

Article

Not peer-reviewed version

Unique Super-Secondary Structures for Novel Leucine-Rich Repeats in Many Proteins from the Bacterial PVC Superphylum

[Norio Matsushima](#)^{*}, [Dashdavaa Batkhishig](#), Purevjav Enkhbayar

Posted Date: 27 January 2026

doi: 10.20944/preprints202601.2081.v1

Keywords: SDS22-like leucine-rich repeat; cysteine-containing leucine-rich repeat; duality; composite helix; $\beta - \alpha/3(10) - \beta$ motif; $\beta - 3(10) - \beta$ motif; PVC group; protein kinase domain



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Unique Super-Secondary Structures for Novel Leucine-Rich Repeats in Many Proteins from the Bacterial PVC Superphylum

Norio Matsushima ^{1,*}, Dashdavaa Batkhishig ² and Purevjav Enkhbayar ³

- ¹ Center for Medical Education, Sapporo Medical University, Sapporo 060-8556, Japan
 - ² Department of Physics, School of Mathematics and Natural Sciences, Mongolian National University of Education, Ulaanbaatar 210648, Mongolia
 - ³ Department of Biology, School of Arts and Sciences, National University of Mongolia, Ulaanbaatar 14201, Mongolia
- * Correspondence: norio_irreko@outlook.jp; Tel.: +81-1-4383-1187

Abstract

The sequence, structure, function relationship in proteins is still significant and meaningful to understanding structural principles Tandem leucine-rich repeats (LRRs) have been well classified into eleven canonical LRR types. We identified novel LRR motifs having a dual characteristic of two canonical types in over 3,600 proteins, mainly from the bacterial PVC superphylum, by using sequence similarity searches. We studied their structures using AlphaFold and their biological functions by using protein domain searches. The novel LRR motifs are classified into three groups. The first and second groups are characterized by the consensuses of LxxLxLxx(C/T)xzI TDxxLxx(L/F)xx(L/C)xx, LxxLxLxxCxxI TDxxLxxLxxLP in which “z” shows a deletion. A new canonical type with the consensus sequence LxxLxLxx(T/N/C)xzV xxLxPLxxMx was identified as the third group. AlphaFold predicts that all of the LRR domains form a solenoid structure by the parallel stacking of β -strands. The LRR domains are mostly characterized by two unique super-secondary structures (SSSs) of β -strand – α -helix adjoining 3(10)-helix – β -strand motif and β -strand – 3(10)-helix – β -strand motif and a well-known SSS of β -strand – α -helix – β -strand motif. Many LRR proteins containing a kinase domain or an F-box domain might be involved in bacterial immunity or ubiquitination. We suggest that the former two SSSs are a fundamental motif in protein structure.

Keywords: SDS22-like leucine-rich repeat; cysteine-containing leucine-rich repeat; duality; composite helix; β – α /3(10) – β motif; β – 3(10) – β motif; PVC group; protein kinase domain

1. Introduction

Leucine-rich repeats (LRRs) occurring in tandem are present in a huge number of proteins [1]. The LRR-containing proteins lead to a variety of functions such as immune response, ubiquitin related process, apoptosis, and neuronal development through protein-protein interactions [2–4]. The repeating unit length (RUL) that consists generally of 20 – 30 amino acids is divided into a highly conserved segment (HCS) and a variable segment (VS). For ease of explanation, we use the following abbreviations. The HCS part with eleven residues is abbreviated as HCS = 11 [5–7]. Correspondingly, the VS part with eleven residues is abbreviated as VS = 11. The same applies hereafter.

Eleven canonical types of LRR have been widely recognized; they include RI-like, Cysteine containing (CC), GALA, SDS22-like, plant specific (PS), and Leptospira-like LRRs [8–12]. All their structures have been determined. In LRRs, positions 3 – 5 in the HCS form short β -strands. The parallel stacking of β -strands produces a superhelix arrangement (called a solenoid structure). The overall shape resembles a horseshoe (arc), a prism, or a circle (Baumkuchen ring-like shape) [13]. The

solenoid structures consist of four parts which are a concave surface, ascending loops, a convex surface, and descending loops [12].

The VS parts adopt various secondary structures including α -helix, 3(10)-helix, polyproline II helix, multiple β -turns, and an extended conformation. α -Helix characterized by consecutive ($i \rightarrow i+4$) hydrogen bonds [14] is observed in CC, RI-like and GALA LRRs [7], while 3(10)-helix characterized by consecutive ($i \rightarrow i+3$) hydrogen bonds [14,15] in SDS22-like, PS, and Leptospira like LRRs [6]. The hydrophobic cores of the solenoid structures are formed by the conserved hydrophobic residues in the consensus sequences; these residues are mainly Leu, Ile, or Val, while Phe, Cys, or Met sometimes appear [16]. Position 9 in HCS = 11 is highly conserved as Asn [2–4]. The side chain forms a continuous hydrogen bond network (called asparagine ladder) [2,17].

The sequence, structure, function correlation in proteins is still significant and meaningful to understanding structural principles and to designing proteins with desired structures and functions. Super-secondary structures (SSSs) comprise several secondary structure elements with unique and compact folding of a polypeptide chain [18]. The most common SSSs are helix-turn-helix, helix-loop-helix, coiled coil, beta-hairpin, Greek key, beta barrel, beta-alpha-beta (β - α - β) and zinc finger in which “helix” shows α -helix and “beta” shows β -strand [19]. Moreover, a composite helix, made up of contiguous α - and 3(10)-helices ($\alpha/3(10)$ composite), is certainly present in protein structures [20].

The LRR domains of RI-like, CC and GALA types form consecutive β -strand – ascending loop – α -helix – descending loop – β -strand (β - α - β) motif (Figure 1A) [6]. This β - α - β motif is a typical SSS seen in many other protein structures. We previously indicated that three types of SDS22-like, Leptospira-like and PS LRRs adopt an SSS consisting of 3(10)-helix – β -turn [6]. The β -turn constitutes a part of the descending loop in the LRR solenoid structures and therefore we take a new look at the secondary structure pattern. We notice that these three types of LRR domains consist of a unique SSS consisting of β -strand – ascending loop – 3(10)-helix – descending loop – β -strand (β – 3(10) – β) motif (Figure 1B).

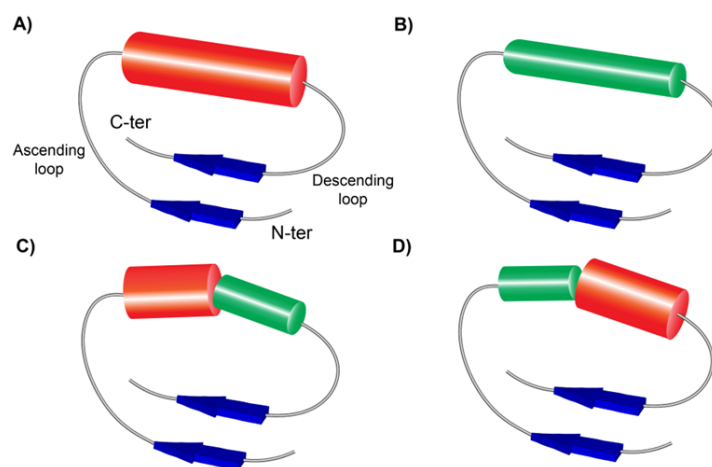


Figure 1. Four super-secondary structure in CC/SDS22-like, short CC/SDS22like and short SDS22-like LRRs. (A) β – α – β motif. (B) β – 3(10) – β motif. (C) β – $\alpha/3(10)$ – β motif. (D) β – 3(10)/ α – β motif. “ β ” indicated by dark blue arrows is β – strand. “ α ” indicated by red tube is α -helix. “3(10)” indicated by green tube shows 3(10)-helix. “ $\alpha/3(10)$ ” shows a composite helix made up made up of contiguous α - and 3(10)-helices. “3(10)/ α ” shows a composite helix made up made up of contiguous 3(10)- and α -helices. The α -helix tube is thicker than the 3(10)-helix tube. The composite helices have a kink between the two helices.

A very recent sequence analysis indicated that a novel LRR is present in over 280 proteins from microorganisms (protists, fungi, and bacteria) [21]. Its consensus is represented by LxxLDLxxTxV SGxLxxLxxLxx (HCS = 11 and VS=12). This type shows a dual characteristic of two canonical types of SDS22-like and PS LRRs, because the consensus of SDS22-like LRR is LxxLxLxxNxI xxLxxLxxLxx with RUL =22 (HCS = 11 and VS = 11) and the consensus of PS LRR is LxxLxLxxNxL SGxIPxxLxxLxx with RUL = 24 (HCS = 11 and VS = 13). We call it PS/SDS22-like LRR [21]. Naturally, both VS parts of SDS22-like and PS/SDS22-like LRRs have in common the sequence of LxxLxxL.

We hypothesized that other types of LRRs with the sequence of LxxLxxLxx in the VS parts are present in many proteins. In fact, a viral protein (UniProtKB: A0A6C0H0T7) with 351 residues contains tandem ten LRRs [15]. The sequence alignment shows that most of the LRRs (8/10 = 0.8) are well represented by the consensus of LxxLxLxxCxxI TDxxLxxLxxLxx with RUL = 25 (HCS=12 and VS=13) (Table 1). The consensus sequence of Cysteine-containing (CC) LRR is LxxLxLxxCxxL TDxGLxxLxxCxx with RUL = 26 (HCS=12 and VS = 14) or LxxLxLxxNxL TDxGLxxLxxCxx (Table 1) [7]. Thus, this novel LRR type appears to have a dual characteristic of CC and SDS22-like LRRs. We call it CC/SDS22-like LRR. We performed sequence similarity searches by using the protein database. We did the analysis of sequence and predicted the structure by AlphaFold. The purpose of this study is to identify new LRR types in many proteins, to deeply understand the sequence – structure correlation, and to discuss the biological functions of proteins identified here.

