Available online at www.sciencedirect.com
SCIENCE DIRECT®
Carbohydrate Research 339 (2004) 2281–2295
Carbohydrate RESEARCH

Pectin methylesterases: sequence-structural features and phylogenetic relationships

Oskar Markovič[a] and Štefan Janeček[b,*]

[a]*Institute of Chemistry, Slovak Academy of Sciences, Dúbravská cesta 9, SK-84538 Bratislava, Slovakia*
[b]*Institute of Molecular Biology, Member of the Centre of Excellence for Molecular Medicine, Slovak Academy of Sciences, Dúbravská cesta 21, SK-84551 Bratislava, Slovakia*

Received 30 January 2004; accepted 25 June 2004

Abstract—Pectin methylesterases (PMEs) are enzymes produced by bacteria, fungi and higher plants. They belong to the carbohydrate esterase family CE-8. This study deals with comparison of 127 amino acid sequences of this family containing the five characteristic sequence segments: 44_GxYxE, 113_QAVAL, 135_QDTL, 157_DFIFG, 223_LGRPW (*Daucus carota* numbering). Six strictly conserved residues (Gly44, Gly154, Asp157, Gly161, Arg225 and Trp227) and six conservative ones (Ile39, Ser86, Ser137, Ile152, Ile159 and Leu223) were identified. A set of 70 representative PMEs was created. The sequences were aligned and the evolutionary tree based on the alignment was calculated. The tree reflected the taxonomy: the fungal and bacterial PMEs formed their own clusters and the plant enzymes were grouped into eight separate clades. The plant PME from *Vitis riparia* was placed in a common clade with fungi. Three plant clades (Plant 1, 2 and 3) were relatively homogenous reflecting high degree of mutual sequence identity. The clade Plant 4 contained PMEs from flower parts (mostly form pollen) and was heterogenous, like the clades Plant 1a and 2a, which moreover exhibit an intermediate character. The clades Plant X1 and X2 were situated in the tree close to microbial clades and represented atypical plant PMEs. Taking into account the remaining plant PMEs, an expanded plant alignment and tree (with most *Arabidopsis thaliana* and *Oryza sativa* enzymes), were prepared. An exclusive *Arabidopsis* alignment and tree indicated the existence of a new plant clade X3. In the *pre pro* region of most plant enzymes a longer conserved segment containing basic dipeptide, R(K)/R(K), that precedes the N-terminal end of PME was revealed. This was not observed in the clade Plant X1 and majority of the clade Plant X2. This study brings further the description of occurrence of potential glycosylation sites in *pre pro* sequences and in mature enzymes as well as important amino acid residues, such as aspartates, cysteines, histidines and other aromatic residues (Tyr, Phe and Trp), with discussion of their possible function in the activity of PMEs.

Keywords: Pectin methylesterase; Carbohydrate esterase family 8; Aspartyl esterase; Sequence comparison; Active site; Evolutionary relationships; Clades

1. Introduction

Pectic esterases (pectinesterases) comprise enzymes that hydrolyze the esters present in the pectin backbone and currently three classes of esterases have been identified:[1] (i) the pectin methylesterases (EC 3.1.1.11); (ii) the pectin acetylesterases (EC 3.1.1.6); and (iii) the rhamnogalacturonan acetylesterases (EC 3.1.1.-). Pectin methylesterase (PME) catalyses the de-esterification of O6 methyl-esterified D-galactosiduronic acid units in pectic compounds. In the frame of CAZy classification system[2] it belongs to the carbohydrate esterase family CE-8. The substrate, pectin, is heteropolysaccharide, located in the primary cell wall of dicotyledonous plants, constituting the main component of the middle lamella.[3] The backbone consists of homogalacturonan (smooth regions), rhamnogalacturonan I, rhamnogalacturonan II and xylogalacturonan (hairy regions).[4–6] Homogalacturonan is

Abbreviations: PME, pectin methylesterase; pgs, potential glycosidic site; 3D, three-dimensional; ORF, open reading frame.

* Corresponding author. Tel.: +421-2-5930-7420; fax: +421-2-5930-7416; e-mail addresses: oskar.markovic@savba.sk; stefan.janecek@savba.sk

0008-6215/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2004.06.023

methyl-esterified in position 6 and acetylated at positions 2 or 3. Highly esterified homogalacturonans, when exported into cell walls, are subsequently de-esterified by the mature PMEs.[7–9]

PME genes occur in multigene families and encode isoforms with different action pattern with respect to removal of methyl esters.[9–13] Some of plant PME genes were shown to be ubiquitously expressed,[14–16] the others are expressed during fruit ripening,[10,17] microsporogenesis and germination of the pollen grain,[18,19] or stem elongation.[11,20] On the other hand, in phytopathogenic fungi and bacteria, PME genes are expressed in heterologous lower-eukaryotic and prokaryotic systems,[21,22] respectively.

Two PME three-dimensional (3D) structures have been determined to date: the bacterial PME from *Erwinia chrysanthemi* PemA[23] and the plant PME from carrot.[24] Based on the proposed active sites it was possible to characterize the enzyme PME as aspartyl esterase. Both structures adopt the right-handed parallel β-helix structural domain first found in pectate lyase.[25] The parallel β-helix class has been divided into seven families:[26] (1) extracellular pectate lyase; (2) polygalacturonases and rhamnogalacturonase A; (3) pectin methylesterases; (4) pectate lyases homologous to PelL; (5) the phage P22 tailspike endorhamnosidases; (6) chondroitinase B; and (7) pertactin. Four of the enzymes families degrade pectin. It was suggested by comparing their substrate-binding sites that the right-handed parallel β-helix enzymes might have evolved from a common ancestor.[26]

Currently the family CE-8 has contained more than 100 complete amino acid sequences.[2] Many members are only putative PMEs or PME-like proteins, especially those from the plant genome sequencing projects. In *Arabidopsis thaliana* EST (expressed sequence tag) project there were sequenced nearly 30,000 anonymous genes.[27,28] The *Arabidopsis* metabolome differs from any other organism sequenced to date by the presence of more than 420 genes encoding the enzymes involved in the synthesis and modification of cell wall. Roughly 52 genes encode polygalacturonases, 20 encode pectate lyases and 79 encode PMEs.[29] Only few of the 67 *A. thaliana* sequences similar to PME, available, for example, at the CAZy web-server,[2] were tested for PME activity and biochemically characterized.[16,30,31] With regard to the Rice Genome Program,[32] at least 13 *Oryza sativa* putative PME sequences were added into the family CE-8 of the CAZY web server.[2]

In higher plants the PME genes encode N-terminal extension of different length (from 2 up to 968 amino acid residues), which consists of *pre* and *pro* sequence. The *pre* region or signal peptide, required for protein targeting, is cleaved in the first step of maturation.[33] The *pro* sequence cleavage of the covalently attached mature enzyme may occur in the second step of maturation by monobasic or dibasic endopeptidases processing in analogy to the KEX2-like proteases.[34,35] This occurs close to the RR(K)LL(M) motif present in plant PMEs between the N-terminus of the variable length and the conserved C-terminal domain.[8,15,36] It was hypothesized that the *pro* regions of the plant PME play the role in folding mechanism, in subcellular targeting as an intramolecular chaperone in conformational folding of mature enzyme, or in acting as autoinhibitor during transport through the endomembrane system.[8,10]

The mature enzymes of plant and microbial PMEs share five segments of high sequence similarity[1,18,19,22,37,38] representing the PME signature patterns. Previously only a few phylogenetic studies comparing a limited sample of plant and microbial PMEs were available.[15,31,66] Very recently a comparison of 39 PME sequences was performed yielding a phylogenetic tree with four plant clusters and one microbial group.[39]

The present work is focused on the comparison of 127 amino acid sequences of family CE-8 members. Most of them are, however, only putative or PME-like proteins although all of them exhibit the five sequence segments characteristic of true PMEs. With help of the evolutionary trees based on the sequence alignments it was possible: (i) to categorize the studied PME sequences, especially the putative *Arabidopsis* and *Oryza* PMEs, into corresponding clades; and (ii) to revise the former conclusions resulting from the fact that a limited number of primary structures were studied previously.

2. Material and methods

The enzymes belonging to the family CE-8 used for this study are listed in Table 1. The CAZy web server[2] served as a base. Almost all sequences were retrieved from UniProt[40] as well as from GenBank[41] and the PMEs from *Citrus sinensis* (Navelina orange; REF_42_CITSI1a) and *Ficus awkeotsang* (REF_43_FICAW) were taken from the literature data.[42,43]

Sequence alignments were performed using the program CLUSTALW.[44] The values of sequence identity and similarity of PME sequences to *Daucus carota* PME mature enzyme were calculated by BLAST.[45] The evolutionary trees were calculated by the Neighbour-Joining method.[46] The Phylip format tree output was applied using the bootstrapping procedure;[47] the number of bootstrap trials used was 1000. The trees were drawn with the program TREEVIEW.[48]

A web-site (http://imb.savba.sk/~janecek/Papers/CE-8/) for this article was created containing the links to additional figures. It contains the following figures: sequence alignment of 70 PMEs of plant, fungal and bacterial origin Figure 1W, alignments of selected plant and *Arabidopsis* PMEs (Figs. 3W and 5W), and their evolutionary trees (Figs. 4W and 6W), and a complete list with characterization of PMEs used in the present study (Table 1W).

Table 1. The enzymes used in the present study[a]

