



# Super Secondary Structure Consisting of a Polyproline II Helix and a $\beta$ -Turn in Leucine Rich Repeats in Bacterial Type III Secretion System Effectors

Dashdavaa Batkhishig<sup>1,2</sup> · Khurelbaatar Bilguun<sup>1,3</sup> · Purevjav Enkhbayar<sup>1</sup> · Hiroki Miyashita<sup>4,5</sup> · Robert H. Kretsinger<sup>6</sup> · Norio Matsushima<sup>1,5,7</sup>

Published online: 12 April 2018  
© The Author(s) 2018

## Abstract

Leucine rich repeats (LRRs) are present in over 100,000 proteins from viruses to eukaryotes. The LRRs are 20–30 residues long and occur in tandem. LRRs form parallel stacks of short  $\beta$ -strands and then assume a super helical arrangement called a solenoid structure. Individual LRRs are separated into highly conserved segment (HCS) with the consensus of LxxLx-LxxNxL and variable segment (VS). Eight classes have been recognized. Bacterial LRRs are short and characterized by two prolines in the VS; the consensus is xxLPxLPxx with Nine residues (N-subtype) and xxLPxxLPxx with Ten residues (T-subtype). Bacterial LRRs are contained in type III secretion system effectors such as YopM, IpaH3/9.8, SspH1/2, and SlrP from bacteria. Some LRRs in decorin, fribromodulin, TLR8/9, and FLRT2/3 from vertebrate also contain the motifs. In order to understand structural features of bacterial LRRs, we performed both secondary structures assignments using four programs—DSSP-PPII, PROSS, SEGNO, and XTLSSTR—and HELFIT analyses (calculating helix axis, pitch, radius, residues per turn, and handedness), based on the atomic coordinates of their crystal structures. The N-subtype VS adopts a left handed polyproline II helix (PPII) with four, five or six residues and a type I  $\beta$ -turn at the C-terminal side. Thus, the N-subtype is characterized by a super secondary structure consisting of a PPII and a  $\beta$ -turn. In contrast, the T-subtype VS prefers two separate PPIIs with two or three and two residues. The HELFIT analysis indicates that the type I  $\beta$ -turn is a right handed helix. The HELFIT analysis determines three unit vectors of the helix axes of PPII (**P**),  $\beta$ -turn (**B**), and LRR domain (**A**). Three structural parameters using these three helix axes are suggested to characterize the super secondary structure and the LRR domain.

**Keywords** Bacterial leucine rich repeat family · Polyproline II helix · Type I  $\beta$ -turn · Super secondary structure · Helical parameters · Helix axis · Vector analysis

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s10930-018-9767-9>) contains supplementary material, which is available to authorized users.

✉ Purevjav Enkhbayar  
enkhbayar.p@seas.num.edu.mn

✉ Norio Matsushima  
norio\_irreko@outlook.jp

<sup>1</sup> Laboratory of Bioinformatics and Systems Biology, Department of Information and Computer Science, School of Engineering and Applied Sciences, National University of Mongolia, Ulaanbaatar 14201, Mongolia

<sup>2</sup> Department of Physics, School of Mathematics and Natural Sciences, Mongolian National University of Education, Ulaanbaatar 210648, Mongolia

## Abbreviations

LRR Leucine rich repeat  
PPII Polyproline II helix  
HCS Highly conserved segment

<sup>3</sup> Institute of Physics and Technology, Mongolian Academy of Sciences, Enkhtaivan avenue 54B, Ulaanbaatar 210651, Mongolia

<sup>4</sup> Hokubu Rinsho Co., Ltd, Sapporo 060-0061, Japan

<sup>5</sup> Institute of Tandem Repeats, Sapporo 004-0882, Japan

<sup>6</sup> Department of Biology, University of Virginia, Charlottesville 22904, USA

<sup>7</sup> Sapporo Medical University, Sapporo 060-8556, Japan

VS	Variable segment
$C_{\alpha}$	$\alpha$ -Carbon
CC	Cysteine containing
$P$	Helix pitch
$R$	Helix radius
$N$	Number of repeat unit/residue per turn in helix
$\delta z$	Rise per repeat unit/residue in helix
$\Delta\Phi$	Rotation angle per repeat unit/residue in helix
$n$	Repeat number of leucine rich repeat
PDB	Protein data bank
RMSD	Root mean square deviation
TLR	Toll like receptor
SLRP	Small leucine rich repeat proteoglycan protein
FLRT	Fibronectin leucine rich repeat transmembrane protein
CTD	C-terminal repeat domain of the large subunit of RNA polymerase II
POL II	RNA polymerase II

## 1 Introduction

Leucine rich repeats (LRRs) are unusually rich in leucine [1–6]. The LRRs are composed of 20–30 residues stretches and repeat in tandem. The published repeat numbers range from 2 to 97. LRRs have been reported in over 100,000 proteins from viruses to eukaryotes.

LRR units are divided into a highly conserved segment (HCS) and a variable segment (VS) [1]. Eight classes of LRRs have been recognized [3]. Matsushima and Kretsinger recently proposed twenty-three types of LRRs [1]. Their grouping is based mainly on the difference of the VS parts. The eight classes are RI-like, cysteine containing (CC), SDS22-like, IRREKO, bacterial, plant specific, typical, and TpLRR.

The HCS part consists of an 11 or 12 residue stretch,  $LxxLxLxx(N/C)(x/-)L$ , in which “L” is Leu, Ile, Val, or Phe, “N” is Asn, Thr, Ser, or Cys, “C” is Cys, Ser or Asn, “x” is any amino acid, and “-” is a deletion. Three residues at positions 3–5 in the underlined residues form a short  $\beta$ -strand [4, 6]. These  $\beta$ -strands stack parallel; they have the pattern of H-bonding ( $N-H \rightarrow O=C$ ), and then these tandem repeats of LRRs assume their super helical arrangements. The LRRs fold into a horse shoe, a right handed or left handed helix, or a prism shape [7]. Conserved hydrophobic residues such as leucine, valine, isoleucine, or phenylalanine in the consensus sequences of LRRs contribute to the hydrophobic cores. Capping structures that shield the hydrophobic core of the first LRR unit at the  $N$ -terminus and/or the last unit at the  $C$ -terminus are observed in most of the known LRR structures [1, 5].

Characteristic of each LRR class, the VS parts adopt a variety of secondary structures including the  $\alpha$ -helix,

$3_{10}$ -helix, and an extended conformation or a tandem arrangement of  $\beta$ -turns [1, 6]. Their secondary structures on the convex side are connected to the strands forming the  $\beta$ -sheet on its concave side by two loops [5]. One of the loops is an “ascending loop” which links the  $C$ -terminus of the HCS to the  $N$ -terminus of the VS. The other is a “descending loop” which links the  $C$ -terminus of the VS to the  $N$ -terminus of the HCS of the following unit. Each LRR domain contains a concave surface, a convex surface, an ascending surface, and a descending surface on the opposite side. LRR domains are involved in direct interaction with proteins (including hormones) or ligands (including nucleic acid, lipid, lipo-polysaccharide, and plant steroid hormones) [1]. LRR domains can engage structurally various proteins or ligands using different surfaces of the LRR domains [1, 5].

LRR proteins participate in the plant immune response and in the mammalian innate immune response [1–6]. They are also involved in a broad range of functions including apoptosis, autophagy, ubiquitin related processes, nuclear mRNA transport, and neuronal development [1, 8]. Plant LRR proteins, many of which involve kinases and other receptor like proteins, act as signal amplifiers in tissue damage, in symbiotic relationships, and in developmental processes [1, 9].

Furthermore, LRR proteins are contained in the type III secretion system of many gram-negative bacterial pathogens. The LRR proteins called effectors are delivered into the cytosol of animal or plant cells [10]. Consequently, these effectors enable the bacteria to avoid the immune response of the infected organism by modulating cell functions of the host. The effector proteins include YopM from the bubonic plague bacterium, *Yersinia pestis*, and SspH1, SspH2, and SlrP from *Salmonella enterica*, and IpaH3 and IpaH9.8 from *Shigella flexneri*. These effectors are bacterial LRR proteins [1]. The LRR domain of SspH1 directly interacts with PKN1 [11].

Bacterial LRRs are characterized by two Leu Pro sequences in the VS; the consensus is  $xxLPxLPxx$  with Nine residues (N-subtype) and  $xxLPxxLPxx$  with Ten residues (T-subtype) where “L” is Leu, Val, or Ile and “x” is predominantly occupied by small residues such as Thr, Ser, or Gly. T-subtype is seen in *Salmonella* SlrP. Moreover, LRRs in the subfamily of toll-like receptors (TLR7, TLR8, and TLR9), the small leucine rich repeat proteoglycan (SLRP) family including fibromodulin, decorin and biglycan, and the fibronectin leucine rich repeat transmembrane family (FLRT) contains N- or T-subtype [1]. The LRRs consist of tandem repeats of a super motif of  $STT$  or  $ST$  in which “S” is Bacterial and “T” is Typical [12–14]. We called this the  $STT$  class [1].

PPIIs are known to be observed frequently in proline rich regions [15–19]. The PPIIs are characterized by the

backbone dihedral angles ( $\Phi$ ,  $\Psi$ ) of ( $-75^\circ$ ,  $145^\circ$ ) [20–29]. The PPIIs have helical parameters: 2.9 residues per turn, a pitch value of 8.7 Å per turn, and a helix radius of 1.33 Å. The assignment of PPII is not done in the widely used programs such as DSSP [28] and STRIDE [29]. Consequently, PPIIs in newly solved protein structures are not registered in protein data bank (PDB) [30]. Now there are some tools for assigning PPII number—DSSP-PPII [31], PROSS [32], SEGNO [33], XTLSSTR [34], and ASSP [35].

