Type II protein secretion in gram-negative pathogenic bacteria: the study of the structure/secretion relationships of the cellulase cel5 (formerly EGZ) from Erwinia chrysanthemi

Virginie Chapon, Mirjam Czjzek, Mohammed Hassouni, Beatrice Py, Michel Juy, Frederic Barras

Type II Protein Secretion in Gram-negative Pathogenic Bacteria: The Study of the Structure/Secretion Relationships of the Cellulase Cel5 (formerly EGZ) from *Erwinia chrysanthemi*

Virginie Chapon[1], Mirjam Czjzek[2]*, Mohammed El Hassouni[1]
Béatrice Py[1], Michel Juy[2] and Frédéric Barras[1]*

[1]*Laboratoire de Chimie Bactérienne (LCB) and*

[2]*Laboratoire d'Architecture Fonctionnelle des Macromolécules Biologiques (AFMB), Institut de Biologie Structurale et Microbiologie CNRS-Marseille, 31, chemin Joseph-Aiguier, 13402 Marseille Cedex 20, France*

*Corresponding authors

Erwinia chrysanthemi, a Gram-negative plant pathogen, secretes the cellulase Cel5 (formerly EGZ) *via* the type II secretion pathway (referred to as Out). Cel5 is composed of two domains, a large N-terminal catalytic domain (390 amino acid residues) and a small C-terminal cellulose-binding domain (62 amino acid residues) separated by a linker region. A combination of mutagenesis and structural analysis permitted us to investigate the structure/secretion relationships with respect to the catalytic domain of Cel5. The 3D structure of the catalytic domain was solved by molecular replacement at 2.3 Å resolution. Cel5 exhibits the $(\beta/\alpha)_8$ structural fold and two extra-barrel features. Our previous genetic study based upon tRNA-mediated suppression allowed us to predict positions of importance in the molecule in relation to structure and catalysis. Remarkably, all of the predictions proved to be correct when compared with the present structural information. Mutations of Arg57, which is located at the heart of the catalytic domain, allowed us to test the consequences of structural modifications on the secretion efficiency. The results revealed that secretability imposes remarkably strong constraints upon folding. In particular, an Arg-to-His mutation yielded a species that folded to a stable conformation close to, but distinct from the wild-type, which however was not secretable. We discuss the relationships between folding of a protein in the periplasm, en route to the cell exterior, and presentation of secretion information. We propose that different solutions have been selected for type II secreted exoproteins in order to meet the constraints imposed by their interaction with their respective secretion machineries. We propose that evolutionary pressure has led to the adaptation of different secretion motifs for different type II exoproteins.

© 2001 Academic Press

Keywords: crystal structure; molecular replacement; clan GH-A; glycosyl hydrolase; secretion motif

Introduction

The pathogenicity of a large number of Gram-negative bacteria relies on their ability to secrete extracellular enzymes into the external medium. Examples of these extracellular enzymes include depolymerising enzymes, such as cellulases or pectinases, secreted by the plant-pathogenic bacteria *Erwinia*, *Pseudomonas* and *Xanthomonas*, and various toxins, secreted by animal or human pathogens like *Vibrio cholerae*, *Pseudomonas aeruginosa* and *Aeromonas salmonicida*.[1−3] All of these very diverse proteins cited above are secreted across the

E-mail addresses of the corresponding authors:
czjzek@ibsm.cnrs-mrs.fr; barras@ibsm.cnrs-mrs.fr

outer membrane by using the so-called type II secretory pathway, also referred to as GSP for general secretory pathway. The type II secretion machineries are made up of at least 13 proteins (referred to as Out) in *Erwinia chrysanthemi*[4] the sequences of which exhibit a high level of conservation, throughout the taxonomically diverse Gram-negative bacteria.

Targeting proteins to cell compartments that are different from the site of their synthesis implies the existence of information within the structure of the targeted polypeptides, namely a secretion motif, which is recognized by the specific secretion machinery. Despite their biochemical and structural diversity, it is currently believed that exoproteins secreted by the same machinery exhibit a similar secretion motif. In order to characterize the type II specific secretion motif, we have chosen to study the secreted cellulase Cel5 (formerly EGZ) from *E. chrysanthemi* as a model.

Cel5 is initially synthesized as a precursor containing a classical N-terminal signal peptide that is cleaved upon crossing the inner membrane *via* the well known Sec pathway. Mature secreted Cel5 is composed of two functionally and structurally independent domains.[5] The N-terminal domain is a catalytic domain, belonging to the glycoside hydrolase family 5,[6,7] sub-family 2, which is responsible for the catalytic cleavage of the β-1,4 glycosidic bonds of soluble derivatives of cellulose, while the C-terminal domain is a cellulose-binding domain (CBD) of family V,[8] that promotes binding to crystalline cellulose. The two domains are separated by a typical Thr-rich, flexible linker region of 35 residues. Surprisingly, all three regions were found necessary for Cel5 to be secreted across the outer membrane.[9] We concluded that the secretion motif is constituted by residues located in all three domains that are brought in close vicinity upon folding of Cel5.

Structural and functional studies of the CBD indicated that this C-terminal domain of Cel5 folds before Cel5 is secreted.[10–12] Recently, we proposed that, prior to secretion, a folded form of the CBD interacts physically with the catalytic domain, resulting in a transient secretion-specific conformer.[13] However, information about the folding status of the catalytic domain within this putative secretion-specific conformer was still lacking. Therefore, we decided to investigate to what extent the catalytic domain folds prior to secretion.

In this study, we report a detailed analysis of the catalytic domain by combining 3D structure analysis and mutagenesis study of a structural position to test the relationships between structure and secretion. Overall, our results indicate that secretion imposes a strong constraint on the structure of the catalytic domain that must fold prior to secretion.

Results

Overall structure of the catalytic domain of Cel5

The structure of three non-crystallographic symmetry-related copies of the catalytic domain of Cel5, belonging to family 5 of the glycoside hydrolase enzymes, have been determined by X-ray crystallography at 2.3 Å resolution using the molecular replacement method. All three chains are visible from Ser1 through Ala291 (Figure 1(a)). Since the crystallization experiments were performed starting with a full-length protein solution, the true C-terminal residue is not known. No electron density for residues beyond Ala291 was visible for either of the three copies. In contrast, all ten arginine residues are clearly defined in electron density
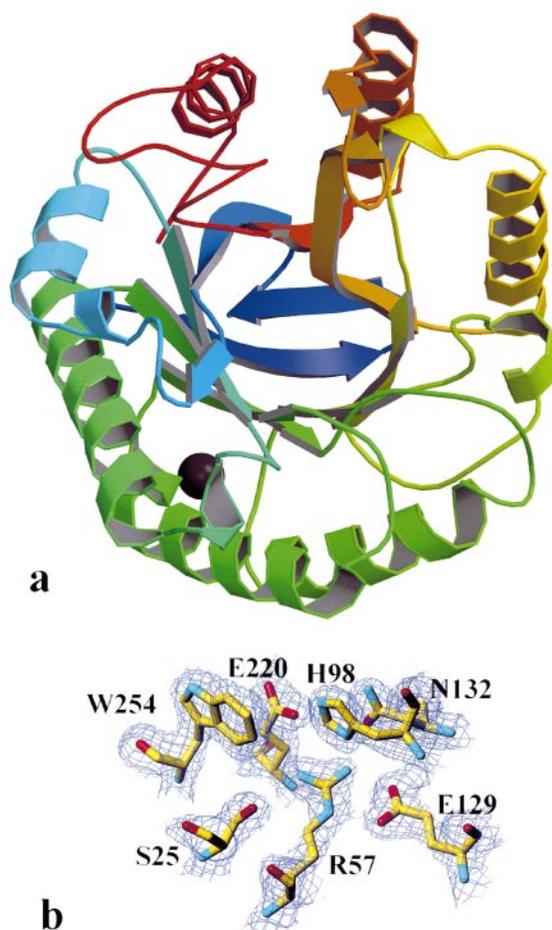
Figure 1. (a) Ribbon representation of the $(\beta/\alpha)_8$ fold of the catalytic domain of Cel5 from *E. chrysanthemi*. The ribbon representation is colour-coded in a gradual manner, from blue (N-terminal) to red (C-terminal). The black ball represents the structural ion (Ca^{2+}, Mg^{2+}). An extra-barrel two stranded β-sheet in blue is located at the N-terminal end of the barrel. The Figure was produced with MOLSCRIPT[51] and Raster3D.[52] (b) Final $2F_o - F_c$ electron density map, contoured at a 1σ level, showing Arg57 and the surrounding residues at the catalytic active centre.