Abbreviation[b]	Source[c]
	Plant PMEs
Q9SRX4_ARATH1	*A. thaliana*, ch. 1, gn. F22D16.20
O23038_ARATH2	*A. thaliana*, ch. 1, gn. YUP8H12.7
Q9LPX8_ARATH4	*A. thaliana*, ch. 1, gn. T23J18_33
Q9LPX7_ARATH5	*A. thaliana*, ch. 1, gn. T23J18.25
O49298_ARATH6	*A. thaliana*, ch. 1, gn. T26J12.4
Q42534_ARATH8	*A. thaliana*, ch. 1, gn. T18A20.6, AtPME2
Q9CAS7_ARATH9	*A. thaliana*, ch. 1, gn. T17F3.3
O64479_ARATH10	*A. thaliana*, ch. 2, gn. T20K24.17
Q9SIJ9_ARATH11	*A. thaliana*, ch. 2, gn. F2G1.12
O48711_ARATH12	*A. thaliana*, ch. 2, gn. T9J22.11
O48712_ARATH13	*A. thaliana*, ch. 2, gn. T9J22.2
Q9ZQA4_ARATH14	*A. thaliana*, ch. 2, gn. At2g36700
Q84WQ3_ARATH15	*A. thaliana*, ch. 2, gn. At2g36710
Q9SKX2_ARATH16	*A. thaliana*, ch. 2, gn. MFL8.9
O22149_ARATH17	*A. thaliana*, ch. 2, gn. F4L23.27
Q9TOP8_ARATH18	*A. thaliana*, ch. 2, gn. AtPME4
O22256_ARATH20	*A. thaliana*, ch. 2, gn. T30B22.15
Q9M9W6_ARATH21	*A. thaliana*, ch. 3, gn. F18C1.12
Q9M9W7_ARATH22	*A. thaliana*, ch. 3, gn. F18C1.11
Q9M7Y9_ARATH23	*A. thaliana*, ch. 3, gn. F3E22.3
Q9SG77_ARATH24	*A. thaliana*, ch. 3, gn. T7M13.21
Q9SG78_ARATH25	*A. thaliana*, ch. 3, gn. T7M13.20
Q9LUL8_ARATH26	*A. thaliana*, ch. 3, gn. MLN21.9
O49006_ARATH27	*A. thaliana*, ch. 3, gn. MLN21_9, AtPME3
Q9LRN4_ARATH29	*A. thaliana*, ch. 3, gn. MUJ8.3
NP_189437_ARATH30	*A. thaliana*, ch. 3, gn. At3g27980
Q9LVQ0_ARATH31	*A. thaliana*, ch. 3, gn. MXE2.4
Q9LXK7_ARATH32	*A. thaliana*, ch. 3, gn. F7K15_120
Q9STY3_ARATH33	*A. thaliana*, ch. 3, gn. T21L8.150
Q9M3B0_ARATH34	*A. thaliana*, ch. 3, gn. F2K15.80
Q9LYT5_ARATH35	*A. thaliana*, ch. 3, gn. F17J16_60
Q9LZZ0_ARATH36	*A. thaliana*, ch. 3, gn. T4C21_140
Q9M1Q7_ARATH37	*A. thaliana*, ch. 3, gn. T17J13.130, pollen sp.
O81320_ARATH38	*A. thaliana*, ch. 4, gn. F6N15.23
O81415_ARATH39	*A. thaliana*, ch. 4, gn. T2H3.6
O81301_ARATH40	*A. thaliana*, ch. 4, gn. T14P8.1
O81300_ARATH41	*A. thaliana*, ch. 4, gn. T14P8.14
O81516_ARATH42	*A. thaliana*, ch. 4, gn. T24M8.6
O23447_ARATH43	*A. thaliana*, ch. 4, gn. dl4030c
Q9SMY7_ARATH44	*A. thaliana*, ch. 4, gn. F4/10.150
Q9SMY6_ARATH45	*A. thaliana*, ch. 4, gn. F4/10.160
Q9FF78_ARATH46	*A. thaliana*, ch. 5, gn. MUG13.18
Q9FF77_ARATH47	*A. thaliana*, ch. 5, gn. At5g04970
Q9LY19_ARATH48	*A. thaliana*, ch. 5, gn. T2/1_120
Q9LY18_ARATH49	*A. thaliana*, ch. 5, gn. T2/1_130
Q9LY17_ARATH50	*A. thaliana*, ch. 5, gn. T2/_140
Q9LXD9_ARATH51	*A. thaliana*, ch. 5, gn. F7/14_50
Q8VYZ3_ARATH53	*A. thaliana*, ch. 5, gn. At5g19730
NP_197586_ARATH54	*A. thaliana*, ch. 5, gn. At5g20860
Q9FJ21_ARATH58	*A. thaliana*, ch. 5, gn. K21P3.5
Q9FHN5_ARATH59	*A. thaliana*, ch. 5, gn. K17N15.4[d]
Q9FHN4_ARATH60	*A. thaliana*, ch. 5, gn. K17N15.5
Q9FK05_ARATH61	*A. thaliana*, ch. 5, gn. K19E1.17
Q9FM79_ARATH62	*A. thaliana*, ch. 5, gn. MDF20.3
Q9FKF3_ARATH63	*A. thaliana*, ch. 5, gn. K11J9.21
Q8LD76_ARATH64	*A. thaliana*, ch. 5, gn. MUB3.16
Q43867_ARATH65	*A. thaliana*, ch. 1, gn. T18A20.7, AtPME1
O80721_ARATH67	*A. thaliana*, ch. 2, gn. F14M4.13, PME5[d]
P41510_BRANA	*Brassica napus*, cv. Westar, pollen, gn. Bp19
Q42608_BRACM	*Brassica campestris*, anther[d]
Q96548_CARPA	*C. papaya*, cv. Solo
O04888_CITSI1	*C. sinensis*, Valencia orange, gn. PECS1.1[d]
Ref_42_CITSI1a	*C. sinensis*, Navelina orange, gn. OPME1a[d]
O04889_CITSI2	*C. sinensis*, Valencia orange, gn. PECS2.1[d]
Q8GS16_CITSI4	*C. sinensis*, Sweer orange, gn. PME4
P83218_DAUCA	*D. carota*, cv. TipTop[d]
Ref_43_FICAW	*F. awkeotsang*, cv. Makino
Q94FS6_LINUS1	*L. usitatissimum*, cv. Ariane, gn. LuPME1
Q9FVF9_LINUS3	*L. usitatissimum*, cv. Ariane, gn. PME3
Q94FS5_LINUS5	*Linum utisatissimum*, cv. Ariane, gn. LuPME5
Q96577_LYCES1	*L. esculentum*, str. VFNT Cherry, fruit, gn. PME1
Q96575_LYCES2	*L. esculentum*, str. VFNT Cherry, fruit, gn. PME2
Q96576_LYCES3	*L. esculentum*, str. VFNT Cherry, fruit, gn. PME3
P14280_LYCES4	*L. esculentum*, cv. Ailsa Craig, fruit, cl. pB8[d]
P09607_LYCES5	*L. esculentum*, Ailsa Craig, fruit, cl. pB16[d]
Q43143_LYCES6	*L. esculentum*, RioGrande, leaves, PMEU1[d]
Q42920_MEDSA	*Medicago sativa*, str. C2-4, pollen, gn. P65
Q9SC89_MEDTR1	*Medicago truncantula*, cv. Jemalong, gn. Pef1[d]
Q9SC90_MEDTR2	*M. truncantula*, cv. Jemalong, gn. Per[d]
Q84V57_NICBE	*N. benthamiana*, taxon: 4100
Q42935_NICPL	*Nicotiana plumbaginifolia*, somatic embryos[d]
Q9LEB0_TOBAC	*N. tabacum*, cv. Samsun, cell wallls, gn. PME[d]
Q9S767_ORYSA1	*O. sativa*, cv. Nipponbare, ch. 1, gn. P0705D01
Q9FSQ0_ORYSA2	*O. sativa ind.*, cv. Gangluai, ch. 4, gn. H0423H10.13
Q93VX7_ORYSA3	*O. sativa*, cv. Nipponbare, ch. 1, gn. P0682B08
Q9LGX7_ORYSA4	*O. sativa*, cv. Nipponbare, ch. 1, gn. P0702F03
Q8S122_ORYSA5	*O. sativa*, cv. Nipponbare, ch. 1, gn. P0415A04
Q9RYM8_ORYSA6	*O. sativa*, cv. Nipponbare, ch. 1, gn. P0663E10
Q7XEU2_ORYSA7	*O. sativa*, cv. Nipponbare, ch. 10, gn. OSJNBa0060A14
Q8LQA0_ORYSA8	*O. sativa*, cv. Nipponbare, ch. 1, gn. B1011A07
Q8LJK2_ORYSA10	*O. sativa*, cv. Nipponbare, ch. 1, gn. P0018C10
Q8LQ65_ORYSA11	*O. sativa*, cv. Nipponbare, ch. 1, gn. P0439E07
BAC16045_ORYSA12	*O. sativa*, cv. Nipponbare, ch. 1, gn. P0597G07
CAD40826_ORYSA13	*O. sativa*, cv. Nipponbare, ch. 4, gn. OSJNBa0006B20
Q43043_PETIN	*Petunia inflata*, pollen, gn. PPE1
Q9M5J0_PHAAU1	*P. aureus*, mung bean Wilczek hypocot, gn. PME2[d]
Q43234_PHAAU2	*P. aureus*, mung bean Wilzeck hypocot, gn. PME[d]
Q43111_PHAVU	*P. vulgaris*, green bean pods, cv. Masai, MPE3[d]
O24298_PEA	*Pisum sativum*, cv. Purple podded, gn. PMEE
Q9FY03_POPTN1	*P. tremula*, poplar cambial region, gn. pme1
Q9FEU1_POPTN2	*P. tremula*, poplar cambial region, gn. pme2
Q9FVU0_POPTN3	*P. tremula*, poplar cambial region, gn. pme3
Q9FET9_POPTN4	*P. tremula*, poplar cambial region, gn. pme4

(*continued on next page*)

Table 1 (*continued*)

Abbreviation[b]	Source[c]
Q43062_PRUPE	*Prunus persica*, cv. Coronet, peach fruit, cl. PPE8B
Q9MBB6_SALGI	*S. gilgiana*, male flower pollen, gn. SgPME1[d]
Q8RVX1_SESRO	*Sesbania rostrata*, developing stem, gn. pme1
Q96497_SILPR	*S. pratensis*, flower bud, cl. lambda GEM2
Q9SEE7_SOLTU1	*S. tuberosum*, cv. Desiree, leaves, gn. Pest1[d]
Q9SEE6_SOLTU2	*S. tuberosum*, cv. Desiree, epidermal Pest2[d]
Q9SW71_VITRI	*V. riparia*, flower bud, gn. PME[d]
Q94B16_VITVI	*Vitis vinifera*, cv. Shiraz, gn. PME1[d]
O24596_MAIZE	*Zea mays*, str. A 188, pollen, gn. C 5
	Fungal PMEs
Q12535_ASPAC	*Aspergillus aculeatus*, str. KSM510, gn. PME1[d]
O94162_ASPOR	*Aspergillus oryzae*, str. KBN616, gn. TEFI, PMEA[d]
P17872_ASPTU	*Aspergillus tubigensis* (*A. niger*), str. RH5344, PMEA[d]
Q9C2Y1_BOTCI1	*Botrytis cinerea*, str. T4, gn. bcpme1[d]
Q8X116_BOTCI2	*B. cinerea*, str. Bd90, gn. bcpme2[d]
Q9Y881_COCCA	*C. carbonum*, str. SB111, gn. PME1[d]
	Bacterial PMEs
Q47474_ERWCH1	*E. chrysanthemi*, str. B3937, gn. PemB[d]
P07863_ERWCH2	*E. chrysanthemi*, str. B374, (str. 3937), PemA[d]
Q9AN15_BRAJA	*B. japonicum*, str. 110spc4, gn. ID637
Q97DU8_CLOAB	*Clostridium acetobutylicum*, Atcc 824, gn. CAC3373
P24791_RALSO2	*Ralstonia solanacearum*, str. DSM 50905, gn. PME2
P58601_RALSO1	*R. solanacearum*, str. GMI 1000, gn. PME1
Q93RU7_STRCO	*S. coelicolor*, str. A3 (2), gn. SCI39.26
Q8PE60_XANCA1	*Xanthomonas campestris*, ATCC33913, gn. XCC0121
Q8P8H6_XANCA2	*X. campestris*, ATCC33413, gn. XCC2265
Q8ZIR5_YERPE	*Yersinia pestis*, str. CO 92, gn. YPO 0424

[a] The extended version of this table containing the accession numbers and several characteristics is available at http://imb.savba.sk/~janecek/Papers/CE-8/table1.htm.