Super secondary structures with several adjacent elements of a secondary structure are also observed in protein structures [36, 37]. Examples include  $\beta$ -hairpins,  $\alpha$ -helix hairpins, and  $\beta$ - $\alpha$ - $\beta$  motifs. Adzhubei and Sternburg [38] identified super secondary structures consisting of PPII and  $\alpha$ -helix and of  $3_{10}$ -helix and PPII. Kumar and Bansal [27] also identified those consisting of  $\beta$ -strand and PPII, of  $\beta$ -strand, PPII, and  $\alpha$ -helix, and of  $\beta$ -strand, PPII, and  $\beta$ -strand.

Evdokimov et al. [39] noted that the VS parts in the YopM LRRs adopt  $3_{10}$ -helices. Matsushima et al. [6, 7], Bella et al. [5], and Park et al. [40] proposed that the VS parts in the bacterial LRR adopt left handed polyproline II helices (PPII). A review article by Adzhubei et al. [41] noted PPIIs in LRRs. However, it appears that PPII in LRR structures has not yet been well characterized based on the consensus sequence. Structural data of proteins containing bacterial LRRs have increased. The crystal structures of YopM, SspH1, SspH2, SlrP, IpaH3, and IpaH9.8 have been determined [39, 42–46]. The structures of TLR8, TLR9, fibromodulin, decorin, biglycan, FLRT2, and FLRT3 are also available [47–56].

The purpose of this study is to understand structural features of bacterial LRRs. We performed both the secondary structures analyses using secondary structures prediction programs of DSSP-PPII, PROSS, SEGNO, and XTLSSTR and the HELFIT analyses that calculate helix axis, helix pitch, helix radius, repeat/residue number per turn, based on the atomic coordinates of the crystal structures [57].

This present analysis demonstrates that the N-subtype VS adopts PPII consisting of 4–6 residues and type I  $\beta$ -turn at the C-terminal side. Thus, the VS part is characterized by super secondary structure consisting of PPII and a  $\beta$ -turn. In contrast, the T-subtype VS frequently prefers two separate PPIIs consisting of two or three and of two residues. The HELFIT analysis indicates that the type I  $\beta$ -turn is a right handed helix and consequently determines three unit vectors of the helix axes of PPII (**P**),  $\beta$ -turn (**B**), and LRR domain (**A**). We propose three structural parameters which are two angles between the two helix axes of PPII and  $\beta$ -turn, between the two helix axes of PPII and LRR domain, and between the helix axis of LRR domain and the vector product of **P**  $\times$  **B**. These three angles are suggested to characterize the super secondary structure and the LRR domain.

## 2 Methods

### 2.1 Structure Data

We collected the structure data of proteins containing bacterial LRRs from the PDB. We performed sequence alignments in LRR proteins from the PDB by LRRpred that recognizes and aligns LRR motifs that predict the repeat number and “phasing” of LRRs with greater reliability [58] and identified bacterial LRR based on the consensus sequence. Bacteria LRR proteins are YopM, SspH1, SspH2, SlrP, IpaH3, IpaH9.8, TLR8, TLR9, fibromodulin, decorin, biglycan, FLRT2, and FLRT3 (Table 1) [39, 42–56]. LRRs in YopM, SspH1, SspH2, SlrP, IpaH3, and IpaH9.8 belong to bacterial LRR class, while TLR8, TLR9, fibromodulin, decorin, biglycan, FLRT2, and FLRT3 belong to the *STT* class (Table 1). Eighteen PDB files solved at resolution  $\leq 3.4$  Å were used; the sequence identity of the 18 different chains shows that the maximum is 48% and the average is 7% (Supplementary Table S1). The structure data of mouse FLRT2 at resolution 4.0 and 6.0 Å were not used for analyses.

### 2.2 Secondary Structures Analysis

Secondary structures assignments were made from the atomic coordinates of the LRR structures using four programs—DSSP-PPII [31], PROSS [32], SEGNO [33], and XTLSSTR [34]. The assignment of the DSSP program is based on the identification of precise hydrogen bond patterns corresponding to regular secondary structures [28]. In DSSP-PPII based on DSSP, PPII are assigned solely in the coil region for at least two consecutive residues in coil with  $\Phi = -75^\circ \pm 29^\circ$  and  $\Psi = +145^\circ \pm 29^\circ$ . The PROSS program assigns secondary structures, based mainly on  $\Phi$  and  $\Psi$  dihedral angles. SEGNO utilizes the  $\Phi$  and  $\Psi$  dihedral angles coupled with other angles. XTLSSTR uses two angles and three distances. DSSP-PPII, PROSS, and XTLSSTR assign  $\beta$ -turns; while SEGNO does not. The secondary assignments were performed using the PolyprOnline web interface [59]. Types of  $\beta$ -turn were also identified by the programs of PROMOTIF [60] and STRIDE [61]. Furthermore, the root-mean-square deviation (RMSD) of the VS part between within the N-subtype and within the T-subtype and between these two subtypes using the coordinates of the backbone atoms of each residue were evaluated by the CHIMERA program [62].

**Table 1** Known structures of proteins containing bacterial LRRs

Number of protein	LRR class	Protein name	Repeat number of LRR <sup>a</sup>	Number of N-subtype	Number of T-subtype	PDBID	Chains	Resolution (Å)
1	Bacterial	<i>Y. pestis</i> yopM	16 (16)	13	0	1JL5	A	2.10
	Bacterial	<i>Y. pestis</i> yopM	16 (16)	13	0	1G9U	A	2.35
2	Bacterial	<i>Y. enterocolitica</i> yopM	21 (20)	19	0	4OW2	A,B,C,D	3.20
3	Bacterial	<i>S. enterica</i> SspH1	10 (9)	6	1	4NKH	A,B,C,D,E,F	2.75
	Bacterial	<i>S. enterica</i> SspH1	10 (9)	6	1	4NKG	A,C	2.90
4	Bacterial	<i>S. enterica</i> SspH2	13 (13)	10	1	3G06	A	1.90
5	Bacterial	<i>S. flexneri</i> ipaH3	9 (9)	5	1	3CVR	A	2.80
6	Bacterial	<i>S. flexneri</i> ipa9.8	8 (8)	5	1	5B0N	A,B	1.80
	Bacterial	<i>S. flexneri</i> ipa9.8	8 (8)	5	1	5B0T	A	2.00
7	Bacterial	<i>S. enterica</i> SlrP	12 (12)	0	10	4PUF	A,B	3.30
8	STT	Human fibromodulin	13 (12)	2	2	5MX0	A,B	2.21
9	STT	Horse TLR9	27 (27)	1	1	3WPC	A,B	1.60
	STT	Horse TLR9	27 (27)	1	1	3WPB	A	2.40
	STT	Horse TLR9	27 (27)	1	1	3WPD	A	2.75
10	STT	Bovine TLR9	27 (27)	1	1	3WPE	A	2.38
11	STT	Mouse TLR9	27 (27)	1	1	3WPF	A	1.96
	STT	Mouse TLR9	27 (27)	1	1	3WPG	A	2.25
	STT	Mouse TLR9	27 (27)	1	1	3WPI	A	2.25
	STT	Mouse TLR9	27 (27)	1	1	3WPH	A	2.35
12	STT	Human TLR8	27 (27)	0	3	3WN4	A	1.81
	STT	Human TLR8	27 (27)	0	3	3W3J	A,B	2.00
	STT	Human TLR8	27 (27)	0	3	3W3N	A,B	2.10
	STT	Human TLR8	27 (27)	0	3	3W3G	A,B	2.30
	STT	Human TLR8	27 (27)	0	3	3W3K	A,B	2.30
	STT	Human TLR8	27 (27)	0	3	3W3L	A,B,C,D	2.33
13	STT	Human TLR8	27 (27)	0	3	3W3M	A	2.70
	STT	Bovine decorin	12 (12)	0	3	1XKU	A	2.15
	STT	Bovine decorin	12 (12)	0	3	1XEC	A,B	2.30
14	STT	Bovine decorin	12 (12)	0	3	1XCD	A	2.31
	STT	Bovine biglycan	12 (12)	0	3	2FT3	A,B,C,D,E,F	3.40
	STT	Human TLRT2	13 (13)	0	3	4V2D	A	2.50
15	STT	Human TLRT2	13 (13)	0	3	5FTT	B,F	3.40
16	STT	Mouse TLRT2	13 (13)	0	3	5CMP	A,B,C,D	2.60
17	STT	Human FLRT3	13 (13)	0	3	4V2E	A,B	2.50
	STT	Mouse FLRT3	12 (12)	0	3	2YEB	B,F	3.19

<sup>a</sup>The number in the parentheses indicates the number of variable segment of LRRs

### 2.3 HELFIT Analysis

We have developed a total least squares program for fitting a helix to data points—HELFIT [57]. A helix consisting of  $n$  repeat units may be characterized by helix axis, helix pitch ( $P$ ), helix radius ( $R$ ), and number of repeat units/residue per turn ( $N$ ). HELFIT computes these parameters in which the helix axis is represented by the unit vector. These parameters also yield the rise per repeat unit/residue ( $\Delta z = P/N$ ) and the rotation per repeat unit/residue in

the helix ( $\Delta\Phi = 360^\circ/N$ ). Moreover, HELFIT gives  $rmsd$ : where  $d_i$  is the closest distance from the data point to the trace of the helix.

$$rmsd = \left[ \left( \text{the minimum of } \sum d_i \right) / N \right]^{1/2} \quad (1)$$

Here  $p = rmsd / (n - 1)^{1/2}$  gives the regularity of helical structures independent of the number of data points or helix length. The criterion for regular PPII helices is  $p \leq 0.10 \text{ \AA}$ . This same test is used for  $\alpha$ -helices,  $\omega$ -helices,

and  $3_{10}$ -helices in proteins [63, 64]. The HELFIT analysis requires only four data points: the coordinates of  $\alpha$ -carbon ( $C_\alpha$ ) of each residue. LRRs form a  $\beta$ -strand of three residues at positions 3–5 in the HCS part. Thus, in LRRs, the  $C_\alpha$  coordinates of the consensus leucine residue at position four in individual LRR repeat units are used. The repeat number of individual LRR domains was defined as the number that participates in the parallel  $\beta$ -sheet. This definition means that the first LRR is sometimes contained in the capping structures.  $\beta$ -Turns consist of four amino acid residues (labelled  $i$ ,  $i + 1$ ,  $i + 2$ , and  $i + 3$ ). We also estimate the helical parameters of  $\beta$ -turns using the  $C_\alpha$  coordinates of each residue.

