The Three-Dimensional Structure of Pectate Lyase E, a Plant Virulence Factor from *Erwinia chrysanthemi*¹

Susan E. Lietzke, Marilyn D. Yoder, Noel T. Keen, and Frances Jurnak*

Department of Biochemistry (S.E.L., M.D.Y., F.J.), and Department of Plant Pathology (N.T.K.), University of California, Riverside, California 92521

---

The three-dimensional structure of pectate lyase E (PelE) has been determined by crystallographic techniques at a resolution of 2.2 Å. The model includes all 355 amino acids but no solvent, and refines to a crystallographic refinement factor of 20.6%. The polypeptide backbone folds into a large right-handed cylinder, termed a parallel β-helix. Loops of various sizes and conformations protrude from the central helix and probably confer function. A putative Ca²⁺-binding site as well as two cationic sites have been deduced from the location of heavy atom derivatives. Comparison of the PelE and recently determined pectate lyase C (PelC) structures has led to identification of a putative polygalacturonate-binding region in PelE. Structural differences relevant to differences in the enzymatic mechanism and maceration properties of PelE and PelC have been identified. The comparative analysis also reveals a large degree of structural conservation of surface loops in one region as well as an apparent aromatic specificity pocket in the amino-terminal branch. Also discussed is the sequence and possible functional relationship of the pectate lyases with pollen and style plant proteins.

PelE consists of 355 amino acids and has a mol wt of 38,069, as calculated from the amino acid sequence. The PelE structure folds into a parallel β-helix in which the strands of three parallel β-sheets fold into a large central coil. The first X-ray structure of a protein in which the parallel β-helix topology was observed is *E. chrysanthemi* PelC (Yoder et al., 1993a). Although PelE has only 22% sequence identity with PelC, the core structure of both proteins is strikingly similar (Yoder et al., 1993b). There are significant structural differences in the various loops that protrude from and cover the parallel β-helix core of each isozyme. There are also pronounced similarities in the loop conformations in one region.

Abbreviations: B factor, isotropic temperature factor; crystallographic R factor, agreement factor between observed structure factor amplitudes, |F₀|, and calculated structure factor amplitudes, |Fₐ|, based on atomic model; Fₐ, structure factor; F₀, calculated structure factor; MIR, multiple isomorphous replacement; Pel, pectate lyase; PGA, polygalacturonic acid; r.m.s., root mean square; φ, rotation angle about the carbon α-nitrogen bond; ψ, rotation angle about the carbon α-carbon bond; χ, rotation angle about the carbon α-carbon β bond; σ, standard deviation.

¹ F.J. was supported by the United States Department of Agriculture (award No. MCB940899).
and in an apparent aromatic specificity pocket. The differences between the structures of the Pels may explain differences in the enzymatic and maceration properties of the isozymes, and the similarities may be related to folding or secretion mechanisms.

**MATERIALS AND METHODS**

**Materials**

PEG 4000 was obtained from Koch-Light, Ltd. (Buckinghamshire, UK), and PEG 3350 was from J.T. Baker (Phillipsburg, NJ). The heavy atom compound LaCl₃ was purchased from Alfa (Ward Hill, MA), and UO₂(NO₃)₂ was from Fisher. All other reagents were obtained from Sigma.

**Crystal Preparation**

For the present study, PelE was purified in quantity by scaling up the protocol described by Keen and Tamaki (1986). PelE was isolated from the periplasm of E. chrysanthemi EC16 pelE gene. The mature, secreted recombinant protein has the same mol wt, isoelectric point, and maceration properties as PelE isolated from E. chrysanthemi. Small crystals of PelE were grown initially from PEG 4000 by vapor diffusion techniques similar to the conditions reported by Kim et al. (1989), with the exception that CaCl₂ and maceration properties as PelE isolated from heavy atom derivatives were prepared by diffusing the heavy atom compound into crystals grown in 2 weeks at 4°C from a solution of 18 to 19% PEG 4000, 17 mM Li₂SO₄, and 50 mM Tris-HCl, pH 8.0. The diffraction patterns of PelE crystals were consistent with the one molecule per asymmetric unit.

The solvent content was estimated to be 48% using Matthews’ method (1968) with recombinant protein has the same mol wt, isoelectric point, and in an apparent aromatic specificity pocket. The differences between the structures of the Pels may explain differences in the enzymatic and maceration properties of the isozymes, and the similarities may be related to folding or secretion mechanisms.

**Summary of x-ray crystallographic data collection**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Native</th>
<th>UO₂⁺</th>
<th>Pt⁺⁺</th>
<th>La⁺⁺</th>
<th>F⁻</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>2.2</td>
<td>2.2</td>
<td>2.6</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td>Total observations</td>
<td>65,204</td>
<td>50,955</td>
<td>64,239</td>
<td>63,239</td>
<td>64,235</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>18,548</td>
<td>30,982</td>
<td>31,328</td>
<td>31,373</td>
<td>31,403</td>
</tr>
<tr>
<td>Total possible (%)</td>
<td>96.5</td>
<td>87.7</td>
<td>93.6</td>
<td>96.8</td>
<td>93.1</td>
</tr>
<tr>
<td>(I/σ(I))</td>
<td>14.11</td>
<td>11.00</td>
<td>8.49</td>
<td>14.23</td>
<td>11.23</td>
</tr>
<tr>
<td>R&lt;sub&gt;sym&lt;/sub&gt; (unedited)</td>
<td>5.39</td>
<td>8.84</td>
<td>7.58</td>
<td>5.82</td>
<td>7.48</td>
</tr>
<tr>
<td>R&lt;sub&gt;sym&lt;/sub&gt; (edited)</td>
<td>4.68</td>
<td>6.11</td>
<td>5.58</td>
<td>4.23</td>
<td>6.88</td>
</tr>
<tr>
<td>R&lt;sub&gt;scale&lt;/sub&gt;</td>
<td>29.51</td>
<td>12.38</td>
<td>12.36</td>
<td>17.19</td>
<td></td>
</tr>
</tbody>
</table>

