Scuola Internazionale Superiore di Studi Avanzati – Trieste

Structural and Functional Studies of a Polygalacturonase from the Phytopathogen *Burkholderia cepacia*

Thesis submitted for the degree of
Doctor Philisophiae

PhD in *Structural and Functional Genomics*

Candidate:
Claudia Massa

Supervisors:
Dr. Doriano Lamba
Dr. Cristiana Campa

December, 2006

SISSA - via Beirut 2-4 - 34014 Trieste ITALY

Preface

This thesis is submitted in partial fulfillment of the requirements for the academic title of Doctor Philosophiae at the International School for Advanced Studies (Trieste, Italy). The work was carried out under the supervision of Dr. Doriano Lamba and Dr. Cristiana Campa within the PhD program in *Structural and Functional Genomics*, at the Structural Biology Laboratory, Elettra, Synchrotron Light Facility (Trieste, Italy), at the Bacteriology Group Laboratory, International Centre for Genetic Engineering and Biotechnology (Trieste, Italy) and at the Laboratory of Bracco Imaging S.p.a. (AREA Science Park, Trieste, Italy).

Table of contents

List of abbreviations

4-ABN: 4-Aminobenzonitrile

Acc. Num.: Protein Accession Number to Swiss-Prot Database

β–ME: β–mercaptoethanol

C-terminus: Carboxy-terminus

CD: Circular Dichroism

CE: Capillary Electrophoresis

DLS: Dynamic Light Scattering

DP: Degree of Polymerization

DTT: Dithiothreitol

EndoPG: Endopolygalacturonase

EDTA: Ethylendiaminetetraacetic acid tetrasodium salt

ES complex: Enzyme-Substrate complex

ESI-MS: Electrospray Mass Spectrometry

ESI-TOF-MS: Electrospray Mass Spectrometry with a Time-Of-Flight analyzer

Gdn-HCl: Guanidine Hydrochloride

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid

HGA: Homogalacturonan

HPAEC-PAD: High-Performance Anion-Exchange chromatography with Pulsed Amperometric Detection

HTP: High-Throughput

IBs: Inclusion Bodies

IEF: Isoelectrofocusing

IPTG: 1-thio-β-D-galactopyranoside

LB: Luria-Bertani broth

M9CA: Minimal medium supplemented with casamino acid

MALDI-TOF-MS: Matrix-Assisted Laser Desorption/Ionisation Mass Spectrometry with a Time-Of-Flight analyser

MEKC: Micellar Electrokinetic Capillary Chromatography

OGAs: Oligogalacturonides

N-terminus: Amino-terminus

PehA: Pectic Enzyme Hydrolase

per-PehA: PehA purified from the periplasmic extract of *E. coli* host cells

PGA: Polygalacturonic Acid

pI: Isoelectric Point

ref-PehA: PehA, after the refolding procedure and the TEV cleavage

RG-I: Rhamnogalacturonan I

RG-II: Rhamnogalacturonan II

SDS-PAGE: Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis

TEV protease: Tobacco Etch Virus protease

Tris-HCl: 2-Amino-2-(hydroxymethyl)-1,3-propanediol hydrochloride

TrxA-PehA: PehA as fusion protein with thioredoxin A

wt-PehA: PehA purified from the supernatant of *B. cepacia* culture

The molecular scenario

Endopolygalacturonases (endoPGs) are a class of enzymes that are involved in the hydrolysis of the α−1,4 glycosidic bond within non-esterified regions of pectins, which are one of the major component of the primary plant cell walls. Plant endoPGs are expressed during periods of cell wall turnover and expansion, since they are involved in the remodeling and developmental processes. Fungal and bacterial endoPGs are instead secreted by pathogenic species and their catalytic activities are responsible for a range of serious plant diseases. Therefore, an investigation of the mode of action of these enzymes that act as virulence factors, supported by structural and functional studies, is of crucial importance for the comprehension, at the molecular level, of the pathogenic mechanisms.

Chapter 1. Introduction

1.1 The plant cell walls

1.1.1 Physiological role

The cell walls are important structures of the plant cells, that are involved in a variety of physiological processes. Beside the protection of the intracellular contents against the extracellular environment, these structures determine the cell shape and confer mechanical strength and rigidity to the plant (1). They provide a porous medium for the circulation and distribution of water, minerals, and other nutrients: moreover, they contain specialized molecules that regulate plant growth and defence mechanisms (2). The walls that surround growing plant cells are called "primary cell walls", and they are typically extensible in response to cell growth. Many plant cells have also a 'secondary cell wall', that is built up once a plant cell has achieved its final shape and size. The secondary cell wall is found between the cell membrane and the primary cell wall. A specialized region associated with the cell walls of plants is the middle lamella (Figure 1.1). It is shared by neighbouring cells and joints them firmly together, allowing cells to communicate and exchange nutrients. These structures are characterized by a specific composition of high molecular weight polysaccharides that define their different physiological properties (3).
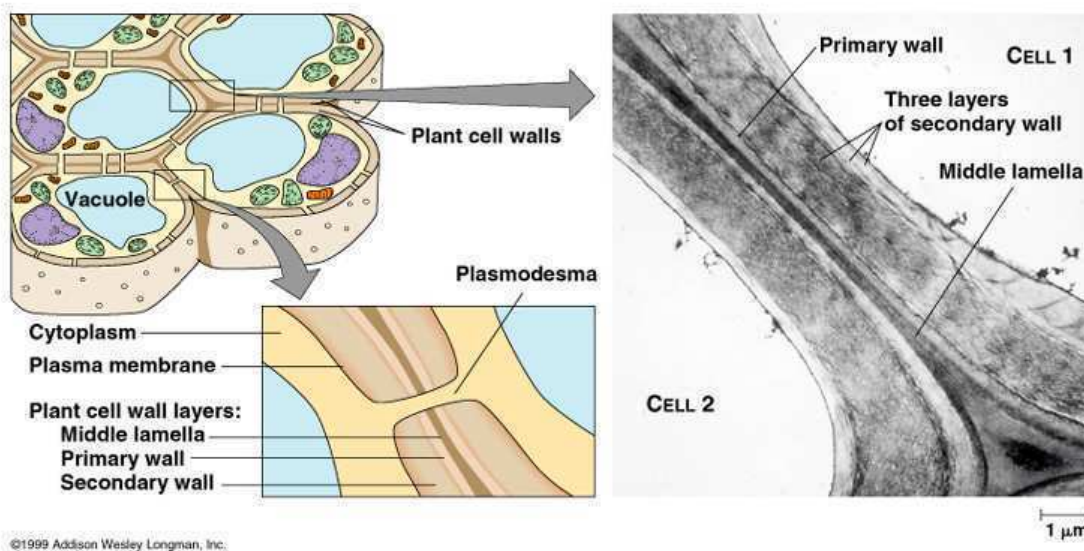
Figure 1.1. A picture of the plant cell wall. The picture is taken from url: http://fig.cox.miami.edu/~cmallery/150/cells/plasmodesmata.jpg.

1.1.2 Chemical composition

The main chemical components of the primary plant cell walls are cellulose microfibrils. These microfibrils are composed of 30 to 36 chains of β-1,4-linked glucose, that are hydrogen bonded to form an insoluble and inelastic crystalline scaffold. Other abundant components are hemicelluloses. Despite the name, they bear no chemical resemblance to cellulose; as hemicelluloses are referred to many type of heteropolysaccharides, like xylans, arabino- and glucurono-xylans, arabinogalactans, gluco- and galacto-mannans and xyloglucans. Together with cellulose microfibrils, they form a cohesive network through non-covalent lateral associations and physical entanglements. The cellulose-hemicellulose network is embedded in a soluble matrix of polysaccharides, glycoproteins, proteoglycans, low-molecular weight compounds and ions (3). Within this matrix, pectic substances represent the most abundant component, accounting for about one third of the all primary cell wall macromolecules (4) (Figure 1.2).
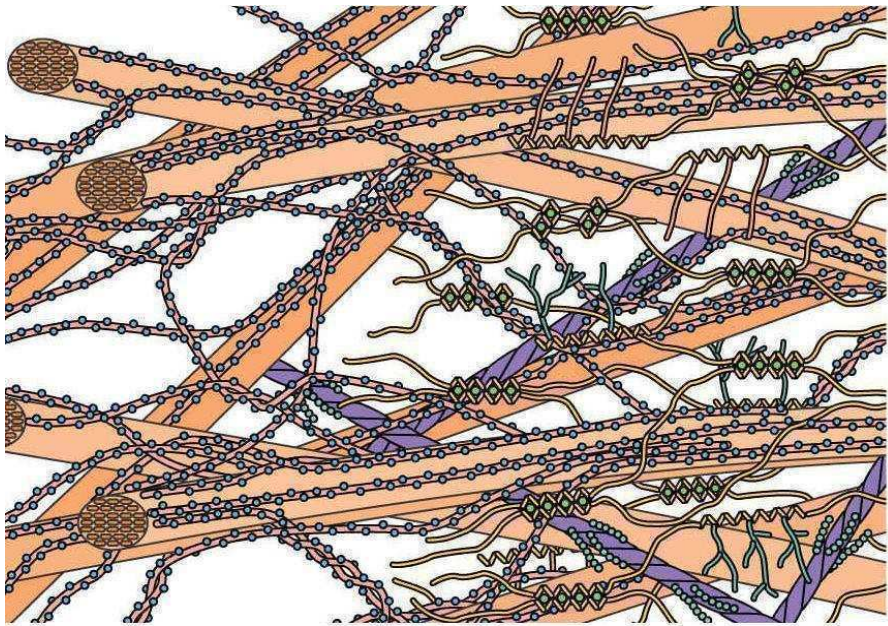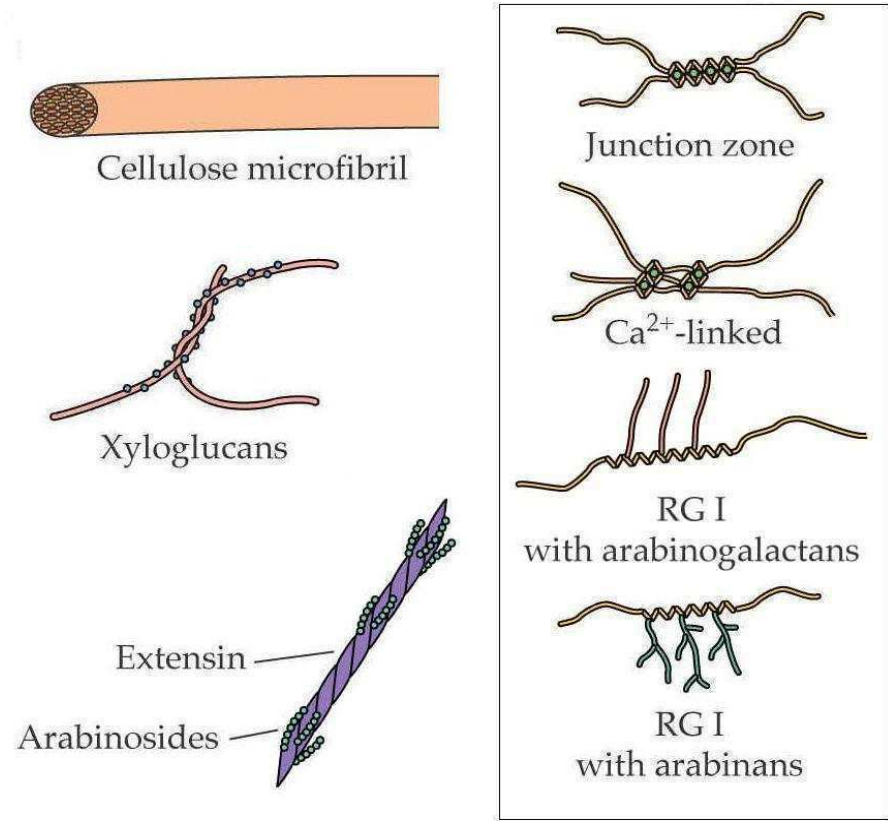
Figure 1.2. A schematic representation of the components of plant cell walls. The image is taken from url: http://www.nrsl.umd.edu/courses/plsc400.

1.2 Insights in the pectic substances

1.2.1 Chemical structures

Pectic substances can be divided into three classes based on their different backbone configurations: homogalacturonan (HGA), rhamnogalacturonan I (RG-I), and rhamnogalacturonan II (RG-II) (4) (Figure 1.3). HGA is a linear homopolymer of 100 to 200 1,4-linked α-D-galactosyluronic acid (GalA) residues, in which some of the carboxyl groups are methyl esterified. HGA may, depending on the plant source, also be partially O-acetylated at C-3 or C-2 (5). RG-I consists of a backbone of up to 100 repeats of alternating rhamnose and galacturonic acid residues. The rhamnose residues can be substituted at C-4 with neutral and acidic oligosaccharide side chains, like α-D-arabinofuranosyl, β-D-galactopyranosyl, α-D-fucosyl, β-D-glucuronosyl, 4-*O*-methyl β-D-glucuronosyl residues. Ferulic acid and coumaric acid may also be present (6). RG-II is composed of at least seven 1,4-linked α-D-GalA residues, to which four side chains are attached. The side chains incorporate 11 different monosaccharides; among these, apiose, acetic acid, and 2-keto-3-deoxy-D-manno-octulosonic acid are only found in this
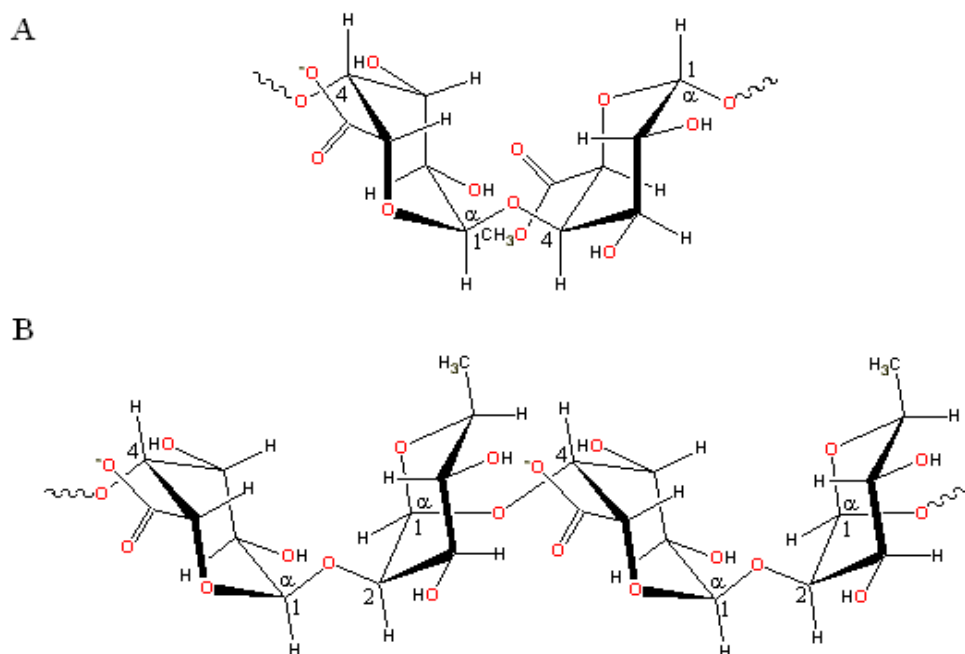
Figure 1.3. (A) Typical constituents of HGA, linear polymer made by 1,4-linked α-D-galactosyluronic acid residues. (B) Typical constituents of RG-I, made by repetition of the disaccharide (1→2)-α-L-rhamnose-(1→4)-α-GalA. The Figures are taken at the url: http://www.lsbu.ac.uk/water/hypec.html.

molecule. RG-II can dimerize by means of borate diester links through the apiosyl residues (7).

The linear regions of pectate networks composed of HGA are generally referred to as "smooth" regions, while regions composed of highly branched RGI are referred to as "hairy" regions.

1.2.2 The macromolecular organization of pectic substances in the primary cell wall

The macromolecular organization of pectic components in the primary cell wall is at present poorly understood. The limited experimental evidences suggest that these components exist in the primary cell wall as a continuous three-dimensional cross-linked network, that involves borate-ester cross-linking as well as anionic cross-linking (8-10).

1.2.3 Physico-chemical investigation

Generally, pectic molecules adopt an extended and curved conformation in solution, with a large amount of flexibility, referred to as "worm like" conformation (11). The carboxylate groups tend to expand the structure of pectins as a result of their charge. Their pK_a of about 2.9 ensures a considerable negative charge under most circumstances (12). The gelling properties of pectins depend on the degree of esterification, which is normally about 70%. Low methoxyl-pectins (< 40% esterified) can produce gels by interacting with divalent cations (mainly Ca^{2+}, but also $Mg^{2+}, Sr^{2+}, Ba^{2+}$) that bridge two adjacent chains and form the so-called "egg-box" junction zone structures (13). Two adjacent carboxylate groups cooperate together in prizing the bound water away from the ions to form the salt links that make up these junction zones, encompassing a minimum of 14-20 residues (Figure 1.4). If the methoxyl esterified content is greater than about 50%, pectins show some interaction with divalent cations but gel do not form. *In vivo*, the formation of gels is probably important for maintaining the integrity of cell walls, both within and between cell layers and across the middle lamella (3).

From a conformational and structural point of view, HGA is the best understood pectic
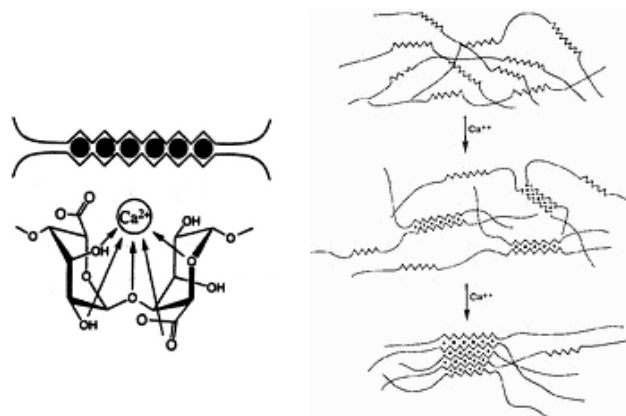
Figure 1.4. The "egg box" model. The Figure is taken from url: www.fao.org/docrep/X5822E/x5822e0i.gif.

domain. Experimental studies (X-ray fibre diffraction, circular dichroism, NMR spectroscopy) have extensively demonstrated that in the solid/gel state, HGA may occur as 2_1 and/or 3_1 helices depending on factors such as the degree of hydration and the nature of the counter ion (11, 14, 15) (Figure 1.5).
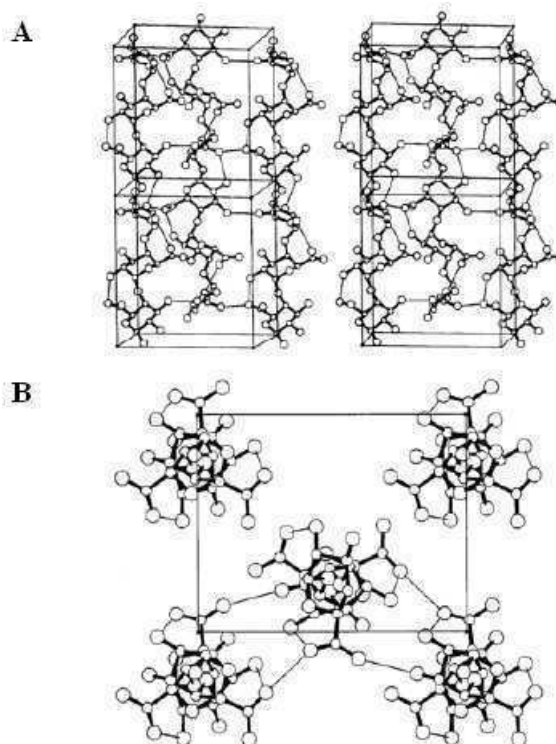
Figure 1.5. Antiparallel packing arrangement of 3-fold pectic acid helices. (A) Stereo view of two unit cells roughly normal to the *bc*-plane. The helix at the center (open bonds) is antiparallel to the two in the front (filled bonds). Intrachain hydrogen bonds (thin lines) stabilize each helix. Association of helices is through direct hydrogen bonds involving the carboxyl groups. (B) A view of the unit cell contents down the *c*-axis highlights the interactions between the three helices near the bottom. The Figure is taken from ref. (14).

1.3 Pectin degrading enzymes

1.3.1 An overview

Being pectic substances a complex substrate, several enzymatic activities are required for their complete degradation (16, 17). The pectin degrading-enzymes include exo- and endo-polygalacturonases (endoPGs), pectin and pectate lyases, rhamnogalacturonases and rhamnogalacturonan lyases. Other accessory enzymes are needed to degrade the branched side groups, like acetylesterases and pectin methylesterases (Figure 1.6). Among them, endoPGs are the most extensively studied.
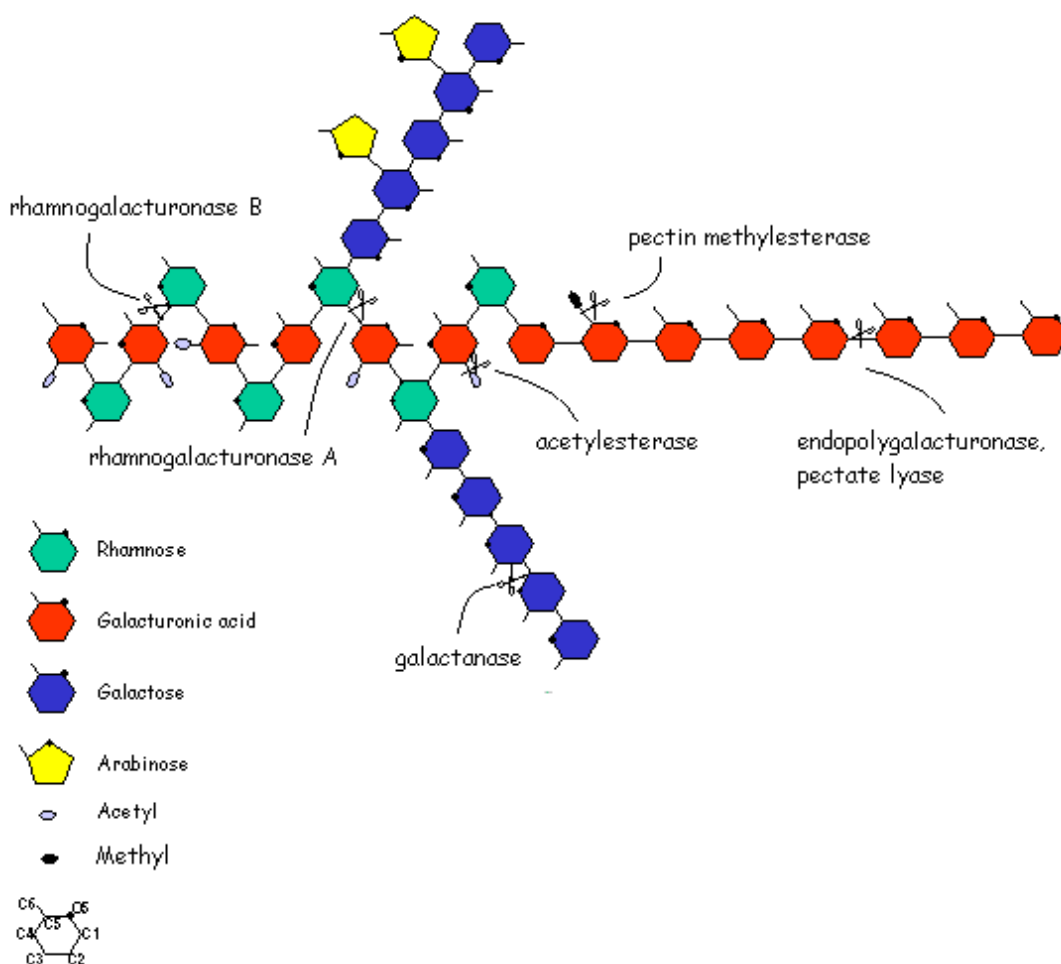
Figure 1.6. The enzymatic activities involved in the degradation of pectins.

1.3.2 Endopolygalacturonases

1.3.2.1 Classification

According to the nomenclature recommended by the Union of Biochemistry and Molecular Biology (IUBMB) (http://www.chem.qmul.ac.uk/iubmb/enzyme), endoPGs are referred to as E.C. 3.2.1.15. This classification does not provide informations about their structural, evolutionary and mechanicistic aspects. A more comprehensive way to describe enzymes able to create, modify or cleave glycosidic bonds, has been proposed by Henrissat, based on amino acid sequence similarities (18). Namely, four classes of so called "carbohydrate-active enzymes" (CAZymes) have been defined: glycoside hydrolases, glycoside transferases, polysaccharides lyases and carbohydrate-esterases (19) (Figure 1.7).
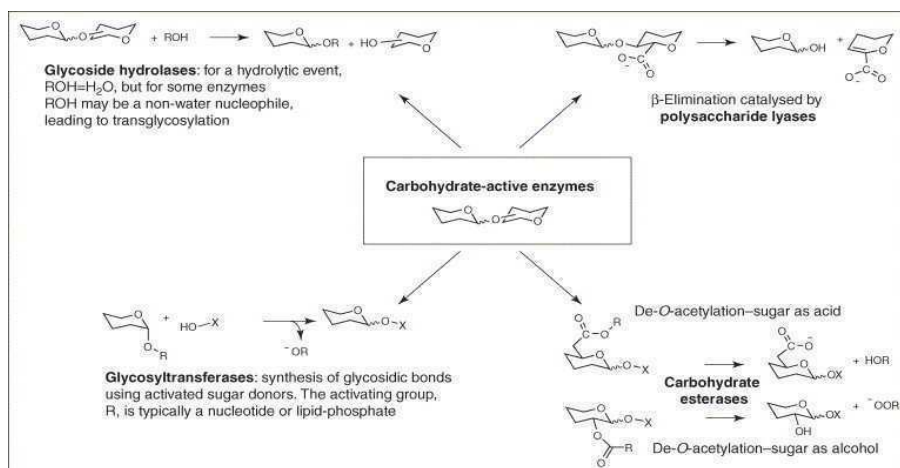
Figure 1.7. The world of carbohydrate-active enzymes. The Figure is taken from G.J. Davies *et al.*, (19).

A further classification into families has been proposed within each of the four classes, based again on amino acid sequence similarities. According to this classification, endoPGs belong to the family 28 of the glycoside hydrolases class, together with exoPGs (EC 3.2.1.67; EC 3.2.1.82), rhamnogalacturonases and endo-xylogalacturonan hydrolases. Furthermore, taking into account that the protein folds are better conserved than their amino acidic sequences, the families of glycoside hydrolases have been grouped into clans based on their three dimensional structure similarity.

Family 28 and family 49 forms the GH-N clan. Family 49 contains dextranases (EC 3.2.1.11), isopullulanases (EC 3.2.1.57), dextran 1,6-isomaltotriosidases (EC 3.2.1.95).

Enzymes belonging to those two families exhibit their catalytic function on polyanionic substrates, like pectate, pullalans and dextranes.

<u>1.3.2.2 Reaction mechanism</u>

Enzymatic hydrolysis of the glycosidic bond may occur *via* general acid/base catalysis. Two critical residues are required acting as proton donor and as nucleophile/base. The hydrolysis can proceed according to two different mechanisms, depending on the retention or inversion of the anomeric configuration of the oxygen of the scissile bond (20).

In the case of retaining enzymes one carboxyl residue acts as general acid and general base while the other one acts as nucleophile and as leaving group. The hydrolysis proceeds *via* a double displacement mechanism. In the first step, one of the carboxylic groups acts as a general acid, protonating the oxygen of the scissile bond, while the second one instead acts as a nucleophile, allowing the release of the aglycon and the formation of a glycosil-enzyme intermediate. In the second step, this intermediate is hydrolysed by a water molecule, previously activated by the other carboxylate residue, that functions as base catalyst. This second nucleophilic substitution at the anomeric carbon generates a product with the same stereochemistry as the substrate (Figure 1.8A). In the case of inverting enzymes, one residue acts as a general acid and the other as a general base, allowing the concomitant protonation of the glycosidic oxygen and the aglycon release. The aglycon release occurs at the same time due to the presence of a water molecule activated by the base residue (single displacement mechanism). This single nucleophilic substitution yields a product with opposite stereochemistry to the substrate (Figure 1.8B) (21).

Retaining and inverting enzymes have an active-site architecture apparently similar, with two essential carboxylic acids on opposite faces of the substrate-binding cavity. These catalytic groups are separated by approximately 5 Å in the retaining enzymes and approximately by 9-10 Å in the inverting enzymes. The larger separation for the inverting enzymes is ascribed to the presence of a water molecule that, as well as the substrate, must be located in between the two carboxyl groups (20).
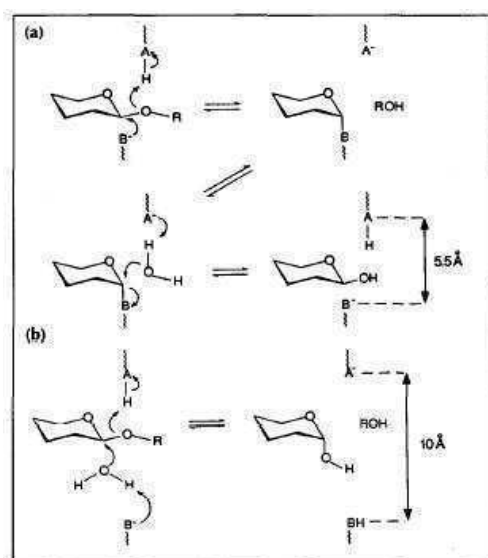
Figure 1.8. The two major mechanisms of enzymatic glycosidic bond hydrolysis. (a) The retaining mechanism. (b) The inverting mechanism. The Figure is taken from G.J. Davies *et al.*, (21).

Enzymes belonging to the family 28 of glycoside hydrolases are inverting enzymes. The sterochemistry of the catalytic action of endoPGs (22) and of rhamnogalacturonases (23) was investigated analysing the anomeric configuration of the hydrolysis products by ^{1}H-NMR. Nevertheless, as clearly showed by the three-dimensional structure determination of endoPGs (24-29), their active site architecture do not seem to be compatible with a conventional inverting mechanism, being the distance between conserved aspartate residues shorter than 10 Å. To explain how this geometry can fit with the inverting glycosidase mechanism, the comparison with the catalytic mechanism of the *Salmonella typhimurium* P22 tailspike protein (TSP) has been widely proposed (24, 25). This protein shows the same fold of the enzymes belonging to family 28, as will be discussed later in more details. The TSP functions as a viral adhesion protein and binds to the O-antigenic repeating units of *Salmonella* host lipopolysaccharide (LPS), its cellular receptor. The O-antigen is composed by a trisaccharide repeating unit formed by α-D-mannose-(1,4)-α-L-rhamnose-(1,3)-α-D-galactose. The TSP displays a receptor destroying endorhamnosidase activity by cleaving the α–(1,3)-*O*-glycosidic bond between rhamnose and galactose of the O-antigenic repeats (30, 31). The crystal structure of the C-terminal fragment of TSP, that resembles the intact protein as concerns its enzymatic activity, receptor binding properties and temperature stability, was initially solved (32). Then, crystal structures of the TPS in complex with

oligosaccharides, comprising two O-antigenic repeating units from *Salmonella typhiumurium, Salmonella enteritidis*, and *Salmonella typhi 253Ty* were determined (33). These crystal structures revealed that in the active site three acidic residues are present. Analysing the relative position between these residues and the terminal rhamnose of the O-antigen fragment from *Salmonella typhi253Ty*, steric evidences led to the conclusion that the reaction proceeds according to an inverting mechanism (34). Moreover, to explain the shorter separation between carboxylates, it has been proposed that the protonation of the anomeric carbon and the nucleophilic attack must occur at the same side with respect to the plane defined by the nucleophilic water, the anomeric carbon and the glycosidic oxygen (35-37).

1.3.2.3 Active site characterization

The analysis of the primary structure of enzymes belonging to family 28 resulted to be useful in identifying those conserved residues that are involved in the catalysis and in the substrate binding. Markovic *et al.* (38) have reported the alignment of 115 amino acid sequences of exo-, and endoPGs. The alignment of a selected group of endoPGs, whose three dimensional structures are known, is showed in Figure 1.9.

```
endoPGI__A_niger        -----------------------ASTCTFTS---ASEASESI---SSCSDVVLSS-IEVP 30

endoPG__A_aculeatus     -----------------------ATTCTFSGSNGASSASKSK---TSCSTIVLSN-VAVP 33
endoPGII__A_niger       -----------------------DSCTFTT---AAAAKAGK---AKCSTITLNN-IEVP 29
endoPG__F_moniliforme   -----------------------DPCSVTE---YSGLATAV---SSCKNIVLNG-FQVP 29
endoPG__S_purpureum     -----------------------ATCTVKS----VDDAKDI---AGCSAVTLNG-FTVP 28
endoPG__E_carotovora    SDSRTVSEPKTPSSCTTLKADSSTATSTIQKALNNCDQGKAVRLSAGSTSVFLSGPLSLP 60
                                               ..:.             : .. : *.. . :*


endoPGI__A_niger        AGETLDLSD-----AADGSTITFEGTTSFGYKEWKGP-------LIRFGGKDLT----VT 74
endoPG__A_aculeatus     SGTTLDLTK-----LNDGTHVIFSGETTFGYKEWSGP-------LISVSGSDLT----IT 77
endoPGII__A_niger       AGTTLDLTG-----LTSGTKVIFEGTTTFQYEEWAGP-------LISMSGEHIT----VT 73
endoPG__F_moniliforme   TGKQLDLSS-----LQNDSTVTFKGTTTFATTADNDFN------PIVISGSNIT----IT 74
endoPG__S_purpureum     AGNTLVLN------PDKGATVTMAGDITFAKTTLDGP-------LFTIDGTGIN----FV 71
endoPG__E_carotovora    SGVSLLIDKGVTLRAVNNAKSFENAPSSCGVVDKNGKGCDAFITAVSTTNSGIYGPGTID 120
                        :*   * :          ..:     .  .  :      .      .   .  :     .


endoPGI__A_niger        MADGAVIDGDGSRWWDSKGTNGG-KTKPKFMYIHDVE---DSTFKGINIKNTPVQAISVQ 130
endoPG__A_aculeatus     GASGHSINGDGSRWWDGEGGNGG-KTKPKFFAAHSLT---NSVISGLKIVNSPVQVFSVA 133
endoPGII__A_niger       GASGHLINCDGARWWDGKGTSG--KKKPKFFYAHGLD---SSSITGLNIKNTPLMAFSVQ 128
endoPG__F_moniliforme   GASGHVIDGNGQAYWDGKGSNSNSNQKPDHFIVVQKTTG-NSKITNLNIQNWPVHCFDIT 133
endoPG__S_purpureum     GAD-HIFDGNGALYWDGKGTNNG-THKPHPFLKIKGS----GTYKKFEVLNSPAQAISVG 125
endoPG__E_carotovora    GQGGVKLQDKKVSWWELAADAKVKKLKQNTPRLIQINKSKNFTLYNVSLINSPNFHVVFS 180
                         .   ::  .    :*:  .     . *  .      ..: * *   . .
```

```
endoPGI__A_niger        ATN--VHLNDFTIDNSDGDDN--------GGHNTDGFDISESTGVYISGATVKNQDDCIA 180
endoPG__A_aculeatus     GSDY-LTLKDITIDNSDGDDN--------GGHNTDAFDIGTSTYVTISGATVYNQDDCVA 184
endoPGII__A_niger       AND--ITFTDVTINNADGDTQ--------GGHNTDAFDVGNSVGVNIIKPWVHNQDDCLA 178
endoPG__F_moniliforme   GSSQ-LTISGLILDNRAGDKPNAKSGSLPAAHNTDGFDISSSDHVTLDNNHVYNQDDCVA 192
endoPG__S_purpureum     PTDAHLTLDGITVDDFAGDTKN-------LGHNTDGFDVSAN-NVTIQNCIVKNQDDCIA 177
endoPG__E_carotovora    DGDG-FTAWKTTIKTPS-----------TARNTDGIDPMSSKNITIAYSNIATGDDNVA 227
                            .   .        :.               .:***.:*    .  : :    :   . ** :*

endoPGI__A_niger        INSGE------SISFTGGTCSGGHGLSIGSVGGRDDNTVKNVTISDSTVSNSANGVRIKT 234
endoPG__A_aculeatus     VNSGE------NIYFSGGYCSGGHGLSIGSVGGRSDNTVKNVTFVDSTIINSDNGVRIKT 238
endoPGII__A_niger       VNSGE------NIWFTGGTCIGGHGLSIGSVGDRSNNVVKNVTIEHSTVSNSENAVRIKT 232
endoPG__F_moniliforme   VTSGT------NIVVSNMYCSGGHGLSIGSVGGKSDNVVDGVQFLSSQVVNSQNGCRIKS 246
endoPG__S_purpureum     INDGN------NIRFENNQCSGGHGISIGSIATG--KHVSNVVIKGNTVTRSMYGVRIKA 229
endoPG__E_carotovora    IKAYKGRAETRNISILHNDFGTGHGMSIGSETMG----VYNVTVDDLKMNGTTNGLRIKS 283
                        :.           .* .        ***:****     * .*.     :  :   . ***:

endoPGI__A_niger        I-YKETGDVSEITYSNIQLSGITDYGIVIEQDYENGSPTGTPSTGIPITDVTVDG--VTG 291
endoPG__A_aculeatus     N-IDTTGSVSDVTYKDITLTSIAKYGIVVQQNY--GDTSSTPTTGVPITDFVLDN--VHG 293
endoPGII__A_niger       I-SGATGSVSEITYSNIVMSGISDYGVVIQQDYEDGKPTGKPTNGVTIQDVKLES--VTG 289
endoPG__F_moniliforme   N-SGATGTINNVTYQNIALTNISTYGVDVQQDYLNGGPTGKPTNGVKISNIKFIK--VTG 303
endoPG__S_purpureum     QRTATSASVSGVTYDANTISGIAKYGVLISQSYP--DDVGNPGTGAPFSDVNFTGGATTI 287
endoPG__E_carotovora    D-KSAAGVVNGVRYSNVVMKNVAKP-IVIDTVYEKKEGSNVP----DWSDITFKD----- 332
                             :. :. : *.   :..::   : :.  *     . *        :. .

endoPGI__A_niger        TLEDDATQVYILCGDGSCSDWTWSGVDLSGGKTSDKCENVPSGASC--- 337
endoPG__A_aculeatus     SVVSSGTNILISCGSGSCSDWTWTDVSVSGGKTSSKCTNVPSGASC--- 339
endoPGII__A_niger       SVDSGATEIYLLCGSGSCSDWTWDDVKVTGGKKSTACKNFPSVASC--- 335
endoPG__F_moniliforme   TVASSAQDWFILCGDGSCSGFTFSGNAITGGGKTSSCNYPTNTCPS--- 349
endoPG__S_purpureum     KVNNAATRVTVECGN-CSGNWNWSQLTVTGGKAGTIKSDKAKITGGQYL 335
endoPG__E_carotovora    -VTSETKGVVVLNGENAKKPIEVTMKNVKLTSDSTWQIKNVNVKK---- 376
                             :  .     : *. .             :.            .
```

Figure 1.9. Sequence alignment preformed with ClustalW server, available online at the following url: http:// www.ebi.ac.uk/clustalw. The sequences correspond to: endoPG I of *Aspergillus niger* (acc. num.: P26213); endoPG of *Aspergillus aculeatus* (acc. num.: O74213); endoPG II of *Aspergillus niger* (acc. num.: P26214); endoPG of *Fusarium moniliforme* (acc. num.: Q07181); endoPG of *Stereum purpureum* (acc. num.: Q400N6); endoPG of *Erwinia carotovora* (acc. num.: P26509). The sequences are aligned without N-terminal signal peptide, as indicated by the SwissProt server for each entry. The strictly conserved sequences are shown in red, while several invariant residues are labeled in blue.

As suggested from the alignment, there are four strictly conserved sequence segments: NTD, DD, GHG, RIK. Several invariant residues have been also observed. To provide a clear explanation for the role of all conserved amino acid residues, chemical amino acid modifications experiments (39, 40) and site-directed mutagenesis studies (25, 41) have been reported. These data, together with the crystal structure determination of some bacterial and fungal endoPGs (24-29), have allowed the identification of the residues that are involved in catalysis. Therefore, the first aspartate in the segment **D**D is considered to be the general acid catalyst while a water molecule, which is found in a conserved position in the crystal structures (24, 26), is supposed to be the nucleophile. The activation of the water molecule is ascribed to two conserved aspartate residues that act as general bases. These residues are the aspartate residue belonging to the NT**D** segment and the second aspartate residue of the segment D**D**. According to site-directed mutagenesis studies (26, 41), the mutations of these aspartate strictly affect the

catalytic activity, resulting in a significant reduced activity if compared to the wild-type enzyme. These mutations do not affect the K_M values, indicating that the aspartates are therefore not involved in the substrate binding. Interestingly, both conservative and non conservative mutations result in a complete enzymatic inactivation, indicating that not only the charge but also the relative distance between these carboxylate groups must be maintained (26).

The crystal structures determination of an endoPG from *Stereum purpureum* in complex with monogalacturonate molecules resulted to be essential to confirm the inverting mechanism (28). In order to describe the architecture of the active site of endoPGs, a conventional nomenclature has been adopted (42): accordingly, the enzymatic regions that interact with the polymeric substrate are referred to as *subsites*. Precisely, the term *subsite* indicates the binding site of each saccharide unit on the protein scaffold. By convention, subsites are labeled from *-n* to *+n*, where *-n* represents the non-reducing end and *+n* the reducing end of the substrate. The cleavage occurs between the subsites -1 and +1.

In the first complex of the *S. purpureum* endoPG with monogalacturonate, a β-D-galactopyranuronic acid molecule (Gal*p*A), a hydrolytic product of the enzymatic reaction, was found in the proximity of the conserved aspartate residues. In particular, the O4 of Gal*p*A is engaged in a hydrogen bond with OD2 of the first aspartate of the segment **D**D (Asp 173), the supposed proton donor residue. This interaction strongly suggest that O4 is involved in the glycosidic bond that will be cleaved in the course of the reaction. Thus Gal*p*A binding site is believed to be the subsite +1, because of its location at the reducing end of the substrate moiety.