[b] The abbreviations consist of the UniProt Accession numbers[40] and UniProt species code (http://www.expasy.org/cgi-bin/speclist), except for NP_189437_ARATH30, NP_197586_ARATH54, BAC16045_ORYSA12, CAD40826_ORYSA13 (protein_ids from GenBank;[41] not available in UniProt) and Ref_42_CITSI1a, Ref_43_FICAW (references, available in neither UniProt nor GenBank).

[c] ch., chromosome; gn., gene; cl., clone; cv. cultivar.

[d] Sequences are marked in data base as EC 3.1.1.11.

3. Results and discussion

This study brings as complete as possible evolutionary picture of pectin methylesterases, the family CE-8, based on the comparison of 127 amino acid sequences of true (experimentally confirmed) and/or putative PMEs and some fragments (listed in Tables 1 and 1W) by analyzing their sequence alignments and evolutionary trees. Concerning the accuracy of the alignments, all the functionally important residues of PMEs were aligned correctly. However, there can be a low background noise in the alignment, especially in the parts where more gaps were introduced.

First the separate sequence alignments and phylogenetic trees of potential mature enzymes for selected 58 *A. thaliana* and 12 *O. sativa* sequences (mostly putative PMEs) were prepared. Twelve *Arabidopsis* and seven *Oryza* PMEs were then selected from the clustered groups. These 19 *Arabidopsis* and *Orysa* PMEs together with 35 additional plant, 10 bacterial and 6 fungal PME sequences were used for the final alignment (Figs. 1 and 1W) and evolutionary tree (Fig. 2). The tree clearly supports the division of PMEs into several phylogenetic clades.[39] Thus the present set of studied PMEs (Tables 1 and 1W) is proposed to consist of Plant clades 1, 1a, 2, 2a, 3, 4, X1 and X2, the clade Fungi (including the plant PME from *Vitis riparia*) and the clade Bacteria (Figs. 1, 1W and 2).

In the group of plant enzymes, 48 *A. thaliana* sequences were compared with 9 *O. sativa* sequences and 36 plant PMEs sequences without the one from *V. riparia* (Fig. 3W). A phylogenetic tree (Fig. 4W) was constructed also based on this alignment.

The alignment of exclusive *A. thaliana* PME sequences (Fig. 5W) yielded a few putative *Arabidopsis* PMEs that are proposed to form in their tree the clade Plant X3 (Fig. 6W).

Unless otherwise specified, all amino acid numbering throughout the text corresponds to the mature PME of *D. carota* (P83218_DAUCA)[39] anchoring the alignments (Figs. 1W and 3W).

3.1. General comparison of amino acid sequences

3.1.1. Conserved and conservative residues; segments typical for PMEs. In 123 compared sequences of complete mature enzymes there were found only six strictly conserved residues: Gly44, Gly154, Asp157, Gly161, Arg225 and Trp227, as well as six conservative residues in positions: Ile39, Ser86, Ser137, Ile152, Ile159 and Leu223. Furthermore the following residues with more than 90% occurrence were found: Tyr46, Glu48, Gly63, Phe96, Thr137, Leu138, Phe158, Ile159, Phe160, Leu223, Gly224, Glu265, Gly270, Gly272 and Trp307. The residues Ala187, Gly224 and Arg278 are conserved in all plant PMEs except of *V. riparia* (Q9SW71_VITRI), which exhibits unambiguous sequence similarities with fungal PMEs. Most of these highly conserved and conservatively residues are involved in the five segments characteristic for PMEs (Fig. 1).

The first segment (GxYxE) is conserved in all PMEs with only two replacements of Tyr46: in NP_197586_ARATH54 (Gly) and O23038_ARATH2 (a gap) and four replacements of Glu48: in Q47474_ERWCH1 (Gly), Q9RYM8_ORYSA6 (Arg), Q9LVQ0_ARATH31 (Gln) and NP_197586_ARATH54 (Tyr). Gly44 is an invariantly conserved residue.

The second segment (QAVAL) is present in unmodified form in 60.4% of all PMEs and in 70% of plant

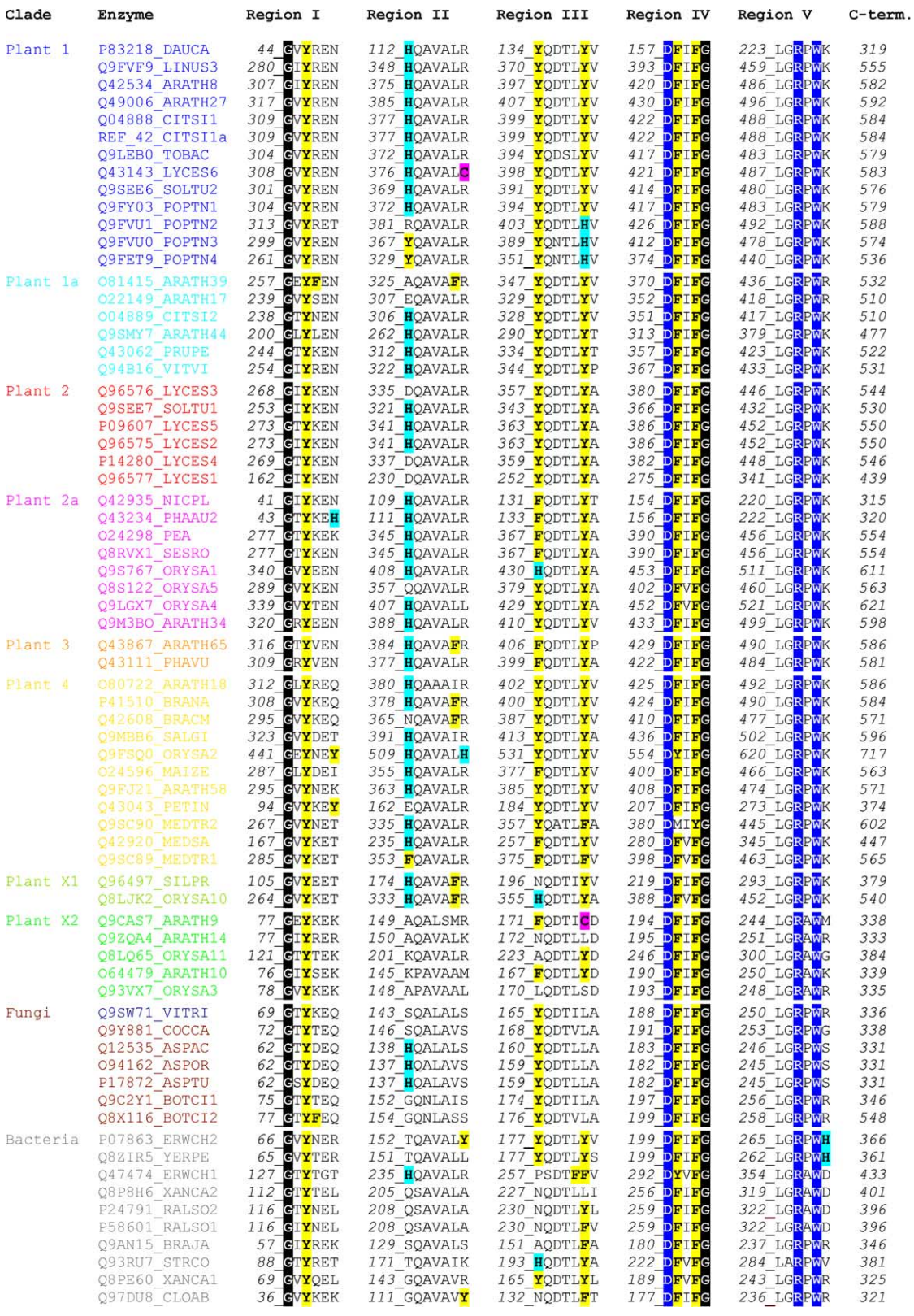

Figure 1. Amino acid sequence alignment of conserved segments of 70 selected PMEs. Abbreviations used are listed in Table 1. The sequences are arranged into the clades. Since the carrot PME was taken as the leading member it is placed in the first line. The division and colouring of clade members is based on the evolutionary tree (Fig. 2). The PME from *V. riparia* (Q9SW71_VITRI coloured indigo) exhibits the unambiguous sequence similarity to fungal PMEs. The three CE-8 catalytic residues, Asp157, Arg225 and Trp227 (P83218_DAUCA numbering), are highlighted in blue. Two of the three additional strictly conserved residues, Gly44, Gly154 and Gly161, are highlighted in black. Cys, His, and Phe (Tyr) residues are highlighted in pink, turquoise and yellow, respectively.