* Bijvoets merged for native only.  
* R<sub>sym</sub> = Σ |I<sub>avg</sub> - I<sub>obs</sub>| / ΣI<sub>avg</sub> where I is the average (avg) or the observed (obs) intensity of the reflection. Unedited is before manual rejection of outliers. Edited is after manual rejection of outliers and scaling of I by a factor of a reflection. 
* R<sub>scale</sub> = 2Σ | I<sub>native</sub> - I<sub>deriv</sub> | / Σ I<sub>native</sub> + I<sub>deriv</sub> where F is the structure factor of the native (nat) or derivative (der) data.

**Data Collection and Reduction**

X-ray diffraction data were collected at 20°C using a dual-chamber San Diego Multiwire Systems (San Diego, CA) area detector system installed on a Rigaku (Danvers, MA) rotating-anode x-ray generator with a Supper (Natick, MA) graphite monochromator operated at 45 kV × 120 to 150 mA. Each x-ray data set was collected from a single crystal at helium-path distances of 615 mm and 560 mm between the crystal and each detector and processed with the San Diego Multiwire Systems software package (Howard et al., 1985). Derivative x-ray diffraction data sets were scaled, using the Fourier-Bessel program of Terwilliger and Eisenberg (1983), to the native x-ray data set, which had been placed on an absolute scale by Wilson’s method (Wilson, 1942). The x-ray data collection statistics are summarized in Table I.

**MIR Phase Determination**

Four heavy atom derivative data sets were used to determine the phases by the MIR method. All sites were located by difference Pattersons using the program FFT ("Ten Eyck et al., 1976) and confirmed by cross-phasing the difference Fourier of each derivative by alternate derivatives. Heavy atom parameters were refined by HEAVY (Terwilliger and Eisenberg, 1983) and the absolute configuration of the molecule was determined based on cross-Fourier calculations as described by Blundell and Johnson (1976). The MIR phases were applied to the native structure factor amplitudes to calculate an initial electron density map at 3.2- to 50.0-Å resolution. Although the parallel β helix topology was very clear in the initial map, the connectivity of the electron density in the polypeptide loop regions was improved by solvent density modification techniques (Wang, 1985). A partial αC backbone was traced on minmaps and a polya-
anine model composed of 207 residues was constructed on an Evans and Sutherland (Salt Lake City, UT) PS390 with the program FRODO (Jones, 1985). Partial model phases were combined with heavy atom derivative phases using SIGMAA (Read, 1986) and applied to the structure factors at 2.5 Å; the procedure was reiterated three times. From the series of improved maps, the amino acid side chains were fitted to the electron density and an initial model consisting of 351 of the 355 amino acids was constructed.

Model Refinement

The model was refined with molecular dynamics techniques using the method of simulated annealing (Brünger et al., 1987) with X-PLOR (Brünger, 1993b). After each cycle of crystallographic refinement, the model was checked and manually rebuilt. The final four residues were built into the model from a 2Fo-Fc map, calculated from model phases, after the first cycle of refinement. Three additional cycles of crystallographic and individual isotropic temperature factor refinement were carried out. Data between 2.2- and 5.0-Å resolution were used in the first two cycles of refinement, and data between 2.2 and 8.0 Å were used in the last two cycles of refinement. Throughout the refinement process, clear density was observed for most of the amino acids in a 2Fo-Fc electron density map, contoured at 1σ. There were two exceptions, which were located on surface loops with weak density. The entire model was checked by inspection of annealed omit maps calculated by X-PLOR (Bhat and Cohen, 1984; Hodel et al., 1992).

The atomic coordinates are deposited with the Brookhaven Protein Data Bank under the file name 1PCL.

Model Analysis

The program PROCHECK (Laskowski et al., 1993) was used to calculate stereochemical parameters of the refined PelE structure. Secondary structural assignments were initially made with the Kabach and Sander algorithm (Kabsch and Sander, 1983) in the PROCHECK package, followed by visual inspection of the main-chain hydrogen bonds. The criteria for a hydrogen bond was a donor-acceptor distance of less than 3.5 Å and an angle greater than 120° (Baker and Hubbard, 1984). The PelE and PelC models were compared with the program O (Jones et al., 1991; Jones and Kjeldgaard, 1993) on an Evans and Sutherland ESV graphics system. A least-squares refinement method in the LSQ option was used to superimpose the αC atoms in both structures. The initial alignment was made with the αCs of amino acids in the β strands of the parallel β helix core only.

The distance between the La3+ heavy atom site, assumed to be a putative Ca2+-binding site, and an important invariant region, 144-ViDH-148, was estimated in two ways. First, a minimum distance was estimated by measuring the distance between La3+ and the αC of Asp147, maintaining approximate van der Waals distances to protein atoms. The coordinates for trigalacturonate were obtained from Walkinshaw and Arnott (1981).

Loops were characterized according to the nomenclature proposed by Ring et al. (1992). Linear loops are referred to as strap loops, and nonlinear loops are subdivided into two categories. Omega loops are nonlinear, but planar loops. This definition of omega loops differs from Leszczynski and Rose’s (1986) in that the ends of omega loops are not necessarily in close proximity. The second type of nonlinear loops are zeta loops, which are nonplanar.