In the second complex, two residues of monogalacturonic acid were identified: one molecule was located at the same position as that of the corresponding Gal*p*A binding site in the first complex. The second one was identified as β-D-galactofuranuronic acid (Gal*f*A). This molecule was found on the opposite site of the proposed subsite +1, across the catalytic aspartate residues, and its binding site was considered as the subsite -1. The Figure 1.10 shows the interactions between the enzyme and the ligands in the ternary complex.
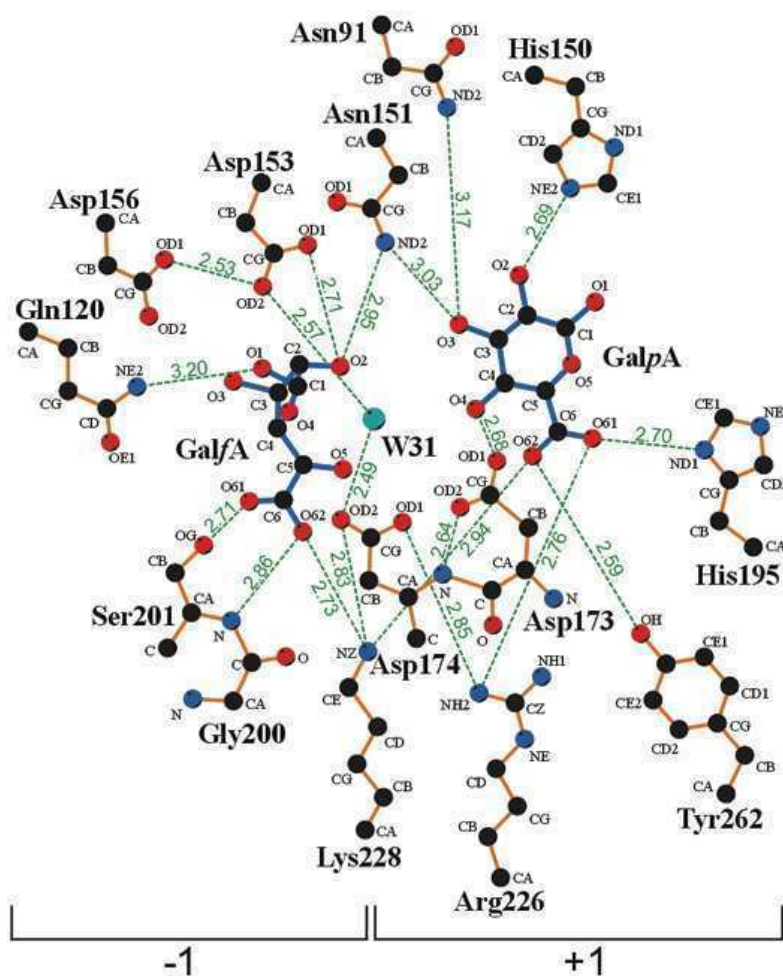
Figure 1.10. Schematic representation of the interactions between the enzyme and the ligands, Gal*p*A and Gal*f*A, in the ternary complex. The dotted lines show hydrogen bonds and electrostatic interactions, and their distances are indicated in angstroms. The Figure is taken from T. Shimizu *et al.*, (28).

The binding of Gal*p*A at subsite +1 involves several conserved residues: the histidine residue belonging to segment G**H**G (H195), the invariant tyrosine residue (**Y**262) and the arginine and the lysine residues belonging to **R**I**K** segment (R226 and K228). The site-directed mutagenesis studies of *Aspergillus niger* endoPG II (41, 43) showed that the replacement of these residues led to an increases in the K_M value, thus indicating that they play an essential role in the binding of the substrate. The presence of arginine and lysine confers to the substrate-binding cleft an overall positive electrostatic potential, thus favouring the binding of the carboxyl group of the substrate.

A water molecule (W31) has been found in the crystal complex, in the proximity of the

O4 atom of Gal*p*A. Its interaction with the carboxy groups of aspartate residues of the NT**D** segment (D153) and of the second aspartate of segment D**D** (D174) strongly suggests that W31 is the nucleophile while the aspartates are the general bases. In the *A. niger* endoPG II, the water molecule at the equivalent position is considered to be the nucleophilic water in the hydrolytic reaction (25).

Concerning the subsite - 1, the crystal complexes of the endoPG from *S. purpureum* with ligands revealed that the binding of the carboxylic group of Gal*f*A involves a unique nonprolyl *cis*-peptide bond between the invariant residue glycine (**G**200) and the following serine residue (**S**201) (28). The flexible glycine residue is necessary to maintain this *cis*-peptide bond. Moreover, in that position, no space for bulkier side chain is available. This nonprolyl *cis*-peptide bond is conserved in all the known structures of endoPGs and in the only known structure of a rhamnogalacturonase (44). The carboxylic group of Gal*f*A also interacts with the lysine of the RI**K** segment (K228). It has been proposed that this lysine together with the invariant glycine and serine residues form a carboxylic group recognition motif at the subsite - 1, that displays a structural role in the definition of the substrate specificity. Lastly, in the model for substrate binding of the endoPG from *Aspergillus aculeatus* (27), the authors suggested that the galacturonate bound in the subsite -1 could adopt a distorted half-chair conformation, referred to as $^4H_3$. This conformation, in which C2, C1, O5 and C5 are coplanar, is proposed to be important in the stabilization of the oxocarbenium-ion-like transition state (37).

The description of the subsites – 1 and + 1 led to define the orientation of the substrate molecule with respect to the protein backbone: the reducing end of the substrate molecule results to be directed toward the carboxy-terminal (C-terminal) portion of the enzyme, while the non -reducing end appeared to be orientated toward the amino-terminal (N-terminal) of the protein (Figure 1.11).
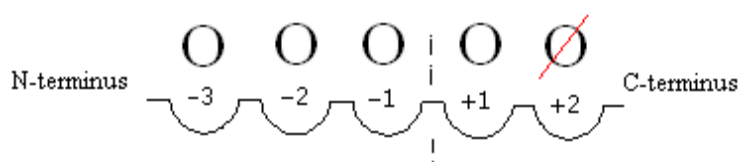
Figure 1.11. A graphical view of the interaction between enzyme and its oligomeric substrate. The red line indicates the reducing end of the substrate.

To complete the description of the role of the conserved amino acids some observations on the histidine residue belonging to the G**H**G segment and on the invariant **Y** can be drawn out. The involvement of the histidine residue in the catalytic machinery has been elucidated by the means of chemical modifications studies (45, 46) and of site-directed mutagenesis studies (25, 47). Therefore, an indirect role of the histidine residue in the catalysis has been proposed. The conserved histidine residue could be involved in the maintaining of the proper ionization state of the general acid catalyst carboxylate by sharing a proton. The role of the invariant residue **Y** is instead very controversial. On the basis of the three dimensional structure of *A. niger* endoPG II (25) and of *S. purpureum* endoPG (28), the authors suggest that the invariant residue **Y** could be included at the subsite +1. In contrast, in the model of the *A. aculeatus* endoPG in complex with an octagalacturonate (27), the authors indicate that the corresponding residue Y could directly interact with the galacturonate moiety at subsite -1. Actually, as reported for other glycoside hydrolases like cellulase (48) and xylanase (49), an invariant tyrosine is proposed to be involved in the stabilization of the oxocarbenium-like ion in the transition state. Since a distortion of the sugar ring in the enzyme-substrate (ES) complex of endoPGs has been proposed to occur at the subsite -1, most probably the invariant tyrosine belongs to the subsite -1.

1.3.2.4 The mode of action

The mode of action of endoPGs on the polymeric substrate allows their classification as *single attack* or *multiple attack* enzymes. Concerning the *single attack* enzymes, the substrate molecules are cleaved only once after the formation of the ES complex, thus the enzyme releases oligomers of different chain length. On the other hand, the *multiple attack* enzymes, called also processive enzymes, can cleave the same substrate molecule several times, and small oligomers, even monomer, are released in the initial stage of the reaction. Therefore, the detection of the hydrolysis products generated in the early stage of the HGA hydrolysis has been proven to be an useful tool to investigate the mode of action of the endoPGs (50-55). As previously mentioned, the interaction between endoPGs and its substrate involves multiple binding sites, referred to as subsites. The subsites at position -1 and +1 are those that have been better characterized,

since they include amino acids directly involved in catalysis. Until now, no crystal structures of endoPGs in complex with oligogalacturonides (OGAs) have been described, preventing the clear definition of the amino acids involved in the binding and in the recognition of the oligogalacturonate moiety. Nevertheless, beside site-directed mutagenesis studies (38), the analysis of the products released from the hydrolysis of OGAs has been widely reported, in order to map the number and the affinity of the subsites of endoPGs (50-55).

1.3.2.5. The β-helix fold

All the six endoPGs, whose crystal structures are known, exhibit a right handed β-helix fold (24-29). The β-helix fold was first described in the three dimensional structure of *Erwinia chrysanthemi* pectate lyase C (56) (Figure 1.12).
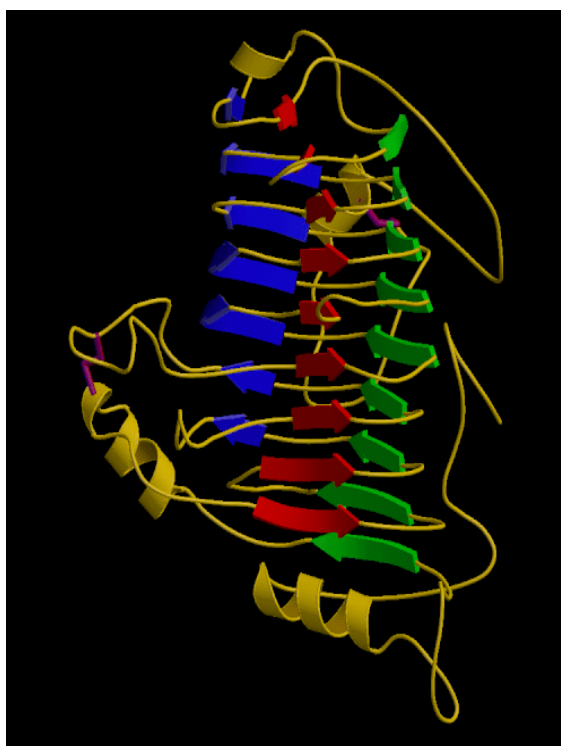
Figure 1.12. The three-dimensional structure of pectate lyase C from *Erwinia chrysanthemi* (51)*.* The two disulphide bonds are shown in violet. The predominant secondary structure is represented by three parallel β-sheets, shown here in blue, red and green.

The β-helix fold consists of parallel β-strands, coiled around a vertical axis to form a large right-handed cylinder. The right-handed β-helices structures are stabilized by an extensive hydrogen bond network between parallel β-strands as well as by highly ordered stacking of side chains in the interior of the cylinder (57). As concerns the extremes of the β-helices, the N-terminal end of the parallel β-helix domain is capped by an α-helix, weakly amphipathic in character, that is structurally conserved. No structural conservation instead occurs in the capping of the C-terminal end of the cylinder although this part of the protein is often found folded in extended conformation. In addition, the presence of disulphide bridges further stabilizes the β-helix fold. They occur in variable number in endoPGs of different sources (38). Both fungal and bacterial endoPGs exhibit the conservation of a disulphide bond in their N-terminal α-helix caps. Other examples of carbohydrate-binding proteins sharing this type of fold include a rhamonogalacturonase from *Aspergillus aculeatus* (44), a chondroitinase B from *Pedobacterium heparium* (58), a ι-carrageenase from *Alteromonas fortis* (59) and a dextranase from *Penicillium linoleum* (60). The functional feature common to these proteins is their interaction with polymeric molecules. Actually, most β-helix proteins are involved in the binding and in the degradation of polyanionic polysaccharides with high linear charge densities, that compose the extracellular matrix of plants and animals, like pectins, ι-carrageenans and dermatan sulphate. Most probably the elongated shape of β-helix proteins is particularly suitable to maximize the interactions with their linear polymeric substrates reported above.

Right handed parallel β-helix folds have been found in P22 tailspike endorhamnosidase that degrades oligosaccharides found on the cell surface receptors (32-34), and in *Bordetella pertussis* P.69 pertactin (61), whose biological function is still under investigation.

Very recently, the first crystal structure of a cedar allergen, *Jun a 1*, from the pollen of the mountain cedar *Juniperus ashei*, has been determined (62). This structure revealed a parallel β-helix fold, which is nearly identical to that found in the bacterial pectin/pectate lyases; nevertheless, *Jun a 1* lacks any hydrolytic activity (63). This is the first description of this type of structure in a higher plant or animal.

Parallel β-helices, which fold into a left-handed cylinder, have been also found (64-66). The left-handed parallel β-helices differ significantly from the right-handed ones. In the left-handed parallel β-helix the sheets are extremely flat and the connections between adjacent β-strands are longer. The cross-section of the left-handed parallel β-helix is much closer to an equilateral triangle and the sequence much more repetitive.

Due to the structural regularity, the parallel β-fold has been proposed as a structural model in the investigation on the assembly of amyloid fibrils (67, 68). In particular, the left-handed β-helix provides a coherent framework to explain the seeded growth characteristics of β-sheet structure in amyloid fibrils and it has been found consistent with many structural, biochemical and immunological features of prions (69).

1.3.2.6. Physiological context

Microbial and fungal endoPGs are important virulence factors in plant diseases (70-74), secreted in the early stage of infection (Figure 1.13) (75). Their catalytic activity may expose underlying components of cell wall to other cell wall degrading enzymes leading pathogens to penetrate plant tissues. Nevertheless, it has been shown that the OGAs derived from enzymatic hydrolysis of pectins are able to evoke the plant defence response, promoting the synthesis of a variety of molecules (76, 77). Among these, specific polygalacturonase-inhibiting proteins (PGIPs) are secreted by plants (78, 79). In the presence of PGIPs, endoPGs retain about 0.3% of their initial catalytic activity and a modified distribution of the hydrolysis products has been observed (80). *In vitro*, endoPGs hydrolyse polygalacturonic acid into monomer, dimer and trimer (50-55): in presence of PGIPs, the end products are oligomers with a degree of polymerization higher than nine. The length of the oligomers is the key point to convert the OGAs from elicitor-inactive molecules into elicitor-active molecules, which are able to trigger the plant defence (80). Interestingly plant PGIPs, shown to be active on fungal PGs, do not seem to be effective on bacterial and plant PGs (81). The interaction between endoPGs and PGIPs has been also extensively studied from a structural point of view, since the crystal structure of a PGIP form *Phaseoul vulgaris* has been elucidated (82). The structural analysis has proven to be an essential tool to identify the structural determinants involved in the interaction and in the recognition between endoPGs and

PGIPs (26, 83). Recently, surface $H/^2H$ exchange in combination with mass spectrometry has been used to study the interaction of *A. niger* PG II with its substrate and with bean PGIP (84).
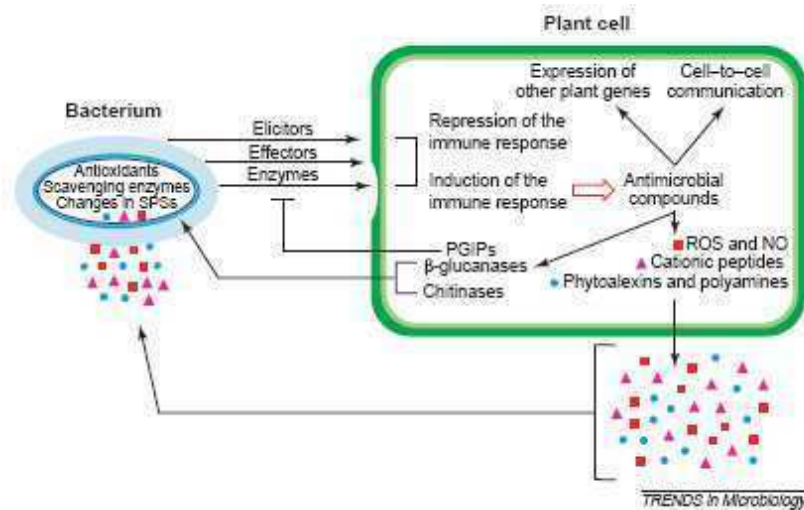
Figure 1.13. Schematic representation of a Gram-negative bacterium interacting with a plant cell. The bacterial cell envelope is composed of an outermost, thick extracellular polysaccharide layer (thick light blue), the outer membrane with lipo- and capsular polysaccharides (thin blue), the periplasmic space with peptidoglycan and b-glucans (greenish blue), and the inner membrane (thin dark blue). The plant cell is surrounded by a cell wall (thick dark green) and a cytoplasmic membrane (thin light green). The bacterium secretes elicitors, effectors (such as AvrPtoB (Genbank accession number AE016867)), and cell-wall-degrading enzymes that might repress or induce the plant immune response. In the latter case, plants produce a set of antimicrobial compounds, including polygalacturonase-inhibiting proteins (PGIPs), β-glucanases and chitinases. PGIPs inhibit the action of certain cell-wall-degrading enzymes; β-glucanases and chitinases can degrade the peptidoglycan of the bacterial cell envelope. Reactive oxygen species (ROS) and nitric oxide (NO; red squares), cationic peptides (such as defensins; pink diamonds), and phytoalexins and polyamines (blue circles) might be generated to attack the bacterium. Various bacterial surface polysaccharides (SPSs) form a diffusion barrier, or sequestrate or repel antimicrobial compounds. The bacterium produces antioxidants (such as glutathione) and ROS-scavenging enzymes (such as catalases), and can alter gene expression to change the properties of the SPSs ensuring an enhanced protection. The antimicrobial compounds produced by the plant might induce the expression of other plant genes and establish intercellular communication leading to systemic-acquired resistance to protect against the spread of the invading bacterium. The Figure and the Legend are taken from W. D'Haeze *et al*., (74).

1.4 Bibliography

(1) C. Somerville, S. Bauer, G. Brininstool, M. Facette, T. Hamann, J. Milne, E. Osborne, A. Paredez, S. Persson, T. Raab, S. Vorwerk, H. Youngs, Toward a Systems Approach to Understanding Plant Cell Walls, Science 306 (2003), 2206-2211.

(2) M. McNeil, A.G. Darvill, S.C. Fry, P. Albersheim, Structure and function of the primary cell walls of plants, Annu. Rev. Biochem. 53 (1984) 625-663.

(3) W.G. Willats, L. McCartney, W. Mackie, J.P. Knox, Pectin: cell biology and prospects for functional analysis. Plant Mol Biol 47 (1-2) (2001) 9-27.

(4) B.L. Ridley, M.A. O'Neill, D. Mohnen, Pectins: structure, biosynthesis and oligogalacturonide-related signaling, Phytochemistry 57(6) (2001) 929-967.

(5) T. Ishii, O-acetylated oligosaccharides from pectins of potato tuber cell walls, Plant Physiol. 113 (1997)1265–1272.

(6) L. Saulnier and J.-F. Thibault , Ferulic acid and diferulic acids as components of sugar-beet pectins and maize bran heteroxylans, J. Sci. Food Agric. 79 (1999) 396–402.

(7) M. Kobayashi, H. Nakagawa, T. Asaka and T. Matoh , Borate-rhamnogalacturonan II bonding reinforced by Ca^{2+} retains pectic polysaccharides in higher plants cell walls. Plant Physiol. 119 (1999) 199–203.

(8) T. Matoh and M. Kobayashi, Boron and calcium, essential inorganic constituents of pectic polysaccharides in higher plant cell walls, J. Plant Res. 111 (1998) 179–190.

(9) D.A. Powell, E.R. Morris, M.J. Gidley and D.A. Rees, Conformations and interactions of pectins. II. Influence of residue sequence on chain association in calcium pectate gels, J. Mol. Biol. 155 (1982) 517–531.

(10) T. Ishii, T. Matsunaga, P. Pellerin, M.A. O'Neill, A. Darvill, P. Albersheim, The plant cell wall polysaccharide rhamnogalacturonan II self-assembles into a covalently cross-linked dimer, J. Biol. Chem. 274 (1999) 13098–13104.

(11) S. Pérez, K. Mazeau and C. Hervé du Penhoat, The three-dimensional structures of the pectic polysaccharides, Plant Physiol. Biochem. 38 (2000) 37-55.

(12) M-C. Ralet, V. Dronnet, H. C. Buchholt and J.F Thibault, Enzymatically and chemically de-esterified lime pectins: characterisation, polyelectrolyte behaviour and calcium binding properties, Carbohydr. Res. 336 (2001) 117-125.

(13) G.T. Grant, E.R. Morris, D.A. Rees, P.J.C. Smith and D. Thom , Biological interactions between polysaccharides and divalent cations: the "egg box" model, FEBS Lett. 32 (1973) 195–198.

(14) R. Chandrasekaran, P.V. Balaji, P.K. Qasba, V.S.R. Rao, Conformation of Carbohydrates, Eds.:Taylor & Francis,1998.

(15) E.R. Morris, D.A. Powell, M.J. Gidley, D.A. Rees, Conformation and interactions of pectins I. Polymorphism between gel and solid states of calcium polygalacturonate, J. Mol. Biol. 155 (1982) 507–516.

(16) O.P. Ward, M. Moo-Young, Enzymatic degradation of cell wall and related plant polysaccharides, Crit. Rev. Biotechnol. 8 (1989) 237-274.

(17) R.P. de Vries, J. Visser, *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides, Microbiol. Mol. Biol. Rev. 65 (2001) 497-522.

(18) B. Henrissat, A classification of glycosyl hydrolases based on amino acid sequence similarities, Biochem. J. 280 (1991) 309-316.

(19) G.J. Davies, T.M. Gloster, B.Henrissat, Recent structural insights into the expanding world of carbohydrate-active enzymes, Curr. Opin. Struct. Biol. 15 (2005) 637-645.

(20) G.D. McCarter, S.G. Withers, Mechanisms of enzymatic glycoside hydrolysis, Curr. Opin. Struct. Biol. 4 (1994) 885-892.

(21) G.J. Davies, T.M. Gloster, B.Henrissat, Structures and mechanisms of glycosyl hydrolases, Structure 3 (1995) 853-859.

(22) P. Biely, J. Benen, K. Heinrichova, H.C.M. Kester, J. Visser, Inversion of configuration during hydrolysis of α-(1,4)- galacturonidic linkage by three *Aspergillus polygalacturonases*, FEBS Lett. 382 (1996) 249–255.

(23) S.M. Pitson, M. Mutter, L.A. Van der Broek, A.G. Voragen, G. Beldman, Stereochemical course of hydrolysis catalysed by α-L-rhamnosyl and α-D-galacturonosyl hydrolases from *Aspergillus aculeatus*, Biochem. Biophys. Res. Commun. 242 (1998) 552-559.

(24) R. Pickersgill, D. Smith, K. Worboys, J. Jenkins, Crystal Structure of Polygalacturonase from *Erwinia carotovora* ssp. *Carotovora*, J. Biol. Chem. 273 (1998) 24660-24664.

(25) Y. Van Santen, J.A.E. Benen, K.-H. Schroter, K.H. Kalk, S. Armand, J. Visser,

B.W. Dijkstra, 1.68 a crystal structure of endopolygalacturonase II from *Aspergillus Nige*r and identification of active site residues by site-directed mutagenesis, J. Biol. Chem. 274 (1999) 30474– 30480.

(26) L. Federici, C. Caprari, B. Mattei, C. Savino, A. Di Matteo, G. De Lorenzo, F. Cervone, D. Tsernoglou, Structural requirements of endopolygalacturonase for the interaction with endoPGIP (polygalacturonase-inhibiting protein), Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 13425– 13430.

(27) S.W. Cho, S. Lee, W. Shin, The X-ray structure of *Aspergillus aculeatus* Polygalacturonase and a modeled structure of the polygalacturonase–octagalacturonate complex, J. Mol. Biol. 314 (2001) 863– 878.

(28) T. Shimizu, T. Nakatsu, K. Miyairi, T. Okuno, H. Kato, Active-site architecture of endopolygalacturonase I from *Stereum purpureum* revealed by crystal structures in native and ligand-bound forms at atomic resolution, Biochemistry 41 (2002) 6651– 6659.

(29) G. van Pouderoyen, H. Snijder, J. Benen, B. Dijkstra, Structural insights into the processivity of endopolygalacturonase I from *Aspergillus niger*, FEBS Letters 554 (2003) 462-466.

(30) S. Iwashita, S. Kanegasaki, Enzymatic and molecular properties of base-plate parts of bacteriophage P22, Eur. J. Biochem. 65 (1976) 87-94.

(31) U. Eriksson, S.B. Svenson, J. Lönngren, A.A. Lindberg, *Salmonella* phage glycanases- substrate specificity of the phage P22 endo-rhamnosidase, J. Gen. Virol. 43 (1979) 503-511.

(32) S. Steinbacher, R Seckler, S. Miller, B.Steipe, R. Huber, Reinemer, Crystal structure of P22 tailspike protein: interdigitated subunits in a thermostable trimer, Science 265 (1994) 383–386.

(33) S. Steinbacher, U. Baxa, S. Miller, A. Weintraub, R. Seckler, R.Huber, Crystal-structure of phage-P22 tailspike protein complexed with *Salmonella spp.* o-antigen receptors, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 10584–10588.

(34) S. Steinbacher, S. Miller, U. Baxa, N. Budisa, A. Weintraub, R. Seckler, R. Huber, Phage P22 tailspike protein: Crystal structure of the head-binding domain at 2.3Å, fully refined structure of the endorhamnosidase at 1.56Å resolution, and the molecular basis of O-antigen recognition and cleavage, J. Mol. Biol. 267 (1997) 865–880.

(35) W. Nerinchx, T. Desmet, K.Piens, M. Claeyssens, An elaboration on the syn-anti proton donor concept of glycoside hydrolases: electrostatic stabilisation of the transition state as a general strategy, FEBS Lett. 579 (2005) 302-12.

(36) A.Vasella, G.J. Davies, M. Bohm, Glycosidase mechanism, Curr. Opin. Chem. Biol. 6 (2002) 619-29.

(37) C.S. Rye, S.G. Withers, Glycosidase mechanisms, Curr. Opin. Chem. Biol. 4 (2000) 573-80.

(38) O.Markovic, S. Janecek, Pectin degrading glycoside hydrolases of family 28: sequence-structural features, specificities and evolution, Protein Eng. 14 (2001) 615-631.

(39) E. Stratilova, M. Dzurova, O. Markovic, H. Jornvall, An essential tyrosine residue of *Aspergillus polygalacturonase*, FEBS Lett. 382 (1996) 164-166.

(40) S.K. Niture, A. Pant, A.R. Kumar, Active site characterization of the single endo-polygalacturonase produced by *Fusarium moniliforme* NCIM 1276, Eur J Biochem. 268 (2001) 832-840.

(41) S. Armand, M.J.M. Wagemaker, P. Sanchez-Torres, H.C.M. Kester, Y. van Santen, B.W. Dijkstra, J. Visser, J.A.E. Benen, The active site topology of *Aspergillus niger* endopolygalacturonase II as studied by site-directed mutagenesis, J. Biol. Chem. 275 (2000) 691– 696.

(42) G.J. Davies, K.S. Wilson, B. Henrissat, Nomenclature for sugar-binding subsites in glycosyl hydrolases, Biochem. J. 321 (1997) 557-559.

(43) S. Pages, W.H. Hejine, H.C.M. Kester, J. Visser, J.A.E. Benen, Subsite mapping of *Aspergillus niger* endopolygalacturonase II by site-directed mutagenesis, J. Biol Chem. 275 (2000) 29348-29535.

(44) T.N. Petersen, S. Kauppinen, S. Larsen, The crystal structure of rhamnogalacturonase A from *Aspergillus aculeatus*: a right-handed β-helix, Structure 5 (1997) 533–544.

(45) M.n. Rao, A.A. Kembhavi, A. Pant, Implication of a tryptophane and histidine in the active site of endo-polygalacturonase from *Aspergillus ustus*: elucidation of the reaction mechanism, Biochim. Biophys. Acta 1296 (1996) 167-173.

(46) T.C. Jyothi, S.A.Singh, A.G. Appu Rao, The contribution of ionic interactions to the conformational stability and function of polygalacturonase from *A. niger*, Int. J.

Biol. Macromol. 36 (2005) 310-317.

(47) C. Caprari, B. Mattei, M.L. Basile, G. Salvi, V. Crescenzi, G. De Lorenzo, F. Cervone, Mutagenesis of endopolygalacturonase from *Fusarium moniliforme*: histidine residue 234 is critical for enzymatic and macerating activities and not for binding to polygalacturonase-inhibiting protein (PGIP), Mol. Plant-Microbe Interact. 9 (1996) 617-624.

(48) G. Sulzenbacher, H. Driguez, B.Henrissat, M. Schülein, G.J. Davies, Structure of the *Fusarium oxysporum* endoglucanase I with a nonhydrolyzable substrate analogue: substrate distortion gives rise to the preferred axial orientation for the leaving group, Biochemistry 35 (1996) 15280-15287.

(49) G. Sidhu, S.G. Withers, N.T. Nguyen, L.P. McIntosh, L. Ziser, G.D. Brayer, Sugar Ring Distortion in the Glycosyl-Enzyme Intermediate of a Family G/11 xylanase, Biochemistry 38 (1999) 5346-5354.

(50) L. Parenicova, J.A.E. Benen, H.C.M. Kester, J. Visser, pgaE encodes a fourth member of the endopolygalacturonase gene family from *Aspergillus niger*, Eur. J. Biochem. 251 (1998) 72-78.

(51) J.A.E. Benen, H.C.M. Kester, J. Visser, Kinetic characterization of *Aspergillus niger* N400 endopolygalacturonases I, II and C, Eur. J. Biochem. 259 (1999) 577– 585.

(52) L. Parenicova, H.C.M. Kester, J.A.E. Benen, J. Visser, Characterization of a novel endopolygalacturonase from *Aspergillus niger* with unique kinetic properties, FEBS Lett. 467 (2000) 333–336.

(53) L. Parenicova, J.A.E. Benen, H. Kester, J. Visser, endoPGaA and endoPGaB encode two constitutively expressed endopolygalacturonases of *Aspergillus niger*, Biochem. J. 345 (2000) 637–644.

(54) E. Bonnin, A. Le Goff, R. Kfrner, G.J.W.M. van Alebeek, T.M.I.E. Christensen, A.G.J. Voragen, P. Roepstorff, A. Caprari, J.-F. Thibault, Study of the mode of action of endopolygalacturonase from *Fusarium moniliforme*, Biochim. Biophys. Acta 1526 (2001) 301–309.

(55) I. Kars, G.H. Krooshof, L. Wagemakers, R. Joosten, J.A. Benen, J.A. van Kan, Necrotizing activity of five *Botrytis cinerea* endopolygalacturonases produced in *Pichia pastoris*, Plant J. 43 (2005) 213-225.

(56) M.D. Yoder, N.T. Keen, F. Jurnak, New domain motif: structure of pectate lyase C,

a secreted plant virulence factor, Science 260 (1993) 1503–1507.

(57) J. Jenkins, O. Mayans, R. Pickersgill, Structure and evolution of parallel β-helix proteins, J. Struct. Biol. 122 (1998) 236-246.

(58) W. Huang, A. Matte, Y. Li, Y.S. Kim, R.J. Linhardt, H. Su, M. Cygler, Crystal structure of chondroitinase B from *Flavobacterium heparinum* and its complex with a disaccharide product at 1.7 A resolution, J. Mol. Biol. 294 (1999) 1257-1269.

(59) G. Michel, L. Chantalat, E. Fanchon, B. Henrissat, B. Kloareg, O. Dideberg, The ι-Carrageenase of *Alteromonas fortis* A β-helix fold-containing enzyme for the degradation of a highly polyanionic polysaccharide, J. Biol. Chem. 276 (2001) 40202-40209.

(60) A.M. Larsson, R. Andersson, J. Stahlberg, L. Kenne, T.A. Jones, Dextranase from *Penicillium minioluteum*: reaction course, crystal structure, and product complex, Structure 11(9) (2003) 1111-21.

(61) P. Emsley, I.G. Charles, N.F. Fairweather, N. W. Isaacs, Structure of *Bordetella pertussis* virulence factor P.69 pertactin, Nature 381 (1996) 90–92.

(62) E..W. Czerwinski, T. Midoro-Horiuti, M.A.White, E.G.Brooks, R.M. Goldblum, Crystal structure of *Jun a 1*, the major cedar pollen allergen from *Juniperus ashei*, reveals a parallel β-helical core, J. Biol. Chem. 280 (2005) 3740-3746.

(63) T. Midoro-Horiuti, V. Mathura, C.H. Schein, W. Braun, S. Yu, M. Watanabe, J.C. Lee, E.G. Brooks, R.M. Goldblum, Major linear IgE epitopes of mountain cedar pollen allergen *Jun a 1* map to the pectate lyase catalytic site, Mol. Immunol. 40 (2003) 555-562.

(64) C.R.H. Raetz, S.L. Roderick, A left-handed parallel β-helix in the structure of UDP-N-acetylglucosamine acyltransferase, Science 270 (1995) 997–1000.

(65) C. Kisker, H. Schindelin, B.E. Alber, J.G. Ferry, D.C. Rees, A left-handed β-helix revealed by the crystal structure of a carbonic anhydrase from the archaeon *Methanosarcina thermophila*, EMBO J. 15 (1996) 2323–2330.

(66) T.W. Beaman, D.A. Binder, J.S. Blanchard, S.L. Roderick, Three-dimensional structure of tetrahydrodipicolinate N-succinyltransferase, Biochemistry 36 (1997) 489–494.

(67) D. Zanuy, K. Gunasekaran, A.M. Lesk, R. Nussinov, Computational study of the

fibril organization of polyglutamine repeats reveals a common motif identified in β-helices, J. Mol. Biol. 358 (2006) 330-345.

(68) H.H. Tsai, K. Gunasekaran, R. Nussinov, Sequence and structure analysis of parallel β-helices: implication for constructing amyloid structural models, Structure 14 (2006) 1059-1072.

(69) C. Govaerts, H. Wille, S.B. Prusiner, F.E. Cohen, Evidence for assembly of prions with left-handed β-helices into trimers, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 8342-3347.

(70) A. Collmer, N.T. Keen, The role of pectic enzymes in plant pathogenesis, Annu. Rev. Phytopahtol. 24 (1986) 383–409.

(71) R.A. Prade, D. Zhan, P. Ayoubi, A.J. Mort, Pectins, pectinases and plant-microbe interactions, Biotechnol. Genet. Eng. Rev. 16 (1999) 361-391.

(72) P.K. Durrands, R.M. Cooper, The role of pectinases in vascular wilt disease as determined by defined mutants of *Verticillium albo-altrum*, Physiol. Mol. Plant. Pathol. 32 (988) 363–371.

(73) S.R. Herron, J.A. Benen, R.D. Scavetta, J. Visser, F. Jurnak, Structure and function of pectic enzymes: Virulence factors of plant pathogens, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 8762-8769.

(74) W. D'Haeze, M. Holsters, Surface polysaccharides enable bacteria to evade plant immunity,Trends Microbiol.12 (2004) 555-561.

(75) B. Dumas, S. Centis, N. Sarrazin, M-T Esquerré-Tugayé, Use of Green Fluorescent Protein To Detect Expression of an Endopolygalacturonase Gene of *Colletotrichum lindemuthianum* during Bean Infection, Appl. Environ. Microbiol. 65 (1999) 1769-1771.

(76) A. Darvill, C. Augur, R.W. Carlson, S.J.J.Cheong, S. Eberhard, M.G. Hahn, V.M. Lo, F. Marfa, B. Meyer, D. Mohnen, M.A. O'Neil, M.D. Spiro, H. Van Halbeek, W.S.York, P. Albersheim, Oligosaccharins: oligosaccharides that regulate growth, development and defence responses in plants, Glycobiology 2 (1992) 181–198.

(77) E. A. Nothnagel, M. McNeil, P. Albersheim, A. Dell, Host-Pathogen Interactions XXII. A Galacturonic Acid Oligosaccharide from Plant Cell Walls Elicits Phytoalexins, Plant Physiol. 71 (1983) 916-926.

(78) G. De Lorenzo, S. Ferrari, S. Polygalacturonase-inhibiting proteins in defense against phytopathogenic fungi, Curr. Opin. Plant Biol. 5 (2002) 295-299.

(79) R. D'Ovidio, B. Mattei, S. Roberti, D. Bellicampi, Polygalacturonases, polygalacturonase-inhibiting proteins and pectic oligomers in plant-pathogen interactions, Biochim. Biophys. Acta. 1696 (2004) 237-244.

(80) F. Cervone, M.G. Hahn, G. De Lorenzo, A. Darvill, P. Albersheim, Host-Pathogen Interactions XXXIII. A Plant Protein Converts a Fungal Pathogenesis Factor into an Elicitor of Plant Defense Responses, Plant Physiol. 90 (1989) 542–548.

(81) F. Cervone, G. De Lorenzo, R. Pressey, P. Albersheim, A.G. Darvill, Host-Pathogen Interactions XXXV. Can *Phaseolus* PGIP inhibit pectic enzymes from microbes and plants?, Phytochemistry 29 (1990) 447-449.

(82) A. Di Matteo, L. Federici, B. Mattei, G. Salvi, K.A. Johnson, C. Savino, G. de Lorenzo, D. Tsernoglou, F. Cervone, The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 10124-10128.

(83) A. Di Matteo, D. Bonivento, D. Tsernoglou, L. Federici, F. Cervone, Polygalacturonase-inhibiting protein (PGIP) in plant defence: a structural view, Phytochemistry 67 (2006) 528-533.

(84) D. King, M. Lumpkin, C. Bergmann, R. Orlando, Studying protein-carbohydrate interactions by amide hydrogen/deuterium exchange mass spectrometry, Rapid Commun. Mass Spectrom. 16 (2002) 1569-1574.

Aims and Outline of the Thesis

The aim of this Thesis is to contribute to the functional and structural characterization of an endopolygalacturonase (PehA) secreted by the plant pathogen *Burkholderia cepacia* ATCC 25416.

PehA has been isolated from the supernatant of a *B. cepacia* ATCC 25416 culture and a preliminary biochemical characterization has been carried out. The N-terminal amino acid sequence has been determined and revealed that the isolated protein matched a previously identified endopolygalacturonase. These results are reported in Chapter 2.

In order to obtain adequate amounts of protein required for structural and functional studies, the *pehA* gene was PCR-amplified and heterologously expressed in *E. coli* host cells, testing several expression strategies and systems. The isolation, the purification and the biochemical and biophysical characterization of the recombinant PehA are reported in Chapter 2.

In order to elucidate the hydrolysis mechanism of PehA, its mode of action on different types of substrates has been investigated. The study on the polymeric substrate has been aimed to verify whether PehA displays a processive or a non-processive mechanism of action on long substrate chains. In addition, the study of the hydrolysis mechanism on shorter substrate molecules that differ in their overall length and in their methylation pattern allowed to exploit and to shed lights on the architecture of the active site. The results of the mode of action of PehA are reported in Chapter 3.

A structural characterization of PehA has been undertaken following different approaches. Efforts have been made to obtain crystals of PehA of suitable quality and size for X-Ray diffraction studies using synchrotron radiation. The occurrence of six cysteine residues in the primary sequence of PehA suggests that they might be involved in the formation of disulphide bonds. The topology of the disulphide bridges has been assigned by MALDI-TOF mass-spectrometry analysis. This represents the first report of a disulphide bond mapping of a bacterial endoPG.

A three dimensional molecular model of PehA has been generated by computational methods based on profile-profile sequence alignments and fold recognition algorithms. The results are reported in Chapter 4.

Chapter 2. Isolation, heterologous expression and characterization of an endopolygalacturonase from the phytopathogen *B. cepacia*

2.1 Introduction

The plant cell wall provides and maintains cell shape and serves as a protective barrier (1). Plants can have two types of cell walls, a primary and a secondary. The primary walls of two cells are joined together by a common layer called the middle lamella. Primary walls are composed predominantly of polysaccharides such as cellulose, hemicellulose and pectic substances. Lignin is a major component of secondary walls. Generally, pectic substances exist in the cell wall as either smooth or hairy regions. Smooth regions are built up by linear chains of α-1,4 galacturonic acid residues. The carboxyl groups of the galacturonic acid residues can be methyl esterified. In addition, galacturonic acid can be O-acetylated at C-2 and/or C-3. The hairy regions contain rhamnogalacturonan I and II. Rhamnogalacturonan I is the second major type of polysaccharide. The backbone consists of a strictly alternating sequence of (1,4)-linked α-D-galacturonic acid units and α-(1,2)-linked L-rhamnose residues. The latter can carry side chains at C-4 consisting of arabinans, galactans or two different types of arabinogalactans (Figure 1.3, Chapter 1) (2, 3).

Pectic substances are subjected to enzymatic degradation not only during the course of important physiological processes, like plant senescence and ripening (4), but also

during infection events by plant pathogens (5, 6). In order to invade plant tissues, pathogens secrete a variety of cell wall degrading-enzymes that are responsible for the breakdown of different types of polysaccharides present in the cell walls (7, 8). The pectin degrading-enzymes include exo- and endo-polygalacturonases (endoPGs), pectin methylesterases, pectin and pectate lyases, acetylesterases and xylanases. Among them, endoPGs are the most extensively studied. These enzymes belong to the glycoside hydrolase family 28 (9) and are responsible for the hydrolysis of the α-1,4 glycosidic bond in the smooth regions of pectin.

Several endoPGs from pythopathogenic bacteria, such as *Ralstonia solanacearum*, *Erwinia carotovora* and *Agrobacterium vitis* have been isolated and characterized (10-12). EndoPGs are extracellular enzymes and their involvement in the plant infections strictly depends on the export efficiency into the extracellular environment (13). The molecular basis of the extracellular export of bacterial endoPGs have been therefore extensively studied (14-16). The export occurs via the general secretion pathway (GSP) type II, a two step process, that requires the presence of a signal peptide at the N-terminus of the protein (17) (Figure 2.1).
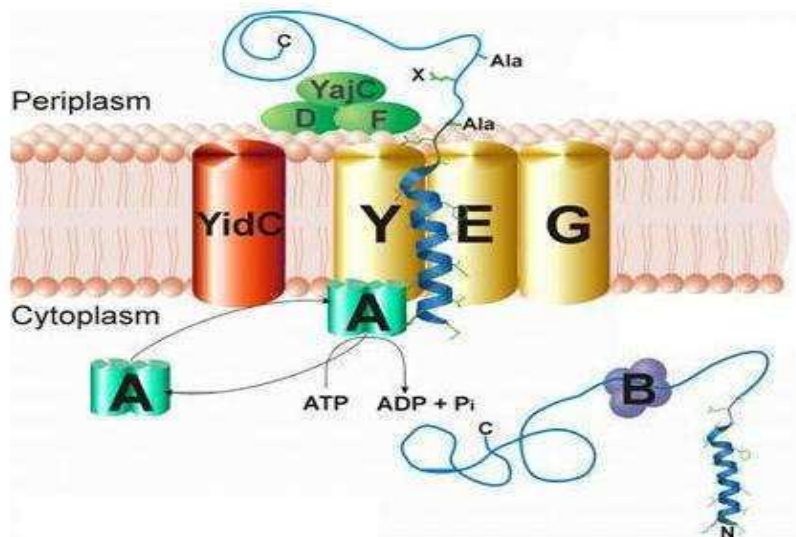
Figure 2.1. The Sec-system. The Sec-system is made up of the proteins SecA, SecB, SecD, SecE, SecF, SecG, SecY, YajC, and YidC. The molecular chaperone SecB interacts with the newly synthesized preprotein in the cytoplasm and targets the protein to the SecA-SecYEG translocase at the membrane surface. The homodimer SecA, in an energy dependant event (ATP), aids in the partial translocation of the presumably unfolded protein across the membrane. The pore through which the protein passes is thought to be formed by the integral membrane proteins SecY, SecE and SecG. The role of three additional membrane bound proteins, SecD, SecF and YajC is still under investigation. After the targeting and translocation steps have occurred, a signal peptidase cleaves off the signal peptide, released the protein in the periplasm. The Figure and the Legend are taken from: www.sfu.ca/~mpaetzel/Research.html.

In the first step, the polypeptide chains of the nascent proteins are kept in an unfolded state and are translocated into the periplasm by intimate interaction with the Sec proteins, that recognize the N-terminal signal peptide (18). In the second step, after the acquisition of their native conformation, proteins are translocated across the outer membrane by interacting with the enzymatic system that vary among gram-negative bacteria. Several endoPGs of phytopathogenic bacteria have been cloned in *E. coli* with their own N-terminal signal peptide. It has been reported that they mainly accumulate into the periplasm, namely because they are not recognized by the enzymatic machinery responsible for the translocation across the outer membrane of *E. coli* (14-16).

This work has been focussed on an endoPG secreted by the plant pathogen *Burkholderia cepacia*, strain ATCC 25416. The *B. cepacia* strain ATCC 25416 belongs to the *B. cepacia* complex (19) and it was first described as the casual agent of soft rot disease of onions (20). This strain produces a variety of pectin-degrading enzymes that are involved in the maceration of bulb scale and leaf tissue (21). Among them, the physiological role of an endoPG, named as pectic enzyme hydrolase (PehA), has been specifically investigated (22). Gonzales *et al.* established that the *pehA* gene is localized on a plasmid and reported the mobilization of a *pehA*-encoding plasmid in non-pathogenic *B. cepacia* strains of soil and clinical origin, proving that PehA is a virulence factor. Therefore the phytopathogenicity of *B. cepacia* spp. has been closely associated with a polygalacturonase activity.