The HELFIT analysis indicates that the  $\beta$ -turn is regarded as a right handed helix, as noted later. Consequently, HELFIT determines three unit vectors of the helix axes of LRR domain (**A**), PPII (**P**), and type I  $\beta$ -turn (**B**). We estimate three structural parameters. One is the angle between the two helix axes of PPII and  $\beta$ -turn ( $\Omega_1$ ). The second is the angle between the two helix axes of PPII and LRR domain ( $\Omega_2$ ). The third is the angle between the helix axis of LRR domain and the vector product of  $\mathbf{P} \times \mathbf{B}$ . ( $\Omega_3$ ). The three angles of  $\Omega_1$ ,  $\Omega_2$ , and  $\Omega_3$  are represented by the following equations.

$$\mathbf{P} \cdot \mathbf{B} = |\mathbf{P}||\mathbf{B}| \cos \Omega_1 \quad (2)$$

$$\mathbf{P} \cdot \mathbf{A} = |\mathbf{P}||\mathbf{A}| \cos \Omega_2 \quad (3)$$

$$(\mathbf{P} \times \mathbf{B}) \cdot \mathbf{A} = |\mathbf{P} \times \mathbf{B}||\mathbf{A}| \cos \Omega_3 \quad (4)$$

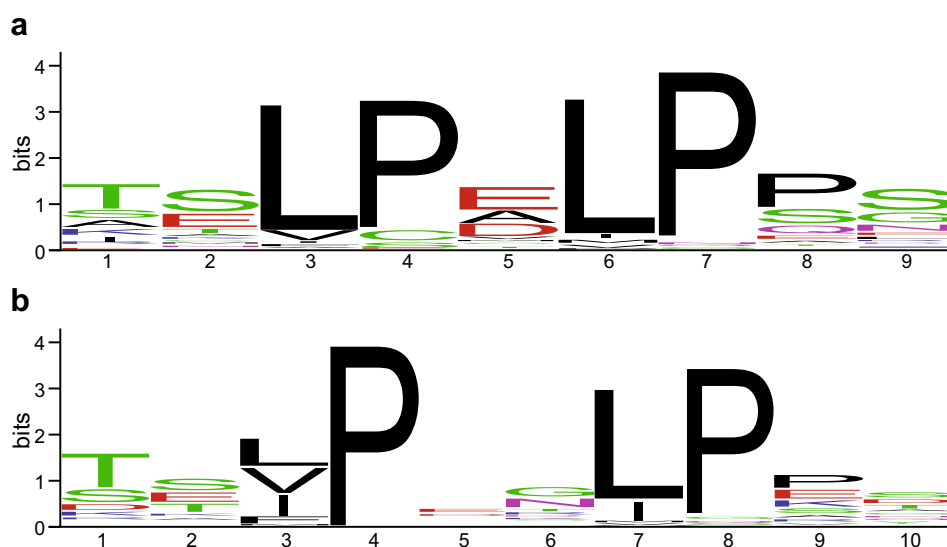
## 3 Results

### 3.1 Two Subtypes of Consensus Sequences

Bacterial LRR is 20 or 21 residues long and is classified into two subtypes. The N-subtype has the VS consensus of xxLPxLPxx with nine residues and the T-subtype has xxLPxxLPxx with ten residues where “L” is Leu, Val, or Ile and “x” is predominantly occupied by small residues such as Thr, Ser, or Gly (Fig. 1). Thus bacterial LRR is characterized by two Leu Pro sequences in the VS parts; although variable VS that lack one of the two conserved prolines is observed. The characteristics are not seen in other LRR classes. For examples, ribonuclease inhibitors and Nod-like receptors contain RI-like LRRs which of the consensus is LxxLxLxx(N/C)xLxxxgoxxLxxoLxxzxxx with typically 28 or 29 residues [65].

The N-subtype appears sixty-three times in the known structures (Table 1). The VS consensus is xxLPxLPxx, as expected, in which “x” positions at the N- and C-terminal sides are frequently occupied by relatively small residues such as Thr or Ser; while the central “x” position is rich in Glu. Fifty-two of the sixty-three VS are completely consistent with the consensus (Fig. 1a). The remaining VSs are xxLxxLPxx in YopM; the conserved Pro at position four is replaced by Cys or Ser. The T-subtype appears forty-three times (Table 1). The conserved Leu at position three in the VS consensus is frequently occupied by other hydrophobic residues such as Val or Ile (Fig. 1b).

**Fig. 1** Consensus sequence of variable segment (VS) of bacterial LRRs in the known structures. **a** N-subtype; **b** T-subtype. The graphical sequence diagrams were generated with WebLogo [78], representing sixty-three LRR units for the N-subtype and forty-three for the T-subtype



### 3.2 Secondary Structures

The assignment of PPII patterns differs among the four programs. Bacterial LRR proteins form not only monomers but also homo-dimers, -tetramers, and -hexamers in crystals (Table 1). The PPII patterns assigned also differ among their individual molecules. We therefore analyzed all chains of the known structures.

The four programs for secondary structures assignment indicate that the HCS parts adopt short  $\beta$ -strands in three

underlined residues of LxxLxLxxNxL. In addition, the assignment indicates that the VS parts are rich in PPII conformations. However, the secondary structures show a difference between the two subtypes.

At least one of the four programs of secondary structures assignment indicates that the N-subtype VS adopts PPIIs consisting of four, five or six residues (Fig. 2a; Supplementary Table S2). Four, five or six residue PPIIs are observed in the underlined residues of xxLPxLPxx, xxLPxLPxx, and xxLPxLPxx, respectively. For example, all four

**a**

**N-subtype**

Protein	SspH2 <i>LRR2</i>	SspH2 <i>LRR3</i>	SspH2 <i>LRR4</i>	SspH2 <i>LRR5</i>	SspH2 <i>LRR6</i>
VS sequence	<b>TSLPALPPE</b>	<b>TSLPVLPPG</b>	<b>THLPALPSG</b>	<b>TSLPVLPPG</b>	<b>ASLPALPSE</b>
DSSP-PPII	S--PPPPTT	S----PPTT	----PPPTT	S--PPPPTT	S--PPPPTT
PROSS	---PPPPTT	---P-PPTT	-EEEEPPPTT	---PPPPTT	---PPPPTT
SEGNO	--PPPPp--	--PPPPp--	--PPPPp--	--PPPPp--	--PPPPp--
XTLSSTR	EEePPPPTT	EEePPPPTT	EEePPPPTT	EEePPPPTT	EEePPPPTT

Protein	SspH2 <i>LRR7</i>	SspH2 <i>LRR8</i>	SspH2 <i>LRR9</i>	SspH2 <i>LRR10</i>	SspH2 <i>LRR11</i>
VS sequence	<b>TSLPMLPSG</b>	<b>ASLPTLPSE</b>	<b>TSLPALPSG</b>	<b>TSLPVLPESE</b>	<b>TSLPMLPSG</b>
DSSP-PPII	S--PPPPTT	S---PPPTT	SS-PPPPTT	S---PPPTT	S---PPPTT
PROSS	---PPPPTT	----PPPTT	---PPPPTT	----PPPTT	----PPPTT
SEGNO	--PPPPp--	--PPPPp--	-pPPPPp--	--PPPPp--	--PPPPp--
XTLSSTR	EEePPPPTT	EEePPPPTT	EEePPPPTT	---PPPPTT	EEePPPPTT

**b**

**T-subtype**

Protein	SspH2 <i>LRR1</i>	Bovine TLR9 <i>LRR7</i>	Human TLR8 <i>LRR1</i>	Human TLR8 <i>LRR4</i>	Human TLR8 <i>LRR7</i>
VS sequence	<b>TTLPDCLPAH</b>	<b>TTVPRSLPPS</b>	<b>QEVPTQTVGKY</b>	<b>PQIPSGLPES</b>	<b>SHVPPKLPSS</b>
DSSP-PPII	S---S---TT	SSPPSSPPTT	SSPPS-PPTT	SSPPTTPPTT	SSPPS-PPTT
PROSS	---P---PTT	--PPP-PPTT	--PPTTPPTT	--PPTTPPTT	--PPTTPPTT
SEGNO	-----pPp--	-pPPP-----	-----	--Pp-----	-pPPP-----
XTLSSTR	EEe-EEeNNN	--PPPPPTT	-----PpTT	---Pp-PpTT	EePPPPPTT

**Fig. 2** Secondary structure assignment of the variable segment (VS) of representative bacterial LRRs by the four programs (DSSP-PPII, PROSS, SEGNO, and XTLSSTR). **a** N-subtype, SspH2 (PDB ID: 3G06\_A). **b** T-subtype, SspH2 (*LRR1*) (PDB ID: 3G06\_A), Bovine

TLR9 (*LRR7*) (PDB ID:3WPC\_A) and Human TLR8 (*LRR1*, *LRR4*, and *LRR7*) (PDB ID: 3WN4\_A). A one letter code is used to represent a specific conformation; P and p, PPII; E and e,  $\beta$ -strand; T,  $\beta$ -turn; N, non-hydrogen-bonded  $\beta$ -turn; H,  $\alpha$ -helix; and S, bend



program assign PPII in underlined residues of **KKLPDL-PLS** (*LRR7*) in *Y. pestis* YopM ( $n = 12$ ) [3G06\_A] (Fig. 2a). The variable VS of **xxLxxxLPxx** frequently adopt four or five PPIIs. The HELFIT analysis demonstrated that all of the PPIIs assigned are definitely left handed polyproline helices, as noted later.