We identified CC/SDS22-like LRR and other motifs closely associated with its type in over 3,600 proteins that mainly come from the bacterial PVC group (Table S1). The new LRR motifs identified are classified in three groups. The first group (CC/SDS22-like LRR) is characterized by the consensuses of LxxLxLxxCxxI TDxxLxxLxxLxx, LxxLxLxxTxV TDxxLxxLxxLxx, or LxxLxLxxTxV TDxGLxHFxxCxx with VS = 13. The consensus of the second (called short CC/SDS22-like LRR) is LxxLxLxxCxxI TDxxLxxLxxLP with VS = 12. The consensuses of the third (called short SDS22-like LRR) are LxxLxLxxTxV xxLxPLxxMx, LxxLxCxx(T/N)xV xxLxPLxxMP or LxxLxLxxCxxL xxLxPLxxMP with VS = 10. AlphaFold structures reveals that the LRR domains form a solenoid structure by the parallel stacking of β -strands as seen in the known structures. The short SDS22-like LRR domains consist of consecutive $\beta - 3(10) - \beta$ motifs. Here we first recognized the $\beta - 3(10) - \beta$ motif as a unique SSS. The LRR domains of the CC/SDS22-like and short CC/SDS22-like types are mainly characterized by a typical SSS of $\beta - \alpha - \beta$ motif, as seen in proteins such as ribonuclease inhibitor and FBXL17, and a quirky SSS of $\beta - \alpha$ -helix adjoining 3(10)-helix – β ($\beta - \alpha/3(10) - \beta$) motif. The $\beta - 3(10) - \beta$ and $\beta - 3(10)/\alpha - \beta$ motifs rarely occur. The SSSs of the $\beta - \alpha/3(10) - \beta$, $\beta - 3(10) - \beta$, $\beta - 3(10)/\alpha - \beta$ motifs are also observed in CC/SDS22-like LRRs that occur occasionally in the known structures. A dual characteristic of CC and SDS22-like LRRs and the feature of a chameleon-like sequence of CC/SDS22-like LRR would provide simultaneous occurrence of four SSSs. We consider that the SSSs of the $\beta - 3(10) - \beta$ and $\beta - \alpha/3(10) - \beta$ motifs are fundamental motifs in protein structure as well as the $\beta - \alpha - \beta$ motif and that the flexibility of the ascending loop and descending loop in the LRRs structures provides the common structure-preservation. Many proteins containing the new types of LRRs from the bacterial PVC superphylum are linked to an F-box domain, a kinase domain, or a TIR domain. These proteins might be involved in ubiquitination and bacterial immunity. It appears that the novel LRRs types identified possibly confer species specificity of the function. The sequence analysis suggests that LRR rapidly evolved.

2. Materials and Methods

Metagenome data of nucleo-cytoplasmic large DNA viruses (NCLDV) were reported by Schulz et al., [22]. We identified 547 LRR proteins from 199,021 proteins. Disease resistance R13L4/SHOC-2-like LRR domain-containing protein (UniProtKB: A0A6C0H0T7) with 351 residues contains tandem ten LRRs [16]. The sequence alignment shows that most of the LRRs (8/10 = 0.8) are well represented by the consensus of LxxLxLxxCxxI TDxxLxxLxxLxx with RUL = 25 (HCS=12 and VS=13). Cysteine-containing LRR (CC LRR) is characterized by the consensus of LxxLxLxxCxxI TDxGLxxLxxCxx with RUL = 26 (HCS = 12 and VS =14) [7]. Consequently, the above LRR type with the consensus of LxxLxLxxCxxI TDxxLxxLxxLxx has a dual characteristic of two types of CC and SDS22-like LRRs. Thus, we called it CC/SDS22-like LRR (Table 1, Figure 2).

We searched for proteins containing CC/SDS22-like LRRs using the sequence of the CC/SDS22-like LRR domain (UniProtKB: A0A6C0H0T7) as a query sequence. First, sequence similarity searches were performed by the programs of FASTA and BLAST on at the Bioinformatic Center, Institute for Chemical Research, Kyoto University on March 02, 2024 (<http://www.genome.jp>) and then candidate proteins were detected. The amino acid sequences of the candidate proteins were taken from the

UniProt database [1]. Secondly, LRR domains including a CC/SDS22-like LRR in the candidate proteins detected in the first step were assigned by LRRpred [23]. Thirdly, the sequence similarity search by using the LRR domains identified was iterated.

We found that one bacterial protein with 2,121 residues (UniProtKB: A0A3L7QJC1) detected by the above procedure contains two LRR domains (Table S1). One consists of six tandem repeats of CC/SDS22-like LRR, as expected. The other is ten tandem repeats with the consensus of LxxLxCxxTxV xxLxxLxxMx with HCS = 11 and VS = 10. The VS part is one residue shorter than VS = 11 of canonical SDS22-like LRR. We call it short SDS22-like LRR. Therefore, we also performed the sequence similarity search using the sequence of this short SDS22-like LRR domain as a query sequence. Furthermore, we identified other type in three proteins. The other type was called short CC/SDS22-like LRR, because the VS is twelve residues long and thus one residue shorter than VS = 13 in the CC/SDS22-like LRR.

The consensus sequences of their new LRR types were determined by utilizing the occurrence probability of an amino acid residue at each position by WebLogo (<https://weblogo.berkeley.edu/logo.cgi>) [24]. Bold uppercase letters indicate more than 70% occurrence of a given residue in a certain position; normal letters indicate 40 - 70% occurrence, and lowercase letters indicate 20 - 40% occurrence.

Three procedures were applied. First, we collected the structures of the new LRR motifs. One is a screening of the CC/SDS22-like LRR in the solved structures. Six structure data of coupling factor (PDB:ID 3E4G), FBXL17 (6W66), LRRC8 (6FNW), InlK (4L3F), lmo2027 (5KZS), and LMOF2365_1397 (4EZG) were useful for analysis (Table 2) [25–29]. The other is the predicted structures by AlphaFold which is a deep learning-based method [30,31].

Secondly, we performed the secondary structure assignments of the structures obtained by the above approaches; the program of DSSP-PPII activated in the PolyprOnline web interface was used [32]. Types of β -turn were also identified by using the PROMOTIF program [33]. DSSP-PPII which is based on the identification of precise hydrogen bond patterns corresponding to regular secondary structures defines eight types of secondary structures [34]. We used rule 6.3 which was described in a previous paper [6]. Thus, the “-GGG-” assignment shows the five-residue 3(10)-helix. The PROMOTIF program elucidates that residue i of a β -turn is located in one N-terminal residue before of “TT” or “SS”. The “-TT-TT-” assignment shows the tandem array of two β -turns. A detailed explanation of the above is given in the previous paper [6].

Thirdly, HELFIT analysis was performed for the predicted structures by AlphaFold. The HELFIT analysis determines helical parameters of helix structures [13]. The parameters are helix axis, helix pitch (P), helix radius (R), number of repeats per turn (N), and handedness; the rise per repeat unit (Δz) is given by P/N and the rotation per repeat unit ($\Delta\phi$) in the helix by $360^\circ/N$. The solenoid structures of LRRs are represented as a right- or left-handed helix. In the LRR solenoids, the $C\alpha$ coordinates of the consensus hydrophobic residues such as leucine at position 4 in the HCS part (located in the center of short β -strands) in individual units are used for the HELFIT analysis.

3. Results

We identified and analyzed 3,858 proteins (Table S1). The LRRs within the 2,801 proteins were divided into three types of CC/SDS22-like, short CC/SDS22-like, and short SDS22-like, LRRs (Table 1, Figs. 2 and 3). The remaining proteins are F-box/LRR-repeat protein 14 (FBXL14) from the phyla Viriplantae.

3.1. Consensus Sequences of the New Types of LRR

3.1.1. Three Subtypes of CC/SDS22-like LRR with HCS = 11 or 12 and VS = 13

The CC/SDS22-like LRR may be divided into three subtypes (Table 1, Figure 2, Table S2). This LRR is characterized by (T/S)DxxLxx(L/F)xx(L/C)xx with VS = 13. The first subtype of CC/SDS22-like LRR was identified in two hundred forty-six proteins. Ninety-six proteins containing LRR domains

consisting mainly of CC/SDS22-like LRR with RUL = 25 (HCS = 12 and VS = 13) are encoded from the bacterial PVC superphylum of the Chlamydiae (81 proteins), the Planctomycetes (14), and the Verrucomicrobia (1). Two proteins from *Oxalobacteraceae bacterium* (the phylum Proteobacteria) and three proteins from metagenome data contain CC/SDS22-like LRR domains.

The Genome sequence of *Protochlamydia amoebophila* UWE25, a member of the phylum Chlamydiae, has been determined by two groups of Horn et al., and Domann et al., [35,36]. Seventy-three proteins from *P. amoebophila* contain the CC/SDS22-like LRR domains. The dual LRR motif occupies 91% of the repeat number (= 976/1073) (Table 1). The consensus sequence is well represented by LQHL(n/d)LSxCNxL TDAGLAHLxPLxA. Fourteen proteins from four species of the phylum Planctomycetes contain 195 LRRs. The four-fifths (154/195 = 0.79) are CC/SDS22-like LRR with the consensus sequence of LqxLdLxgCxxI TDAGLAHLxxLxx (Table 1).

One hundred forty containing LRR domains consisting mainly of CC/SDS22-like LRR with RUL = 25 (HCS = 12, VS = 13) are encoded from two phyla of Stramenopile and Discoba (only *Bodo Saltans*). The fifty proteins from thirty species of the phylum Stramenopile contain 967 LRRs; the forty-four proteins are homologous. More than three-quarters of the 967 repeats (752/967 = 0.78) are CC/SDS22-like LRR with RUL = 25 (HCS = 12 and VS = 13) (Table 1). The consensus is LtsL(n/d)LxgCnx(i/lv) TDxG(l/i)xxLaxLxx.

Three-quarters of total LRR repeats (641/857 = 0.75) in the *B. Saltans* ninety proteins are the CC/SDS22-like type with the consensus of LQxLdLxgCxxI TDAGLxx(i/v/l)AxLxq (Table 1).

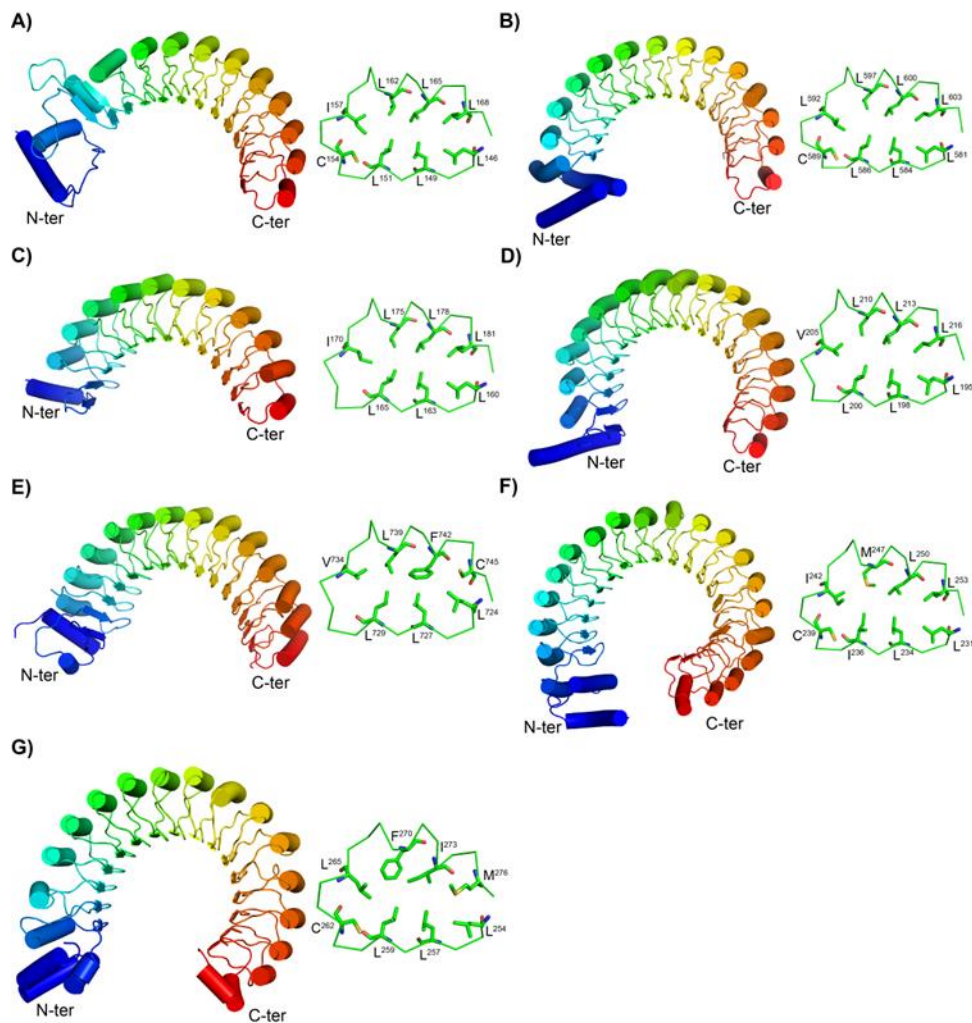
Eleven species of the class Chlorophyta (green algae) encode fifty-two CC/SDS22-like LRR proteins. Thirty-two proteins from *Ostreobium quekettii* contain CC/SDS22-like LRR

