

1998

# Sequence Profile of the Parallel $\beta$ Helix in the Pectate Lyase Superfamily

Susan Heffron

*University of California - Irvine*

Gregory R. Moe

*Children's Hospital Oakland Research Institute*

Volker Sieber

*University of Bayreuth*

Jerome Mengaud


*Pasteur Institute*

Pascale Cossart

*Pasteur Institute*

*See next page for additional authors*

Follow this and additional works at: [https://engagedscholarship.csuohio.edu/sciphysics\\_facpub](https://engagedscholarship.csuohio.edu/sciphysics_facpub)

 Part of the [Biochemistry, Biophysics, and Structural Biology Commons](#)

**How does access to this work benefit you? Let us know!**

## Repository Citation

Heffron, Susan; Moe, Gregory R.; Sieber, Volker; Mengaud, Jerome; Cossart, Pascale; Vitali, Jacqueline; and Jumak, Frances, "Sequence Profile of the Parallel  $\beta$  Helix in the Pectate Lyase Superfamily" (1998). *Physics Faculty Publications*. 152.  
[https://engagedscholarship.csuohio.edu/sciphysics\\_facpub/152](https://engagedscholarship.csuohio.edu/sciphysics_facpub/152)

---

**Authors**

Susan Heffron, Gregory R. Moe, Volker Sieber, Jerome Mengaud, Pascale Cossart, Jacqueline Vitali, and Frances Jumak

# SEQUENCE PROFILE OF THE PARALLEL $\beta$ HELIX IN THE PECTATE LYASE SUPERFAMILY

Susan Heffron, *University of California*

Gregory R. Moe, *Children's Hospital Oakland Research Institute*

Volker Sieber, *University of Bayreuth*

Jérôme Mengaud and Pascale Cossart, *Pasteur Institute*

Jacqueline Vitali, *University of Texas*

Frances Journak, *University of California*

## ABSTRACT

The parallel  $\beta$  helix structure found in the pectate lyase superfamily has been analyzed in detail. A comparative analysis of known structures has revealed a unique sequence profile, with a strong positional preference for specific amino acids oriented toward the interior of the parallel  $\beta$  helix. Using the unique sequence profile, search patterns have been constructed and applied to the sequence databases to identify a subset of proteins that are likely to fold into the parallel  $\beta$  helix. Of the 19 families identified, 39% are known to be carbohydrate-binding proteins, and 50% belong to a broad category of proteins with sequences containing leucine-rich repeats (LRRs). The most striking result is the sequence match between the search pattern and four contiguous segments of internalin A, a surface protein from the bacterial pathogen *Listeria monocytogenes*. A plausible model of the repetitive LRR sequences of internalin A has been constructed and favorable 3D-1D profile scores have been calcu-

lated. Moreover, spectroscopic features characteristic of the parallel  $\beta$  helix topology in the pectate lyases are present in the circular dichroic spectrum of internalin A. Altogether, the data support the hypothesis that sequence search patterns can be used to identify proteins, including a subset of LRR proteins, that are likely to fold into the parallel  $\beta$  helix.

**Key Words:** parallel  $\beta$  helix; internalin; LRR proteins; pectate lyases.

---

## INTRODUCTION

In the first 30 years of protein crystallography, only four types of domain structure had been observed, including all helical domains, alternating  $\alpha$ /parallel  $\beta$  domains, antiparallel  $\beta$  domains, and small domains stabilized by disulfide bonds or metal ion coordination (Levitt and Chothia, 1976; Richardson, 1981). In 1993, a new type of domain, one that had not been predicted *a priori*, was reported for *Erwinia chrysanthemi* pectate lyase C (PelC) (Yoder *et al.*, 1993a). The same structural fold has now been observed in other members of the pectate lyase

<sup>1</sup>To whom correspondence and reprint requests should be addressed. Fax: (949) 824-8540. E-mail: journak@uci.edu.

superfamily, including *E. chrysanthemi* pectate lyase E (PelE) (Lietzke *et al.*, 1994), *Bacillus subtilis* pectate lyase (*B.s.* Pel) (Pickersgill *et al.*, 1994), *Aspergillus niger* pectin lyase A (PLA) (Mayans *et al.*, 1997), and *A. niger* pectin lyase B (PLB) (Vitali *et al.*, 1998). Each structure consists of a single domain of parallel  $\beta$  strands folded into a large right-handed cylinder. The domain fold, termed the parallel  $\beta$  helix, is compatible with all accepted structural rules, albeit in a unique manner. The central cylinder consists of seven to nine complete helical turns, each of which has a rise of 4.86 Å and a minimum of 22 amino acids per turn. The cross-section of the parallel  $\beta$  helix is not circular, but rather L-shaped due to the unique arrangement of three parallel  $\beta$  strands in each turn of the helix. The parallel  $\beta$  strands are connected by three different types of  $\beta$  bends. One of the  $\beta$  bends represents a new type of repetitive structural unit that is stabilized by an unusual stacking arrangement of asparagines, termed the asparagine ladder. Loops of various size and conformation protrude from the remaining two types of  $\beta$  bends and appear to confer functional properties upon the enzymes. As a consequence of the loop variability, the amino acid sequence does not contain any repeating segments that would suggest a periodic structure *a priori*. The structure is stabilized by an extensive hydrogen bond network between parallel  $\beta$  strands as well as by highly ordered stacking of side chains in the interior of the cylinder. The observed stability of the pectate lyases in solution and their general resistance to proteases are macroscopic properties that are consistent with the domain structure at the atomic level.

In addition to the pectate lyase family members, the parallel  $\beta$  helix topology and various deviations have subsequently been observed in a number of proteins (Yoder and Jurnak, 1995a). Right-handed parallel  $\beta$  helices with a greater number of coils have been observed in the *Salmonella typhimurium* P22 tailspike protein (TSP), which has 13 coils (Steinbacher *et al.*, 1994), and *Bordetella pertussis* P.69 pertactin, which has 16 coils (Emsley *et al.*, 1996). Right-handed coils, in which each turn contains two or four  $\beta$  strands, rather than the typical three, have also been reported. A coil containing two  $\beta$  strands and termed a  $\beta$  roll is present as a subdomain in *Pseudomonas aeruginosa* alkaline protease (Baumann *et al.*, 1993). A coil containing four  $\beta$  strands is found in *Aspergillus aculeatus* rhamnogalacturonase A (RGase A) (Petersen *et al.*, 1997). Parallel  $\beta$  helices, which fold into a left-handed cylinder, have been observed in *Escherichia coli* UDP-*N*-acetylglucosamine acyltransferase (Raetz and Roderick, 1995) and *Methanosarcina thermophila* carbonic anhydrase (Kisker *et al.*, 1996). The left-handed parallel  $\beta$

helices differ significantly from the right-handed ones in that the cross-section resembles an equilateral triangle and each turn can accommodate a minimum of 17 amino acids. With the exception of *B. pertussis* P.69 pertactin and *M. thermophila* carbonic anhydrase, the remaining proteins with parallel  $\beta$  helix domains bind to oligosaccharides; thus the domain structure suggests, but does not necessarily establish a common function or substrate type.

A multiple sequence alignment of the pectate lyase superfamily suggests that all members are likely to fold into right-handed parallel  $\beta$  helices (Henrissat *et al.*, 1995). With the appearance of the topology in other diverse protein classes, it becomes of interest to explore whether there are characteristic features that can serve as predictors of the new fold. One such feature is found in a unique CD spectroscopic profile in which the magnitudes of the absorption bands are much greater than that of a typical  $\beta$  sheet (Sieber *et al.*, 1995). Detecting characteristic features from an amino acid sequence is more difficult, in large part because prediction methods are less accurate for  $\beta$  structure than for  $\alpha$  helices. The task is further complicated for parallel  $\beta$  helices because the number of amino acids in each coil is not a fixed length. The loops, which protrude at the turn regions of each rung of the parallel  $\beta$  helix, vary in size, from 4 to as many as 50 amino acids. Not only does the variability in length obscure repetitive structural features that might otherwise be identified in the sequence, but the allowance of gaps in some regions, but not others, is a difficult task for most search programs. To better understand the defining features of the right-handed parallel  $\beta$  helix, a profile of the minimal coil size was extracted from the pectate lyase structures. The coil profile revealed unique features that were used to develop a repetitive sequence pattern for database searches of potential right-handed parallel  $\beta$  helix structures. Independent confirmation of the parallel  $\beta$  helix structure for one protein identified by the search was sought using modeling and CD spectroscopic analysis. The analyses and results are described herein.