In addition, the programs identified  $\beta$ -turns at the C-terminal side in the VS parts. The sequences in the underlined residues of **xxLPxLPxxLxxLxLxxNxL** adopt  $\beta$ -turns; the second conserved Pro corresponds to the residue,  $i$ , of  $\beta$ -turns. The types are distinguished by the  $\Phi$ ,  $\Psi$ , angles of residues  $i + 1$  and  $i + 2$ . The average  $C\alpha(i) - C\alpha(i + 3)$  is 5.48 (0.22 Å); the numbers in parenthesis are standard deviations; they are reasonable [66]. The  $\Phi$ ,  $\Psi$  angles of residues  $i + 1$  and  $i + 2$  of the  $\beta$ -turns have the average angles of  $\Phi_{i+1} = -60.7$  (8.8°),  $\Psi_{i+1} = -22.6$  (8.4°),  $\Phi_{i+2} = -94.0$  (13.8°),  $\Psi_{i+2} = 3.4$  (14.0°); the numbers in parentheses are standard deviations. These values are close to  $-60^\circ$ ,  $-30^\circ$  and  $-90^\circ$ ,  $0^\circ$  which define the type I  $\beta$ -turns. The  $\beta$ -turns assignments by PROMOTIF [60] and STRIDE [61] give the same results and indicate that most of the  $\beta$ -turns are type I (Table 2). Types IV and VIII rarely appear. Also very rarely  $\beta$ -turns are not assigned.

In the T-subtype VS PPIIs assigned may be divided into three patterns. Many VSs adopt two separate PPIIs with two or three and two residues, which are observed in the underlined residues of **xxLPxxLPxx** and **xxLPxxLPxx** (Fig. 2b; Supplementary Table S2). The second is one PPII with three or four residues in the underlined residues of **xxLPxxLPxx**, **xxLPxxLPxx** or **xxLPxxLPxx**. The third pattern is six residue PPII in the underlined residues of **xxLPxxLPxx**, which are seen in decorin, TLR9, and TLR8 by the XLTSSSTR program. Moreover, the C-terminal two residues of the T-subtype VS are assigned to adopt mostly type I  $\beta$ -turn as does that of the N-subtype VS.

The average RMSD of the VS parts of the N-subtypes and of the T-subtypes is 0.589 (0.315 Å) and 0.993 (0.337 Å), respectively; all Bacterial VSs show the RMSD of 1.087 (0.549 Å).

In conclusion the N-subtype VS is characterized by a super secondary structure consisting of PPII with four, five, or six residues and a type I  $\beta$ -turn (Fig. 3), while the T-subtype VS strongly prefers one or two separate PPIIs and adopts a type I  $\beta$ -turn at the C-terminal side as does the

N-subtype VS. It appears that the structure of the T-subtype is more variable.

### 3.3 Helical Parameters of PPIIs, Type I $\beta$ -turns, and LRR Domains

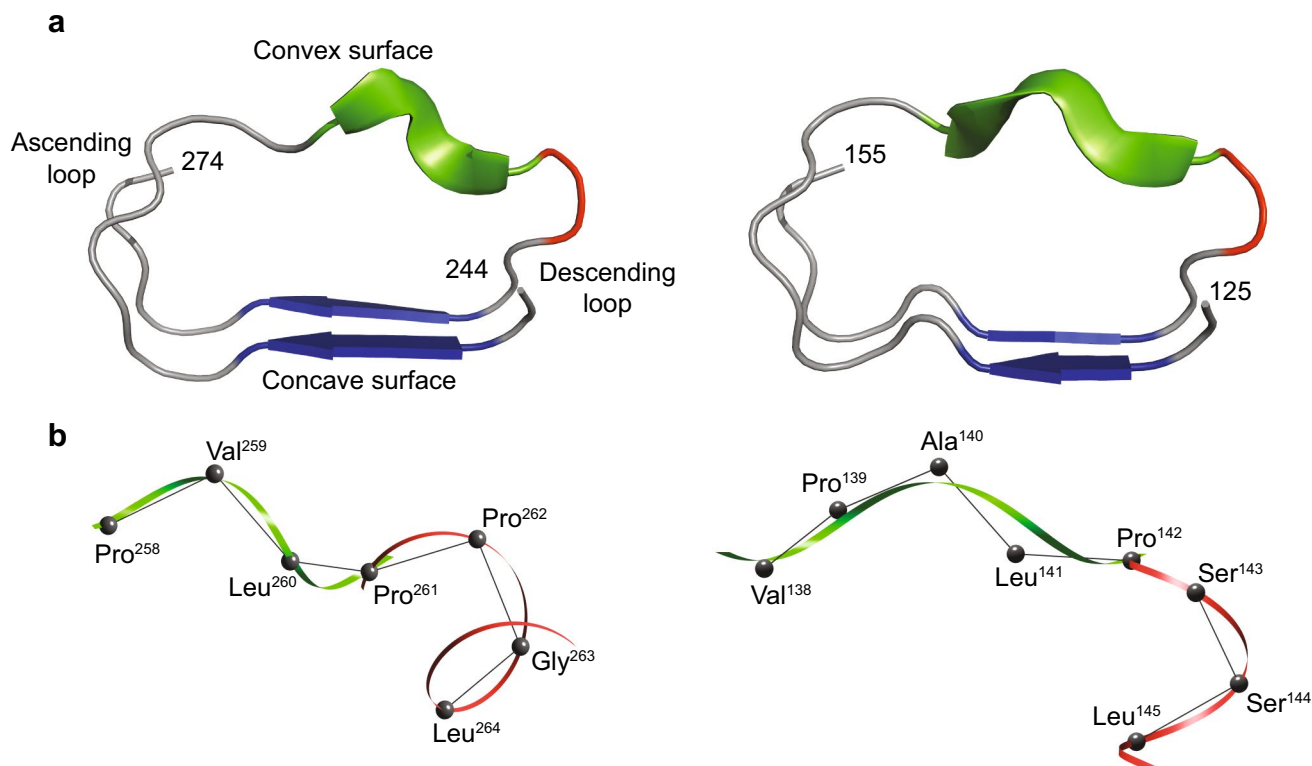
The number of assigned PPII of which the helix length is longer than three residues increase in order of DSSP-PPII < PROSS < XLTSSSTR < SEGNO (Table 3). SEGNO assigns longer PPIIs. Four residue PPIIs are regular and thus are a near ideal form (Table 2). All of five and six residue PPIIs are irregular. The helix regularity decreases with increasing helix length. The deviation of helix parameters from ideal values increases with increasing helix length. The helix irregularities in the five and six residue PPIIs are mostly due to larger deviations of the  $\Phi$ ,  $\Psi$  angles from ideal:  $-75^\circ$ ,  $145^\circ$  in the first and/or second residues of N-terminal side in the sequence of **xLPxLP**. The sequence of VP(A/R)LP in TLR9 (*LRR4*), which corresponds to the underlined residues of **xxLPxLPxx**, adopt five residue PPIIs with large helix irregularity;  $p = 0.41 - 0.47$  Å. In this case, the irregularity comes from large deviation of the  $\Phi$ ,  $\Psi$  angles of the conserved Leu at the C-terminal side. Taking account of the helical parameters, the five residue PPIIs are regarded as a highly deformed form. Consequently, the HELFIT analysis demonstrated that all four, five, and six residue PPIIs assigned by the secondary structures analyses are regular or irregular left handed polyproline helices.

In the T-subtype only the XLTSSSTR program identifies long PPIIs with six residues in *LRR7* of horse/bovine TLR9, *LRR4* of human TLR8, and *LRR4* of decorin, as noted. The  $p$  values are very large;  $p = 0.38 - 0.50$  Å (Table 3). The high irregularity comes from the  $\Phi$ ,  $\Psi$  angle of any residue at position six in the T-subtype VS consensus of **xxLPxxLPxx**; the  $\Phi$ ,  $\Psi$  angles are in regions of a left handed  $\alpha$ -helix. In addition, the helical parameters deviate highly from those of ideal form (Table 3). The six residue PPIIs assigned are not recognized as a PPII. Alternatively, the SEGNO program identifies regular, four residue PPIIs in the underlined residues of **xxLPxxLPxx** (Supplementary Table S2). Thus, the six residue PPIIs may be divided into two separate PPIIs with three and two residues. This supports the conclusion by the secondary structure assignment that the T-subtype VS contains two separate PPIIs.

The average helix parameters of the type I  $\beta$ -turns are;  $P = 6.11 \rightarrow 6.14$  Å,  $N = 3.65 \rightarrow 3.68$  residues/turn,  $R = 2.26 \rightarrow 2.28$  Å, and  $\Delta z = 1.68 \rightarrow 1.69$  Å (Table 4). It appears that the helical parameters are close to those of  $\alpha$ -helix as it has  $P = 5.4$  Å,  $N = 3.6$  residues/turn,  $R = 2.4$  Å, and  $\Delta z = 1.5$  Å. The average  $p$  value is  $0.02 \rightarrow 0.03$  Å. The HELFIT analysis

**Table 2** Types of  $\beta$ -turns at the C-terminal side in the bacterial LRRs

	DSSP-II	PROSS	XLTSSSTR
Type I	66	87	137
Type IV	5	5	5
Type VIII	1	0	1
Total	72	92	143



**Fig. 3** Super secondary structure consisting of a PPII and a  $\beta$ -turn in bacterial LRRs. **a** Secondary structures. Left panel. Sequence 244-LRTLEVSGNQLTSLPVLPPGLLELSIFSNPL-274 in SspH2 (*LRR3* and HCV of *LRR4*) (PDB ID: 3G06\_A). Right panel. 125-LEELNLSYNGITTPALPSSLVSLILSRNTNI-155 in bovine TLR9 (*LRR4* and HCS of *LRR5*) (PDB ID: 3WPC\_A). Blue arrows represent  $\beta$ -strands, green ribbons PPIIs, and red tubes  $\beta$ -turns. **b** Super secondary structure (HELFIT). Left panel. Sequence 258-PVLPPGL-264 in SspH2 (PDB ID: 3G06\_A). The sequence PVLPP is a part of the *LRR3* VS which correspond to the underlined residues of xxLPxLPxx (the consensus of the N-type VS). The

sequence PVLPP adopt PPII and the PPGL adopt type I  $\beta$ -turn. The sequence PVLPP is a part of the *LRR3* VS which correspond to the underlined residues of xxLPxLPxx (the consensus of the N-type VS). The sequence PVLPP adopts four residue PPII and the PPGL adopts type I  $\beta$ -turn. Right panel. Sequence 138-VPALPSSL-145 in bovine TLR9 (PDB ID: 3WPC\_A). The sequence VPALP is a part of the *LRR4* VS which correspond to the underlined residues of xxLPxLPxx. The sequence VPALP forms five residue PPII and the sequence PPSL forms one type I  $\beta$ -turn. Best fitted lines by HELFIT are colored green for PPIIs and red for  $\beta$ -turns, and  $\alpha$ -carbons grey. (Color figure online)

indicates that these type I  $\beta$ -turns form a regular, right handed helix.