with RUL = 25 (HCS = 12 and VS = 13). Four-fifths (423/531=0.80) are this type (Table 1). The consensus is LtxLdLsgCxx(l/v) TDxGLxx(l/v)gxLtx.

Table 1. Consensus sequences of CC/SDS22-like, short SDS22-like, and short CC/SDS22-like leucine-rich repeats (LRRs).

LRR types	Sub-type	RUL ^a	Consensus sequence														Pro.	LRR											
			HCS ^a												VS ^a														
			1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	13	14	No. ^b
CC/SDS22-like	1	25	L	Q	H	L	n	L	S	x	C	N	x	L	T	D	A	G	L	A	H	L	x	P	L	x	x	73	976(1073)
"	1	25	L	q	x	L	d	L	x	g	C	N	x	I	T	D	A	G	L	A	H	L	x	x	L	x	x	15	154(195)
"	1	25	L	t	s	L	n	L	x	g	C	n	x	i	T	D	A	G	L	A	H	L	x	x	L	x	x	55	752(967)
"	1	25	L	Q	x	L	d	L	x	g	C	n	x	I	T	D	A	G	L	x	x	i	A	x	L	x	q	90	641(857)
"	1	25	L	t	x	L	d	L	x	g	C	x	x	I	T	D	x	G	L	x	x	l	g	x	L	t	x	52	423(531)
"	1	25	L	t	x	L	d	L	x	g	C	x	x	v	T	D	x	G	L	x	x	L	x	x	L	x	x	12	104(181)
"	2	24	L	x	x	L	x	L	x	x	T	q	V	i	s	D	a	G	L	e	h	L	r	g	L	t	n	26	246(265)
"	2	24	L	e	x	L	x	L	x	g	T	q	I	v	T	D	a	G	L	e	h	L	k	g	L	t	s	82	804(933)
"	2	24	L	r	e	L	x	L	d	g	t	x	I		t	D	x	G	L	x	x	L	x	x	L	x	x	8	208(298)
"	3	24	L	T	x	L	x	L	x	x	T	x	V		T	D	A	G	L	A	H	F	K	d	C	K	n	28	204(239)
Short CC/SDS22-like		24	L	Q	x	L	D	L	S	G	C	x	G	I	T	D	a	G	L	A	H	L	k	x	L	P		3	34(58)
Short SDS22-like	1	21	L	x	x	V	d	L	s	g	T	p	V	i		s	D	L	S	P	L	k	G	M	P		364	2733(3737)	
"	2	21	L	t	t	L	x	C	s	g	N	r	I	l		x	S	L	E	P	L	R	G	M	P		28	294(337)	
"	3	21	L	T	x	L	n	C	x	g	T	p	V			S	D	L	S	P	L	K	G	M	P		27	160(223)	
"	4	22	L	T	x	L	d	L	x	g	C	x	q	V		r	D	L	s	P	L	k	G	M	P		12	88(117)	
CC		26	L	T	x	L	x	L	x	x	C	x	x	I	T	D	x	G	L	x	x	L	A	x	C	x	x	Reference [7]	
CC		25	L	T	x	L	x	L	x	x	N	x	L		T	D	x	G	L	x	x	L	A	x	C	x	x	Reference [7]	
SDS22-like		22	L	T	x	L	n	L	x	x	N	x	L			x	x	L	x	x	L	x	x	L	x	x	x	Reference [6]	

Note. (a) “RUL” indicates the length of repeat unit of LRR. “HCS” and “VS” indicate its highly conserved segment and its variable segment, respectively. (b) “Pro No.” is the number of proteins used for analysis. “LRR No.” is the number of LRR repeat units characterizing respective LRR patterns in the proteins. The numbers in parentheses denote the total number of LRR units in the proteins. Bold uppercase letters in the consensus sequences indicate more than 70% occurrence of a given residue in a certain position; normal letters indicate 40-70% occurrence and lowercase letters indicate 20-40% occurrence. Shaded areas indicate highly conserved, characteristic amino acid residues.



Blastopirellula cremea (the phylum Planctomycetes) encodes eighty-two CC/SDS22-like LRR proteins in which most of the repeats are this type (804/933 = 0.86) with the consensus of L(e/k)xLxLxgTqx(I/v) TDaGLhLkgLts (Table 1).

Candidatus Melainabacteria, *Candidatus Hydrogenedentes*, and *Candidatus Poribacteria*, and *Candidatus Stahlbacteria* that are members of Candidatus phyla encode proteins containing CC/SDS22-like LRR with RUL = 24 (HCS = 11 and VS = 13). Candidatus bacteria (*Candidatus Obscuribacter phosphatis* and *Candidatus Obscuribacter sp*) and two unclassified Bacteria encode eight CC/SDS22-like LRR proteins that are likely homologous. Of 298 repeats, 73% (= 208/298) are CC/SDS22-like with the consensus of L(r/k)eLxL(d/n)txxI (t/s)DxGLxxLxxLxx (Table 1).

An euryarchaeota archaeon encodes one CC/SDS22-like LRR protein. It contains six LRR repeats with the consensus of LxxLxLxxTxI TDxGLxEVxxLxx. Marine metagenome data encode CC/SDS22like LRR proteins.

The third subtype of CC/SDS22-like LRR is identified in twenty-eight proteins from eight species of the bacterial phylum Planctomycetes (Table 1). Most of the LRRs (204/239 = 0.85) are CC/SDS22-like LRR with the consensus of LTxLxLxxTxV TDAGLAHFkDCKn (Table 1). The first unit of the LRR domains is frequently represented by not HCS = 11 but HCS = 12; the consensus is LT(a/e)(V/L)dLsgNpqV TDAGLAhFKdCKn.

3.1.2. F-box/LRR-Repeat Protein 14 (FBXL14)

Homologs of F-box/LRR-repeat protein 14 (FBXL14) are identified from many species of the phylum Viridiplantae. The homolog is present in over seven hundred eighty proteins at least. Most individual LRR domains in the FBXL14 homologs contain both the first subtype and the second subtype of CC/SDS22-like LRR; the number of the second type is higher than that of the first type. For the consensus sequence, we used typical FBXL14 from fifty-one plant species. The consensus of the second type (663/969 = 0.68) is LksLnLsg(c/s/t)x(I/v) TDaGLxhLkgLxn; position 9 prefers not only Thr but also Cys and Ser. The consensus of the first type (202/969 = 0.21) is L(t/e)SLN(L/f)(n/s)x(C/N)(n/d)x(I/l) TDxG(L/M)(e/k)x(I/L)SGLtN; position 9 prefers Cys and Asn. Similarly, at least seventy-six proteins have the two subtypes of CC/SDS22-LRR in individual LRR domains; most of the seventy-six proteins (87 % = 65/75) come from thirty-four species of the phylum Sar.

3.1.3. Short CC SDS22-like LRR with HCS = 12 and VS = 12

This type is seen in only three proteins from *Oxalobacteraceae bacterium*. The consensus is LqxLDLSGCxGI TDaGLAHLkx(L/m)P (34/58 = 0.59) (Table 1). A short CC/SDS22-like LRR with HCS = 11 is seen in one protein from *Bdellovibrionales bacterium*; all of the six LRRs are represented by the consensus of LRxxLxxTxV TDxGLxxLKGLP.

3.1.4. Four Subtypes of Short SDS22-like LRR with HCS = 11 or 12 and VS = 10

Short SDS22-like LRR may be grouped into four subtypes (Table 1). The short SDS22-like LRR is characterized by LxxLx(L/C)xx(N/T/C)x(x/z)(V/I) xxLxPLxxMx. The remarkable feature is the occupation of Met at position 8 in the VS part. Most proteins are identified from the bacterial PVC superphylum.

The first subtype is the most abundant. Almost all of a total of 364 proteins (347/364 = 0.95) are identified from the bacterial PVC superphylum. Two proteins are from archaeal *Candidatus Woesearchaeota*. Over 70% of the LRRs (2733/3737 = 0.73) are well represented by the consensus of LxxLxLxgTpV sDLSPLkGMP (Table 1). The feature is that position 9 is occupied by not Asn but Thr.

The consensus of the second and third subtypes are LttLxCsGnrI xSLEPLRGMP and LTxLnCxgTpV SDLSPLKGMp, respectively. Position 5 is occupied by Cys instead of Leu. The former is seen in twenty-eight proteins (294/337 = 0.87), while the latter is seen in twenty-seven proteins (160/223 = 0.72) (Table 1).

The fourth subtype with HCS = 12 and VS = 10 is identified in twelve proteins from the bacterial PVC superphylum (Planctomycetota and Verrucomicrobiota). Three-quarters of the LRRs (88/117 = 0.75) are well represented by the consensus of LTxLdLxgCx(q/r)(V/l/i) rDL(s/t)PLkGMP (Table 1).

Table 2. Secondary structures of the variable segment (VS) part of CC/SDS22like LRR motifs occurred occasionally in the solved structures of LRR proteins.