## METHODS

*Construction of idealized helical coil and idealized parallel  $\beta$  helix.* The atomic coordinates of the three-dimensional structures of PelC and PelE, both refined to a resolution of 2.2 Å (Yoder and Jurnak, 1995b; Lietzke *et al.*, 1996), were used. A visual examination of the two structures revealed that one helical coil in each protein approximates the minimal coil. The coil contains 23 amino acids with only a small  $\beta$  bulge, but no protruding loops. With the exception of the  $\beta$  bulge at Gly173, the PelC sequence from Val169 through Val191 exhibits repetitive structural elements shared by segments from three or more helical coils in each of the parallel  $\beta$  helices PelC or PelE. For example, the segments from Ile105 through Ile120 and Ile133 through Leu150 in PelC have identical backbone geometries with Val169 through Ile185, excepting a small region around Gly183. To excise the glycine

bulge, Ile107 through Val114 were superimposed upon Ile171 through Thr179, and a least-squares minimization of the distances between all  $\alpha$  carbons,  $\beta$  carbons, carbonyl oxygens, and amide nitrogens was carried out using the program O (Jones and Kjeldgaard, 1993). The atomic coordinates for Ile171 through Thr179 were deleted and those for the superimposed segment, Ile107–Val114, were kept. Thus, the minimal idealized helical coil consists of Val169–Asp170, superimposed Ile107–Val114, and Val180–Val191. The idealized coil was renumbered from 1 through 22.

To create an idealized parallel  $\beta$  helix with multiple coils, the asparagine at position 14 of the idealized coil was superimposed upon each asparagine of the asparagine ladder in PelC. Least-squares minimization of the distances between all  $\alpha$  carbons,  $\beta$  carbons, carbonyl oxygens, and amide nitrogens, which shared the same backbone geometry with the idealized helical coil, were minimized using the program O. The peptide bond at the gap between one idealized helical coil and the next was closed and refined to ideal geometry.

**Construction of amino acid search patterns.** A superposition of the three-dimensional structures of PelC, PelE, and PLB was used to correct the multiple sequence alignment of the pectate lyase superfamily. The amino acids that participated in repetitive secondary structural elements of the parallel  $\beta$  helix topology were grouped according to their position in the idealized coil. By structural analogy, *B.s.* Pel and PLA were included in the grouping. All search patterns were based upon those amino acids with side-chains oriented toward the interior of the parallel  $\beta$  helix. No restrictions were placed upon the amino acid type in those coil positions with an outward orientation of the side chain. Given the limited sampling size, the amino acid preference was weighted equally in the search pattern, rather than statistically according to the frequency of occurrence. The FINDPATTERNS program of the GCG package (Genetics Computer Group, Copyright 1982–1997) was used to search the Swiss-Prot database, Release 33.0, and the Protein Information Resource (PIR) database, Release 53.0. Multiple search patterns were developed and tested. Criteria for selecting the sequence patterns reported herein included sequence matches with multiple members of the pectate lyase superfamily and a relatively short list of protein families for further characterization.

**Model construction, refinement, and assessment.** The amino acid sequence from five repetitive regions of internalin A, from Ala170 through Ile279, was substituted onto an idealized parallel  $\beta$  helix of five turns. The most favorable rotamer conformations were selected for the nonconserved amino acids. The model was visually inspected and adjusted in the program O to minimize unfavorable nonbonded intramolecular contacts. The potential energy of the model was minimized using the conjugate gradient algorithm in the computer program X-PLOR (Brünger, 1988; Brünger, 1993). The validity of the model was assessed by the program PROCHECK (Laskowski *et al.*, 1993) and the 3D-1D profile method of Lüthy and colleagues (Lüthy *et al.*, 1992).

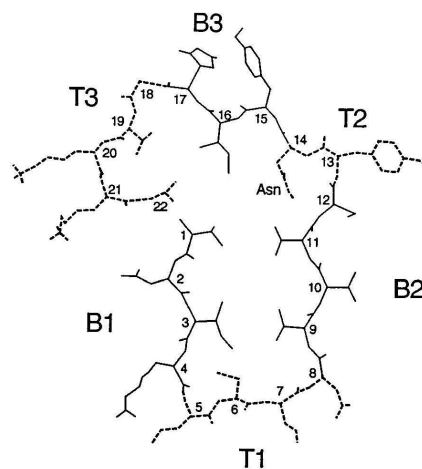
**Circular dichroism.** Internalin A was purified according to the procedure of Mengaud *et al.* (1996) from the culture supernatant of BUG 531, a recombinant *Listeria innocua* strain (CLIP 11262) harboring pGM4. The protein concentration was determined in triplicate from the absorbance at 280 nm in 6 M guanidine-HCl using an extinction coefficient of 80,650 M<sup>-1</sup> cm<sup>-1</sup> (Edelhoch, 1967). The CD spectrum of internalin A (2.63  $\mu$ M) was recorded with the protein suspended in 50 mM Tris-HCl, pH 7.5, buffer containing 120 mM NaCl on a Jasco J-710 CD spectrometer. The spectrum is the average of five scans recorded at a rate of 20 nm/min using a 0.1-cm pathlength quartz cell (Helma Cells, Jamaica, NY) after subtracting the spectrum of the buffer alone recorded using the same conditions. The spectrometer was calibrated with (+)-10-camphorsulfonic acid (Johnson, 1990). The internalin A CD spectrum was deconvoluted using a linear

combination of basis spectra for  $\alpha$  helix, antiparallel  $\beta$  sheet, random coil, and the average parallel  $\beta$  helix spectrum derived from pectate lyases C and E (Sieber *et al.*, 1995).

## RESULTS

**Properties of the idealized coil.** The idealized helical turn, extracted from the PelC structure, is shown in Fig. 1. The idealized rung is composed of 22 amino acids organized into three short parallel  $\beta$  strands, designated B1, B2, and B3. The B1 and B2 strands each contain four amino acids and are connected by a narrow bend, T1, which has a minimum length of four amino acids. A very short  $\beta$  bend, T2, with two amino acids, connects B2 to the B3 strand, with three amino acids. The B3 and B1 strands are connected by a wide bend, T3, which has a minimum length of five amino acids. In all pectate or pectin lyase structures, loops frequently protrude from the parallel  $\beta$  helix at T1 and T3, but never at T2. The T2 bend is a new type of repetitive secondary structure (Yoder *et al.*, 1993b), stabilized by interactions of the asparagine at position 14 of the idealized coil. Although the second amino acid in the T2 bend has  $\phi$  and  $\psi$  angles comparable to an extended  $\beta$  conformation, it is not included in the B3 strand because the intrastrand hydrogen bond pattern does not conform to a  $\beta$  conformation.