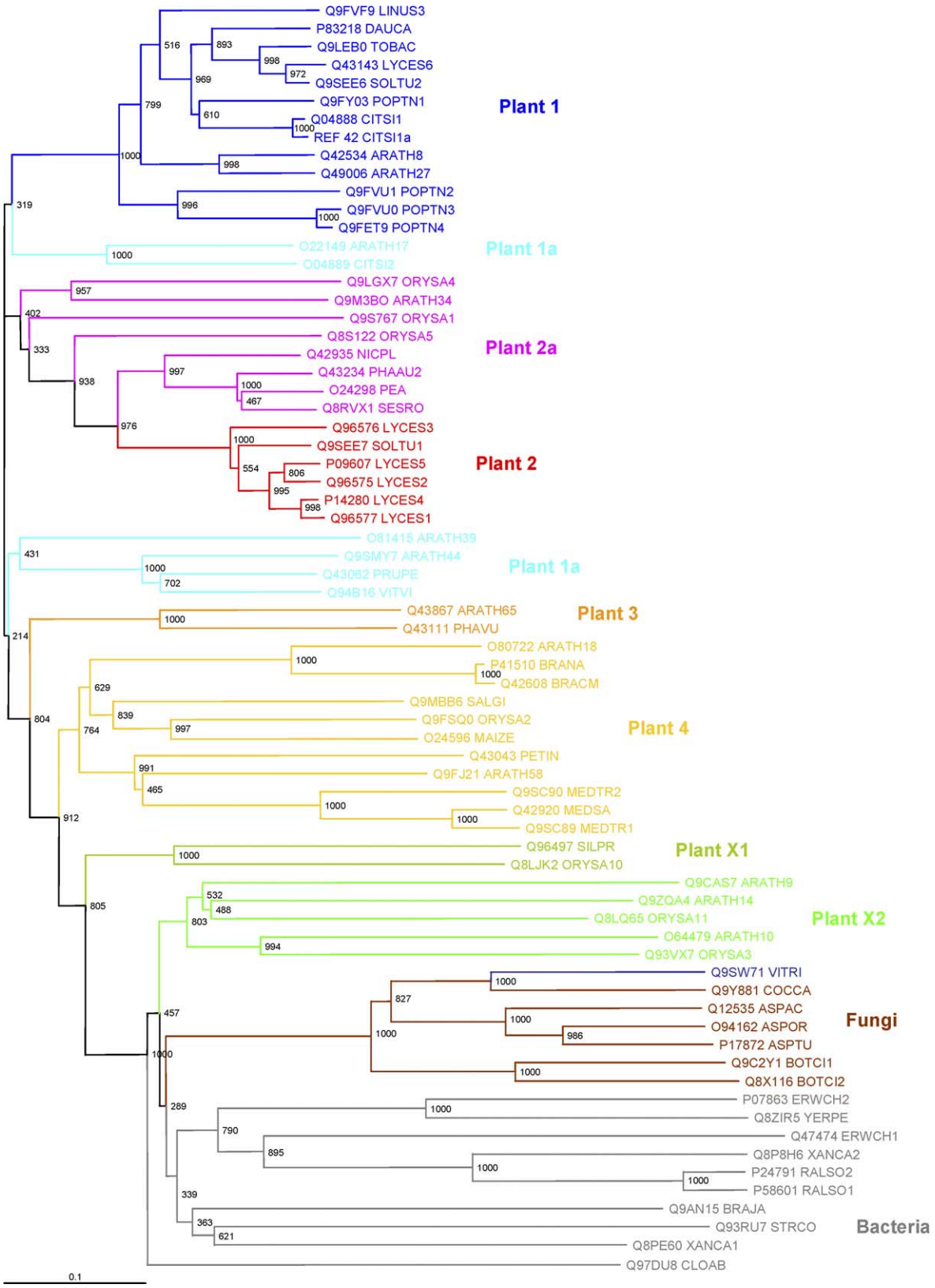

Figure 2. Evolutionary tree of 70 selected PMEs. The tree is based on the alignment of mature enzyme sequences (Fig. 1W). Branch lengths are proportional to sequence divergence. The individual PME clades are distinguished from each other by different colours, except for the plant PME from *V. riparia* (Q9SW71_VITRI) that belongs to the clade Fungi. The abbreviated member names are defined in Table 1. The numbers along branches are bootstrap values (1000 replicates).

PMEs. The residue Gln113 is replaced by Ser in three bacterial PMEs and by Pro and once by Arg in five plant enzymes.

In the third segment (QDTL) the residue Gln135 is conserved in fungal and plant enzymes, while in one bacterial PME (Q47474_ERWCH1) is replaced by Ser. The Asp136, with proposed function in the active site of bacterial and plant PMEs,[23,24] was found to be non-conserved in all studied PMEs: in Q9FVU0_POPTN3 and Q9FET4_POPTN4 it is replaced by Asn and in Q9SC90_MEDTR2 by Ala. It is worth mentioning that in most plant PMEs at both sides of the third segment there are Tyr or Phe residues, except for the Q9FEU1_POPTN2, Q9FVU0_POPTN3 and Q9FET4_POPTN4, as well as for six *Arabidopsis* ARATH PMEs, where His and Cys can be found (Figs. 1, 3W and 5W). In bacterial enzymes Tyr is present on the left side of this segment in three members only, while the remaining bacterial PMEs contain on left side a nonaromatic residue. In all fungal PMEs the aromatic residue on the right side is replaced by Leu.

The fourth segment (DFIFG) contains the strictly conserved Asp159 playing the role in the PME active site, and the aromatic residue Phe160, which is involved in the substrate binding.[23,24] Remarkably in all investigated PMEs the Phe158 is only once substituted by a nonaromatic residue (Met in Q9SC90_MEDTR2) and the Phe160 is replaced once with Asp (in Q42608_BRACM) and four times with Cys (Q9LPX7_ARATH5, Q9SIJ9_ARATH11, NP_189437_ARATH30 and O81516_ARATH42).

The fifth segment (LGRPW) contains two strictly conserved residues Arg225 and Trp227, which are involved in the PME active site.[23,24] The Leu223 and Gly224 are in two plant PMEs (Q9LXD9_ARATH51 and Q9FHN4_ARATH60) replaced by a conservative Met and only once replaced by Ala (Q93RU7_STRCO), respectively, while the Pro226 is in all fungal (including the Q9SW71_VITRI), 4 bacterial and 11 plant PMEs replaced by Ala. There is an aromatic residue on the left side of this segment in most cases.

3.1.2. Cysteine residues. No invariantly conserved Cys residue was found throughout the mature enzymes of 123 PMEs studied. The number of cysteines in a single PME varies between 1 (in Q9PE60_XANCA1) and 14 residues (in Q9LVQ0_ARATH31). Several cysteines are conserved only in the individual plant clades (Table 2) or only in some branches of clades Plant 4, Plant X2, Fungi and Bacteria (Figs. 1W, 3W and 5W). The other cysteines are distributed irregularly in the sequences.

Cysteine in position of Thr102, which is conserved only in the clade Plant 2, is present also in *F. awkeotsang*[43] (an acidic PME with pI = 4.39) and in the three *Arabidopsis* putative PMEs (Q9LPX3_ARATH5, NP_189437_ARATH30 and O81516_ARATH42) that, in the *Arabidopsis* phylogenetic tree, were grouped into one clade Plant X3 (Figs. 5W and 6W). The members of the clade Plant X1 contain two strictly conserved cysteines (Ala12 and Ser238) (Fig. 3W).

Cysteine in position of Ser204 is replaced in 123 PMEs 37 times by Ser and only twice by Ala or Gly residues (in bacterial PMEs: Q93RU7_STRCO and Q8PE60_XANCA1).

In fungal and bacterial PMEs no conserved cysteines were found.

The presence of two disulfide bridges was described in one form of tomato fruit PME (P14280_LYCES4).[37] By similarity it is possible to assume the presence of the same bridges in all members of the clade Plant 2.

The 3D structure of PME from *E. chrysanthemi* PemA[23] has roughly equal amounts of the disulfide-stacked and disulfide-bridged forms (Cys150 and Cys170), while in the 3D structure of carrot PME[24] no disulfide bridge exists and the three cysteines (Cys129, Cys150 and Cys170) form the internal stacking ladder. It should be possible to generalize this fact to all members of the clade Plant 1 possessing the three conserved Cys residues corresponding to those of carrot PME.

In the bacterial PME group two cysteines, corresponding to those of P07863_ERWCH2, are present in Q8ZIR5_YERPE, Q9AN15_BRAJA, Q93RU7_STRCO and Q97DU8_CLOAB. In remaining bacterial

Table 2. Cysteine residues conserved in particular clades[a]

Position[b]	All PMEs	Plants	Fungi	Bacteria	Conserved in clades[c]
Ala12	3.2	3.7	0.0	0.0	Plant X1
Thr102	8.1	9.3	0.0	0.0	Plant 2, X3
Cys129	86.2	92.5	66.6	30.0	Plant 1, 2, 3, 4, X3
Cys150	78.0	82.2	33.3	60.0	Plant 3, X1, X2, X3
Phe160	3.2	3.7	0.0	0.0	Plant X3
Cys170	82.1	85.0	50.0	70.0	Plant 1, 2, 3, X3
Ser204	69.1	77.6	0.0	20.0	Plant 2, X1, X2, X3
Ser238	5.7	6.5	0.0	0.0	Plant X1
Asn268	14.6	16.8	0.0	0.0	Plant X2

[a] The presence of Cys residues in 123 PMEs, 107 plant sequences (except for fragments), 6 fungal and 10 bacterial sequences expressed in %.
[b] The *D. carota* PME numbering.
[c] The conservation of Cys residues in the particular clades was assigned based on sequence alignments.

PMEs Cys150 is replaced by Ser and Cys170 by Val and Ala.

3.1.3. Histidine residues. The negative effects of esterase inhibitors as well as Ellman reagent on the activity of purified tomato PME revealed that this enzyme is neither a serine-type esterase, nor SH-enzyme; while the strong inhibitory effect of iodine suggested the possible role of His or Tyr residues in the action of PME.[49] The 3D structures showed that PME is a new type of aspartyl esterase containing neither the α/β hydrolaze fold, nor a catalytic Ser-His-Asp triad.[23,24] The H137A mutation in *Aspergillus niger* PME (His112 in P83218_DAUCA) resulted in the complete loss of enzyme activity indicating that histidine could be located in the active site,[50] however the modification of His residues in tomato and *A. niger* PMEs by diethyl pyrocarbonate suggested that accessible histidines do not have the active-site functions in these PMEs, but contribute to their overall structural stability.[51]

In the set of 123 PME sequences of mature enzymes it was not possible to find any conserved His residues, the fact recognised already from previous alignments.[15,18,19,37] The histidine number in a single sequence ranging from 0 (*V. riparia*) to 11 (*Bradyrhizobium japonicum*). Two histidines are nevertheless present in PME sequences in larger amount: (i) His112 conserved in clades Plant 3 and X1, and present in 55.1% in the set of 123 PMEs; and (ii) His141, which is present only in plant enzymes (in 71.2%), is conserved in clades Plant 1, 2, 2a, 3, X1 and X3 (Figs. 1W, 3W and 5W).

3.1.4. Aromatic residues. As shown by 3D structures of bacterial and plant PMEs,[23,24] aromatic residues are involved in the substrate-binding site by forming a cluster of aromatic side-chains. The PME from *E. chrysanthemi* (P07863_ERWCH2) contains three aromatic residues (Tyr158, Tyr181 and Phe202) on the exposed surface of parallel β-sheet PB1 forming an external aromatic stack near the strictly conserved Trp269 (Trp227 in carrot PME) on a T1 loop.[23] This finding is not possible to generalize for all PMEs, because the first aromatic residue Tyr158 (Arg118) is present only in P07863_ERWCH2 and Q97DU8_CLOAB; in fungal enzymes there is a conserved serine in this position, while in most plant PMEs arginine is present (Figs. 1W, 3W and 5W). The second one, Tyr181 (Tyr139), is conserved in most plant clades (Table 3), while in all fungal PMEs and in Q8P8H6_XANCA2 it is replaced by Leu; in four bacterial PMEs it is replaced by a phenylalanine (Fig. 1W). The third residue, Phe202 (Phe160) is highly conserved in all PMEs except for the clade Plant X3 where it is replaced by cysteine residues (Figs. 3W and 5W).

In the 3D structure of carrot PME,[24] the central part of the long shallow cleft across the molecule is lined by aromatic residues: Phe84, Tyr139, Phe160, Tyr222, Trp227, Phe250 and Trp252. Most of these residues are conserved in several plant clades (Table 3), except for the Phe250, which is present only in P83218_DAUCA. It is worth mentioning, however, that Phe250 is preceded by a tryptophane (Fig. 1) that is well conserved in plant and fungal PMEs. It could eventually play the role of Phe250 in substrate binding.

There are 55 aromatic residues in PME sequences conserved in individual clades (Table 3). The function of most of them has still to be elucidated, except of those mentioned above and of Tyr46, which is member of the first characteristic conserved segment (GxYxE; Fig. 1). This Tyr46 is located far from active site, is internal and involved in two hydrogen bonds thus stabilizing that part of the structure.[52]

Most aromatic residues are present alternatively as Tyr, Phe or Trp. Only Trp227 has been found to be conserved strictly.