with none of them being disordered (Figure 1(b)). However, Arg81, Arg111 and Arg118 display high B-factor values (around 40 Å^2) for atoms of the guanidinium groups that are nevertheless visible in density.

The structure of the Cel5 catalytic domain consists of a single domain having the $(\beta/\alpha)_8$ fold in which a short irregular chain segment replaces helix 5 (Figure 1(a)). Another extra-barrel element is present: it includes residues 3-8 and 11-15, forming a small anti-parallel β-sheet, that closes the β-barrel at its N-terminal side (Figure 1(a)). With the exception of the two elements noted above, the structure of *E. chrysanthemi* Cel5 is essentially the same as that of the other family 5 enzymes described earlier.[14-16] The closest structure is that of the endoglucanase of *Bacillus agaradherans*[17] (pdb code: 1A3H), which belongs to subfamily 5-2 as Cel5 does. Therefore, the coordinates of this latter structure have been used as the starting model for the molecular replacement calculations. The refinement of the three independent copies (28,266 degrees of freedom) against the 32,807 unique reflections at 2.3 Å resolution led to a final R-factor of 18.7 % and an R_{free}-factor of 22.6 % (Table 1).

Comparison to the other family 5 enzymes

The final structural model of *E. chrysanthemi* Cel5 has approximately 260 out of 290 residues in structurally equivalent positions in *B. agaradherans* Cel5A,[17] which superimposes with an r.m.s C$^\alpha$ deviation of about 0.98 Å. The structures of three other family 5 cellulase enzymes, belonging to sub-family 5-1 (endocellulase E1 from *Acidothermus cellulolyticus*),[16] 5-3 (endoglucanase CelC from *Clostridium thermocellum*)[14] and 5-4 (endoglucanase CelCCA from *Clostridium cellulolyticum*),[15] respectively, have been determined in the past and superimpose with an r.m.s C$^\alpha$ deviation of 1.84 Å (221 out of 358 atoms), 1.87 Å (191 out of 340 atoms) and 1.87 Å (207 out of 380 atoms), respectively. This reflects the low degree of similarity observed for the details of the structures, also apparent when aligning the sequences of these different enzymes. The sequence identities can be as low as 10 % for these enzymes, displaying the same overall 3D structural fold.[17]

Significant differences are observed mostly in the length, the conformation and the flexibility of the loops connecting the β-strands to the adjacent α-helices (Figure 1(a)). In *E. chrysanthemi* Cel5, all the loops are shorter than in any of the other structures reported so far. The long groove or active-site cleft, in which numerous aromatic residues line up to form the sugar-binding sub-sites is therefore extremely open to the solvent, compared to other family 5 cellulases.

A number of differences, however, between Cel5 and the other enzymes of the same family are located at the N-terminal ends of the β-strands (Figure 1(a)). For instance, a short extra-barrel β-sheet reminiscent of that described above, exists in Cel5A.[17] However, neither primary nor tertiary structural similarity is shared by extra-barrel β-sheets of *B. agaradherans* Cel5A or *E. chrysanthemi* Cel5. A similar feature is observed in the cellulase of subfamily 5-1,[16] but the sheet is composed of three β-strands. Furthermore, in Cel5 the connective loops between α-helix 3 and β-sheet 4 and between α-helix 4 and β-sheet 5 coordinate a calcium (or magnesium) ion by the carbonyl oxygen atom of Gly121 and the aspartic acid residues and the asparagine of the sequence DPDN, comprising residues 158-161. The ion is present in all three non-crystallographic symmetry-related copies of the catalytic domain of Cel5, with the same occupation. This ion-binding site is unique to this type of catalytic domain. Other calcium ions have been observed in cellulase and xylanase catalytic domains,[18–20] but they are located in large, flexible surface loops surrounding the active-site cleft, and probably stabilize the loop structures. In contrast, the calcium ion, present in *E. chrysanthemi* Cel5, is located on the N-terminal side of the barrel in tight turns between the main structural elements. The question therefore arises of whether the presence of this ion is due to the crystallization conditions, or is biologically significant.

Role of conserved residues with respect to catalytic activity

As has been reported earlier, five strictly conserved residues form the catalytic site. These residues constitute the basis of the classification of *E. chrysanthemi* Cel5 in glycosyl hydrolase family 5.[6,21] In Cel5 all these residues have been substituted for at least 13 other residues and the mutation pattern and activity have been reported.[22]

Our previous genetic analysis identified residues Glu133 and Glu220 as the nucleophile and acid/base, respectively.[22] In the active site of the 3D structural model, these two residues are separated by a distance of 5.5 Å, which is typical for retaining enzymes (Figure 2(a)). Furthermore, two histidine residues (98 and 192 in Cel5) were found to be crucial for catalysis and are conserved equally throughout all the family 5 enzymes (Figure 2(a)). The geometry of the arrangement of these four residues is almost identical for all family 5 enzymes, with the exception of the enzyme CelC of family 5-3, for which, in the native form of the enzyme, the nucleophile was turned outside of the active site.[14] The fifth strictly conserved residue, Arg57, is buried in the heart of the protein, in close proximity to the catalytic active site (Figure 2(a)). It is involved in numerous hydrogen bonds, many of which are formed with the catalytic residues. Arg57 however does not seem to be involved in the catalytic reaction directly but rather be very important for the stability of the structure of domain (see below and Figure 3(b)).[22]

Table 1. Statistics on data collection, molecular replacement and refinement

Data
Space group	C2				
Unit cell parameters (Å)	$a = 168.8$	$b = 47.2$	$c = 120.1$	$\beta = 105\,^\circ$	
Resolution range (Å)	29.0-2.3	Last shell	2.34-2.3		
Unique reflections	38,365		1749		
Completeness (%)	98		97		
$I/\sigma(I)$	11.4		2.2		
R_{sym} (%)	8.6		37.3		

Molecular replacement

						Corr. coef.	R-factor (%)
Resolution range (Å)	8.0-3.0						
Rotation function							
Sol 1:	133.3	163.1		206.5		14.0	
Sol 2:	8.0	36.9		322.9		13.1	
Sol 3:	51.2	168.7		43.8		10.0	
Rotation and translation							
133.3	163.1	206.5	0.0425	0.000	0.2752	21.9	50.2
28.0	36.9	322.9	0.2821	0.407	0.3497	26.7	49.1
51.2	168.7	43.8	0.8621	0.542	0.0424	35.1	46.3
After one cycle of rigid body refinement						**43.1**	**43.9**