There are two distinctive features of the idealized helical coil. The first is the L-shaped, rather than circular, appearance of the cross-section, which is a consequence of the unusual arrangement of the parallel  $\beta$  strands. B1 and B2 are paired in a  $\beta$  sandwich formation, separated by a distance of 9.1 Å between  $\alpha$ Cs. Although the side chains of B1 and B2



**FIG. 1.** Cross-sectional view of a single helical coil of the idealized parallel  $\beta$  helix. The  $\beta$  strands, B1, B2, and B3, are represented by solid lines and the  $\beta$  turns, T1, T2, and T3, are indicated by dotted lines. The numbering system is indicated at the  $\alpha$ C of each position and the asparagine at coil position 14 is labeled.

appear to abut each other across the interface of the  $\beta$  sandwich, the side chains avoid steric clashes by the 1.3 to 2.2 Å rise resulting from the helical nature of the motif. The second notable feature is the unique profile of the orientations of the amino acid side chains toward the interior or toward the exterior of the parallel  $\beta$  helix. In the  $\beta$  strand regions, the orientation of the side chains alternate between inward and outward. At the initiation and termination of the bends, adjacent side chains share the same orientation. Starting with position 1, the idealized helical coil has the following pattern: ioioioioioioioioioioioi where "i" represents an inward orientation and "o," an outward orientation of the side chain. The side-chain orientation pattern is quite distinct from either  $\alpha$  or standard  $\beta$  geometry. In the latter motifs, the side-chain orientation pattern gives rise to surfaces that are either hydrophobic or hydrophilic in character. In contrast, the interior and exterior surfaces of the parallel  $\beta$  helix of the pectate lyase structures are not readily categorized. On the interior of the parallel  $\beta$  helices, approximately 80% of the amino acids are hydrophobic with most of the remaining groups, polar. On the exterior surface, all types of amino acids are found. The polar or charged groups in the exterior orientation are generally exposed to solvent and the hydrophobic groups are usually covered by surface loops, which protrude from the central cylinder at the T1 and T3 bends.

A summary of the amino acids oriented toward the interior of 30 coils in the parallel  $\beta$  helices of the five known pectate and pectin lyase structures is listed in Table IA. A similar compilation has been extracted from the structure-based multiple sequence alignment of 31 members of the pectate lyase superfamily (Henrissat *et al.*, 1995) and the results are shown in Table IB. Both tables reveal a strong preference of amino acids for specific positions of the coil. Hydrophobic amino acids dominate in positions 1, 3, 9, 11, and 19. Asparagine is strongly favored in position 14, aromatic groups dominate in position 16, and small amino acids are favored in position 22. A visual inspection of the parallel  $\beta$  helix structures provides a partial explanation. Asparagine in position 14 stabilizes the unusual T2 turn, by forming hydrogen bonds to the backbone of the preceding residue and to other asparagine side chains above and below it. Positions 1, 3, 9, and 11 represent the inward positions of the pseudo- $\beta$  sandwich formed by B1 and B2; thus the amino acids are constrained by close packing interactions with neighboring groups. In contrast, there is a greater void to fill at position 16, the inward position of the B3 strand; thus aromatic residues appear to be favored. Using a similar argu-

ment, position 22 appears to favor smaller groups because there is a smaller internal volume around the sharp bend. The pattern is less clear for positions 6 and 19, which are the inward positions of the T1 and T3 turns. T1 and T3 turns vary in length and many do not have an amino acid in an orientation comparable to positions 6 or 19. Thus, Table I includes only amino acids observed for a subset of T1 and T3 turns which form a distinctive, nonrepetitive type of secondary structure. Two types of regular T1 turns are observed, one with a  $\beta$  bulge involving a glycine, and one without, utilizing a serine side-chain to stabilize a sharp bend via hydrogen bonding. Thus, glycine and serine/threonine are favored in position 6. Only one type of repetitive T3 turn is observed, having a minimum length of five amino acids and a preference for hydrophobic amino acids in position 19.

*Amino acid sequence search patterns.* The primary objective was to develop a search pattern that would identify protein sequences likely to fold into a parallel  $\beta$  helix. Multiple search patterns were tested in order to generate a short list of potential candidates that could be further characterized by spectroscopic or structural studies to confirm the presence of the parallel  $\beta$  helix fold. Because all patterns were developed from sequences found in members of the pectate lyase superfamily, a successful pattern was expected to generate sequence matches that included a high proportion of the latter proteins. Many of the initial search patterns generated extensive lists of sequence matches. To reduce the list to a manageable size, several additional, but reasonable restrictions were incorporated into the search pattern. The first was that position 14 be limited to the dominant amino acid, asparagine. Such a restriction reduced the possibilities of sequence matches to a subset of proteins that included the asparagine ladder motif at the T2 turn. Although sequences with known parallel  $\beta$  helix structures lacking an asparagine ladder, such as TSP, pertactin P.69, and RGase A, were eliminated *a priori*, the probability of correctly identifying potential candidates was increased. The second restriction was based upon the observation that the B2-T2-B3 region is the most conserved segment in all parallel  $\beta$  helices. Thus, a search pattern was constructed to require not one, but three occurrences of the B2-T2-B3 profile by specifying amino acids at coil positions -9, -11, -14, -16 as well as 9, 11, 14, 16 and +9, +11, +14, +16. Together, the latter two restrictions greatly reduced the number of sequence matches, allowing subsequent tests of the pattern to focus on the best way to accommodate the expected variability in the length of the T1 and T3 turns. Of the three search

**TABLE I**  
Positional Preference of Amino Acids in Coil

A. Amino Acids Oriented Toward Interior of Parallel $\beta$ Helix of Known Structures									
Inward position	1	3	6	9	11	14	16	19	22
	8I	12I	8G	19V	9I	19N	11F	7I	10A
	5L	8V	8S	8I	6A	3I	8I	5V	3G
	5P	6L	3A	2L	5V	3V	6Y	2A	3S
	4F	1A	2D	1A	4L	2S	2V	2H	3V
	2V	1D	2T		2M	1C	2T	2L	2H
	1A	1F	1C		1C	1L	1L	1F	2M
	1G	1Q	1V		1F	1M		1T	2T
	1S				1S				1I
	1T				1Y				1L
	1W								1N
	1Y								1P
									1Q
Predominant AA	16	27	16	30	24	19	17	16	15
	VAIL	VAIL	GS	VAIL	VAIL	N	FY	VAIL	VAIL

B. Amino Acids Oriented Toward Interior of Postulated Parallel $\beta$ Helix in Members of the Pectate Lyase Superfamily									
Inward position	1	3	6	9	11	14	16	19	22
	6II	55I	47S	82V	73I	107N	80F	21I	57A
	32P	36L	43G	63I	43V	23C	36I	21V	22L
	31L	33V	13V	13F	18S	15I	31Y	20H	21G
	17F	13G	25A	12L	17L	11L	13L	12W	20V
	16V	12A	9T	10A	14A	11V	6M	11L	15M
	10M	12F	1C	3Y	9Y	S	6V	7A	7P
	6G	11C	1P	1C	7F	6M	6W	5M	7T
	3Y	6Q		1G	3M	2T	4T	4P	6S
	2A	5S		1M	2C	1G	2H	4T	4H
	2C	1D			2H			2N	3I
	2S	1M						1C	3N
	1H	1T						1F	1Q
	1N							1Q	
	1T								
	1W								
Predominant AA	110	136	90	167	137	109	116	102	103
	VAIL	VAIL	GS	VAIL	VAIL	N	FYW	VAIL	VAIL

*Note.* (A) The occurrence of an amino acid is listed for each coil position at which the side-chain is oriented toward the interior of the parallel  $\beta$  helix. The coil numbering system is the same as that described in the legend of Fig. 1. In all, six coils from five known structures from the pectate lyase superfamily have been used in the tabulation. The structures include PelC, PelE, *B.s.* Pel, PLA, and PLB. Although individual amino acids are listed separately, the predominant amino acids are grouped according to some common feature, such as propensity to form structural turns (GS), aromaticity (FY or FYW), or hydrophobicity (VAIL). (B) The occurrence of an amino acid is listed for each coil position at which the side-chain is oriented toward the interior of the parallel  $\beta$  helix. The coil numbering is the same as described in the legend of Fig. 1. The frequency of occurrence is derived from the structure-based multiple sequence alignment of 31 proteins in the pectate lyase superfamily (Henrissat *et al.*, 1995). The location of each coil is deduced by analogy to the five known structures listed in A.

patterns reported in Table II, gaps in the idealized coil length of 22 were not permitted in the first two. In Search A, the pattern required a perfect sequence match at the inward positions of the idealized coil. No sequence matches were found in a subsequent search of the databases. In Search B, the amino acid constraints at position 6 and 19, representing the inward position of the T1 and T3 turn, respectively, were eliminated. With the relaxation of the T1 and T3 requirements, several sequence matches were found but none included members of the pectate

lyase superfamily. Analysis revealed that the latter proteins were excluded because all have a coil length that exceeds the idealized length of 22 as a consequence of size variations in the T1 and T3 turns. To address the problem, the pattern for Search C allowed T1 turns, varying from 4 to 6 unspecified amino acids between coil positions 3 and 9 as well as T3 turns of 5 to 15 in length between coil positions 16 and 22. The list of sequence matches increased to a reasonable number and included many members of the pectate lyase superfamily.