Recent studies suggest that, in *B. cepacia,* ATCC 25416, the expression of the genes coding for virulence factors is controlled by the quorum sensing (QS) system (23). It has been also demonstrated that mutations in genes involved in QS regulation lead to the attenuation of tissue maceration if compared to the parental strain and to be responsible of a 40% reduction in the amount of PehA detected in the extracellular media (24).

Here we report on the isolation and the purification of PehA from the supernatant of the *B. cepacia* ATCC 25416 culture. In addition, the *pehA* gene was PCR-amplified and cloned in *E. coli*, testing several expression strategies and systems. Therefore, a biochemical and biophysical characterization of the recombinant enzymes has been carried out.

2.2 Materials and methods

2.2.1 Isolation, purification and characterization of PehA from *B. cepacia* culture (wt-PehA)

B. cepacia ATCC 25416 was grown in M9CA medium (25) at pH 5.2, except for the addition of 0.2% polygalacturonic acid (PGA) and 0.2% glycerol instead of glucose as carbon source. The culture was grown at 30 °C for two days on a rotary shaker. The induction of the enzymatic activity was estimated by growing *B. cepacia* in the presence and in the absence of PGA sodium salt from citrus fruit (Sigma Aldrich). Samples from the culture supernatant were taken periodically and tested for polygalacturonase activity. The enzymatic activity was measured using PGA as substrate and determined by the following two methods. Quantitative reducing-end formation was performed according to Bernfeld (26). A qualitative determination of the polygalacturonase activity was done according to the ruthenium red staining protocol (27). Namely, either the culture supernatant or the enzyme samples were spotted onto a 0.75 mm thick 1% agarose gel, containing 0.5% PGA in 50 mM sodium acetate pH 5.2, incubated at 30 °C for 30 minutes, stained with 0.5 % ruthenium red solution and destained with water. The enzymatic activity appeared as an unstained spot against the red background. The wt-PehA was purified at room temperature with a low-pressure liquid chromatography system (Biologic, BioRad), as follows. The cell-free culture supernatant from 800 mL was dialyzed against 50 mM sodium acetate pH 3.5 and concentrated to 200 mL using an Amicon concentration cell (Millipore) equipped with a YM10 membrane (10 kDa cut-off). The sample was then loaded onto a cation exchange S Sepharose FF column (Amersham Biosciences) equilibrated with 50 mM sodium acetate pH 3.5 and eluted with a 200 mL linear gradient from 0.0 to 0.4 M sodium chloride. Positive fractions were pooled, dialyzed against 50 mM sodium acetate pH 5.0 and concentrated to 1 mL. The sample was loaded onto a gel filtration column (Sephacryl HR200, column 16XK, Amersham Biosciences), previously equilibrated with 50 mM sodium acetate acetate pH 5.2, 150 mM sodium chloride. Proteins were eluted at a flow rate of 0.5 mL/min and fractions of 2.5 mL were collected. The column was calibrated with an MW-GF-200 kit

(Sigma Chemical Co.) for the molecular weight estimation. Positive fractions from the previous step were pooled and dialyzed against 50 mM sodium acetate pH 3.5 and the concentration of ammonium sulphate adjusted to 1.7 M. The sample was then applied to a hydrophobic interaction Phenyl Sepharose HR column (Amersham Biosciences), previously equilibrated with 50 mM sodium acetate pH 3.5, 1.7 M ammonium sulphate. The column was washed with 40 ml of the same buffer before eluting the proteins with a 240 ml linear gradient from 1.7 to 0 M ammonium sulphate. Flow rate was 3 mL/min and fractions of 5 mL each. Protein concentration of samples taken at different purification steps was determined using the BioRad protein assay kit I, with bovine serum albumin as reference. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to Laemmli (28) using 12% acrylamide gels. To estimate the apparent molecular weight of the purified wt-PehA, the broad range molecular weight marker kit (New England Bioscience) was used. Protein bands were stained with Coomassie Brilliant Blue R250. Analytical isoelectric focusing of the purified enzymes was performed with an Ampholine PAG-plate precast polyacrylamide gel (Amersham Biosciences), in the pH interval 3-10, using the broad isoelectric point (pI) calibration kit as pI marker (Amersham Biosciences).

In order to perform the N-terminal amino acid sequence analysis, the purified wt-PehA was resolved on SDS-PAGE, blotted onto a PVDF membrane and subjected to N-terminal analysis by automated Edman degradation on a pulsed liquid sequencer (model 470A; Applied Biosystems).

2.2.2 Heterologous expression

2.2.2.1 Heterologous expression of PehA in *E. coli* periplasm (per-PehA)

The coding region of PehA, starting from the triplet GCC corresponding to Ala28, was PCR-amplified from *B. cepacia* strain ATCC 25416 genomic DNA, using the Expand High Fidelity PCR System (Roche). The genomic DNA of *B. cepacia* cells was purified as previously reported (23). The forward (F1: 5'-ATAATCCATGGCCACCTGCACGCCG-3') and the reverse (R1: 5'-TATATAAGCTTAATCTGCGCAGGGATCGT-3') primers were designed to introduce NcoI and HindIII restriction sites at the N-terminus and at the C-terminus of the protein, respectively. The reverse primer also contains a stop codon. The NcoI and HindIII digested and amplified fragment was ligated into the pET22b plasmid (Novagen) using T4 ligase (Figure 2.2).
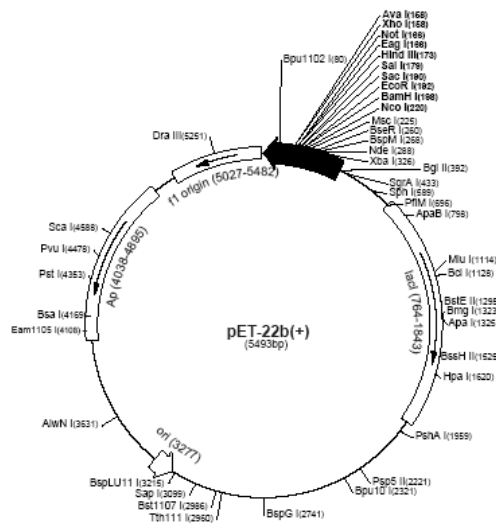
Figure 2.2. (A) The vector map of pET22b (Novagen). (B) A portion of the MSC (multiple cloning site) of pET22b containing the *pelB* signal peptide. The Figure is taken from: www.merckbiosciences.co.uk.

The *E. coli* host strain DH5α cells (Clontech) were transformed with the pET22_NcoI_HindIII_*pehA* plasmid. Transformants were selected on Luria-Bertani (LB) plates (25), containing 100 µg/mL ampicillin. The recombinant plasmid was isolated and it was shown by restriction analysis and automated DNA sequencing that it harbors the correct gene sequence.

For expression tests, *E. coli* BL21(DE3) cells (Novagen) were transformed with the pET22_NcoI_HindIII_*pehA* plasmid and plated on LB plates with 50 µg/mL ampicillin. Well grown single colonies from the selected plates were inoculated in 10 mL of the following culture media: LB medium (25), containing 100 µg/mL ampicillin for selection, or M9CA medium (25), containing 100 µg/mL ampicillin for selection. The minimal medium was supplemented with different carbon sources, either 0.4% (w/v) glycerol or 0.4% (w/v) glucose. The 10 mL starter cultures were grown over night at 37 °C and then added to 100 mL of fresh media. When the cultures reached an $OD_{600} = 1.0$, 10 mL of each culture were taken and left as non-induced control, while the remaining volumes were induced using 0.5 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG). Two cultures, induced and non-induced, for each culture media tested, were kept for additional 15 hours at 12 °C. Then the cells were harvested by centrifugation for 20 minutes at 4200 g at 4 °C. The recombinant cells were submitted to an osmotic shock in order to release the periplasmic proteins (29). Briefly, cells were washed twice with 30 mM Tris-HCl buffer (pH 8.0). The pellet was resuspended in the same buffer containing 20 % w/v sucrose and 1 mM EDTA and kept 15 minutes at room temperature with gentle shaking. The cells were recovered by centrifugation for 10 minutes at 10000 g at 4 °C and the pellet dissolved in ice-cold water containing 1 mM magnesium chloride and 1 tablet of Complete Protease Inhibitor Cocktail (Roche). The final re-suspension volume was 1/40 with respect to the initial culture volume. The extract was then centrifuged at 20000 g for 10 minutes at 4 °C. The supernatant corresponds to the periplasmic fraction. The enzymatic activity of the periplasmic extracts was tested using the pectate-agarose gel method (27), as described in Materials and Methods (Par. 2.2.1). For the large scale expression, *E. coli* BL21(DE3) cells (Novagen) harboring the pET22_NcoI_HindIII_*pehA* plasmid were inoculated in 50 mL of M9CA medium, 0.4 (w/v) glycerol as carbon source and 100 µg/mL ampicillin for selection. The 50 mL

starter culture was grown over night at 37 °C and then added to 3.5 L of fresh media. When the culture reached an OD_{600} = 1.0 abs, the protein expression was induced for 15 hours at 12 °C, using 0.5 mM IPTG. Then, cells were harvested by centrifugation for 20 minutes at 4200 g at 4 °C. The periplasmic fraction was subsequently isolated as described above.

The pH of the periplasmic fraction was lowered by the addition of 1 M sodium-acetate pH 3.5 up to a final concentration of 50 mM. The solution was spun down at 20000 g for 10 minutes at 4 °C and then applied to a SP-Sepharose 26/10 column (Amersham Biosciences) equilibrated with 50 mM acetate buffer, pH 3.5. The elution was performed by applying a 15 column volumes linear gradient of sodium chloride from 0.0 to 0.6 M in the same buffer. Polygalacturonase activity was eluted at a sodium chloride concentration of 140 mM. Protein purity and molecular weight of the per-PehA were verified by SDS-PAGE, as reported in Materials and Methods (Par. 2.2.1). Samples for non-reducing SDS-PAGE were mixed with an equal volume of 10 mM Tris-HCl buffer, pH 8.0, containing 1 mM EDTA, 2% SDS, 10% glycerol. Protein concentration was estimated from the measurement of the optical density at 280 nm, using the extinction coefficient and the molecular weight calculated from the protein sequence, using the online ProtParam server (www.expasy.ch/tools/protparam.html).

pI determination and titration curve of the per-PehA were performed using Phast System (Pharmacia) equipment. Isoelectrofocusing (3-9 and 5-8 pH range) precasted gels (Pharmacia) were used according to the manufacturer's instructions. The pI was calculated using the broad pI calibration kit (Pharmacia). Titration curve of wt-PehA was obtained as follows: after the prefocusing step, the gel was rotated clockwise 90° and the samples were then applied across the middle of the gel, perpendicular to the pH gradient. Electrophoresis perpendicular to the first dimension axis was run at 1000 V for 60 Vh. The gal was stained as reported above.

2.2.2.2 Heterologous expression of PehA in *E. coli* cytoplasm

2.2.2.2.1 Small scale analysis for protein solubility. Two constructs were prepared to check the solubility of the PehA in *E. coli* cytoplasm. In the first (referred to as

construct 1), the coding region of the mature wt-PehA was PCR-amplified from *B. cepacia* strain ATCC 25416 genomic DNA, using Expand High Fidelity PCR System (from Roche). Total DNA from *B. cepacia* cells was prepared as previously reported (Par. 2.2.2.1). The forward (F2: 5'-ATAATCATATGGCCACCTGCACGCCG-3') and reverse (R1: 5'-TATATAAGCTTAATCTGCGCAGGGATCGT-3') primers were designed to introduce NdeI and HindIII restriction sites at the N-terminus and at the C-terminus of the protein, respectively. The NdeI and HindIII digested and amplified fragment was ligated into plasmid pET22b (Novagen) using T4 ligase. The *E. coli* host strain DH5α cells (Clontech) were transformed with the pET22_NdeI_HindIII_*pehA* plasmid. Transformants were selected on LB plates with 100 μg/mL ampicillin. Recombinant plasmid was isolated and was shown by restriction analysis and automated DNA sequencing to harbour the correct gene sequence.

On the other hand, in the second construct (referred to as construct 2), the coding region of the mature wt-PehA was PCR-amplified as previously described (Par 2.2.2.1). The NcoI and HindIII digested and amplified fragment was ligated into the pETM20 plasmid (EMBL) (Figure 2.3), using ligase T4 .
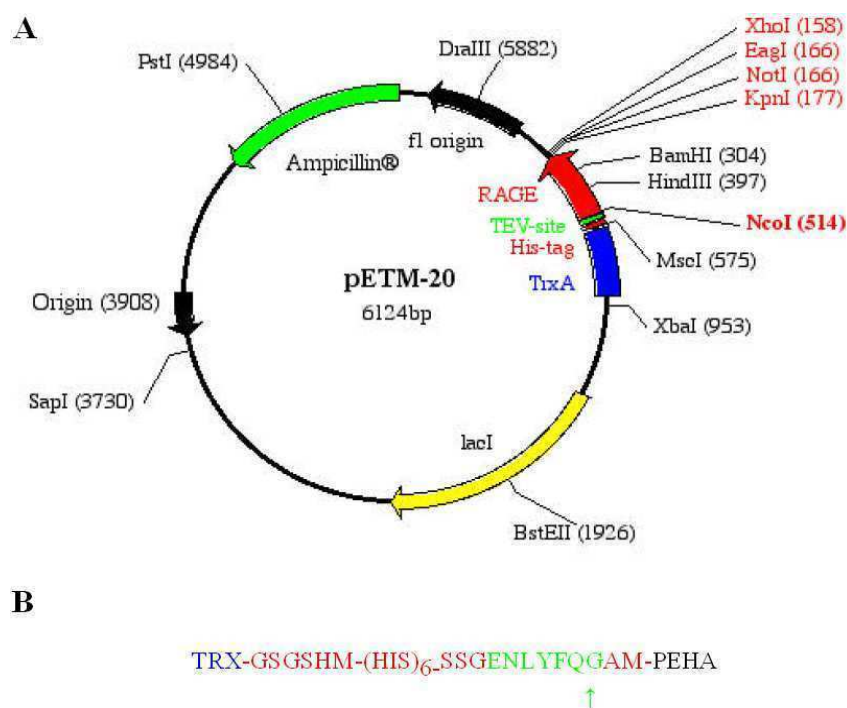
Figure 2.3. (A) The vector map of pETM20 (EMBL). (B) A schematic view of TrxA-pehA fusion ptotein. The green arrow indicates the TEV cleavage site. The Figure 3A is taken from url: www.embl.de.

This vector provides a thioredoxin A (*trxA*) gene followed by a 6×His tag and a tobacco etch virus (TEV) protease cleavage site as an N-terminal fusion partner. The *E. coli* host strain DH5α cells (Clontech) were transformed with pETM20-*trxA-pehA* plasmid. Transformants were selected on LB plates with 100 μg/mL ampicillin. Recombinant plasmid was isolated and was shown by restriction analysis and automated DNA sequencing to harbour the correct gene sequence.

For protein solubility studies, the *E. coli* host strain OrigamiB (DE3) cells (Novagen) were transformed with pET22_NdeI_HindIII_*pehA* and pETM20-*trxA-pehA* plasmids, respectively, and plated on LB plates with 50 μg/mL ampicillin, 15 μg/mL kanamycin and 12.5 μg/mL tetracyclin. Well grown single colony from the selected plates was inoculated in 10 mL of LB medium containing 50 μg/mL ampicillin, 15 μg/mL kanamycin and 12.5 μg/mL tetracyclin and incubated overnight at 37 °C in a rotary shaker. The starter culture was added to 100 mL of LB and grown in a shaker at 37 °C, until an $OD_{600} =$ 0.7 abs. Then, 10 mL of culture were taken and left as non-induced control, while the remaining volume was induced using 0.5 mM IPTG. Both cultures, induced and non-induced, of both plasmids, were kept for further 4 hours at 20 °C. The cell pellets were collected by centrifugation at 4200 g for 20 minutes and resuspended in 4.0 mL of 50 mM sodium acetate buffer (pH 5.0) containing 150 mM sodium chloride. The resuspensions were then sonicated 5 times for 10s each at maximum power, using a Soniprep 150 Sonicator (Sanyo) and then centrifuged at 16000 g for 20 minutes at 4 °C. The supernatant was the soluble fraction and its enzymatic activity was estimated in a qualitative way with the pectate-agarose method, as previously described (Par 2.2.1). The insoluble fraction was resuspended in 4.0 mL of 30 mM Tris-HCl pH 8.0. Soluble and insoluble fractions were analysed by SDS-PAGE as already described (Par. 2.2.1). The non-induced cultures were treated in a similar manner, using 0.4 mL as resuspension volume.

2.2.2.2.2 Large scale expression and refolding of TrxA-PehA fusion protein. The *E.coli* host strain OrigamiB (DE3) cells (Novagen) were transformed with the pETM20-*trxA-pehA* plasmid and inoculated in 20 mL of LB medium containing 50 μg/mL ampicillin, 15 μg/mL kanamycin and 12.5 μg/mL tetracyclin. The starter culture was grown over

night at 37 °C and then added to 1.0 L of fresh medium. When the culture reached an OD_{600} = 0.7, it was induced using 0.5 mM IPTG and kept for 4 hours at 20 °C. Then, cells were collected by centrifugation at 4200 g for 35 minutes at 4 °C and frozen. The bacterial pellets were resuspended in 20 mM sodium phosphate buffer (pH 7.8), 0.5 M sodium chloride, 10% glycerol, 1% Tween 20, 1 mM β-mercaptoethanol (β-ME) and 1 tablet of Complete Protease Inhibitor Cocktail (Roche). Lysozyme at final concentration of 1 mg/mL was added and the solution was kept on ice for 1 h. The lysate was sonicated 3 times for 30s each at max power, using a Soniprep 150 Sonicator (Sanyo), and then centrifuged at 20000 g for 35 minutes at 4°C. In order to isolate the inclusion bodies (IBs), the cell pellet was resuspended in 50 mM Tris-HCl buffer (pH 8.0), 0.1 M sodium chloride, 0.5 % v/v Triton X-100, 5 mM β-ME and washed twice. The resuspension was then washed three times with 50 mM Tris-HCl buffer (pH 8.0), 0.1 M sodium chloride and 5 mM β-ME. IBs were harvested by centrifugation at 25000g for 15 minutes at 4 °C and dissolved in 5 mL of 20 mM sodium phosphate buffer (pH 7.8), 0.5 M sodium chloride, 5 mM β-ME, 6.0 M guanidine hydrochloride (GdnHCl) and 20 mM imidazole (buffer RefA), by stirring overnight at room temperature. The homogenate was centrifuged at 25000 g for 15 minutes at 4 °C to remove insoluble materials and filtered onto a 0.22 mm filter. The purity of TrxA-PehA recovered from the solubilized IBs was checked with SDS-PAGE, as already reported (Par. 2.2.1). Protein concentration was estimated from the measurement of optical density at 280 nm using the extinction coefficient and the molecular weight calculated from the TrxA-PehA protein sequence with the ProtParam tool (www.expasy.ch/tools/protparam.htmL). Purified IBs were store in 20 mg (total protein) aliquots at -20 °C.

The solubilized TrxA-PehA was refolded using an on-column refolding procedure, by metal chelate affinity chromatography. An Hi-Trap Chelating (Amersham Biosciences) column with a bed volume of 1 mL, previously charged by flowing 0.5 mL of 0.1 M nickel sulphate, was equilibrated with 10 column volumes of buffer RefA. The solubilized TrxA-pehA at the final concentration of 2.0 mg/mL was loaded and the column was washed with 10 mL of buffer RefA. To achieve protein refolding, a linear decreasing gradient of GdnHCl was applied: the concentration of GdnHCl was

reduced from 6.0 M to 0.0 M in buffer RefA in a total volume of 40 column volumes, at a flow rate of 0.2 mL/min. In order to eliminate the interference in the UV absorbance of imidazole present in the refolding buffer, the column was washed with 10 column volumes of 20 mM sodium phosphate buffer pH 7.8, 0.5 M sodium chloride, 5 mM β-ME. Then, refolded TrxA-PehA was eluted by pH drop, using as elution buffer 20 mM sodium phosphate (pH 4.0), 0.5 M sodium chloride, 1 mM β-ME. The enzymatic activity of the refolded TrxA-PehA was estimated using the Bernfeld method, as already described (Par. 2.2.1). The concentration of purified and refolded TrxA-PehA was estimated from the measurement of optical density at 280 nm, as previously described.

The fusion protein was cleaved at room temperature with an excess of TEV protease (Invitrogen), during overnight dialysis against 50 mM acetate buffer, pH 5.4, containing 1 mM EDTA and 1 mM DTT. 1 unit of TEV protease was used to cleave 2 μg of fusion protein. The cleaved protein (ref-PehA) was separated from the uncleaved TrxA-PehA, its fusion partner, and the TEV protease in a further chromatographic step. The sample was applied to a MonoS (Amersham Biosciences) column with a bed volume of 1 mL, previously equilibrated with 50 mM acetate buffer (pH 3.5). The ref-PehA was eluted with a 15 column volumes linear gradient of NaCl from 0.0 to 0.6 M in 50 mM acetate buffer (pH 3.5), at a salt concentration of 140 mM. Protein purity and molecular weight of the ref-PehA were checked by SDS-PAGE. The protein concentration of ref-PehA was estimated from the measurement of optical density at 280 nm, as previously described.

2.2.3 Characterization of the recombinant PehA

2.2.3.1 Characterization of per-PehA

2.2.3.1.1 Biochemical characterization of per-PehA. The enzymatic activity of the per-PehA was estimated using the Bernfeld method (26), as already described (Par. 2.2.1). The optimal pH and temperature values were determined in the pH interval 2.0-8.0 and in the temperature range between 15 and 70 °C. To estimate pH stability, samples were incubated in buffers of different pH values at 4 °C for 24 hours. To define thermal stability, protein samples were incubated at different temperatures in acetate buffer pH 3.5, for 2 hours. The residual activity was detected according to the Bernfeld method. In order to determine the Michaelis-Menten parameters of the per-PehA, the highly sensitive neocuproine assay was used for the estimation of the number of the released reducing ends sugars (30). Briefly, protein samples were assayed at 37 °C in a total reaction volume of 0.2 mL of 20 mM sodium acetate pH 3.5, using PGA as substrate at concentrations ranging from 0.01 to 3.0 mg/mL. At different incubation times the reaction was stopped by the addition of 0.4 mL of mix A (3.8 M sodium carbonate, 200 mM glycine, 2 mM cupric sulphate) and 0.4 mL of mix B (5.7 mM neocuproine hydrochloride in water). Samples were kept for 12 minutes at 100 °C and then their absorbance at 450 nm was measured. Determination of reducing ends was done using a reference curve, obtained with glucose as standard. Enzymatic activities were assayed in triplicate and expressed in unit, as already reported for endoPGs (31). One enzymatic unit was defined as the amount of enzyme that catalyzes the release of one micromole of reducing end sugars per minute.

2.2.3.1.2 Biophysical characterization of per-PehA. The molecular mass of the per-PehA was analyzed by electrospray mass spectrometry (ESI-MS), using an API 150EX spectrometer (Applied Biosystem), and compared with the molecular mass of wt-PehA. Analytical size-exclusion chromatography was performed on a Superdex G75 HR 10/30, equilibrated with 50 mM sodium acetate pH 3.5, 150 mM sodium chloride. The column was previously calibrated using a broad molecular weight gel filtration kit (Amersham Pharmacia). Protein sample in 50 mM sodium acetate pH 3.5, 150 mM

sodium chloride was loaded onto the column at a concentration of 0.1 mg/mL and eluted at a flow rate of 0.5 mL/min.

Dynamic Light Scattering (DLS) measurements were performed with a 5.6 mg/mL per-PehA solution in distilled water containing 270 μM hydrochloric acid in order to keep the pH acidic, at room temperature on a DynaPro MS/X instrument (Proteins Solutions). Sample was filtered through a 0.02 μm Anodisc membrane using the NanoFilter kit (Wyatt Technology) to remove dust particles and precipitate and then injected in a 12 μL quartz microCUVETTE (Wyatt Technology). Data collection and deconvolution were performed using the DYNAMICS 6.2.05 software (Proteins Solutions).

Circular dichroism (CD) studies were performed on a Jasco J-810 spectropolarimeter (Jasco, Japan). CD spectra were recorded from 250 nm to 190 nm in a 1.0 mm path-length quartz cuvette at a protein concentration of 5 μM (~ 0.2 mg/mL). Sample was dissolved in distilled water containing 270 μM hydrochloric acid in order to keep the pH acidic. Spectra were measured with a scanning speed of 10 nm/min, setting a bandwidth of 1 nm. Five scans were averaged for the buffer and protein solutions, respectively. Buffer spectra were subtracted from the protein spectra. The results are expressed as mean residue molar ellipticity $(\Theta)_{MRW} = \Theta/(c \times d \times N)$, where Θ, is the observed ellipticity; c, is the protein concentration; d, is the optical path length; N, are the number of residues. Analysis of the CD data for secondary structure determination was carried out, using algorithms (32-34) that are available online at Dichroweb server (url: http://www.cryst.bbk.ac.uk/cdweb) (35).

2.2.3.2 Preliminary characterization of ref-PehA

The enzymatic activity of the ref-PehA was estimated using the Bernfeld method, as described in Materials and Methods (Par 2.2.1). The molecular mass of ref-PehA was estimated by mass spectrometry. Mass spectrometry with electrospray ionization and time-of-flight analyzer (ESI-TOF) was performed on a microTOF Focus system (Bruker Daltonics), equipped with a nitrogen generator N2LCMS1(Claind). The samples were diluted five times with a mixture of methanol/water/formic acid (49/49/2) and injected with a syringe pump in the mass spectrometer. The injection flow rate was

180 mL/h; the capillary voltage was 4500 V and the end plate offset was 500V (positive mode); the dry temperature was 200°C, the dry gas flow was 4 L/min and the nebulizer pressure was 0.4 bar.

Circular dichroism studies were performed as described in Materials and Methods (Par 2.2.3.1.2). CD spectra were recorded using a 0.2 mm path-length quartz cuvette at protein concentration of 5 µM (~ 0.2 mg/mL). The superimposition of CD spectra was performed using the Spectra Manager software (Jasco).

2.3 Results

2.3.1 Isolation, purification and characterization of PehA from *B. cepacia* culture (wt-PehA)

The polygalacturonase activity in the *B. cepacia* culture supernatant could only be detected when the bacterium was grown in the presence of PGA, thus suggesting that the substrate plays a major role in inducing the enzyme production. The use of a low pH buffer (pH 3.5) was crucial for the successful purification of the PehA. At this acidic pH the enzyme was able to interact with both the cation exchanger and the hydrophobic interaction chromatography media. The enzyme was eluted using a sodium chloride and an ammonium sulphate gradient, at 170 mM sodium chloride and 0.8 M ammonium sulphate concentrations, respectively (data not shown). Protein purification steps are summarized in Table 2.1. The yield of wt-PehA protein isolated from the culture supernatant was 0.2 mg/L.

The molecular weight of the purified enzyme was determined by SDS-PAGE and size exclusion chromatography. The SDS-PAGE showed a single sharp band corresponding to an estimated molecular mass of 47 kDa (Figure 2.1A). The molecular mass of the wt-PehA estimated by gel filtration chromatography was approximately 40 kDa, thus suggesting that the enzyme is monomeric in solution. The pI of the purified wt-PehA was determined under native conditions and found to be approximately 7.8 (Figure 2.1B).

When subjected to N-terminal amino acid sequence determination by Edman degradation, the wt-PehA showed the following sequence: ATCTPQWSSS. This information suggested that the enzyme was encoded by the genetic determinant designed *pehA* (AAB46984) sequenced and previously reported by Gonzales *et al.*, (22). The signal peptide and the corresponding cleavage site of the secreted protein were deduced and it matched with the predicted one obtained using the SignalP 3.0 Server program (http://cbs.dtu.dk/services/SignalP/). The signal peptide was found to be a 27-amino acid N-terminal extension of the mature protein with the following sequence: MKGKSSTRLVLRLSTLAALAVQASAQA.
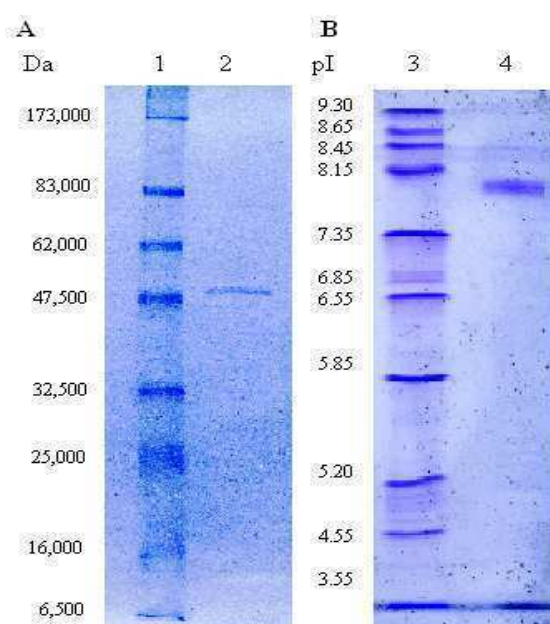
Figure 2.4. SDS-PAGE (**A**) and analytical isoelectric focusing (**B**) of the wt-PehA purified from *B. cepacia* culture supernatant. Lane 1: molecular mass standard; lane 2, 2.5 µg of enzyme; lane 3: pI markers, broad range; lane 4: 2.5 µg of enzyme.

Table 2.1. Summary of the protein purification steps of wt-PehA from 800 mL of *B. cepacia* culture.

Purification step [a]	Total protein (mg)	Total activity (U)[b]	Specific activity (U/mg)	Purification factor	Yield (% of activity)
Supernatant	17.60	63.8	3.6	1	100
Concentration and dialysis	1.60	14.4	9.0	2.5	22.5
S Sepharose FF	0.65	7.5	11.5	3.2	11.7
Sephacryl HR200	0.22	3.5	16.0	4.4	5.5
Phenyl Sepharose HP	0.13	2.8	21.1	5.8	4.3

[a] See Materials and Methods for details
[b] Measured with PGA as the substrate

2.3.2 Heterologous expression

2.3.2.1 Heterologous expression of PehA in *E. coli* periplasm (per-PehA)

The *pehA* gene is localized on a plasmid of approximately 200 kb (22), that co-purified with genomic DNA of *B. cepacia* cells. The cloning of the coding region without the signal peptide was performed by PCR. Due to the use of the NcoI restriction site in the design of the forward primer, a methionine in position 1 has been added. To target the expression of PehA in the periplasmic space of *E. coli*, the signal sequence of an endogenous pectate lyase B (*pelB)* has been inserted at the N-terminal end using the commercial vector pET22b (Figure 2.2). A broad screening of the culture media conditions has been exploited to more efficiently explore the periplasmic expression of PehA. A qualitative estimation of the enzymatic activity in the periplasmic extract isolated from cultures grown on different media was performed by using the pectate-agarose gel method (27) (Figure 2.5).
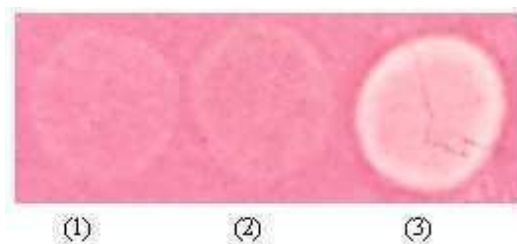
Figure 2.5. Enzymatic activities of the periplasmatic extracts isolated from recombinant *E. coli* cells, grown on LB (1), on minimal medium supplemented with glucose (2) or glycerol (3) as carbon source, were estimated using the pectate-agarose gel method (27).

Enzymatic activity was detected in the periplasmic extract of the *E. coli* cells, grown on minimal medium, using glycerol as carbon source. Interestingly, in the periplasmic fractions isolated from the cultures grown either on LB broth or on minimal medium supplemented with glucose as carbon source, polygalacturonase activity was not detected.

The periplasmic extract (Figure 2.6, lane 1) was then submitted to pH precipitation. Among the proteins soluble at pH 3.5, a major band of approximately 45 kDa, corresponding to the expected molecular weight of PehA, was identified (Figure 2.6, lane 2). After the pH precipitation, the sample was submitted to a cation-exchange

chromatography, that was sufficient to obtain the pure protein, as shown by the single band observed on SDS-PAGE (Figure 2.6, lane 5).

The use of low pH buffer resulted to be critical in the protein purification protocols of both the wild-type and the recombinant enzymes. The yield of the per-PehA was about 1.0 mg/L of culture, approximately five times more than the yield of the protein purified from *B. cepacia* culture.
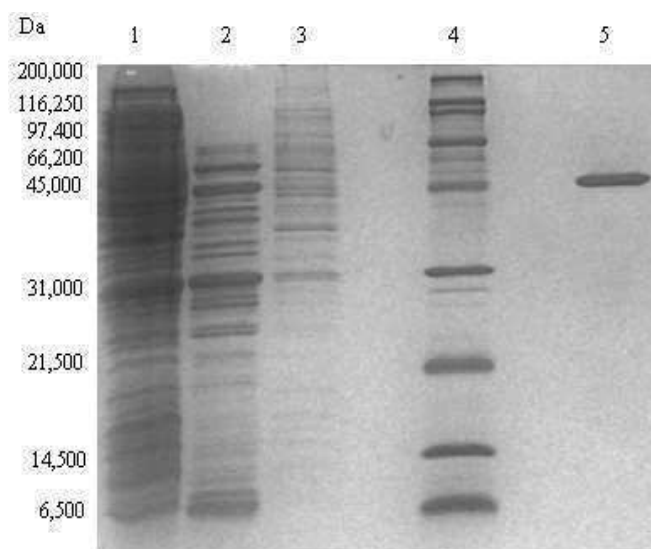
Figure 2.6. The purification of per-PehA monitored by SDS-PAGE. Lane 1: periplasmic extract; lane 2: soluble fraction after pH precipitation; lane 3: insoluble fraction after pH precipitation; lane 4: molecular mass standard; lane 5: purified per-PehA.

The presence of six cysteine residues in the primary sequence of PehA leads us to hypothesize that they might be involved in the formation of disulphide bonds. The occurrence of disulphide bonds was investigated comparing the position of the protein band on denaturing acrylamide gels, run under reducing and non-reducing conditions. In reducing condition a band shift was observed, thus supporting the presence of disulphide bonds in the active form of the enzyme (Figure 2.7).
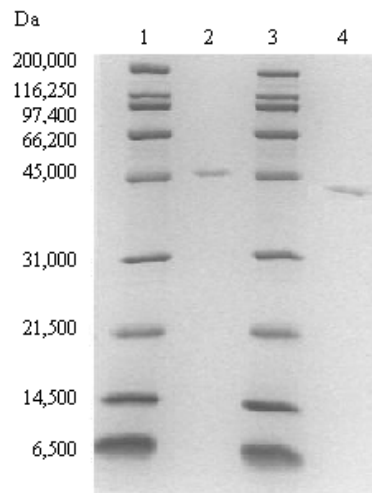
Figure 2.7. SDS-PAGE of per-PehA run under reducing and non-reducing conditions. Lanes 1 and 3: molecular mass standard; lane 2: reducing conditions; lane 4: non-reducing conditions.

The pI of the per-PehA was determined under native conditions and found to be approximately 8.0 (data not shown), in agreement with the value determined for the wt-PehA. In order to investigate the charge properties of the enzyme, a titration curve was performed (Figure 2.8). The protein does not move across the IEF gel (5.0-8.0), with respect to the loading position, indicating that it acquires a net charge only at pH values lower than 5.0 or higher that 8.0. This behavior may explain the difference between the experimentally determined isoelectric point near 8.0 and the value predicted by the online ProtParam server (www.expasy.ch/tools/protparam.html), that was near 5.0.
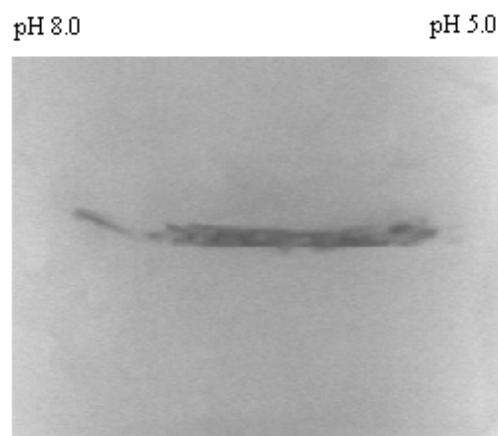
Figure 2.8. The titration curve of the per-PehA.

2.3.2.2 Heterologous expression of PehA in *E. coli* cytoplasm

2.3.2.2.1 Small scale analysis for protein solubility. Two constructs were prepared to achieve the cytoplasmic expression of PehA. As first attempt, the *pehA* gene without the proper signal peptide was cloned in the same vector used for the periplasmic expression, pET22b, but a different restriction enzyme was chosen. The use of the NdeI restriction enzyme allowed the elimination of the *pelB* signal peptide, costitutively present in the multiple cloning site of the plasmid pET22b (Figure 2.2B). Then, another plasmid was chosen for the cytoplasmic expression, the plasmid pETM20 (36). This vector provides a thioredoxin A (trxA) gene followed by a 6×His tag and a tobacco etch virus (TEV) protease cleavage site as an N-terminal fusion partner (Figure 2.3).

The presence of disulphide bonds in per-PehA led us to choice a different host strain to achieve the expression of the enzyme in its active form. Therefore, construct 1 and construct 2 were introduced in the *E. coli* OrigamiB (DE3) host strain. In this strain mutations in thioredoxin and glutaredoxin pathways interfere with the cytoplasmic reduction pathway, thus facilitating the formation of disulphide bonds in proteins (37, 38). The enzymatic activity of the soluble fractions derived from the host cells transformed with construct 1 and construct 2 was estimated in a qualitative way, using the pectate-agarose gel method, and was detected in soluble fractions of both cell lysates (data not shown). The expression level of the recombinant protein produced without and with N-terminal fusion partner was analyzed by SDS-PAGE (Figure 2.9). Concerning the expression of wt-PehA without tags, a thin band of approximately 45 kDa, corresponding to the expected molecular weight of PehA (Figure 2.9A, lane 2), was observed in the soluble fraction of the induced cell lysate. Instead, concerning the expression of the fusion protein, a band of 58 kDa, corresponding to the expected molecular weight of TrxA-PehA, was found mainly in the insoluble fraction of the induced cell lysates (Figure 2.9B, lane 4). The SDS-PAGE analysis revealed also that a small amount of fusion protein was present in the soluble fraction of the induced cell lysate (Figure 2.9B, lane 2).
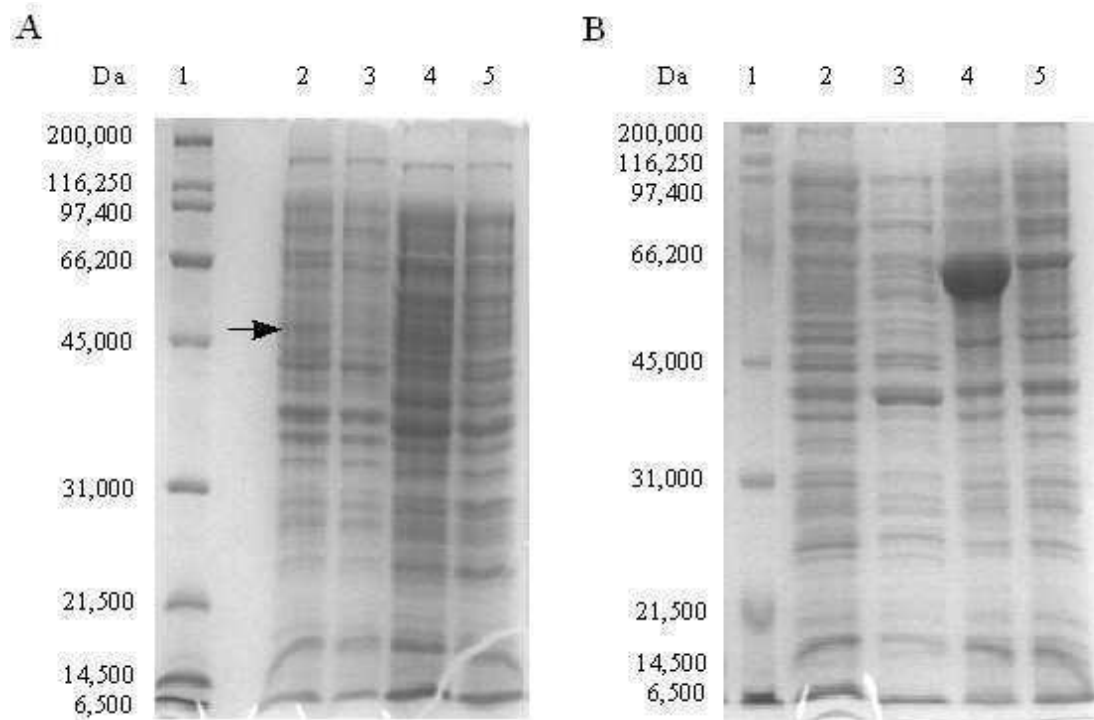
Figure 2.9. SDS-PAGE analysis of PehA produced in *E. coli* OrigamiB (DE3), harbouring construct 1 (A) and construct 2 (B). (A) Lane 1: molecular mass standard; lane 2: soluble fraction of cell lysate, induced with 0.5 mM IPTG; lane 3: soluble fraction of cell lysate, not induced; lane 4: insoluble fraction of cell lysate, induced with 0.5 mM IPTG; lane 5: insoluble fraction of cell lysate, not induced. (B) Lane 1: molecular mass standard; lane 2: soluble fraction of cell lysate, induced with 0.5 mM IPTG; lane 3: soluble fraction of cell lysate, not induced; lane 4: insoluble fraction of cell lysate, induced with 0.5 mM IPTG; lane 5: insoluble fraction of cell lysate, not induced.