The helix parameters of LRR domains were determined for IpaH9.8 ( $n=8$ ), SspH1, ( $n=10$ ), SspH2 ( $n=13$ ), *Y. pestis* YopM ( $n=16$ ), *Y. enterocolitica* YopM ( $n=21$ ), and SlrP ( $n=12$ ). The bacterial LRR domains are represented by a right handed helix (Table 5). The helix parameters range over:  $P=47.3 \rightarrow 115$  Å,  $N=28.4 \rightarrow 41.3$  units/turn,  $R=18.9 \rightarrow 24.6$  Å,  $\Delta z=1.67 \rightarrow 3.46$  Å, and  $\Delta\Phi=8.7^\circ \rightarrow 12.7^\circ$ ;  $p=0.03 \rightarrow 0.19$  Å.

Figure 4 shows a plot of  $2 \cdot R \cdot \sin(\Delta\Phi/2)$  versus  $\Delta z$ . The values fall on a circle with radius  $D$  (circle plot) [7].  $D$  is a function of  $\Delta z$ ,  $\Delta\Phi$ , and  $R$  [7].

$$D = \left[ \left\{ 2R \sin \left( \frac{\Delta\Phi}{2} \right) \right\}^2 + (\Delta z)^2 \right]^{1/2} \quad (5)$$

$D$  is the average  $C_\alpha(i) - C_\alpha(i+1)$  distance between adjacent repeats— $i$  and  $i+1$ ;  $D$  corresponds to the inter-strand distance. Equation 5 gives  $D=4.97 \pm 0.10$  Å; this allows the formation of hydrogen bonds between parallel strands. This circle plot shows that the helix pitch,  $P$ , and rise per turn,  $\Delta z$ , of bacterial LRR is comparable to those of SDS22-like and Plant specific LRRs; while it is larger than those of RI-like and CC LRRs [7].

### 3.4 Geometrical Analysis

Figure 5 shows the frequency distributions of three angles of  $\Omega_1$ ,  $\Omega_2$ , and  $\Omega_3$ . The  $\Omega_1$  angle shows an asymmetrical distribution (Fig. 5b). The  $\Omega_1$  angle ranges from  $70^\circ$  to  $120^\circ$ ; the average value is  $\sim 103^\circ$  (Table 4). The  $\Omega_2$  and  $\Omega_3$  have the average values of  $\sim 33^\circ$  and  $\sim 99^\circ$ , respectively.



**Table 3** Helix parameters of PPIIs in bacterial LRRs

	Programs	Number of PPII	$P$ (Å) <sup>a</sup>	$N$ <sup>a</sup>	$R$ (Å) <sup>a</sup>	$\Delta z$ (Å) <sup>a</sup>	$p$ (Å)	$V_c$ (Å <sup>3</sup> ) <sup>a</sup>
PPII <sup>b</sup>	–	–	8.96	2.99	1.36	3.00	–	17.41
PPII <sup>c</sup>	–	–	8.69	2.90	1.33	3.00	–	16.70
PPII <sup>d</sup>	–	–	8.58	2.99	1.45	2.87	–	18.95
PPII <sup>e</sup>	–	–	8.62	3.08	1.52	2.80	–	20.30
Four residue PPII	DSSP-PPII	50	8.61 (0.40)	2.82 (0.13)	1.25 (0.05)	3.06 (0.05)	0.10 (0.03)	15.07 (1.07)
	PROSS	67	8.52 (0.51)	2.78 (0.21)	1.23 (0.09)	3.07 (0.08)	0.10 (0.04)	14.70 (1.74)
	SEGNO	21	8.78 (0.31)	3.05 (0.10)	1.47 (0.05)	2.87 (0.05)	0.05 (0.04)	19.49 (0.05)
	XLTSSTR	130	8.45 (0.38)	2.75 (0.15)	1.23 (0.07)	3.07 (0.07)	0.10 (0.04)	14.49 (1.49)
Five residue PPII	DSSP-PPII	49	9.83 (0.50)	3.54 (0.25)	1.53 (0.11)	2.78 (0.10)	0.26 (0.04)	20.50 (2.39)
	PROSS	32	9.94 (0.43)	3.64 (0.16)	1.58 (0.06)	2.73 (0.05)	0.27 (0.03)	21.46 (1.39)
	SEGNO	132	10.05 (0.37)	3.65 (0.15)	1.57 (0.07)	2.75 (0.07)	0.27 (0.04)	21.38 (1.62)
	XLTSSTR	43	10.17 (1.22)	3.54 (0.39)	1.47 (0.16)	2.88 (0.18)	0.29 (0.09)	19.39 (3.10)
Six residue PPII	SEGNO	21	10.45 (1.21)	3.82 (0.50)	1.50 (0.09)	2.74 (0.07)	0.30 (0.04)	19.56 (2.18)
	XLTSSTR	14	13.91 (1.64)	6.10 (0.71)	2.40 (0.09)	2.28 (0.05)	0.45 (0.04)	41.36 (2.92)

The number in the parentheses indicates standard deviations

<sup>a</sup> $P$  helix pitch,  $N$  residue number per turn,  $R$  helix radius,  $\Delta z$  helix rise per turn;  $V_c = \pi R^2(\Delta z)$

<sup>b</sup> $(\Phi, \Psi) = (-75, 150)$  by Jha [67]

<sup>c</sup> $(\Phi, \Psi) = (-75, 145)$  by Hopfinger [25]

<sup>d</sup> $(\Phi, \Psi) = (-65, 140)$  by Adzhubei [68]

<sup>e</sup> $(\Phi, \Psi) = (-60, 140)$  by Schulz and Schirmer [26]

## 4 Discussion

### 4.1 Structural Role of PPIIs in Bacterial LRRs

The backbone dihedral angles ( $\Phi, \Psi$ ) of ideal PPII is  $(-75^\circ, 145^\circ)$  [23]. Other dihedral angles has been also proposed;  $(\Phi, \Psi) = (-75^\circ, 140^\circ)$ ,  $(-65^\circ, 145^\circ)$ , and  $(-60^\circ, 140^\circ)$  (Table 3) [26, 67, 68]. It appears that four residue PPIIs in proteins are a near ideal form with  $(\Phi, \Psi) = (-75^\circ, 145^\circ)$  or  $(-75^\circ, 140^\circ)$ .

The secondary structure assignment and the HELFIT analysis indicate that the N-subtype VS adopts one stretch of PPII of four, five, or six residues. In contrast, the T-subtype prefers two separate PPIIs consisting two or three and of two residues. Consequently, there is a clear difference in the PPII patterns between the two subtypes. The two hydrophobic residues in the VS part of the two subtypes are concentrated on the side that is oriented toward the hydrophobic core as well as other conserved hydrophobic residues in the HCS part (Fig. 6). This structural restriction makes the difference. The assignments by the four programs show different PPII patterns in most cases. This observation mainly comes from high flexibility of PPII due to no intrachain hydrogen bond.

The variable N-type VSs are seen in YopM—RSLCDLPPS and SGLSELPPN (Supplementary Table S2). The first repeat of twelve LRRs in chondroadherin is also

a variable N-subtype VS of QKIPKYSEK; the structure of human chondroadherin which forms tetramers in crystal has been determined at 2.1 Å resolution [69]. Their secondary structure assignment sometimes shows no PPII or only short PPII of two or three residues. These observations indicate that two conserved prolines in the N-subtype VS parts are strongly required for the super secondary structure consisting of PPII and  $\beta$ -turn.

It may be significant that Src tyrosine kinases SH3 domain binds to short proline rich sequence of xPxxP that forms PPII [70]. This sequence is very similar to LPxLP in the N-subtype.

### 4.2 Super Secondary Structure

The present analyses demonstrate that the N-subtype adopts a super secondary structure consisting of a PPII and a  $\beta$ -turn. Ananthanarayanan et al. [71] first described this super secondary structure. Tandem repeats of the super secondary structure form a novel helical structure called the polyproline,  $\beta$ -turn helix [72]. This structure is observed in tandem repeats of the hepta-peptide, YSPSPSPS, in the C-terminal repeat domain (CTD) of the large subunit of RNA polymerase II (POL II) [73–75]. Many factors involved in RNA processing bind the CTD [76].