**TABLE II**  
Permitted Amino Acids in Search Patterns

Search pattern	-9	-11	-14	-16	-19	-2	1	3	6	9	11	14	16	19	22	+1	+3	+6	+9	+11	+14	+16
A	V A I L	V A I L C F M S Y	N	F Y V I L L T	V A I L H F T	V A I L G S T H M N P Q	V A I L G S T H M N W P	V A I L L F Q	G S T C V A	V A I L C F M S Y	N	F Y V I	V A I L L T	V A I L L H F T	V A I L L G S T H M N P Q	V A I L L G S T H F Y W P	G S T C F Q	V A I L L V A	V A I L L C F A	N	F Y V I L L T	
B	V A I L	V A I L C F M S Y	N	F Y V I L T	X	V A I L G S T H M N W P Q	V A I L L G S T H M N W P	V A I L L F Q	X	V A I L	V A I L C F M S Y	N	F Y V I	X	V A I L L H F T	V A I L L G S T H M N W P Q	V A I L L G S T H F Y W P	X	V A I L L V A	V A I L L C F M S Y	N	F Y V I L L T
C	V A I L	V A I L C F M S Y	N	F Y V I L T	Gap of 5 to 15	V A I L G S T H M N W P Q	V A I L L G S T H M N W P	V A I L L F Q	Gap of 4 to 6	V A I L	V A I L C F M S Y	N	F Y V I	Gap of 5 to 15	V A I L L H F T	V A I L L G S T H M N W P Q	V A I L L G S T H F Y W P	Gap of 4 to 6	V A I L L V A	V A I L L C F M S Y	N	F Y V I L L T

*Search pattern results.* The three search patterns were applied to the Swiss-Prot and PIR amino acid sequence databases and the results are summarized in Table III. No sequence matches were found for the most restrictive pattern, represented in Search A. By releasing the amino acid constraints at coil positions 6 and 19, two members of a protein family, internalin, were identified in Search B. Moreover, one member, internalin A, was found to have four contiguous sequence matches, suggesting that amino acids 156 through 339 fold into a parallel  $\beta$  helix with a minimum of eight coils. When a length variation in the T1 and T3 turns was permitted, as in Search Pattern C, 19 protein families, including many members of the pectate lyase superfamily, were identified. With the exception of the pectate lyases, no proteins with known structure were identified; thus it is not possible to determine the percent-

age of false positive results in Search C. However, like the pectate lyases, 39% of the characterized protein families are known to interact with carbohydrates, possibly indicating a common function. Approximately 50% of the characterized protein families contain sequences with leucine-rich repeats (LRR). Several protein families with LRR sequences have multiple repeats of the search pattern. In addition to internalin, these include the Garp precursor, with four sequence matches; the plant disease resistant factors, tomato Cf-9 and rice Xa21 proteins with two and three sequence matches, respectively; and the *ras*-interacting proteins, adenylate cyclase and Rsp-1 protein, with two and three sequences matches, respectively. Although many LRR proteins with related functions have been grouped into distinctive protein families in Table III, the LRR region is not necessarily involved in the function.



*Internalin A.* Only two proteins were identified in searches in which no gaps were permitted in the T1 and T3 turns: internalin A (MW 80,289) and internalin B (MW 71,144). Both proteins are produced by the gram-positive intracellular pathogen *Listeria monocytogenes*. Internalin A has four contiguous repeats of the search pattern and additional sequence matches can be identified in the LRR region, if the restriction for three consecutive repeats of the B2-T2-T3 motif is removed. The N-terminal region of internalin A, which contains the repeated segments, is involved in the attachment of the bacteria to *E-cadherin* in cultured epithelial cells while the C-terminal portion is anchored in the bacterial membrane (Gaillard *et al.*, 1991; Mengaud *et al.*, 1996a; Mengaud *et al.*, 1996b). Attempts to identify a structural motif, using 1D-3D threading algorithms (Jones *et al.*, 1995; Alexandrov *et al.*, 1996; Rice and Eisenberg, 1997), failed to identify a probable fold for the repetitive sequence region of internalin A. To determine if the parallel  $\beta$  helix is a plausible topology for the contiguous repetitive regions of internalin A, a model was constructed using as a template the idealized parallel  $\beta$  helix that had been generated from PelC. The repetitive asparagines of internalin A were aligned with the asparagine ladder of the idealized parallel  $\beta$  helix and the remaining amino acids of the internal five repetitive units were superimposed upon the rest of the idealized structure. Model construction was simplified by the lack of structural insertions because the 22-amino acid sequence repeats in internalin A coincided with the length of the idealized coil. The refined internalin A model is shown in Fig. 2. One striking feature was the right-handed superhelical twist of the extended parallel  $\beta$  helix. Not unexpectedly, all of the repetitive leucines were found at coil positions 3, 6, 9, 11, and 22, all of which have side chains oriented toward the interior of the parallel  $\beta$  helix. The program PROCHECK (Laskowski *et al.*, 1993) was used to verify the stereochemical quality. No amino acids were found in disallowed regions of the Ramachandran plot and none had prohibitive contacts. Furthermore, a 3D-1D profile score was calculated for the model and the value of 49.3 lies at the upper range of scores, approximately 22-50, for other correct structures with 110 amino acids. Also, the profile window plot (Fig. 3) was comparable to correct protein structures.

To further evaluate the plausibility of the model, three additional models were constructed, refined, and scored. In two models, the sequence was shifted by +1 or +3, respectively, from the position of the asparagine ladder. These shifts represent the minimum (6) and the maximum (16) number of amino acids, which have the same inward, or outward,

orientation of the amino acids in the original search pattern. The calculated 3D-1D score of 4.7 for the +1 model is even much lower than those reported for erroneous protein structures with 110 amino acids. The score of 24.1 for the +3 model lies at the lower and questionable range of correct structures. In the third model, the composition of amino acids was fixed but the sequence was randomized. The 3D-1D profile score of 7.3 for the randomized sequence also was indicative of an incorrect model. Together, the modeling results suggest that the parallel  $\beta$  helix topology, including the novel asparagine ladder motif, is a plausible structural model for the LRR region extending from amino acids 156 through 339 in internalin A.

As an independent and additional confirmation of the internalin A structure, the circular dichroism spectrum for the full-length protein was recorded, analyzed, and compared to the CD spectra of several pectate lyases. The CD spectra of the latter proteins resemble those for antiparallel  $\beta$  sheets except that the amplitudes at the maximum and the minimum are unusually large (Sieber *et al.*, 1995). The magnitudes of the amplitudes as well as the ratio of the extrema appear to be characteristic signatures for the parallel  $\beta$  helix fold. The CD spectrum of internalin A is shown in Fig. 4. The spectrum was deconvoluted using a set of basis spectra that contained the average parallel  $\beta$  helix spectrum of PelC and PelE. Superimposed upon the measured spectrum is the fitted spectrum as well as the component spectra of the fitted spectrum for full-length internalin A. The excellent agreement between the measured and fitted spectrum could not otherwise be achieved without the inclusion of the parallel  $\beta$  helix basis spectrum. Based on the percentage of each component spectrum contributing to fitted spectrum, the secondary structural content of internalin A is 4%  $\alpha$  helix, 1% antiparallel  $\beta$  sheet, 73% aperiodic conformations, and 22% parallel  $\beta$  helix. The composition of the fitted spectrum, including the large percentage of the aperiodic conformations, is very similar to that found for PelC and for PelE (Sieber *et al.*, 1995). Thus, the CD spectral analysis is consistent with the presence of a parallel  $\beta$  helix topology in internalin A.