Figure Preparation

The Ramachandran plot in Figure 1 was prepared with the program PROCHECK. In Figure 2, the 2Fo-Fc map was calculated by X-PLOR and displayed on O; the x-utility program Oplot (G.J. Klegwegt, unpublished data) was then used to create the illustration. Figures 5, 6, 7, and 8 were produced with the program MOLSCRIPT (Kraulis, 1991). The space-filling models in Figure 9 were created with InSight (Biosym Technologies, Inc., San Diego, CA).

<table>
<thead>
<tr>
<th>Table II. Refined heavy atom derivative parameters for PelE in the resolution range of 2.2 to 50.0 Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>Derivative</td>
</tr>
<tr>
<td>UO2²⁺</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Pt²⁺</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>La³⁺</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

* Heavy atom derivative was refined only to 3.2 Å.

Downloaded from on October 15, 2017 - Published by www.plantphysiol.org
Copyright © 1994 American Society of Plant Biologists. All rights reserved.
RESULTS

Structural Determination

Four heavy atom derivative data sets produced nine heavy atom sites, eight of which were unique. The refined heavy atom parameters are listed in Table II. Although none of the individual sites were strongly substituted, the phasing power, \( F_o/E \), was sufficient for an unambiguous interpretation of the electron density. The overall figure of merit for the MR refinement was 0.54 for 12,589 reflections in the 2.5- to 50.0-Å resolution range. The MR phasing statistics are presented in Table III. Each heavy atom position substituted at a chemically reasonable site. The three uranyl (\( UO_2^{2+} \)) sites substituted near carboxylic acid residues: uranyl at Asp134, Asp177, and Asp177; uranyl at Glu126; and uranyl at Glu126. \( La^+ \) occupied the same site as uranyl, and the four iodine sites substituted at Tyr125 (two sites), Tyr169, and Tyr352. The platinum derivative substituted at the amino terminus with low occupancy and was used only for phasing to a resolution of 3.2 Å.

The PelE model presented herein consists of 2,693 non-hydrogen atoms in all 355 amino acids, but no solvent molecules. With the exception of the tips of two extended loops, the main chain and most side chains have good geometry. The Pro232, which is analogous to cis-Pro220 at the same position in the PelE model also has good geometry. The average \( \chi_1 \) torsion angles agree well with the preferred conformers \( g^- (+60^\circ) \), \( t (+180^\circ) \), and \( g^+ (-60^\circ) \). The average \( \chi_1 \) angles for 289 residues are \( 61.1^\circ \pm 10.6^\circ \), \( 183.5^\circ \pm 10.2^\circ \), and \( -62.9^\circ \pm 9.9^\circ \), respectively, which correlate well with the values calculated by Morris et al. (1992) for well-refined structures at high resolution.

With the exception of the ends of two surface loops previously mentioned, the final 2\( F_o-F_c \) electron density map, contoured at 1σ, shows continuous density for all main-chain atoms. A representative section of the 2\( F_o-F_c \) map is shown in Figure 2. The density of most amino acid side chains, with the exception of some surface side chains, mostly Lys’s, is well resolved. The ambiguous regions of the map are the tips of two extended loops, region 1 consisting of residues 124 to 127 and region 2 consisting of residues 164 to 167. There is consistently weak density in the two regions in all maps, indicating that these regions may have considerable flexibility. The main-chain atoms for residues 164 to 167 are reasonably clear in a \( \sigma_A \) weighted 2\( F_o-F_c \) electron density map (Read, 1986). In contrast, there is minimal density for Ser125 and Gly126, making it difficult to determine the orientations of the main chain as well as the side chains of Glu126 and Asp177. Ser125 and Gly126 have ultimately been modeled on the basis of proper backbone dihedral angles and the minimization of potential close contacts with a symmetry-related molecule. The lack of density may be a result of multiple conformations for the tip of the loop. Confirmation of the correct direction for both loops is provided by the substitution of iodine at