The VSs of RI-like and CC LRRs adopt an  $\alpha$ -helical conformation ( $\beta$ - $\alpha$  structural units). Typical LRR VSs

**Table 4** Helix parameters of  $\beta$ -turn and three structural parameters in bacterial LRRs

Programs	Number of secondary structure	Helical parameters of the $\beta$ -turns				Structural parameters				
		$P$ ( $\text{\AA}$ ) <sup>a</sup>	$N^b$	$R$ ( $\text{\AA}$ ) <sup>a</sup>	$\Delta z$ ( $\text{\AA}$ ) <sup>a</sup>	$p$ ( $\text{\AA}$ )	$V_c$ ( $\text{\AA}^3$ ) <sup>a</sup>	$\Omega_1$ ( $^\circ$ ) <sup>b</sup>	$\Omega_2$ ( $^\circ$ ) <sup>b</sup>	$\Omega_3$ ( $^\circ$ ) <sup>b</sup>
DSSP-PPII	72	6.14 (0.32)	3.68 (0.23)	2.28 (0.15)	1.68 (0.16)	0.03 (0.02)	16.36 (2.16)	99.8 (9.0)	30.6 (6.9)	100.6 (8.2)
PROSS	92	6.11 (0.34)	3.66 (0.22)	2.26 (0.14)	1.68 (0.15)	0.03 (0.02)	26.90 (1.65)	101.3 (10.2)	30.8 (7.1)	99.9 (7.6)
XLTSSSTR	143	6.14 (0.31)	3.65 (0.22)	2.26 (0.14)	1.69 (0.14)	0.03 (0.02)	26.93 (1.58)	103.7 (8.1)	33.4 (6.5)	99.3 (8.9)

The number in the parentheses indicates standard deviations

<sup>a</sup> $P$  helix pitch,  $N$  residue number per turn,  $R$  helix radius;  $\Delta z$  helix rise per turn;  $V_c = \pi R^2(\Delta z)$

<sup>b</sup> $\Omega_1$  is the angle between the two helices of the PPII and the  $\beta$ -turn;  $\Omega_2$  is the angle between the two helices of the PPII and the LRR domain;  $\Omega_3$  is the angle between the helix of the LRR domain and the vector product of  $\mathbf{P} \times \mathbf{B}$  in which  $\mathbf{P}$  and  $\mathbf{B}$  is the unit vectors of the helices of the PPII and the  $\beta$ -turn, respectively

prefer tandem  $\beta$ -turns. The SDS22-like LRR VSs strongly prefer  $3_{10}$ -helix ( $\beta$ - $3_{10}$ ). The Plant specific LRR structural unit is  $\beta$ - $\beta$ - $3_{10}$ . The occurrence of  $\beta$ -turns at the C-terminal sides of the VS parts are also observed in Plant specific, SDS22-like, and Typical LRRs as well as bacterial LRR [7]. Consequently, a super secondary structure consisting of  $3_{10}$ -helix and  $\beta$ -turns is present in Plant specific and SDS22-like LRRs. The unique super secondary structures consisting of  $\beta$ -turns and PPII, and of  $\beta$ -turns and  $3_{10}$ -helix should be recognized as structural elements in proteins.

Here we propose a structural parameter that characterizes the super secondary structure; the parameter is the angle between the two helix axes of PPII and type I  $\beta$ -turn ( $\Omega_1$ ). Its average value is  $\sim 103^\circ$ . The  $\Omega_1$  angle also helps to characterize other super secondary structures including those consisting of PPII and an  $\alpha$ -helix, and of PPII and a  $3_{10}$ -helix.

### 4.3 Solenoid Structure of Bacterial LRR Domains

The circle plot of bacterial LRR is comparable to those of SDS22-like and Plant specific LRR; while it differs from those of RI-like and CC LRRs (Fig. 4). We recently calculated helical parameters of 642 LRRs of known structures of 114 proteins by the HELFIT program [7]. The results indicate that the helical parameters are influenced by the structures of the ascending loops rather than of the descending loops, helical elements on the convex face, and the uniformity of parallel strand stacking on the concave face [1, 7]. The helix radius of PPII is the smallest between  $\alpha$ -helix,  $3_{10}$ -helix,  $\pi$ -helix,  $\omega$ -helix, and PPII. This partly contributes to a relatively large helix pitch for the bacterial LRR domains.

We determined the  $\Omega_2$  angle between the helix axes of PPII and of the bacterial LRR domain. The VSs of SDS22-like, Plant-specific, CC, and RI-like adopt  $3_{10}$ -helix or  $\alpha$ -helix instead of PPII. The comparison of the  $\Omega_2$  angle with the angles between their helices and LRR domains may identify fundamental features of individual LRR classes.

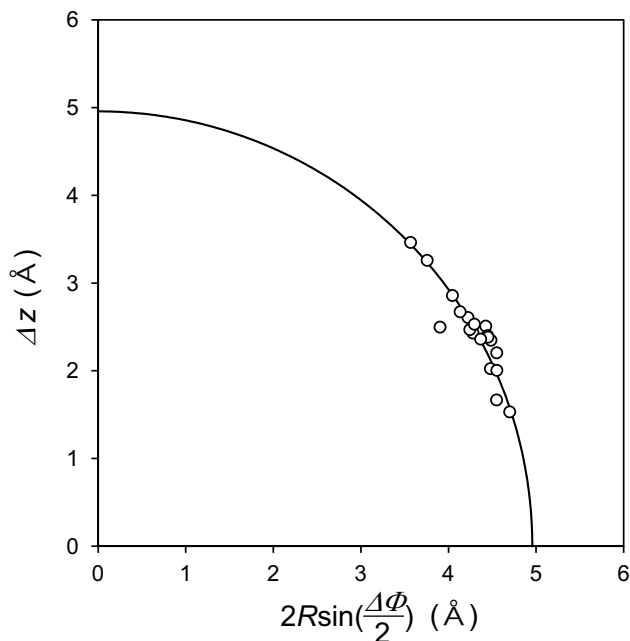
### 4.4 The PPII Assignment by the Four Programs

In many methods for assignments of secondary structures from atomic coordinates, the termini of the segments are frequently ill-defined and it is difficult to decide unambiguously which residues at the edge of the segments have to be included [77]. In this study the PPII pattern assigned also differs between the four programs in most cases. The number of four, five, or six residue PPIIs assigned is larger in SEGNO and XLTSSSTR than in DSSP-PPII and PROSS. However, the HELFIT analyses indicate that all PPIIs

**Table 5** Helix parameters of LRR domains in bacteria LRR proteins

	Protein	$n^a$	$P(\text{\AA})^a$	$N^a$	$R(\text{\AA})^a$	$\Delta z(\text{\AA})^a$	$\Delta\Phi(^{\circ})^a$	$p(\text{\AA})^a$	PDB_chanis
1	IpaH9.8	8	114.94	33.19	18.87	3.46	10.85	0.14	5B0N_A
	IpaH9.8	8	111.48	34.21	20.48	3.26	10.52	0.13	5B0N_B
2	SspH2	13	76.72	30.57	21.57	2.51	11.78	0.03	3G06
3	YopM	16	47.31	28.37	20.58	1.67	12.69	0.08	1JL5A
4	YopM	21	71.27	30.35	21.70	2.35	11.86	0.12	4OW2_A
	YopM	21	67.16	30.46	22.10	2.20	11.82	0.14	4OW2_B
	YopM	21	71.42	29.75	21.09	2.40	12.10	0.08	4OW2_C
	YopM	21	70.50	29.58	20.99	2.38	12.17	0.09	4OW2_D
5	SspH1	10	84.44	34.74	23.69	2.43	10.36	0.14	4NKH_A
	SspH1	10	88.78	35.95	24.31	2.47	10.01	0.13	4NKH_B
	SspH1	10	83.55	35.39	24.63	2.36	10.17	0.13	4NKH_C
	SspH1	10	89.67	34.39	23.14	2.61	10.47	0.16	4NKH_D
	SspH1	10	103.20	32.90	20.47	2.50	10.94	0.16	4NKH_E
	SspH1	10	83.66	41.32	29.48	2.02	8.71	0.19	4NKH_F
	SspH1	10	89.67	33.55	22.10	2.67	10.73	0.16	4NKG_A
	SspH1	10	87.89	30.75	19.83	2.86	11.71	0.15	4NKG_C
6	IpaH3	9	89.22	35.25	24.13	2.53	10.21	0.07	3CVR
7	slrP	12	70.03	34.93	25.34	2.00	10.31	0.11	4PUF_A
	slrP	12	54.07	35.31	26.44	1.53	10.20	0.15	4PUF_B

<sup>a</sup> $n$  repeat number of LRRs,  $P$  helix pitch,  $N$  residue number per turn,  $R$  helix radius,  $\Delta z$  helix rise per turn,  $\Delta\Phi$  rotation per repeat unit,  $p$  helix regularity