No	Protein Name	LRR No. ^a	Position No. ^b	Sequence of Variable Segment (VS) ^c	Secondary structure ^d	PDB_chain	Resol. (Å) ^e
1	Coupling	5	3	TDKGI I ALHH F RN	-HHHHHHGGG-TT	3E4G_A	0.96
2	FBXL17	12	7	TSKG V IHLTK L RN	-HHHHGGGGG-TT	6W66_B	3.21
3	LRRC8	15	7	KLIV L NS L KK M VN	PPB-TTGGGG-TT	6FNW_A	1.80
4	Coupling	5	1	MSIG F DHMEGLQY	-GGGGGGGTT-S-	3E4G_A	0.96
5	InlK	7	4	DDSI L TKINT L PK	BGGGGGGTTT-SS	4L3F_A	2.39
6	InlK	7	3	TSDK I PN L SG L TN	EGGGS---TT-TT	4L3F_A	2.39
7	lmo2027	9	5	TSDD V PD F HT I PK	-GGG-----S-TT	5KZS_A	2.30
8	lmo2470	9	5	TSDE L PN F NN I PL	BGGG-----SS-TT	5HL3_A	1.40
9	LMOF2365_1397	7	3	TSDK I PN L SG L TS	BGGGSPP-TT-TT	4EZG_A	1.50
10	LMOF2365_1397	7	4	DDSI L TKINT L PK	BGGGHHHHTT-SS	4EZG_A	1.50

Note: (a) The number of repeat units in LRR domains; (b) The position number of a CC/SDS22-like LRR motif in LRR domain; (c) Bold uppercase letters indicate highly conserved hydrophobic residues of a CC/SDS22-like LRR; (d) The program of DSSP-PPII was used [34]. One letter code represents a specific conformation; 3(10) – helix (G), PPII (P), extended β -strand in parallel and or anti-parallel β -sheet conformation (E), isolated β -strand (B), turn (T), bend (S) and coil ('-'); (e) "Resol." indicates the resolution in x-ray crystallography.

3.2. Structures of the New Types of LRR Motifs

3.2.1. Secondary Structures of CC/SDS22-Like and Short CC/SDS22-Like LRRs

In the solved structures, ten CC/SDS22-like LRRs occur irregularly (Table 2), while short CC/SDS22-like and short SDS22-like LRRs were not found. The secondary structure assignment shows that the VS parts are divided into three types. The first type is observed in "-HHHHHHGGG-TT" (in the 3rd repeating units in bovine mitochondrial factor B) which consists of an eight-residue α -helix adjoining a five-residue 3(10)-helix and a type I β -turn. The former helix part is a composite helix made up of α - and 3(10)-helices (called $\alpha/3(10)$ composite) [20]. It is also "-HHHHGGGGG-TT" (in the 7th repeating unit in FBXL17) which consists of a six-residue α -helix adjoining a seven-residue 3(10)-helix and a type I β -turn (Table 2). The β -turns constitute a part of the descending loop. The CC/SDS22-like LRRs are characterized by a quirky SSS consisting of β -strand – ascending loop – $\alpha/3(10)$ composite – descending loop and β -strand (β – $\alpha/3(10)$ – β) motif.

Table 3. Secondary structures of CC/SDS22-like, short CC/SDS22-like LRRs, and short SDS22-like LRRs in eleven representative proteins in the predicted structures by AlphaFold.

(1) CC/SDS22-like LRR

LRR	Sequence	Secondary structure ^a	SS ^b	UniProtKB
Subtype 1	LxxLxLxxCxxI TDxGLxHLxxLxx			A0A6C0H0T7
1	IDNLVIKVPDIT NDELNQLLSKLNRNK	EEEEEE--TTPP HHHHHHHHTB-TTT	α	
2	LQTLNLHGCGQI TNAGLAHLASLTS	--EEE-TT-TT- BHHHHHGGGG-TT	$\alpha/3(10)$	
3	LQNLDLTRCYQI TDVGLAHLASLTS	--EEE-TT-TB- BHHHHHTTTT-TT	α	
4	LQNLDLTRCYLI TNAGLVHLGNLTS	--EEE-TT-TB- -HHHHGGGT-TT	$\alpha/3(10)$	
5	LQTLWLVCNKI TDVGLVHLSHLTS	--EEE-TT-TT- BHHHHGGGT-TT	$\alpha/3(10)$	
6	LETLYLSRCDQI TDVGLGHLGNLTS	--EEE-TT-TB- -HHHHHTTTT-TT	α	
7	LQHLSLAWCFRI TDAGLEHLTNLRS	--EEE-TT-TT- -HHHHHTTTT-TT	α	
8	LTTFNLAGCFRI TDAGLRYLTNLTS	--EEE-TT-TT- -HHHHGGGG-TT	$\alpha/3(10)$	
9	LQTLDLFVCYQI TDAGLGHLTNLRS	--EEE-TT-TT- -HHHHHTTTT-TT	α	
10	LTTFNLAGCFRI TDEGVAQLRESLP	--EEE-TT-TTP PHHHHHHHHHST	α	
Subtype 1	LxxLxLxxCxxL TDxxLxHLxPLxx			Q6MEE0
1	IEALNFSKNAHL TDAQLLTLKNCKN	--EEE-TT-TT- BHHHHHGGGG-TT	$\alpha/3(10)$	
2	LKVLQQLQACHNF TDAGLAHLTPLMA	--EEE-TT--B- -HHHHHTTTT-TT	α	
3	LQHLNLSYCKNL TDAGLAHLAPLVV	--EEE-TT-TT- -HHHHHTTTT-TT	α	
4	LQHLNLSSSHNL TDAGLAHLTPLVA	--EEE-TT-TT- -HHHHHTTTT-TT	α	
5	LTHLNLSWCNKL TDAGLAHLTPLVA	--EEE-TT-TT- -HHHHHTTTT-TT	α	
6	LTHLDLRECDKL TNRGLAHLALLT	--EEE-TT-TT- -HHHHHHHTT-TT	α	
7	LQYLDLNYCRNL TDAGLAHLSSLVA	--EEE-TT-TT- -HHHHHTTTT-TT	α	
8	LQHLKLCCCVSL TDAGLAHLAPLVA	--EEE-TT-TT- -HHHHHTTTT-TT	α	
9	LTHLDLSWCFNI TDAGLAHLTPLVT	-SEEE-TT-TT- -HHHHHHHTT-TT	α	
10	LQHLGLSGCRRL TDVGLAHLTRLVA	--EEE-TT-TT- -HHHHHHHTT-TT	α	
11	LQHLGLNRCRDL TDAGLAHLTPLIN	--EEE-TT-TT- -HHHHHTTTT-TT	α	
12	LQHLDLSECRKL TNAGLAHLTPLVA	--EEE-TT-TT- -TTGGGGGT-TT	3(10)	
13	LQRLDLRCCNKL TGARLAHFKFLVA	--EEE-TT-TT- -HHHHHHHHHTT	α	
Subtype 2	LxxLxLxxTxI TDxGLxxLxxLxx			A0A355G5V6
1	ISQVSFSGSKL VDAGLVYLGRLSK	EEEE--TTS-- -GGGGGGGT-TT	3(10)	
2	LRKDLDSGSKV TDDGMVHLKSLKS	--EEE-TTS-- -HHHHGGGT-TT	$\alpha/3(10)$	
3	LREITHGIPV SDTGLGEFKLSN	--EEE-TTS- -BHHHHGGGG-TT	$\alpha/3(10)$	
4	LEILNLSRTKV TDAGLKHLSLDS	--EEE-TTSB- BHHHHGGGT-TT	$\alpha/3(10)$	
5	LKELFTGLEI SADGLAQLSGLKS	--EEE-TT-B- -HHHHHGGG-TT	$\alpha/3(10)$	
6	LETGLSETQI TDEALAHKTLKK	--EEE-TTS-- -HHHHGGGG-TT	$\alpha/3(10)$	
7	LRVLLLRDTQI TDEGLKQIKSLTR	--EEE-TTS-- -HHHHHHHTT-TT	α	
8	LQRLWLRNTQI TDEGLKYLAKMKD	--EEE-TTS-- BHHHHHTTTT-TT	α	
9	MEWLELNDTKI TNAGISEIKVLEN	--EEE-TTSB- BHHHHHHHTT-TT	α	
10	IVDMNLRNTDV NDKCISSLKMMKN	--EEE-TTSB- BGGGHHHHHTT-TT	3(10) α	
11	LGTLYIDGTEI TEDGIAKLEKSLP	--EEE-TTSBP PHHHHHHHHH-T	α	
Subtype 2	LTxLxLxxTxV TDxGLKELxxLxx			A0A225EG43
1	VIGADLRDAPV TDAGLKELAPFQH	EEEE--TTS-- -HHHHGGGT-TT	$\alpha/3(10)$	
2	LTTDLNSTRV TDAGLKDVASFRH	--EEE-TTS-- -HHHHHHHTT-TT	α	
3	LTTGLSGTKV TDAGLKELAPLQH	-SEEE-TTS-- -HHHHGGGT-TT	$\alpha/3(10)$	
4	LTTDLVHTPV TDAGVTHLAALKN	--EEE-TT--- -HHHHHHHTT-TT	α	
5	LTTNLNGITPV TDAGVKHLAALPH	--EEE-TTS-- -HHHHHHHTT-TT	α	
6	LTTGLSRTRV TGVLKELATLEQ	--EEE-TTS-- -HHHHHHHTT-TT	α	
7	LTTDLDSFTKV TAPSLKELGTLRH	--EEE-TTS-- BGGGGGGGG-TT	3(10)	
8	LTSLDLYSTRV TDDGLKELGTLQR	--EEE-TTSB- BHHHHHHHTT-TT	α	
9	LNTLDLSHTDV TGTGLKGLAPLQD	--EEE-TTSB- -TTGGGGGT-TT	3(10)	
10	LTILTFDTPV TDAGLKELAPLQN	--EEE-TT--- BHHHHGGGT-TT	$\alpha/3(10)$	
11	LISLCLNSTRV TDAGMKEVAALKN	--EEE-TTSB- -HHHHHHHTT-TT	α	
12	LTSNLNINTGV TGAGMKELAGLPQ	--EEE-TTSS- BHHHHHHHTT-TT	α	
13	LANLYLDRTRV TDAGAKELAALTQ	--EEE-TTSB- -HHHHHHHTT-TT	α	
14	LTALQLQNTQL TDAGLRELAALKK	-SEEE-TTS-- BHHHHHHHTT-TT	α	
15	LSRLYLSGTKI SDAGAKELAAFRQ	-SEEE-TTSB- BHHHHHGGG-TT	$\alpha/3(10)$	
16	LAILDQLNTQV TDAGLKDLQKALP	-SEEE-TTSBP PHHHHHHHHHST	α	

continued.