### Table III. PelE MIR phasing statistics

<table>
<thead>
<tr>
<th>Resolution Range (Å)</th>
<th>( F_o/E )</th>
<th>( R_{cuff} )</th>
<th>( F_o/E )</th>
<th>( R_{cuff} )</th>
<th>( F_o/E )</th>
<th>( R_{cuff} )</th>
<th>( F_o/E )</th>
<th>( R_{cuff} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.04</td>
<td>1.52</td>
<td>0.55</td>
<td>0.79</td>
<td>0.55</td>
<td>1.38</td>
<td>0.61</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>5.69</td>
<td>1.96</td>
<td>0.52</td>
<td>1.75</td>
<td>0.47</td>
<td>1.18</td>
<td>0.69</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>4.45</td>
<td>1.23</td>
<td>0.61</td>
<td>1.18</td>
<td>0.67</td>
<td>0.73</td>
<td>0.73</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>3.77</td>
<td>1.05</td>
<td>0.62</td>
<td>1.06</td>
<td>0.64</td>
<td>0.60</td>
<td>0.72</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>3.33</td>
<td>1.01</td>
<td>0.63</td>
<td>0.91</td>
<td>0.63</td>
<td>0.51</td>
<td>0.76</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>3.01</td>
<td>1.09</td>
<td>0.66</td>
<td>0.79</td>
<td>0.74</td>
<td>0.54</td>
<td>0.72</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>2.77</td>
<td>1.09</td>
<td>0.69</td>
<td>0.70</td>
<td>0.72</td>
<td>0.53</td>
<td>0.70</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>2.58</td>
<td>1.06</td>
<td>0.69</td>
<td>0.58</td>
<td>0.73</td>
<td>0.51</td>
<td>0.74</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1.20</td>
<td>0.60</td>
<td>0.94</td>
<td>0.61</td>
<td>0.67</td>
<td>0.71</td>
<td>0.54</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( F_o/E )</th>
<th>( R_{cuff} )</th>
<th>( F_o/E )</th>
<th>( R_{cuff} )</th>
<th>( F_o/E )</th>
<th>( R_{cuff} )</th>
<th>( F_o/E )</th>
<th>( R_{cuff} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.04</td>
<td>1.52</td>
<td>0.55</td>
<td>0.79</td>
<td>0.55</td>
<td>1.38</td>
<td>0.61</td>
<td>0.75</td>
</tr>
<tr>
<td>5.69</td>
<td>1.96</td>
<td>0.52</td>
<td>1.75</td>
<td>0.47</td>
<td>1.18</td>
<td>0.69</td>
<td>0.82</td>
</tr>
<tr>
<td>4.45</td>
<td>1.23</td>
<td>0.61</td>
<td>1.18</td>
<td>0.67</td>
<td>0.73</td>
<td>0.73</td>
<td>0.69</td>
</tr>
<tr>
<td>3.77</td>
<td>1.05</td>
<td>0.62</td>
<td>1.06</td>
<td>0.64</td>
<td>0.60</td>
<td>0.72</td>
<td>0.61</td>
</tr>
<tr>
<td>3.33</td>
<td>1.01</td>
<td>0.63</td>
<td>0.91</td>
<td>0.63</td>
<td>0.51</td>
<td>0.76</td>
<td>0.51</td>
</tr>
<tr>
<td>3.01</td>
<td>1.09</td>
<td>0.66</td>
<td>0.79</td>
<td>0.74</td>
<td>0.54</td>
<td>0.72</td>
<td>0.45</td>
</tr>
<tr>
<td>2.77</td>
<td>1.09</td>
<td>0.69</td>
<td>0.70</td>
<td>0.72</td>
<td>0.53</td>
<td>0.70</td>
<td>0.42</td>
</tr>
<tr>
<td>2.58</td>
<td>1.06</td>
<td>0.69</td>
<td>0.58</td>
<td>0.73</td>
<td>0.51</td>
<td>0.74</td>
<td>0.37</td>
</tr>
<tr>
<td>Total</td>
<td>1.20</td>
<td>0.60</td>
<td>0.94</td>
<td>0.61</td>
<td>0.67</td>
<td>0.71</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Table IV. PelE refinement statistics

<table>
<thead>
<tr>
<th>Resolution range (Å)</th>
<th>Number of reflections ( (F &gt; 2\sigma) ) 16,846</th>
<th>Number of non-H atoms/asymmetric unit 2,693</th>
<th>Crystallographic R factor ( (F &gt; 2\sigma, 90% \text{ data}) ) 21.0%</th>
<th>r.m.s. deviations from ideality</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2-8.0 Å</td>
<td>16,846</td>
<td>2,693</td>
<td>21.0%</td>
<td>Bond length 0.006 Å</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bond angle 1.56°</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Improper angle 1.090°</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Average refined B factor (isotropic)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Main chain atoms 11.5 Å²</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>All non-H atoms 12.5 Å²</td>
</tr>
</tbody>
</table>

\( F_P/E \), Ratio of heavy atom structure factor and the residual lack of closure. \( R_{cuff} \), Cullis R for centric reflections = \( \Sigma |F_{par} - F_p - F_R|/ \Sigma |F_{par} - F_P| \) where \( F_{par} \), \( F_p \), and \( F_R \) are the structure factors of the protein, derivative, and heavy atom, respectively. FOM, Figure of merit.
Tyrr\(^{123}\) and Tyr\(^{189}\), both near the most poorly defined regions of electron density. Two iodine positions were substituted near Tyr\(^{123}\), possibly reflecting alternate conformations of the residue due to flexibility of the side chain and/or region 1.

Because the topology for PelE is unusual, additional measures were considered to substantiate the validity of the model. The average real space correlation coefficient, calculated by O, for main-chain atoms is 0.905. Figure 3 shows that the real space correlation coefficient is good for all parts of the model except the two flexible loop regions, residues 124 to 127 and 164 to 167. The average temperature factor for main-chain atoms is 12.5 \(\text{Å}^2\); a plot of the average B factor for each residue is shown in Figure 3. The highest peaks in the plot correspond to the loop regions, and the valleys correspond to the core of the parallel \(\beta\) helix. Amino acids within the two surface loops with ill-defined electron density have the highest average main-chain B factors, 55 and 45 \(\text{Å}^2\). Finally, the model was characterized by a three-dimensional window profile analysis (Lüthy et al., 1992). Figure 4 shows that the PelE model has fairly high three-dimensional/one-dimensional profile scores. The overall three-dimensional/one-dimensional window profile score per number of residues was 167.3, which is comparable to the value for correct models reported by Lüthy et al. (1992).

**Structural Results**

The PelE molecule is a single structural domain with approximate dimensions of 53 × 52 × 41 Å. The core of the protein consists of three parallel \(\beta\) sheets whose individual parallel \(\beta\) strands coil into a large right-handed cylinder (Fig. 5). The folding topology is similar to that first observed for PelC (Yoder et al., 1993a) and is termed a parallel \(\beta\) helix.
Within PelE, there are seven complete turns in the parallel β helix. A schematic of the PelE polypeptide backbone, highlighting the β structural features, is shown in Figure 5B. Only those amino acids with repetitive φ and ψ angles of classical β structure and maximal hydrogen bond formation between the amide and carbonyl groups on adjacent strands are considered to be β structure. The three parallel β sheets consist of 8, 10, and 6 strands. Within each β sheet, the β strands are relatively short, ranging from two to five residues in length and generating only a small, right-handed twist. As a consequence of the packing arrangement of the parallel β sheets, the cross-section of the parallel β helix is not circular but L-shaped. Two of the β sheets form a parallel β sandwich, with the third parallel β sheet oriented approximately 110° to the parallel β sandwich.