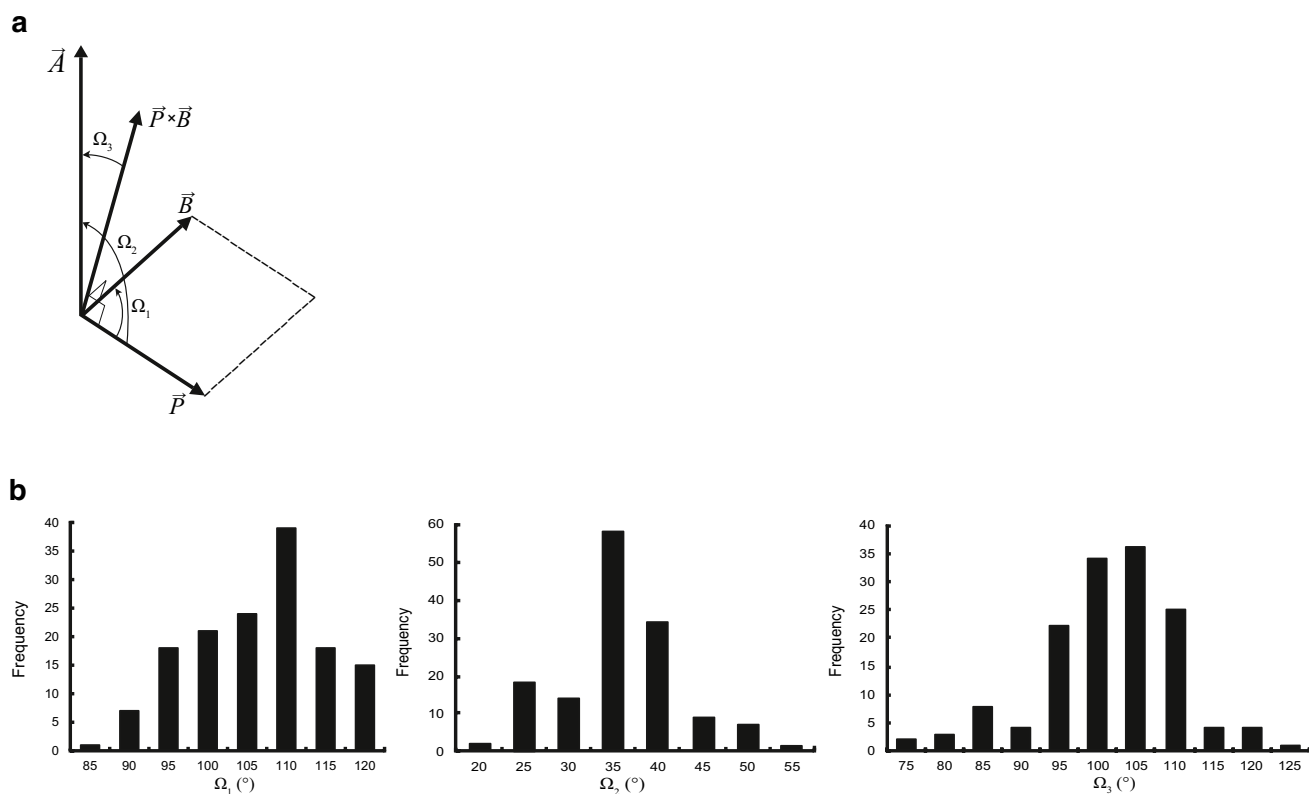


**Fig. 4** The correlation of  $\Delta z$  and  $2R \cdot \sin(\Delta\Phi/2)$  in the helix parameters of LRR domains in the seven bacterial LRR proteins

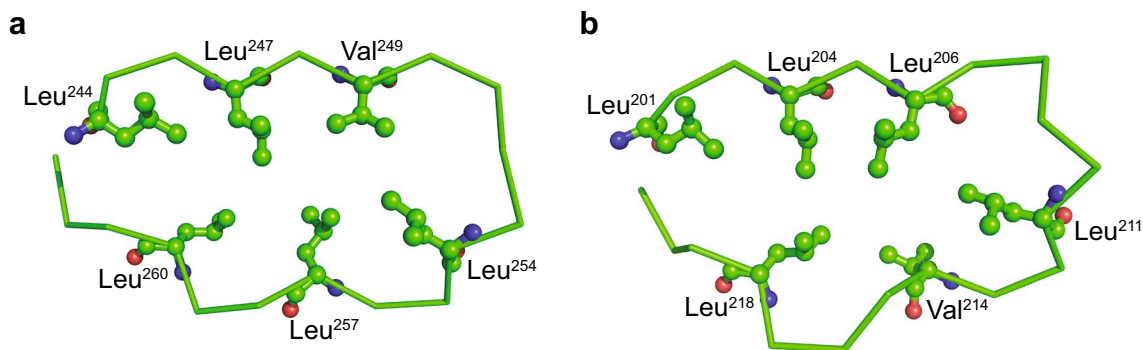
assigned are unambiguously regular or irregular left handed polyproline helices with only a few exceptions. The combination of the secondary structure assignment programs (SEGNO and XLTSSTR) and the HELFIT analysis is useful for PPII assignment in proteins.

## 5 Conclusions

The present study shows that the N-subtype bacterial LRRs are characterized by a unique super secondary structures consisting of PPII helices and a  $\beta$ -turn. In contrast, the T-subtype VS prefers two separate PPIIs with two or three or with only two residues. The type I  $\beta$ -turns can be regarded as regular, right handed helices. We propose three important structural parameters: the three angles between the two helix axes of PPII and  $\beta$ -turn, between two helix axes of PPII and LRR domain, and between the helix axis of LRR domain and the vector product of  $\mathbf{P} \times \mathbf{B}$ . These three angles characterize the super secondary structure and the LRR domain.



**Fig. 5** Geometrical analysis of super secondary structure. **a** The definition of the three angles of  $\Omega_1$ ,  $\Omega_2$ , and  $\Omega_3$ . **b** Frequency distributions of the  $\Omega_1$ ,  $\Omega_2$ , and  $\Omega_3$  angles (by the XLTSSSTR program)



**Fig. 6** Hydrophobic cores in the N-subtype (**a**) and T-subtype (**b**) of bacterial LRR. **a** Sequence 244-LRTLEVSGNQLT-SLPVLPPG-263 (*LRR3*) in *SspH2* (PDB ID: 3G06\_A). **b** Sequence

211-LTHLSLKYNNLTTVPRSLPPS-221 (*LRR7*) in bovine TLR9 (PDB ID:3WPC\_A). Green is  $\alpha$ -carbon, red oxygen, and blue nitrogen. (Color figure online)

**Funding** This study was funded by National University of Mongolia (FELLOWSHIP GRANT-P2016-1173) (to P. E.)

### Compliance with Ethical Standards

**Conflict of interest** All authors declare that they have no conflicts of interest.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Matsushima N, Kretsinger RH (2016) Leucine rich repeats: sequences, structures, ligand—interactions, and evolution. LAMBERT Academic Publishing, Saarbrücken
- Kobe B, Deisenhofer J (1994) The leucine-rich repeat: a versatile binding motif. *Trends Biochem Sci* 19:415–421
- Kobe B, Kajava AV (2001) The leucine-rich repeat as a protein recognition motif. *Curr Opin Struct Biol* 1:725–732
- Matsushima N, Tachi N, Kuroki Y, Enkhbayar P, Osaki M, Kamiya M, Kretsinger RH (2005) Structural analysis of leucine-rich-repeat variants in proteins associated with human diseases. *Cell Mol Life Sci* 62:2771–2791
- Bella J, Hindle KL, McEwan PA, Lovell SC (2008) The leucine-rich repeat structure. *Cell Mol Life Sci* 65:2307–2333
- Matsushima N, Enkhbayar P, Kamiya M, Osaki M, Kretsinger RH (2005) Leucine-rich repeats (LRRs): structure, function, evolution and interaction with ligands. *Drug Design Rev* 2:305–322
- Enkhbayar P, Miyashita H, Kretsinger RH, Matsushima N (2014) Helical parameters and correlations of Tandem Leucine rich repeats in proteins. *J Proteom Bioinform* 7:139–150
- Ng AC, Eisenberg JM, Heath RJ, Huett A, Robinson CM, Nau GJ, Xavier RJ (2011) Human leucine-rich repeat proteins: a genome-wide bioinformatic categorization and functional analysis in innate immunity. *Proc Natl Acad Sci USA* 108(Suppl 1):4631–4638
- Tör M, Lotze MT, Holton N (2009) Receptor-mediated signaling in plants: molecular patterns and programmes. *J Exp Bot* 60:3645–3654
- Coburn B, Sekirov I, Finlay BB (2007) Type III secretion systems and disease. *Clin Microbiol Rev* 20:535–549
- Haraga A, Miller SI (2006) A Salmonella type III secretion effector interacts with the mammalian serine/threonine protein kinase PKN1. *Cell Microbiol* 8:837–846
- Matsushima N, Ohyanagi T, Tanaka T, Kretsinger RH (2000) Super-motifs and evolution of tandem leucine-rich repeats within the small proteoglycans-biglycan, decorin, lumican, fibromodulin, PRELP, keratocan, osteoadherin, epiphycan, and osteoglycin. *Proteins* 38:210–225
- Matsushima N, Kamiya M, Suzuki N, Tanaka T (2000) Super-motifs of leucine-rich repeats (LRRs) proteins. *Genome Inform* 11:343–345
- Matsushima N, Tanaka T, Enkhbayar P, Mikami T, Taga M, Yamada K, Kuroki Y (2007) Comparative sequence analysis of leucine-rich repeats (LRRs) within vertebrate toll-like receptors. *BMC Genom* 8:24–143
- Siligardi G, Drak AF (1995) The importance of extended conformations and, in particular, the PII conformation for the molecular recognition of peptides. *Biopolymers* 37:281–292
- MacArthur MW, Thornton JM (1991) Influence of proline residues on protein conformation. *J Mol Biol* 218:397–412
- Rath A, Davidson AR, Deber CM (2005) The structure of “unstructured” regions in peptides and proteins: role of the polyproline II helix in protein folding and recognition. *Biopolymers* 80:179–185
- Vitagliano L, Berisio R, Mastrangelo A, Mazzarella L, Zagari A (2001) Preferred proline puckerings in cis and trans peptide groups: implications for collagen stability. *Protein Sci* 10:2627–2632
- Adzhubei AA, Sternberg MJ (1993) Left-handed polyproline II helices commonly occur in globular proteins. *J Mol Biol* 229:472–493
- Stapley BJ, Creamer TP (1993) A survey of left-handed polyproline II helices. *Protein Sci* 8:587–595
- Cubellis MV, Caillez F, Blundell TL, Lovell SC (2005) Properties of polyproline II, a secondary structure element implicated in protein–protein interactions. *Proteins* 58:880–892
- Berisio R, Loguercio S, De Simone A, Zagari A, Vitagliano L (2006) Polyproline helices in protein structures: a statistical survey. *Protein Pept Lett* 13:847–854
- Barlow DJ, Thornton JM (1988) Helix geometry in proteins. *J Mol Biol* 201:601–619
- Creamer TP, Campbell MN (2012) Determinants of the polyproline II helix from modeling studies. *Adv Protein Chem* 62:263–282
- Hopfinger A (2012) Conformational properties of macromolecules. Elsevier, Amsterdam
- Schulz GE, Schirmer RH (1979) Principles of protein structure. Springer, Berlin
- Kumar P, Bansal M (2016) Structural and functional analyses of PolyProline-II helices in globular proteins. *J Struct Biol* 196:414–425
- Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637
- Frishman D, Argos P (1995) Knowledge-based protein secondary structure assignment. *Proteins* 23:566–579
- Berman HM, Kleywegt GJ, Nakamura H, Markley JL (2014) The protein data bank archive as an open data resource. *J Comput Aided Mol Des* 28:1009–1014
- Mansiaux Y, Joseph AP, Gelly JC, de Brevern AG (2011) Assignment of PolyProline II conformation and analysis of sequence–structure relationship. *PLoS ONE* 6:e18401
- Srinivasan R, Rose GD (1999) A physical basis for protein secondary structure. *Proc Natl Acad Sci USA* 96:14258–14263
- Cubellis MV, Cailliez F, Lovell SC (2005) Secondary structure assignment that accurately reflects physical and evolutionary characteristics. *BMC Bioinform* 6(Suppl 4):S8
- King SM, Johnson WC (1999) Assigning secondary structure from protein coordinate data. *Proteins* 35:313–320
- Kumar P, Bansal M (2005) Identification of local variations within secondary structures of proteins. *Acta Crystallogr* 71:1077–1086
- Richards FM, Kundrot CE (1988) Identification of structural motifs from protein coordinate data: secondary structure and first-level supersecondary structure. *Proteins* 3:71–84
- Chiang YS, Gelfand TI, Kister AE, Gelfand IM (2007) New classification of supersecondary structures of sandwich-like proteins uncovers strict patterns of strand assemblage. *Proteins* 68:915–921
- Adzhubei AA, Sternberg MJ (1994) Conservation of polyproline II helices in homologous proteins: implications for structure prediction by model building. *Protein Sci* 3:2395–2410
- Evdokimov AG, Anderson DE, Routzahn KM, Waugh DS (2001) Unusual molecular architecture of the Yersinia pestis cytotoxin YopM: a leucine-rich repeat protein with the shortest repeating unit. *J Mol Biol* 312:807–821
- Park H, Huxley-Jones J, Boot-Handford RP, Bishop PN, Attwood TK, Bella J (2008) LRRCE: a leucine-rich repeat cysteine capping motif unique to the chordate lineage. *BMC Genom* 9:599
- Adzhubei AA, Sternberg MJ, Makarov AA (2013) Polyproline-II helix in proteins: structure and function. *J Mol Biol* 425:2100–2132
- Keszei AF, Tang X, McCormick C, Zeqiraj E, Rohde JR, Tyers M, Sicheri F (2014) Structure of an SspH1-PKN1 complex reveals the basis for host substrate recognition and mechanism of activation for a bacterial E3 ubiquitin ligase. *Mol Cell Biol* 34:362–373
- Quezada CM, Hicks SW, Galan JE, Stebbins CE (2009) A family of Salmonella virulence factors functions as a distinct class of autoregulated E3 ubiquitin ligases. *Proc Natl Acad Sci USA* 106:4864–4869