Refinement
Resolution range (Å)	29.0-2.3
Number of atoms	7142
Final R-factor (%) No. of reflect.	18.7/32,807
Free R-factor (%) No. of reflect.	22.6/1616
rmsd from ideal	
Bond distances (Å)	0.007
Bond angles (deg.)	1.65
Average B-factor (Å^2)	
Overall	22.4
Mol A	19.4
Mol B	20.1
Mol C	25.3
Water molecules	39.2

### Role of His100 and His198 residues for activity

Remarkably, some of the residues, supposed to be implied in stacking interactions with the substrate in the aglycon sub-sites further away from the cleavage site, are not always conserved, even within one sub-family. Thus, His100, does not have a counterpart in Cel5A of *B. agaradherans* (Figure 2(a)). This was surprising, since mutations at this position in Cel5 of *E. chrysanthemi* led to an important decrease of the catalytic activity.[22] Moreover, it is located in close proximity to the active site, forming hydrogen bonds to the catalytic residues. As a second example, one can mention Tyr66 of *B. agaradherans* that is not conserved in Cel5 of *E. chrysanthemi* but replaced by a tryptophan residue Trp28, which comes to lie approximately in the same position as Tyr66 but is donated by a topologically inequivalent position (Figure 2(b)). These findings underline the arguments put forward by Davies *et al.*,[17] proposing a possible convergent evolution for functionally equivalent, but structurally unrelated, aglycon-binding sites. Here, the evolution appears even within one sub-family.

The His198 residue is also possibly involved in an aglycon-binding sub-site. This position exhibited a semitolerant pattern in the mutagenesis studies of Cel5.[22] In the structure of Cel5A of *B. agaradherans*, the analogous residue (i.e. His206) was found on the surface of the molecule, totally exposed to solvent and not directed towards the active-site cleft. The authors opened the possibility, that after a conformational change, the residue could play a role at the extreme reducing end of the substrate-binding cleft.[17] A different crystal form of Cel5A of *B. agaradherans* indeed confirmed that this histidine residue could undergo a conformational change (G. Davies, personal communication). In the structure of *E. chrysanthemi* Cel5, His198 is lining the substrate-binding cleft at its extreme end, in which the sugar rings are supposed, to stack in order to be directed towards the cleavage site. This geometry supports the proposition based on the mutagenesis study of *E. chrysanthemi* Cel5, that it is involved in, but not crucial for, substrate binding (Figure 2(c)).
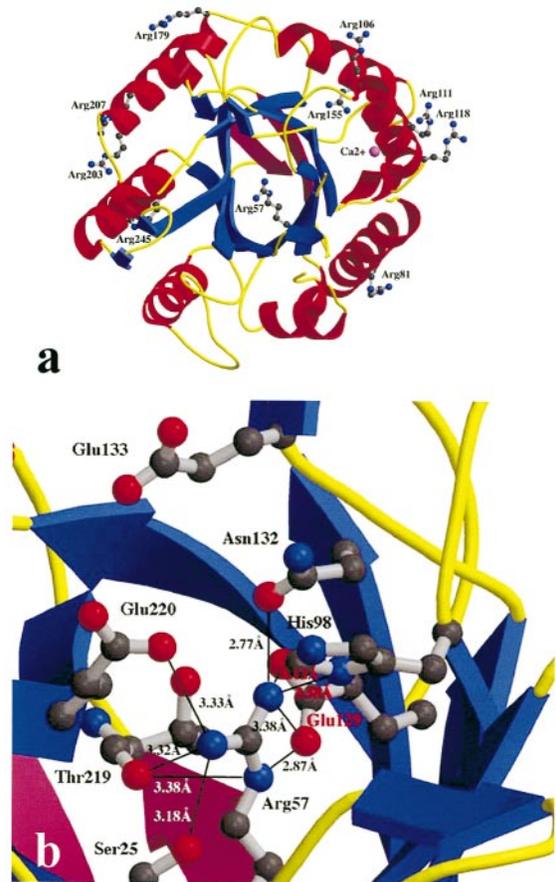
Comparison of Cel5 structure to that of other proteins secreted by the Out system

The 3D structure of Cel5 and its CBD were compared to the 3D structure of the pectate lyases from *E. chrysanthemi*,[23] secreted by the same pathway. The molecular surfaces and/or envelopes of
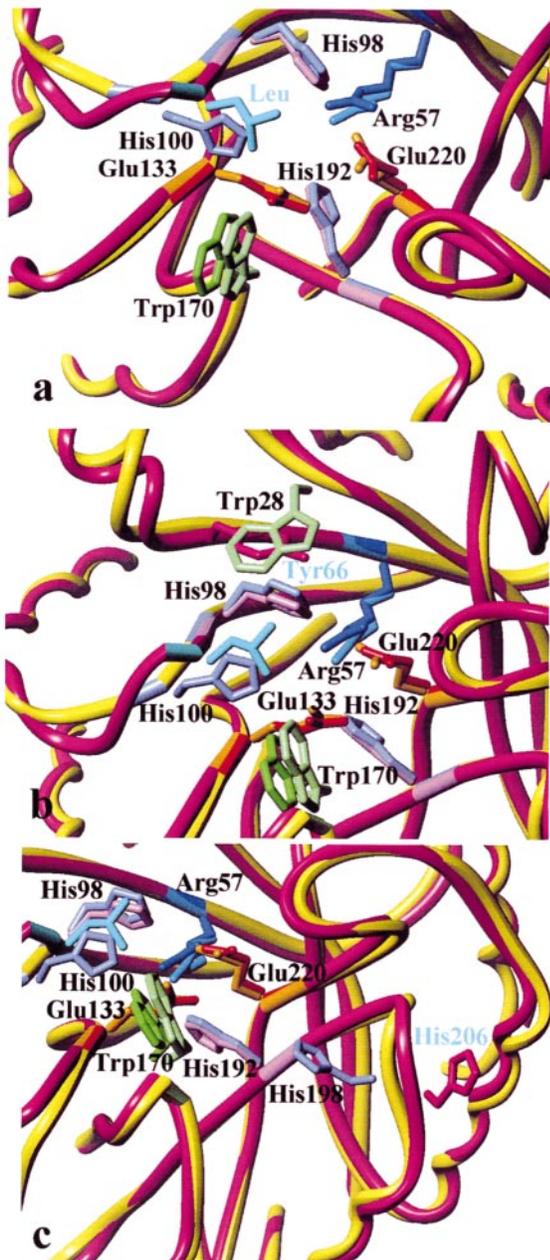
Figure 2. Close-up view of the catalytic site and the highly conserved residues of Cel5 from *E. chrysanthemi* (yellow), and of Cel5A from *B. agharadherans* (magenta). The numbers correspond to the residues of Cel5 from *E. chrysanthemi*. (a) His100 is replaced by a leucine residue in Cel5A from *B. agaradherans*. (b) The NH group of Trp28 in Cel5 superposes well with the hydroxyl group of Tyr66, although they are donated by different stretches of the β-barrel. (c) His198 is turned towards the active site in Cel5 from *E. chrysanthemi*, whereas His206 points into the solvent in the structural model of Cel5A of *B. agaradherans*. The Figures were produced with TURBO-FRODO.[45]

Figure 3. (a) Ribbon representation of the $(\beta/\alpha)_8$ barrel highlighting the positions of all ten arginine residues (balls-and-sticks representation) of Cel5. (b) A close-up view of Arg57 in the heart of the catalytic domain, displaying the hydrogen bonding network in which Arg57 is involved. Connected residues are His98, Glu129, Asn132 and Glu220, which are all related directly to catalysis, and Ser25 and Thr219, which participate in structural stabilization. The Figures were produced using the programs MOLSCRIPT[51] and Raster3D.[52]

the two molecules were analyzed using the program GRASP[24] in order to detect patches or regions having similarity with respect to hydrophobicity, surface charge or overall form. However, no particular resemblance with respect to either of these criteria could be detected.