3.2. Sequence similarities and evolutionary relationships within the PME clades

The phylogenetic tree (Fig. 2) shows the relationships among the selected 70 PMEs. The tree is based on the alignment of sequences (Fig. 1W). The plant PMEs are grouped into eight clades marked as Plant 1, 2, 3, 4, 1a, 2a, X1 and X2. The fungal enzymes form their own clade including the plant PME from *V. riparia*. The bacterial PMEs are also well separated with the exception of PME from *Clostridium acetylobutylicum*.

For illustration, the expanded trees focused more on the plant PMEs as well as on the *A. thaliana* PMEs were calculated and these are presented on the web-site as Figures 4W and 6W, respectively.

It should be pointed out that clades Plant 1a and Plant 2a are less compact since these are formed by more clusters that are not rooted in a single node (Figs. 2, 4W and 6W). They contain the members that are closely as well as more distantly related to the clades Plant 1 and Plant 2. The clade Plant X2 seems to be most distantly related to all plant counterparts due to its positioning in the microbial part of the tree (Fig. 2). There has been no PME from *A. thaliana* as yet that would belong to the clade Plant 2, but a new clade, Plant X3, has emerged (Figs. 5W and 6W).

3.2.1. The clade Plant 1. This clade contains 13 PMEs and the *pre pro* sequence of 10 members (without separately branched three *Populus* PMEs) share high identity and similarity and contain five conserved cysteines and two conserved pgs. The region preceding the mature enzyme possesses a conserved segment WPxWxxxxDRRLLQ, except for the Q9FEU1_POPTN2 (having Asp → Asn), and Q9FVU0_POPTN3 and Q9FET4_POPTN4 (having Leu → Phe). The mature enzyme starts immediately after this conserved

Table 3. Conservation of aromatic residues in mature enzymes[a]

Position[b]	All PMEs	Plants	Fungi	Bacteria	Conserved in clades[c]
Tyr19	88.6	87.8	100.0	90.0	Plant 1, 2, X3, Fungi
Tyr37	60.1	68.2	0.0	30.0	Plant 1, 2, X1, X3
Arg40	56.9	58.9	100.0	10.0	Plant 2, 3, X3, Fungi
Tyr46	98.4	98.2	100.0	100.0	Plant 1, 2, 3, 4, 1a, 2a, X1, X3, Fungi, Bacteria
Asp51	8.1	2.8	100.0	10.0	Fungi
Lys57	7.3	8.4	0.0	0.0	Plant 3
Phe61	26.0	28.9	0.0	10.0	Plant 1
Ile71	8.1	3.7	100.0	0.0	Fungi
Ala73	22.7	19.6	0.0	70.0	Plant X2
Phe84	73.2	81.3	0.0	30.0	Plant 1, 2, 3, X1, X3
Val89	30.1	32.7	33.3	0.0	Plant 3
Phe96	91.8	96.2	0.0	100.0	Plant 2, 3, 4, 1a, X1, X3, Bacteria
Arg99	7.3	2.8	100.0	0.0	Fungi
Phe103	67.5	72.9	0.0	50.0	Plant 1, X3
Leu117	7.3	8.4	0.0	0.0	Plant X1
Phe126	66.6	65.4	100.0	60.0	Plant 1, 3, Fungi
Tyr127	65.0	69.1	100.0	0.0	Plant 1, 3, X3, Fungi
Ile131	45.5	44.8	100.0	0.0	Fungi
Tyr134	82.1	86.0	100.0	30.0	Plant 1, 2, 3, 2a, X3, Fungi
Tyr139	81.3	85.0	0.0	90.0	Plant 2, 3, 4, 2a, X1, X3
Phe146	93.5	98.1	0.0	100.0	Plant 1, 2, 3, 4, 1a, 2a, X2, X3, Bacteria
Phe147	97.6	98.1	100.0	90.0	Plant 1, 2, 3, 4, 1a, 2a, X1, X3, Fungi
Phe151	39.0	35.5	50.0	70.0	Plant 2, X3
Phe158	98.4	98.1	100.0	100.0	Plant 1, 2, 3, 1a, 2a, X1, X2, X3, Fungi, Bacteria
Phe160	95.9	95.3	100.0	100.0	Plant 1, 2, 3, 1a, 2a, X1, Fungi, Bacteria
Val166	6.5	0.9	100.0	10.0	Fungi
Leu167	73.2	69.1	100.0	100.0	Plant 2, X3, Fungi, Bacteria
Asp169	2.4	2.8	0.0	0.0	Plant X3
Gly198	5.7	0.9	100.0	0.0	Fungi
Ile199	41.4	34.6	100.0	80.0	Plant X2, X3, Fungi
Ile201	23.6	23.3	16.6	30.0	Plant X2
Lys203	8.9	10.3	0.0	0.0	Plant 2
Leu212	7.3	8.4	0.0	0.0	Plant X1
Pro214*	3.2	3.7	0.0	0.0	Plant X1
Phe219	25.2	28.0	16.6	0.0	Plant 1, 2
Tyr222	91.8	97.2	100.0	30.0	Plant 1, 2, 3, 4, 1a, 2a, X1, X3, Fungi
Trp227	100.0	100.0	100.0	100.0	Plant 1, 2, 3, 4, 1a, 2a, X1, X2, X3, Fungi, Bacteria
Tyr230	87.0	91.6	100.0	30.0	Plant 1, 3, 4, 1a, 2a, X1, X3, Fungi
Val235	42.3	38.3	100.0	50.0	Plant X1, Fungi
Ser239	45.5	46.7	33.3	40.0	Plant 2, X3
Trp249	94.3	100.0	100.0	30.0	Plant 1, 2, 3, 4, 1a, 2a, X1, X2, X3, Fungi
Trp252	88.6	91.6	100.0	50.0	Plant 1, 2, 3, 2a, X1, X3, Fungi
Phe256	55.2	62.6	0.0	10.0	Plant 1, 2, X1
Tyr262	69.1	75.7	0.0	40.0	Plant 2, 2a, X1
Tyr263	94.3	99.1	100.0	40.0	Plant 1, 2, 3, 4, 1a, 2a, X1, X3, Fungi
Tyr266	87.8	90.6	100.0	50.0	Plant 2, 3, 1a, 2a, X3, Fungi
Trp281	89.4	98.1	0.0	50.0	Plant 1, 2, 3, 4, 1a, 2a, X1, X3
Phe284	61.8	65.4	100.0	0.0	Plant 2, X3, Fungi
Phe296	83.7	92.5	16.6	30.0	Plant 1, 2, 3, 1a, 2a, X1, X3
Phe301	68.3	75.7	0.0	30.0	Plant 1, X1, X2
Ile302	11.4	6.5	100.0	10.0	Fungi
Trp307	93.5	97.2	100.0	50.0	Plant 1, 2, 3, 1a, 2a, X1, X2, X3, Fungi
Thr312	5.7	0.9	100.0	0.0	Fungi
Phe313	14.6	15.9	16.6	0.0	Plant 1
Phe315	63.4	72.8	0.0	0.0	Plant 1, 2, 1a, 2a, X3

[a] The presence of aromatic residues in 123 PMEs, 107 plant sequences (except for fragments), 6 fungal and 10 bacterial sequences expressed in %.
[b] The *D. carota* PME numbering. Except for the residues in positions Tyr46, Asp51 (Tyr in Fungi), Lys57 (Trp in Plant 3), Phe61, Val89 (Tyr in Plant 3), Phe96, Arg99 (Tyr in Fungi), Phe103, Leu117 (Phe in Plant X1), Tyr127, Ile131 (Phe in Fungi), Phe160, Val166 (Trp in Fungi), Asp169 (Phe in Plant X3), Lys203 (Phe in Plant 2), Leu212 (Tyr in Plant X1), Pro214* (second gap succeeding P214: Tyr/Phe in Plant X1), Trp227 and Thr312 (Tyr in Fungi), all remaining aromatic residues are alternatively present as Tyr, Phe or Trp.
[c] The conservation of aromatic residues in the particular clades was assigned based on sequence alignments.

segment as documented in *C. sinensis* PME.[42] The sequence identity of mature enzymes of the clade Plant 1 members to the carrot PME sequence varies between 73% and 87% (sequence similarity is 82–91%). The overall mutual identity and similarity is 55.3% and 69.7%, respectively. With regard to pgs in the mature enzyme region, again only three *Populus tremula* PMEs contain one to three pgs, while the remaining 10 members including Q9FY03_POPTN1 do not contain any. The entire clade Plant 1 can thus be at present divided into two groups: the three PMEs from *Populus* and the 10 members including the Q9FY03_POPTN1. The PME from *D. carota*[39,53] can be considered to be the leading member of this clade since also the 3D structure is known for this PME.[24] PME from *C. sinensis*; Navelina orange[42] differs from Q04888_CITSI1 (Valencia orange)[36] in nine residues only. The enzyme Q43143_LYCES6 (*Lycopersicon esculentum*; leaves) is ubiquitously expressed enzymatically functional PME,[14] as well as *Nicotiana tabacum* cell walls PME,[54] which was proposed to be a host cell receptor involved in the cell-to-cell movement of the tobacco mosaic virus. The PME from *Nicotiana benthamiana* (Q84V57_NICBE) exhibits similar amino acid composition as Q9LEB0_TOBAC, differing in ORF only in three residues, is situated in the same branch with Q9SEE6_SOLTU2 (*Solanum tuberosum*; epidermal gene *Pest2*). Q9FY03_POPTN1 PME1 (*P. tremula*) is distributed across the cambial region along with the other isoforms differentiating from each other by pI.[55] This PME is placed among the subgroup of 10 PMEs of this clade Plant 1, while three other *P. tremula* PMEs are positioned separately at the adjacent branch; they differ by presence of conserved histidine in position of Tyr139 as well as by presence of pgs discussed above. Q9FVF9_LINUS3 and Q94FS5_LINUS5 (*Linum usitatissimum*) are two of three genes coding for PME[56] and possess great common identity and similarity, both clustered in the clade Plant 1 (however, were not included in Figs. 1 and 2). Q42534_ARATH8 (*A. thaliana*, AtPME2) represents a basic PME isoform with specific biological function.[31] This PME together with the O49006_ARATH27 (AtPME3)[16] are the only 2 of 67 *Arabidopsis* PME sequences (Table 1) located in the clade Plant 1.