2.3.2.2.2 Large scale expression and refolding of TrxA-PehA fusion protein. The *E. coli* OrigamiB (DE3) host strain cells containing the pETM20-*trxA-pehA* plasmid were grown at large scale, in order to produce a significant amount of fusion protein as insoluble aggregates. The IBs were treated as described in Materials and Methods (Par 2.2.2.2.2). As shown by SDS-PAGE analysis, the purity of TrxA-PehA in the isolated IBs was of 80% (Figure 2.10). The protein content of the solubilized IBs derived from 1.0 L of culture was found to be of approximately 80 mg.
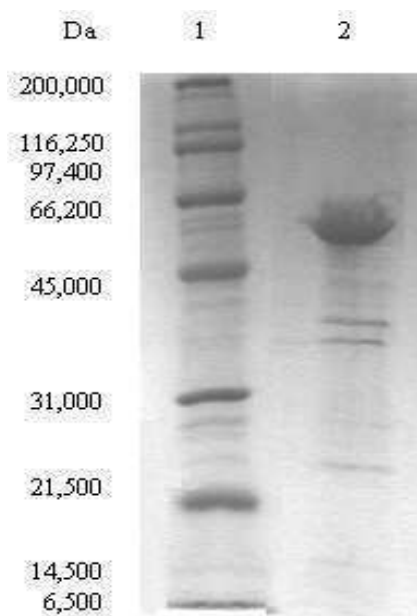
Figure 2.10. SDS-PAGE analysis of the purified IBs. Lane 1: molecular mass standard; lane 2: IBs, at a concentration of 1.8 mg/mL

Aliquots of purified and solubilized IBs were applied to a metal chelating affinity column. The on-column refolding process consists in a linear gradient from 6.0 to 0.0 M GdnHCl in 40 column volumes. Two different elution protocols were tested, including the elution with a stepwise increase of imidazole concentration (data not shown) and the elution driven by a pH drop. In the former approach, the pH of the elution buffer was 7.8, far from the pH range in which enzymatic activity of per-PehA can be detected, as reported below. The refolded protein eluted at pH 7.8 was submitted to an extensive dyalysis against acetate buffer, pH 3.5, but the enzymatic activity was not recovered. Thus, the second approach was chosen, using a low pH phosphate buffer (pH 4.0) as elution buffer. The peak fractions were collected and subjected to the SDS-PAGE analysis, that revealed the presence of a large band of 58 KDa (Figure 2.11).
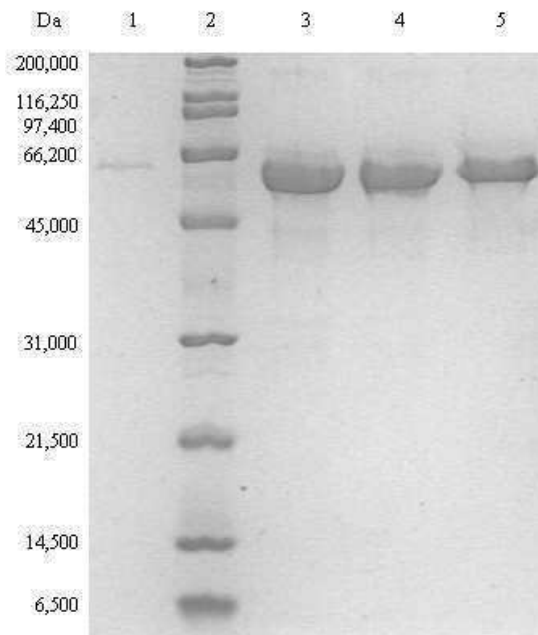
Figure 2.11. SDS-PAGE analysis of the fractions eluted by the affinity chromatographic step of TrxA-PehA. Lane 1: flowthrough; lane 2: molecular mass standard; lane 3, 4, 5: eluted fractions.

The enzymatic activity of the fusion protein could be detected only in a qualitative manner, using the pectate-agarose gel method. The yield of the refolding was calculated as percentage of the mg of the TrxA-PehA eluted in active form, with respect to the mg of proteins bound to the affinity column and was found to be of approximately 20%.

The refolded TrxA-PehA was incubated with TEV protease, in order to remove the fusion partner. As shown in the vector map (Figure 2.3B), the pETM20 vector contains a TEV cleavage site located between the histidine tag and the sequence coding for the recombinant protein. The reaction was performed at pH 5.4, even if the optimum pH of TEV protease is 8.0. A low pH buffer was chosen to preserve the enzymatic activity of the ref-PehA. As a consequence, the efficiency of the TEV cleavage was very low. Nevertheless, a small amount of TrxA-PehA was cleaved and the ref-PehA could be isolated from the uncleaved form and the other components, using a cation exchange chromatography (Figure 2.12). The yield of the cleavage was expressed as percentage of the mg of cleaved and purified ref-PehA with respect to the mg of fusion protein submitted to TEV proteolysis and was found to be of approximately 2 %.
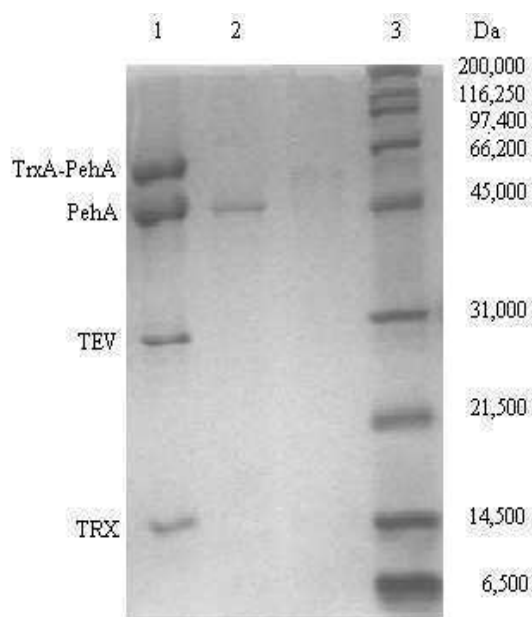
Figure 2.12. SDS-PAGE analysis of the cleavage of TrxA-PehA, after incubation with the TEV protease. Lane 1: TEV cleavage mixture, after 24 hours of incubation at room temeprature; lane 2: the purified ref-PehA, after the cation exchange chromatography; lane 3: molecular mass standard.

The final yield of ref-PehA, after refolding and TEV cleavage, was approximately of 0.3 mg/L of culture, three times less than the yield of the enzyme purified from the periplasm of the host cells.

2.3.3 Characterization of the recombinant PehA

2.3.3.1 Characterization of per-PehA

2.3.3.1.1 Biochemical studies on per-PehA. The specific activity of the purified recombinant enzyme was estimated by the Bernfeld method and was found to be of 19.5 U/mg, comparable to the enzymatic activity of wt-PehA, namely 21 U/mg (Table 1). The enzyme exhibited an optimal activity at pH 3.5 (Figure 2.13A) and in the temperature interval 40-55 °C (Figure 2.13C). Monitoring the stability of per-PehA in

the pH interval 2.0-8.0, a drop in the enzymatic activity at pH 6.0 was observed (Figure 2.13B).
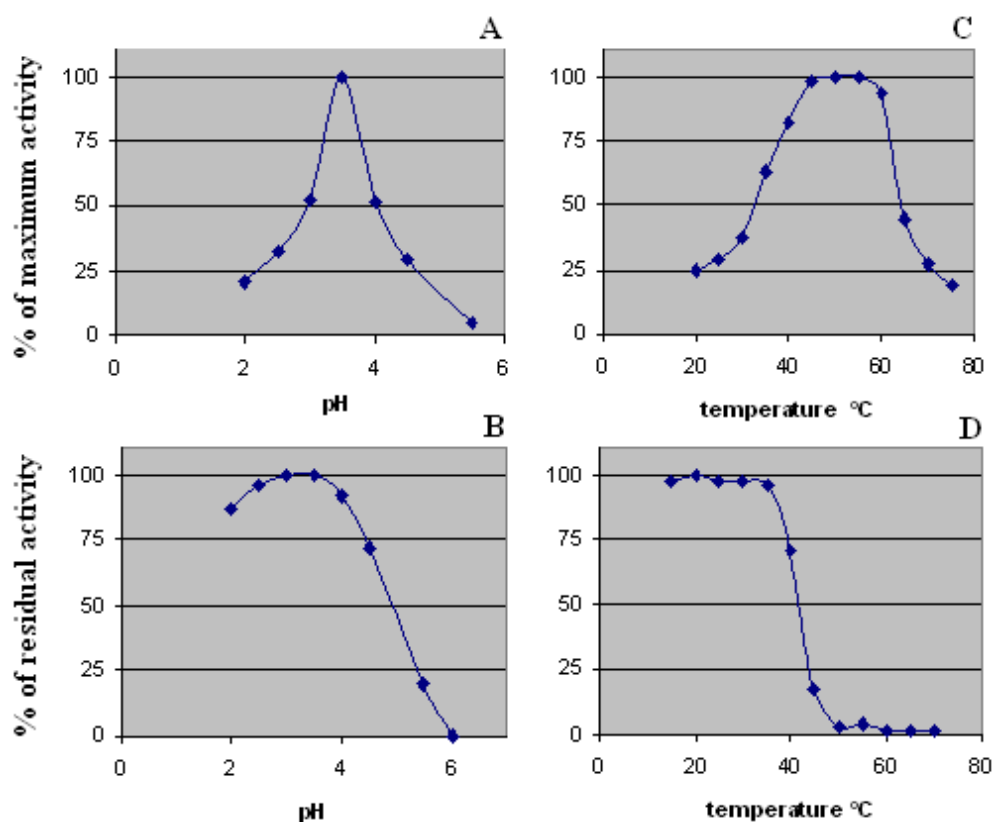
Figure 2.13. Enzymatic activity and stability. Panel A, determination of the optimal pH. Panel B: Enzymatic stability with respect to pH. Panel C, determination of the optimal temperature. Panel D, stability with respect to temperature. The enzymatic activity was tested in the pH interval 2.0-8.0 and in the temperature range between 15 and 70 °C, using the Bernfeld method (26).

Concerning the effect of the temperature on the enzymatic activity, per-PehA exhibits a relatively flat activity profile in a broad temperature interval, displaying an optimal activity in the temperature interval 40-55 °C (Figure 2.13C). As illustrated in Figure 2.13D, the enzyme resulted quite thermolabile, being completely inactivated if kept at 50 °C for 2 hours.

Michaelis-Menten parameters were defined measuring the initial reaction rates using PGA as substrate in the concentration range of 0.1-3.0 mg/mL. To determine K_M and V_{MAX} values, the Michaelis-Menten curve was linearized using Lineweaver-Burk and Hanes plots. The estimated values of K_M and V_{MAX} were of 0.487 mg/mL and of 240 U/mg, respectively. It is worthy of note that the K_M is usually given in molar concentration. In this case, since the molecular weight of the commercial PGA has not

been provided by the manufacturer nor it has been estimated by us, it has been chosen to present the concentration of the polymeric substrate as mg/mL. Therefore this unit represents also the unit of the K_M value.

2.3.3.1.2 Biophysical studies on per-PehA. The molecular mass of per-PehA was determined by ESI mass spectrometry and was found to be in agreement with the molecular mass of the mature enzyme, considering the additional methionine at the N-terminus. The molecular mass of the per-PehA was also checked by analytical gel filtration chromatography. The elution volume of the per-PehA was of 10.2 mL, corresponding to an estimated molecular weight of 45 kDa. This finding suggests that the recombinant protein in solution is a monomer, as already observed for wt-PehA.

In order to verify the homogeneity of the protein solution in view of subsequent crystallization trials, DLS measurements were performed. The low degree of polydispersity estimated for per-PehA protein solution, 15.9%, indicates that the protein preparation is suitable for crystallization studies. The DLS measurement allowed also the estimation of the hydrodynamic radius of per-PehA that was found to be of 2.9 nm.

In order to verify that if per-PehA is a well folded enzyme, CD spectra were recorded (Figure 2.14). A common feature of the endoPGs is the conservation of their overall three dimensional structure, referred to as β-helix fold (39). The β-helix motif consists of parallel β-strands joined together by highly variable loops. In addition, a conserved helical coil is present at the N-terminal end. The far-UV CD spectra of β-helix proteins show a broad minimum at 216 nm, characteristic of the β-sheets secondary structure elements (40).

Three different algorithms, Selcon, CDSSTR, and Contin (32-34), were used for the determination of the secondary structure. The results are summarized in Table 2.2. The main secondary structure estimated by all algorithms is the β-strand, as expected. The determination of unordered region could be the result of limited accuracy, since the β-helix fold is not very well represented in the reference sets used for the secondary structure analysis.
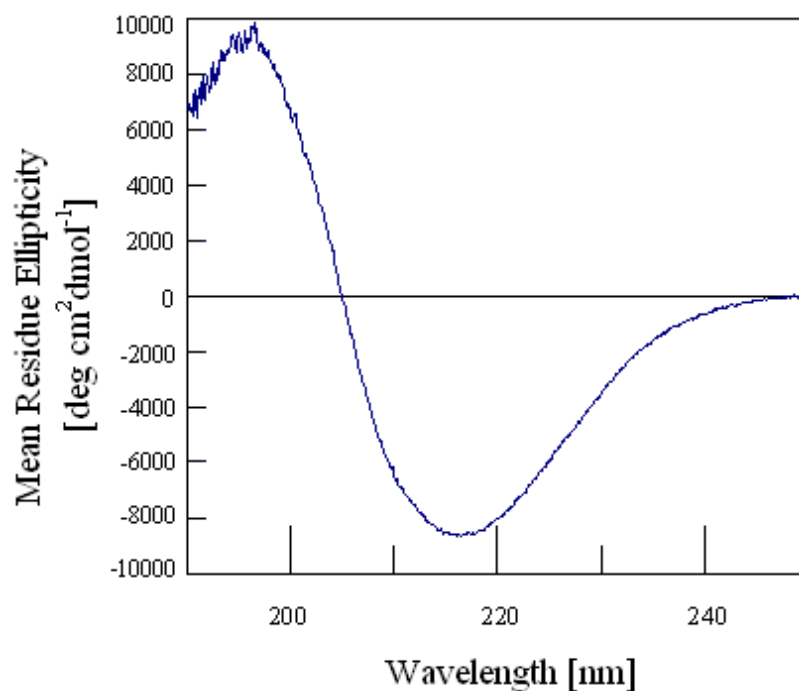
Figure 2.14. CD spectrum of per-PehA. The CD spectra were collected in the wavelength range of 190-250 nm at room temperature. The concentration of protein was 0.2 mg/mL, in aqueous solution pH 3.5. The CD data are presented in terms of mean residue ellipticity, as described in Par. 2.2.3.1.2.

Table 2.2. Secondary structure determination of per-PehA, obtained by the quantitative analysis of CD spectra, carried out by selected methods (32-34).

Program	Helix	Strand	Turns	Unordered
Selcon	0.131	0.460	0.202	0.216
Contin	0.129	0.360	0.223	0.288
CDSSTR	0.080	0.320	0.220	0.280

2.3.3.2 Preliminary characterization of ref-PehA. The enzymatic activity of ref-PehA was quantitatively estimated using the Bernfeld method and was found to be comparable to that of the wt-PehA. The molecular mass of the ref-PehA was estimated by ESI-TOF-MS and was found to be of 43464.54 Da. This value corresponds to the molecular mass of the mature enzyme, with additional amino acids that have been included by the cloning, as shown in Figure 2.3B. The mass spectrum is reported in Figure 2.15. The results of the mass spectrum analysis are shown in Table 2.3.
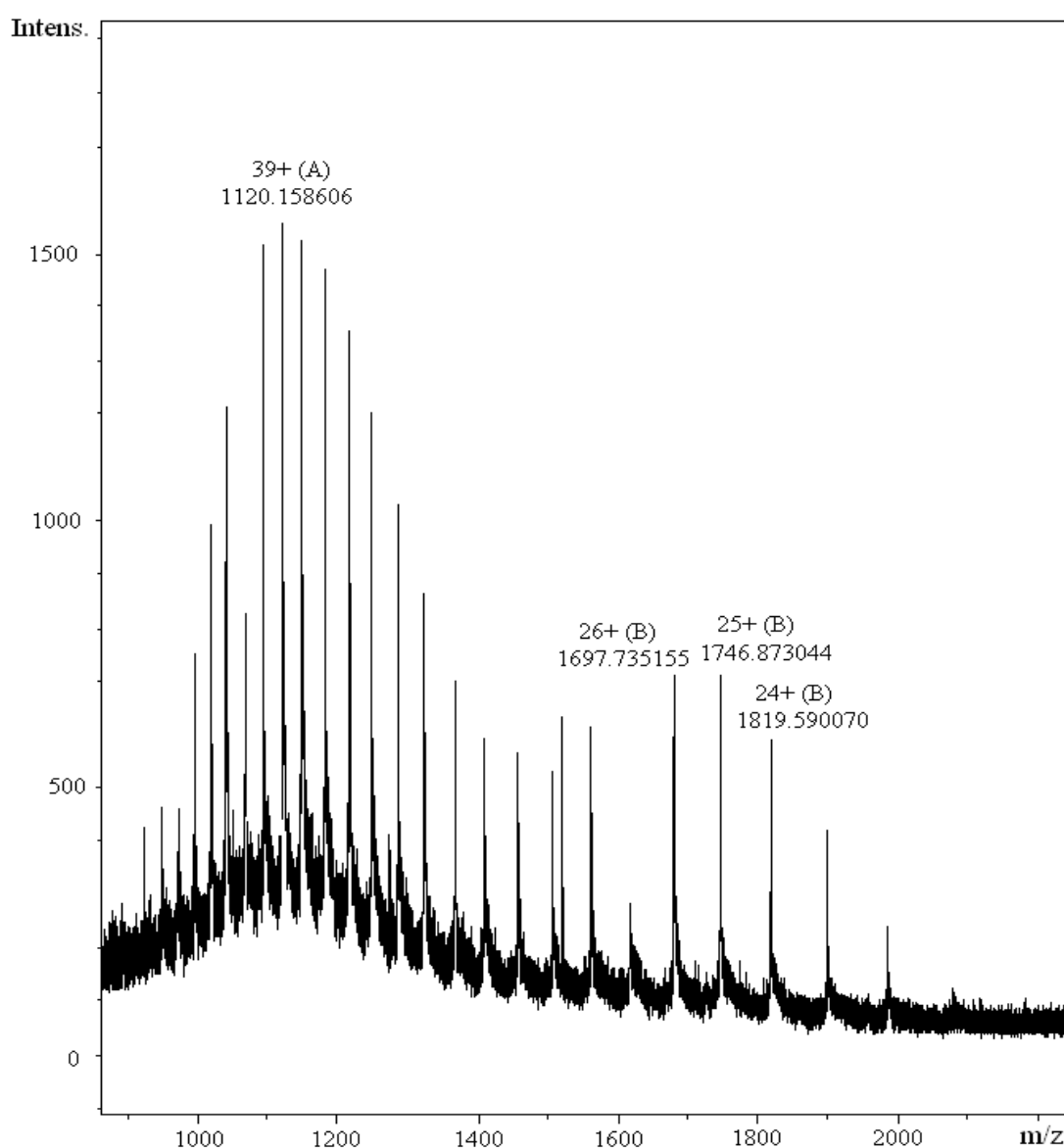
Figure 2.15. The ESI-TOF mass spectrum of ref-PehA.

Table 2.3. Ion signals in the ESI-TOF mass spectrum of ref-PehA.

Component A

Actual Peak	Charge	Isotopic mass ([M + H]+)	Predicted peak
1016.064905	43+	43640.285496	1016.053534
1040.208137	42+	43639.236155	1040.197399
1065.571834	41+	43639.939326	1065.567958
1092.180099	40+	43639.698586	1092.156878
1120.158606	39+	43639.681184	1120.135059
1149.582265	38+	43638.623460	1149.585776
1180.635182	37+	43639.001090	1180.628423
1213.415646	36+	43639.465228	1213.395662
1248.039885	35+	43638.901616	1248.035314
1284.703210	34+	43638.419181	1284.712593

Component B

Actual Peak	Charge	Isotopic mass ([M + H]+)	Predicted peak
1617.531056	27+	43638.907921	1617.521305
1679.735155	26+	43639.697352	1679.694901
1746.873044	25+	43639.425631	1746.842384
1819.590070	24+	43638.779660	1819.585490

Results

Molecular Mass A:	43639.327765
Std. Deviation A:	0.45345
Molecular Mass B:	43639.229634
Std. Deviation B:	0.375334

In order to verify the success of the refolding process, CD spectra were recorded. As shown by the overlay of CD spectra of the per-PehA and the ref-PehA (Figure 2.16), the two spectra exhibited a very similar trend, therefore we concluded the ref-PehA was correctly folded.
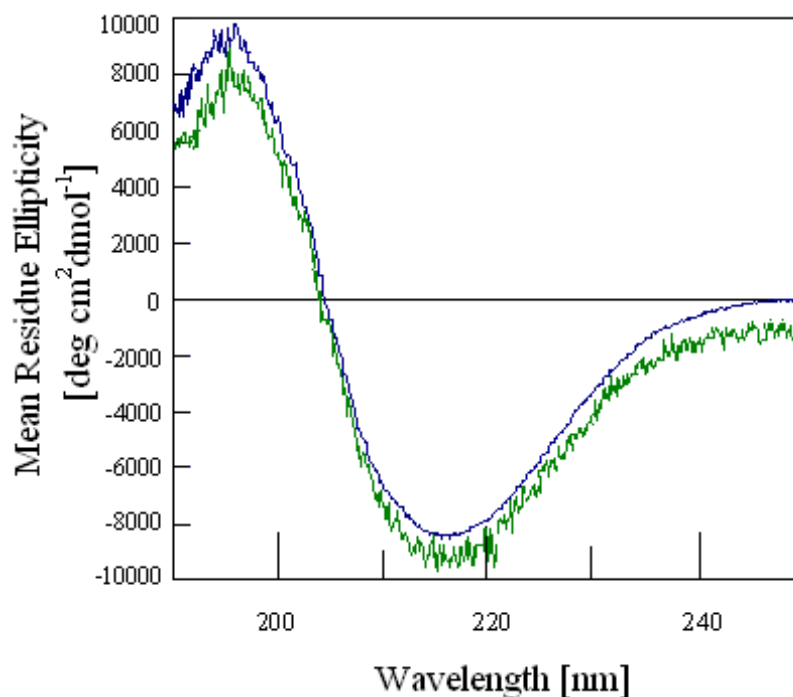
Figure 2.16. Overlay of CD spectra of per-PehA (----) and ref-PehA (----). The CD spectra were collected in the wavelength range of 190-250 nm at room temperature. The CD data are presented in terms of mean residue ellipticity, as described in Par. 2.2.3.1.2. In both cases, the protein concentration was 0.2 mg/mL (0.5μM), in aqueous solution pH 3.5.

2.4 Discussion

An endopolygalacturonase was isolated from *B. cepacia* ATCC25146 culture media and purified to homogeneity. Its N-terminal amino acid sequence revealed that this protein corresponds to an endoPG gene previously identified by Gonzales et *al.* (22). Evidences presented in (22) proved that the *pehA* gene is localized on a large plasmid of 200 kb, pPEC320, and that the PehA is a virulence factor involved in serious plant diseases. The elucidation at the molecular level of the enzymatic reaction mechanism awaits the determination of the PehA three dimensional structure. Therefore, in order to produce adequate amounts of protein suitable for the subsequent structural and functional studies, PehA was heterologously expressed in *E. coli* cells, testing different expression strategies.

As first attempts, PehA was expressed in the periplasmic space of the host cells, according to the heterologous expression strategy already reported for other bacterial endoPGs (14-16). On the other hand, since the cytoplasmic expression systems are associated with higher yields of heterologous protein production (41), several approaches for cytoplasmic expression of wt-PehA were tested.

Concerning the periplasmic expression, a broad screening of the culture media conditions has been exploited to more efficiently explore the protein expression. Enzymatic activity could only be detected in the periplasmic extract of cells grown using glycerol as carbon source. As elsewhere reported (42, 43), the presence of glycerol in the culture medium affects the expression level of a cytoplasmic translocation-specific chaperone, SecB. Thus, the exclusive detection of enzymatic activity in the culture grown on glycerol may reflect the improvement in the periplasmic secretion as a result of the SecB overexpression.

From the results of the biochemical studies, it is worthy of note that the per-PehA enzymatic activity is very sensitive to pH (Figure 2.13). The enzyme exhibited an optimal activity at pH 3.5 (Figure 2.13A), similar to that reported by J.M. Ulrich (21). Such acidic pH value can be easily explained keeping in mind the physiological conditions of the PehA activity. The enzyme is produced during the infection of the onion tissue. The onion juice leaked out from infected tissue has a pH near 4.0, that is lower than that of the healthy tissue (pH 5.5). Monitoring the stability of per-PehA in

the pH interval 2.0-8.0, a drop in the enzymatic activity at pH 6.0 was observed (Figure 2.13B). The loss of activity at neutral and basic pH values may be ascribed to the ionization state of amino acidic residues involved in substrate binding and in catalysis, as already reported for other glycoside hydrolases (44). With regard to endoPGs, the ionization state of a conserved histidine residue has been shown to be crucial for their enzymatic activity (45, 46). This histidine is thought to be implicated in maintainance of the proper ionization state of a carboxylate involved in catalysis, by sharing a proton, as discussed also in Chapter 1 (Par. 1.3.2.3). Also the substrate conformation may influence the enzymatic activity. The pKa of the PGA is 3.7 (47), thus at very acidic pH values the substrate has a null net charge and would be conformationally more flexible and adaptable to the binding sites of the enzyme. Regarding the enzyme activity and stability with respect to temperature, it is important to note that also temperature changes may also affect the conformation of the substrate. However, the three dimensional structure of PGA in solution is currently under study, since it is influenced by many factors including, beside the temperature and the pH, also the presence of cations and the degree of esterification (48, 49).

From a functional point of view, per-PehA exhibits an enzymatic activity comparable to that of the wild-type protein. The observed low K_M value (0.487 mg/mL) points out the per-PehA high affinity toward PGA if compared to the K_M value estimated for the endopolygalacturonase E from *A. niger* (2.5 mg/mL) (50).

On the other hand, by the means of the biophysical characterization of per-PehA, CD analysis has been shown the prevalence of β-strands among secondary structure elements (see Table 2.2). As elsewhere reported for the circular dichroism of the parallel β-helical proteins pectate lyase C and E, two features are characteristic in the CD spectra of these proteins: a minumum at 216 nm and a crossover at 208 nm (51). The CD spectra of per-PehA exhibits these elements, in agreement with the expected β-helix fold.

A further conserved feature of the β-helix fold is the occurrence of disulphide bonds that most likely play an important role in stabilizing the overall structure (39). The existence of disulphide bonds in per-PehA was supported by the band shift on the SDS-PAGE gels that were run under reducing and non-reducing conditions. A more detailed

study will be reported in the Chapter 4 and it has been aimed to verify if all the six cysteines found in the primary sequence of PehA are involved in disulphide bridges.

As concerns the expression in the cytoplasm of the host cells, a different strain type was chosen. As elsewhere reported (52), the OrigamiB (DE3) strain is eligible for the heterologous expression of proteins that contain disulphide bonds, since mutations in the thioredoxin and glutaredoxin pathways force its cytoplasm to function under a non-reduced state. The cytoplasmic expression tests were performed using two different constructs, in order to produce the recombinant enzyme without or with an N-terminal fusion partner. As well known, fusion tags are widely used in order to enhance the solubility of the target protein and to facilitate the purification (53). Therefore, the heterologous expression of PehA in fusion with thioredoxin has been performed. It has been reported that the thioredoxin fusion tag not only promotes the formation of disulphide bond in the target protein (54), but also improves the solubility of proteins isolated from the inclusion bodies (55).

With regard to protein solubility, PehA either alone or as fusion protein, was expressed as active enzyme in the cytoplasm of ORIGAMIB (DE3) cells, even if its expression level was quite low. An overexpression of the TrxA-PehA was instead observed in the insoluble fraction of the cell lysate. Therefore, the purification from the inclusion bodies and the refolding of the fusion protein were attempted. The TrxA-PehA was on-column refolded and subsequently incubated with the TEV protease, in order to remove the fusion partner. The enzymatic activity of the refolded fusion protein was detected only in a qualitative manner, by using the pectate-agarose gel method. A comparison of the enzymatic activity between the cleaved and the uncleaved forms of PehA revealed that the presence of the thioredoxin tag reduced the enzymatic activity, probably interfering with the substrate binding.

The refolded and cleaved PehA (ref-PehA) was also characterized using CD analysis, in order to verify if it displays the expected parallel β-helix fold. As shown in Figure 2.16, the CD spectra of per-PehA and ref-PehA exhibited a very similar trend. Therefore it is possible to conclude that the ref-PehA is properly folded. Our finding is in good agreement with statement elsewhere reported (40), according which the native structure of the parallel β-helix proteins can be completely regained, after the denaturation with caothropic agents.

The final yield of the ref-PehA was quite low, resulting three times lower than the yield of the PehA, expressed according to the periplasmic expression system described above. The striking influence of the pH on the enzymatic activity of the endoPGs (45) left us with the choice of buffer conditions that were significantly unfavoured for the cleavage with TEV protease.

In conclusions, it has been reported the isolation and the purification of an endoPG from *Burkholderia cepacia* supernatant and its preliminary characterization. Then, two strategies for the heterologous expression of PehA have been developed. Among these, the strategy for the periplasmic expression of PehA has been revealed to be the more convenient, in order to produce in adequate amounts a functional and correctly folded enzyme.

Bibliography

(1) S. Vorwerk, S. Somerville, C. Somerville, The role of plant cell wall polysaccharide composition in disease resistance, Trends Plant. Sci. 9 (2004) 203-209.

(2) M. McNeil, A.G. Darvill, S.C. Fry, P. Albersheim, Structure and function of the primary cell walls of plants, Annu. Rev. Biochem. 53 (1984) 625-663.

(3) W.G. Willats, L. McCartney, W. Mackie, J.P. Knox, Pectin: cell biology and. prospects for functional analysis, Plant Mol. Biol. 47 (2001) 9-27.

(4) B.L. Ridley, M.A. O'Neill, D. Mohnen, Pectins: structure, biosynthesis and oligogalacturonide-related signaling, Phytochemistry 57 (2001) 929-967.

(5) R.A. Prade, D. Zhan, P. Ayoubi, A.J. Mort, Pectins, pectinases and plant-microbe interactions, Biotechnol. Genet. Eng. Rev. 16 (1999) 361-391.

(6) A. Collmer, N.T. Keen, The role of pectic enzymes in plant pathogenesis, Annu. Rev. Phytopahtol. 24 (1986) 383–409.

(7) O.P. Ward, M. Moo-Young, Enzymatic degradation of cell wall and related plant polysaccharides, Crit. Rev. Biotechnol. 8 (1989) 237-274.

(8) R.P. de Vries, J. Visser, .Aspergillus enzymes involved in degradation of plant cell wall polysaccharides, Microbiol. Mol. Biol. Rev. 65 (2001) 497-522.

(9) O. Markovic, S. Janecek, Pectin degrading glycoside hydrolases of family 28: sequence-structural features, specificities and evolution, Protein Eng. 14 (2001) 615-631.

(10) C. Allen, Y. Huang, L. Sequeira, Cloning of genes affecting polygalacturonase production in *Pseudomonas solanacearum*., Mol. Plant-Microbe Interact. 4 (1991) 147-154.

(11) S.P. Lei, H.C. Lin, S.S. Wang, P. Higaki, G. Wilcox, Characterization of the *Erwinia carotovora* peh gene and its product polygalacturonase, Gene 177 (1992) 119-124.

(12) P. Rodriguez-Palenzuela, T.J. Burr, A. Collmer, Polygalacturonase is a virulence factor in *Agrobacterium tumefaciens* biovar 3, J. Bacteriol. 173 (1991) 6547-6552.

(13) H.T. Saarilahti, M. Pirhonen, M.B. Karlsson, D. Flego, E.T. Palva, Expression of *pehA-bla* gene fusions in *Erwinia carotovora* subsp. *carotovora* and isolation of regulatory mutants affecting polygalacturonase production. Mol. Gen. Genet. 234

(1992) 81-88.

(14) J.H. Huang, M.A. Schell, DNA sequence analysis of pglA and mechanism of export of its polygalacturonase product from *Pseudomonas solanacearum*, J. Bacteriol. 172 (1990) 3879–3887.

(15) H. Hemilä, R. Pakkanen, R. Heikinheimo, E.T. Palva, I. Palva, Expression of the *Erwinia carotovora* polygalacturonase-encoding gene in *Bacillus subtilis*: role of signal peptide fusions on production of a heterologous protein, Gene 116 (1992) 27–33.

(16) T.C. Herlache, A.T. Jr Hotchkiss, T.J. Burr, A. Collmer, Characterization of the *Agrobacterium vitis* PehA gene and comparison of the encoded polygalacturonase with the homologous enzymes from *Erwinia carotovora* and *Ralstonia solanacearum*, Appl. Environ. Microbiol. 63 (1997) 338-46.

(17) A.P. Pugsley, The complete general secretory pathway in gram-negative bacteria, Microbiol. Mol. Biol. 57 (1993) 50-108.

(18) A.J.M. Driessen, P. Fekkes. J.P.W. Van der Wolk, The Sec-system, Curr. Opin. Microbiol. 1 (1998) 216-222.

(19) L. Chiarini, A. Bevivino, C. Dalmastri, S. Tabacchioni, P.Visca, *Burkholderia cepacia* complex species: health hazards and biotechnological potential, Trends in Microbiol. 14 (2006) 277-286.

(20) W.H. Burkholder, Sour skin, a bacterial rot of onions bulbs, Physiopathology 40 (1950) 115-117.

(21) J.M. Ulrich, Pectic enzymes of *Pseudomonos cepacia* and penetration of polygalacturonase into cells, Physiol. Plant Pathol. 5 (1975) 37-44.

(22) C.F. Gonzales, E.A. Pettit, V.A. Valadez, E.M. Provin, Mobilization, cloning, and sequence determination of a plasmid-encoded polygalacturonase from phytopathogenic *Burkholderia (Pseudomonas) cepacia*, Mol. Plant-Microbe Interact. 10 (1997) 840-851. Erratum in: Mol. Plant-Microbe Interact. 11 (1998) 580.

(23) V. Venturi, A. Friscina, I. Bertani, G. Devescovi, C. Aguilar, Quorum sensing in the *Burkholderia cepacia* complex, Res. Microbiol. 155 (2004) 238-244.

(24) C. Aguilar, I. Bertani, V. Venturi, Quorum-sensing system and stationary-phase sigma factor (rpoS) of the onion pathogen *Burkholderia cepacia* genomovar I type strain, ATCC 25416, Appl. Environ. Microbiol. 69 (2003) 1739-1747.

(25) T. Maniatis, E.M. Fritsch, J. Sambrook, Molecular cloning: a laboratory

manual 2nd Ed. Cold Spring Harbor Press, New York, 1989.

(26) P. Bernfeld, Amylases α and β, Methods Enzymol. 1 (1955) 149–158.

(27) J.L. Ried, A. Collmer, An activity stain for the rapid characterization of pectic enzymes in isoelectric focusing and sodium dodecyl sulfate-polyacrylamide, Appl. Environ. Microbiol. 50 (1985) 615-622.

(28) V.K. Laemmli, Cleavage of structural protein during the assembly of the head of bacteriophage T4, Nature 227 (1970) 680-685.

(29) C. Manoil, J. Beckwith, A genetic approach to analyzing membrane protein topology, Science 233 (1986) 1403-1408.

(30) M.F. Chaplin, Reducing sugar-neocuproine assay, in: M.F. Chaplin, J. Kennedy (Eds), Carbohydrate analysis: a practical approach, Oxford University Press, United Kingdom, 1986.

(31) H.C.M. Kester, J. Visser, Purification and characterization of polygalacturonases produced by the hyphal fungus *Aspergillus niger*, Biotechnol. Appl. Biochem. 12 (1990) 150-160.

(32) N. Sreerama, R.W. Woody, A self-consistent method for the analysis of protein secondary structure from circular dichroism, Anal. Biochemistry 209 (1993) 32-44.

(33) W.C. Johnson, Analyzing protein circular dichroism spectra for accurate secondary structures, Proteins: struct. funct. genet. 35 (1999) 307-312.

(34) S.W. Provencher, J. Glockner, Estimation of globular protein secondary structure from circular dichroism, Biochemistry 20 (1981) 33-37.

(35) L. Whitmore, B.A. Wallace, DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data, Nucleic Acid Res. 32 (2004) W668-W673.

(36) A. Dümmler, A.M. Lawrence, A. de Marco, Simplified screening for the detection of soluble fusion constructs expressed in E. coli using a modular set of vectors , Microb. Cell Fact. 4 (2005) 34.

(37) W.A. Prinz, F. Aslund, A. Holmgren, J. Beckwith, The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm, J. Biol. Chem. 272 (1997) 15661-15667.

(38) A.I. Derman, W.A. Prinz, D. Belin, J. Beckwith, Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*, Science 262 (1993) 1744-1747.

(39) J. Jenkins, O. Mayans, R. Pickersgill, Structure and evolution of parallel β-helix proteins, J. Struct. Biol. 122 (1998) 236-246.

(40) D.E. Kamen, Y. Griko, R.W. Woody, The stability, structural organization, and denaturation of pectate lyase C, a parallel β-helix protein, Biochemistry 36 (2000) 15932-15943.

(41) H.P. Sørensen, K.K Mortensen, Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*, Microb. Cell Fact. 4 (2005) 1.

(42) R.S. Ullers, J. Luirink, N. Harms, F. Schwager, C. Georgopoulos, P. Genevaux, SecB is a bona fide generalized chaperone in *Escherichia coli*, Proc. Natl. Acad. Sci. U.S.A. 10 (2004) 7583-7588.

(43) H.K. Seoh, P.C. Tai, Carbon source-dependent synthesis of SecB, a cytosolic chaperone involved in protein translocation across *Escherichia coli* membranes, J. Bacteriol. 179 (1997) 1077-1081.

(44) A. Olivera-Nappa, B.A. Andrews, J.A. Asenjo, A mixed mechanistic-electrostatic model to explain pH dependence of glycosyl hydrolase enzyme activity, Biotechnol. Bioeng. 86 (2004) 573-586.

(45) T.C. Jyothi, S.A.Singh, A.G. Appu Rao, The contribution of ionic interactions to the conformational stability and function of polygalacturonase from *A. niger*, Int. J. Biol. Macromol. 36 (2005) 310-317.

(46) C. Caprari, B. Mattei, M.L. Basile, G. Salvi, V. Crescenzi, G. De Lorenzo, F. Cervone, Mutagenesis of endopolygalacturonase from *Fusarium moniliforme*: histidine residue 234 is critical for enzymatic and macerating activities and not for binding to polygalacturonase-inhibiting protein (PGIP), Mol. Plant-Microbe Interact. 9 (1996) 617-624.

(47) R.W. Stoddart, I.P. Spires, K.F. Tipton, Solution properties of polygalacturonic acid, Biochem. J. 114 (1969) 863-70.

(48) I. Braccini, R.P. Grasso, S. Pérez, Conformational and configurational features of acidic polysaccharides and their interactions with calcium ions: a molecular modeling investigation, Carbohydr. Res. 317 (1999) 119–130.

(49) I. Braccini, S. Pérez, Molecular Basis of Ca^{2+}-Induced Gelation in Alginates and Pectins: The Egg-Box Model Revisited, Biomacromolecules 2 (2001) 1089 -1096.

(50) L. Parenicova, J.A.E. Benen, H.C.M. Kester, J. Visser, pgaE encodes a fourth member of the endopolygalacturonase gene family from *Aspergillus niger*. Eur. J. Biochem. 251 (1998) 72-78.

(51) V. Sieber, F. Jurnak, G.R. Moe, Circular dichroism of the parallel β-helical proteins pectate lyase C and E, Proteins: struct. funct. Genet. 23 (1995) 32-37.

(52) E.J. Stewart, F. Aslund, J. Beckwith, Disulfide bond formation in the *Escherichia coli* cytoplasm: an in vivo role reversal for the thioredoxins, EMBO J. 17 (1998) 5543-5550.

(53) K. Terpe, Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems, Appl. Microbiol. Biotechnol. 60 (2003) 523-533.

(54) E.R. La Vallie, Z. Lu, E.A. Diblasio-Smith, L.A. Collins-Racie, J.M. McCoy, Thioredoxin as a fusion partner for production of soluble recombinant proteins in Escherichia coli, Methods Enzymol. 326 (2000) 322-340.

(55) D. Sachdev, J.M. Chirqwin, Solubility of proteins isolated from inclusion bodies is enhanced by fusion to maltose-binding protein or thioredoxin. Protein Expr. Purif. 12 (1998) 122-132.

Chapter 3. Study of the mode of action of PehA

3.1 Introduction

As discussed in the Chapter 1, endopolygalacturonases (endoPGs) are responsible for the enzymatic degradation of pectic substances that constitute the cell walls of the higher plants. These enzymes hydrolyse the $\alpha-1,4$ glycosidic bond in the non-esterified smooth region of pectins, built up by homogalacturonan (HGA). In the well-studied plant infection process caused by pathogenic fungi, a battery of enzymes able to degrade pectic substances has been identified (1-3). Their enzymatic activities have been extensively characterized (4-9). The identification of a large number of pectin-degrading enzymes that exhibit distinct biochemical properties suggests that pathogenic fungi have evolved catalytic specificities, in order to guarantee the maximum efficiency during the infection process (10).

To elucidate the mode of action of endoPGs and to define their substrate specificity, the most commonly used experimental approach consists in the identification of the products released following the hydrolysis of polygalacturonic acid (PGA) and of its fragments, oligogalacturonides (OGAs) (4-9). The detection of the hydrolysis products generated in the early stage of the PGA hydrolysis allows the endoPGs to be classified as *single attack* or *multiple attack* enzymes. In the case of *single attack* enzymes, the substrate molecules are cleaved only once after the formation of the enzyme-substrate

(ES) complex, resulting in the release of different chain length OGAs. The *multiple attack* enzymes, called processive enzymes, instead cleave the same substrate molecule several times, and small oligomers, even monomer, are released in the initial minutes of the reaction. Moreover, the mode of action on OGAs that differ in the degree of polymerization (DP) and in their esterification pattern has been also extensively studied (4-9, 11), in order to describe the architecture of the active site of endoPGs. A conventional nomenclature has been adopted to describe the ES complex (12): the enzymatic regions that interact with the polymeric substrate are referred to as *subsites*. Precisely, the term *subsite* indicates the binding site of each saccharide unit on the protein scaffold. By convention, subsites are labeled from *-n* to *+n*, where *-n* represents the non-reducing end and *+n* the reducing end of the OGAs. The cleavage occurs between the subsites -1 and +1. The orientation of the OGAs in the ES complex has been previously defined by means of biochemical studies on endoPGs of *Aspergillus niger* (5) and further confirmed by crystallographic evidences revealed by the crystal structures determination of an endoPG from *Stereum purpureum* in complex with monogalacturonate molecules (13). Therefore, it has been proposed that OGAs are bound to the enzyme in such a way that their non reducing ends are oriented toward the N-terminal of the protein, while the reducing ends are directed toward the C-terminal portion (Figure 3.1).
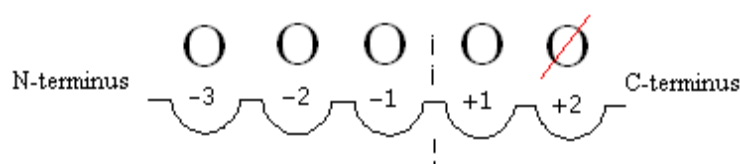
Figure 3.1. A schematic representation of the interaction between endoPGs and OGAs. The reducing end is indicated with the red line.

The number of subsites involved in the substrate recognition and the evaluation of their affinity allow a detailed characterization of the ES complex, as it has been elegantly reported for the endoPG II of *A. niger*. In this work, on the basis of the three dimensional structure determination (14), several mutants have been produced, leading to the identification of the amino acids involved in catalysis (15) and in substrate

binding (16). The three dimensional structure of a second endoPG of *A. niger*, the endoPG I, that exhibits a processive mode of action on PGA (5), has been also reported (17). Their structural comparison allowed the identification of the structural determinants that are relevant in discriminating between the processive and non-processive behaviour of endoPGs (18).

At present, very little is known about the mode of action of bacterial endoPGs (19). In this chapter, the mode of action of PehA on polygacturonic acid (PGA) and on pure OGAs that differ in their DP and in their methylation pattern will be discussed.

3.2 Materials and Methods

3.2.1 Production and purification of PehA

The endopolygalacturonase of *Burkholderia cepacia* (PehA) was heterologously expressed and purified as previously reported (Chapter 2, Par. 2.2.2.1).

3.2.2 Substrates

3.2.2.1 Polygalacturonic acid

The PGA sodium salt from citrus fruit was purchased from Sigma Aldrich. According to information provided by the Sigma Aldrich Co., the raw material derived from the citrus fruit was de-esterified with NaOH and the final level of ester was about 5%, provided not as a specification. Concerning the molecular weight, the Manufacturer declares that this product has been purified from high-grade pectin which has a molecular weight distribution of 50,000-150,000 Da: however, no molecular weight testing has been performed on the final product.

3.2.2.2 Oligogalacturonides

A mixture containing OGAs with DP ranging from 1 to 8 was kindly provided by Dr. G. Salvi, Dipartimento di Biologia Vegetale, Universita' La Sapienza di Roma, Italy, and was used as standard in the analysis by capillary electrophoresis (CE) and high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD). The mixture was obtained as described in (20).