Polar and hydrophobic side chains fill the interior of the protein core and many are involved in extensive stacking interactions. Short stacks of two or three aliphatic amino acids are frequently found on the interior of the parallel β helix. One long stack of Ile's, Ile100, Ile138, Ile178, and Ile201, is found in the interior (fig. 3c, Yoder et al., 1993b). Another extended interior stack is composed of aromatic residues, Phe192, Phe224, Tyr246, and Phe276 (fig. 9a, Yoder et al., 1993b). The aromatic planes are oriented in a manner that resembles base-pair stacking interactions in double-stranded DNA structures. The inter-ring distance is 3.6 Å and the planes of the aromatic side chains are nearly parallel. A third type of stack involves the polar residues Ser and Asn. One example is a short, interior stack composed of Ser144 and Ser148. The second example is the ladder composed of Ser170, Asn222, Asn144, Asn274, and Ser207 (fig. 7b, Yoder et al., 1993b). Both of the stacks form extensive networks of hydrogen bonds that stabilize tight turn connections between β strands. The latter stack is analogous to the six-residue Asn ladder observed in Pea and described in more detail by Yoder et al. (1993b).

Although the structural core of the protein is very regular, no repetition in the primary sequence is detected as a consequence of the variation in the size of the loops that protrude from the central core in the peptide connections between two β strands. One of the three peptide segments connecting the β strands in each helical turn forms a regular secondary structural element consisting of two amino acids. There are seven examples of the unique bend in PelE, in which the first amino acid has average φ and ψ angles of 55.7° and 39.5° and the second amino acid has angles of −106.09° and 160.75°, respectively. The bend resembles a distorted γβ2 turn, with the exception that the αC1−αC1+2 is 9.15 Å in the PelE structure and 7 Å or less in the rare γβ2 turn categorized by Wilmot and Thornton (1990). The remaining peptide connections between the β strands in each helical turn form loops of various sizes and conformations. The loops range from 4 to 23 residues in length. Some loops contain short segments of α or 3_10 helices and others fold into β arches that form many main- and side-chain hydrogen bonds with the main-chain atoms in the core regions of the parallel β helix. The loops cover approximately 25% of the surface of the parallel β helix and cap the ends of the parallel β helix. The N-terminal end of the parallel β helix is covered by an α helix and the C-terminal end is capped by two loops that are connected by a disulfide bond between Cys291 and Cys320. The disulfide bridge has a left-hand conformation with a χ3 angle of −87.7°.

The structurally conserved N- and C-terminal loops run parallel to the axis of the parallel β helix and meet in a...
midregion section. The N terminus, residues 5 to 22, is a zeta-type loop. The C terminus folds into a large omega loop, residues 323 to 339, followed by a 10-residue α helix and a short hook composed of the last 6 residues. Not only do the surface loops protect hydrophobic regions from solvent interactions, but the loops appear to form the functional regions. For example, the most distinctive feature of the PelE surface is a large groove (Fig. 5) formed by five loops, residues 114 to 135, 153 to 175, 201 to 216, 248 to 260, and 297 to 301. The groove contains two cationic binding sites and is analogous to the putative saccharide-binding region in PelC (Yoder et al., 1993a).

Ca²⁺ is required for in vitro activity of all extracellular Pels, but whether Ca²⁺ binds the substrate or the protein has not been established. The locations of heavy atom sites on PelE suggest several potential cationic sites, including a possible Ca²⁺ site. Although Ca²⁺ is not present in the PelE crystals, La³⁺ and uranyl were selected as derivatives because they frequently substitute at Ca²⁺ sites. The single La³⁺ site substituted at the same position as one of the uranyl sites and at a location analogous to the putative Ca²⁺-binding site in PelC. The site lies in the major surface groove and is closely coordinated to the carboxylic acid groups of three Asp residues, Asp¹⁷⁴, Asp¹⁷⁵, and Asp¹⁷⁷, at distances ranging from 2.5 to 2.8 Å. The first two are invariant in all Pels. The third, Asp¹⁷⁷, is invariant in the pelADE subfamily but is a conserved Glu in the pelBC subfamily. The second, strongly substituted uranyl site also lies in the major surface groove, approximately 12.8 Å from the putative Ca²⁺ site. Because only one ligand, O₂ of Glu¹⁷⁹, directly coordinates the second uranyl...
site at a distance of 2.4 Å, it is more likely to be a general cationic site than a Ca\(^{2+}\) site. The third uranyl site is only weakly substituted, with an occupancy of 28.9%. The site is located 2.5 Å from the coordinating atom, Oe2, of Glu\(^3\), suggesting another potential cationic binding site.