3.2.2. The clade Plant 2. This small and compact cluster contains six members (all belonging to the family *Solanaceae*); five of them are from tomato fruit and one is from potato leaf. In both the *pre pro* region as well as mature enzyme of these members exhibit a high degree of overall mutual identity and similarity (more than 70%). The lengths of ORFs are in the range 446–550 residues and all members possess a conserved segment DRKLMESSGKD preceding the mature enzyme. In contrast to the clade Plant 1, where the mature enzyme starts in a distance of two or three residues after a pair of basic residues (Arg-Arg), the mature enzyme in the clade Plant 2 starts eight residues after that dipeptide (Arg-Lys). The *pre pro* region of all clade Plant 2 members possesses conserved one pgs, four cysteines, eight aromatic residues and two histidines. Mature enzymes of three tomato PMEs do not contain any pgs, while Q96577_LYCES1 and Q96576_LYCES3, as well as Q9SEE6_SOLTU2 have one to three pgs. The members of clade Plant 2 differ from the other PMEs by the presence of four conserved cysteine residues in mature enzymes (Table 2), which form in P14280_LYCES4 two disulfide bridges.[37] Furthermore in this clade, there are conserved 2 histidines and 30 aromatic residues (Table 3). Of these, two tyrosines (Phe151 and Ser239) and one phenylalanine (Lys203) are characteristic of only this clade. The clade Plant 2-specific conserved segment (124_SVINRC) is present also in four members of the clade Plant 2a. On the other hand, the Phe301, characteristic of most plant PMEs, is in the clade Plant 2 and in six members of the clade Plant 2a replaced by conserved Leu residue. The most deeply studied member of the clade Plant 2 is the tomato PME P14280_LYCES4, the product of the clone pB8 as one of the eight isozymes detected as the major isoform in tomato fruit and originally called PE2.[10,57] The N-terminal isoleucine residue of this major form was identified almost 30 years ago[58] and later was confirmed by amino acid sequencing.[59] This PME sequence was completed and modified by sequencing of corresponding cDNA.[60] The final characterization and classification of two tomato fruit PME isoforms was performed after isolation and sequencing of cDNA clones pB8 (P14280_LYCES4) and pB16 (P09607_LYCES5).[10] The next three tomato fruit PMEs represent three tandemly arranged PMEs from specific genes[61] with great similarity to both P14280_LYCES4 and P09602_LYCES6.

3.2.3. The clade Plant 3. This clade contains only two PMEs (Figs. 1W and 2), those from *A. thaliana* (Q43867_ARATH65; AtPME1) and *Phaseolus vulgaris* (Q43111_PHAVU; MPE3). In the extended version focused on plant PMEs three additional *Arabidopsis* members were included (Figs. 3W and 4W).

The ORF lengths of PMEs of this clade range between 561 and 586 residues, except for the Q9LUL8_ARATH26 containing 968 residues (the longest putative PME). The *pre pro* region exhibits 14.0% internal identity and 15.7% similarity, contains four conserved cysteines, two conserved aromatic residues (Phe and Trp), and one conserved pgs. In contrast to the clades Plant 1 and Plant 2, the members of this clade possess the conserved segment RRLL twice in the region close to potential cleavage site of the mature enzyme. The mature PMEs exhibit the sequence identity and similarity to carrot PME between 49–54% and 63–68%,

respectively. The overall mutual identity and similarity was recognised as 37.0% and 53.8%, respectively. The mature enzymes of this clade contain one strictly conserved proline (Phe258), three conserved cysteines (Cys129, Cys150 and Cys170), two conserved histidines (His112 and His141), and 24 conserved aromatic residues (Table 3). Interestingly, the tryptophane (Lys57) is present, except for this clade, only in four additional members of the clade Plant 1a (Figs. 1W and 3W).

3.2.4. The clade Plant 4. It covers 11 PME sequences from flower parts, mostly from pollen (Figs. 1W and 2). There are 16 PMEs in the extended plant alignment and phylogenetic tree in the clade Plant 4, clustering into four main groups (Figs. 3W and 4W). Their lengths vary from 374 (Q43043_PETIN) to 717 (Q9FSQ0_ORYSA2) amino acid residues. The *pre pro* region, in contrast to the previous clades Plant 1, Plant 2 and Plant 3, exhibits a significant heterogeneity, marginal identity and/or similarity being found only within the members belonging to the four clusters mentioned above. In the *pre pro* sequences four Cys and two Trp residues are conserved (except for the Q9FSQ0_ORYSA2, O24596_MAIZE and Q43043_PETIN). Preceding the potential mature enzyme the conserved segment RK/R/DLL/M is present twice. Multiplications of this segment in a modified form can be found in the *pre pro* region of PMEs from *Salix gilgiana* PME[62] Q9MBB6_SALGI (three times) and Q9FSQ0_ORYSA2 (11 times). The degree of sequence identity of potential mature enzymes of this clade to that of carrot PME is 42–57% (similarity 58–71%). The overall intraclade sequence identity is 16.9% and similarity 30.3%. The mature enzymes of this clade contain 1 conserved cysteine (Cys129) and 11 conserved aromatic residues (Tables 2 and 3). The members belonging to the individual groups (four branches in the trees; Figs. 2 and 4W) can be discriminate from each other by characteristically conserved residues. The group of three *Arabidopsis* and two *Brassica* PMEs possess strictly conserved two cysteines (Ala29 and Tyr37), one tryptophane (Asp100) and one histidine (Leu258). The group including the Q9MBB6_SALGI, Q9FSQ0_ORYSA2 and O24596_MAIZE has one cysteine (Ala98), whereas the cluster of three other *Arabidopsis* PMEs shares one histidine (Asn58) and one cysteine (Asn183), the latter is strictly conserved also in the group of *Petunia* and three *Medicago* PMEs.

3.2.5. The clade Plant 1a. This clade contains six members separated in two main branches (Figs. 1W and 2). However, after taking into account further PME sequences, especially those from *A. thaliana* genome,[29] in the extended plant alignment and tree, this clade has become the largest one with 24 members (Figs. 3W and 4W). There were included in this clade also two fragment PMEs (Q96548_CARPA and Q9M5J0_PHAAU1) as well as the putative *Oryza* PME (Q9S767_ORYSA1), which was originally placed into the clade Plant 2a (Fig. 2). The entire clade Plant 1a is thus divided into six main branches (Fig. 4W). The length of 22 members (without fragment PMEs from *Carica papaya* and *Phaseolus aureus*) varies from 510 to 624 residues. Similar to clade Plant 1, in the *pro* sequence of 12 members there is a conserved segment PxWxxxxRK/RLLQ/E/D preceding the potential mature enzyme. The enzymes Q9LYT5_ARATH35 and Q84WQ3_ARATH16 contain in addition a segment RKLL preceding the long conserved segment. The remaining 10 members (two *Oryza* and eight *Arabidopsis* PMEs) have only a dipeptide of two basic residues, or even a single basic residue (Lys or Arg). The *pre pro* region contains 1–6 pgs, two conserved cysteine-containing dipeptides (DC and TC). The rest of sequences are, however, in contrast to the clade Plant 1, heterogeneous. In the clade Plant 1a also Q94FS6_LINUS1 is clustered together with Q9SKX2_ARATH16 and Q9LYT5_ARATH35 PMEs (not presented in Figs. 1W and 2). The potential mature enzymes exhibit the 51–55% sequence identity and 69–78% similarity to carrot PME. The mutual sequence similarities are quite low (14.9% for the identity and 29.9% for similarity). Nine members of this clade do not possess any pgs, while the remaining 15 members contain 1–10 pgs (Table 1W). Sixteen aromatic residues are conserved in this clade (Table 3).

3.2.6. The clade Plant 2a. Eight members of this clade (Fig. 1W) grouped in the phylogenetic tree (Fig. 2) into three clusters positioned on both sides of the clade Plant 2. Two members were added in the extended plant alignment (Fig. 3W): Ref_43_FICAW (*F. awkeotsang*) and Q9LPX8_ARATH4, while, based on the extended plant alignment, the Q9S767_ORYSA1 PME was re-clustered in the tree into the clade Plant 1a (Fig. 4W). The lengths of these PMEs range between 545 and 621 residues. In the *pre pro* sequences of seven members of this clade, a few residues are conserved: one cysteine, one leucine, as well as one dipeptide (TC) and one tripeptide (WLS). All members of this clade contain a conserved segment PxWxxxxDRR/K/ELL preceding the mature enzyme; Q9M3B0_ARATH34 and Q9LGX7_ORYSA4 PMEs have in addition a segment RR/KLL. The mature enzyme begins eight residues after dipeptide of basic nature, like in the clade Plant 2 (in *F. awkeotsang* PME[43] with the N-terminal Ile229). The PMEs of clade Plant 2a exhibit 51–67% sequence identity and 64–77% similarity to DAUCA PME. The intraclade identity and similarity is 27.3% and 45.0%, respectively. Mature enzymes contain one conserved histidine (His141) and 19 conserved aromatic residues (Table 3).

3.2.7. The clade Plant X1. This clade together with the clade Plant X2 represent atypical plant PMEs, which

both are in the phylogenetic tree (Fig. 2) positioned close to microbial clades. There are two members in the clade Plant X1 in final alignment and tree (Fig. 1W and 2). In the extended plant alignment and tree are together four members (Figs. 3W and 4W). The sequence length varies from 540 to 602 residues, except for the putative PME from *Silene pratensis* flower buds (it is a fragment with 379 residues). The *pre pro* sequences of the three members (without Q96497_SILPR) contain in general three pgs, one dipeptide (SM), one tripeptide (TER), and two Cys, one Asn and one Trp residues. In contrast to the previous plant clades, there are not any basic dipeptides preceding potential mature enzyme. The degrees of identity and similarity of mature enzymes to carrot PME are between 43–52% and 58–64%, respectively. The internal identity and similarity within the clade is 39.4% and 59.6%, respectively. In mature enzymes of this clade one cysteine and two histidines (Ala12, and Phe219 and Ser289, respectively) are conserved strictly. Furthermore there are three additional conserved cysteines (Cys150, Ser204 and Ser238), 24 conserved aromatic residues (Table 3), and one pgs (Gly207). A specific feature of this clade is the L117F substitution in the second segment characteristic for PMEs (Figs. 1 and 3W).

3.2.8. The clade Plant X2. This five-member clade (Fig. 1W) covers in the extended plant version (Fig. 3W) 17 putative PMEs, exclusively all being taken only from the *A. thaliana* and *O. sativa* genomes. The lengths of these PMEs vary from 294 to 407 residues. The *pre pro* sequence is rather short (20–80 residues), however, five *Arabidopsis* members of one of the clade Plant X2 branches contain dibasic motif RK preceding the potential mature enzyme. The remaining 10 members do not possess the motif, while the two members (Q9LVQ0_ARATH31 and Q9RYM8_ORYSA6) do not have the *pre pro* sequence at all. The degrees of identity and similarity of mature enzymes to carrot PME are between 28–37% and 43–54%, respectively. The internal identity and similarity within the clade is only 8.4% and 16.8%, respectively. Members of this clade have one strictly conserved cysteine residue (Asn268), whereas two more cysteines (Cys150 and Ser204) are conserved (Table 2). The residue Gln145, conserved in the other plant and all fungal PMEs, is in the clade Plant X2 replaced by strictly conserved His residue, except for the exceptional member Q9LVQ0_ARATH31 that contains the in this place glutamine and has leucine residue instead of conserved aspartate (Val140). The members of the clade Plant X2 contain nine conserved aromatic residues (Table 3). With regard to particular four branches of this clade, several strictly conserved residues can be found (corresponding to carrot PME: Alal73, Gln78, Tyr139, Lys203, Tyr230, Ala274, Arg36 and Ser74; Figs. 3W and 4W).

