.

Nackley A.G., Shabalina S.A., Tchivileva I.E., Satterfield K., Korchynskyi O., Makarov S.S., Maixner W. & Diatchenko L. (2006) Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science* **314,** 1930–1933.

Nicot N., Hausman J.-F., Hoffmann L. & Evers D. (2005) Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *Journal of Experimental Botany* **56,** 2907–2914.

PGSC (2011) Genome sequence and analysis of the tuber crop potato. *Nature* **475,** 189–195.

Pressey R. & Shaw R. (1966) Effect of temperature on invertase, invertase inhibitor, and sugars in potato tubers. *Plant Physiology* **41,** 1657–1661.

Reddy V.A. & Maley F. (1990) Identification of an active-site residue in yeast invertase by affinity labeling and site-directed mutagenesis. *Journal of Biological Chemistry* **265,** 10817–10820.

Riesmeier J.W., Willmitzer L. & Frommer W.B. (1992) Isolation and characterization of a sucrose carrier cDNA from spinach by functional expression in yeast. *EMBO Journal* **11,** 4705–4713.

Roe M.A., Faulks R.M. & Belsten J.L. (1990) Role of reducing sugars and amino-acids in fry color of chips from potatoes grown under different nitrogen regimes. *Journal of the Science of Food and Agriculture* **52,** 207–214.

Roitsch T. & González M.-C. (2004) Function and regulation of plant invertases: sweet sensations. *Trends in Plant Science* **9,** 606–613.

Ronaghi M., Karamohamed S., Petterson B., Uhlen M. & Nyren P. (1996) Real-time DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry* **242,** 84–89.

Rorem E.S. & Schwimmer S. (1963) Double pH optima of potato invertase. *Experientia* **19,** 150–151.

Ross H.A., McRae D. & Davies H.V. (1996) Sucrolytic enzyme activities in cotyledons of the Faba Bean (developmental changes and purification of alkaline invertase). *Plant Physiology* **111,** 329–338.

Royo J.L., Hidalgo M. & Ruiz A. (2007) Pyrosequencing protocol using a universal biotinylated primer for mutation detection and SNP genotyping. *Nature Protocols* **2,** 1734–1739.

Sali A. & Blundell T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* **234,** 779–815.

Shallenberger R.S., Smith O. & Treadway R.H. (1959) Food color changes, role of the sugars in the browning reaction in potato chips. *Journal of Agricultural and Food Chemistry* **7,** 274–277.

Söding J., Biegert A. & Lupas A.N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research* **33,** W244–W248.

Sturm A., Montreuil J., Vliegenthart J.F.G. & Schachter H. (1995) Chapter 9 N-glycosylation of plant proteins. In *New Comprehensive Biochemistry* (eds J. Montreuil, H. Schacter & J.F.G. Vliegenthart), pp. 521–541. Elsevier, Kidlington, UK.

Tymowska-Lalanne Z., Kreis M. & Callow J.A. (1998) The plant invertases: physiology, biochemistry and molecular biology. *Advances in Botanical Research* **28,** 71–117.

Veitia R.A., Bottani S. & Birchler J.A. (2008) Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends in Genetics* **24,** 390–397.

Verhaest M., Lammens W., Le Roy K., De Coninck B., De Ranter C.J., Van Laere A., Van den Ende W. & Rabijns A. (2006) X-ray diffraction structure of a cell-wall invertase from *Arabidopsis thaliana*. *Acta Crystallographica D Biological Crystallography* **62,** 1555–1563.

Zhou D., Mattoo A., Li N., Imaseki H. & Solomos T. (1994) Complete nucleotide sequence of potato tuber acid invertase cDNA. *Plant Physiology* **106,** 397–398.

Zrenner R., Salanoubat M., Willmitzer L. & Sonnewald U. (1995) Evidence of the crucial role of sucrose synthase for sink strength using transgenic potato plants (*Solanum tuberosum L.*). *Plant Journal* **7,** 97–107.