**Comparison of PelE and PelC**

The three-dimensional structure of PelE was superimposed upon the PelC structure by first aligning the \(\alpha\)C of the parallel \(\beta\) strands within the parallel \(\beta\) helix core and then minimizing the r.m.s. distances between \(\alpha\)C in both models. The results of the superposition, illustrating the alignment of secondary structural elements and \(\alpha\)C backbones, are shown in Figures 6 and 7. As evident in Figure 6, the core parallel \(\beta\) helix aligns very well, within an r.m.s. deviation of 1.02 Å for the \(\alpha\)C of 81 amino acids (Yoder et al., 1993b). The r.m.s. deviation of the \(\alpha\)C is 2° or less for 197 residues and 3° or less for 232 residues. Sixty eight of 198 residues with \(\alpha\)C r.m.s. deviations of 2° or less are invariant between PelE and PelC. Both Pels form seven complete turns of the parallel \(\beta\) helix. The number of amino acids per \(\beta\) strand and the number of parallel \(\beta\) strands per \(\beta\) sheet vary only slightly between the two enzymes. As noted previously, most of the amino acids found on the interior of the core of PelE and PelC are involved in extensive side-chain stacking interactions. However, the length of the homologous stacks differs: the Asn ladder of six residues in PelC is only three in PelE, Asn\(^{222}\), Asn\(^{244}\), and Asn\(^{274}\); the Ser stack of three residues in PelC is only two in PelE, Ser\(^{141}\) and Ser\(^{182}\); and the aromatic stack of three residues in PelE is four in PelE, Phe\(^{192}\), Phe\(^{224}\), Tyr\(^{246}\), and Phe\(^{276}\). The heterologous hydrophobic stack of Ile\(^{107}\), Val\(^{135}\), and Ile\(^{178}\) in PelC is a homologous Ile stack of four residues in PelE, Ile\(^{100}\), Ile\(^{130}\), Ile\(^{178}\), and Ile\(^{201}\).

As illustrated in Figures 6 and 7, many of the loops that protrude from and cover the central parallel \(\beta\) helix occur at analogous positions in PelE and PelC. Eight loops are structurally conserved. Four of these loops are found in the N- or C-terminal regions, and include the N-terminal zeta loop, residues 9 to 22; the \(\alpha\) helix covering the amino end of the parallel \(\beta\) helix, residues 35 to 41; a large omega loop, residues 328 to 339; and the C-terminal \(\alpha\) helix and hook, residues 345 to 355. The remaining structurally conserved loops are short loops that protrude from the parallel \(\beta\) helix core. The degree of structural conservation is somewhat surprising because both the N-terminal and C-terminal loops have unusual conformations. Three of these four conserved loops are spatially adjacent to one another and meet in a midregion that is located on the opposite external surface of the parallel \(\beta\) helix as the putative Ca\(^{2+}\) site. In the N-terminal loop, the similarities between PelE and PelC extend beyond the peptide conformation to a structural cavity that appears to be a specificity pocket for an aromatic residue. In contrast to the results of primary sequence alignments (Hinton et al., 1989), Trp\(^{11}\) in PelE is structurally aligned with Tyr\(^{7}\) in PelC. The side chains of both the aromatic residues are located in pockets surrounded by a network of seven ringed residues, five of which are identical in PelE and PelC as well as highly conserved in the \(pelADE\) and \(pelBC\) subfamilies. The conserved amino acids in the network include His\(^{220}\), His\(^{240}\), Tyr\(^{242}\), Tyr\(^{331}\), and Tyr\(^{333}\). Tyr\(^{325}\) is similar to a Tyr in PelC; however, its centroid is positioned 3 Å farther away from the centroid of the Tyr in PelC. Tyr\(^{184}\) is not conserved in PelC. As shown in Figure 8, most of the ringed residues interact in the edge-to-face manner that is common among proteins (Burley and Petsko, 1985; Singh and Thornton, 1985) rather...
than in the base stacking mode observed within the core of the parallel \( \beta \) helix.

The remaining loops that protrude from the parallel \( \beta \) helix are conserved neither in length nor in conformation. Most of the nonconserved loops are located around the putative Ca\(^{2+}\) site and form a cleft on the surface of PelE or PelC. The difference in the number, size, and conformation of the loops that constitute the cleft is the most significant structural difference between the two proteins. Five loops form the cleft region in PelE and consist of residues 114 to 135, 153 to 175, 201 to 216, 248 to 260, and 297 to 301. The cleft in PelC is formed by residues in six loops, 36 to 51, 57 to 80, 126 to 130, 150 to 168, 267 to 274, and 292 to 306. The shape of the PelC cleft differs substantially from PelE. The PelC groove is longer as a consequence of the loop composed of residues 36 to 51 at the amino end of the parallel \( \beta \) helix. Also, the Cys\(^{72}\)-Cys\(^{105}\) bond, which connects two loops in PelC, residues 57 to 80 and 150 to 168, presumably constrains the loops into a more rigid conformation with close interaction. In contrast, the analogous loops in PelE are conforma-

Figure 7. A, Stereo view of the superposition of the ribbon backbones of PelE (in dark gray) and PelC (in light gray). The single PelE disulfide bond is illustrated by the thickest black line, and the two disulfide bonds in PelC are illustrated by thick dark gray lines. The putative Ca\(^{2+}\) ions are marked by spheres, PelE in dark gray and PelC in light gray. The side chains of 59 invariant residues between PelE and PelC are shown in black for PelE only. The nine invariant Gly's are illustrated with small dark gray spheres at their \( \alpha C \) position. B, The same figure as in A, rotated 80° in the vertical axis.
tionally flexible. The disulfide bond in the PelC cleft is one of two that are invariant in the pelBC subfamily; the second invariant disulfide bond is Cys335-Cys352. Neither of the two invariant disulfide bonds in the PelC subfamily are analogous to the single invariant disulfide bond in the PelE subfamily. In PelE, the Cys291-Cys320 bond connects two loops that cap the C-terminal end of the parallel β helix. In PelC, the second disulfide bond appears to stabilize a C-terminal loop, which is five residues longer than in PelE and extends back up to the carboxy end of the parallel β helix.