Subtype 3	LxxLxLxxTxV SDxGLxxFxxCxx			A0A925DAG5
1	LNTVILEGNQQVSDEGLASFKDCKN	EEEE- TT- TT- -HHHHGGGT- -	$\alpha/3(10)$	
2	LVYLNLFGTKV TDGGLAHFKDCKD	-SEEE- TTS- - -HHHHGGGT- TT	$\alpha/3(10)$	
3	LRELYLSGTTT SDAGLAHFKDCKN	--EEE- TTS- - -HHHHGGGT- TT	$\alpha/3(10)$	
4	LSVLLLSETQV GDAGIGAFKDLTK	-SEEE- TT- S- BHHHHGGGT- -	$\alpha/3(10)$	
5	LTQLNLVRTRV SDAGLAAFKDCKN	--EEE- TT- B- -TTGGGGGT- TT	3(10)	
6	LTFLNLFGTQV SDAGLANFKDCEN	-SEEE- TTS- - -HHHHGGGT- TT	$\alpha/3(10)$	
7	LTSLLLLYTQV SDKGLAHFKNCKK	-SEEE- TTS- - -HHHHGGGT- TT	$\alpha/3(10)$	
8	LQSLNLSATQV SDAGLVHFKDCKN	--EEE- TTSS- BHHHHTTTT- TT	α	
9	LLELVLNDTKL SDAGLAHLQNCKK	-SEEE- TT- B- -HHHHGGGT- TT	$\alpha/3(10)$	
10	VTVLSLINTPV SDAGLAHFKECKN	-SEEE- TT- - -HHHHHTTT- TT	α	
11	LAHLFLQGTQV SGAGLAHFKDYPN	--EEE- TT- - -BGGGGGGGG- TT	$\alpha/3(10)$	
12	LSVLNLDGTL SDAGLAHLVDCKK	--EEE- TTSS- BHHHHTTTT- TT	α	
13	LTHLTLTKEV TTAGIDELKKALP	--EEEETTSS- -HHHHHHHHH- T	α	
FBXL14	LxxLxLxxCxzV xxxLxxLxxLxx			A0A178WET7
1	IQDLYLGEYPGV NDDWMDVISSQSTS	--EEE- TT- TT- BGGGHHHHHTTTT	3(10)/ α	
2	LLSVDFSGSD I TDSGLVSLKGCTN	--EEE- TT -B- -HHHHHGGG- TT	$\alpha/3(10)$	
3	LESNLFNFCQI SNRGLVHLSGLSN	--EEE- TT- TT- -HHHHGGGG- TT	$\alpha/3(10)$	
4	LTSLSFRRNAI TAQGMRLSNLVN	--EEE- TT- SS- BTGGGGGGG- TT	3(10)	
5	LKKLDLEKCPGI DGGLVHLRALTK	--EEE- TT- TB- TTGGGGGT- TT	3(10)	
6	LESNLKWCNCI TDADMEPLSVLTN	--EEEEET- TT- -HHHHGGGG- TT	$\alpha/3(10)$	
7	LRSLQCCSK I TDIGISYLGKLNK	--EEEEES- - - BHHHHHTTTT- TT	α	
8	LNLLNLEGCRHV TAACLDLTALAG	--EEE- TT- TB- -GGGHHHTT- TT	3(10)/ α	
9	LMYLNLRNCN F SDSGCEKFSDLIN	--EEE- TTS- - -STGGGGGG- TT	3(10)	
10	LKILNLGMNN I TNSCLVHLKGLTK	-SEEE- -SS- - BHHHHGGGT- TT	$\alpha/3(10)$	
11	LESNLNDSR I GDEGLVHLSGMLE	--EEE- TTSS- - BHHHHGGGT- TT	$\alpha/3(10)$	
12	LKSLELSDTE V GSNGLRHLSGLSN	--EEE- TTSS- - BTGGGGGT- TT	$\alpha/3(10)$	
13	LESINLSFTV V TDSGLRKLSGLTS	--EEE- TTSS- - -HHHHHTTT- TT	α	
14	LRTLNLDAH V TDAGLSALTSLTG	--EEEEET- TT - BHHHHGGGT- TT	$\alpha/3(10)$	
15	LTHLDLFGAR I TDSGTNHLRLNKK	--EEEEET -B- -HHHHGGGG- TT	$\alpha/3(10)$	
16	LQSLEICGGG L TDGVKNIKDLSS	-SEEE- -S S- - BHHHHHTTTT- TT	α	
17	LTLNLNSQNSNL TDKTLELISGLTG	-SEEE- TT- TB- -HHHHHTTTT- TT	α	
18	LVSLNVNSR V SSSGLRHLKPLKN	-SEEE- -S S- - BTGGGGGT- TT	3(10)	
19	LRSLTLESCK L SANDIRKLQATDL	--EEE- TT SB- -HHHHHHHHHT-	α	

continued.

The second type is observed in six LRRs (the first repeating units in bovine mitochondrial factor B, the third and fourth units in InlK, the 5th units of lmo2027 and lmo2470, and the third units of LMO2365_1397) (Table 2). The VS parts consist of a five, eight or nine-residue 3(10)-helix and two tandem β -turns constituting a part of the descending loop. These CC/SDS22-like LRRs are characterized by a unique SSS consisting of β -strand - ascending loop - 3(10)-helix - descending loop - β -strand (β - 3(10) - β) motif, as seen in SDS22-like, PS, and Leptospira-like LRRs [6].

The third type is observed in only one LRR (the third and fourth units of LMO2365_1397). The VS part consists of a five-residue 3(10)-helix adjoining a six-residue α -helix and two tandem β -turns (Table 2). This CC/SDS22-like LRR is characterized by a quirky SSS consisting of β -strand – ascending loop – 3(10)-helix adjoining α -helix – descending loop – β -strand (β – 3(10)/ α – β) motif

We analyzed AlphaFold structures of CC/SDS22-like LRR domains in fifteen representative proteins (Figure 2, Table 3, Tables S2 and Table S3). The parallel β -strand stacking in the HCS parts produces a solenoid structure in all representative proteins (Figs. 2 and 3). The conserved hydrophobic residues participate in the hydrophobic core.

(2) Short CC/SDS22-like LRR

LRR	Sequence	Secondary Structure ^a	SS ^b	UniProtKB
	LxxLxLxxCxxI TDxGIxxIxxMx			A0A9E5EXX2
1	LSDEIEDEDDI EDDVAVFARFPEKFPN	-TT-B---GGGH HHHHHHTTHHHH-TT		
2	LVLDLDSHCD I SDDFLKQIAKMK	--EEE-TT --- -HHHHHHHTTS-	α	
3	LQTINLIGCKQI TDEGIKHLKDMF	-SEEE-TT-TT- BHHHHHTTTTS-	α	
4	LRSLYLSGCVQI TDAGMEHLSDMQ	-SEEE-TT-TB- -HHHHHHHTT--	α	
5	LEKLVLERCDQI TDEGIKHLKDMF	-SEEE-TT-TT- -HHHHHTTTT--	α	
6	MRHMLSNCTKI TDEGLKSLSGMP	-SEEE-TT-TT- -HHHHHHHTT--	α	
7	LQVLGLRGCNQI TDEGITNIKHLR	-SEEE-TT-TT- -HHHHHHHTTS-	α	
8	LISLSLIGCGEL TDESFKFIKLP	-SEEE-TT-TT- -GGGHHHTTS-	α	
9	LRSLYLGGCQOI TDEALNHIKDLQ	-SEEE-TT-TT- -HHHHHHHTTS-	3(10)/ α	
10	LIDLNLSTCYRI SSTRGLAYLSKMP	-SEEE-TT-TTS -TTHHHHTTTTS-	α	
11	LRSLDLSSMGWKITNESLAIISHLF	-SEEE-TTS-TTTSBHHHHHHHTTS-	α	
12	LQYLNLTESI TDQGLVHLSKMP	-SEEE--SEE E EHHHHGGGTTS-	α	
13	LQSLILALCG I SDEGLAHLSCLR	-SEEE-TTEE - -HHHHGGGTTS-	α /3(10)	
14	LLHDLSECWNI DSGGFAHISHAQ	-SEEE-TT-TT- -SGGGGGTTTS-	α /3(10)	
15	LQSLVVRDCG FDERIFEIFPDLAGVLRI	-SEEE-TT-S S-TTHHHH-HHHHTTB--		

(3) Short SDS22-like LRR

LRR	Sequence	Secondary Structure ^a	SS ^b	UnProtKB
Subtype 1	LxxLxLxxTxL xxLxxLxxMx			A0A517WG16
1	IISVNLNFMNV ENISALKGLE	EEEE-TTS-- ---GGGT--	3(10)	
2	LEYLDLTNCPV SDLSPIKGMK	-SEEE-TT--- ---GGGT--	3(10)	
3	LEQLFLEGTYV TDLRPLQGMP	-SEEE-TT--- ---GGGT--	3(10)	
4	LQVLMEHTPV SDIAPLEGMP	-SEEE-TT--- ---GGGT--	3(10)	
5	LNQLNLFDTKV KDLSLINTLP	-SEEE-TT-S- ---TTGGGS-	3(10)	
6	LRTLWIPKTEI SDISPLKGM	-SEEE-TTS-- ---GGGT--	3(10)	
7	LESLDPEKTKV SDLTPLRGMQ	-SEEE-TTS-- ---GGGT--	3(10)	
8	ILRLNMTGSEV TDLTPLKGM	-SEEE-TTSS- -B-GGGT--	3(10)	
9	LQRLIFTPSKI TKGIEVIRDSP	-SEEE--GGG- -B-HHHHT-S	α	
Subtype 2	LxxLxLxxNxL xxLxxLxxMx			A0A351WLM1
1	GISLEIPNAGI ADLEPLRGIP	EEEE-TTS-- ---GGGT--	3(10)	
2	LTSLVCGWNR I RSLEPLCGMP	EEEE-TTS-- ---GGGT--	3(10)	
3	LVELVCSGNPI ETLEPLRGMP	-SEEE-TTS-- ---GGGT--	3(10)	
4	LKTLRCERTGI RTLDALKGMQ	-SEEE-TT-S- ---GGGT--	3(10)	
5	LTMFSCGENKLDGLEALAGMP	-SEEE-TT---SSTGGGT--	3(10)	
6	LTWLSICCCGL TTLSPLRGLP	-SEEE-TTS-- ---GGGT--	3(10)	
7	LNRLFCDGNSI EDLSPLKRLP	-SEEE-TTS-- ---GGGT--	3(10)	
8	LIEFSLAGNRV KDLESIRGAS	-SEEE--SS-- ---GGGT--	3(10)	
9	LNSLQCGSNLI ESLEPLRGMP	-SEEE--SS-- ---GGGT--	3(10)	
10	LSTLFCMNNQI RNLEPLRGMP	-SEEE--SS-- ---GGGT--	3(10)	
11	LNALICGNPL KDIGPFVKNP	-SEEE--SS-- S--GGGTSP	3(10)	
Subtype 3	LxxLxCxxTxL xxLxxLxxMx			A0A354GSR5
1	LVTFCYGTGPV SDLSPLKDMF	-SEEE-TTS-- ---GGGT--	3(10)	
2	LTYLHCDDTQV SDLSSLRGMK	-SEEE-TTSS- ---GGGT--	3(10)	
3	LESLDSCGTAV SDLSPLKDMF	-SEEE-TTS-- ---GGGT--	3(10)	
4	LTRLSCGGTQI TDLSPKDMF	-SEEE-TTS-- ---GGGT--	3(10)	
5	LTYLNCGGTKV SDLSPLKGMK	-SEEE-TTS-- ---GGGT--	3(10)	
6	LDMLLCNTLV SDLSLLKDMF	-SEEE-TTS-- ---GGGGGS-	3(10)	
7	LKELFCDFKPE RDADLLRSIK	-SEEE---HH HHHHHHHH-T	α	
Subtype 4	LxxLxLxxCxxL xxLxxLxxMx			A0A518E2T1
1	LTAIHLQQPGL VDLSPKKFDLSA	EEEE-TT-TT- ---GGGGGS--TT	3(10)	
2	VTSVSLYRATNV SDLSPLKGCP	--EEE-TT-TT- ---GGGT--	3(10)	
3	LTTLNLERCGKI QDLSPLKGMP	-SEEE-TT-TT- ---GGGT--	3(10)	
4	LTWFRMYACGPV GDLSPLKGMP	-SEEE-TT-TT- ---GGGT--	3(10)	
5	LEHLDIGLNDQI DDLSPLARMP	-SEEE-TT-SS- ---GGGT--	3(10)	
6	LHDLRLDSCPQV TDIGVLRGMP	-SEEE-TT-TT- ---GGGT--	3(10)	
7	LDFLSLFACPAI QDYSPLAELP	-SEEE-TT-TT- ---GGGGGS-	3(10)	
8	LETLYFTPSLLS PAEIQQVEAI	--EEE--GGGS- HHHHHHHHT-	α	