Relationships between structure and secretion: study of Arg57

The structural analysis described above identified Arg57 as a key residue in the structure of Cel5 catalytic domain (Figure 3(b)). Hence, Arg57 appeared as an excellent residue to probe relationships between the structure of the catalytic domain and the secretability of Cel5. Therefore, variants were constructed that contained Lys, His or Gln residues at position 57, yielding R57K, R57H and R57Q proteins, respectively. Secretability of these variants was analyzed by using an *Escherichia coli* strain that was made secretion-proficient by intro-

ducing the *E. chrysanthemi out* genes.[25] Immunoblot analysis showed that the R57K protein was secretable (Figure 4). In contrast, no cross-reacting material was to be seen in the supernatant culture of cells synthesizing R57H or R57Q variants. While the R57H protein appeared stable, degradation products were visible when analyzing the R57Q protein (Figure 4). Hence, a possibility was that changing Arg to His altered specifically the secretion information, thereby hampering recognition of the protein by the Out machinery, while substitution by Gln yielded an unstable protein that was degraded prior to secretion. To investigate this issue, we carried out a kinetic analysis using the protocol as described.[13] Again, R57K protein was found to exhibit the same kinetic pattern as the wild-type (Figure 5). Results were consistent with previous similar analysis that showed that extracellular forms of Cel5 appeared in less than five minutes in the supernatant.[10,13] In contrast, no extracellular forms of R57H protein were to be seen in the supernatant fraction but for very late time-points (i.e. 60 minutes). Of particular interest was the fact that R57H species remained stable in the cells. This confirmed that R57H was secreted, if at all, with a drastically reduced efficiency. In contrast, the R57Q protein proved to be unstable, since at early time-points two species could be seen in roughly similar amount, namely a species migrating with the full-length molecular mass and a faster-migrating species that is likely to be a degradation product (Figure 5). Moreover, we could see concomitant decrease in intensity of both forms throughout the experiment, since at the latest time-point, neither one was visible. This confirmed that changing Arg to Gln had a profound effect on the stability of the protein and was consistent with the view that this protein had lost secretability as a consequence of its reduced half-life and/or structural modification. Last, it should be added that cell fractionation experiments were carried out to check that the R57H and R57Q proteins were localized in the periplasm of *E. coli* strain expressing the *E. chrysanthemi out* genes (data not shown).

## Stability of purified variants

In order to get further insight into the consequences of the mutations, the wild-type and the three mutated species of Cel5 were purified from *E. coli* cell lysate. Purified proteins were then subjected to proteinase K treatment during various times (Figure 6). Wild-type and R57K proteins exhibited identical proteolysis patterns. A fast-migrating species appeared in both cases after incubation with proteinase K for 1.5 minutes. Scanning the electrophoresis gel allowed us to show that, by starting with samples at 7 μg ml^{-1} concentration, 50 % of the full-length form was degraded in approximately 2.5 minutes. R57H was found to be slightly less resistant to proteinase K than the wild-type (Figure 6). By starting with samples at 20 μg ml^{-1} concentration, 50 % of full-length wild-type and R57H species were degraded in nine and eight minutes, respectively, while after 20 minutes of incubation, 28 % of full-length wild-type Cel5 and 12 % of full-length R57H remained visible. Surprisingly, the R57Q protein proved to be extremely resistant to proteolysis since, after 20 minutes incubation, only 30 % of full-length R57Q (20 μg ml^{-1}) had undergone degradation.

Discussion

This work was initiated in order to carry out a structure/secretion relationship study of the *E. chrysanthemi* cellulase Cel5 (formerly EGZ), a model for the numerous proteins secreted by the type II system. Earlier studies indicated that disulfide bridge formation in the C-terminal CBD domain of Cel5 had a crucial role for secretion.[10,12]

wt				R57K				R57Q				R57H			
−		+		−		+		−		+		−		+	
c	s/n	c	s/n	c	s/n	c	s/n	c	s/n	c	s/n	c	s/n	c	s/n
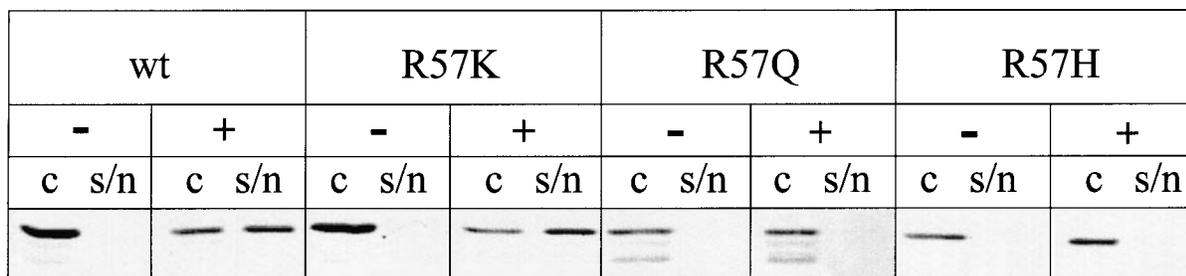
Figure 4. Differential effect of Arg57 mutations for Cel5 secretion. Secretion from *E. coli* carrying (+) or not (−) the *E. chrysanthemi out* genes (carried on pCPP2006 cosmid) of wild-type Cel5 and of the R57 K, R57Q and R57H mutants was analysed. Competent cells of *E. coli* DH5α (containing (+) or not (−) the cosmid pCPP2006) were prepared and transformed with pSZ (which contains the *E. chrysanthemi cel5* gene) or with pSZ-encoded cellulase Arg57 variants. The cells were grown in LB medium at 32 °C for four to five hours ($A_{600} \sim 1.5$), starting with an inoculum of $A_{600} \sim 0.05$. Cultures (0.2 ml) were centrifuged for five minutes at 13,000 rpm and the supernatant (0.05 ml) was collected and added with loading buffer (0.05 ml). Cells were washed, centrifuged again and resuspended in loading buffer (0.4 ml). Detection of Cel5 was determined using homogeneous 10 % (w/v) polyacrylamide SDS-PAGE with a Bio-Rad mini ProteanII system followed by immunoblotting.[10] C, cells; s/n, supernatant.
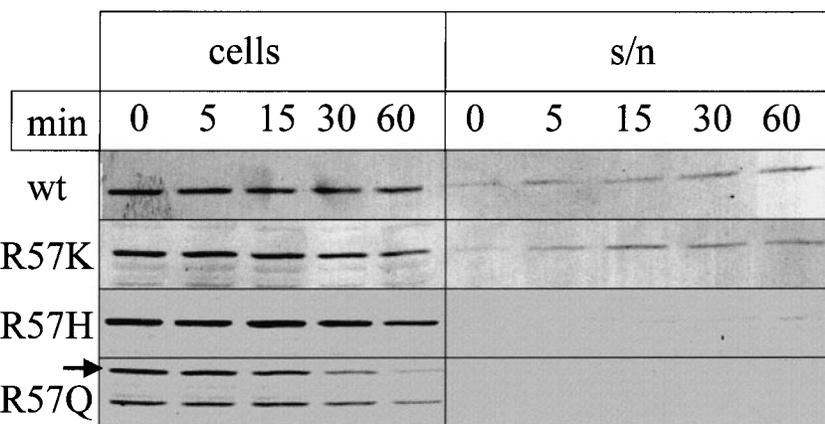
Figure 5. Kinetic analysis of Cel5 and Arg57 mutants secretion. Cultures of *E. coli* DH5α carrying the pCPP2006 cosmid and the pSZ-encoded Cel5 R57K, R57Q or R57H were grown in LB at 32 °C until $A_{600} \sim 1$. A 3 ml sample was centrifuged, the supernatant discarded and the pellet resuspended in 1 ml of LB + chloramphenicol (40 μg ml^{-1}), centrifuged at 32 °C for five minutes. The pellet was resuspended in 0.5 ml of LB + chloramphenicol (40 μg ml^{-1}) and incubated at 32 °C. Aliquots were taken at desired time intervals, centrifuged at 4 °C, yielding cell (C) and supernatant (s/n) fractions that were analysed by immunoblotting as described above. The arrow indicates the full-length form of R57Q.