3.2.9. The clade Plant X3. It contains three *Arabidopsis* putative PMEs (Q9LPX7_ARATH5, NP_189437_ARATH30 and O81516_ARATH42), which clustered separately only in the *Arabidopsis* tree (Fig. 6W), indicating thus a new plant clade. The length of these members varies from 497 to 551 residues and the *pre pro* sequence beside many identical segments contain a long conserved segment PSWLxHVDKKDLL/Y preceding the potential mature enzyme. The sequence identity and similarity of mature enzymes to carrot PME is 47–50% and 61–64%, respectively. These three PMEs exhibit high degree of intraclade sequence identity (84.2%) and similarity (88.3%). As conserved they contain two histidines (Asp32 and Tyr262), six cysteines (Thr102, Cys129, Cys150, Phe160 Cys170 and Ser204) and three pgs. Of the 30 conserved aromatic residues (Table 3) the phenylalanine (Asp169) is strictly conserved. The following substitutions were found in the three segments characteristic for PMEs (Figs. 1 and 5W): (i) Q113P in the second segment QAVAL; (ii) T137A in the third segment QDTL; and (iii) F160C in the fourth segment DFIFG.

3.2.10. *A. thaliana* PMEs. A separate set of available PME sequences from *A. thaliana* was collected as a complement to the plant alignment and tree (Figs. 3W and 4W). The *Arabidopsis* tree (Fig. 6W) revealed a few differences with respect to both the tree shown in Figure 2 and the plant tree (Fig. 4W) that concern the clustering of three members of the clades Plant 1a and 2a: Q9SKX2_ARATH16, Q9LYT5_ARATH35 and Q9LPX8_ARATH4. While the members Q9SKX2_ARATH16 and Q9LYT5_ARATH35 were clustered in plant tree in the clade Plant 1a (Fig. 4W), in the *Arabidopsis* tree these were grouped into the clade Plant 2a (Fig. 6W). With regard to the member Q9LPX8_ARATH4 placed in plant tree into the clade Plant 2a (Fig. 4W), this one was found in the clade Plant X3 in the *Arabidopsis* tree (Fig. 6W). These observations indicate the intermediate character of both clades Plant 1a and Plant 2a, which was possible to see also in the case of different clustering of Q9FSQ0_ORYSA1 PME found in the clade Plant 2a (Fig. 2) or Plant 1a (Fig. 4W).

3.2.11. The clade Fungi. This clade contains, in addition to six fungal PMEs, one plant PME (*V. riparia*) possessing all sequence features characteristic of fungal enzymes. They form three main branches in the phylogenetic tree (Fig. 2) respecting, in fact, taxonomy: the two groups of *Aspergillus* and *Botryotinia* PMEs, and the PME from *Cochliobolus carbonum* grouped with the plant enzyme from *V. riparia*. The lengths of these sequences vary between 331 and 348 residues, the putative start of the mature protein being not preceded by a typical von Heijne signal peptidase cleavage site.[21,33] The sequence identity and similarity of their mature enzymes to carrot PME is 28–34% and 44–47%, respec-

tively. The sequences of this clade exhibit overall mutual identity 30.1% and similarity 43.6%. All members possess 2–5 pgs. Nine strictly conserved residues were found: three arginins (Ser3, Ala90, Trp281) and one glutamine (Tyr37), aspartate (Phe84), serine (Arg118), leucine (Tyr139), tryptophane (Val166) and tyrosine (Thr312). Several sequence characteristics were found that discriminate the members of the individual groups of this clade: (i) two cysteines (Val10 and Val38) present in Q9SW71_VITRI and Q9Y881_COCCA PMEs; (ii) two histidines (seventh gap succeeding the Ala73 and Ala 163) and two cysteines (Ala107 and Lys11) present in all *Aspergillus* PMEs; and (iii) two tyrosines (Asp64 and seventh gap succeeding the Ala73), histidine (Ala110) and tryptophane (Gly198) in both *Botrytis* PMEs. In the second characteristic segment QAVAL the Val115 is replaced by Leu in all members of this clade.

3.2.12. The clade Bacteria. This 10-member clade is the most heterogenous one of all. The ORF lengths range from 321 (Q97DU8_CLOAB) to 433 (Q47474_ERWCH1) residues. The length of signal peptide (*pre pro* sequence) is with 18–80 residues similar to that of fungal PMEs. Similarly it does not contain the dibasic-dipeptide motif, present in most plant PMEs. The N-terminus was identified only in *E. chrysanthemi* PmeA[22,23] with cleavage site between residues Ala24-Ala25, the same as in *Aspergillus* PMEs.[21,63,64] There are no pgs in the *pre pro* sequence region of the members of this clade. However, the mature enzymes contain 1–13 pgs, except for *Streptomyces coelicolor* PME, which does not possess any pgs. The 21–36% sequence identity and 35–50% similarity of mature enzymes to carrot PME were found. The above-mentioned heterogeneity of this clade is supported also by the values of the mutual sequence identity (3.1%) and similarity (8.6%) that are the lowest values of all PME clades. Bacterial PMEs were placed at two main branches in the phylogenetic tree, the most different Q97DU8_CLOAB PME being positioned separately (Fig. 2). The first group is formed by pair of P07863_ERWCH2 and Q8ZIR5_YERPE as well as the one of P58601_RALSO1 and P24791_RALSO2 with closely related Q8P8H6_XANCA2 plus the Q47474_ERWCH1 PME positioning separately. The second group contains three PMEs (Q9AN15_BRAJA, Q93RU7_STRCO and Q8PE60_XANCA1) with long branches, although the last two PMEs exhibit a closer relatedness. Identical residues or segments were found only in the members belonging to individual groups (or subgroups) of this clade (without q47474_ERWCH1). Thus for the P07863_ERWCH2 and Q8ZIR5_YERPE, there are strictly conserved two tryptophanes (14th gap succeeding the Ala73 and in 6th gap succeeding the Leu319) and a histidine (Lys228), and conserved two cysteines (Cys150 and Cys170), histidine (Thr260) and two tryptophanes (Gln267 and Trp280). Extended and modified version of the second PME conserved segment QAVAL (Fig. 1W) is highly characteristic of the P58601_RALSO1, P24791_RALSO2 (*Pseudomonas solanacearum*)[65] and Q8PE60_XANCA2 (ADNNQSAVALAV_119), as well as four strictly conserved cysteines (Asp51, Ser74, Gly277 and Trp281) and a conserved tryptophane (Gln267), which is present also in P07863_ERWCH2 and Q8ZIR5_YERPE. The second main group of this clade (Q9AN15_BRAJA, Q93RU7_STRCO and Q8PE60_XANCA1, also including Q97DU8_CLOAB) exhibits conserved cysteine (Cys150) and tryptophane (Trp302). Q47474_ERWCH1 (*E. chrysanthemi* PemB[66]) is a specific member of this clade: it is an outer membrane lipoprotein PME with a blocked N-terminus and with higher activity on methylated oligogalacturonides than on pectin. Its position in the tree in the frame of the clade Bacteria (Fig. 2) reflects the sequence differences distinguishing it from the rest of the members of this clade.

3.3. Glycosylation of PMEs

In order to complete the sequence and evolutionary picture of PMEs, that is the family CE-8, it could be convenient to have a look at the potential glycosylation sites (pgs). Eucaryotic enzymes and proteins are prone to N- and O-linked glycosylation occurring by attachment of the glycans to the protein via the nitrogen of asparagine (N-linked) or the oxygen of serine and threonine (O-linked). The pgs NxT or NxS are observed in the *pre pro* sequence and mature enzyme of most PMEs.

It has already been suggested that glycosylation of PMEs seems to affect enzyme properties (e.g., thermostability and activity to pectin) of kiwi[67] and Valencia orange.[68] The analysis of *F. awkeotsang* N-linked glycoprotein PME revealed its monosaccharide composition (mannose, galactose and N-acetylglucosamine).[69] A new putative thermostable PME protein of 36 kDa (CsPME4) was prepared from heat treated citrus pulp (*C. sinensis*);[70] the *pre pro* sequence contains four pgs, while the mature enzyme is without any pgs. However, it is difficult to ascribe the thermostability to the presence of some degree of glycosylation.

By comparing the number of pgs in *pre pro* sequences and mature enzymes of individual plant PMEs, 16% of *pre pro* sequences (most of them were in the clades Plant X2 and Plant X3) and more than 30% of the mature enzymes were found without any pgs.

All mature enzymes of six fungal PMEs of this study possess 1–6 pgs (Table 1W). Site-directed mutagenesis of three pgs of *A. niger* PME (N95Q, N283Q and N302Q; P17872_ASPTU numbering) suggested that glycosylation does not play a major role in its activity and protease resistance.[71]

With regard to 10 bacterial PMEs studied here, 1–14 pgs were found in the mature enzymes only, except for the Q93RU7_STRCO (*S. coelicolor* PME). This bacterial PME belongs together with the three plant PMEs: O24298_PEA, Q9LPX8_ARATH4 and P41510_BRANA, to four PMEs without pgs in the ORF in the set of PMEs in this study.

3.4. Conclusion

Using the set of more than 100 available primary structures (with both experimentally confirmed and putative proteins), the present study showed that pectin methylesterases contain six strictly conserved residues (Gly44, Gly154, Asp157, Gly161, Arg225 and Trp227), three of them (Asp, Arg and Trp) being involved in the active site. The revealed sequence differences support the view that the conclusions from 3D structures of *E. chrysanthemi* and carrot PMEs could be valid only for particular clades or specific groups of PMEs. Significant differences between plant, fungal and bacterial PMEs were found, especially in presence of histidines, cysteines and aromatic residues conserved in particular clades. This could enable one to find the eventual targets suitable for future site-directed mutagenesis.

Acknowledgements

This work was supported by the grants VEGA 2/3160/24 and 2/2057/24.