Note. (a) The program of DSSP-PPII was used [34]. One letter code represents a specific conformation; α -helix (G), 3(10)-helix (G), extended β -strand in parallel and or anti-parallel β -sheet conformation (E), isolated β -strand (B), turn (T), bend (S) and coil ('-'); (b) Secondary structure of the variable segment of LRR. (b) Secondary structures of the variable segment of LRR. " α " shows α - helix. "3(10)" shows 3(10)-helix. " α /3(10)" shows a composite helix made up of contiguous α - and 3(10)-helices. "3(10)/ α " shows a composite helix made up of contiguous 3(10)- and α -helices.

The above-mentioned three SSSs observed in the known structures were observed. In addition, a typical SSS of the $\beta - \alpha - \beta$ motif was observed, as seen in many proteins such as ribonuclease inhibitor, FBXL17, and LegL1 containing RI-like, CC, and GALA LRRs [6]. The occurrence frequency of the $\beta - \alpha - \beta$ and $\beta - 3(10)/\alpha - \beta$ motifs is much higher than the $\beta - 3(10) - \beta$ and $\beta - 3(10)/\alpha - \beta$ motifs.

Here we describe three examples in more detail. In a viral protein (UniProtKB: A0A6C0H0T7) (Figure 2A, Table 3), the secondary structure of eight complete CC/SDS22-like LRRs (subtype 1) in the VS parts is grouped into two types. The first type is “BHHHHHTTTT-TT” (LRR3) or “-HHHHHTTTT-TT” (LRR6, 7 and 9) which adopt a seven-residue α -helix and multiple β -turns constituting a part of the descending loop. It is clearly characterized by an SSS of $\beta - \alpha - \beta$ motif. The second type is “BHHHHHGGGG-TT” (LRR2 and LRR5) or “-HHHHHGGGT-TT (LRR4 and LRR8)” which consists of a six-residue α -helix adjoining a five-residue 3(10)-helix and one or two tandem β -turns. It is characterized by an SSS of $\beta - \alpha/3(10) - \beta$ motif.

A bacterial protein (A0A925DAG5) contains thirteen CC/SDS22-like LRRs (subtype 3) (Figure 2F, Table 3A). The secondary assignments indicate that the VS parts adopt a six-residue α -helix adjoining a five-residue 3(10)-helix and two β -turns in LRR1-4, 6, 7, 9, and 11, a six-residue α -helix and two β -turns in LRR8, 10, and 11, and a seven-residue 3(10)-helix and two β -turns only in LRR5. Thus, the $\beta - \alpha/3(10) - \beta$, $\beta - \alpha - \beta$, and $\beta - 3(10) - \beta$ motifs occur eight, three, and one times, respectively. The consecutive conserved phenylalanine at position 8 in the VS part forms a phenylalanine spine which is a hydrophobic spine, as seen in LRR proteins such as Nogo receptor, TLR3, LGI1, and Crov588 [11,37–39].

A bacterial protein (A0A9E5EXX2) contains fifteen LRRs of which most are short CC/SDS22-like LRR (Figure 2G, Table 3A). The $\beta - \alpha - \beta$ motif occurs nine times, while the $\beta - \alpha/3(10) - \beta$, $\beta - 3(10)/\alpha - \beta$, and $\beta - 3(10) - \beta$ motifs occur once each.

3.2.2. Structures of short SDS22-like LRR

We analyzed AlphaFold structures in eight representative proteins (Figure 3, Table 3B, Tables S2 and S3). The secondary structure assignment indicates that the sequence of xxLxxLxxMx in the VS part is mostly assigned as “---GGGT---” in all four subtypes. The VS parts adopt a five-residue 3(10)-helix and a type I β -turn. The short SDS22-like LRR domains consist of consecutive $\beta - 3(10) - \beta$ motifs, as seen in SDS22-like LRR domains in many proteins such as internalin and lmo2027 [6].

3.3. HELFIT Analysis

The HELFIT analysis shows that CC/SDS22-like LRR domains have the helical parameters of $R = 18.53 \pm 0.57$ (Å), $\Delta z = 1.44 \pm 0.24$ (Å), and $\Delta\varphi = 14.95 \pm 0.30$ (°) and short SDS22-like LRR domains have those of $R = 18.46 \pm 1.19$ (Å), $\Delta z = 1.84 \pm 0.39$ (Å), and $\Delta\varphi = 14.62 \pm 1.11$ (°) (Table S3). Typical SDS22-like LRR domains show $R = 18.00 \pm 1.54$ (Å), $\Delta z = 2.23 \pm 0.47$ (Å), and $\Delta\varphi = 14.20 \pm 0.89$ (°) in eleven known structures [6], while CC LRR domains show $R = 16.37 \pm 0.88$ Å, $\Delta z = 1.12 \pm 0.41$ (Å), and $\Delta\varphi = 16.37 \pm 0.88$ (please check the numbers) (°) in seven known structures and typical RI-like LRR domains $R = 17.92 \pm 1.04$ (Å), $\Delta z = 0.93 \pm 0.31$ (Å), and $\Delta\varphi = 16.37 \pm 0.88$ (°) in twelve known structures [7]. It appears that the Δz parameter decreases in the order of typical SDS22-like > short SDS22-like > CC/SDS22-like > CC > RI-like. The comparison of the helical parameters with the known structures indicates that the AlphaFold structures are fully reasonable.

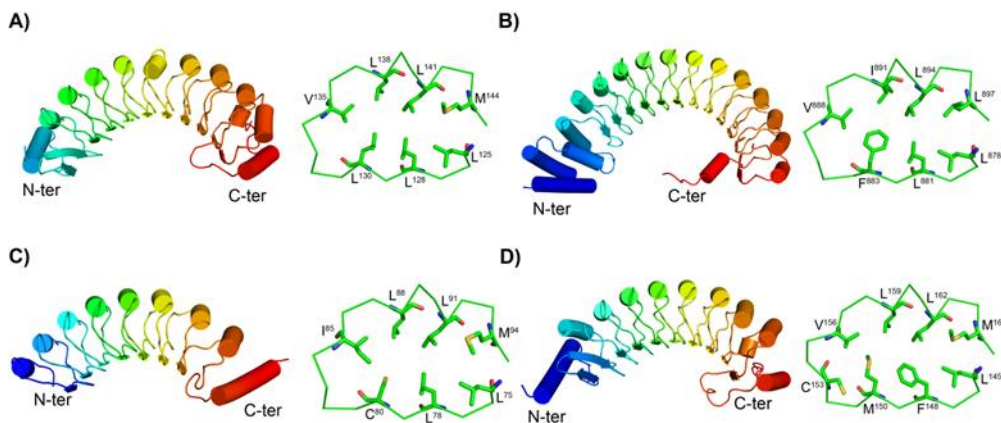


Figure 3. Structures of short SDS22-like LRR domains in four representative proteins. The left side shows the overall structure of the LRR domains. The right side shows a single LRR unit to clearly specify hydrophobic core from the conserved residues. (A) Internalin-A from *Gimesia chilikensis* (A0A517WG16). (B) A2017_15240 from *Lentisphaerae bacterium* GWF2_44_16 (A0A1G0YWT4). (C) A leucine-rich repeat domain-containing protein from *Planctomycetales bacterium* (A0A354GSR5). (D) Pla8534_62660 from *Lignipirellula cremea* (A0A518E2T1). α -Helices or 3(10)-helices are shown as thick tubes, β -sheets as arrows. The sequence alignments of LRRs in these proteins are shown in Table S3.

4. Discussion

4.1. A novel SSS of β - 3(10) - β Motif in Short SDS22-like LRR

The sequence search in the present study indicates that over 300 proteins contain a novel type of short SDS22-like LRR (Table 1, Table S1). The AlphaFold structures indicate that short SDS22-like LRRs adopt a five-residue 3(10)-helix in the VS part (Table 3, Table S2), as seen in SDS22-like, Leptospira-like, and PS LRRs [5]. For the first time we recognized that these LRR domains consist of consecutive β - 3(10) - β motif (Figure 1). It appears that the flexibilities of the ascending loop and descending loop provide the common structure-preservation. The high flexibility of the ascending loop in the β - α - β motif in RI-like LRRs has been observed [7]. The stability of β - 3(10) - β motif may be less than that of β - α - β motif.