## DISCUSSION

The parallel  $\beta$  helix is a new type of structural domain that has been observed in six different protein families since its original discovery in 1993. Identification of the structural motif from the amino acid sequence is challenging because an obvious sequence repeat does not exist, the  $\beta$  strands are unusually short, and the coils within the parallel  $\beta$  helix are variable in length. In the present research,

**TABLE III**  
Search Results

Protein family	Database entry	Type <sup>a</sup>	Sequence no. of match	Search ID
1. ATP-Dependent nuclease subunit	SW:ADDA_BACSU		607–658	C
2. Cell adhesion proteins				
a. Choptin precursor	SW:CHAO_DROME	Carb, LRR, ME	230–290	C
	SW:CHAO_DROME		531–588	C
b. Slit protein precursor	SW:SLIT_DROME	LRR	399–454	C
c. Toll protein precursor	SW:TOLL_DROME	Carb, LRR, ME	478–534	C
3. Garp precursor	PIR:S42799	LRR	54–109	C
	PIR:S42799		102–161	C
	PIR:S42799		223–277	C
	PIR:S42799		344–398	C
4. Genome polyproteins				
a. B	SW:VGNB_CPMV	Protease Seq.	1451–1511	C
b. N	SW:POLN_FCWF4	Protease Seq.	132–197	C
	SW:POLN_FCWF6	Protease Seq.	1038–1103	C
	SW:POLN_FCWF9	Protease Seq.	1514–1579	C
c. M Precursor	SW:VGLM_INSV		280–343	C
d. P	SW:POLG_PVYC		416–476	C
	SW:POLG_PVYHU		416–476	C
	SW:POLG_PVYN		416–458	C
	SW:POLG_PVYO		416–458	C
5. Geranylgeranyltransferase $\alpha$ subunit	SW:PGTA_RAT		468–520	C
6. Hypothetical proteins				
a. MJ1341 protein	PIR:D64467		23–86	C
b. Silkworm protein 2	PIR:S08405		661–722	C
c. Wheat protein	PIR:S43889		13–65	C
d. YOR353c protein	PIR:S67265		94–149	C
e. 72.1-kDa protein in ACS1-PTA1 intergenic region	SW:YAE7_YEAST		482–543	C
f. 118.2-kDa protein f43c1.1	SW:YR71_CAEEL		542–596	C
g. 153.4-kDa protein b0523.5	SW:YKF5_CAEEL		289–344	C
7. Insulin-like growth factor binding protein complex acid labile chain precursor	SW:ALS_HUMAN	Carb, LRR	463–518	C
8. Internalin				
a. Internalin A	SW:INLA_LISMO	LRR	156–207	B, C
	SW:INLA_LISMO		200–251	B, C
	SW:INLA_LISMO		244–295	B, C
	SW:INLA_LISMO		288–339	B, C
b. Internalin B	SW:INLB_LISMO	LRR	147–198	B, C
9. Major ring-forming surface protein	PIR:S41525	ME	462–519	C
10. Pectate/pectin lyase precursors	PIR:A44852	Carb	359–318	C
	PIR:JC1313	Carb	26–278	C
	SW:PEL_BACSU	Carb	287–346	C
	SW:PEL1_ERWCA	Carb	226–278	C
	SW:PEL3_ERWCA	Carb	226–278	C
	SW:PELB_ERWCA	Carb	226–278	C
	SW:PELC_ERWCA	Carb	226–278	C
	SW:PELA_ERWCH	Carb	260–319	C
	SW:PELC_ERWCH	Carb	227–279	C
	SW:PELD_ERWCH	Carb	253–312	C
	SW:PELE_ERWCH	Carb	247–306	C
	SW:PELF_ERWCH	Carb	265–324	C
	SW:PLYA_ASPNG	Carb	243–296	C
11. Plant disease resistance proteins				
a. Cf-9 protein—tomato	PIR:A55173	LRR	364–419	C
	PIR:A55173		555–612	C
b. Xa21 protein—rice	PIR:A57676	LRR, ME	108–163	C
	PIR:A57676		381–436	C
	PIR:A57676		429–484	C
	PIR:A57676		550–605	C
12. Polygalacturonase inhibitor precursor	SW:PGIP_PYRCO	Carb	197–251	C
13. Pollen-specific genes with extensin-like domains	PIR:S49915	Carb	275–330	C