3.2.2.3 Decagalacturonide

A fully unmethylated decagalacturonide was a kind gift of Prof. K.J. Jensen, Department of Natural Sciences, The Royal Veterinary and Agricultural University, Denmark. The decamer was purified from a pectin digest as reported in (21).

3.2.2.4 Hexagalacturonides

Hexagalacturonides that differ in their methylation pattern were a kind gift of Prof. R. Madsen, Department of Chemistry, Technical University of Denmark, Lyngby, Denmark. These molecules were synthesized according to the procedure reported in (22).

3.2.3 Time-course experiments

3.2.3.1 Product progression profile

PehA (0.5 μg/mL) was incubated at different temperatures, 37 °C and 20°C , in 2.0 mL of 50 mM sodium acetate, pH 3.5, in the presence of 0.25% g/L PGA as substrate. Aliquots of 50 μl were taken at regular intervals and the reaction was stopped by adding an equal volume of derivatization solution, as described below (Par. 3.2.5.1). The samples were then boiled at 95 °C for 20 minutes. The identification and molar number distribution of OGAs in the hydrolysis mixture has been performed by CE.

3.2.3.2 Substrate specificity

K_{cat}/K_M values of oligomers of different chain lengths were determined according to the following equation:

$$v_A = (K_{cat}/K_M) \; [E_{tot}][A_i]$$

where: v_A is the rate of disappearance of the substrate A, $v_A = -([A]_{t_2} - [A]_{t_1})/(t_2 - t_1)$;

$[E_{tot}]$ is the initial concentration of the enzyme;

$[A_i]$ is the concentration of the substrate A at time t_1.

This equation is used to define the substrate specificity when more substrates of an enzyme are present in the reaction (23). In this case, multiple substrates molecules are generated by the primary enzymatic cleavage of the initial substrate, the PGA.

3.2.4 Mode of action on OGAs

3.2.4.1 Hydrolysis of decagalacturonide

PehA (0.075 µg/mL) was incubated at 37 °C with decagalacturonide (50 µM) in 50 mM acetate buffer, pH 3.5. The reaction was stopped at regular interval by the addition of an equal volume of 2.0 mM Tris-HCl, 50 mM sodium hydroxide, which resulted in a final pH of 8.3-8.5. In order to avoid the spontaneous breakdown of OGAs reported to occur at high pH (4), the samples were then immediately submitted to HPAEC-PAD. The identification of the hydrolysis products was done in accordance to the retention time of a standard mixtures containing OGAs with a DP ranging from 1 to 8.

3.2.4.2 Hydrolysis of hexagalacturonides

PehA (1.0 µg/mL) was incubated at 37 °C with 25 µM of hexagalacturonides in 50 mM acetate buffer pH 3.5, for different incubation times. The enzymatic digest was analysed by electrospray mass spectrometry with a time of flight analyser (ESI-TOF-MS).

3.2.5 Analytical techniques

3.2.5.1 Capillary electrophoresis (CE)

Capillary electrophoresis experiments were carried out using a Hewlett-Packard HP3D Capillary electrophoresis system, equipped with a diode array UV detector. Fused capillaries were from Agilent Technologies (total length: 104 cm; effective length: 95.5 cm; i.d. 50 µm; extended light path). Before sample injection, a 4-min conditioning of the capillary with the running buffer was necessary, preceded by a 2-min wash with 0.1 M sodium hydroxide (pressure equal to 950 mbar). Signal acquisition was made by continuously monitoring UV absorbance at 195 nm and 285 nm. All analyses were done at 25 °C. The separation potential was equal to 27 kV (applied at cathode). Samples were injected at 50 mbar for 6 sec. Monomer, dimer and trimer of galacturonic acid purchased from Sigma Aldrich were used as standard species. A mixture containing OGAs with DP ranging from 1 to 8 was used as standard too. To allow a sensitive UV detection, the saccharidic compounds were derivatized by reaction with 4-amino-

benzonitrile (4-ABN) in the presence of sodium cyanoborohydride (NaCNBH$_3$) (24). The derivatization solution is composed of 0.16 M NaCNBH$_3$ and 0.5 M 4-ABN in methanol/acetic acid (95/5 v/v). Only one 4-ABN molecule chromophore is attached to each oligomer, leading to have an UV-response independent of the length of OGAs. This represents an advantage over chromatographic techniques such as HPAEC-PAD where the electrochemical response must be determined for each OGA (25). Derivatization solution was diluted 1:1 with an aqueous solution of the uronic acid(s) and the derivatization mixture was heated for 20 minutes at 95 °C. Prior to injection into the CE system, samples were spun down for 1 minute at 13000 g. The analysis was performed by micellar electrokinetic capillary chromatography (MEKC) as described in (25). The buffer used was 600 mM boric acid, containing 75 mM SDS (pH 8.0). The peak areas divided by migration time (A/t) were used for quantitative purposes (26). To obtain a calibration curve, electrophoregrams of different amounts of galacturonic acid were recorded and their corresponding values of A/t were plotted as a function of nanomoles of galacturonic acid. Calibration was linear in the concentration range comprised between 1.0 and 14.0 nanomoles.

3.2.5.2 High-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD)

The analysis of oligomers released from the hydrolysis of decagalacturonide was done with a HPAEC-PAD system (Dionex Corporation), equipped with a GP50 gradient pump, an ED50 electrochemical detector, and a LC25 chromatography oven. The separation was performed using a CarboPac PA-200 analytical anion exchange column (3x250 mm) with Guard column (3x50 mm). The volume of the injection loop was equal to 25 μL, and the column temperature was set to 35 °C. The flow-through cell of the pulsed amperometric detector consisted of an Au working electrode (1.0-mm diameter) and a pH-Ag/AgCl combination reference electrode; the titanium body of the cell served as the counterelectrode. The sequence of potentials applied to the electrode was set as recommended by the manufacturer for carbohydrate detection, as reported in Table 3.1.

Table 3.1. The sequence of potentials applied to the electrode to achieve the amperometric detection of OGAs in the HPAEC-PAD analysis.

Time (s)	Potential(V)
0.00	0.10
0.20	0.10, Integration = Begin
0.40	0.10, Integration = End
0.41	-2.00
0.42	-2.00
0.43	0.60
0.44	-0.10
0.50	-0.10

Acquisition and processing of the chromatograms was accomplished by the software Chromeleon, version 6.60 (Dionex Corporation). The column was equilibrated with 0.1 M sodium hydroxide, 0.2 M sodium acetate. The samples were eluted applying a linear gradient of 0.2-0.8 M sodium acetate in 0.1 M sodium hydroxide at 0.5 mL/min for 60 minutes. The identification of oligomers was done according to the retention time of a mixture containing OGAs with DP ranging from 1 to 8.

3.2.5.3 Electrospray mass spectrometry with a time of flight analyser (ESI-TOF-MS)

Mass spectrometry with electrospray ionization and a time-of-flight analyzer (ESI-TOF) was performed on a microTOF Focus system (Bruker Daltonics), equipped with a nitrogen generator N2LCMS1(Claind). The samples were diluted five times with a mixture of methanol/water/formic acid (49/49/2) and injected with a syringe pump in the mass spectrometer. The injection flow rate was 180 µL/h; the capillary voltage was 4500 V and the end plate offset was 500V (positive mode); the dry temperature was 180°C, the dry gas flow was 4 L/min and the nebulizer pressure was 0.4 bar.

3.3 Results

3.3.1 Time-course experiments

<u>3.3.1.1 Product progression profile</u>

In order to investigate the mode of action of PehA on polymeric substrate, the products released upon the hydrolysis of PGA were followed in a time course experiment by collecting CE electropherograms. Figure 3.2 shows the electrophoretic profile of a standard mixture containing OGAs with DP 1-8, supporting the suitability of CE for the separation of OGAs (25).
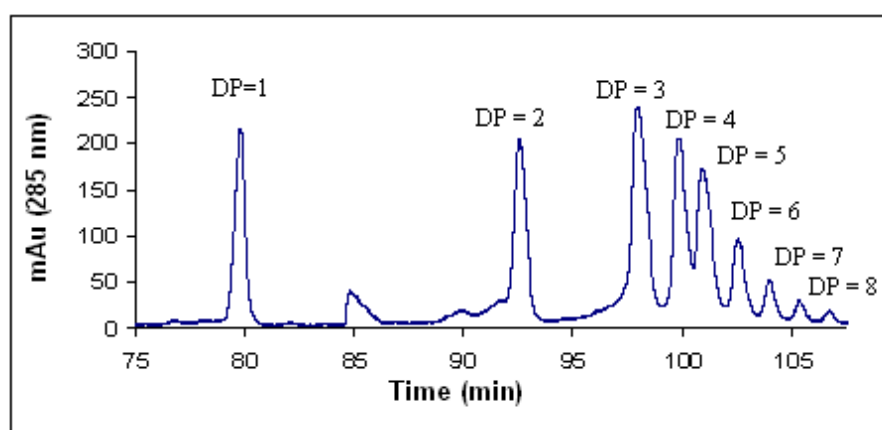

Figure 3.2. Electropherogram of the standard mixture of OGAs with DP 1-8.

As indicated by product progression profile at 37 °C (Figure 3.3), in the first minutes of the reaction, OGAs with a DP ranging from 2 to 8 are detected. This confirms that PehA is an endo-acting enzyme, cleaving the polymeric chain in a random way. A typical electropherogram of the reaction mixture in shown in Figure 3.4, giving an immediate visual impression of the molar product distribution in the enzymatic digest.

The detection of OGAs of different chain length clearly indicates that PehA can be classified as *single attack* enzyme. In fact, the *multiple attack* enzymes, also called processive enzymes, would exhibit a product progression profile in which smaller OGAs, even monomer, are the most abundant species released either in the early stage of the reaction or among the end products (see Figure 2C and 2E in Kars *et al.*, (9)).
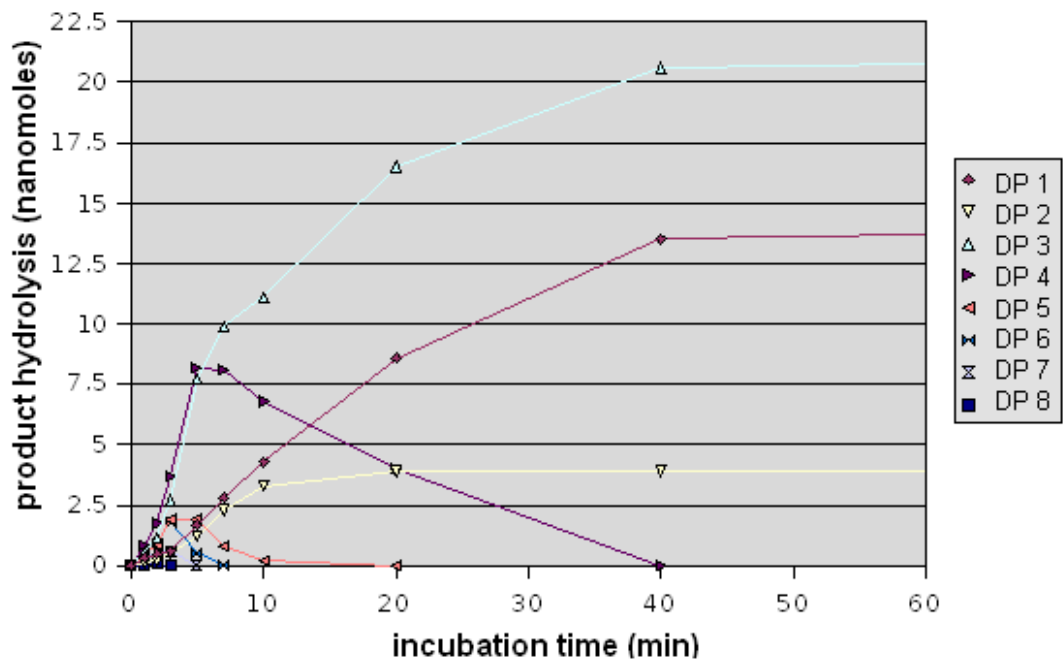
Figure 3.3. The progression profile of products released by PehA after 1 hour of incubation with 0.25 g/L PGA, at 37 °C.
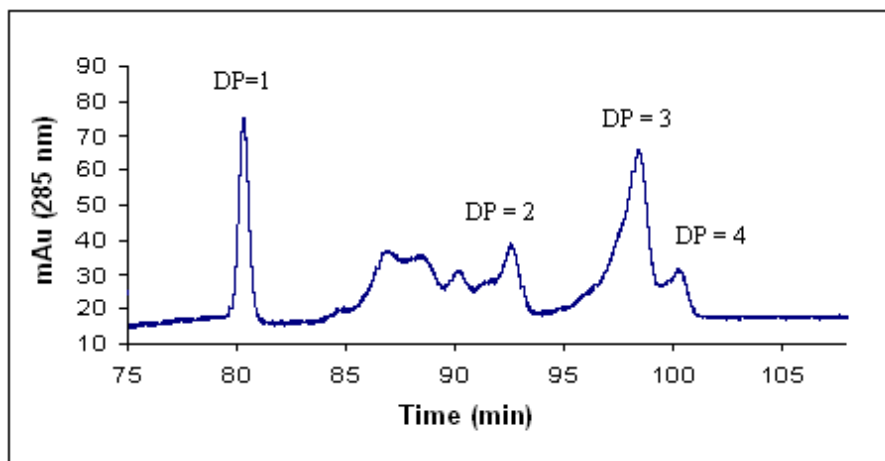
Figure 3.4 Electropherogram of the reaction mixtures, after 30 minutes of incubation at 37 °C.

To better investigate the mode of action of the enzyme on the polymeric substrate, the same experiment was performed at 20 °C. As indicated by the thermoactivity curve, less than 20% of the enzymatic activity is left at 20 °C (Figure 2.13, Chapter 2), allowing a more detailed detection of the oligomers released in the first minutes of the reaction. In

fact, being the cleavage slower, an accumulation of OGAs could be observed (Figure 3.5).

As clearly indicated in Figure 3.6A, in the first minutes of the hydrolysis, OGAs with DP ranging from 4 to 8 are transiently accumulated and further converted into shorter ones. The reaction plateau is reached when all OGAs that are substrates of the enzyme are degraded. The minimum length of oligomers released varies among pectin-degrading enzymes and provide further information about the architecture of the active site of the enzyme. Concerning the PGA hydrolysis catalysed by PehA, the end products of the reaction are monomer, dimer and trimer, with a clear excess of trimer (Figure 3.6B). To confirm the finding that the trimer cannot be degraded by PehA, the enzyme was incubated with a commercial sample of trigalacturonic acid, which was not cleaved (data not shown).

OGAs with DP 4-8 can be either products of the primary enzymatic cleavage of the polymeric substrate or substrates for the following cleavages. Their amount is the net sum between the amount of molecules that are generated by continuous degradation of longer substrate molecules (positive contribution) and the amount of molecules that are themselves hydrolysed (negative contribution). In the first minutes of the reaction, the positive contribution is larger than the negative one, thus explaining the product accumulation that we experimentally observed. Then, at the highest amount detected for each oligomer, these two contributions can be considered equivalent: an equal amount of the oligomer is produced and degraded. Beyond this equilibrium condition, the negative contribution becomes predominant and OGAs are hydrolysed.
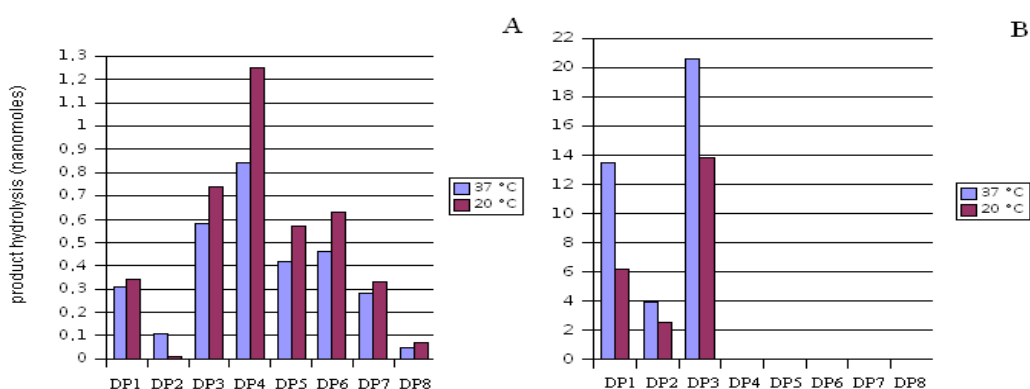
Figure 3.5. Differential accumulation of the products released by PehA in the time-course experiments run at 37 °C and 20 °C, after 1 minute (A) and 45 minutes (B) of incubation with 0.25 g/L PGA.
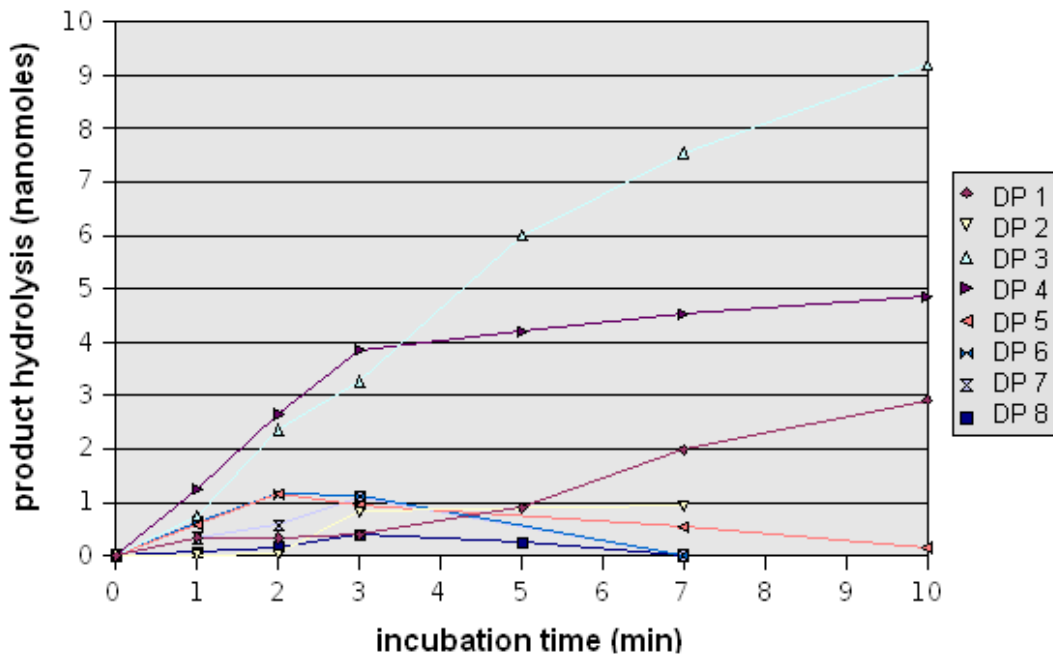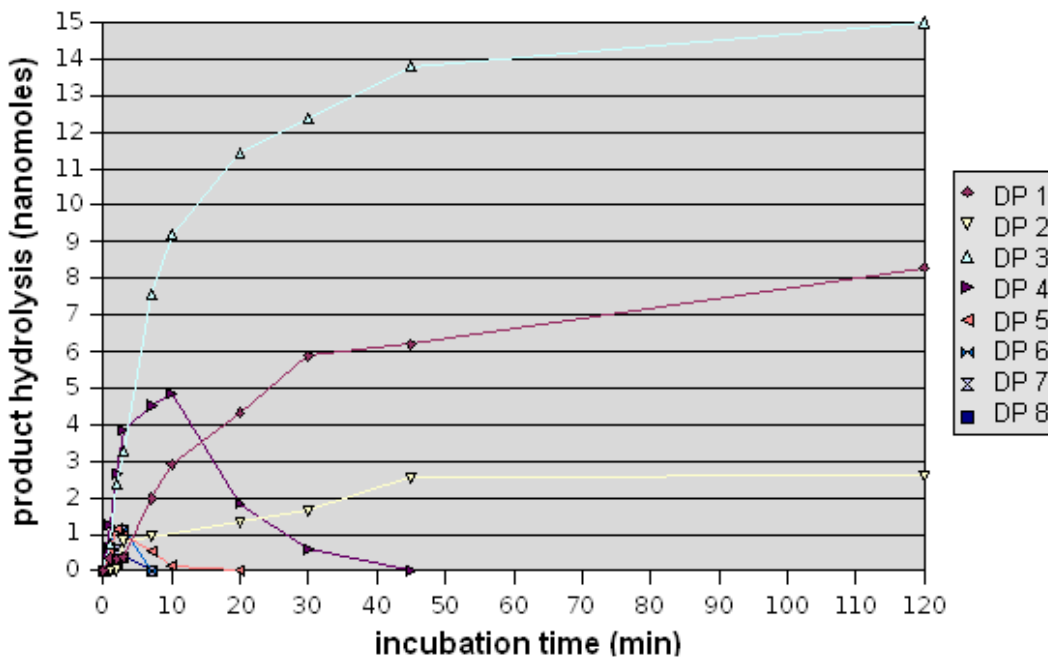
Figure 3.6. The progression profile of products released by PehA after 10 minutes (A) and 2 hours (B) of incubation with 0.25 g/L PGA, at 20 °C.

3.3.1.2 Substrate specificity

The rate of hydrolysis of each OGA is determined by the enzyme specificity. In order to have a quantitative way to describe the enzyme specificity, the K_{cat}/K_M values were calculated for each of the OGAs that could be detected under these reaction conditions. Higher values of K_{cat}/K_M ratio indicate higher substrate specificity. Interestingly, as suggested by Figure 3.7, the substrate specificity seems to depend on the length of the substrates. The affinity dramatically decreases at decreasing length of the substrates.
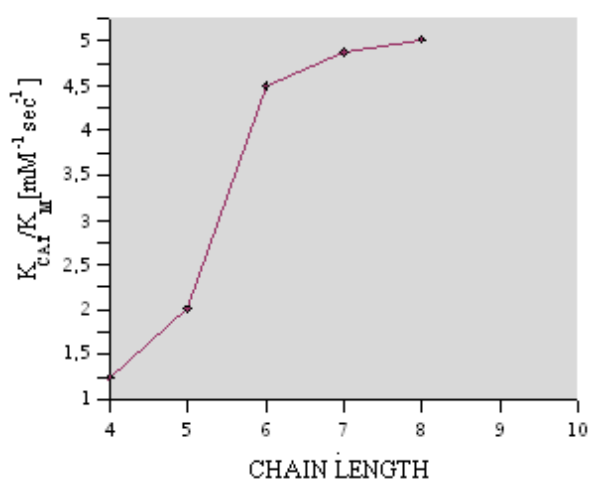
Figure 3.7. K_{cat}/K_M values are reported as a function of the chain length of the OGAs.

3.3.2 Mode of action on OGAs

3.3.2.1 Hydrolysis of decagalacturonide

To investigate if PehA displays a preferential mode of cleavage, its mode of action on a pure decagalacturonide was investigated. In this case, the hydrolysis products were analysed with a highly sensitive technique, the HPAEC-PAD. The amount of the initial substrate was in fact under the sensitivity limit for OGAs detection with MEKC-UV. To check the purity of the substrate, the decamer alone, prior to the enzymatic treatment, was subjected to HPAEC-PAD analysis. As shown by Figure 3.8A, small amounts of octamer and nonamer were detected. In the presence of the enzyme (Figure 3.8B), the decamer is hydrolysed, giving rise to OGAs with DP ranging from 3 to 7. To stop the enzymatic reaction, 2.0 mM Tris-HCl, 50 mM sodium hydroxide were added, as described in Materials and Methods (Par. 3.2.4.1). This buffer generated a large peak that masked the signal of monomer and dimer, not allowing their detection. The HPAEC-PAD analysis of the decamer digestion indicates that PehA is an endo-acting enzyme that displays a non-processive behaviour also on shorter substrates. As for the PGA hydrolysis, the trimer was also detected as the main end product, after the complete hydrolysis of decagalacturonides (data not shown).
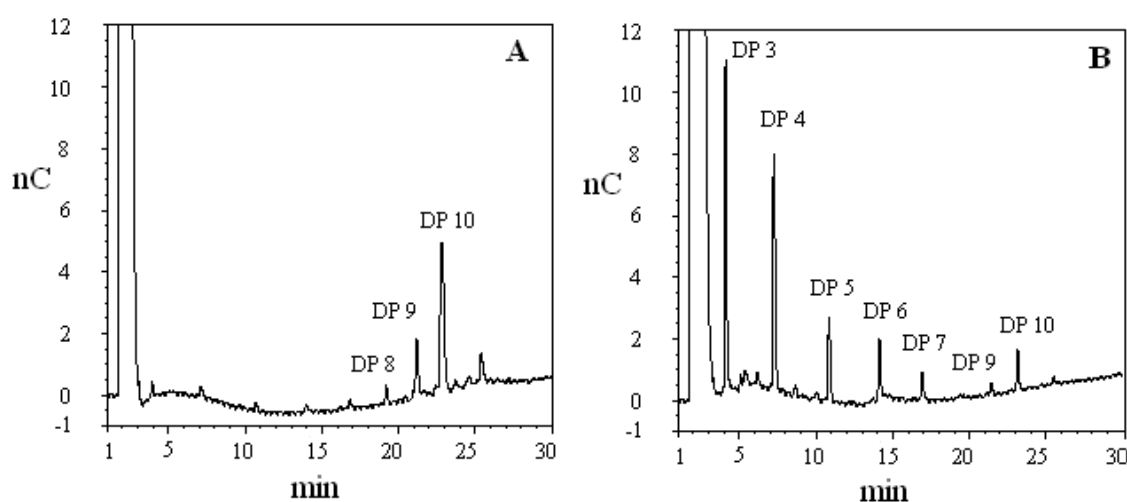
Figure 3.8. HPAEC-PAD analysis of the hydrolysis of decamer. In panel A, the chromatogram of the decamer prior to the enzymatic treatment is reported. In panel B, the hydrolysis products after 5 minutes of incubation at 37 °C are shown.

3.3.2.2 Hydrolysis of hexagalacturonides

The mode of action of PehA on hexagalacturonides that differ in the methylation pattern (Figure 3.9A and 3.9B) had been then investigated, in order to achieve a deeper investigation of the interaction between enzyme and substrate. In particular, the hydrolysis of the compound 1 and compound 2, that exhibit an asymmetric methylation pattern, is of interest in order to establish the orientation of the OGAs in the ES complex.

The end products of the hydrolysis reaction were analysed using either HPAEC-PAD or ESI-TOF-MS. The HPAEC-PAD analysis was not exhaustive to characterize the hydrolysis products, since, as elsewhere reported (27, 28), the methylated OGAs undergo a spontaneous de-methylation at the pH value requested for the HPAEC-PAD analysis (with the available instrumental setup, we could not perform any post-column sodium hydroxide addition for the electrochemical detection). As a consequence, although typical chromatograms showed peaks attributable to both methylated and de-methylated oligomers, it was not possible to quantitatively assess their amounts in the sample due to the occurring de-methylation during the analysis.
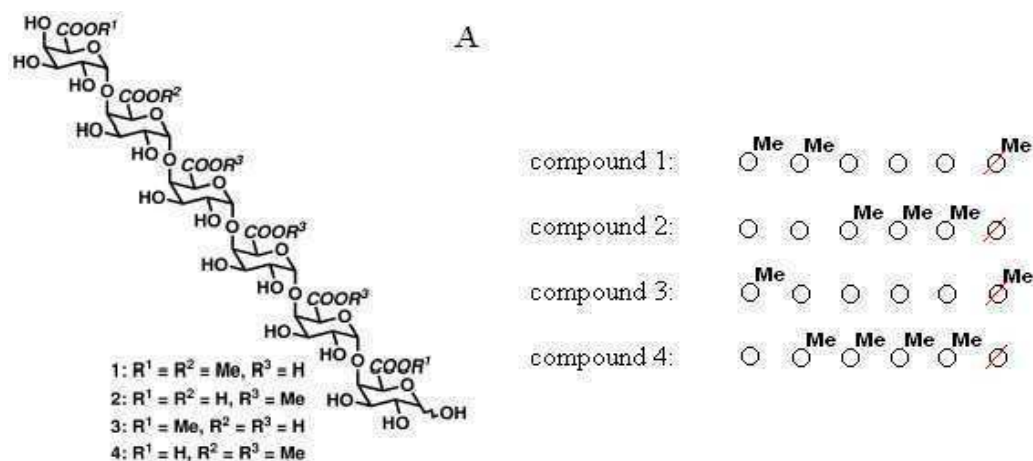

Figure 3.9. (A) Partial methyl-esterified hexagalacturonates. The Figure is taken from M.H. Clausen *et al*, (22). (B) Methylation patterns of the hexagalacturonides, the reducing end (+*n*) is indicated by the red line.

Nevertheless, HPAEC-PAD was used to determine the relative concentrations of enzyme and substrate at which the hydrolysis could occur. In fact, since methylated

pectins are not a good substrate for the enzyme, a higher concentration of the enzyme was chosen to trigger the reaction, while a smaller amount of substrate was used, as described in Materials and Methods (Par. 3.2.4.2). The HPAEC-PAD analysis revealed that compound 2 and compound 4 were not cleaved by PehA, as illustrated in Figure 3.10A and 3.10B for the compound 2.
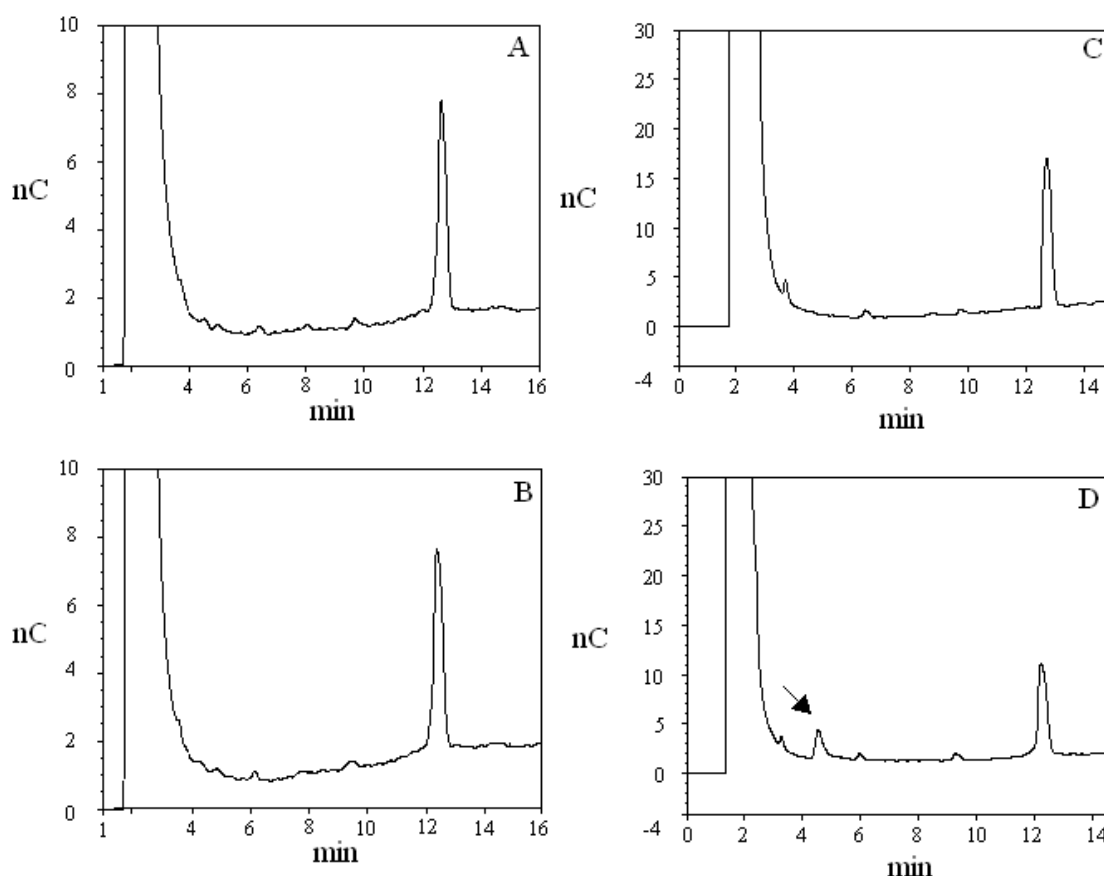
Figure 3.10. HPAEC-PAD analysis of compound 2 (panel A) and compound 1 (panel C), prior to the enzymatic treatment. Panel B and panel D show the HPEAC-PAD profile of compound 2 and compound 1, after the enzymatic treatment. The arrow indicates the formation of a product released upon the enzymatic hydrolysis of compound 1. Its identity has been later determined by ESI-TOF-MS (see Figure 3.12).

On the other hand, compound 1 (Figure 3.10C) and compound 3 are substrates of PehA, since new peaks corresponding to methylated hydrolysis products were present in the chromatograms, as indicated for the hydrolysis of compound 1 by the black arrow in Figure 3.10D. The identity of the hydrolysis products was assessed by ESI-TOF-MS. Figure 3.11 reports the mass spectrum of the compound 1, after 1 hour of incubation at 37 °C in the presence of the enzyme. The identification of a peak corresponding to

compound 1 (m/z = 1139.24, [M-2H+3Na]$^+$) suggested that it was not completely hydrolysed. The monomethylated dimer (m/z = 407.08, [M+Na]$^+$) and the dimethylated tetramer (m/z = 773.16, [M+Na]$^+$) were identified as hydrolysis products
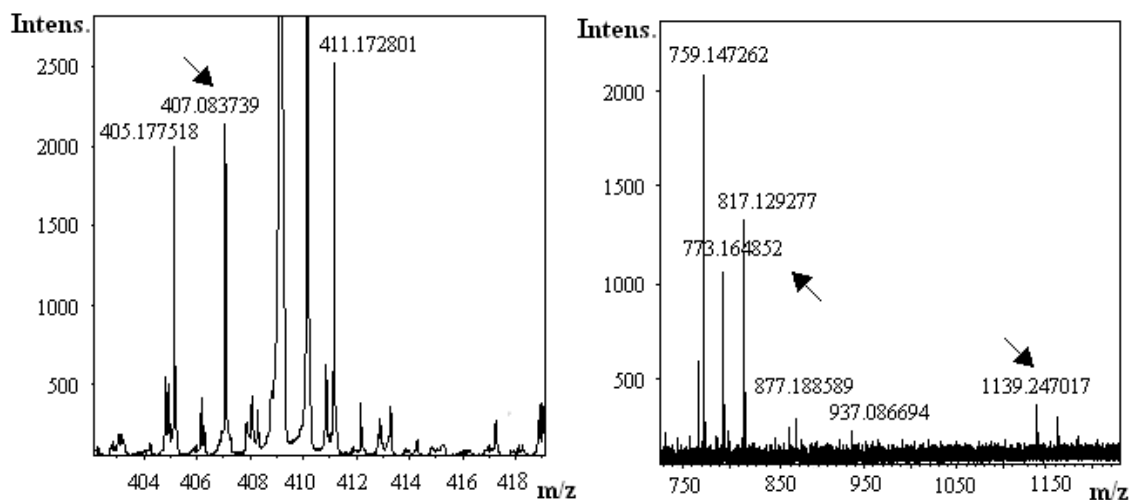
.Figure 3.11. Mass spectrum of the hydrolytic products released after the enzymatic hydrolysis of the compound 1.

The mass spectrum of the products released upon the enzymatic hydrolysis of compound 3 is shown in Figure 3.12. No peak attributable to compound 3 was found after an incubation time of 20 minutes at 37 °C. The monomethylated trimer (m/z = 583.11, [M+Na]$^+$) and the monomethylated dimer (m/z = 407.08, [M+Na]$^+$) were identified as end products.
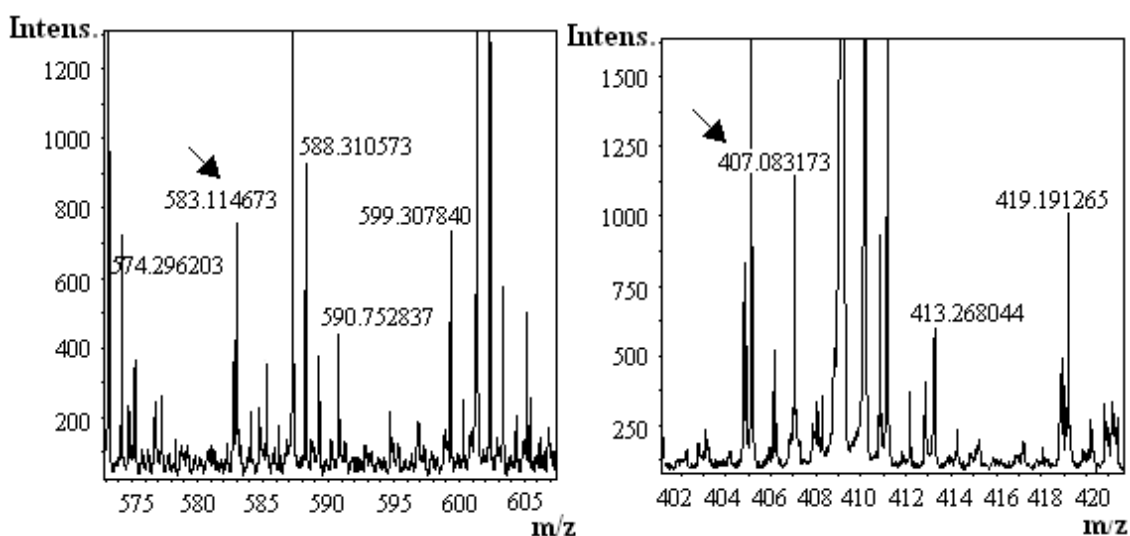

Figure 3.12. Mass spectrum of the hydrolytic mixture, originating from the enzymatic degradation of compound 3.

3.4 Discussion

Different experimental approaches have been used to shed lights on the mode of action of the PehA from *B. cepacia*. From the analysis of the products released upon the enzymatic hydrolysis of the PGA, it clearly appears that PehA is an endo-acting enzyme, that cleaves the polymeric chain in a random way. Moreover, since OGAs of different chain length were detected already in the first minutes of the reaction, the enzyme can be classified as a *single attack* enzyme. The PehA displays the same behaviour also on shorter substrates, as suggested by its mode of action on pure decagalacturonide. Among the end products, the trimer has been found in large excess. This result suggests that the minimum OGAs length that PehA can degrade is DP 4. DP 4 and DP 5 are hydrolysed slowly with respect to the other OGAs (Figure 3.3 and Figure 3.6).

To explain their accumulation, it is useful to introduce the concept of binding modes. As already mentioned, several enzymatic regions, referred as subsites, are involved in the binding to the monomeric units that compose a polymeric substrate (Figure 3.1). The binding modes refer to the multiple modes of interaction between the subsites and the saccharide units, with respect to the glycosidic bond that undergoes the enzymatic cleavage that is located between subsites -1 and +1. Multiple binding modes imply the occurrence of a variety of ES complexes. Not all of these complexes lead to product release, allowing a distinction between productive and non-productive binding modes. The number of productive binding modes increases with increasing lengths of OGAs. This observation also explains the higher K_{cat}/K_M values found for longer substrate molecules. Indeed, a lower number of productive binding modes may occur for shorter substrate molecules, like DP4 and DP5, thus explaining their slow degradation and their small K_{cat}/K_M values.

In a deeper description of the ES complex, the orientation of the OGAs bound to the enzyme must be defined. In our studies, the investigation of the PehA mode of action on methylated hexagalacturonides that exhibit an asymmetric methylation pattern, confirms the orientation of the OGAs in the ES complex to be in accordance with what has previously been reported for other endoPGs (5). Among the four methylated

hexagalacturonides that have been tested as substrates (Figure 3.9), only compound 1 and compound 3 were cleaved by PehA. The description of the productive and the non-productive binding modes of compound 1, the only hexagalacturonide with an asymmetric methylation pattern that has been hydrolysed by PehA, are reported in Table 3.2.

Table 3.2. The binding modes of compound 1 to the enzyme. The red line indicates the reducing end of the hexagalacturonides. The numbers in parenthesis refer to each mode of binding, while the asterisks mark those binding modes that are referred as productive, since their corresponding hydrolysis products have been experimentally detected. The binding mode 1.1, 1.2 and 1.3 refer to ES complexes in which the substrate molecules are bound to the enzyme in a such way that their non reducing ends are oriented toward the N-terminal of the protein, while the reducing ends are directed toward the C-terminal portion. The binding mode 1.4, 1.5 and 1.6 refer to ES complexes in which the substrates are oriented in the opposite manner.

Several binding modes are considered to be productive, since their corresponding hydrolysis products have been experimentally detected. According to the productive binding mode 1.1, the subsites -2, -1 and +1 of the enzyme accommodate an unmethylated saccharide unit. In the non-productive binding mode 1.2, a methylated saccharide unit should instead occupy the subsite -2. Therefore, it is possible to suggest that the presence of a methylated saccharide unit at subsite -2 would exclude the occurrence of any ES complex. In the non-productive binding modes 1.3, beside the location of a methylated saccharide unit at the subsite -2, a methylated galacturonate moiety should occupy the subsite -1. Such binding can be also considered very unlikely, since, as shown by the crystal structures of the endoPG of *S. purpureum* with galacturonates (13), the carboxylate of galacturonic moiety located at position -1 is involved in the electrostatic interactions with several conserved, positively charged amino acids. These interactions significantly contribute to the stability of the ES complex, since the crystallographic evidences suggest that the sugar ring bound at the subsite -1 should adopt a distorted half-chair conformation. This distorted conformation is thought to be involved in the stabilization of the oxocarbenium-like ion supposed to be present in the transition state (29). Therefore, a methylated saccharide unit cannot be located at subsite -1, confirming that the binding mode 1.3 does not occur.

Considering the binding mode of compound 1 to the enzyme in the opposite orientation (Table 3.2), the unique productive binding mode could be the one referred as 1.4. In fact, the hydrolysis products related to the binding modes 1.5 and 1.6 have not been detected. According to the binding mode 1.4, a methylated saccharide unit should be accommodated at the subsite -2, in a similar manner to that described in the binding mode 1.2. Since it was shown that the subsite -2 does not accommodate a methylated saccharide unit, it is possible to conclude that the binding mode 1.4 does not occur.

Taken together these findings, we conclude that the OGAs are bound to PehA in such a way that their non reducing ends are orientated toward the N-terminal of the protein, while the reducing ends are directed toward the C-terminal portion.

With regard to the hydrolysis of compound 3, three productive binding modes can be described (Table 3.3). The binding modes labeled as 3.1 and 3.2 are supposed to be favoured, since they allow to accommodate three consecutive unmethylated saccharide units at the subsites -2, -1 and +1. Instead, the binding mode 3.3, that implies the

binding of a methylated saccharide unit at the subsite -2, is considered very unlikely to occur.

The ESI-TOF-MS analysis revealed that compound 1 was hydrolysed more slowly than the compound 3, since it was present in the hydrolytic mixture after 1 hour of enzymatic incubation at 37 °C (Figure 3.11). This finding suggests that the compound 3 is a preferred substrate compared to compound 1. In order to explain their different susceptibility to the enzymatic hydrolysis, their binding modes to the enzyme can be compared (see Table 3.1 and Table 3.2). A larger number of the productive ES complexes can be described for compound 3, thus explaining its faster degradation. It should be emphasized that compound 1, that has three methyl groups, is more volatile than compound 3, which displays only two methylated carboxylates. In the ESI-TOF-MS analysis, their detection depends on the efficiency of the ionization, which is influenced by the volatility. Therefore, ESI-TOF-MS was reliable only for a qualitative assessment of the mixtures analysed.

Table 3.3. The binding modes of compound 3 to the enzyme. The red line indicates the reducing end of the hexagalacturonides. The numbers in parenthesis refer to each mode of binding while the asterisks mark those binding modes that are referred as productive, since their corresponding hydrolysis products have been experimentally detected

The results regarding the mode of action of PehA on compounds 1 and 3 and the inability of the enzyme to hydrolyse compounds 2 and 4 lead us to conclude that at least three consecutive unmethylated saccharide units must be present in the OGAs, in order to have a productive ES complex.