4.2. Simultaneous Occurrence of four SSSs in CC/SDS22-like LRR

The sequence search indicates that over 2,200 proteins contain novel types of CC/SDS22-like and short CC/SDS22-like LRRs (Table S1). The present structural analysis indicates that the CC/SDS22-like LRR occurred occasionally in the known structures and the AlphaFold structures indicates that the CC/SDS22-like and short CC/SDS22-like LRRs simultaneously adopt four SSSs of β - α - β , β - $\alpha/3(10)$ - β , β - 3(10)/ α - β , and β - 3(10) - β ; the latter two SSSs are rare (Tables 2 and 3, Table S2). CC and GALA LRRs as well as RI-like LRRs are characterized by the β - α - β motif (Figure 1) [6]. The conserved aliphatic residue at position 8 in their VS parts adopts α -helix conformation. The conserved aliphatic residue plays a crucial role to preserve two parallel β -strands [6]. Position 8 in the VS part of CC/SDS22-like and short CC/SDS22-like LRRs is also occupied by aliphatic or aromatic residues (Table 2, Table S3). The conserved residues adopt α -helix with high frequency. A dual characteristic of CC and SDS22-like LRRs is reflected in a composite helix of α -helix adjoining 3(10)-helix. The $\alpha/3(10)$ and 3(10)/ α composites produce a kink between the two helices in the known structures [20]. Pal et al., [20] calculated the bending angle which is defined as the axes of the adjacent 3(10)- and α -helices. The average bending angles for the $\alpha/3(10)$ and 3(10)/ α composites were 22 ± 15 ($^\circ$) and 37 ± 19 ($^\circ$), respectively, in the known structures. Chameleon sequences refer to the same sequences accepting different secondary structures in protein structures [40,41]. A BTB domain-containing protein from *Protochlamydia amoebophila* (Q6MEE0) contains thirteen complete CC/SDS22-like LRRs (Table S2). The VS sequences of LRR11

and LRR12 that are TDAGLAHLTPLIN and TNAGLAHLTPLVA, respectively, are almost identical. However, the secondary structure is different from each other. LRR11 adopts 3(10)-helix, while LRR12 adopts α -helix (Table S2). Thus, it appears that the CC/SDS22-like LRRs are a chameleon-like sequence. This property was seen in PS/SDS22-like LRR [18]. Consequently, the dual characteristic and the feature as chameleon-like sequence of CC/SDS22-like LRR would provide simultaneous occurrence of four SSSs.

4.3. Evolutionary Insights

A big protein from *Planctomycetota bacterium* (A0A3L7QJC1) contains two LRR domains with multiple other domains (Figure 4O). The LRR type of individual LRR domains is clearly different; the first domain consists of CC/SDS22-like LRR with RUL = 24, while the second domain consists of short SDS22-like LRR with RUL = 21. Two LRR proteins from *P. bacterium* (A0A932SUJ7) and *Limnoglobus roseus* (A0A5C1AHQ6) are homologs, which contain protein kinase and FGE-sulfatase. However, surprisingly, the LRR type in the two respective LRR domains is clearly different from each other; one is CC/SDS22-like LRR and the other is short SDS22-like LRR. These observations suggest that CC/SDS22-like and short SDS22-like LRRs rapidly evolved from a common ancestor.

4.4. Implications of Multiple Functions

So far, little is known about the functions of proteins containing CC/SDS22-like LRR and short SDS22-like LRR. These proteins are linked to many diverse domains with high frequency (Table S1). One of them is the F-box domain. FBXLs are present in eukaryotes, prokaryotes, and even viruses [42]. FBXLs function as part of the SCF ubiquitin ligase complex that confers ubiquitination and degradation. Protein domain search indicates that many CC/SDS22-like LRR proteins are FBXLs. The FBXLs identified here might be implicated in the functions.

The structures of five human FBXLs (FBXL1, 2, 3, 5, and 17) are available [43–47]. The LRRs are the CC type or similar to it, in which the VS part adopts mainly α -helix. An NCLDV, Mimiviridae, encodes many FBXLs in which LRR is a short CC type with RUL = 22 [42]. These LRR types clearly differ from CC/SDS22-like LRR in bacterial FBXLs. It appears that this observation confers species specificity of the functions.

Sixty-four CC/SDS22-like LRR proteins from the phylum Chlamydiae, *Candidatus Protochlamydia* that are FBXLs contain BTB-domain at the N-terminus of the F-box domain (Figs. 4A and 4B). The structure revealed that the BTB domain complexed with FBXL17 is related to target degradation signals, i.e., “degrons” in human [26,47]. The CC/SDS22-like LRR proteins might form a degron (i don't understand).

CC/SDS22-like LRR proteins from the phylum Planctomycetes, *Verrucomicrobiaceae bacterium* and the phylum *Proteobacteria*, *Woeseia* sp. contain the VHL domain. The VHL domain acts as the substrate recognition component of an E3 ubiquitin ligase complex [48] (Figure 4C). The LRR proteins might be closely related to ubiquitination.

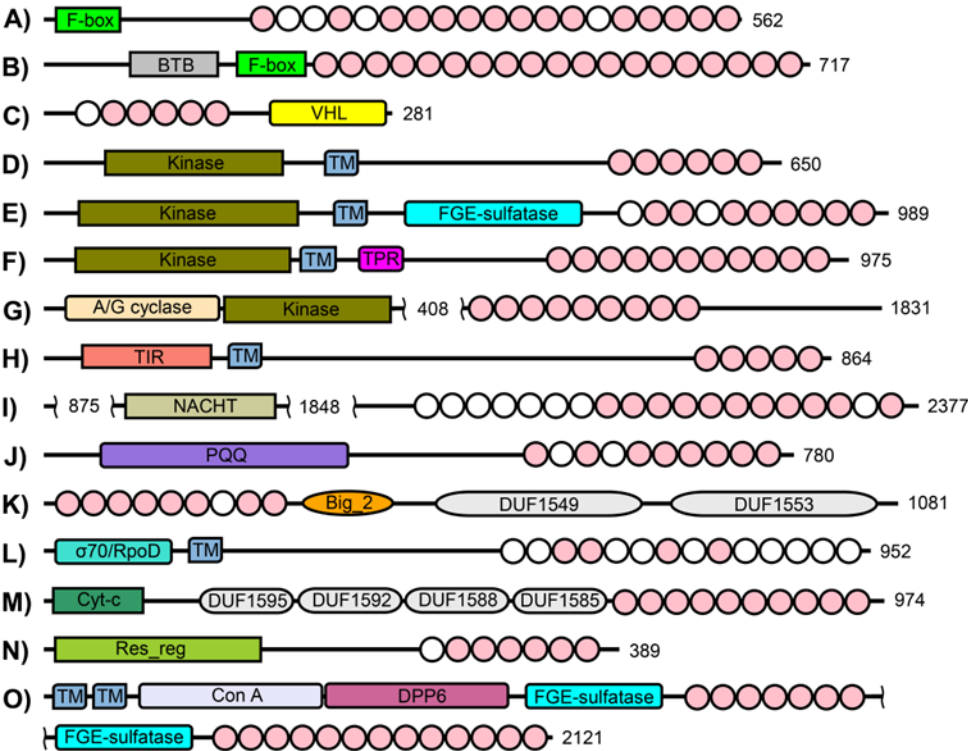


Figure 4. Protein domain architecture of fifteen representative proteins containing CC/SDS22-like LRR proteins (remove?) or short SDS22-like LRR domains and other characteristic domains. (A) PHPALM_29341 from *Phytophthora palmivora* var. *palmivora* (UniProtKB: A0A2P4X7T7). (B) fbxL14_3 from *Candidatus Protochlamydia amoebophila* (A0A0C1H8F1). (C) CMO38_05550 from *Verrucomicrobiaceae bacterium* (A0A2E4AB07). (D) GCJUQL4_47850 from *Gemmataceae bacterium* (A0A4V0I9Z6). (E) PX52LOC_03625 from *Limnoglobus roseus* (A0A5C1AHQ6). (F) prkC_4 from *Planctomycetes bacterium ADurb.Bin12* (A0A1V6D218). (G) DB346_24525 from *Verrucomicrobia bacterium* (A0A2T6CPM7). (H) ENR53_04930 from *Gemmataceae bacterium* (A0A7C4Y7X7). (I) BSAL_51985 from *Bodo saltans* (A0A0S4IL11). (J) CMJ62_18345 from *Planctomycetaceae bacterium* (A0A2E8B5T4). (K) V6x_36850 from *Gimesia chilikensis* (A0A517WFF4). (L) D6739_03220 from *Candidatus Hydrogenedentes bacterium* (A0A7C8AK04). (M) EXR98_13575 from *Gemmataceae bacterium* (A0A964PF13). (N) D6739_03220 from *Nitrospirae bacterium* (A0A3M1R2E1). (O) DWH91_05750 from *Planctomycetota bacterium* (A0A3L7QJC1). Abbreviations: Transmembrane region (TM), von Hippel-Lindau disease tumour suppressor β domain (VHL), Sulfatase-modifying factor enzyme (FGE-sulfatase), Tetratricopeptide repeats (TPR), Guanylate cyclase domain (A/G cyclase), Toll/interleukin-1 receptor domain (TIR), Pyrroloquinoline quinone repeat (PQQ), Bacterial immunoglobulin (Ig)-like domain (Big_2), RNA polymerase sigma-70-like domain (σ/RpoD), Cytochrome c domain (Cyt-c); Response regulator receiver domain (Res_reg), Concanavalin A-like lectin/glucanase (Con A), DPP6 N-terminal domain (DPP6).

Many short SDS22-like LRR proteins and CC/SDS22-like proteins are linked to a protein kinase domain (PK) that confers protein phosphorylation (Figs. 4D, 4E, 4F, and 4G). The domain architecture completely corresponds to that of plant LRR receptor kinases (LRR-RKs). Therefore, it appears that these proteins are regarded as bacterial LRR-PKs. Plant LRR-RKs play critical roles in plant growth, development, defense, reproduction, and symbiosis [49]. Similarly, the bacterial LRR-PKs might contribute to defense and reproduction. The LRR type in plant LRR-RKs is “plant specific”. This confers species specificity of the functions, as seen in FBXLs.

Two transmembrane CC/SDS22-like LRR proteins have a TIR domain (Figure 4H). The protein domain architecture corresponds to vertebrate Toll-like receptors (TLRs) [50]. Eight CC/SDS22-like LRR proteins from the phylum Discoba and *Bodo saltans* contain the NACHT domain (Figure 4I). Bacterial NACHT proteins containing the LRR domain provide immunity against both DNA and RNA phages [51]. Thus, their CC/SDS22-like LRR proteins might be bacterial immunity proteins that

protect bacteria from bacteriophages. The LRR type in TLRs is “Typical”. This also confers species specificity of the functions. CRISPR-Cas is a bacterial immune system that provides adaptive immunity against foreign DNA and RNA [52]. The present observation indicates that there might be a natural immunity system in bacteria.

CC/SDS22-like LRR proteins and short SDS22-like LRR proteins contain various enzymatic domains (Figure 4J) and enzyme-related domains including the PQQ repeat that is a redox coenzyme [53]. The LRR domains in the proteins might regulate their enzymatic activity through protein-protein interactions.

Short SDS22-like LRR proteins from the phylum Planctomycetes contain bacterial Ig-like domain (Big_2) with DUF1549 and DUF1553 (Figure 4K). The listerial proteins, internalin (InlA and InlB) mediate invasion of host cells. The structure of InlB reveals that the central LRR is flanked by an EF-hand-like cap and an Ig-like domain and the LRR type is typical SDS22-like [54]. The short SDS22-like LRR proteins might be closely related to host cell invasion.