44. Zhu Y, Li H, Hu L, Wang J, Zhou Y, Pang Z, Liu L, Shao F (2008) Structure of a Shigella effector reveals a new class of ubiquitin ligases. *Nat Struct Mol Biol* 15:1302–1308
45. Takagi K, Kim M, Sasakawa C, Mizushima T (2016) Crystal structure of the substrate-recognition domain of the Shigella E3 ligase IpaH9.8. *Acta Crystallogr* 72:269–275
46. Zouhir S, Bernal-Bayard J, Cordero-Alba M, Cardenal-Munoz E, Guimaraes B, Lazar N, Ramos-Morales F, Nessler S (2014) The structure of the Slrp-Trx1 complex sheds light on the autoinhibition of the type III secretion system effectors of the NEL family. *Biochem J* 464:135–144
47. Kokatla HP, Sil D, Tanji H, Ohto U, Malladi SS, Fox LM, Shimizu T, David SA (2014) Structure-based design of novel human Toll-like receptor 8 agonists. *ChemMedChem* 9:719–723
48. Ohto U, Shibata T, Tanji H, Ishida H, Krayukhina E, Uchiyama S, Miyake K, Shimizu T (2015) Structural basis of CpG and inhibitory DNA recognition by Toll-like receptor 9. *Nature* 520:702–705
49. Tanji H, Ohto U, Shibata T, Miyake K, Shimizu T (2013) Structural reorganization of the toll-like receptor 8 dimer induced by agonistic ligands. *Science* 339:1426–1429
50. Paracuellos P, Kalamajski S, Bonna A, Bihan D, Farndale RW, Hohenester E (2017) Structural and functional analysis of two small leucine-rich repeat proteoglycans, fibromodulin and chondroadherin. *Matrix Biol* 63:106–116
51. Scott PG, McEwan PA, Dodd CM, Bergmann EM, Bishop PN, Bella J (2004) Crystal structure of the dimeric protein core of decorin, the archetypal small leucine-rich repeat proteoglycan. *Proc Natl Acad Sci USA* 101:15633–15638
52. Scott PG, Dodd CM, Bergmann EM, Sheehan JK, Bishop PN (2006) Crystal structure of the biglycan dimer and evidence that dimerization is essential for folding and stability of class I small leucine-rich repeat proteoglycans. *J Biol Chem* 281:13324–13332
53. Seiradake E, del Toro D, Nagel D, Cop F, Härtl R, Ruff T, Seyit-Bremer G, Harlos K, Border EC, Acker-Palmer A, Jones EY, Klein R (2014) FLRT structure: balancing repulsion and cell adhesion in cortical and vascular development. *Neuron* 84:370–385
54. Lu YC, Nazarko OV, Sando R 3rd, Salzman GS, Südhof TC, Araç D (2015) Structural basis of latrophilin-FLRT-UNC5 interaction in cell adhesion. *Structure* 23:1678–1691
55. Ranaivoson FM, Liu Q, Martini F, Bergami F, von Daake S, Li S, Lee D, Demeler B, Hendrickson WA, Comoletti D (2015) Structural and mechanistic insights into the latrophilin3-FLRT3 complex that mediates glutamatergic synapse development. *Structure* 23:1665–1677
56. Jackson VA, Mehmood S, Chavent M, Roversi P, Carrasquero M, Del Toro D, Seyit-Bremer G, Ranaivoson FM, Comoletti D, Sansom MS, Robinson CV, Klein R, Seiradake E (2016) Super-complexes of adhesion GPCRs and neural guidance receptors. *Nat Commun* 17:11184
57. Enkhbayar P, Damdinsuren S, Osaki M, Matsushima N (2008) HELFIT: helix fitting by a total least squares method. *Comput Biol Chem* 32:307–310
58. Matsushima N, Miyashita H, Mikami T, Yamada K (2011) A new method for the identification of leucine-rich repeats by incorporating protein secondary structure prediction. Tuteja R (ed) *Bioinformatics: genome bioinformatics and computational biology*. NOVA Science Publishers, Hauppauge, pp 61–88
59. Chebrek R, Leonard S, de Brevern AG, Gelly JC (2014) PolyprOnline: polyproline helix II and secondary structure assignment database. *Database* 2014:bau102
60. Hutchinson EG, Thornton JM (1996) PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Sci* 5:212–220
61. Heinig M, Frishman D (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res* 32(WebServer issue):W500–W502
62. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612
63. Enkhbayar P, Hikichi K, Osaki M, Kretsinger RH, Matsushima N (2006) 3(10)-helices in proteins are parahelices. *Proteins* 64:691–699
64. Enkhbayar P, Boldgiv B, Matsushima N (2010) Omega-Helices in proteins. *Protein J* 29:242–249
65. Matsushima N, Miyashita H, Enkhbayar P, Kretsinger RH (2015) Comparative geometrical analysis of leucine-rich repeat structures in the nod-like and toll-like receptors in vertebrate innate immunity. *Biomolecules* 5:1955–1978
66. Wilmot CM, Thornton JM (1988) Analysis and prediction of the different types of beta-turn in proteins. *J Mol Biol* 203:221–232
67. Jha AK, Colubri A, Zaman MH, Koide S, Sosnick TR, Freed KF (2005) Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library. *Biochemistry* 44:9691–9702
68. Adzhubei AA, Eisenmenger F, Tumanyan VG, Zinke M, Brodzinski S, Esipova NG (1987) Approaching a complete classification of protein secondary structure. *J Biomol Struct Dyn* 5:689–704
69. Rämisch S, Pramhed A, Tillgren V, Aspberg A, Logan DT (2017) Crystal structure of human chondroadherin: solving a difficult molecular-replacement problem using de novo models. *Acta Crystallogr* 73:53–63
70. Kay BK, Williamson MP, Sudol M (2000) The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J* 14:231–241
71. Ananthanarayanan VS, Soman KV, Ramakrishnan C (1987) A novel supersecondary structure in globular proteins comprising the collagen-like helix and beta-turn. *J Mol Biol* 198:705–709
72. Matsushima N, Creutz CE, Kretsinger RH (1990) Polyproline, beta-turn helices. Novel secondary structures proposed for the tandem repeats within rhodopsin, synaptophysin, synexin, gliadin, RNA polymerase II, hordein, and gluten. *Proteins* 7:125–155
73. Meredith GD, Chang WH, Li Y, Bushnell DA, Darst SA, Kornberg RD (1996) The C-terminal domain revealed in the structure of RNA polymerase II. *J Mol Biol* 258:413–419
74. Kumaki Y, Matsushima N, Yoshida H, Nitta K, Hikichi K (2001) Structure of the YSPTSPS repeat containing two SPXX motifs in the CTD of RNA polymerase II: NMR studies of cyclic model peptides reveal that the SPTS turn is more stable than SPSY in water. *Biochim Biophys Acta* 1548:81–93
75. Cramer P, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292:1863–1876
76. Hsin JP, Manley JL (2012) The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* 26:2119–2137
77. Martin J, Letellier G, Marin A, Taly JF, de Brevern AG, Gibrat JF (2005) Protein secondary structure assignment revisited: a detailed analysis of different assignment methods. *BMC Struct Biol* 5:17
78. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190