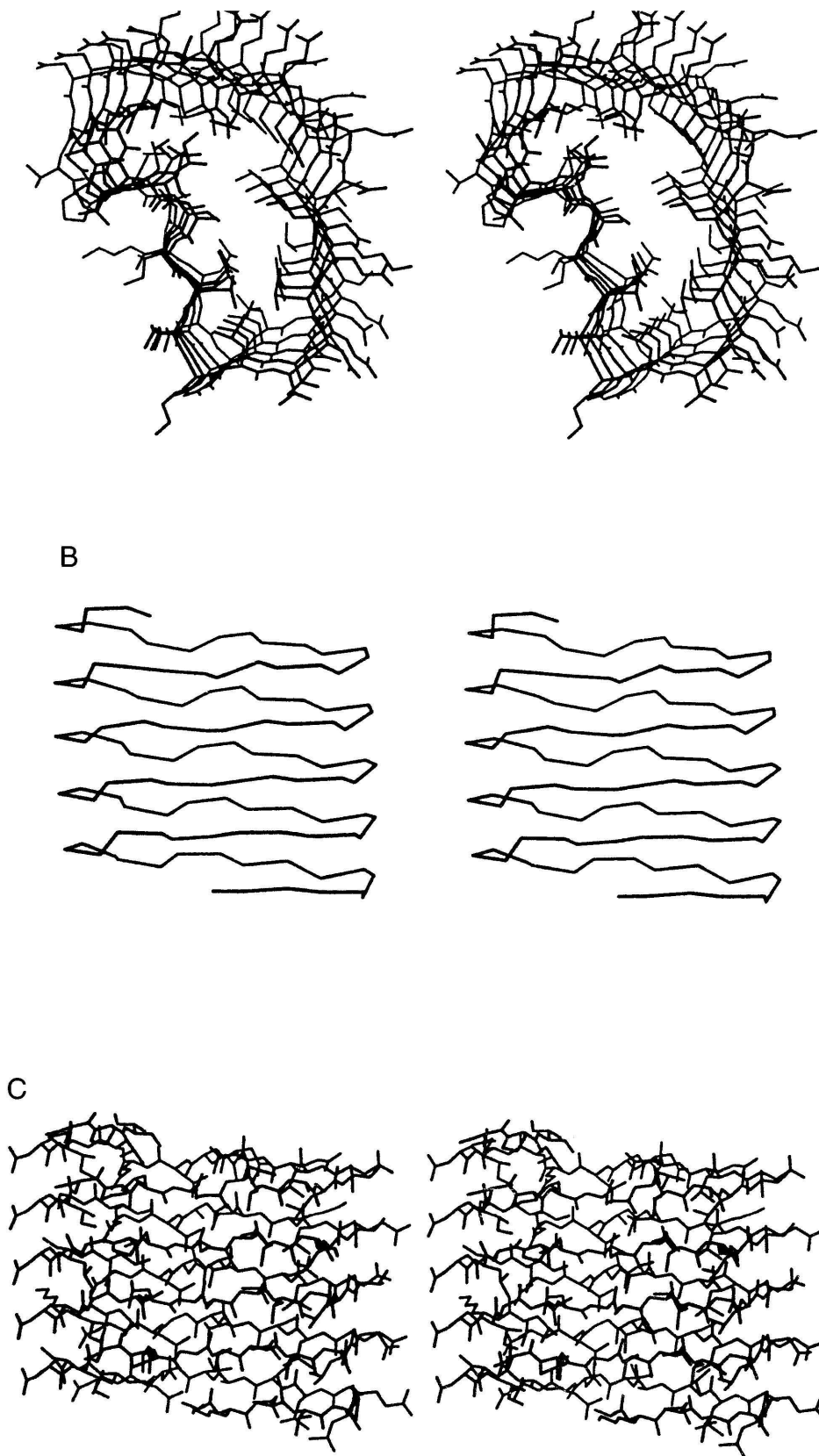
TABLE III—Continued

Protein family	Database entry	Type <sup>a</sup>	Sequence no. of match	Search ID
14. <i>Ras</i> -interacting proteins				
a. Adenylate cyclase	PIR:JC4747	LRR	877–930	C
	SW:CYAA_NEUCR	LRR	944–998	C
	SW:CYAA_SACKL	LRR	954–1017	C
	SW:CYAA_SCHPO	LRR	530–583	C
	SW:CYAA_USTMA	LRR	1367–1420	C
	SW:CYAA_USTMA		1563–1617	C
	SW:CYAA_YEAST	LRR	1069–1122	C
b. Flightless-I protein	PIR:S60461	LRR	60–115	C
Flightless-I homolog	PIR:A49674	LRR	60–115	C
c. LRR47 protein	PIR:S45361	Carb, LRR, ME	182–239	C
d. Rsp-1 protein	SW:RSP1_Mouse	LRR	45–98	C
	SW:RSP1_Mouse		139–192	C
	SW:RSP1_Mouse		726–785	C
15. Receptor proteins				
a. Biglycans	SW:PGS1_BOVIN	Carb, LRR	222–276	C
	SW:PGS1_HUMAN	Carb, LRR	258–312	C
	SW:PGS1_MOUSE	Carb, LRR	259–313	C
	SW:PGS1_RAT	Carb, LRR	259–313	C
b. Decorins	SW:PGS2_CHICK	Carb, LRR	224–310	C
	SW:PGS2_MOUSE	Carb, LRR	245–299	C
	SW:PGS2_RAT	Carb, LRR	245–199	C
c. Epidermal growth factor receptors	PIR:A45558	ME	95–158	C
	PIR:B45558	ME	95–158	C
	PIR:C45558	ME	95–158	C
	PIR:D45558	ME	95–158	C
d. RP105 cell surface protein	PIR:I56258	LRR, ME	525–579	C
e. TMKL1 protein	SW:TML1_ARATH	Carb, LRR, Me	204–258	C
16. <i>sds22</i> <sup>+</sup> regulatory subunit of protein phosphatases PP1	SW:SD22_SCHPO	LRR, NU	242–297	C
17. Transport system permease				
a. Galactoside MGLC	SW:MGLC_ECOLI	Carb, ME	197–260	C
	SW:MGLC_HAEIN	Carb, ME	197–260	C
b. Putative glutamine	SW:GLNP_RICPR	ME	51–115	C
18. U2 small nuclear ribonucleoprotein A'	SW:RU2A_ARATH	LRR, NU	69–125	C
19. Wheat proteins				
AWJL172	PIR:S49299	LRR	241–296	C
AWJL236	PIR:S49300	LRR	224–279	C
AWJL218	PIR:S49302	LRR	191–246	C
AWJL175	PIR:S49301	LRR	180–235	C
	PIR:S49301		303–358	C

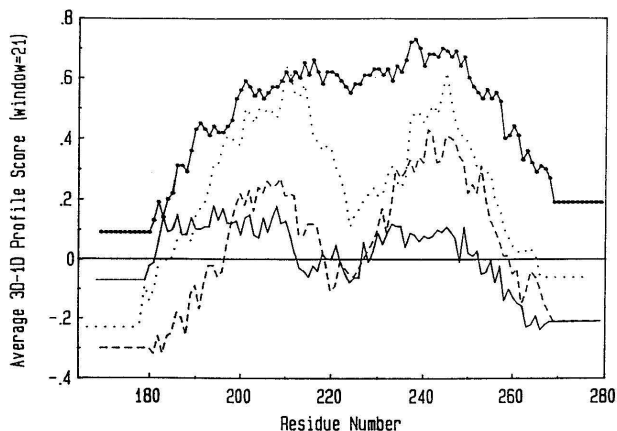
<sup>a</sup> Codes for protein types are as follows: Carb, carbohydrate binding protein; LRR, protein containing leucine-rich repeats; ME, transmembrane, membrane-anchored or membrane-associated and NU, nuclear protein. SW prefix denotes the entry from the Swiss-Prot database; and PIR is the prefix for the Protein Information Resource database.

an idealized coil, consisting of 22 amino acids in one rung of the parallel  $\beta$  helix, has been extrapolated from the structure of PelC in order to characterize any novel features. A comparison of five pectate lyase structures reveals a striking preference for certain amino acids in specific positions of the coil. Most significant, the sequence profile of the favored amino acids in a single coil of the parallel  $\beta$  helix is very different from the profile of either an  $\alpha$  helix or a longer  $\beta$  strand. The unique profile forms the basis of sequence patterns developed herein to search databases in order to identify protein families that might share the parallel  $\beta$  helix topology. Possible search

patterns have ultimately been limited to a few that identify a significant number of the pectate lyase superfamily, without generating an extensive list of proteins to characterize by alternate means. As a consequence of the restrictions, some proteins known to contain a parallel  $\beta$  helix, but lacking certain features such as the asparagine ladder, were not identified in the search. Moreover, it is not possible to estimate the percentage of false positives because, with the exception of the pectate lyases, the three-dimensional structures are not known for any other proteins identified by the searches. Nevertheless, it is notable that 39% of the protein families identified



**FIG. 2.** Stereo views of the internalin A model. The model includes five helical coils containing 110 amino acids, from residues 170 to 279 in the repetitive region of internalin A. (A) Cross-sectional view of the internalin A model. The stacking interactions of the side-chains in the interior are similar to that observed in PelC and PelE. (B)  $\alpha$ C trace of the internalin A model. The view is perpendicular to the axis of the parallel  $\beta$  helix. (C) Same view as in B, but with all atoms of the internalin A model shown.



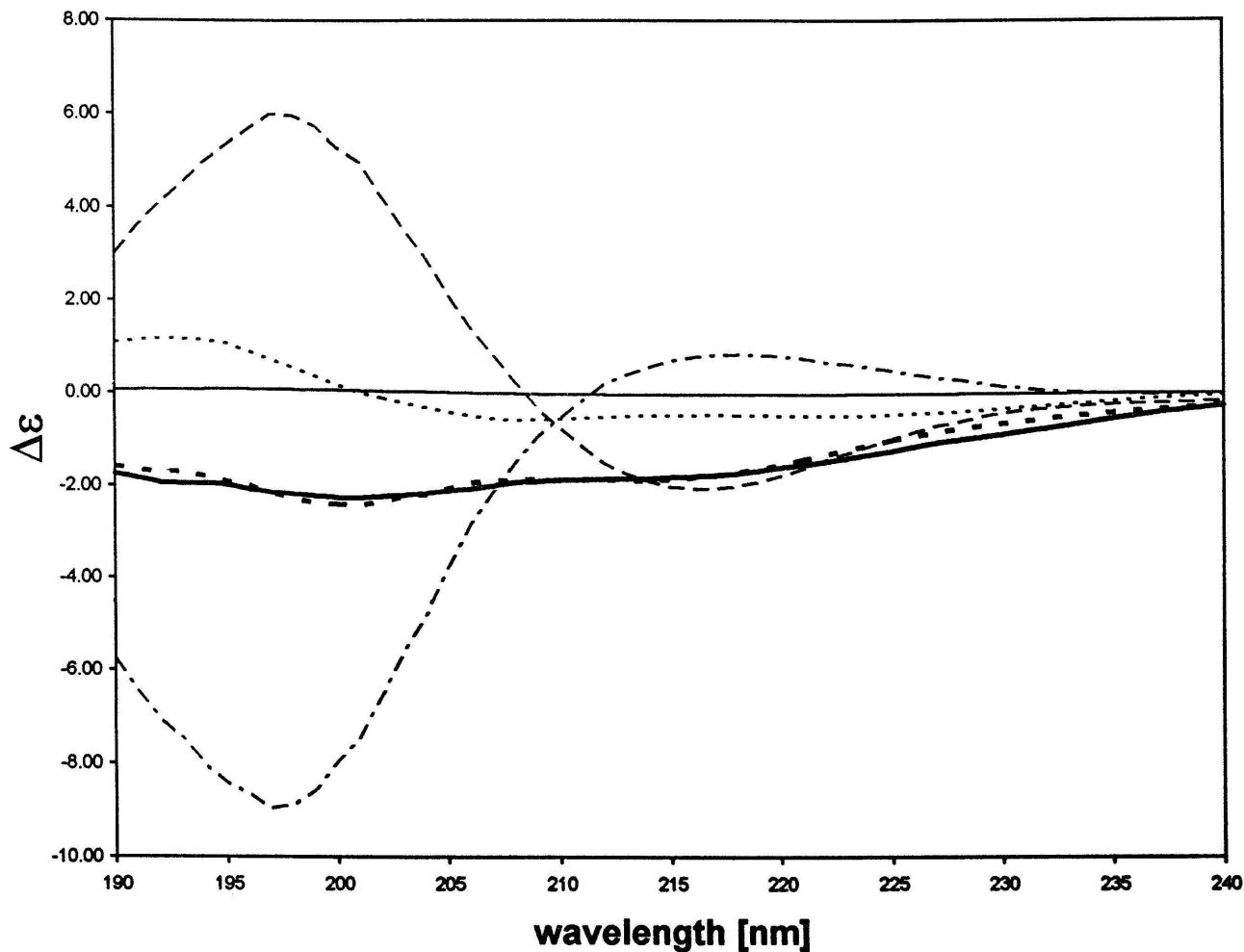
**FIG. 3.** Profile window plot of the average 3D-1D profile score versus the amino acid sequence number for four models of internalin A, using a window length of 21. The model with the repetitive asparagines in internalin A correctly aligned to position 14 of the idealized parallel  $\beta$  helix structure is shown by the line with filled circles at the data points. The overall 3D-1D profile score of 49.3 lies in the range of correct structures of proteins with 110 amino acids. Moreover, the window-averaged 3D-1D profile score over all sequence ranges is indicative of a correct structural model. Three models of internalin A with erroneous alignments to the idealized parallel  $\beta$  helix structure were intentionally constructed. The solid line indicates the model in which the repetitive asparagine was shifted by +1; the dotted line indicates a shift of +3 in the repetitive asparagine; and the dashed line indicates a model in which the amino acid composition was fixed but the sequence was randomized. The overall 3D-1D profile scores as well as the averaged values over the range of residues are indicative of incorrect structural models.