Overall the study of the mode of action indicates that the catalytic action of PehA involves a minimum of four subsites. Among them, the three subsites referred to as -2, -1 and +1 are responsible for crucial contacts between the enzyme and the substrate. Pertubation of these interactions excludes the occurrence of any ES complex, as demonstrated by the fact that methylated saccharide units cannot be accommodated at these subsites. The fourth subsite may be either -3 or +2. The interactions between the enzyme and the substrate that occur at both subsites -3 and +2 are less critical for the formation of a productive ES complex, since they can accommodate a methylated saccharide unit, as demonstrated by the hydrolysis product of compound 3 (see Table 3.3).

It is possible to suggest that the preferred fourth subsite is -3, since the trimer was found in a large excess among the end products of the PGA and of decagalacturonide hydrolysis. This conclusion is supported also by the model of substrate binding proposed for the endoPGs of *S. purpureum* (13). As already pointed out, the galacturonate bound to the subsites -1 adopts a distorted half-chair conformation and therefore the occupancy of the subsite –3 may contribute to the stabilization of the overall ES complex.

3.5 Bibliography

(1) E.S. Martens-Uzunova, J.S.Zandleven, J.A.Benen, H. Awad, H.J. Kools, G. Beldman, A.G. Voragen, J.A. van den Berg, P.J. Schaap, A new group of exo-acting family 28 glycoside hydrolases of *Aspergillus niger* that are involved in pectin degradation, Biochem. J. 5 (2006) (Epub ahead of print)

(2) H.J. Bussink, F.P. Buxton, B.A. Fraaye, L.H. de Graaff, J. Visser, The polygalacturonases of *Aspergillus niger* are encoded by a family of diverged genes, Eur. J. Biochem. 208 (1992) 83-90.

(3) J.P. Wubben, W. Mulder, A. ten Have, J.A. van Kan, J. Visser, Cloning and partial characterization of endopolygalacturonase genes from *Botrytis cinerea*, Appl. Environ. Microbiol. 65 (1999) 1596-1602.

(4) L. Parenicova, J.A.E. Benen, H.C.M. Kester, J. Visser, pgaE encodes a fourth member of the endopolygalacturonase gene family from *Aspergillus niger*, Eur. J. Biochem. 251 (1998) 72-78.

(5) J.A.E. Benen, H.C.M. Kester, J. Visser, Kinetic characterization of *Aspergillus niger* N400 endopolygalacturonases I, II and C, Eur. J. Biochem. 259 (1999) 577– 585.

(6) L. Parenicova, H.C.M. Kester, J.A.E. Benen, J. Visser, Characterization of a novel endopolygalacturonase from *Aspergillus niger* with unique kinetic properties, FEBS Lett. 467 (2000) 333–336.

(7) L. Parenicova, J.A.E. Benen, H. Kester, J. Visser, endoPGaA and endoPGaB encode two constitutively expressed endopolygalacturonases of *Aspergillus niger*, Biochem. J. 345 (2000) 637–644.

(8) E. Bonnin, A. Le Goff, R. Körner, G-J.W.M. van Alebeek, T.M.I.E. Christensen, A.G.J. Voragen, P. Roepstorff, A. Caprari, J.-F. Thibault, Study of the mode of action of endopolygalacturonase from *Fusarium moniliforme*, Biochim. Biophys. Acta 1526 (2001) 301–309.

(9) I. Kars, G.H. Krooshof, L. Wagemakers, R. Joosten, J.A. Benen, J.A. van Kan, Necrotizing activity of five Botrytis cinerea endopolygalacturonases produced in Pichia pastoris, Plant J. 43 (2005) 213-225.

(10) R.P. de Vries, J. Visser, *Aspergillus* enzymes involved in degradation of plant cell wall polysaccharides, Microbiol. Mol. Biol. Rev. 65 (2001) 497-522.

(11) H.C. Kester, D. Magaud, C. Roy, D. Anker, A. Doutheau, V. Shevchik, N. Hugouvieux-Cotte-Pattat, J.A. Benen, J. Visser, Performance of selected microbial pectinases on synthetic monomethyl-esterified di- and trigalacturonates. J Biol. Chem. 274 (1999) 37053-37059.

(12) G.J. Davies, K.S. Wilson, B. Henrissat, Nomenclature for sugar-binding subsites in glycosyl hydrolases, Biochem. J. 321 (1997) 557-559.

(13) T. Shimizu, T. Nakatsu, K. Miyairi, T. Okuno, H. Kato, Active-site architecture of endopolygalacturonase I from *Stereum purpureum* revealed by crystal structures in native and ligand-bound forms at atomic resolution, Biochemistry 41 (2002) 6651–6659.

(14) Y. Van Santen, J.A.E. Benen, K.-H. Schroter, K.H. Kalk, S. Armand, J. Visser, B.W. Dijkstra, 1.68 Å crystal structure of endopolygalacturonase II from *Aspergillus niger* and identification of active site residues by site-directed mutagenesis, J. Biol. Chem. 274 (1999) 30474– 30480.

(15) S. Armand, M.J.M. Wagemaker, P. Sanchez-Torres, H.C.M. Kester, Y. van Santen, B.W. Dijkstra, J. Visser, J.A.E. Benen, The active site topology of *Aspergillus niger* endopolygalacturonase II as studied by site-directed mutagenesis, J. Biol. Chem. 275 (2000) 691– 696.

(16) S. Pages, W.H. Hejine, H.C.M. Kester, J. Visser, J.A.E. Benen, Subsite mapping of *Aspergillus niger* endopolygalacturonase II by site-directed mutagenesis, J. Biol. Chem. 275 (2000) 29348-29553.

(17) G. van Pouderoyen, H.J. Snijder, J.A. Benen, B.W. Dijkstra, Structural insights into the processivity of endopolygalacturonase I from *Aspergillus niger*, FEBS Lett. 554 (2003) 462-466.

(18) S. Pages, H.C. Kester, J. Visser, J.A. Benen, Changing a single amino acid residue switches processive and non-processive behaviour of *Aspergillus niger* endopolygalacturonase I and II, J. Biol. Chem. 276 (2001) 33652-33656.

(19) T.C. Herlache, A.T. Jr Hotchkiss, T.J. Burr, A. Collmer, Characterization of the *Agrobacterium vitis* PehA gene and comparison of the encoded polygalacturonase with the homologous enzymes from *Erwinia carotovora* and *Ralstonia solanacearum*,

Appl. Environ. Microbiol. 63 (1997) 338-346.

(20) D. Bellicampi, G. Salvi, G. De Lorenzo, F. Cervone, V. Marfà, S. Eberhard, A. Darvill, P. Albersheim, Oligogalacturonides inhibit the formation of roots on tobacco explants, Plant J. 4 (1993) 207-213.

(21) F. Guillaumie F, S.F. Justesen, K.E. Mutenda, P. Roepstorff, K.J. Jensen, O. R. Thomas, Fractionation, solid-phase immobilization and chemical degradation of long pectin oligogalacturonides. Initial steps towards sequencing of oligosaccharides, Carbohydr. Res. 341 (2006) 118-129.

(22) M.H. Clausen, R. Madsen, Synthesis of Hexasaccharide Fragments of Pectin, Chem. Eur. J. 9 (2003) 3821-3832.

(23) N.C. Price, L. Stevens, Fundamentals of Enzymology – Second Edition, Oxford University Press, New York, 1989.

(24) H. Schwaiger, P.J. Oefner, C. Huber, E. Grill, G.K. Bonn, Capillary zone electrophoresis and micellar electrokinetic chromatography of 4-aminobenzonitrile carbohydrate derivatives , Electrophoresis 15 (1994) 941-952.

(25) C. Campa, A. Oust, G. Skjak-Braek, B.S. Paulsen, S. Paoletti, B.E. Christensen, S. Balance, Determination of average degree of polymerisation and distribution of oligosaccharides in a partially acid-hydrolysed homopolysaccharide: a comparison of four experimental methods applied to mannuronan, J. Chromatogr. A 1026 (2004) 271-278.

(26) D.R. Baker, Capillary Electrophoresis, John Wiley & Son (Eds.), New York, 1995.

(27) G.J. van Alebeek, K. van Scherpenzeel, G. Beldman, H.A. Schols, A.G. Voragen, Partially esterified oligogalacturonides are the preferred substrates for pectin methylesterase of *Aspergillus niger*, Biochem. J. 372 (2003) 211-218.

(28) M.A. Williams, J.A. Benen, A novel enzyme activity involving the demethylation of specific partially methylated oligogalacturonides, Biochem. J. 367 (2002) 511-515.

(29) C.S. Rye, S.G. Withers, Glycosidase mechanisms, Curr. Opin. Chem. Biol. 4 (2000) 573-80.

Chapter 4. Insights in the tertiary structure of PehA

4.1 Introduction

The six fungal and bacterial endopolygalacturonases (endoPGs), whose crystal structures have been determined (Table 4.1), exhibit a right handed β-helix fold (1-6). The β-helix fold was first found in the three dimensional structure of the *Erwinia chrysanthemi* pectate lyase C (7). Since then, the same structural motif has been observed in the structures of other carbohydrate-binding proteins, including a rhamonogalacturonase from *Aspergillus aculeatus* (8), a chondroitinase B from *Pedobacterium heparium* (9),a ι-carrageenase from *Alteromonas fortis* (10) and a dextranase by *Penicillium linoleum* (11). These structures are characterized by parallel β-strands, coiled around a vertical axis to form a large right-handed cylinder. A canonical right-handed β-helix protein includes a number of complete helical rungs ranging from 7 to 10, containing a minimum of 22 amino acids *per* turn (12). Each rung displays three short parallel β-strands, designated B1, B2, and B3, connected by turns, referred as T1, T2 and T3.

Table 4.1. The list of the endoPGs, whose three dimensional structures have been determined.

Source	*PDB code*	*Reference*
endoPG from *Erwinia carotovora*	1BHE	1
endoPGII from *Aspergillus niger*	1CZF	2
endoPG from *Fusarium moniliforme*	1HG8	3
endoPG from *Aspergillus aculeatus*	1IA5; 1IB4	4
endoPGI from *Aspergillus niger*	1NHC	5
endoPG from *Stereum purpureum*	1K5C; 1KCC; 1KCD	6

A schematic description of the rung found in right handed parallel β-helices is depicted in Figure 4.1.
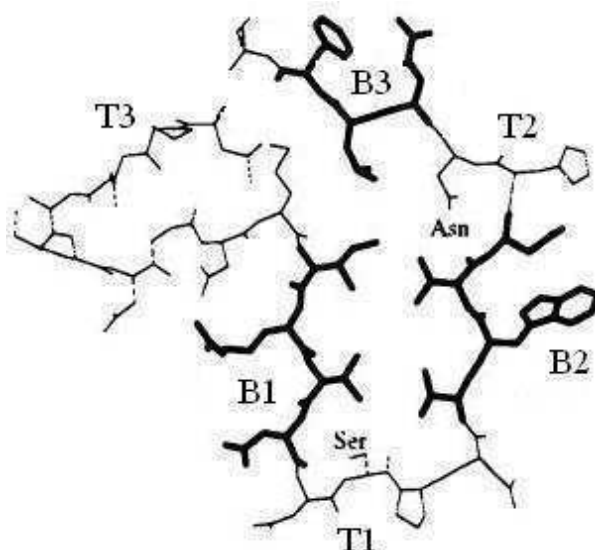
Figure 4.1. Cross-sectional view of a single helical coil of the right-handed β-helices of pectate lyase C (pdb code: 1pcl). B1 and B2 strands contain four amino acids and are connected by a narrow bend, T1, which has a minimum length of four amino acids. The B3 consists of 3 amino acids and it is connected to B1 strand by a wide bend, T3, which has a minimum length of five amino acids. A very short β-turn, T2, composed of an asparigine and a small hydrophobic residue, connects B2 to the B3 strand. B1 and B2 form an antiparallel β-sandwich, while B3 is approximately perpendicular to B2. Due to the unusual pairing of B1 and B2, the cross section of right-handed parallel helices is L-shaped, rather than circular. The Figure is taken from M. Yoder *et al.* (21).

It is worthy of note that right-handed parallel β-helices, that differ from the above described canonical fold, have been also found. For example, right-handed helices that contains two or four β-strands for each turn have been reported (13, 8). Moreover, β-helices longer than canonical ones have been observed in the crystal structures of *Salmonella typhimurium* P22 tailspike protein (TSP), which has 13 coils (14), and *Bordetella pertussis* P.69 pertactin, which has 16 coils (15). Although the known parallel β-helices vary in the number of complete rungs and in the lengths of the turn regions, the β-strand portions of the rungs have a basic pleating patterns and hydrogen bonding networks that are well conserved across the superfamily. Therefore, computational tools that predict the β-helices fold based on the protein sequence have been developed (16-18).

The analysis of the molecular structures of β-helices proteins revealed a preferential

amino acidic arrangement, inward or outward the core of the β-helix. The amino acids found in the interior of the parallel β-helices have mainly an hydrophobic character, although polar groups have been found too. On the exterior, all types of amino acids occur. The polar or charged groups are generally exposed to the solvent, while the hydrophobic ones are usually covered by surface loops, which protrude from the central cylinder. These loops display a functional role, being involved in the substrate binding. Since loops are the most variable secondary structure elements, their sequence divergence is responsible for the evolution of the substrate specificity (19).

The β-helices structures are stabilized by an extensive hydrogen bond network between parallel β-strands as well as by highly ordered stacking of side chains in the interior of the cylinder (20). There are three predominant types of stacks: stacks of planar aromatic residues, stacks of aliphatic residues, and hydrogen bonded stacks such those involved in the asparagine ladders. Tyrosine, tryptophan and phenylalanine, that are packed face to face, form stacks within and outside the parallel β-helix. Aliphatic stacks are mostly found inside the β-helix and are stabilized by hydrophobic interactions and efficient packing. Stronger stacks are composed by the same residue or by diverse residues with similar side chain conformation. An unusual stacking arrangement of asparagines, termed the asparagine ladder, is present in the turn T2 (Figure 4.2).
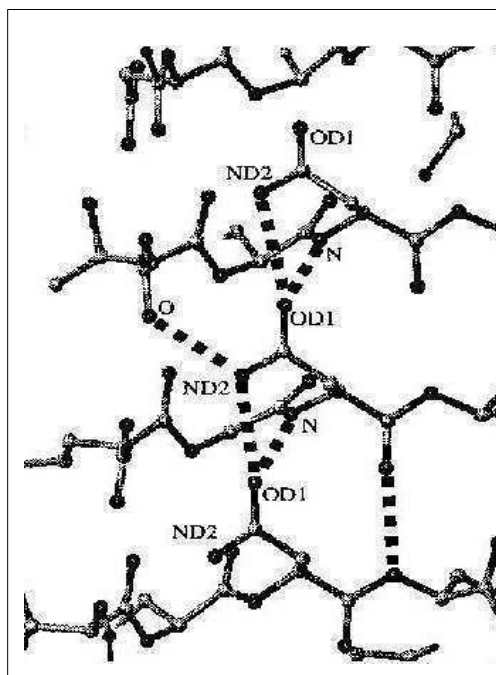

Figure 4.2. The hydrogen bonding pattern in the lyases at the asparagine ladder. The Figure is taken from J. Jenkins *et al.*, (20).

In this new type of repetitive secondary structure, all the potential hydrogen bonding donors and acceptors of either the main chains or the side chains of asparagine residues that belong to stacked β-strands, are engaged (21).

As concerns the extremes of the β-helices, the N-terminal end of the parallel β-helix domain is capped by an α-helix, weakly amphipathic in character, that is structurally conserved. On the other hand, no structural conservation occurs in the capping of the C-terminal end of the cylinder although this part of the protein is often folded in extended conformation.

To complete the description of the conserved elements that stabilize the β-helix fold, the presence of disulphide bonds must be taken into account. The amino acid sequence alignment of 115 endoPGs of different sources revealed that in bacterial endoPGs the cysteine residues are not conserved in their respective position. On the contrary, the position of the cysteine residues among fungal endoPGs are very well conserved (22). The endoPG of *Erwinia carotovora*, that is the only bacterial endoPG whose three dimensional structure has been determined, has two disulphide bonds (1). Indeed, the known crystal structures of fungal endoPGs known revealed the presence of three (5) o four disulphide bonds (2-4, 6). Both fungal and bacterial endoPGs exhibit the conservation of a disulphide bond in their N-terminal regions.

The putative β-helix fold of the PehA was investigated using a variety of approaches. Efforts to obtain suitable crystals for determining three dimensional structure of PehA by X-ray diffraction will be discussed. Namely, crystallization studies and diffraction tests will be reported (Section 1). Then, the topology of the disulphide bonds has been assigned using a mass spectrometry technique (Section 2). Lastly, a three dimensional molecular model of PehA has been generated by computational methods based on profile-profile sequence alignments and fold recognition algorithms (Section 3).

Section 1. Crystallization and diffraction tests

4.1.1 Materials and Methods

4.1.1.1 Principles of protein crystallization

4.1.1.1.1 Crystal definition

Crystal is a solid having regularly repeating internal arrangements of atoms, molecules or ions and is a result of their cooperative, self-promoted, three-dimensional ordering (23). Protein crystals are macromolecular crystals, that rarely exceed a millimeter on an edge and are generally much smaller. They are extremely fragile and have very weak mechanical properties.

4.1.1.1.2 Theory of protein crystallization

If a protein-solvent system is at a concentration of the solute below its solubility limit for a particular set of conditions, it is considered to be in the undersaturation state (Figure 4.3). If it is exactly at the limit, then the solution is called saturated. At this unique concentration of the material in solution, the rates of loss and gain of both solid and solution phases are equal, and the system is at the equilibrium. Crystals, therefore, cannot grow from a solution that is simply saturated. When there are more protein molecules in a solution than the solubility limit would allow, the solution is called supersaturated. We would expect molecules in solution to rejoin the solid state due to a thermodynamic driving force that pushes the system back to the solubility limit. In this way the system will tend to re-establish the equilibrium. This is from non-equilibrium, saturated solutions that crystals are grown. If conditions are adjusted in a tranquil manner, in the absence of any pre-existing solid state, and if no sudden impulses of energy are provided to the system, then the supersaturation will result. Certain amount of activation energy is required to initiate the formation of the solid state. There are two approaches to helping the system to overcome the activation barrier: putting energy into the system by increasing the supersaturation of the solution, or effectively lowering the

energy barrier by seeding the solution with pre-existing crystals. If no adequate amount of energy becomes available, the system will remain in metastable, non-equilibrium state of supersaturation (Figure 4.3). No nuclei will be form and no crystals will grow. In the metastable zone crystals may grow only if the nuclei are present. The high probability of spontaneous critical nuclei formation in solution is present in the extreme supersaturation zone called labile region. The ideal approach for protein crystallization would be to induce nuclei at the lowest level of the labile region. As these few nuclei begin to grow, the supersaturation will fall and the solution will gradually enter the metastable region where the crystal growth will be slow and ordered. Therefore, few and large crystals will be formed.
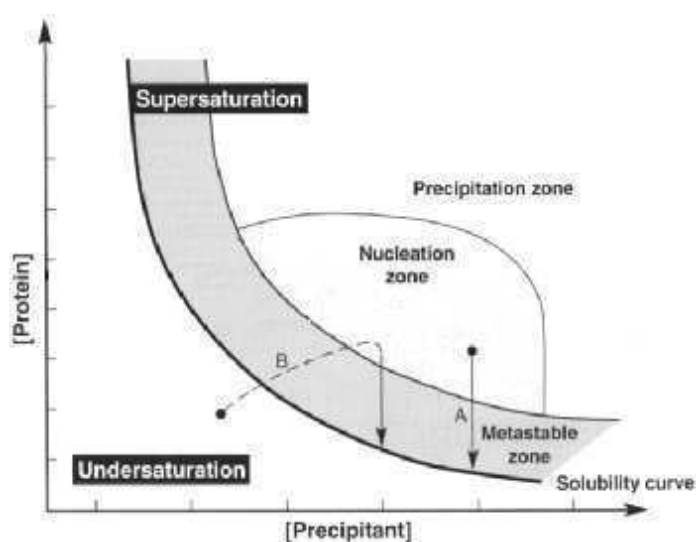
Figure 4.3. The phase diagram. (A) Vapor diffusion. (B) Microbatch.

4.1.1.1.3 Methods for promoting supersaturation

4.1.1.1.3.1 Vapor diffusion. The hanging drop vapor diffusion technique is the most popular method for the crystallization of macromolecules (23). A drop composed of a mixture of sample and reagent is placed in vapor equilibration with a liquid reservoir of reagent. The sample drop is dispensed onto a siliconized surface, in this way it is spatially separated from the reservoir solution. When the drop is suspended from some surface, the method in named hanging drop (Figure 4.4A); when the drop is supported

by some surface, the method is called sitting drop (Figure 4.4B). The crystallization agent concentration into the drop is lower than that required for the formation of the crystal. Over time, equilibrium takes place through the vapor phase. The vapor phase equilibrium is reached when the reagent concentration in the drop is approximately the same as that in the reservoir. Because the volume of the reservoir is much bigger than that of the droplet, the ultimate concentration of the precipitant in the drop of mother liquor will correspond to the initial, experimentally established conditions of the reservoir. As water vapor leaves the drop and eventually ends up in the reservoir, the volume of the drop decreases with an increase of protein-reagent concentration. Therefore, vapor diffusion simultaneously provides at least two ways to drive the system toward supersaturation: increase in precipitant concentration and increase in protein concentration (Figure 4.3A).
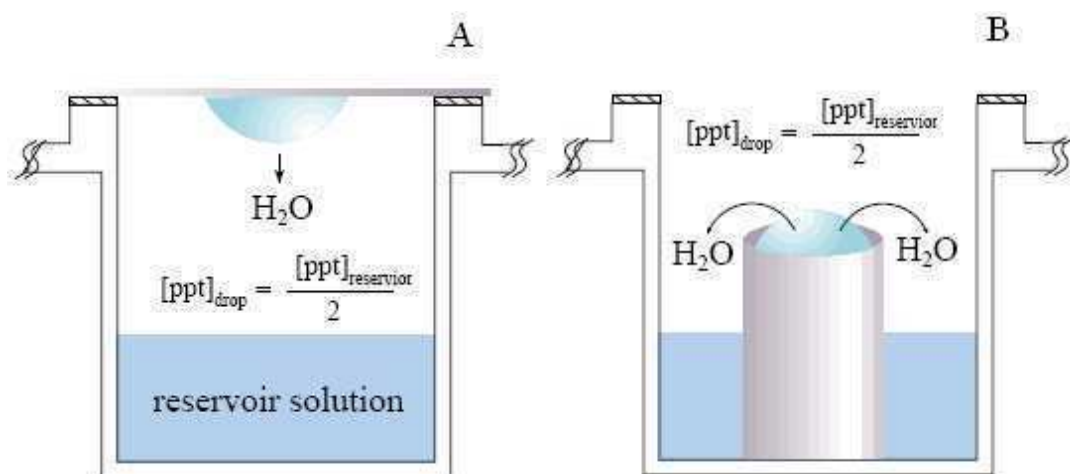
Figure 4.4. Process of vapor diffusion. (A) Hanging drop. (B) Sitting drop. The images are taken from www.hamptonresearch.com.

4.1.1.1.3.2 Microbacth crystallization technique. The crystallization of proteins under a thin layer of paraffin oil was originally described by Chayen *et al.* (24). In this technique a drop of undersaturated protein solution combined with a precipitant is pipetted under a small layer of paraffin oil. The precipitant alters the protein solubility or the electrolyte properties of the solution giving immediately saturated mother liquor. Thus, the protein and the precipitant agent are mixed at their final concentrations at the start of the experiment. There is a less exploration of the phase diagram (Figure 4.3B) and several microbatch trials may be required to replace a single vapor diffusion

experiment (25). The oil generally used is a mineral oil of branched paraffins in the C20+ range and allows for little to no diffusion of water through the oil (Figure 4.5A). All of the reagents involved in the crystallization are present at a specific concentration and no significant variation of the initial concentration of the protein nor of the reagents can occur in the drop. Since oil and water are essentially immiscible, evaporation of water from the drop is negligible. However, slow evaporation occur (there is not absolute immiscibility). In order to control the rate of vapor diffusion in the microbatch setup, it had been proposed the use of different types of oil (26). Silicon oil consists of a polymer of repeating dimethylsiloxane units, that permit a freely diffusion of water. A mixture of paraffin and silicon oils allows partial diffusion. This is what is called modified microbatch (Figure 4.5B). Combination of the microbatch and diffusion methods are emerging as useful tool in the optimization of the crystallization conditions. A possible combination includes an oil barrier (usually, a mixture of paraffin and silicon oils) over a reservoir of a vapor diffusion trial in order to slow down the equilibrium rate and thus to approach supersaturation more slowly (27) (Figure 4.5C).
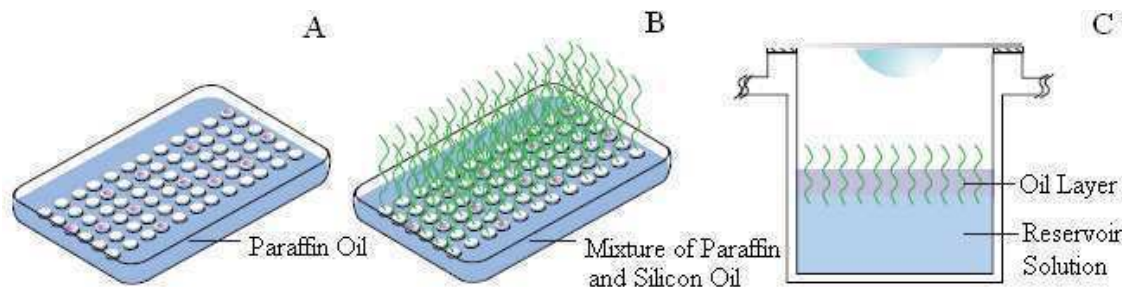

Figure 4.5. Microbatch crystallization technique. (A) Microbatch setup. (B) Modified microbatch setup. (C) Combination of microbatch and diffusion methods. The images are taken at the following url: www.hamptonresearch.org.

4.1.1.2 Crystallization experiments

4.1.1.2.1 Production and purification of PehA

The endopolygalacturonase of *Burkholderia cepacia* (PehA) was heterologously expressed and purified as previously reported (Chapter 2, Par. 2.2.2.1).

4.1.1.2.2 In house crystallization screening

Crystallization was performed by the hanging drop vapor diffusion method using the 24-well plates Crystallization Tool (Nextal, Canada). Initial crystallization conditions were established at 20 °C and 4 °C, using crystal screening kits Crystal Screen I and II (Hampton Research, USA) (28). Each drop was prepared by mixing equal volumes, 1 µL, of the protein sample at different concentrations, 5.6 mg/mL or 2.8 mg/mL, and the precipitant buffer. Each hanging drop was equilibrated against 750 µL precipitant buffer. In order to improve crystal quality, seeding techniques were attempted. In these techniques, previously nucleated crystals were introduced into new drops equilibrated at lower levels of supersaturation. More precisely, based on the size of the seeds, seeding techniques are classified into microseeding, that implies the transfer of submicroscopic seeds, and macroseeding, where the transfer involves a single crystal of 5-50 µm (29). Moreover, other optimization strategies have been explored including changes of the ratio of precipitant and protein concentrations into the drop, pH adjustment and screening of additives.

The microbatch crystallization screening was performed at 20 °C, using a 96 wells microbatch plate (Hampton Research). 1 µL of protein solution at 5.6 mg/mL and 1 µL of the precipitant buffer were mixed under 250 µL of paraffin oil.

4.1.1.2.3 The high-throughput crystallization screening

The high-throughput (HTP) crystallization screening approach was exploited at the high-throughput crystallization laboratory of the   Hauptman Woodward Medical Research Institute, Buffalo NY, USA. The crystal-growth screening experiment was performed in 1536-well microassay plate, using the microbatch crystallization technique. The total volume of drop was 0.2 µL, obtained mixing an equal volume of

the precipitant buffer and the PehA solution at a concentration of 2.8 mg/mL. The images of each crystallization condition were provided at regular times: prior and after the drops setup, weekly for a month period.

4.1.1.3 Diffraction tests

Diffraction tests on the obtained crystals were performed at the following beam lines: the XRD1 beam line of the ELETTRA Synchrotron Light Laboratory (Trieste, Italy); the ID14 beam line of the European Synchrotron Radiation Facility (Grenoble, France); the X10SA beam line of the Swiss Light Source (Villigen, Swiss).
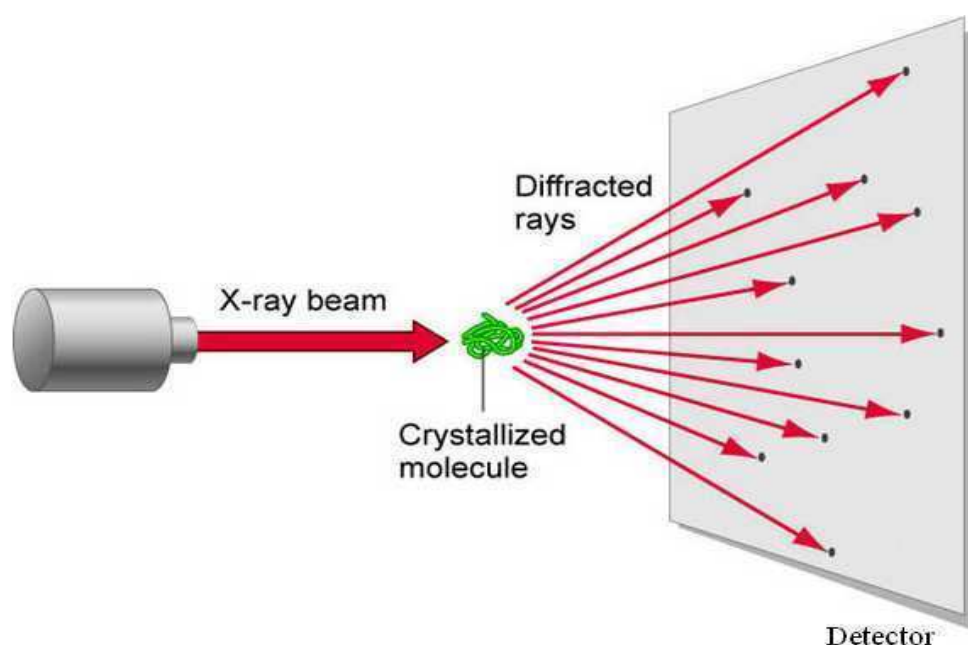

Figure 4.6. A graphical view of a crystal protein diffraction experiment.

The crystals were mounted on nylon loops, transferred to a cryo-protecting solution and subsequently flash cooled in a liquid nitrogen stream at – 173 °C, using an Oxford Cryosystems cooling device (Oxford Cryosystems Ltd., UK).

A graphical view of a crystal protein diffraction experiment is reported in Figure 4.6.

4.1.2 Results

4.1.2.1 Crystallization

<u>4.1.2.1.1 In house crystallization screening</u>

From a preliminary crystallization screening, spherulites were obtained after three days from 0.2 M di-ammonium hydrogen citrate, 20% PEG 3350 pH 5.0 at a protein concentration of 5.6 mg/mL (Figure 4.7A). To decrease the number of crystals in the drop, the same crystallization conditions were tested at a lower temperature, leaving the plate at 4 °C (figure 4.7B). Single needle-shaped microcrystals appeared after one week. Their dimensions were approximately of 100 × 5 × 5 μm (Figure 4.7C). To improve the crystal size, microseeding experiments were performed. In these experiments either protein or precipitant concentrations were lowered up to 3.0 mg/mL and 15%, respectively. In addition, two different pH conditions of the pre-equilibrated reservoir solutions were tested, pH 5.0 (Figure 4.7D) and pH 3.0 (Figure 4.7E). Better crystals, even if still small in size (100 × 30 × 30 μm), were obtained after 15 days from seeding experiments, performed at 4 °C using 0.2 M di-ammonium hydrogen citrate, 15% PEG 3350 pH 3.0, as reservoir solution (Figure 4.7E).

A different crystallization condition that provided PehA crystals was also found. Crystals of 500 × 50 × 50 μm were obtained after 10 days from 0.1 M citric acid, 30% PEG 6000, 1.0 M lithium chloride, pH 5.0, at a protein concentration of 4.0 mg/mL (Figure 4.7F and Figure 4.7G). An extended screening around these crystallization conditions was performed, however these crystals resulted to be not reproducible.

<u>4.1.2.1.2 The high-throughput crystallization screening</u>

A sample of PehA at a concentration of 2.0 mg/mL was sent to the highly automated crystallization laboratory at the Hauptman Woodward Medical Research Institute, Buffalo NY, USA. A new crystallization conditions, not yet tested in house, was found as result of the HTP: a 0.5 M ammonium sulphate, 30 % (v/v) MPD, 0.1 M HEPES pH 7.5. In fact, as shown in Figure 4.7H, a large crystal is present. The image of the same drop just after the drop deposition is reported in Figure 4.7I. It clearly indicates that
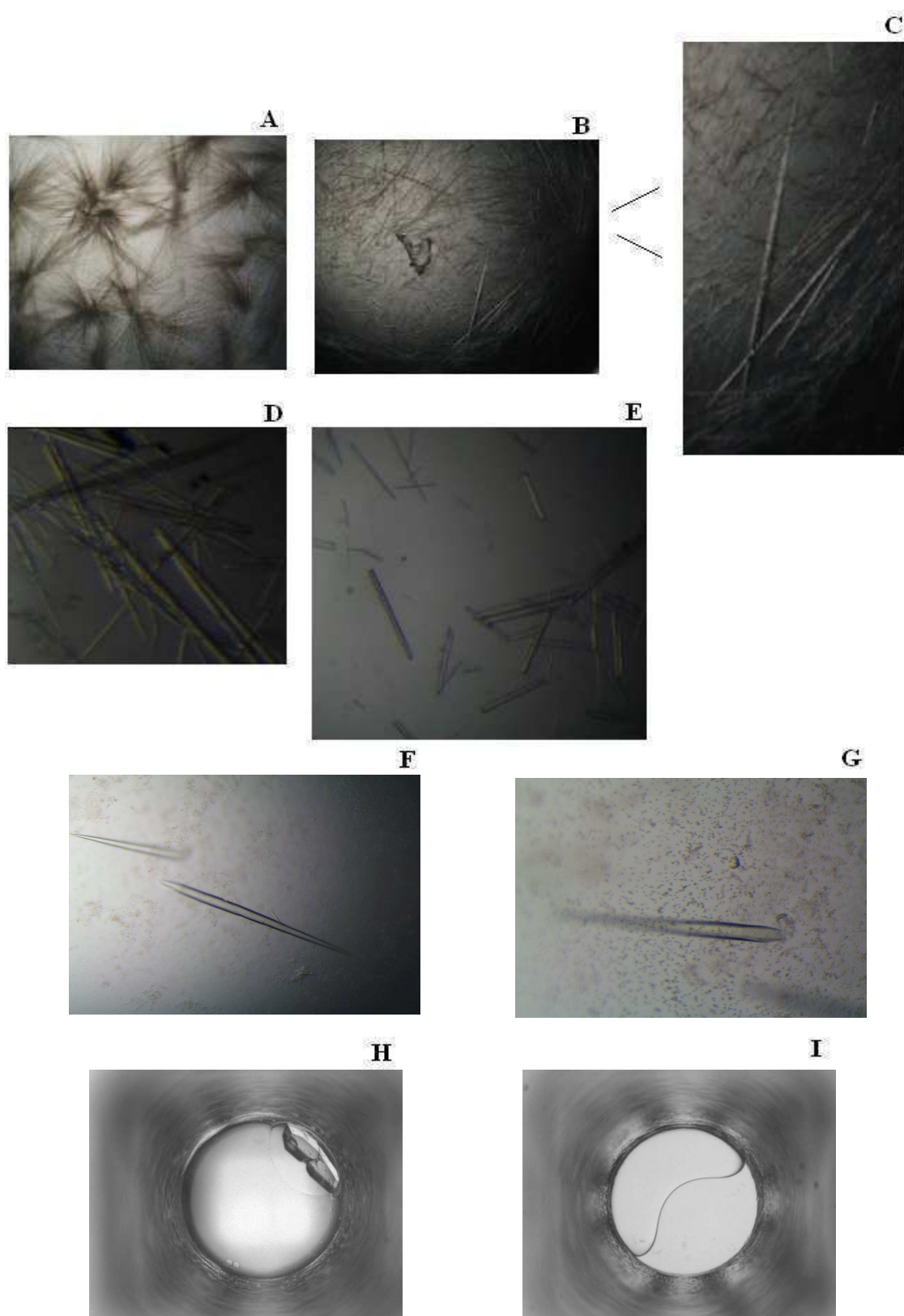
Figure 4.7. The images of the obtained PehA crystals, as described in the text.

no contaminants were initially present into the well. The crystallization condition suggested by the HTP crystallization screening was tested in our laboratory either using the hanging drop vapor diffusion technique or the microbatch technique. Unfortunately, no crystals could be obtained.

4.1.2.2 Diffraction tests

The microcrystals obtained from 0.2 M di-ammonium hydrogen citrate, 15% PEG 3350 pH 3.0 (Figure 4.7E) were fished out from the drop using nylon loops. Then they were transferred to a cryo-protecting solution, containing 0.2 M di-ammonium hydrogen citrate, 30% PEG 3350, 20% glycerol pH 3.0, and immediately flash-cooled in a liquid nitrogen gas stream at -173 °C. Their diffraction quality was tested at several beam lines, as reported in Materials and Methods (Par. 4.1.1.3). Diffraction tests at SLS were performed during the 4th NCCR Practical Course in Synchrotron Data Acquisition Techniques in Macromolecular Crystallography, that I attended in October 2005. The diffraction quality of PehA microcrystals was quite poor, in all the tests done. The diffraction pattern confirmed that the obtained microcrystals are protein microcrystals diffracting weakly up to 5.0 Å.

The crystals obtained from 0.1 M citric acid, 30% PEG 6000, 1.0 M lithium chloride, pH 5.0 were tested on the beam line XRD1 beam line of the ELETTRA Synchrotron Light Laboratory (Trieste, Italy). They weakly diffracted up to 4.5 Å.

Section 2. Disulphide bond mapping

4.2.1 Materials and Methods

The mapping of disulphide bonds was done according to procedures reported by Shimizu *et al.* (30). The endopolygalacturonase of *Burkholderia cepacia* (PehA) was prepared as previously reported (Chapter 2, Par. 2.2.2.1). In this experiment, PehA (30 μg) in 100 μL of 0.5 M Tris-HCl buffer (pH 8.5), containing 2 μL of 4-vinylpyridyne and 10 mM EDTA was incubated for 4 hours in the dark at room temperature. Then, the reaction mixture was purified by reverse-phase HPLC, in order to exclude reagents, and lyophilized. PehA fraction was redissolved in 8 M urea, containing 50 mM ammonium hydrogen carbonate (NH_4HCO_3) and denatured for 2 hours at 37 °C. After denaturation, a sevenfold volume of 50 mM NH_4HCO_3 containing 1 μg trypsin was added to the reaction mixture. The reaction was incubated for 4 hours at 37 °C and then lyophilized. The tryptic digest of pyridyethylated PehA was dissolved in water to a total volume of 20 μL and the solution containing tryptic peptides was loaded on $ZipTipC_{18}$ pipette tips (Millipore). The peptides were washed with water to remove salts and then were eluted with 3 μL of MALDI matrix solution (10 mg of α-cyan-4-hydroxycinnamic acid in 1 mL of 70% acetonitrile and 0.1% trifluoroacetic acid in water) and aliquots of 1 μL were spotted on a MALDI plate. Matrix assisted laser desorption ionization mass spectrometry (MALDI-MS) spectra were acquired on an Applied Biosystems/MDS Sciex 4800 MALDI TOF/TOF Analyzer operated in Linear or Reflectron Positive Ion Mode.

4.2.2 Results

To determine the disulphide bridges pattern, the tryptic digest of pyridylethylated PehA was analysed by MALDI-MS. On the other hand, the prediction of the peptides produced by the tryptic digestion of PehA was carried out online, using the server PeptideMass (http://www.expasy.ch/tools/peptide-mass.html). The result is reported in Table 4.2.

Table 4.2. The list of peptides produced by the tryptic digestion of PehA with all cysteine residues in reduced form, predicted by the online server PeptideMass . The cysteine residues are labeled in red.

No.	AVERAGE MASSES	POSITION	PEPTIDE SEQUENCE
T1	6327.0075	-1-62	MATCTPQWSSSASTNTTNLQNAIQQCAASGTSSSPGLVDLASNNGISTAVITSVNLANNIVLK
T2	389.4717	63-65	LEK
T3	565.6898	66-70	GFTLK
T4	4847.2401	71-119	GSPAQPSSGAMLTGSNLSNLTITGTGAIDGDGQDYWPAAVGQNNTARPK
T5	2021.2772	120-139	LIAITGSNLQIGSNFTDAGK
T6	2107.3708	140-160	SQSIVAFPSSSNATGSALIIR
T7	445.4955	161-164	NSPK
T8	1090.2182	165-174	EQLVIESGSK
T9	1732.9359	175-189	NVTIDGVWIYANPNR
T10	2516.72	190-214	NASGDDLAPNTDAIDIIGTQTATIK
T11	1506.6672	215-228	NCLLDTGDDDIAIK
T12	5715.2156	229-287	SNAGGAATSSVNVSHCVVGGGHGISIGGQEAAGTTLAKPGVSQVTVDTMQFSGTDYGYR
T13	260.3562	288-289	IK
T14	663.7051	290-295	TDQTAK
T15	1128.1835	296-306	DSGATTGVTYR
T16	624.7505	309-311	NTCMR
T17	3842.38	312-348	NVQQPFLFTYAYASGTGGALPIIANVTIDNVIATATK
T18	1714.0324	349-365	QQGAIIGLSNSLMGVPK
T19	1622.7287	366-382	SGDTGISITNSQISGGK
T20	4763.1976	383-431	AFSVTDGELQLGSHSSATTSTGSNGQVVGIPDTGATLSCPSSITIPAQI

The prediction was generated based on the hypothesis that all the cysteine residues are in the reduced form. The triptic fragments (T) produced by the proteolytic digestion of PehA will be named as reported in Table 4.2. Peaks T6 (m/z = 2107.41), T9 (m/z = 1732.13) and T10 (m/z = 2516.59) were observed in the MALDI-TOF spectrum of the pyridylethylated and digested PehA. Additional ion signals that could not be addressed as single tryptic peptides were observed. Two of these peaks (m/z = 5385.51; m/z = 7219.44) have been addressed as disulphide-bonded peptides (Figure 4.8). The first peak corresponds to the peptide T16 (m/z = 624.75), bonded through a disulphide bridge with T20 (m/z = 4763.19). The second peak corresponds to the peptides T11 (m/z = 1506.67) and T12 (m/z = 5715.21), that are joined together by an intermolecular disulphide bond. Therefore, the disulphide bridges are formed between Cys216 (T16) and Cys244 (T20) and between Cys309 (T11) and Cys421 (T12), respectively.
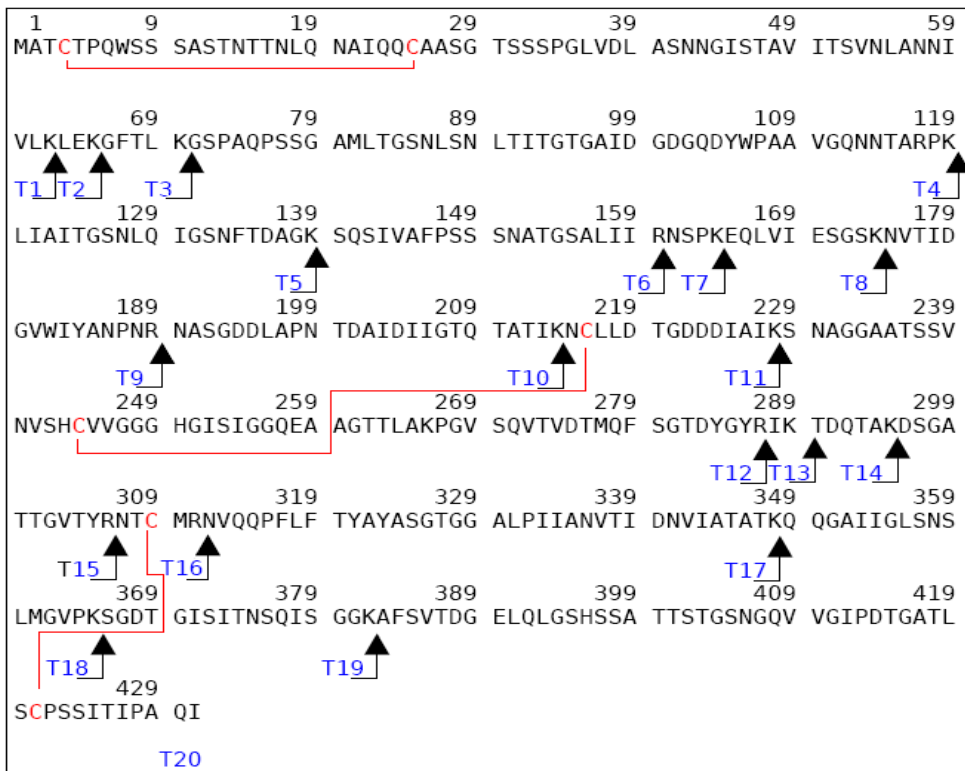
Figure 4.8. amino acid sequence of the per-PehA. The numbering is referred to sequence of mature PehA, excluding the first methionine added by the cloning. Trypsin cleavage sites are indicated by the arrows. The naming of the tryptic fragment is in accordance with that reported in Table 2. Disulphide bonds are indicated by the red lines.