Concerning the N-terminal catalytic domain, the picture remained unclear as long as the mutational results could not be coupled to the three-dimensional structure analysis of the domain. The present analysis establishes the existence of extremely tight links between structure of the catalytic domain and secretability of Cel5. In particular, we identified a position wherein the nature of the lateral chain dictates the fate of the protein. Since this position is located at the heart of the domain, we do not believe that it participates in the direct recognition of Cel5 by the secretion machinery but, instead, controls secretability *via* its control of the conformation of the domain.

Arginine residues are potentially important as stabilizing elements in protein structure.[26,27] Systematic mutagenesis study of all ten arginine positions of Cel5 catalytic domain was done previously with the aim of identifying structurally important positions.[22] Out of the ten arginine residues, six (namely Arg81, Arg106, Arg111, Arg118, Arg179 and Arg245) were predicted to be exposed at the surface and to be important neither for the structure nor for activity. This is fully confirmed upon examination of the 3D structure (Figure 3(a)). Two other arginine residues, Arg203 and Arg207, were predicted to have a structural role, possibly related to their location in an α-helix, as predicted both by secondary structure algorithms and by the destabilizing effect of introducing Pro residues at these positions. This prediction is also fully confirmed upon examination of the 3D structure, wherein one can locate both Arg203 and Arg207 in the seventh α-helix of the $(\beta/\alpha)_8$-barrel (Figure 3(b)). Arg155 was proposed to be essential to protein stability with no direct role in catalysis. Interestingly, the position equivalent to Arg155 in Cel5 of *E. chrysanthemi* (Arg163 of Cel5A from
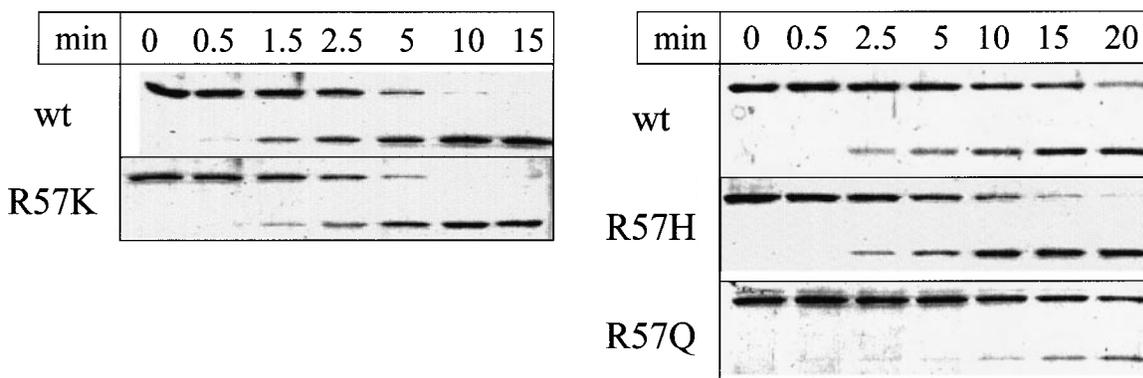
Figure 6. Proteinase K sensitivity of Cel5 and the Arg57 mutants. Purified Cel5, R57Q and R57H mutants were diluted to a final concentration of 20 μg ml^{-1} and incubated at 32 °C with proteinase K at a final concentration of 1 μg ml^{-1}. Aliquots (10 μl) were collected at desired time intervals after addition of proteinase K. Samples were then analyzed on 10 % (w/v) polyacrylamide SDS-PAGE by Coomassie blue staining and subsequent quantification of each band. For each sample, the quantity of full-length variants was expressed as the percentage of total protein in the sample, yielding a kinetic of the full-length form degradation. For the R57K mutant analysis, the experiment was done exactly as described before except that the final concentration of protein used in the assay was 7 μg ml^{-1}.

B. agaradherans), is implied in a strong salt-bridge with an aspartic acid residue in Cel5A.[17] In the structure of *E. chrysanthemi* Cel5, the corresponding aspartic acid residue is not present and Arg155 mainly makes hydrogen bonds to main-chain carbonyl groups. However, in the structure of Cel5 it is this arginine residue that stabilizes the turn containing the sequence DPDN (158-161), which in turn binds a calcium/magnesium ion. One can easily imagine that changing Arg155 to another residue prevents calcium binding and, as a consequence, the catalytic domain is destabilized. This would account for the effect previously observed when substituting Arg155 by a series of different residue.[22]

Arg57 was proposed to have a structural role, since mutations yielded thermosensitive proteins.[22] This is fully consistent with the structural results, as shown in Figure 3(b), since a tight hydrogen bonding network links Arg57 to six residues in the center of the catalytic active site. Changing Arg57 to Lys yielded a protein with wild-type stability both *in vivo* and *in vitro*. This was consistent with the physicochemical similarities of Arg and Lys amino acids. Changing Arg57 to Gln yielded a protein that exhibited seemingly contradictory properties, since it was unstable *in vivo* but resisted proteinase K treatment *in vitro* much more efficiently than the wild-type did. Although we do not know the basis for these properties, both changes in protease sensitivity patterns, as compared with that of the wild-type, indicated that the mutation altered the protein conformation quite drastically. Changing Arg57 to His yielded a protein that was slightly more sensitive to proteinase K as compared with the wild-type but was stable *in vivo*. It seems reasonable to assume that this mutant adopts a conformation close but different from that of the wild-type.

Analysis of R57K protein indicated that this mutant species is non-native, since it is not active, but retains wild-type conformation and is secretable at the same rate as compared with the wild-type. In line with this, R57Q, a species that possesses a conformation different from the wild-type was not secreted. Despite being stable *in vivo*, R57H was not secreted. Since *in vitro* analysis suggested that R57H adopts a conformation slightly different from the wild-type, we suggest that this subtle change might be sufficient to block secretion. An alternate hypothesis is that this position is part of the secretion motif. Such an hypothesis is not supported by (i) the location of Arg57 that is buried inside the protein, (ii) the fact that changing Arg to either Lys or Ser (not shown) does not alter secretion, and (iii) its strict conservation throughout all family 5 cellulases, most of which are not secreted by type II-like systems.