Two views of space-filling models of PelE and PelC are compared in Figure 9, with the charged amino acids coded by color. Surprisingly, the distribution of the surface charge and the overall shape of the two proteins are very different. In PelE, the charged groups are more uniformly distributed over the entire surface of the protein than is observed in PelC. One exception is a small region around the putative Ca2+ site. The majority of charged amino acids in PelC cluster in a narrow groove that includes the putative Ca2+ site and extends 51 Å. The size and shape of the groove in PelC approximate the dimensions of a dodecamer of galacturonate, leading Yoder et al. (1993a) to speculate that the region is the saccharide-binding site in PelC. The remainder of the PelC surface is predominantly neutral. Given the surface charge distribution in PelE, a potential saccharide-binding site is ambiguous. In addition to the charge distribution, another pronounced difference is the overall shape. PelE is quite globular, whereas PelC is more elongated.

DISCUSSION

The three-dimensional structure of *E. chrysanthemi* PelE is the most recent example of the newest structural class of proteins in which the predominant tertiary feature is an all-parallel β motif (Cohen, 1993; Sprang, 1993; Yoder et al., 1993a). The structural family also includes the related Pels *E. chrysanthemi* PelC (Yoder et al., 1993a) and *Bacillus subtilis* PelE (Pickersgill et al., 1994), as well as a partial domain of the *Pseudomonas aeruginosa* alkaline protease (Baumann et al., 1993). In all the structures, the parallel β strands are folded into a right-hand coil. In addition to the novel tertiary structure, the proteins have special interactions that contribute to their stability. In the Pels, the side chains of the amino acids on the interior of the core form extensive linear arrays of stacking interactions; in the protease structure, five interior Ca2+ ions stabilize short polypeptide bends between β strands through coordination to the protein. A major question left unanswered is the functional role, if any, of the novel topology. It is quite clear that the extensive network of intrastrand H bonding between parallel β sheets and the special types of side-chain or Ca2+ interactions are responsible for the observed stability of the proteins in solution. Such stability is, no doubt, required of extracellular proteins secreted into a hostile environment. What is more intriguing is the possibility that the novel parallel β motifs may aid in the secretion of each protein through the outer membrane of its respective organism or in its subsequent pathogenic function. Answers to these questions must await further investigation.

*E. chrysanthemi* PelE and PelC are isozymes sharing 22% sequence identity. Their genes are independently regulated (Collmer and Keen, 1986), yet both have similar enzymatic, secretory, and pathogenic properties. Although there are limited biochemical data regarding the mode of action of Pels, some details are known to differ. PelE and PelC cleave α-1,4-PGA in vitro in the presence of Ca2+ by a β-elimination reaction, but the end product of cleavage is a dimer for PelE.
and a trimer for PelC (Preston et al., 1992). The same study suggests that the minimal substrate length for normal cleavage by either PelE or PelC is a hexamer of galacturonic acid. The pH of plant surfaces and tissues is 6 or lower (Yang et al., 1992); however, the pH optimum for both Pels is much higher: 9.0 for PelE and 9.5 for PelC (Kotoujansky, 1987; M. Garrett, F. Jurnak, N.T. Keen, unpublished observations). *E. chrysanthemi* PelE and PelC appear to have the same host specificity, but the maceration rate of recombinant PelE on potato tubers is 10-fold that of recombinant PelC (Tamaki et al., 1988). Such subtle differences in functional properties are likely to be correlated with small but detectable structural differences.

A comparison of the PelE and PelC structures was undertaken at atomic resolution to elucidate any distinguishing features relevant to their functional properties. Not surprisingly, the overall three-dimensional structures of PelE and PelC are very similar. The major differences between the two enzymes appear to be in the size and conformation of the loops that protrude from the core of the parallel β helix. The
surface loops dramatically alter the surface charge distribution and overall shape of the two enzymes. The cluster of charged amino acids in PelC in a narrow groove led Yoder et al. (1993a) to speculate that this region is a binding site for a negatively charged polygalacturonate substrate in PelC. A similar inference is not possible with only the PelE structure because the distribution of the surface charges is more random. The lack of an extensive charge cluster on the PelE surface suggests that the tentative identification of the pectinolytic active site in PelC must be re-evaluated and other possibilities considered.

The most plausible pectinolytic active site still remains the region around the putative Ca\(^{2+}\)-binding site. Ca\(^{2+}\) is essential for in vitro Pel activity, but its role is undetermined. A putative Ca\(^{2+}\) site toward one end of the charged groove in PelC was identified by the substitution results of three Ca\(^{2+}\)-like heavy atom derivatives. Uranyl, Pb\(^{2+}\), and Lu\(^{3+}\) substituted at the same atomic location in PelC, each with a reasonable coordination to invariant or highly conserved amino acids in all Pels. In PelE, La\(^{3+}\) and one of three uranyl sites substitute at an analogous location to that in PelC with a similar coordination sphere. A second uranyl site, observed only in PelE, substitutes strongly 12.8 Å from the putative Ca\(^{2+}\) site. The coordination to a single, nonconserved amino acid as well as the lack of substitution by the La\(^{3+}\) derivative suggests that the second uranyl site is a cationic binding site, but not specific for Ca\(^{2+}\). The significance of a second cationic site in PelE, but not in PelC, lies in a possible difference in the mode of saccharide binding between the two enzymes. PGA is a saccharide that contains repetitive, negatively charged uronate groups. Potentially, a charged substrate could bind to the protein via one or more ionic interactions. In the Pels, ionic interactions with the substrate are likely to occur via Lys's or Arg's. In the putative saccharide-binding regions, PelE and PelC share four positively charged amino acids: Lys\(^{101}\), Lys\(^{197}\), Arg\(^{230}\), and Arg\(^{275}\) in PelE nomenclature. Another potential source of an ionic interaction is the essential Ca\(^{2+}\) found in all Pels. In addition, the putative cationic site in PelE, spaced 12.8 Å from a Ca\(^{2+}\) site, is suggestive of another strong ionic interaction between PelE and the substrate, thereby reducing the need for many of the positively charged amino acids found in the groove of PelC. Alternatively, PelE may have a greater affinity for a different saccharide substrate in vivo than PelC. Given the relatively low number of positively charged amino acids in this region of PelE compared to that in PelC, it is quite plausible that the in vivo PelE substrate may not consist entirely of galacturonic acid units. A difference in saccharide specificity in vivo may be responsible for some of the differences in maceration rates between the two enzymes. Unfortunately, no studies have been carried out that define the optimal saccharide substrate for any of the Pels.