are carbohydrate-binding proteins, a function shared by 80% of the proteins with known parallel  $\beta$  helix structures.

Approximately 50% of the characterized protein families identified by the parallel  $\beta$  helix search pattern C contain amino acid sequences known as leucine-rich repeats (LRR) (for review, Kobe and Deisenhofer, 1994; Buchanan and Gay, 1996). As the name implies, LRRs are multiple repeats of a consensus sequence containing many leucines. The consensus sequence varies from 20 to 29 amino acids and many contain an invariant asparagine as well. Proteins with LRR segments have different cellular locations and diverse functions, ranging from RNA processing to cell adhesion. For comparative purposes, the classification of Buchanan and Gay (1996) has been used to group the LRR proteins that have sequence matches with the parallel  $\beta$  helix search patterns. Although more than 70 LRR proteins have been sequenced, the three-dimensional structures of only two have been reported: porcine ribonuclease inhibitor (RINI) with typical LRR repeats (Kobe and Deisenhofer, 1993) and *Azotobacter vinelandii* nitrogen fixation specific Fe-binding protein, an LRV variant (Peters *et al.*, 1996). Neither protein is

identified in the search patterns for parallel  $\beta$  helices as a consequence of amino acid deviations in coil positions 1, 3, and 19. In RINI, each of the LRR segments, either 28 or 29 amino acids in length, folds into an  $\alpha$  helix/ $\beta$  strand unit. The  $\beta$  strands of the  $\alpha/\beta$  units assemble in a parallel manner and with 15 such units, the structure of RINI folds into a horseshoe, rather than the more typical sheet or barrel motifs. However, it is highly unlikely that a similar structural fold will be found if the LRR segment is less than 25 amino acids because the  $\beta$  strands in RINI are very short, averaging only three amino acids in length. The turns in RINI are tight and already use a minimal number of amino acids. Therefore, a decrease in the LRR length, from 28 to 25, would necessitate the removal of two amino acids from the  $\alpha$  helix and one from the  $\beta$  strand, reducing the latter to an improbable length of two. The second protein contains a shorter repeat with 24 amino acids. However, the repetitive sequence is a variant, containing arginines instead of leucines in two key positions of the typical LRR sequence. Thus, the structure of the Fe-binding protein may not be representative of other LRR proteins. The LRV region folds into an  $\alpha/\alpha$  coiled domain, similar to the lipoprotein, lipovitellin (Banaszak *et al.*, 1991; Yoder and Jurnak, 1995a). In the latter fold, alternating helices are equivalent to the  $\alpha$  helices in the  $\alpha/\beta$  units of RINI, and the other set of helices occupy the equivalent location as the  $\beta$  strands. Notably, both the LRR and LRV protein folds contain  $\alpha$  helices, a secondary structural element that can be readily detected by CD spectroscopy.

All of the LRR proteins identified by the parallel  $\beta$  helix search patterns contain repetitive sequences with 25 or fewer amino acids. The most striking sequence match is that for internalin A. Not only does the LRR repeat of 22 coincide with the number of amino acids in the idealized coil, but there are four contiguous sequence regions that match the search patterns. Although threading algorithms failed to identify a probable structural motif for internalin A, the modeling studies herein, based on a parallel  $\beta$  helix motif, demonstrate that all repetitive leucines are readily accommodated in coil positions oriented toward the interior. Moreover, the repetitive asparagines form an asparagine ladder motif at the T2 turn. The high 3D-1D profile score of the primary internalin A model and the low scores for models with known errors support the possibility that the parallel  $\beta$  helix model is a feasible fold for the LRR region of internalin A. Additional support for this conclusion derives from the deconvolution of the CD spectrum, which reveals a parallel  $\beta$  helix component equivalent to 22% of the entire structure. The notion that the LRR regions of internalin A might



**FIG. 4.** Analysis of the internalin A CD spectrum. The measured CD spectrum of internalin A (heavy solid line) is superimposed upon the fitted spectrum from a linear combination of basis spectra (heavy dashed line) composed of 4%  $\alpha$  helix (dotted line), 1% antiparallel  $\beta$ -sheet (solid line), 73% aperiodic conformations (dotted dashed line), and 22% parallel  $\beta$  helix (dashed line). The leucine-rich repetitive region of internalin A spans residues 29 through 375 (MW 35,792) out of 744 amino acids (MW 80,289) in the full-length protein.

fold into a parallel  $\beta$  helix is not entirely new. Internalin A, as well as other LRR proteins, were identified in mid-1993 using early search patterns, based only on the PelC structure. The preliminary results were reported (Jurnak *et al.*, 1994; Yoder and Jurnak, 1995a), albeit without complete documentation, and further elaborated by others (Kobe and Deisenhofer, 1994; Buchanan and Gay, 1996). The current report presents not only the complete rationale for the original hypothesis, but additional experimental data which supports the conclusion that internalin A and possibly other proteins with short LRR segments are likely to fold into a parallel  $\beta$  helix. That pectate lyases and internalins share a similar structural fold is somewhat intriguing. Both protein families are involved in virulence, one targeting plant cells and the other, mammalian tissue. Whether this feature reveals a common origin remains to be established.

The search patterns reported herein represent the first published attempt to identify proteins with primary sequences that are compatible with the parallel  $\beta$  helix fold. Until more examples of the parallel  $\beta$  helix become available, it is not possible to determine the validity of the predictions without additional experimental data. Nevertheless, the success in identifying and characterizing the folding motif in the LRR regions of internalin A support the legitimacy of the present results. However, additional structural examples, improvements in the search patterns, and/or the development of a suitable scoring matrix for profile searches will be needed before the parallel  $\beta$  helix fold can be predicted with confidence from the protein sequence alone.

The authors gratefully acknowledge the support of the National Science Foundation (MCB9408999), the United States Depart-

ment of Agriculture (Award 96-02966), and the San Diego Super-computer Center.