After it was recognized that exoproteins were synthesized first as precursors with typical signal sequences, a two-step model for secretion was proposed (see Introduction). Since primary sequence comparisons failed to identify a putative secretion motif, it was suggested that such a secretion motif would be made of residue located in distinct parts of the primary sequences, that eventually come into close proximity upon folding. That proteins fold prior to secretion received experimental support in different systems: (i) cellulase Cel5 was found to exhibit a disulfide-bridged CBD before crossing the outer membrane,[10] (ii) cholera toxin and aerolysin were found to oligomerise prior to being secreted,[28,29] (iii) a pullulanase form could accumulate in the periplasm for a long period of time, in the absence of the secretion machinery, yet retained its secretion competence.[30] These observations helped to establish the currently accepted view that proteins fold into their so-called "native" state in the periplasm before interacting with the secretion machinery. However, this view was not consistent with various observations. Firstly, chimeric proteins containing β-lactamase and rather small regions of either pullulanase or exotoxin A were found to be secreted.[31,32] Clearly, these chimeric proteins cannot be expected to be in the native conformation supposedly recognized by the machineries. To account for secretability, one can however argue that secretion motifs were retained in those chimeric proteins. Secondly, a pullulanase form, thought to be a periplasmic intermediate, exhibited a susceptibility to protease that appeared to differ from that of the cell-surface exposed (secreted) form.[33] Thirdly, modifications that modify conformation in a very subtle way, if at all, such as N-terminal acylation affected secretion efficiency of pullulanase.[34] Fourthly, elastase is recognized as a proform including an N-terminal extension that is eventually cleaved off and is not part of the so-called "native state".[35] Hence, it appears quite unsafe to use the term native in order to describe the folding status of exoproteins when they interact with the secretion machinery.

In the present study, the identification of Arg57 as a residue having a key role in Cel5 structure allowed us to investigate the relationships between conformation and secretion in some details. The fact that modification of a key conformational controlling position affects secretion makes a very strong case for the idea that, in Cel5, secretability is indeed a specific feature of the wild-type fold. In particular, the fact that R57H protein, which exhibits distinct but close conformation from the wild-type, is not secreted, best illustrates the strong constraints that secretion exerts upon folding. One must however note that non-active species, e.g. R57K and all proteins mutated at the catalytic positions (data not shown), were *bona fide* secretable species, illustrating that secretability is not a feature of the native species but rather of a conformation that can be taken by the wild-type protein. Hence, it seems extremely difficult to extract some general rules that could account for all data collected from studies of different exoproteins. Again, if one accepts that chimeric proteins made of pullulanase and β-lactamase or exotoxin A and β-lactamase retain correctly exposed secretion

information, it is difficult to understand how, in contrast, secretion information is so much disturbed by the Arg-to-His substitution in R57H that it is no longer accessible to the secretion machinery. This leaves open the possibility that each exoprotein has reconciled both folding and secretion motif exposition on its own way. A prediction of this would be that not all exoproteins will share a similar type II secretion motif. In a sense, this could eventually explain the so-called species specificity wherein highly related proteins are not secreted in heterologous systems,[36–38] and might also explain why comparison of 3D structures of exoproteins recognized by the very same machinery, such as *E. chrysanthemi* pectate lyases and cellulase, failed to identify putative secretion regions. Interestingly, very recently, a similar hypothesis was proposed by Shevchik and collaborators after they analyzed secretion efficiencies of numerous chimeric machineries made of Out proteins from *E. chrysanthemi* and *E. carotovora*.[39]

In conclusion, this study indicated that the catalytic domain ought to fold into a wild-type conformation before Cel5 is secreted. We identified a folded non-secretable variant, the structure of which should help in pinpointing those conformational modifications that are responsible for non-recognition by the secretion machinery. Structural analysis of R57H is underway.

Materials and Methods

Strains, plasmids and microbial techniques

Growth media and storage conditions were as described.[13] The plasmid carrying *E. chrysanthemi out* genes was pCPP2006 and was obtained from A. Collmer.[25]

Construction of mutants

Site-directed mutagenesis of plasmid pSZ to generate point mutations in Cel5 was performed with a recombinant PCR method.[40] The R57K and R57Q mutants were produced by PCR amplification of the entire pSZ plasmid by mutagenic primers (mutated nucleotides in boldface) divergently orientated but overlapping (underlined) at their 5′ ends (mutagenic oligonucleotides, 5′-TTCCTgAACgCCCATAgCggC**TTT**AACAATgCTggA 3′ (R57 K), 5′-TTCCTgAACgCgCCCATAgCggC**TT**gAA-CAATgCTggATTTCCA-3′ (R57Q); anchor oligonucleotide, 5′-ATgggCgTTCAggAAAgCggTggTTATCTg CAggACCCg-3′).

PCR was performed with Expand High Fidelity DNA polymerase (Boehringer Mannheim) in the presence of 5% (v/v) dimethyl sulfoxide. Treatment with *Dpn*I destroyed the parental Dam-methylated DNA, whereas PCR-synthetised molecules remained intact. The resulting linear DNA molecules were transformed into the *E. coli* DH5α (a *recA* strain). The presence of the desired mutation and the absence of any secondary ones were verified by sequencing.

The R57H mutant was constructed as described by Bortoli-German *et al.*,[22] with the following oligonucleotide: 5′-CATAgCggCgTgAACAATgCTgg-3′.

Secretion analysis

Secretion was tested by immunoblot analysis of culture supernatant and of whole cells after four to five hours of growth in LB medium at 32 °C as described.[13]

Purification of Cel5 and mutated variants

Competent cells of *E. coli* DH5α were prepared and transformed with the plasmid pSZ (encoded the wild-type Cel5) or with pSZ-encoded cellulase R57K, R57Q and R57H mutants. Cells containing plasmids producing Cel5 wild-type or mutated variants of Cel5 were grown in 5 l of LB medium with 50 μg of ampicillin per ml at 32 °C for five hours ($A_{600} \sim 1.5$). After centrifugation, the cell pellet was washed and suspended in a solution containing TE (20 mM Tris-HCl (pH 7.5), 3 mM EDTA). This suspension was passed twice through a French Press cell and the crude extract was centrifuged at 10,000 *g* for 15 minutes at 13,000 rpm to remove cells debris. The resulting supernatant (80 ml) was made 1 M NaCl and mixed with cotton (6 mg ml^{-1}). The cotton, purchased from a local drugstore, had previously been soaked in TEN (TE containing 1 M NaCl), autoclaved for 30 minutes 110 °C, washed once with TEN and finally resuspended in TEN. The mix was incubated with gentle mixing during 14 hours at 4 °C. Cotton was washed three times with TEN and three times with TE. Elution of cellulase was obtained by washing the cotton under vacuum with 80 ml of cold Milli-Q water. The flow-through was desalted and concentrated by using an Ultrafree-15 centrifugal filter unit with a 5000-NMWL (molecular mass of 5000 cutoff) polysulfone membrane (Millipore). The concentrated sample (1 ml) was loaded onto a FPLC monoQ column (HR 10/10, Pharmacia). The wild-type Cel5 and the variants mutants were eluted with a continuous NaCl gradient from 0 to 0.5 M. Fractions were collected and assayed for the presence of Cel5 or variants by SDS-PAGE followed by immunoblotting. Peak fractions were desalted and concentrated by using an Ultrafree-15 centrifugal filter unit with a 5000-NMWL (Millipore). The protein concentration was estimated by the Bradford method using bovin serum albumin as the standard.