Zrenner R., Schuler K. & Sonnewald U. (1996) Soluble acid invertase determines the hexose-to-sucrose ratio in cold-stored potato tubers. *Planta* **198,** 246–252.

Received 15 February 2012; received in revised form 14 May 2012; accepted for publication 14 May 2012

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

File S1. Theoretical model of potato invertase allele *Pain1-Da*.

File S2. Theoretical model of potato invertase allele *Pain1-DaT*.

File S3. Theoretical model of potato invertase allele *Pain1-Db*.

File S4. Theoretical model of potato invertase allele *Pain1-Dc*.

File S5. Theoretical model of potato invertase allele *Pain1-Sa*.

File S6. Theoretical model of potato invertase allele *Pain1-P40d1*.

File S7. Crystal structure of sucrose from PDB (PDB 2QQU).

Figure S1. Classification of *Pain-1* allelic proteins in phylogenetic subgroups according to Draffehn *et al*. (2010).

Figure S2. Amino acid alignment of *Pain-1* alleles used for 3D protein modelling. As representatives of the phylogenetic subgroups *b*, *c* and *d*, the *Pain-1* alleles *Pain1-Db*, *Pain1-Dc* and *P40d1* were selected. Amino acid exchanges are highlighted in colours. Amino acids specific for the associated *Pain-1* alleles *Sa* and *Da* (Draffehn *et al*. 2010) are coloured in red.

Figure S3. Structural models of nine superimposed *Pain-1* invertase alleles. The modelling was based on superimposition of two allelic sequences. Structural differences manifested in four major domains with variable intensities, indicated by A, B, C and D. The intensity of structural differences is colour coded: blue = no structural difference; red = strong structural difference; colours in between represent transitions of different structural intensities from red (strong) to blue (none). The white dotted line depicts the putative sucrose binding site. The amino acids, which form the putative sucrose binding site, are indicated by red dots.

Figure S4. Electrostatic potential (EP) models of six *Pain-1* invertase alleles. EP models were generated at pH 4.7 mimicking vacuolar conditions. Blue = positively charged, red = negatively charged, white = neutral. The white dotted line depicts the putative sucrose binding site.

Figure S5. Complementation of the yeast *SUC2* mutant with *Pain-1* alleles. Yeast wild-type strain *FY 1679*, invertase mutant strain *SUC2*, and five strains transformed with *Pain-1* alleles *Sa, Da, Db, Dc* and *P40d1*, were grown in yeast minimal media with 2% sucrose as the sole carbon source. Growth was recorded by OD600 and plotted against time in hours. OD600 values represent the means of two replicates. Standard deviations were less than 20% of mean.

Figure S6. Western blot analysis of *Pain-1* alleles expressed in yeast. The blot was probed with an antibody against 58 kDa vacuolar invertase of potato (Burch *et al*. 1992) (top panel). Total protein (15 μg) was loaded. Allele names are given above each lane. The *SUC2* mutant and the wild-type strain *FY 1679* were also blotted. The arrows on the left indicate size bands of the MagicMark Marker (Invitrogen, Karlsruhe Germany); the arrow on the right refers to the 58 kDa band corresponding to the potato invertase protein. The Ponceau S stained blot membrane (bottom panel) is the loading control for the Western blot analysis.

Table S1. Primers used in pyrosequencing.

Table S2. *P* values of pairwise comparisons between Km values of the *Pain-1* alleles at 30 °C. *P* values <0.05 are shown in bold numbers.

Table S3. *P* values of pairwise comparisons between vmax values of the *Pain-1* alleles at 30 °C. *P* values <0.05 are shown in bold numbers.

Table S4. *P* values of pairwise comparisons between Km values of the *Pain-1* alleles at 4 °C. *P* values <0.05 are shown in bold numbers.

Table S5. *P* values of pairwise comparisons between vmax values of the *Pain-1* alleles at 4 °C. *P* values <0.05 are shown in bold numbers.