No peak attributable to the peptide T1, that contains the remaining two cysteine residues, was found.

Section 3. Three dimensional model of PehA

4.3.1 Materials and Methods

4.3.1.1 Computational analysis of the amino acidic sequence

4.3.1.1.1 BetaWrapPro prediction

BetaWrapPro is a program for structural motif prediction and comparative modeling of the pectin lyase like right-handed parallel β-helix and of β-trefoil structural motifs, available online at the following url: http://betawrappro.csail.mit.edu. The BetaWrap algorithm assigns a raw score to each sequence. This score measures the compatibility of the sequence with a β-helical structure. A P-value is associated to this score. It gives a rough estimation of the likelihood that a randomly chosen sequence from the PDB that does not form the template fold would attain a similar score (16-18).

4.3.1.1.2 Sequence alignment

The multiple sequence alignments were generated using the server ClustalW, available online at the following url: http:// www.ebi.ac.uk/clustalw.

4.3.1.2 Modeling

4.3.1.2.1 Principles of threading

The most effective methods of protein structure and function predictions are based on establishing a homology between the protein of interest and an already characterized protein. The standard sequence-sequence comparison methods, however, rapidly lose sensitivity for sequence identity of 30% or less. The sensitivity of homology recognition can be improved by using information present in the families of protein sequences connected with detectable homology. In this approach, profile-profile algorithms use information present in sequences of homologous proteins to amplify the patterns defining the family.

The basic idea of the computational methods based on profile-profile sequence

alignmentsis is that there are a strictly limited number of different protein folds in nature, mostly as a result of evolution but also due to constraints imposed by the basic physics and chemistry of polypeptide chains. There is, therefore, a good chance that a protein which has a similar fold to the target protein has already been studied by X-ray crystallography or NMR spectroscopy. The protein sequence of an unknown structure is threaded through the backbone structures of a collection of template proteins (fold library). For each sequence-structure alignment, a "goodness of fit" score is calculated and expressed in terms of an empirical energy function.

The profile-based fold recognition approach was first described by Bowie, Lüthy and Eisenberg in 1991 (31). A profile of each known structure can be generated analysing each amino acid in the structure and simply labeling it according to whether it is buried in the core of the protein or exposed on the surface. The profile-based fold recognition derives a 1-D profile for each structure in the fold library and align the target sequence to these profiles. The term *threading* was first coined by Jones, Taylor and Thornton in 1992 (32) and originally referred specifically to the use of a set of inter-atomic distances to describe the protein templates. The limiting problem of fold recognition methods is the identification of the best alignment of the target sequence with the templates.

4.3.1.2.2 Model generation

The model of PehA has been generated using the server FFAS03, available on line at the following url: http://ffas.ljcrf.edu/ffas-cgi/cgi/ffas.pl . The FFAS03 server provides a web interface to the third generation of the profile-profile alignment and fold-recognition algorithm and function assignment system FFAS (33). SCRWL (http://www1.jcsg.org/scripts/prod/scwrl/serve.cgi), a side-chain packing program (34), has been used to create a 3D structural model of the predicted fold.

4.3.1.2.3 Model optimisation and validation

The total energy of the model was minimized using CNS (35) (model_minimize.inp protocol) for 2000 steps with all of the atoms non fixed to reduce the likelihood of molecular geometry distortions, spatial conflicts and energetically unfavourable van der Waals contacts. The stereochemical criteria used are those of Engh & Huber (36). In addition, the information of the three selected pairs of cysteine residues that form

disulphide bonds (Cys3-Cys25, Cys216-Cys245, Cys-309-Cys421) has been included. The web service NQ-Flipper (http://flipper.services.came.sbg.ac.at) based on knowledge-based potentials of mean force (37, 38) has been used to automatically detect and correct erroneous asparagine, glutamine and histidine rotamers in the PehA model.

The final energy minimized PehA model has been evaluated by the protein structure quality score (PSQS). The PSQS is an energy-like measure for the quality of a protein structure, and is based on the statistical potentials of the mean interactive force between residue pairs and between single residues and solvent (P. Szczesny, unpublished; http://www1.jcsg.org/psqs/psqs.cgi).The average PSQS for a representative set of structures covering all folds taken from the SCOP data base (39) is of -0.27 and most structures have PSQS less than -0.1.

The graphical representations of the PehA model have been generated using the molecular graphics system PyMOL (40).

4.3.2 Results

4.3.2.1 Computational analysis of the primary sequence

The primary sequence of PehA was analysed using the server BetaWrapPro. The score value (- 20.70) and Pi value (0.0) confirmed that PehA folds into a right-handed β-helix. The sequence alignment of the six rungs predicted by the server is shown in Figure 4.9. As indicated in the Legend of Figure 4.9, a colour code is used to define the secondary structure elements that characterize a *wrap*.

```
144    VAFPSS.............SNATGSALIIRNSP.............KEQ    166
167    LVIESG...........SKNVTIDGVWIYANPNRNASGDDLAPNTDA    202
203    IDIIGT............QTATIKNCLLDTG.............DDD    224
225    IAIKSNAGG......AATSSVNVSHCVVGGG..............HG    251
252    ISIGGQEAAGTTLAKPGVSQVTVDTMQFSGT.............DYG    285
286    YRIKTDQTAK...DSGATTGVTYRNTCMRNV.............QQP    316
317    FLFTYAYASGTGGALPIIANVTIDNVIAT.................    411

          B1         T1         B2  T2  B3           T3
```

Figure 4.9. Results of the BetaWrapPro prediction. The BetaWrapPro server predicts the sequence of *wraps*, from the PehA sequence (acc. num.: P93294), identifying the B2-T2-B3 segments on the basis of conserved pattern of amino acidic stacks and hydrogen bonds pleating (see Figure 4.1). The secondary structure elements are shown according to a colour code. The β-strand B1, B2 and B3 are shown in brown, red and orange, respectively. The β-turns T1, T2 and T3 are labeled in grey, blue and black, respectively.

4.3.2.2 Modeling

A model of the three dimensional structure of PehA was built up using the threading approach. The FFAS03 server predicted the polygalacturonase from *Erwinia caroto-vora* (PDB entry 1BHE, 24% identity, 376 residues, score: -61.30) and the rhamnogalacturonase A from *Aspergillus aculeatus* (PDB entry 1RMG, 15% identity, 422 residues, score: -53.50), covering the PehA residues range (1-393) and (1-431) respectively, as the most suitable templates. An view of the multiple sequence alignment is reported in the Appendix 1. The pairwise alignments of the target protein with the two selected templates are shown in Figure 4.10.
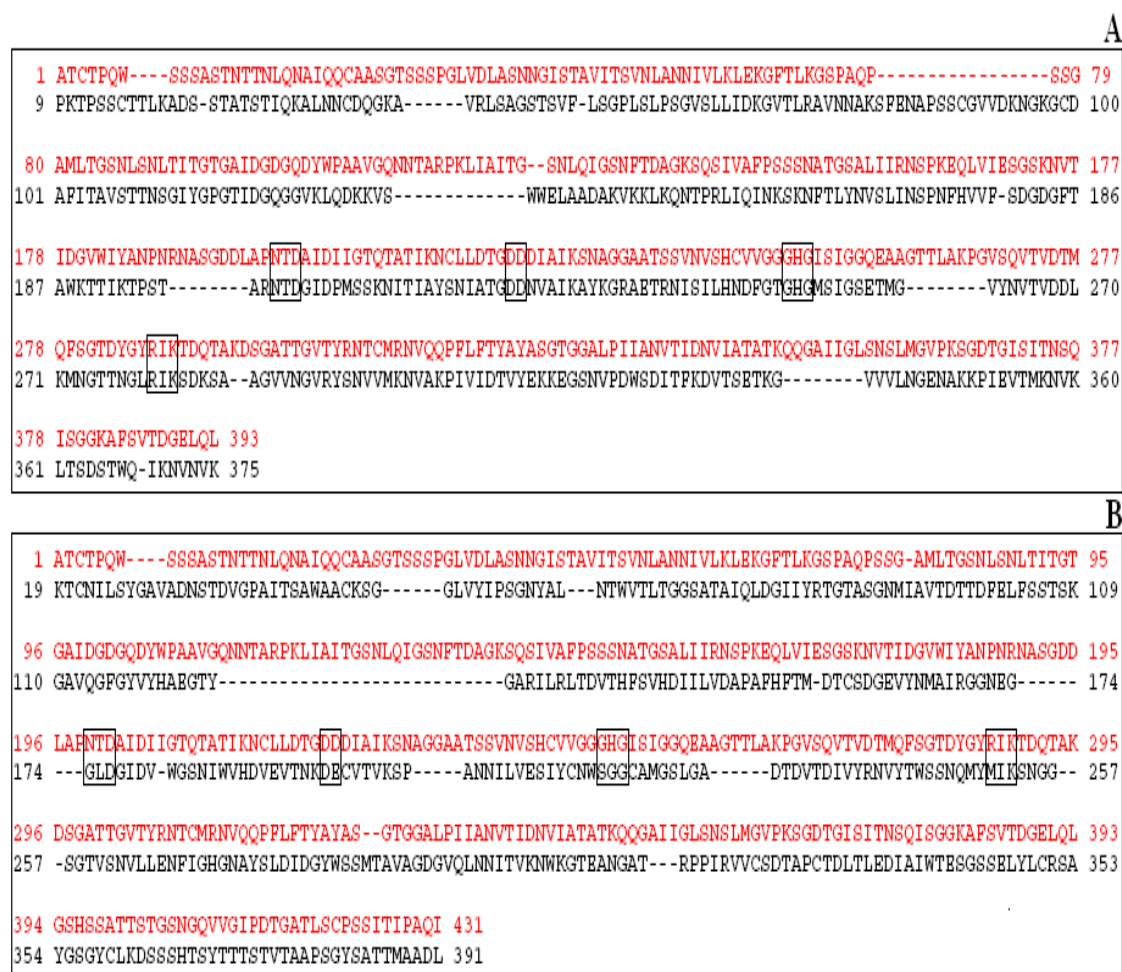
Figure 4.10. Pairwise sequences alignment generated by the FFAS03 server. (A) The alignment involves the target protein, PehA of *B. cepacia* (acc. num.: P93294, labeled in red) and the endoPGII of *E. carotovora* (acc. num.: P26509, labeled in black). (B) The alignment involves the target protein, PehA of *B. cepacia* (acc. num.: P93294, labeled in red) and the rhamnogalacturonase of *A. aculeatus* (acc. num.: Q00001, labeled in black). The boxes include the four strictly conserved amino acid segments involved in the catalysis, as discussed in Chapter 1 (Par. 1.3.2.3).

Since, the latter alignment allows to cover the full range of the PehA target protein (431 residues) the backbone template of rhamnogalacturonase A from *Aspergillus aculeatus* has been selected as template. In fact, the sequence alignment clearly revealed that a model of PehA, including the C-terminal tail, could be built up exclusively using the rhamnogalacturonase structure as template. The superimposition of the template and the target protein backbones is shown in Figure 4.11A. It is worthy of note that several portions of the PehA sequence could be not included in the model (Figure 4.11B). These missed regions have been designated as deleted sequences (D) and have been identified

as shown in Figure 4.12.
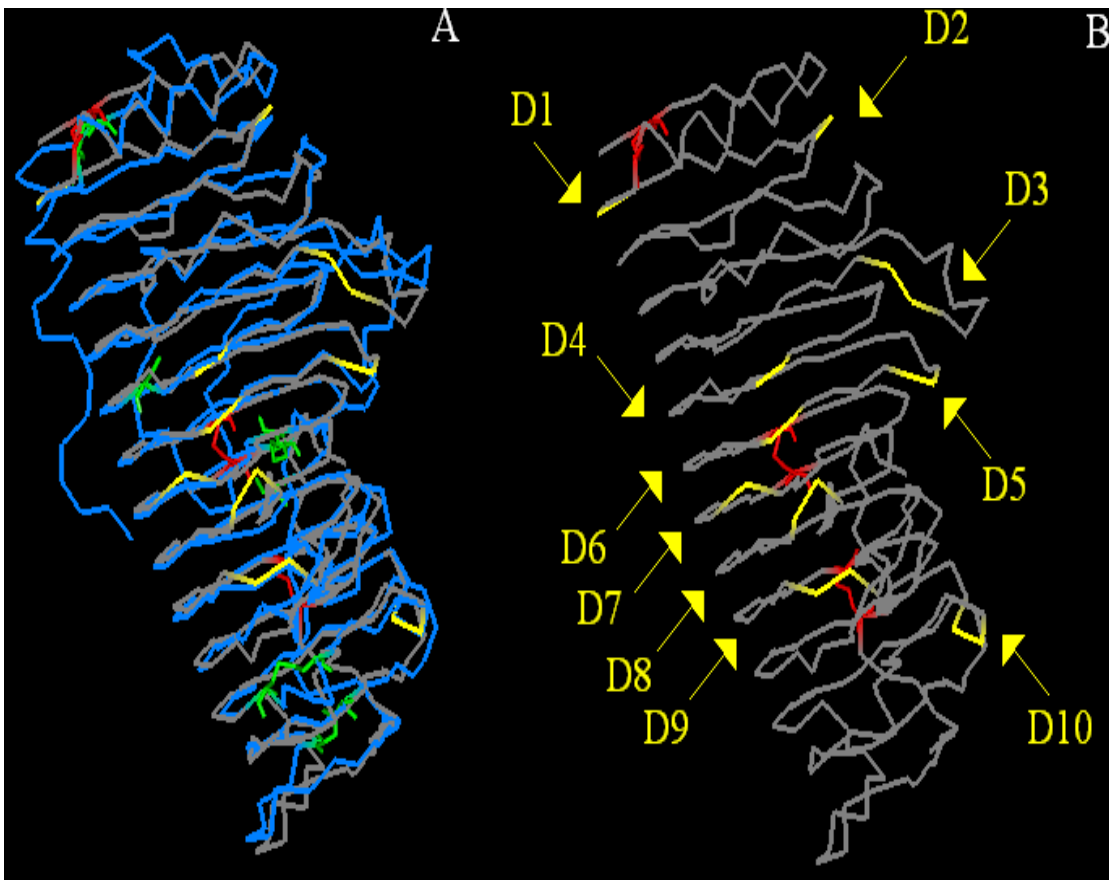
Figure 4.11. PehA model building process. (A) The superimposition of the backbones of the template (pdb code: 1RMG, labeled in blue) and the target protein (labeled in grey). The cysteine residues of the template protein have been labeled in green, while those of the target protein in red. The amino acid residues flanking the deleted sequences are shown in yellow. (B) A representation of the PehA backbone, where cysteine residues have been labeled in red. The amino acid residues flanking the deleted sequences are shown in yellow while the numbering of the deleted sequences (D) is reported in Figure 4.12.
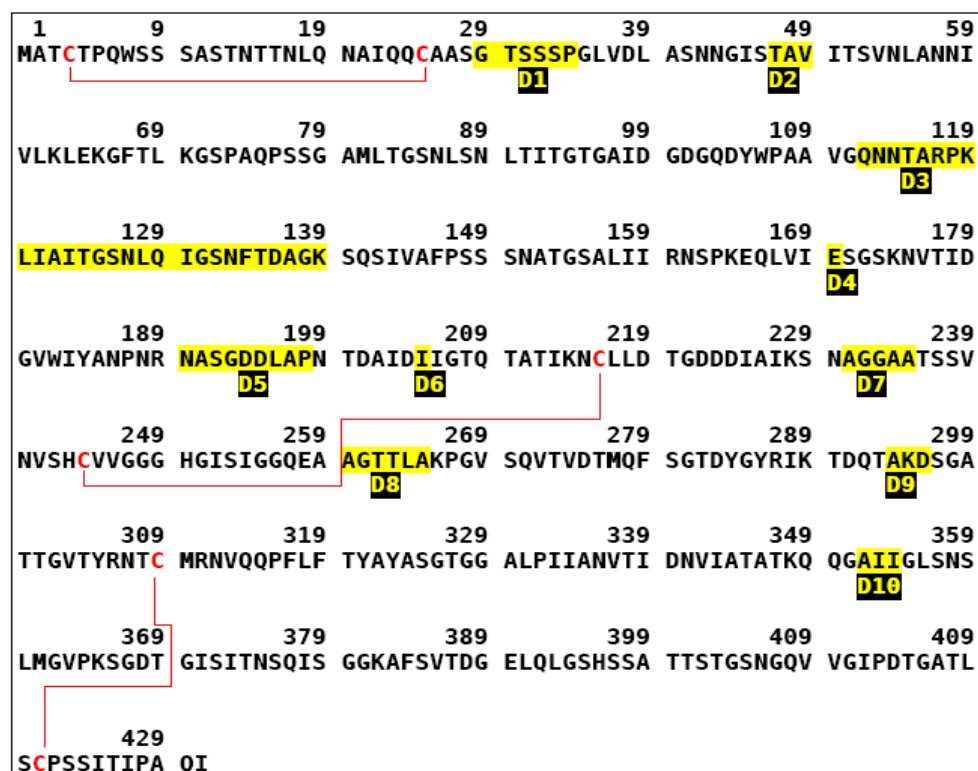
Figure 4.12. Numbering of the deleted sequences (D). The portion of the amino acid sequence of PehA that have been not found in the model are shown in yellow and are referred as D (deleted sequences). The numbering of the amino acid sequence starts with the first residue of the mature protein, an alanine (see Chapter 2, Par. 2.3.1).

The final model of PehA is shown in Figure 4.13. The PSQS value for the PehA model is of - 0.15, indicating the good quality of the model (see Par. 4.3.1.2.3).

The modelled β-helix contains 10 complete *wraps*. The β-helix is capped at the N-terminus end by a short α-helix, that contains a disulphide bond between Cys3 and Cys25. The presence of this disulphide bond has not been experimentally proven, nevertheless it has been included in the model by homology with other β-helix proteins. In fact, a disulphide bond at equivalent position has been found not only in the crystal structure of the template protein (Figure 4.11A) but also in all the six endoPGs crystal structures (1-6). Concerning the C-terminal extreme, a long tail in an extended conformation is found. The tail contains a cysteine residue, Cys421, that forms a disulphide bond with a cysteine very far in sequence, Cys309. The model revealed that the disulphide bond between Cys309 and Cys421 plays a crucial role in the stabilization of the C-terminal

tail, that is bent towards the central core of the β-helix. The third disulphide bond, that involves Cys216 and Cys244, is found in the proximity of the active site.
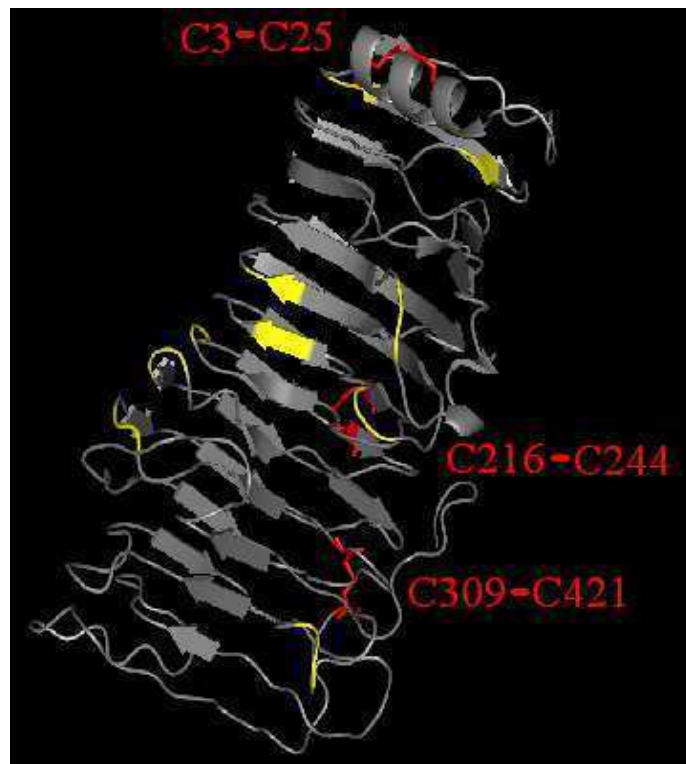
Figure 4.13. The PehA model. The model shows that the PehA β-helix contains 10 complete *wraps*. As concerns the extremes of the helix, a small α-helix is found at the N-terminal end, while a long segment in extended conformation is found at the C-terminus. The cysteine residues involved in the formation of disulphide bonds are shown in red. The amino acid residues flanking the deleted sequences are shown in yellow.

The amino acids involved in catalysis has been identified by homology with other endoPGs, as widely discussed in the Chapter 1 (Par. 1.3.2.3). The active site architecture of PehA is shown in Figure 4.14. The catalytic residues are believed to be Asp201 (belonging to the NT**D** segment), Asp222 and Asp223 (belonging to the **DD** segment). The other strictly conserved amino acids are: His250 (belonging to the G**H**G segment), involved in the regeneration of the acid-base equilibrium of the catalytic aspartates; Arg287, Lys289 (belonging to **R**I**K** segment) and the invariant Tyr323 that partecipate in the substrate binding; Gly255 and Gly256, connected by a nonprolyl *cis*-peptide bond in the known endoPGs crystal structures (1-6), involved in the formation of the substrate binding cleft. The model suggest that a disulphide bond between Cys216 and Cys244 is crucial to preserve the catalytic residues in a functional spacial

arrangement.
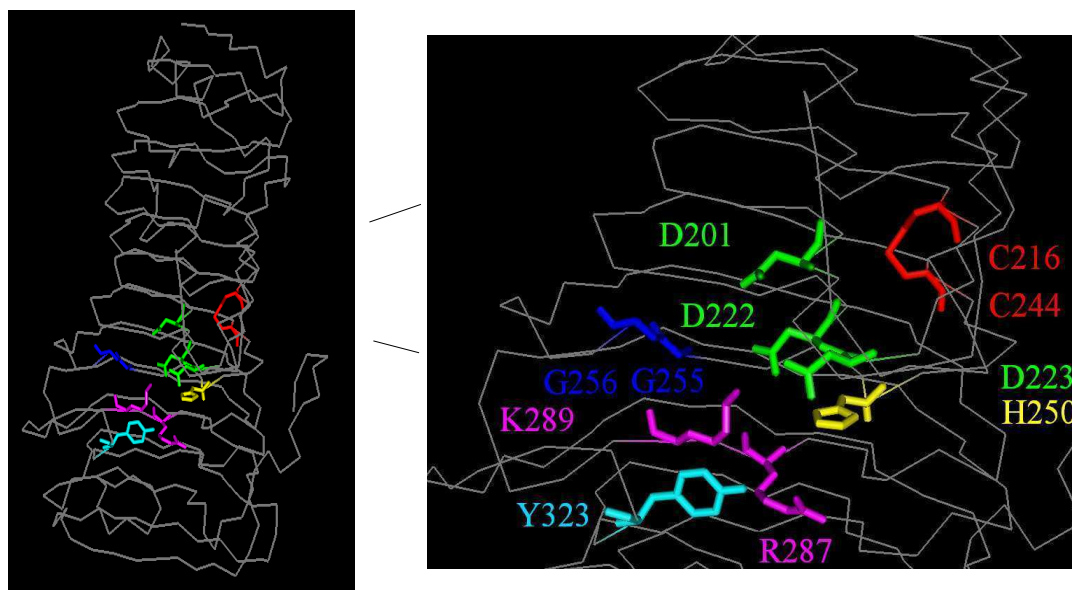
Figure 4.14. Active site of PehA. The aspartate residues involved in the catalysis are labeled in green. D222 is supposed to be the acid catalyst, while D201 and D223 are considered to be the general bases. As concerns the subsite +1, it is supposed to include H250 (labeled in yellow), Y323 (labeled in cyan), R287 and K289 (labeled in magenta). On the other hand, the substrate binding cleft at subsite -1 may involve G255 and G256 (labeled in blue) and K289 (labeled in magenta). The formation of a disulphide bond between C216 and C244 (labeled in red) is thought to be crucial to maintain the functional geometry of the active site.

As already mentioned, the β-helices structures are stabilized by highly ordered stacking of side chains of amino acids, that belong to adjacent parallel β-strands. Several aliphatic stacks have been observed inward the β-helix (Figure 4.15). These stacks often involves the same type of amino acid and are extended over a large number of β-strands (Figure 4.15A). The amino acids involved in the five aliphatic stacks identified are reported in Table 4.3, according to the colour code used in Figure 4.15.

Table 4.3. The amino acids involved in the aliphatic stacks shown in Figure 4.15. The numbering of the amino acids refers to the numbering reported in Figure 4.12.

Stacks	*Amino acids*	*Colour code*
Stack 1	Ile169 – Ile206 – Ile227 – Ile254 – Ile288	yellow
Stack 2	Leu69 – Ile97 – Ile59 – Ile183 – Leu218 – Val246	red
Stack 3	Leu167 – Ile203 – Ile225 – Ile252	blue
Stack 4	Ala152 – Val176 – Ala211 – Val239 – Val272 – Val303 – Val337 – Ile371	magenta
Stack 5	Ile213 – Val241 – Val274 – Tyr305 – Ile338 – Ile373	cyan

Another conserved feature of parallel β-helices proteins is the asparagine ladder (21). The canonical asparagine ladder, found in correspondance of turn T2, is usually present near the C-terminus. In our model, an additional asparagine ladder has been observed in the N-terminal region. Moreover, the model evidences an unusual long stacking of threonine, interrupted by an asparagine residue, that cover a long part of the β-helix. The asparagine ladders and the threonine ladder are oriented outward the β-helix, as ilustrated in Figure 4.16.
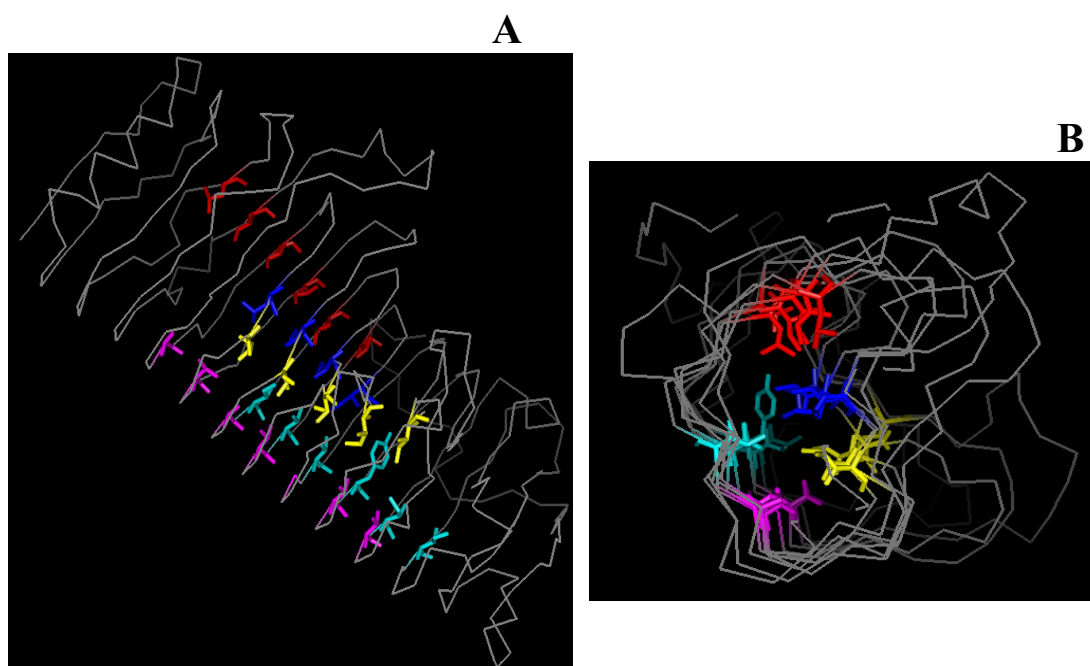
Figure 4.15. Aliphatic stacking, inward the β-helix core. (A) A lateral view. (B) A top view. These stacks involves the amino acids listed in Table 4.3 and they are extended over a large number of β-strands. Aliphatic stacks 1, 2, 3, 4, 5 are shown in red, yellow, blue, magenta and cyan, respectively.

The amino acids involved in the asparagine ladders and in the threonine ladder are listed in Table 4.4.

Table 4.4. The amino acids involved in the asparagine ladders and in the threonine ladder, as shown in Figure 4.16. The numbering of the amino acids refers to the numbering reported in Figure 4.12.

Stack	Amino acids	Colour code
Asparagine ladder	Asn86 – Asn151 – Asn175	cyan
Threonine ladder	Thr153 – Thr177 – Thr212 – Asn240 Thr273 – Thr304 – Thr338	magenta
Asparagine ladder	Asn307 – Asn341 – Asn375	yellow
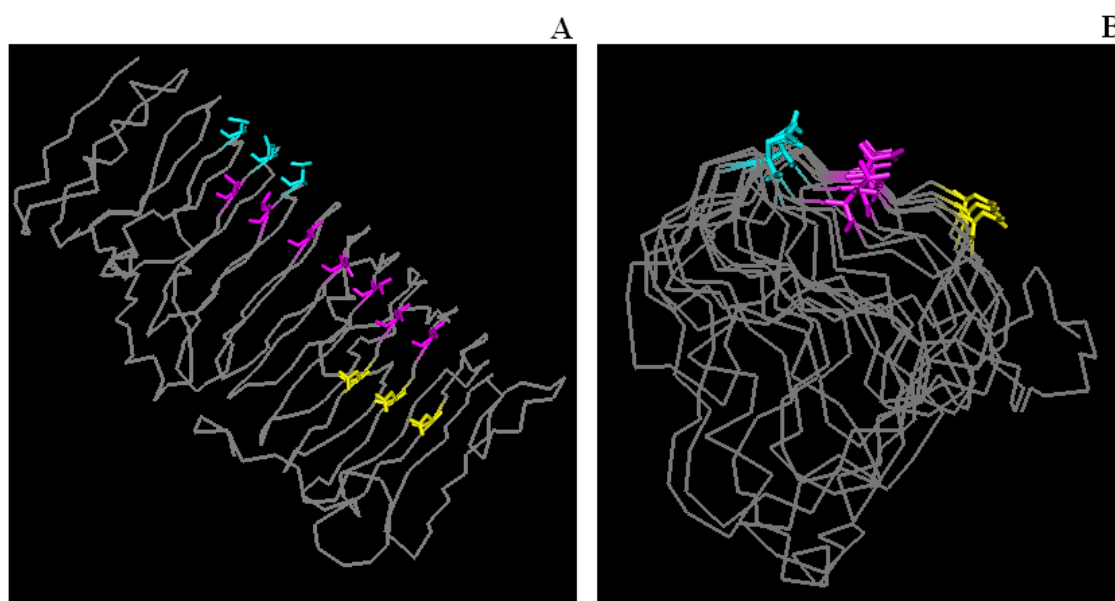
Figure 4.16. The asparagine ladders and the threonine ladder, present outward the β-helix core. (A) A lateral view. (B) A top view. These stacks involves the amino acids listed in Table 4.4 and are coloured according to the same colour code.

Lastly, two aromatic stackings have been found, involving Trp7- Phe67 and Tyr286-Phe317 (Figure 4.17A and 4.17B, respectively). They are oriented inward the β-helix core.
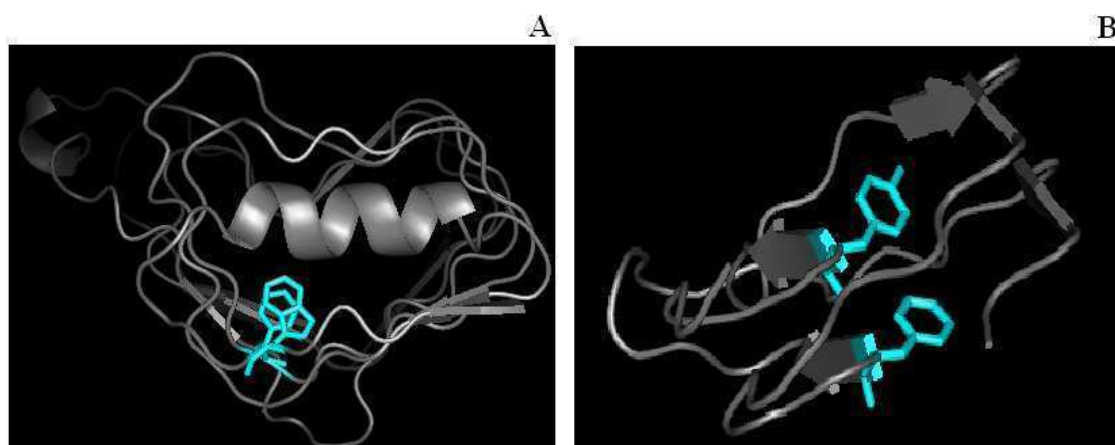
Figure 4.17. Aromatic stacks. (A) Aromatic stack involving Trp 7 – Phe 67. (B) Aromatic stack involving Tyr 268 – Phe 317.

4.4 Discussion

A structural investigation of PehA was undertaken exploiting various methods. At first exaustive crystallization screenings were conducted, in order to determine the three dimensional structure of the protein by X-Ray diffraction. Best crystals were obtained by lowering the pH of the crystallization conditions to pH 3.0 and by using the microseeding technique (Figure 4.7E). Nevertheless, the obtained crystals diffracted poorly, even if their diffraction quality have been checked at diverse synchroton beam lines, including the microfocus synchroton beam line X10SA at Swiss Light Source. This finding indicates that the low diffraction power of the PehA crystals might not be due to their limited dimensions (100 × 30 × 30 μm), but most probably to a disordered lattice.

The HTP crystallization screening approach, performed at the HTP crystallization laboratory of Hauptman Woodward Medical Research Institute, Buffalo NY, USA, revealed a new crystallization condition. When this crystallization condition was reproduced using in-house setups as described in Materials and Methods (Par. 4.1.1.2.2), no crystals could be obtained. The change of the drop size (from nano- to microliters) may have affected the achievement of the saturation equilibrium, thus explaining the non-reproducibility of crystals in the in-house screenings.

The occurrence of six cysteine residues in the primary sequence of PehA led us to speculate that they might be involved in the formation of disulphide bonds. The presence of disulphide bonds was investigated by MALDI-TOF-MS analysis. This is the first report of a disulphide bond mapping in bacterial endoPGs. The topology of two disulphide bridges has been experimentally proven. The first disulphide bond (Cys216-Cys245) is found in the proximity of the active site and it probably plays a structural role in the preserving the correct spacial disposition of the active site (Figure 4.14). The second disulphide bond (Cys309-Cys421) involves a cysteine residue found at the C-terminal tail, that is usually present in extended conformation. Therefore it is an important element in the stabilization of the long tail with respect to the β-helix scaffold. The formation of a dusulphide bond between the two remaining cysteine residues (Cys3-Cys25) has not been experimentally proven. Nevertheless, since a disulphide bond is conserved in the N-terminal end of both fungal and bacterial endoPGs, it is possible suggest that it

should occur also in the N-terminal extreme of PehA.

Since the crystallization screening failed in obtaining diffraction-quality crystals, the investigation of the tertiary structure of PehA was exploited by generating a three dimensional model. Since the β-helix fold is conserved among several endoPGs, threading approach was chosen. The occurrence of three disulphide bonds was included during the energy-minimization process. The model revealed that the PehA β-helix is composed of 10 coils (Figure 4.13). A detailed structural analysis led to the identification of the typical hydrogen bond networks that engage the asparagine residues belonging to parallel β-strands (Figure 4.16) and several ordered stackings of hydrophobic side chains (Figures 4.15 and 4.17).

However, several regions of the PehA sequence could be not modelled using the threading method, as graphically visualized in Figure 4.11B. In order to explain why these segments could be not included in the final model, a detailed analysis of PehA primary sequence has been carried out, taking into account the BetaWrapPro prediction. An extended structure-based sequence alignment of the amino acid sequences corresponding to each *wrap* is shown in Figure 4.18.
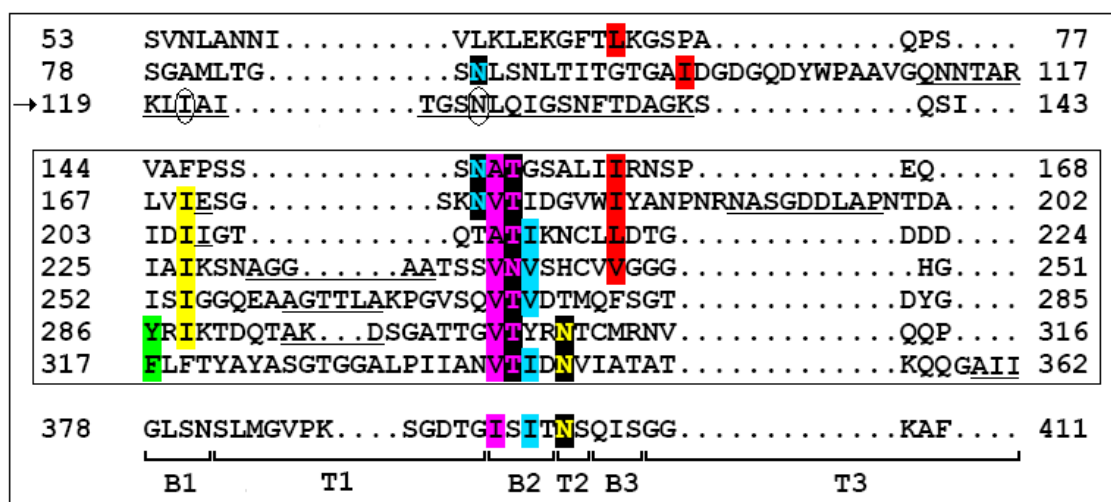
Figure 4.18. An extended structure-based sequence alignment of the parallel *wraps,* in which the amino acids are aligned on the basis of their secondary structure. The box contains the BetaWrapPro prediction. The aliphatic stacks are highlighted according the colour code reported in the Table 3. The asparagine ladders and the threonine ladder are shown according the colour code reported in the Table 4, against a black background. The aromatic stack that involves Tyr 286 and Phe 317 is shown in green. The underlined sequences represent the deleted sequences (D, see Figure 4.12). The arrow indicates the D3 sequence, aligned as a putative additional *wrap*. The circles highlighted two residues of D3 sequence, Ile121 and Asn127.

All stackings previously described have been visualized, as indicated in the Figure 4.18. On the basis of the amino acid conservation, we could include additional *wraps* of the β-helix to those already identified by the BetaWrapPro server (Figure 4.9). For instance, the longer deletion D3, encompassing Glu112-Lys139, could be considered as an additional *wrap*. In fact, several residues have been found in the conserved position with respect to the identified amino acid stacks. In particular, Ile121 could participate to the aliphatic stack 1 (see Table 4.3), while Asn127 could be involved in the N-terminal asparagine ladder (see Table 4.4). From an evolutionary point of view, since the binding of the oligogalacturonides (OGAs) to endoPGs is thought to be roughly parallel to the β-helix axis, the insertion of an additional *wrap* might have been selected in order to allow the binding of a longer substrate. On the other hand, the D3 sequence could also fold as part of an extra-domain, completely unrelated with the β-helix *wraps*. In fact, for example, in the structure of *E. carotovora* endoPG, the regularity of the β-helix is interrupted by a large extra-domain, located in the proximity of the active site (1). The existence of extra-domains that protrude from the β-helix scaffold surely interfere with the substrate binding, leading to define the enzyme specificity.

Most of the other not modelled sequences are insertions within the predicted *wraps* and in particular are identified in the elements T1 and T3 (Figure 4.19). Since the turns usually contain highly variable loops, it is not surprising that the threading method was not able to model these regions.
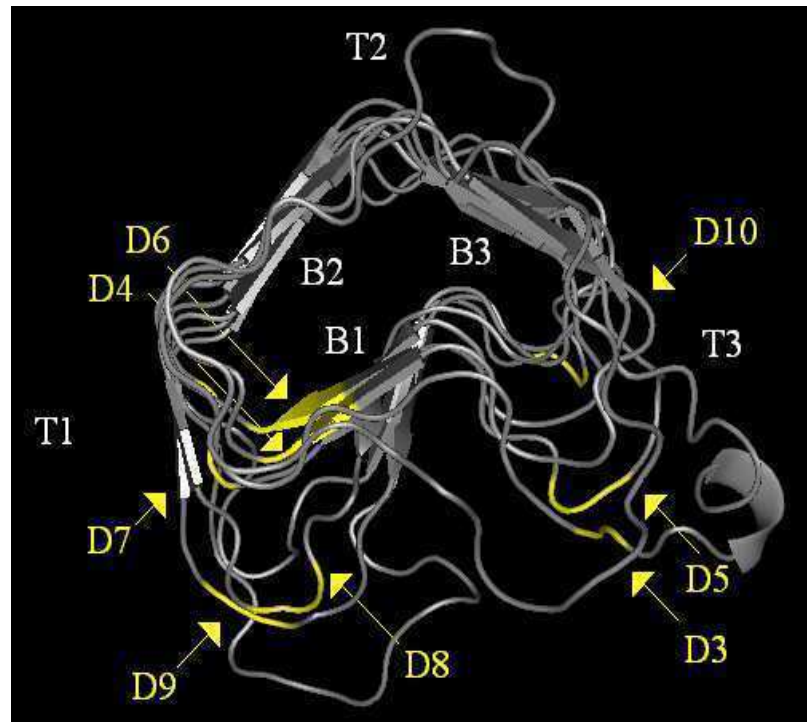
Figure 4.19. A top view of the PehA β-helix. The deleted sequences are shown in yellow. As reported also in Figure 4.18, the deletions have been predominantly found in the turn regions.

4.5 Bibliography

(1) R. Pickersgill, D. Smith, K. Worboys, J. Jenkins, Crystal Structure of Polygalacturonase from *Erwinia carotovora* ssp. *Carotovora*, J. Biol. Chem. 273(38) (1998) 24660-24664.