Limited proteolysis assays

Purified Cel5 or R57Q, R57H and R57K mutants were tested for their sensitivity to proteinase K. Solution of purified proteins were diluted to a concentration of 20 μg ml^{-1} (or 7 μg ml^{-1}; see the Legend to Figure 6) and incubated at 32 °C with proteinase K at a final concentration of 1 μg ml^{-1}. Aliquots (10 μl) were collected at 0.5, 2.5, 5, 10, 15 and 20 minutes after addition of protease. The reaction was stopped on ice in the presence of trichloroacetic acid at a final concentration of 5% (w/v). The samples were then analysed by SDS-PAGE and subsequent staining with Coomassie blue. After staining, the gels were scanned and the intensity of each band was quantified yielding a kinetic of Cel5 and R57Q, R57H and R57K mutants degradation.

Crystallization and X-ray structure determination

Crystallization experiments were performed at 4 °C employing the vapour diffusion technique. The crystals

of the catalytic domain were obtained by attempts at crystallizing the full-length protein. Crystals from the catalytic domain of Cel5 protein (7.5 mg ml^{-1} full-length protein concentration), buffered in 10 mM Tris-HCl (pH 8.0) appeared within one year using 28% (w/v) PEG 4000 as precipitant, buffered in 20 mM Tris-HCl (pH 8.2), 20 mM MgSO$_4$. The crystals belong to the space group $C2$ with lattice constants $a = 168.8$ Å, $b = 47.2$ Å, $c = 120.1$ Å, $\beta = 105°$. The asymmetric unit contains three molecules of the catalytic domain of Cel5. The crystals have a solvent content of 41% (v/v) ($V_M = 2.08$ Å^3 Da^{-1}) and diffract to a resolution of 2.2 Å. Data sets at 2.2 Å resolution were collected on a MarResearch imaging plate system installed on a Rigaku rotating anode generator (CuK$_\alpha$) by taking 1° frames over 120° of an arbitrary orientation of the crystal. The intensity integration and data treatment were performed with the DENZO program package.[41,42] Statistics on the data quality are given in Table 1. The structure was solved by molecular replacement using the program AMoRe.[43] The family 5 endoglucanase of *B. agaradherans*[17] (pdb code: 1A3H) was the only available model of family 5 members, that led to a solution for the structure of the catalytic domain of Cel5A (43% sequence identity). Three solutions, corresponding to the three molecules in the asymmetric unit, were found having correlation coefficients of 0.22, 0.20 and 0.18 for data in the range of 8.0 Å to 3.0 Å resolution. The rigid-body refinement (included in the program AMoRe) led to an overall correlation coefficient of 0.43 and an R-factor of 43.9% for the solution formed by the three symmetry-independent molecules (see Table 1). The resultant map was improved with cycles of density modification, utilising histogram matching and solvent flattening, using the program DM.[44] The residues that had to be replaced with respect to the sequence of Cel5A were clearly visible and were built in the DM modified Fourier-map with the program TURBO-FRODO:[45] 5% of the observations were then set aside for cross-validation analysis[46] and were used to monitor various refinement strategies such as geometric and temperature factor restraint values and the insertion of solvent water molecules. The refinement was performed using cycles of simulated annealing, as is implemented in the program XPLOR3.8.[47] As all observed data from 22 Å resolution were employed in the refinement, a low-resolution bulk solvent correction was applied. Cycles of simulated annealing refinement were interspersed with rebuilding using TURBO-FRODO. The water molecules were inspected manually. No non-crystallographic symmetry restraints were applied during the refinement cycles. The current model, refined to 2.3 Å, comprises 3×2270 non-hydrogen protein atoms, 329 solvent atoms and three calcium ions. The final crystallographic R-factor for this model in the range of 29 Å to 2.3 Å is 18.7% ($R_{free} = 22.6$%). The r.m.s.d. from ideal stereo-chemistry are 0.007 Å for bond lengths and 1.65° for bond angles.[48] The stereochemistry of the model was verified using the program PROCHECK.[49] The dihedral angles of the polypeptide backbone all lie in allowed regions of the Ramachandran diagram.

Protein Data Bank accession numbers

Coordinates for the protein structure described here have been deposited with the RCSB Protein Data Bank,[50] pdbcode: 1egz.

References

1. Filloux, A., Lazdunski, A. & Bally, M. (1998). GSP-dependent protein secretion in Gram-negative bacteria: the Xcp system of *Pseudomonas aeruginosa*. *FEMS Microbiol. Rev.* **22**, 177-198.
2. Russel, M. (1998). Macromolecular assembly and secretion across the bacterial cell envelope: type II protein secretion systems. *J. Mol. Biol.* **279**, 485-499.
3. Nunn, D. (1999). Bacterial type II protein export and pilus biogenesis: more than just homologies? *Trends Cell Biol.* **9**, 402-408.
4. Barras, F., van Gijsegem, F. & Chatterjee, A. K. (1994). Extracellular enzymes and pathogenesis of soft-rot *Erwinia*. *Annu. Rev. Phytopathol.* **32**, 201-234.
5. Py, B., Bortoli-German, I., Haiech, J., Chippaux, M. & Barras, F. (1991). Cellulase EGZ of *Erwinia chrysanthemi*: structural organisation and importance of His98 and Glu133 residues for catalysis. *Protein Eng.* **4**, 325-333.
6. Henrissat, B. & Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **293**, 781-788.
7. Henrissat, B. & Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* **316**, 695-696.
8. Tomme, P., Warren, R. A. J. & Gilkes, N. R. (1995). Cellulose hydrolysis by bacteria and fungi. *Adan. Microb. Physiol.* **37**, 1-81.
9. Py, B., Chippaux, M. & Barras, F. (1993). Mutagenesis of cellulase EGZ for studying the general protein secretory pathway in *Erwinia chrysanthemi*. *Mol. Microhiol.* **7**, 785-793.
10. Bortoli-German, I., Brun, E., Py, B., Chippaux, M. & Barras, F. (1994). Periplasmic disulphide bond formation is essential for cellulase secretion by the plant pathogen *Erwinia chrysanthemi*. *Mol. Microbiol.* **11**, 545-553.
11. Brun, E., Gans, P., Marion, D. & Barras, F. (1995). Overproduction, purification and characterization of the cellulose-binding domain of the *Erwinia chrysanthemi* secreted endoglucanase EGZ. *Eur. J. Biochem.* **231**, 142-148.
12. Brun, E., Moriaud, F., Gans, P., Blackledge, M. J., Barras, F. & Marion, D. (1997). Solution structure of the cellulose binding domain of the endoglucanase Z secreted by *Erwinia chrysanthemi*. *Biochemistry*, **36**, 16074-16086.

13. Chapon, V., Simpson, H., Morelli, X., Brun, E. & Barras, F. (2000). Alteration of a single residue of the cellulose binding domain blocks secretion of the *Erwinia chrysanthemi* Cel5 cellulase (ex-EGZ) *via* the type II system. *J. Mol. Biol.* **303**, 117-123.

14. Dominguez, R., Souchon, H., Spinelli, S., Dauter, Z. N., Wilson, K. S., Chauvaux, S., Béguin, P. & Alzari, P. M. (1995). A common fold and similar active site in two distinct families of β-glycanases. *Nature Struct. Biol.* **2**, 569-576.