The results of the PelE and PelC structural comparison is surprising, not only for the observed differences in the loop structures around the putative Ca\(^{2+}\) site but also for the structural conservation of some loops with unusual conformations and for the discovery of an apparent aromatic specificity pocket. The most striking similarity between PelE and PelC is the conformation of a segment of the N-terminal loop and the large omega loop, neither of which shares extensive sequence homology among the Pels. Both loops fold over and cover one exterior surface of the parallel β helix. The two loops meet in a mid-region that has highly conserved sequences, not only among pectate and pectin lyases but among other plant homologs. Many of the invariant amino acids in this region (Kuster-Van Someren, 1991) cluster around the Asn ladder in PelE and PelC, with several invariant residues lying on the exterior of the parallel β helix. With the exception of the invariant Trp, Asp, and His in a highly conserved vWiDH sequence, most of the invariant amino acids are not exposed to the bulk solvent region. Curiously, a highly conserved aromatic residue is located at the first turn of the N-terminal branch. The side chain of Trp\(^{11}\) in PelE and Tyr\(^{7}\) in PelC is oriented in a cavity surrounded on one side by five invariant ringed amino acids located on the exterior surface of the parallel β helix or on the interior surface of the large omega loop (Fig. 8). The presence of an apparent aromatic specificity pocket suggests an essential role, perhaps an alignment function, for the conserved aromatic residue in the N-terminal branch. The entire region is also a potential candidate for the pectinolytic active site. The catalytic amino acids would most likely include the invariant Asp\(^{147}\) and His\(^{146}\), whose side chains form a hydrogen bond and have a structural orientation reminiscent of the catalytic amino acids in Ser proteases. The concentration of invariant and highly conserved ringed residues in the region also is suggestive of a common mode of saccharide-protein interactions involving aromatic groups. However, if this region were the pectinolytic active site, it would be difficult to postulate an essential role for the Ca\(^{2+}\), whose putative binding site lies at approximately 65 to 95 Å from Asp\(^{147}\) around the circumference of the parallel β helix. Given the large separation between the two regions, it is rather unlikely that both regions are part of the pectinolytic active site. Rather it is more probable that Pels have another, as-yet unidentified enzymatic function that may be related to secretory or pathogenesis.

Several investigators have noted that the Pels share sequence homologies with pollen and style plant proteins (Wing et al., 1989; Kuster-Van Someren, 1991; Rafnar et al., 1991). By comparison with the PelE and PelC structures, the homologies include all amino acids except Lys\(^{178}\) around the putative Ca\(^{2+}\)-binding site, two invariant Arg's in the putative saccharide binding region, the three-residue Asn ladder observed in PelE, and the highly conserved vWiDH region. The homologies suggest that not only are the plant proteins likely to have the parallel β helix topology, but they are also likely to share similar functions, such as some type of saccharidase activity. Such functional similarities have important ramifications for pollination as well as for the allergenic response in which the pollen homologs play a major role.

In the last 30 years, most macromolecular structural determinations have been preceded by a plethora of biochemical studies to facilitate the structural interpretation of functional domains. Because Pels are generally secreted as isoamylases, which pose a purification problem, the enzymes were not suitable for crystallographic analyses until the genes for single isoamylases had been cloned and overexpressed in E. coli. By coupling recombinant DNA technology with advancements in crystallographic methodology, the Pel system now dem-
onstrates that it can be more expedient to obtain detailed information at atomic resolution than to characterize an enzyme by other methods. Unfortunately, the lack of support by biochemical data for Pels has complicated the task of correlating the structures with functional properties. However, comparison of the PelE and PelC structures has revealed unforeseen structural differences in loops and surface properties that may ultimately be correlated with differences in enzymatic and maceration properties of the virulence factors. Even more unexpected is the structural conservation of the N- and C-terminal regions and the discovery of an apparent specificity pocket for an aromatic residue. Just as the initial structure of a Pel was a surprise, the first comparative analysis of two Pels suggests that more surprises are likely to follow when the enzymatic, secretory, and pathogenic functions of these virulence factors are probed.

ACKNOWLEDGMENTS

F.J. gratefully acknowledges the support of the Academic Computing Graphics and Visual Imaging Lab, University of California, Riverside, and the San Diego Supercomputer Center.

Received March 29, 1994; accepted July 1, 1994.
Copyright Clearance Center: 0032-0889/94/106/0849/14.

LITERATURE CITED

Brünger AT (1993b) X-PLOR Manual, Version 3.1. Yale University, New Haven, CT
He SY, Schoedel C, Chatterjee AK, Collmer A (1991a) Cloned Erwinia chrysanthemi out genes enable Escherichia coli to selectively secrete a diverse family of heterologous proteins to its milieu. Proc Natl Acad Sci USA 88: 1079–1083


Wing RA, Yamaguchi J, Larabell SK, Ursin VM, McCormick S (1989) Molecular and genetic characterization of two pollen-expressed genes that have sequence similarity to pectate lyases of the plant pathogen *Erwinia*. *Plant Mol Biol* 14: 17–23