## REFERENCES

- Alexandrov, N. N., Nussinov, R., and Zimmer, R. M. (1996) Fast protein fold recognition via sequence to structure alignment and contact capacity potentials, *in* Hunter, L., and Klein, T. (Eds.), *Biocomputing: Proceedings of the 1996 Pacific Symposium*, pp. 53–72, World Scientific, Singapore.
- Banaszak, L., Sharrock, W., and Timmins, P. (1991) Structure and function of a lipoprotein: Lipovitellin, *Annu. Rev. Biophys. Chem.* **20**, 221–246.
- Baumann, U., Wu, S., Flaherty, K. M., and McKay, D. B. (1993) Three-dimensional structure of the alkaline protease of *Pseudomonas aeruginosa*: A two-domain protein with a calcium binding parallel beta roll motif, *EMBO J.* **12**, 3357–3364.
- Brünger, A. (1988) Crystallographic refinement by simulated annealing: Application to a 2.8 Å resolution structure of aspartate aminotransferase, *J. Mol. Biol.* **208**, 803–816.
- Brünger, A. (1993) XPLOR Manual, Version 3.1, Yale University, New Haven, CT.
- Buchanan, S. G. S., and Gay, N. J. (1996) Structural and functional diversity in the leucine-rich repeat family of proteins, *Prog. Biophys. Molec. Biol.* **65**, 1–44.
- Edelhoch, H. (1967) Spectroscopic determination of tryptophan and tyrosine in proteins, *Biochemistry* **6**, 1948–1954.
- Emsley, P., Charles, I. G., Fairweather, N. F., and Isaacs, N. W. (1996) Structure of *Bordetella pertussis* virulence factor P.69 pertactin, *Nature* **381**, 90–92.
- Gaillard, J.-L., Berche, P., Fréhel, C., Gouin, E., and Cossart, P. (1991) Entry of *L. monocytogenes* into cells is mediated by internalin, a repeat protein reminiscent of surface antigens from gram-positive cocci, *Cell* **65**, 1127–1141.
- Henrissat, B., Heffron, S. E., Yoder, M. D., Lietzke, S. E., and Journak, F. (1996) Functional implications of a structure-based sequence alignment of proteins in the extracellular pectate lyase superfamily, *Plant Physiol.* **107**, 963–976.
- Johnson, W. C. (1990) Protein secondary structure and circular dichroism: A practical guide, *Proteins* **7**, 205–214.
- Jones, T. A., and Kjeldgaard, M. (1993) O: The Manual, Version 5.10.1, Uppsala University, Uppsala, Sweden.
- Jones, D. T., Miller, R. T., and Thornton, J. M. (1995) Successful protein fold recognition by optimal sequence threading validated by rigorous blind testing, *Proteins* **23**, 387–397.
- Journak, F., Yoder, M. D., Lietzke, S. E., and Heffron, S. E. (1994) Novel structural features of pectate lyases and implications for membrane proteins, American Crystallographic Association Meeting, Association Meeting, Atlanta, GA.
- Kajava, A. V., Vassart, G., and Wodak, S. J. (1995) Modeling of the three-dimensional structure of proteins with the typical leucine-rich repeats, *Structure* **3**, 867–877.
- Kisker, C., Schindelin, H., Alber, B. E., Ferry, J. G., and Rees, D. C. (1996) A left-handed  $\beta$ -helix revealed by the crystal structure of a carbonic anhydrase from the archaeon *Methanosarcina thermophila*, *EMBO J.* **15**, 2323–2330.
- Kobe, B., and Deisenhofer, J. (1993) Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats, *Nature* **266**, 751–756.
- Kobe, B., and Deisenhofer, J. (1994) The leucine-rich repeat: a versatile binding motif, *J. Trends Biochem. Sci.* **19**, 415–421.
- Laskowski, R. A., McArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr.* **26**, 282–291.
- Levitt, M., and Chothia, C. (1976) Structural patterns in globular proteins, *Nature* **261**, 552–558.
- Lietzke, S. E., Keen, N. T., Yoder, M. D., and Journak, F. (1994) The three-dimensional structure of pectate lyase E, a plant virulence factor from *Erwinia chrysanthemi*, *Plant Physiol.* **106**, 849–862.
- Lietzke, S. E., Scavetta R. D., Yoder, M. D., and Journak, F. (1996) The refined three-dimensional structure of pectate lyase E from *Erwinia chrysanthemi*, *Plant Physiol.* **111**, 73–92.
- Lüthy, R., Bowie, J. U., and Eisenberg, D. (1992) Assessment of protein models with three-dimensional profiles, *Nature* **356**, 83–85.
- Mengaud, J., Ohayon, H., Gounon, P., Mege, R.-M., and Cossart, P. (1996a) E-cadherin is the receptor for internalin, a surface protein required for entry of *L. monocytogenes* into epithelial cells, *Cell* **84**, 932–932.
- Mengaud, J., Lecuit, M., Lebrun, M., Nato, F., Mazie, J. C., and Cossart, P. (1996b) Antibodies to the leucine-rich repeat region of internalin block entry of *Listeria monocytogenes* into cells expressing E-cadherin, *Infect. Immunol.* **12**, 5430–5433.
- Mayans, O., Scott, M., Connerton, I., Gravesen, T., Benen, J., Visser, J., Pickersgill, R., and Jenkins, J. (1997) Two crystal structures of pectin lyase A from *Aspergillus* reveal a pH driven conformational change and striking divergences in the substrate-binding clefts of pectin and pectate lyases, *Structure* **5**, 677–689.
- Peters, J., Stowell, M. H. B., and Rees, D. C. R. (1996) A leucine-rich repeat variant with a novel repetitive protein structural motif, *Nature Struct. Biol.* **3**, 991–994.
- Petersen, T. N., Kauppinen, S., and Larsen, S. (1997) The crystal structure of rhamnogalacturonase A from *Aspergillus aculeatus*: a right-handed  $\beta$ -helix, *Structure* **5**, 533–544.
- Pickersgill, R., Jenkins, J., Harris, G., Nasser, W., and Robert-Baudouy, J. (1994) The structure of *Bacillus subtilis* pectate lyase in complex with calcium, *Nature Struct. Biol.* **1**, 717–723.
- Raetz, C. R. H., and Roderick, S. L. (1995) A left-handed parallel  $\beta$  helix in the structure of UDP-*N*-acetylglucosamine acyltransferase, *Science* **270**, 997–1000.
- Rice, D., and Eisenberg, D. (1997) A 3D–1D substitution matrix for protein fold recognition that includes predicted secondary structure of the sequence, *J. Mol. Biol.* **267**, 1026–1038.
- Richardson, J. S. (1981) The anatomy and taxonomy of protein structure, *Adv. Protein Chem.* **34**, 167–339.
- Sieber, V., Journak, F., and Moe, G. R. (1995) Circular dichroism of the parallel  $\beta$  helical proteins pectate lyase C and E, *PROTEINS: Struct. Funct. Genet.* **23**, 32–37.
- Steinbacher, S., Seckler, R., Miller, S., Steipe, B., Huber, R., and Reinemer, P. (1994) Crystal structure of P22 tailspike-protein: Interdigitated subunits in a thermostable trimer, *Science* **265**, 383–386.
- Vitali, J., Schick, B., Kester, H. C. M., Visser, J., and Journak, F. (1998) The three-dimensional structure of *Aspergillus niger* pectin lyase B at 1.7 Å resolution, *Plant Physiol.* **116**, 1–13.
- Yoder, M. D., Keen, N. T., and Journak, F. (1993a) New domain motif: structure of pectate lyase C, a secreted plant virulence factor, *Science* **260**, 1503–1507.
- Yoder, M. D., Lietzke, S. E., and Journak, F. (1993b) Unusual structural features of the parallel  $\beta$  helix, *Structure* **1**, 241–251.
- Yoder, M. D., and Journak, F. (1995a) The parallel  $\beta$  helix and other coiled folds, *FASEB J.* **9**, 335–342.
- Yoder, M. D., and Journak, F. (1995b) The refined three-dimensional structure of pectate lyase C from *Erwinia chrysanthemi* at 2.2 Å resolution, *Plant Physiol.* **107**, 349–364.