15. Ducros, V., Czjzek, M., Belaich, A., Gaudin, C., Fierobe, H-P. & Belaich, J. P. *et al.* (1995). Crystal structure of the catalytic domain of a bacterial cellulase belonging to family 5. *Structure,* **3**, 939-949.

16. Sakon, J., Adney, W. S., Himmel, M. E., Thomas, S. R. & Karplus, P. A. (1996). Crystal structure of thermostable family 5 endocellulase E1 from *Acidothermus cellulolyticus* in complex with cellotetraose. *Biochemistry,* **35**, 10648-10660.

17. Davies, G. J., Dauter, M., Brzozowski, A. M., Bjornvad, M. E., Andersen, K. V. & Schulein, M. (1998). Structure of the *Bacillus agaradherans* family 5 endoglucanase at 1.6 Å and its cellobiose complex at 2.0 Å resolution. *Biochemistry,* **37**, 1926-1932.

18. Juy, M., Amit, A. G., Alzari, P. M., Poljak, R. J., Claeyssens, M., Béguin, P. & Aubert, J.-P. (1992). Three-dimensional structure of a thermostable bacterial cellulase. *Nature,* **57**, 89-91.

19. Törrönen, A. & Rouvinen, J. (1995). Structural comparison of two major endo-1,4-xylanases from *Trichoderma reesei. Biochemistry,* **34**, 847-856.

20. Harris, G. W., Jenkins, J. A., Connerton, I. & Pickersgill, R. W. (1996). Refined crystal structure of the catalytic core of xylanase A from *Pseudomonas fluorescens* at 1.8 Å resolution. *Acta Crystallog. sect. D,* **52**, 393-401.

21. Wang, Q., Tull, D., Meinke, A., Gilkes, N. R., Warren, R. A. J., Aeberold, R. & Withers, S. G. (1993). Glu280 is the nucleophile in the active site of *Clostridium thermocellum* CelC, a family A endo-β-1,4-glucanase. *J. Biol. Chem.* **268**, 14096-14102.

22. Bortoli-German, I., Haiech, J., Chippaux, M. & Barras, F. (1995). Informational suppression to investigate structural, functional and evolutionary aspects of the *Erwinia chrysanthemi* cellulase EGZ. *J. Mol. Biol.* **246**, 82-94.

23. Yoder, M. D., Lietzke, S. E. & Jurnak, F. (1993). Unusual structural features in the parallel beta-helix in pectate lyases. *Structure,* **1**, 241-251.

24. Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281-296.

25. He, S. Y., Lindberg, M., Chatterjee, A. K. & Collmer, A. (1991). 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.

26. Mitchell, J. B. O., Thronton, J. M., Singh, J. & Price, S. L. (1992). Towards an understanding of the arginine-aspartate interaction. *J. Mol. Biol.* **226**, 251-262.

27. Flocco, M. M. & Mowbray, S. L. (1994). Planar stacking interactions of arginine and aromatic side-chains in proteins. *J. Mol. Biol.* **235**, 709-717.

28. Hirst, T. R. & Holmgren, J. (1987). Conformation of protein secreted across bacterial outer membranes: a study of enterotoxin translocation of *Vibrio cholerea. Proc. Natl Acad. Sci. USA,* **84**, 7418-7422.

29. Hardie, K. R., Schulze, A., Parker, M. W. & Buckley, J. T. (1995). *Vibrio spp.* secrete proaerolysin as a folded dimer without the need for disulphide bond formation. *Mol. Microbiol.* **17**, 1035-1044.

30. Poquet, I., Faucher, D. & Pugsley, A. P. (1993). Stable periplasmic secretion intermediate in the general secretory pathway of *Escherichia coli. EMBO J.* **12**, 271-278.

31. Lu, H. M. & Lory, S. (1996). A specific targeting domain in mature exotoxin A is required for its extracellular secretion from *Pseudomonas aeruginosa. EMBO J.* **15**, 429-436.

32. Sauvonnet, N. & Pugsley, A. P. (1996). Identification of two regions of *Klebsiella oxytoca* pullulanase that together are capable of promoting beta-lactamase secretion by the general secretory pathway. *Mol. Microbiol.* **22**, 1-7.

33. Pugsley, A. P., Poquet, I. & Kornacker, M. G. (1991). Two distinct steps in pullulanase secretion by *Escherichia coli* K12. *Mol. Microbiol.* **5**, 865-873.

34. Pugsley, A. P. & Kornacker, M. G. (1991). Secretion of the cell surface lipoprotein pullulanase by *Escherichia coli*: collaboration or competition between the specific secretion pathway and the lipoprotein sorting pathway? *J. Biol. Chem.* **266**, 13640-13645.

35. Kessler, E., Safrin, M., Gustin, J. K. & Ohman, D. E. (1998). Elastase and the LasA protease of *Pseudomonas aeruginosa* are secreted with their propeptides. *J. Biol. Chem.* **273**, 30225-30231.

36. de Groot, A., Filloux, A. & Tommassen, J. (1991). Conservation of *xcp* genes, involved in the two-step protein secretion process, in different *Pseudomonas* species and other gram-negative bacteria. *Mol. Gen. Genet.* **229**, 272-284.

37. Py, B., Salmond, G. P. C., Chippaux, M. & Barras, F. (1991). Secretion of cellulases of *Erwinia chrysanthemi* and *E. carotovora* is species-specific. *FEMS Microbiol. Letters,* **79**, 315-322.

38. Michel, L. O., Sandkvist, M. & Bagdasarian, M. (1995). Specificity of the protein secretory apparatus: secretion of the heat-labile enterotoxin B subunit pentamers by different species of gram-bacteria. *Gene,* **152**, 41-45.

39. Bouley, J., Condemine, G. & Schevchik, V. (2001). The PDZ domain of OutC and the N-terminal region of OutD determine the secretion specificity of the type II Out pathway of *Erwinia chrysanthemi. J. Mol. Biol.* **308**, 205-219.

40. Ansaldi, M., Lepelletier, M. & Méjean, V. (1996). Site-specific mutagenesis by using an accurate recombinant polymerase chain reaction method. *Anal. Biochem.* **234**, 110-111.

41. Minor, W. (1993). *XDISPLAYF program*, Purdue University, West Lafayette, Indiana.

42. Otwinowski, Z. (1993). *Proceedings of the CCP4 study Weekend: Data Collection and Processing* (Sawyer, L., Isaacs, N. & Bailey, S., eds), pp. 56-62, SERC Daresbury Laboratory, Warrington, UK.

43. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A,* **50**, 157-163.

44. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. A,* **50**, 760-763.

45. Roussel, A. & Cambillau, C. (1992). TURBO-FRODO, *The Manual*, Biographics LCCMB, Marseille, France.

46. Brünger, A. T. (1992). Free *R*-value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature,* **355**, 472-475.
47. Brünger, A. T. (1996). *X-PLOR, Version 3.8*, Yale University, New Haven, CT.
48. Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallog. sect. A,* **47**, 392-400.
49. Laskowski, K. A., McArthur, M. W., Moss, D. S. & Thronton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.
50. Bernstein, F. C., Koetzle, T. F., Williams, G. I. B., Meye, E. F., Jr, Brice, M. D. & Rodgers, J. R. *et al.* (1977). The Protein Data Bank: a computer-based archival file for macro-molecular structures. *J. Mol. Biol.* **112**, 535-542.
51. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.
52. Meritt, E. A. & Murphy, M. E. P. (1994). Raster 3D version 2.0, a program for photorealistic molecular graphics. *Acta Cryslallog. sect. D,* **50**, 869-873.