References

1. Benen, J. A. E.; van Alebeek, G.-J. W. M.; Voragen, A. G. J.; Visser, J.. In *Handbook of Food Enzymology*; Whitaker, J. R., Voragen, A. G. J., Wong, D. W. S., Eds.; Marcel Dekker: New York, 2003; pp 847–856.
2. Coutinho, P. M.; Henrissat, B. *Carbohydrate-Active Enzymes server*; 1999; http://afmb.cnrs-mrs.fr/~cazy/CAZY/ CE_8.html.
3. Schols, H. A.; Voragen, A. G. J. In *Pectin and Pectinases*; Visser, J., Voragen, A. G. J., Eds.; Elsevier Science B.V.: Amsterdam, 1996; pp 3–19.
4. Albersheim, P.; Darvill, A. G.; O'Neill, M. A.; Schols, H. A.; Voragen, A. G. J. In *Pectin and Pectinases*; Visser, J., Ed.; Elsevier Science B.V.: Amsterdam, 1996; pp 47–55.
5. Van der Vlugt-Bergmans, C. J. B.; Meeuwsen, P. J. A.; Voragen, A. G. J.; van Ooyen, A. J. J. *Appl. Environ. Microbiol.* **2000**, *66*, 36–41.
6. Pérez, S.; Mazeau, K.; Hervé du Penhoat, C. *Plant Physiol. Biochem.* **2000**, *38*, 37–55.
7. Catoire, L.; Pierron, M.; Morvan, C.; du Penhoat, C. H.; Goldberg, R. *J. Biol. Chem.* **1998**, *273*, 33150–33156.
8. Micheli, F. *Trends Plant Sci.* **2001**, *6*, 414–419.
9. Willats, W. G.; Orfila, C.; Limberg, G.; Buchholt, H. C.; van Alebeek, G. J.; Voragen, A. G. J.; Marcus, S. E.; Christensen, J. P.; Mikkelsen, J. D.; Murray, B. S.; Knox, J. P. *J. Biol. Chem.* **2001**, *276*, 19404–19413.
10. Hall, L. N.; Bird, C. R.; Picton, S.; Tucker, G. A.; Seymour, G. B.; Grierson, D. *Plant Mol. Biol.* **1994**, *25*, 313–318.
11. Goldberg, R.; Pierron, M.; Bordenave, M.; Breton, C.; Morvan, C.; du Penhoat, C. H. *J. Biol. Chem.* **2001**, *276*, 8841–8847.
12. Denés, J. M.; Baron, A.; Renard, C. M. G. C.; Pean, C.; Drilleau, J. F. *Carbohydr. Res.* **2000**, *327*, 385–393.
13. Limberg, G.; Korner, R.; Buchholt, H. C.; Christensen, T. M.; Roepstorff, P.; Mikkelsen, J. D. *Carbohydr. Res.* **2000**, *327*, 293–307.
14. Gaffe, J.; Tieman, D. M.; Handa, A. K. *Plant Physiol.* **1994**, *105*, 199–203.
15. Gaffe, J.; Tiznado, M. E.; Handa, A. K. *Plant Physiol.* **1997**, *114*, 1547–1556.
16. Micheli, F.; Holliger, C.; Goldberg, R.; Richard, L. *Gene* **1998**, *220*, 13–20.
17. Harriman, R. W.; Tieman, D. M.; Handa, A. K. *Plant Physiol.* **1991**, *97*, 80–87.
18. Albani, D.; Altosaar, I.; Arnison, P. G.; Fabijansky, S. F. *Plant Mol. Biol.* **1997**, *16*, 501–513.
19. Mu, J. H.; Stains, J. P.; Kao, T. H. *Plant Mol. Biol.* **1994**, *25*, 539–544.
20. Bordenave, M.; Breton, C.; Goldberg, R.; Huet, J. C.; Perez, S.; Pernollet, J. C. *Plant Mol. Biol.* **1996**, *31*, 1039–1049.
21. Christgau, S.; Kofod, L. V.; Halkier, T.; Andersen, L. N.; Hockauf, M.; Dorreich, K.; Dalboge, H.; Kauppinen, S. *Biochem. J.* **1996**, *319*, 705–712.
22. Laurent, F.; Kotoujansky, A.; Labesse, G.; Bertheau, Y. *Gene* **1993**, *131*, 17–25.
23. Jenkins, J.; Mayans, O.; Smith, D.; Worboys, K.; Pickersgill, R. W. *J. Mol. Biol.* **2001**, *305*, 951–960.
24. Johansson, K.; El-Ahmad, M.; Friemann, R.; Jörnvall, H.; Markovič, O.; Eklund, H. *FEBS Lett.* **2002**, *514*, 243–249.
25. Yoder, M. D.; Keen, N. T.; Jurnak, F. *Science* **1993**, *260*, 1503–1507.
26. Jenkins, J.; Pickersgill, R. *Prog. Biophys. Mol. Biol.* **2001**, *77*, 111–175.
27. Newman, T.; de Bruijn, F. J.; Green, P.; Keegstra, K.; Kende, H.; McIntosh, L.; Ohlrogge, J.; Raikhel, N.; Somerville, S.; Thomashow, M.; Retzel, E.; Somerville, C. *Plant Physiol.* **1994**, *106*, 1241–1255.
28. Delseny, M.; Cooke, R.; Raynal, M.; Grellet, F. *FEBS Lett.* **1997**, *403*, 221–224.
29. The Arabidopsis Genome Initiative *Nature* **2000**, *408*, 796–815.
30. Richard, L.; Qin, L. X.; Gadal, P.; Goldberg, R. *FEBS Lett.* **1994**, *355*, 135–139.
31. Richard, L.; Qin, L. X.; Goldberg, R. *Gene* **1996**, *170*, 207–211.
32. Sasaki, T.; Matsumoto, T.; Yamamoto, K., et al. *Nature* **2002**, *420*, 312–316.
33. Nielsen, H.; Brunak, S.; von Heijne, G. *Protein Eng.* **1999**, *12*, 3–9.
34. Calmels, T. P. G.; Martin, F.; Durand, H.; Tiraby, G. *J. Biotechnol.* **1991**, *17*, 51–66.
35. Baker, D.; Shiau, A. K.; Agard, D. A. *Curr. Opin. Cell Biol.* **1993**, *5*, 966–970.
36. Nairn, C. J.; Lewandowski, D. J.; Burns, J. K. *Physiol. Plantarum* **1998**, *102*, 226–235.
37. Markovič, O.; Jörnvall, H. *Protein Sci.* **1992**, *1*, 1288–1292.

38. Ebbelaar, M. E.; Tucker, G. A.; Laats, M. M.; van Dijk, C.; Stolle-Smits, T.; Recourt, K. *Plant Mol. Biol.* **1996**, *31*, 1141–1151.
39. Markovič, O.; Cederlund, E.; Griffiths, W. J.; Lipka, T.; Jörnvall, H. *Cell Mol. Life Sci.* **2002**, *59*, 513–518.
40. Apweiler, R.; Bairoch, A.; Wu, C. H.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M. J.; Natale, D. A.; O'Donovan, C.; Redaschi, N.; Yeh, L. S. *Nucleic Acids Res.* **2004**, *32*, D115–D119.
41. Benson, D. A.; Karsch-Mizrachi, I.; Lipman, D. J.; Ostell, J.; Rapp, B. A.; Wheeler, D. L. *Nucleic Acids Res.* **2002**, *30*, 17–20.
42. Christensen, T. M. I. E.; Nielsen, J. E.; Kreiberg, J. D.; Rasmussen, P.; Mikkelsen, J. D. *Planta* **1998**, *206*, 493–503.
43. Ding, J. L. C.; Lee, T. T. T.; Wang, M. M. C.; Tai, S. S. K.; Tzen, J. T. C. *J. Agric. Food Chem.* **2000**, *48*, 3052–3057.
44. Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucleic Acids Res.* **1994**, *22*, 4673–4680.
45. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215*, 403–410.
46. Saitou, N.; Nei, M. *Mol. Biol. Evol.* **1987**, *4*, 406–425.
47. Felsenstein, J. *Evolution* **1985**, *39*, 783–791.
48. Page, R. D. *Comput. Appl. Biosci.* **1996**, *12*, 357–358.
49. Markovič, O.; Patočka, J. *Experientia* **1977**, *33*, 711–712.
50. Duwe, B.; Khanh, N. Q. *Biotechnol. Lett.* **1996**, *18*, 621–626.
51. Markovič, O.; Stovíčková, J.; Jörnvall, H. *J. Protein Chem.* **1996**, *15*, 127–130.
52. Eklund, H. *Personal communication*; 2002.
53. Stratilová, E.; Markovič, O.; Dzúrová, M.; Malovíková, A.; Capek, P.; Omelková, J. *Biologia, Bratislava* **1998**, *53*, 731–738.
54. Dorokhov, Y.; Mäkinen, K.; Frolova, O. Y.; Merits, A.; Saarinen, J.; Kalkkinen, N.; Atabekov, J. G.; Saarma, M. *FEBS Lett.* **1999**, *461*, 223–228.
55. Micheli, F.; Sundberg, B.; Goldberg, R.; Richard, L. *Plant Physiol.* **2000**, *124*, 191–199.
56. Roger, D.; Lacoux, J.; Lamblin, F.; Gaillet, D.; Dauchel, H.; Klein, D.; Balangé, A. P.; David, A.; Lainé, E. *Plant Sci.* **2001**, *160*, 713–721.
57. Tucker, G. A.; Zhang, J. In *Pectin and Pectinases*; Visser, J., Voragen, A. G. J., Eds.; Elsevier Science B.V.: Amsterdam, 1996; pp 347–354.
58. Markovič, O.; Sajgó, M. *Acta Biochim. Biophys. Acad. Sci. Hung.* **1977**, *12*, 45–48.
59. Markovič, O.; Jörnvall, H. *Eur. J. Biochem.* **1986**, *158*, 455–462.
60. Ray, J.; Knapp, J.; Grierson, D.; Schuch, W. *Eur. J. Biochem.* **1988**, *174*, 119–124.
61. Turner, L. A.; Harriman, R. W.; Handa, A. K. *Plant Physiol.* **1996**, *112*, 1398.
62. Futamura, N.; Mori, H.; Kouchi, H.; Shinohara, K. *Plant Cell Physiol.* **2000**, *41*, 16–26.
63. Kitamoto, N.; Okada, H.; Yoshino, S.; Ohmiya, K.; Tsukagushi, N. *Biosci. Biotechnol. Biochem.* **1999**, *63*, 120–124.
64. Khanh, N. Q.; Albrecht, H.; Rutkowski, E.; Löffler, F.; Gottschalk, M.; Jany, K.-D. *Nucleic Acids Res.* **1999**, *18*, 4262.
65. Spök, A.; Stubenrauch, G.; Schörgendorfer, K.; Schwab, H. *J. Gen. Microbiol.* **1991**, *137*, 131–140.
66. Shevchik, V. E.; Condemine, G.; Hugovieux-Cotte-Pattat, N.; Robert-Boudouy, J. *Mol. Microbiol.* **1996**, *19*, 455–466.
67. Giovane, A.; Quagliuolo, L.; Castaldo, D.; Servillo, L.; Balestrieri, C. *Phytochemistry* **1990**, *29*, 2821–2823.
68. Randall, G. D.; Karel, G. *J. Agric. Food Chem.* **1996**, *44*, 458–462.
69. Ding, J. L. C.; Hsu, J. S. F.; Wang, M. M. C.; Tzen, J. T. C. *J. Agric. Food Chem.* **2002**, *50*, 2920–2925.
70. Arias, C. R.; Burns, J. K. *J. Agric. Food Chem.* **2002**, *50*, 3465–3472.
71. Warren, M. E.; Kester, H.; Benen, J.; Colangelo, J.; Visser, J.; Bergmann, C.; Orlando, R. *Carbohydr. Res.* **2002**, *337*, 803–812.