(2) Y. Van Santen, J.A.E. Benen, K.-H. Schroter, K.H. Kalk, S. Armand, J. Visser, B.W. Dijkstra, 1.68 a crystal structure of endopolygalacturonase II from *Aspergillus niger* and identification of active site residues by site-directed mutagenesis, J. Biol. Chem. 274 (1999) 30474– 30480.

(3) L. Federici, C. Caprari, B. Mattei, C. Savino, A. Di Matteo, G. De Lorenzo, F. Cervone, D. Tsernoglou, Structural requirements of endopolygalacturonase for the interaction with endoPGIP (polygalacturonase-inhibiting protein), Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 13425– 13430.

(4) S.W. Cho, S. Lee, W. Shin, The X-ray structure of *Aspergillus aculeatus* polygalacturonase and a modeled structure of the polygalacturonase–octagalacturonate complex, J. Mol. Biol. 314 (2001) 863– 878.

(5) T. Shimizu, T. Nakatsu, K. Miyairi, T. Okuno, H. Kato, Active-site architecture of endopolygalacturonase I from *Stereum purpureum* revealed by crystal structures in native and ligand-bound forms at atomic resolution, Biochemistry 41 (2002) 6651– 6659.

(6) G. van Pouderoyen, H. Snijder, J. Benen, B. Dijkstra, Structural insights into the processivity of endopolygalacturonase I from *Aspergillus niger*, FEBS Letters 554 (2003) 462-466.

(7) M.D. Yoder, N.T. Keen, F. Jurnak, New domain motif: structure of pectate lyase C, a secreted plant virulence factor, Science 260 (1993) 1503–1507.

(8) T.N. Petersen, S. Kauppinen, S. Larsen, The crystal structure of rhamnogalacturonase A from *Aspergillus aculeatus:* a right-handed β-helix, Structure 5 (1997) 533–544.

(9) W. Huang, A. Matte, Y. Li, Y.S. Kim, R.J. Linhardt, H. Su, M. Cygler, Crystal structure of chondroitinase B from *Flavobacterium heparinum* and its complex with a disaccharide product at 1.7 A resolution, J. Mol. Biol. 294 (1999) 1257-1269.

(10) G. Michel, L. Chantalat, E. Fanchon, B. Henrissat, B. Kloareg, O. Dideberg, The ι–Carrageenase of *Alteromonas fortis* A β–helix fold-containing enzyme for the degradation of a highly polyanionic polysaccharide, J. Biol. Chem. 276 (2001) 40202-40209.

(11) A.M. Larsson, R. Andersson, J. Stahlberg, L. Kenne, T.A. Jones, Dextranase from *Penicillium minioluteum*: reaction course, crystal structure, and product complex, Structure 11 (2003) 1111-21.

(12) S. Heffron, G.R. Moe, V. Sieber, J. Mengaud, P. Cossart, J. Vitali, F. Jurnak, Sequence profile of the parallel β-helix in the pectate lyase superfamily, J. Struct. Biol.122 (1998) 223–235.

(13) U. Baumann, S. Wu, K.M. Flaherty, D.B. McKay, Three-dimensional structure of the alkaline protease of *Pseudomonas aeruginosa:* A two-domain protein with a calcium binding parallel β-roll motif, EMBO J. 12 (1993) 3357–3364.

(14) S. Steinbacher, R. Seckler, S. Miller, B. Steipe, R. Huber, P. Reinemer, P. Crystal structure of P22 tailspike-protein: Interdigitated subunits in a thermostable trimer, Science 265 (1994) 383–386.

(15) P. Emsley, I.G. Charles, N.F. Fairweather, N. W. Isaacs, Structure of *Bordetella pertussis* virulence factor P.69 pertactin, Nature 381 (1996) 90–92.

(16) P. Bradley, L. Cowen, M. Menke, J. King, B. Berger, BETAWRAP: successful prediction of parallel β-helices from primary sequence reveals an association with many microbial pathogens. Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 14819-14824.

(17) L. Cowen, P. Bradley, M. Menke, J. King, B. Berger, Predicting the β-Helix Fold from Protein Sequence Data, J. Comput. Biol., 9 (2002) 261-276.

(18) A.W. McDonnell, M. Menke, N. Palmer, J. King, L. Cowen, B. Berger, Fold recognition and accurate sequence-structure alignment of sequences directing β-sheet proteins, Proteins 63 (2006) 976-985.

(19) O. Mayans, M. Scott, I. Connerton, T. Gravesen, J. Benen, J. Visser, R. Pickersgill, J. Jenkins, Two crystal structures of pectin lyase A from *Aspergillus* reveal a pH driven conformational change and striking divergence in the substrate-binding clefts of pectin and pectate lyases, Structure 5 (1997) 677– 689.

(20) J. Jenkins, O. Mayans, R. Pickersgill, Structure and evolution of parallel β-helix

proteins, J. Struct. Biol. 122 (1998) 236-246.

(21) M.D. Yoder, F. Jurnak, The parallel β-helix and other coiled folds, FASEB J. 9 (1995) 335–342.

(22) O. Markovic, S. Janecek, Pectin degrading glycoside hydrolases of family 28: sequence-structural features, specificities and evolution, Protein Eng. 14 (2001) 615-631.

(23) A. McPherson, Crystallization of Biological Macromolecules, Cold Spring Harbor Laboratory Press, New York (1999).

(24) N. E. Chayen, P. D. Shaw Stewart, D. L. Maeder, D. M. Blow, An automated system for micro-batch protein crystallization and screening. J. Appl. Crystallogr. 23 (1990) 297-302.

(25) N. E. Chayen, comparative studies of protein crystallization by vapour-diffusion and microbatch tecniques, Acta Cryst. D54 (1998) 8-15.

(26) A. D'Arcy, C. Elmore, M. Stihle, J.E. Johnston, A novel approach to crystallizing proteins under oil, J. Crystal Growth 168 (1996) 175-180.

(27) N. E. Chayen, A novel technique to control the rate of vapour diffusion, giving larger protein crystals J. Appl. Cryst 30 (1997) 198-202.

(28) J. Jancarik, S.H. Kim, Sparse matrix sampling: a screening method for crystallization of proteins, J. Appl. Cryst. 24 (1991) 409-411.

(29) T. Bergfors, Seeds to crystals, J. Struct. Biol. 142 (2003) 66-76.

(30) T. Shimizu, K. Miyairi, T. Okuno, Determination of glycosilation sites, disulphide bridges, and the C-terminus of Stereum purpureum mature endopolygalacturonase I by electrospray ionization mass spectrometry, Eur. J. Biochem. 267 (2000) 2380-2389.

(31) J.U. Bowie, R. Lüthy, D. Eisenberg, A method to identify protein sequences that fold into a known three-dimensional structure, Science 253 (1991) 164-170.

(32) D.T. Jones, W.R. Taylor, JM. Thornton, A new approach to protein fold recognition, Nature 358 (1992) 86-89.

(33) L. Jaroszewski, L. Rychlewski, Z. Li, W. Li, A. Godzik, FFAS03: a server for profile-profile sequence alignments, Nucleic Acid Res. 33 (2005) W284-W288.

(34) A. A. Canutescu, A. A. Shelenkov, and R. L. Dunbrack, Jr. A graph theory

algorithm for protein side-chain prediction, Protein Science 12 (2003) 2001-2014.

(35) A.T. Brünger, P. D. Adams, G. M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, M. Nigels, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, Crystallography & NMR system: a new software for macromolecular structure determination. Acta Crystallogr., Sect. D54 (1998) 905-921.

(36) R.A. Engh, R. Huber, Accurate bond and angle parameters for X-ray protein structure refinement. Acta Crystallogr., Sect. A47 (1991) 392-400.

(37) C. X. Weichenberger, M. J. Sippl, NQ-Flipper: Validation and correction of asparagine / glutamine amide rotamers in protein crystal structures, Bioinformatics, 22 (2006) 1397-1398.

(38) C. X. Weichenberger, M. J. Sippl, Self-consistent assignment of asparagine and glutamine amide rotamers in protein crystal structures, Structure, 14 (2006) 967-972.

(39) A. Andreeva, D. Howorth, S.E. Brenner, T.J.P. Hubbard, C. Chothia, A.G. Murzin, SCOP database in 2004: refinements integrate structure and sequence family data, Nucl. Acid Res. 32 (2004) D226-D229.

(40) W.L. DeLano, The PyMOL Molecular Graphics System (2002) DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org).

Final remarks and perspectives

In this Thesis it has been reported the isolation and the purification of an endopolygalacturonase (PehA) from *Burkholderia cepacia* supernatant and its preliminary characterization. In order to produce adequate amounts of protein suitable for the subsequent functional and structural studies, several strategies for the heterologous expression of PehA have been developed. Among these, the strategy for the periplasmic expression of PehA resulted to be the more efficient. A detailed characterization of the biochemical and biophysical properties of the recombinant enzyme has been carried out, leading to the conclusion that the periplasmic expression system is the more efficient in order to produce in adequate amounts a functional and correctly folded enzyme.

The study of the mode of action demonstrated that PehA is an endo-acting enzyme that cleaves the polymeric chains in a random way, displaying a non-processive behaviour. It has also been proven that the enzyme acts in a similar manner on short substrate chains. A thorough investigation on the hydrolysis mechanism of PehA led us to conclude that the formation of a productive enzyme-substrate (ES) complex involves at least four subsites. Among them, it had been shown that the three subsites referred as -2, -1 and +1 are responsible for crucial ES contacts. The fourth subsite may be either the one referred as -3 or as +2, since it has been demonstrated that the involvement of these subsites in the formation of a productive ES complex is less critical. In order to map the affinity of each subsite, further investigations on the mode of action of PehA on substrate molecules of DP 4-8 should be carried out.

Concerning the structural characterization of PehA, it has been reported the disulphide bond mapping by MALDI-TOF mass spectrometry analysis. In addition, since crystallization screening failed in obtaining diffraction-quality crystals, the expected β-helix fold of the PehA has been investigated and a three dimensional model has been generated. Due to the β-helix fold conservation among the three dimensional structures of endoPGs, the threading approach was chosen. The model revealed that the PehA β-helix is composed of 10 coils. A detailed structural analysis led the identification of the

typical hydrogen bond networks that engage the asparagine residues belonging to parallel β-strands and several ordered stackings of side chains. However, several regions of the PehA sequence could not be modeled, leaving open several questions, as largely discussed in Chapter 4. Only the determination of three dimensional structure of PehA by X-ray diffraction would give the proper answers to these questions. Therefore alternative crystallization conditions should be explored.

Appendix 1. An overview of the multiple sequence alignment generated by the FFSA03 server

#	Score	Template	Psi-Blast/align/model	Result vs.	%id
1	-61.300	1bhe_ mol:protein length:376 POLYGALACTURONASE	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	24
2	-60.600	1czf_A mol:protein length:362 POLYGALACTURONASE II	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	17
3	-59.200	1nhc_A mol:protein length:336 Polygalacturonase I	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	20
4	-56.900	1k5c_A mol:protein length:335 ENDOPOLYGALACTURONASE	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	21
5	-56.200	1hg8_A mol:protein length:349 ENDOPOLYGALACTURONASE	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	22
6	-55.000	1ia5_A mol:protein length:339 POLYGALACTURONASE	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	24
7	-53.500	1rmg_ mol:protein length:422 RHAMNOGALACTURONASE A	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	15
8	-44.200	1ogm_X mol:protein length:574 DEXTRANASE	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	13
9	-43.400	1wmr_A mol:protein length:549 Isopullulanase	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	11
10	-38.400	1h80_A mol:protein length:464 IOTA-CARRAGEENASE	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	9
11	-27.100	1pxz_A mol:protein length:346 Major pollen allergen Jun a 1	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	15
12	-24.000	1vbl_A mol:protein length:416 pectate lyase 47	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	10
13	-23.200	1pcl_ mol:protein length:355 PECTATE LYASE E	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	13
14	-21.200	1idj_A mol:protein length:359 PECTIN LYASE A	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	9
15	-21.000	1air_ mol:protein length:353 PECTATE LYASE C	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	11
16	-19.400	1idk_ mol:protein length:359 PECTIN LYASE A	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	10
17	-18.500	1qcx_A mol:protein length:359 PECTIN LYASE B	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	13
18	-17.700	1ru4_A mol:protein length:400 Pectate lyase	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	9
19	-17.300	1bn8_A mol:protein length:420 PROTEIN (PECTATE LYASE)	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	13
20	-15.500	1ooc_A mol:protein length:361 Pectate lyase A	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	14
21	-15.500	1jrg_A mol:protein length:361 Pectate lyase	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	14
22	-13.900	1gq8_A mol:protein length:319 PECTIN METHYLESTERASE	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	10
23	-13.800	1xg2_A mol:protein length:317 Pectinesterase 1	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	10
24	-13.300	1qjv_A mol:protein length:342 PECTIN METHYLESTERASE	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	13
25	-10.200	1dbg_A mol:protein length:506 CHONDROITINASE B	psi-nr85 ncbi ali scwrl	COG0706 JCSG0706 PDB0706 PfamA200 SCOP169	7

```
              .    10    .    20    .    30    .    40    .    50    .    60    .    70    .    80    .    90    .   100    .   110    .   120    .   130    .   140    .   150
First  ATCTPQWSSSASTNTTNLQNAIQQCAASGTSSSPGLVDLASNNGISTAVITSVNLANNIVLKLEKGFTLKGSSPAQPSSGAMLTGSNLSNLITCTGAIDGDGQDYWPAAVGQNNTARPKLIAITGSNLQIGSNFTDACKSQSIVARPSSS
    9  PKTPSSCADS-STATSTIQKALNNCDQGKA-------VRLSAGSTSVF-LSGPLSPSGVSLLIDKGVTLRAVNNAKGCDAFITAVSTTNSGIVPGTIDGQGGVKLQDKKVS----------------ELAADAKVKKLKQNTPRLIQINKSK
   16  ATFASASRDSCTFFTAAAAKAGKAKCSTIT---LNNIEVPAGTTLDLT-----GLTSGTKVIFEGTTFQYEEW--AGPLISMSGEHITVTGASGHLINCDGARWDWGKGTSG-------------KKKPKFFYAHGLD
    1  .......STCTFTSASEASESLSSCSDVVL----SSIEVPAGETLDLS-----DAADGSTTTFEGTTSF-GYKEWKGPLIRFGGKDLIVTMADGAVIDGDGSRWWDSKGTNGGK----------------TRPKFMYIHDVE
    1  ........ATCTVKSVDDAKDI-AGCSAVTLNGFTVPAGNTLV------LNPDKGATVTMAGDITFAKTI---LDGPLFTIDGTGINFVGADIFDGNGALVWDGKTNNGTHKP-----------------HPFLKIKG
    1  ........DPCSVTEYSLATAVSSCKNIV----LNGFQVPTGKQLDLS-----SLQNDSTVFFKGTTTFATTADNDFNPIVLSGSNIT-ITGASGHVIDGNGQAVWDGKGSNSNS---------------NQKPDHFIVVQKTTG
    2  TTCTFSG----SNGASSASKSKTSCSTIVL----SNVAVPSGTTLDLT----KLNDGTHVIFSGETTFGYK----EWSGPLISVSGSDLITGASSINGDGSRWWDGEGGNGGK--------------TRPKFFAAHSLI
   19  KTCNLLSADNSTDVGPALTSAWAACKSG------GLVIIPSGNYAL---NTWVTLTGSSATAIQLDGIIYRTGTASGNMAVTDTTDFELFSSTSKGAVQGFGVVHAEGTY------------GARILRLTDVT
  198  ASPFLPSGMIPHMTPDNTQTMPGPINNGDWGAKSILYFPPGVYWMNQDQSGNSGKLGSNIRLNSNTYWVYLAPGAYVKGAIEYFTKQNFYATCHGILSGENYVVQANAGDNYIAVKSD----------STSLRMWHNNLGGGQ
  176  ....PQPGSPNSIAPAPGRVLGLNTTS---ASTVVFNPGVYFTGHDHMVLSSSVTWVYFAPGAYVKG---------AVEFLSTASEVKASCHGVLSGRQYVWYADDPEGYQKAS--------------GANNNGLRMWRGTLGNSSQ
   21  NVDLVDDGNDTSDDSNALQRALNAISRKPN---GGTLLIPNGTYHF----LGIQMKSNVHIRVESDVIIKPTWNGDGK-NHRLFEVGVNNIVRN-FSFQGLGNGFLVDFKDSRDK----------------NLAVFKLGDVR
    1  .......DNPIDNWDQNRMKLADCAVGFGSSTMGG---KGGDFYVTVTSTDDNPVNPTYGATREKALWIIFSQN---------MNIKLKMPLYVAGHKTIDGRGAD------------------VHLGNGGPCLFMRKVS
   20  ...GTTGGAMASTNRTELIQALGGNNHTNQYNSVPKIIYVKGTIDLNVDDNNQPVGPDFY-ATWGKKEVBGPLEEARVRSQKKQKDRIMYVGSNTSIIGVGKD----------------AKIKGGGFLIKNVD
   15  ..............NGGTTGGAKAAKAVEVKNISDFKKALNGTDSSAKIIKVTGPIDISGGKAVTSFDDQXARSQ-----SIPSNTTIIGVGSNGKFTNGSLVIKGVK
   13  ...GVTGGGDATPVYPDTIDELVSYLGDDE-------ARVIVLIKTF-DFTDSEGTTTGTCGCAPWGTASACQYAIDQDDWCEN--YNAGVLGITVTSNKSLIGEG--------------------------S
   88  ...............................................................................................................................TK
   13  ...GVTGGGSATPVYPDTIDELVSYLGDDE----ARVIVLIKTF-DFTDSEGTTTGTCGCAPWGTASACQYAIDQDDWCEN--YNAGTLGITVTSNKSLIGEG----------------------------S
   13  ...GVTGGGSASTTTDELVSYLGD-NEPRVIILDQTFDFTGETETTGCGAPWGTASQCQVAINLHSWCDNYQASAPKVSVTTDGILPITVNSNKSIVGQG---------------------------T
   15  RIYYYAPGNSSNNGSMSFSAAMAAVNPG-----ELILLKPGTY------TIPVTQGKNTITPNKSGK------DGAPIVVAAANCGRAVFDFSFP---------DSQWVQASYGFYVTTGDYWYFKGVETVTRAGYQGAVVIG-S
   41  ...GTTGGSKASSNRNQLVSALGKEINTT-----PKIIYIKG-DMNVDDNLKPLGLNDYKDPEYDLLDKYLKAYDPSTWGKK-KNQXARVMVDIPANTTIVGSG---------------------------T
   49  ...................................................................SAGAEAKIIQLKGTIDISGGTPYTDFADQXARSQ-----------------INIPA
   49  ...................................................................SAGAEAKIIQLKGTIDISGGTPYTDFADQXARSQ-----------------INIPA
    7  PNVVVAADGSG--DYKTVSEAVTAAAPEDSKTRY--VIRIKAGVY-------RENVDVP-----------------------------KKKKNIMFLGDGIITASKNV----------------QDGSTTFNSATVAAVGA
    3  ANAVVAQDGTG--DYQTLAEAVAAPDKSKTRY---VIVVKRGTY-----------KENVEVA----------------------SNKMNLMIVDGTITGSLNV----------------VDGSTTFRSATLAAVGQ
    4  YNAVVSKSSSDGKTFKTIADALASAPAGSTPF---VILIKNGVY---------NERLTI--------------------TRNNLHKGES--------------------RNGAVIAAATAAGTLKSDGSKWGTAGSSTITISAK
   17  LLNVAPCLGQVVASNETLIYQVVKEVPG------GLVQIADGTYKDVQLIVSNSGKSGLPITIKA----------LNPGKVFFTGDAKVELRGEEHLFKDGNRAIQAWKSHGPCLVAIYGSYNRITACVFDCFDEANSTTSLITEDGKVPQ
```

```
        .    160       .    170       .    180       .    190       .    200       .    210       .    220       .    230       .    240       .    250       .    260       .    270       .    280       .    290       .    300

NATGSALIIRNSPKEQLVIESGSKNVTIDGVWIYANPNRNASGDDLAPNTDAIDIIGTQTATIKNCLLDTGDDDIAIKSNAGGAATSSVNVSHCVVGGHGISIGGQEAAGTTLAKPGVSQVTVDTMQFSGTDYGYRIKTDQTAKDSGAT
NFTLINVSLINSPNFHVVP-SDGDGFTAWKTTIKTPST--------ARNTDGIDPMSSKNITIAYSNIATGDDNVAIKAYKGRAETRNISILHNDFGTGHGMSIGSETMG-------VINVTVDDLKMNGTTNGLRIKSDKSA--AGVV
SSSITGLNIKNTPLMAFSV--QANDITFTDVTINNADGDTQG----GHNTDAFDVGNSVGVNILKPWVHNQDDCLAVNSG------ENIWFTGGTCIGHGLSIGSVGDR----SNNVVKNVTIEHSTVSNSENAVRIKTISGA--TGSV
DSTPKGINIKNTPVQAISV--QATNVHLNDFTIDNSDGDDNG----GHNTDGFDLSESTGVYISGATVKNQDDCIAINSG------ESISFTGGTCSGGHGLSIGSVGGR----DDNTVKNVTISDSTVSNSANGVRIKTIYKE--TGDV
SGTYKKFEVLNSPAQAISVGPTDAHLTLDGITVDDFA---GDTKNLGHNTDGFDV-SANNVTIQNCIVKNQDDCIAINDG------NNIRFENNQCSGGHGLSIGSIAT-------GKHVSNVVIKGNTVTRSMYGVRIKAQRTA-TSASV
NSKITNLNIQNMPVHCFDI-TGSSQLTISGLLDNRAGDKPNSLPAAHNTDGFDLSSSDHVTLDNNHVYNQDDCVAVTS-------GTNIVVSNMYCSGGHGLSIGSVGGK----SDNVVDGVQFLSSQVVNSQNGCRIKSNSGA-TGTI
NSVISGLKIVNSPVQVFSV-AGSDYLITLKDITIDNSDGDDNG----GHNTDAFDIGTSTVVTISGATVTNQDDCVAVNS------GENIYFSGGYCSGGHGLSIGSVGGR----SDNTVKNVTFVDSTIINSDNGVRIKTNIDT--TGSV
HPSVHDIILVDAPAFHFTM-DTCSDGEVYNMAIRGGNEG----------GLDGIDV-WGSNIWVHDVEVTNKDECVTVKSP------ANNILVESIYCNWSGGCAMGSLGA------DTDVTDIVRNVTWSSNQMYMIKSNGG---SGTV
TWYCVGPTINAPPENTMDF-NGNSGISSQISDYKQVGA-------FFFQQTDGPELYPNSV--VHDVFWHVNDDAIKIYY-------DISGVTIDTLNVIHTRPSAIIGASPFYASGMSP
TFVLNGVTVSAPPPNSMDWSGNLITCRVDDYKQVGAR-------YGQTDGLEM--YPGTILQDVFYHTDDDGLKMYYS--------NVTARNIVMVKESVAPVVEFGWTPRNTENVLFDNVDVHQAYANAGNNPGIFGAVNNVLHSTGN
NYKISNFTIDDNKTIFASILVDVTE-------------------NGRLHWSRNGIIERIKQNN--ALFGYGLIQTYG------ADNILFRNLHSEGGIALRMETDNLLMKNYKQGGIRNIFADNIRCSKGLAAVMFGPHFMK------N
HVILHSLHIHGCNT-------SVLGDVIVSESIGVEPVH--------AQDGDAITMRNVTNAWIDHNSLSDCDGLIDVTLG-----STGITISNNHFPNHHKVMLLGHDDTY---DDDKSMKVTVAFNQFGPNACQ-RMPRARYG--LVHV
NVIIRNIEFEAPLDYFPEWDPTDGTLGEWN----------------SEYDSISIEGSSHIWIDHNTFTDGDALDIKNS------SDFITISYNVFTNHDKVTL--IGASDSRMADSGHLRVTLHHNYVKNVTQ----RLPRVRFGQVHI
NVLRNLYIETPVDVAPHYESGDGW-------------------NAEWDAAVIDNSTNVWDHVTISDGSFKYTTKDGEKKKGSDVVTISYSRFELHDKTILIGHSDSN-GSQDSGKLRVTFHNNVFDRVT------ERAPRVRFGSI
SGAIKG------------GLRIVSGAENIIIQNIAVTDINPKY------VWGGDAITLDDCDLIWIDHVTARGRQHYVLGTS----ADNRVSLTNNYIDGVSDYSATCDGYHYWGIYDGDADIVTMKGNYIYHTSNNYFYDISGHAFEIGEG
GITTIGANG-SSANFGIWI-KKSSDVVVQNMRIGYLPGG----------AKDGMIRVDDSPNVWDHNELFAANHECDGTPDNDKGASNTVTVSYNYIHGVKKVGLDGSSS------SDTGRNITHHNYINDVNARLPLQRGGL----VHA
SGAIKG------------GLRIVSGAENIIIQNIAVTDINPKY------VWGGDAITLDDCDLIWIDHVTARGRQHYVLGTS----ADNRVSLTNNYIDGVSDYSATCDGYHYWAIYDGDADLVTMKGNYIYHTGRSPKVQ----DNTLLHA
KGVIKG------------GLRVVSGAKNVIIQNIAVTDIN----------PKYVWGGDAITVDDSDLIWIDHVTARGRQHIVLGTS----ADNRVTISYSLIDGRSDYSATCNGHHYWGVYDGSNDMVTLKGNYFVNLSRMPKVQ----GNTLLHA
HNTFENTAFPHHNRNTGLEINNGGSYNTVINSDAYRNVDPKKNGS-MADGFGPXQKQGPGNRFVGCRAENSDDGFDLFDSPQKVVIENSWAFRNGINYWNDSAFAGNGNGFKLGGNQAVGNHRITRSVAFGNVS-------KGFDQNNNA
NAKVVGGNFQ---------------IKSDNVIIRNIEFQDAYDYFPSSGNWNSQYDNITINGGTHIWIDHCTFNDSRDDSTSPKYYGRKGANYITMSYNYYHDHDKSSI-FGSSDSKTSDDGKLKITLHHNRYKNIV---------QRAPRVRF
NTTVIGAKFINGSLI-IDGTDGTNNVIIRNVIIQTPIDVEPHGDGWNAEWDAMNINGAAHHVWIDHVTISDGNFMYTTKDGETKRGSDVVTISNSLIDQHDKTMLIGHSDSN-GSQDKGKLHVTLFRNNVFNRVT-------ERAPRVRYGSI
NTTVIGAKFINGSLI-IDGTDGTNNVIIRNVIIQTPIDVEPHGDGWNAEWDAMNINGAAHHVWIDHVTISFTDMYTTKDGETVVGSDYVTISNSLIDQHDKTMLIGHNDTN-SAQDKGKLHVTLFRNNVFNRVT-------ERAPRVRYGSI
GFLARDITFQNTAGAAVALRVGSDLSAFYRCDIL-------------AVQDSLIVHSNRQFFINCFIAGTVDFIFGNAAV----LQDCDIHARRPGSGQKNMVTAQGR----TDPNQNTGIVIQKSRIGATSDLQPVQSSFPTYLGKEY
GFILQDICIQNTAGQAVALRVGADMSVINRCRID--------------AVQDTIVAHSQRQFYRDSVYVTGTVDFIFGNAAV----FQKCQLVARKPGKYQQNMVTAQGR----TDPNQATGTSIQFCNIIASSDLEPVLKEPGRPWKEY
DFSAQSLTIRNDQAVALIVTKSGDRAYFKDVSLVGYQDTLVSGGRSFFSDCR-IFGDGTALFNNCDLVSRYRADVKSGNVSGYLTAPSTNINQ----KYGLVITNSRVIRES----------DSVPAKSYGLGRPWHPTTFSDGRYA
HCRIDHCSFTDKITFDQVINLNNTARAIKDGSVGGPGMYHRVDHCFPFSNPIGYYRNDIGRCLVDSNLFMRQDSEAEIITSK----SQENVYGNTYLNCQG-----------TMNFRHGDHQVAINNFYIGNDQRFGYGGMFVWGSRHVI
```

```
            .   310       .   320       .   330       .   340       .   350       .   360       .   370       .   380       .   390       .   400       .   410       .   420       .   430                      Last
TGVTYRNTCMRNVQQPFLFTYAYASGTGGALPIIANVTIDNVIATATKQQGAIIGLSNSLMGVPKSGDTGISITNSQISGGKAFSVTDGELQLGSHSSAATTSTGSNGQVVGIPDTGATLSCPSSITIPAQI   Last
NGVRYSNVVMKNVAKPIVIDTVYEKKEGSNVPDWSDITFKDVTSEIKG-------VVVLNGENAKKPIEVTMKNVKLITSDSTWQ-IKNVNVK................                           375
SEITYSNIVMSGISDGVVIQQDYEDGKPTNGVTIQDVKLESVTGSVDSGATE-----IYILCGSGSCSDWTMDDVKVTGGKKSTACKNFPSVAS...............                           361
SEITYSNIQLSGITDGIVIEQDYENGSPSTGIPITDVTVDGVTGTLEDDATQ-----VVIILCGDGSCSDWTWSGVDLSGGKTISDKCENVPSGAS..............                           335
SGVTYDANTISGIKYGVLISQSYPDDNPGTGAPFSDVNFTGGATTIKVNNAATR-----VTVECGNCSGNWNWSQLTVTGGKA-GTIKSDKAKITGGQ............                           333
NNVTYQNIALTNISTGVDVQQDYLNGGPTNGVKISNKIFIKVTGTVASSAQD------WFILCGDGSCSGFTFSGNAITGGKTSSCNVPTNTC.................                           347
SDVTYKDITLTSIKYGIVVQQNYGDSTPTTGVPITDFVLDNVHGSVVSSGTN-----ILISCGSGSCSDWTWTDVSVSGGKTSSKCTNVPSGAS...............                           338
SNVLLENFIGHGNAYSLDIDGYWSSAVAGDGVQLNNITVKNWKGTEANGAT---RPPIRVVCSDTAPCTDLTLEDIAIWTESGSSELYLCRSAYGSGYCLKDSSHTSYTTTSTVTAAPSGYSATTMAADL      391
DSRKSISMTVSNVVCEGLCPSLFRITPLQN---YKNFVVKNVAFPDGLQTNSIGT-GESIIPAASGLTMGLAISAWTIGGQK.........................                             547
SNMTVRNIIWSNFRAEGSSSALFRINPIQDNISIKNVSIESFEPLSINTTESWMPVWYDLNNGKQIITVTDFSIEGFTVGNTTITASNAASVGRIDG.............                           536
GDVQVTNVSSVSCGSAVRSDSGFVETRQSWKQAVESKLGRGCAQTP.......................................                                                   313
ANNNYDPWNIYAIGGS----------SNPTILSEGNSFTAPSESYKKE......................................                                                  278
YNNYYEFSNLADYDFQYAWGVGV----FSQIYAQNNYFSFDWDIDPS.......................................                                                  339
H--AYNNVYLGDVKHSVYPYLIYSFGLGTSGSILSESNSFTLSNLKSIDGK..................................                                                   287
GYVLAEGNVFQNVDTVLETYEGATTAGEVCSTYLGRDCVINGFGCSGTFSEDSTSFLSDFEGKNIASASAYTSVASRVVANAGQ..................                                  356
YNNLYTNITGSGLNV---------RQNGQALIENNWFEKAINPVTSRYDGKNFGTWVLKGNNITKPADFSTYSITWTADTKPYVNADSWTSTGTFPTVAYNYSPVSAQCVGYAGVGKNLATLTSTACK.       353
VNNYWYDISGHAFEI---------GEGGYVLAEGNVFQNVDTVLETYEG.....................................                                                  288
VNNLFHNFDGHAFEIG---------TGGYVLAEGNVFQDVNVVVETP......................................                                                   286
GGVTVINNTSYKNGINYGFG------SNVQSGQKHYFRNNVSLSASVTV----------SNADAKSNSWDTGPAASASDFVSLDTISLATVSRDNDGTLPETSLFRLSANSNAGTKESNISYSGSAPDL        394
GQVHVYNNYYEGSTSSSSYPFSYAGIGKSSKIYAQNNVIDVPGLSAAK.....................................                                                   355
H---SFNNVFKGDAKDPVYRYQYSGIGTSGSVLSEGNSFTIANLSASKAC..................................                                                    298
H--SFNNVFKGDAKDPVYRYQYSRFGIGTSGSVLSEGNSFTIANLSASKAC.................................                                                    298
SRTVVMQSSITNVINP-LYYGEYQNTGAGAATS-GRVTWKGFKVITSSTEAQGFTPGSFIAGGSWLKATTFPFS.............                                                  316
SRTVVMESYLGDFALKTLYYGEFMNNGPGAGTS-KRVKWPGYHVITDPAKAMPFTVAKLIQGGSWLRSTGVATV.............                                                  312
DPNAIGQTVFLNTSMDNHIYGWDKMSGKDKNGNTIWFNPEDSRFFEYKSYGAGAAVSKDRRQLTDAQAAEYTQSKV...........                                                 333
ACNYFELSETIKSRGNAALGAMASEHALAFDMLIANNAFINVNGYAIHFNPLDERRKEYCAANRFETPHQLMLKGNLFFKDKPYVYPFFKDDYFIAGKNSWTGNVALGVEGIPVNISANRSAYKPVKIKDI      442
```

Appendix 2. Heterologous expression and purification of an endopolygalacturonase from *Burkholderia glumae*

In April 2003, the nucleotidic sequence of an endopolygalacturonase (endoPG) from *Burkholderia glumae* was submitted to the nucleotidic sequence database of the National Center for Biotechnolgy Information (url: http://www.ncbi.nlm.nih.gov), by Fumihiko Suzuki and coworkers, Department of Plant Protection, Nishigoshi Kikuchi, Japan (unpublished data). The gene locus of *B. glumae* endoPG was marked as AB109207, while the genetic determinant as BAC98359.

In order to verify the conservation of the amino acid sequence between the *B. cepacia* PehA and the *B. glumae* endoPG, a pairwise alignment was performed using the ClustalW server (url: http://www.ebi.ac.uk).

```
SeqA Name        Len(aa)   SeqB Name        Len(aa)   Score
===========================================================
1    Bur_cepacia  458      2    Bur_glumae   462       78
===========================================================


Bur_cepacia    MKGKSSTRLVLR-----LSTLAALAVQASAQAATC-TPQWSSSASTNTTNLQNAIQQCAA 54
Bur_glumae     MTPRNRVRTSVTATFGALASLIAFAAPPAAHAASCATPQWSSSPAANTSALQSAINRC-- 58
               *. :. .* :       *::* *:*. .:*:**:* *******.::**: **.**::*

Bur_cepacia    SGTSSSPGLVDLASNNGISTAVITSVNLANNIVLKLEKGFTLKGSPAQPSSGAMLTGSNL 114
Bur_glumae     SGAAGNPGLIDLRANNGVSTAVITSVKLASNIVLKLEKGFTLKGSPGQPSDGAMLTGSGL 118
               **::..***:** :***:********:**.****************.***.*******.*

Bur_cepacia    SNLTITGTGAIDGDGQDYWPAAVGQNNTARPKLIAITGSNLQIGSNFTDAGKSQSIVAFP 174
Bur_glumae     SNVSLTGTGAIDGDGQSYWASAVGKNNTARPRLIKLTGSNLQIGSNFTDAGKPQSIVAFP 178
               **:::*********** .**.:.***:**:** :*************** :*.*******

Bur_cepacia    SSSNATGSALIIRNSPKEQLVIESGSKNVTIDGVWIYANPNRNASGDDLAPNTDAIDIIG 234
Bur_glumae     SASNDPGNALIIRNSPKEQLVIESGSKNVTIDGVWIYANPKRNANGNDLAPNTDAIDIIG 238
               *:** .*.**********************************:***.*.*:**********

Bur_cepacia    TQTATIKNCLLDTGDDDIAIKSNAGGAATSSVNVSHCVVGGGHGISIGGQEAAGTTLAKP 294
Bur_glumae     TQTAHVRNCLLDTGDDDIAIKSNAGSAATSDVDISHCVVGGGHGISIGGQEAAGHTLAKP 298
               ****  ::.****************.****.*::.****************** *****

Bur_cepacia    GVSQVTVDTMQFSGTDYGYRIKTDQTAKDSGATTGVTYRNTCMRNVQQPFLFTYAYASGT 354
Bur_glumae     GVSHVTVDTVQFSGTDFGYRIKTDQTAKDSGATTGVTYRNTCMRNVQQPFLFTYTYASGT 358
               ***:*****:******:***************************************:*****

Bur_cepacia    GGALPIIANVTIDNVIATATKQQGAIIGLSNSLMGVPKSGDTGISITNSQISGGKAFSVT 414
Bur_glumae     GGDLPVIANVSIDNVIATATKAQGAIIGLPNSLIGVPKSGDTGIRITNSRITGGKPFAVS 418
               ** **:****:********** ******.***:********** ****:*:***.*:*:

Bur_cepacia    DGELQLGSHSSATTSTGSNGQVVGIPDTGATLSCPSSITIPAQI 458
Bur_glumae     NGELQVGSHSIVTTSSGTNGQVVPIADTGATLACPASITIPAQR 462
               :****:**** .***:*:***** *.******:**:*:*******
```

Sequence alignment. The amino acid sequence alignment of *B. cepacia* PehA (acc. num.: P94293) and the *B. glumae* endoPG (protein_id: BAC98359). The sequences highlighted in yellow represent the N-terminal signal peptides.

As shown by the **Sequence alignment**, the two enzymes display a high score in sequence identity (78%). The study of the functional and structural properties of the *B. glumae* endoPG and its comparison with those of the *B. cepacia* PehA could provide new insights into the enzymatic substrate specificity and the evolution of the two enzymes.

For this reason, *B. glumae* endoPG was heterologously expressed and purified, according to procedures reported in the following section of **Experimental methods**. A preliminary characterization of the recombinant *B. glumae* endoPG showed that its specific activity resulted to be 10 U/mg, two times lower than that of *B. cepacia* PehA (namely, 19.5 U/mg) (Chapter 2 , Par. 2.2.3.3.1).

As future perspective, further detailed biochemical and kinetic studies of *B. glumae* endoPG should be carried out, in order to address the role of the non-conserved amino acids residues that are responsible for the different properties of the two enzymes.

Experimental methods

A plasmid containing the DNA of *B. glumae* endoPG was kindly provided by Dr. Fumihiko Suzuki, Department of Plant Protection, Nishigoshi Kikuchi, Japan. The coding region of *B. glumae* endoPG, starting from the triplet GCC corresponding to Ala33, was PCR-amplified from the provided plasmid, using the Expand High Fidelity PCR System (Roche). The presence of an N-terminal signal peptide was deduced in analogy with the sequence of *B. cepacia* PehA (Chapter 2, Par. 2.3.1). The forward (F1: 5'-ATAATCCATGGCCAGCTGCGCGACGCCG-3') and the reverse (R1: 5'-TATATAAGCTTTCAGCGTTGCGCCGGAAT-3') primers were designed to introduce NcoI and HindIII restriction sites at the N-terminus and at the C-terminus of the protein, respectively. The reverse primer also contains a stop codon. The NcoI and HindIII digested and amplified fragment was ligated into the pET22b plasmid (Novagen) using T4 ligase (Chapter 2, Figure 2.2). The *E. coli* host strain DH5α cells (Clontech) were transformed with the pET22_NcoI_HindIII_*B_glumae_endoPG* plasmid. Transformants were selected on Luria-Bertani (LB) plates, containing 100 µg/mL ampicillin. The recombinant plasmid was isolated and it was shown by restriction analysis and automated DNA sequencing that it harbors the correct gene sequence. For the large scale expression, *E. coli* BL21(DE3) cells (Novagen) were transformed with the pET22_NcoI_HindIII_*B_glumae_endoPG* and were inoculated in 50 mL of M9CA medium, 0.4 (wt/vol) glycerol as carbon source and 100 µg/mL ampicillin for selection. The 50 mL starter culture was grown over night at 37 °C and then added to 3.5 L of fresh media. When the culture reached an OD_{600} = 1.0 abs, the protein expression was induced for 15 hours at 12 °C using 0.5 mM IPTG. Then, cells were harvested by centrifugation for 20 minutes at 4200 g at 4 °C. The periplasmic fraction was subsequently isolated as previously described (Chapter 2 , Par. 2.2.2.1).

The enzymatic activity of the periplasmic extract was measured using polygalacturonic acid as substrate, according to the Bernfeld method (Chapter 2, Par. 2.2.1). The periplasmic extract was concentrated using an Amicon concentration cell (Millipore) equipped with a YM10 membrane (10 kDa cut-off) and dialysed against 50 mM sodium

phosphate pH 6.0, 1.7 M ammonium sulphate. The sample was spun down at 20000 g for 10 minutes at 4 °C and then loaded on a hydrophobic interaction Phenyl Sepharose HR 26/10 column (Amersham Biosciences), previously equilibrated with 50 mM sodium phosphate pH 6.0, 1.7 M ammonium sulphate. The column was washed with 2 column volumes of the same buffer before eluting the proteins with a 15 column volumes linear gradient from 1.7 to 0 M ammonium sulphate. Positive fraction were collected, dialyzed against 50 mM acetate buffer, pH 3.5 and then applied to a MonoS (Amersham Biosciences) equilibrated with the same buffer. The elution was performed by applying a 15 column volumes linear gradient of sodium chloride from 0.0 to 0.6 M in the same buffer. Polygalacturonase activity was eluted at a sodium chloride concentration of 160 mM.

Protein purity and molecular weight of the *B. glumae* endoPG were verified by SDS-PAGE, as previously described (Chapter 2, Par. 2.2.1).

Protein concentration was estimated from the measurement of the optical density at 280 nm, using the extinction coefficient and the molecular weight calculated from the protein sequence, using the online ProtParam server (www.expasy.ch/tools/protparam.html).

Acknowledgements

Many people contributed to my scientific formation and my nice staying in Trieste and I feel I should gratefully acknowledge all of them.

First of all, I'd like to acknowledge my supervisor Dr. Doriano Lamba for his continuous support, his enthusiasm and his patience. In particular, I feel in debt with him for the interesting discussions on the scientific problems that I encountered during my PhD. These allowed me to develop a critical approach in the experimental design and in the interpretation of the results.

I'd like to gratefully acknowledge my co-supervisor Dr. Cristiana Campa for her optimistic manner to approach science and life. She has always been ready to share with me her valuable practical experiences and her wide scientific interests.

A special thank to all the members of the Bacteriology Group of International Centre for Genetic Engineering and Biotechnology (Trieste, Italy) for their valuable contributions, material supply, and expertise. Particularly, I'd like to acknowledge Dr. Giuliano Degrassi, Dr. Giulia Devescovi and Dr. Vittorio Venturi. They followed me during the initial steps of my PhD with love and patience.

A special thank to Dr. Corrado Guarnaccia (Protein Structure and Bioinformatics, International Centre for Genetic Engineering and Biotechnology, Trieste, Italy) for his skillful assistance in the protein disulphide bond mapping.

I especially feel in debt with Dr. G. Salvi (Dipartimento di Biologia Vegetale, Universita' La Sapienza di Roma, Italy), Prof. K.J. Jensen (Department of Natural Sciences, The Royal Veterinary and Agricultural University, Denmark) and Prof. R. Madsen and Prof. M.H. Clausen (Department of Chemistry, Technical University of Denmark, Denmark) for the precious material supply. Their contributions allowed me to conduct a significant part of this work.

I'd like to acknowledge Fumihiko Suzuki (Department of Plant Protection, Nishigoshi Kikuchi, Japan) for the material supply.

I'd like to acknowledge all the members, present and past, of the Structural Biology Laboratory (Elettra, Synchrotron Light Facility, Trieste, Italia). Particularly, I feel in debt with Ivet, Alessandro and Ljiljana for their friendship and their smiling support. A