RNA polymerase sigma-70-like domain is present in six CC/SDS22-like LRR proteins from *Candidatus hydrogenedentes bacterium*, *Gemmata massiliana*, and *Gemmata sp. SH-PL17* (Figure 4L). The sigma factor has been known to be required for transcription initiation from promoter elements in bacterial RNA polymerase [55]. The LRR domains might play a role for the transcription initiation process.

Cytochrome c-type is present in CC/SDS22-like LRR proteins from the bacterial phyla Planctomycetes and Bacteroidetes, and the phylum Armatimonadetes (Figure 4M). Cytochrome c has a role as electron transfer mediator [56]. The function might be affected by the LRR domain.

Three CC/SDS22-like LRR proteins only from *Nitrospirae bacterium* and *Pseudomonadota bacterium* contain a response regulator receiver domain which receives the signal from the sensor partner in bacterial two-component signal transduction systems [57] (Figure 4N). The LRR domain might affect the signaling. The keyword search detects many homologs. In Response regulatory domain-containing protein of the phylum Viridiplantae, *Mustard* (A0A8S9KMR5), the LRR type is not “CC/SDS22-like” but “Plant specific”. This observation suggests possibly species specificity of the functions.

Three short SDS22-like LRR proteins from the PVC group contain a guanylate cyclase domain. Two of them are linked to PKs (Figure 4G). Fungal adenylate cyclase contains an LRR domain with a cyclase catalytic domain and three other domains [58]. The LRR type is preferentially “Leptospira-like” or “Typical”. This observation confers possibly species specificity of the function.

We suggest that the proteins containing the new LRR types might be involved in bacterial immunity and ubiquitination at least. We insist that the new LRR types confer possibly species specificity of the function. We think that the new LRR types would provide an effective method to prevent, detect, diagnose, treat, and, ultimately, cure cancer. Experimental research is needed.

4.5. The Bacterial PVC Superphylum

The new LRR types are present in over 3,600 proteins. Many of those proteins ($2280/3654 = 0.62$) come from the bacterial PVC superphylum (Table S1). It is composed of the phyla Planctomycetes, Verrucomicrobia, Chlamydiae, Lentisphaerae, and Kiriatiellaeota [59]. The new LRR types might be closely related to the survival of the bacterial PVC superphylum.

5. Conclusion

We identified new LRR types in over 3,600 proteins that mainly come from the bacterial PVC group. The LRR types are classified into three groups consisting of CC/SDS22-like, short CC/SDS22-like, and short SDS22-like LRRs. AlphaFold reveals that the LRR domains form a solenoid structure by the parallel stacking of β -strands as seen in the known structures. The short SDS22-like LRR domains consist of consecutive $\beta - 3(10) - \beta$ motifs. For the first time we recognized the $\beta - 3(10) - \beta$ motif as a unique SSS. The LRR domains of CC/SDS22-like and short CC/SDS22-like types are mainly

characterized by $\beta - \alpha - \beta$ and $\beta - \alpha/3(10) - \beta$ motifs. In addition, the $\beta - 3(10)/\alpha - \beta$ and $\beta - 3(10) - \beta$ motifs are rarely observed. The SSSs of the $\beta - \alpha/3(10) - \beta$ motif with the $\beta - 3(10) - \beta$ and $\beta - 3(10)/\alpha - \beta$ motifs are also observed in CC/SDS22-like LRRs that occur occasionally in the known structures. The dual characteristic of CC and SDS22-like LRRs and the feature as chameleon like sequence would provide simultaneous occurrence of four SSSs. The SSSs of the $\beta - 3(10) - \beta$ and $\beta - \alpha/3(10) - \beta$ motifs are likely fundamental motifs in protein structure as well as typical $\beta - \alpha - \beta$ motif and that the flexibility of the ascending loop and descending loop in the LRR structures provides the common structure-preservation. Many proteins containing the new types of LRR from the bacterial PVC superphylum are linked to an F-box domain, kinase domain or TIR domain. These proteins thus might be involved in ubiquitination and bacterial immunity. It appears that the novel LRR types identified confer possibly species specificity of the functions. The sequence analysis suggests that LRRs evolved very rapidly.

Supplementary Materials: The following supporting information can be downloaded at: Preprints.org, Table S1 : Leucine rich repeat (LRR) proteins identified here ; Table S2. Secondary structures of CC/SDS22-like, short CC/SDS22-like and short SDS22-like LRRs in twenty-six representative proteins in the predicted structures by AlphaFold; Table S3: Helix parameters of CC/SDS22-like LRR domains, short CC/SDS22-like LRR domains, and short SDS22-like LRR domains in twenty-six representative proteins shown in Table 3 and Table S2.

Author Contributions: N.M.: Conceptualization, Investigation, Data Curation, Formal Analysis, Writing - Original Draft Preparation, Review & Editing. D.B.: Writing - Review, Methodology, Data Curation, Analysis. P.E.: Methodology, Visualization, Analysis. All authors have read and agreed to the published version of the manuscript.

Funding: D.B. was supported by Grant-in-Aid for Postdoctoral FELLOWSHIP from Mongolian National University of Education (Funder ID: 100020678, Grant No. MNUE2024F002).

Data Availability Statement: The sequence data that support the findings of this study are available from the corresponding author. All data associated with this study are present in the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

LRR	Leucine-rich repeat
RUL	Repeating unit length of LRR
HCS	Highly conserved segment of LRR repeating unit
VS	Variable segment of LRR repeating unit
CC	Cysteine-containing
PS	Plant specific
SSS	Super-secondary structure
FBXL	F-box/LRR-repeat protein
PK	Protein kinase domain
LRR-RK	LRR receptor kinases
P	Helix pitch
R	Helix radius
N	Number of repeats per turn in helix
Δz	Rise per repeat unit in helix
$\Delta\phi$	Rotation per repeat unit in helix

References

1. UniProt Consortium. UniProt: The universal protein knowledgebase in 2023. *Nucleic Acids Res.* **2023**, *51*, D523–D531.

2. Kobe, B.; Deisenhofer, J. The leucine-rich repeat: A versatile binding motif. *Trends Biochem. Sci.* **1994**, *19*, 415–421.
3. Bella, J.; Hindle, K.L.; McEwan, P.A.; Lovell, S.C. The leucine-rich repeat structure. *Cell. Mol. Life Sci.* **2008**, *65*, 2307–2333.
4. Matsushima, N.; Kretsinger, R.H. *Leucine Rich Repeats: Sequences, Structures, Ligand-Interactions, and Evolution*; Lambert Academic Publishing: Saarbrücken, Germany, **2016**; pp. 1–134.
5. Batkhishig, D.; Bilguun, K.; Enkhbayar, P.; Miyashita, H.; Kretsinger, R.H.; Matsushima, N. Super secondary structure consisting of a polyproline II helix and a β -turn in leucine rich repeats in bacterial type III secretion system effectors. *Protein J.* **2018**, *37*, 223–236.
6. Batkhishig, D.; Enkhbayar, P.; Kretsinger, R.H.; Matsushima, N. A strong correlation between consensus sequences and unique super secondary structures in leucine rich repeats. *Proteins* **2020**, *88*, 840–852.
7. Batkhishig, D.; Enkhbayar, P.; Kretsinger, R.H.; Matsushima, N. A crucial residue in the hydrophobic core of the solenoid structure of leucine rich repeats. *Biochim. Biophys. Acta Proteins Proteom.* **2021**, *1869*, 140631.
8. Kobe, B.; Kajava, A.V. The leucine-rich repeat as a protein recognition motif. *Curr. Opin. Struct. Biol.* **2001**, *11*, 725–732.
9. Kajava, A.V.; Anisimova, M.; Peeters, N. Origin and evolution of GALA-LRR, a new member of the CC-LRR subfamily: From plants to bacteria? *PLoS ONE* **2008**, *3*, e1694.
10. Matsushima, N.; Miyashita, H.; Mikami, T.; Kuroki, Y. A nested leucine rich repeat (LRR) domain: The precursor of LRRs is a ten or eleven residue motif. *BMC Microbiol.* **2010**, *10*, 235.
11. Huyton, T.; Jaiswal, M.; Taxer, W.; Fischer, M.; Görlich, D. Crystal structures of FNIP/FGxxFN motif-containing leucine-rich repeat proteins. *Sci. Rep.* **2022**, *12*, 16430.
12. Matsushima, N.; Takatsuka, S.; Miyashita, H.; Kretsinger, R.H. Leucine rich repeat proteins: Sequences, mutations, structures and diseases. *Protein Pept. Lett.* **2019**, *26*, 108–131.
13. Enkhbayar, P.; Miyashita, H.; Kretsinger, R.H.; Matsushima, N. Helical parameters and correlations of tandem leucine rich repeats in proteins. *J. Proteomics Bioinform.* **2014**, *7*, 139–150.
14. Donohue, J. Hydrogen bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA* **1953**, *39*, 470–478.
15. Enkhbayar, P.; Hikichi, K.; Osaki, M.; Kretsinger, R.H.; Matsushima, N. 3_{10} -helices in proteins are parahelices. *Proteins* **2006**, *64*, 691–699.
16. Matsushima, N.; Kretsinger, R.H. Numerous variants of leucine rich repeats in proteins from nucleocytoplasmic large DNA viruses. *Gene* **2022**, *817*, 146156.
17. Klein, S.A.; Majumdar, A.; Barrick, D. A second backbone: The contribution of a buried asparagine ladder to the global and local stability of a leucine-rich repeat protein. *Biochemistry* **2019**, *58*, 3480–3493.
18. Rudnev, V.R.; Kulikova, L.I.; Nikolsky, K.S.; Malsagova, K.A.; Kopylov, A.T.; Kaysheva, A.L. Current approaches in supersecondary structures investigation. *Int. J. Mol. Sci.* **2021**, *22*, 11879.
19. Super secondary structure. Ditki (Smart Biology). Available online: (accessed 24 January 2026). [CrossRef]
20. Pal, L.; Dasgupta, B.; Chakrabarti, P. 3_{10} -helix adjoining alpha-helix and beta-strand: Sequence and structural features and their conservation. *Biopolymers* **2005**, *78*, 147–162.
21. Matsushima, N.; Batkhishig, D.; Enkhbayar, P.; Kretsinger, R.H. A dual leucine-rich repeat in proteins from the eukaryotic SAR group. *Protein Pept. Lett.* **2023**, *30*, 574–586.
22. Schulz, F.; Roux, S.; Paez-Espino, D.; Jungbluth, S.; Walsh, D.A.; Denev, V.J.; McMahon, K.D.; Konstantinidis, K.T.; Elie-Fadrosh, E.A.; Kyrpides, N.C. Giant virus diversity and host interactions through global metagenomics. *Nature* **2020**, *578*, 432–436.
23. Matsushima, N.; Miyashita, H.; Mikami, T.; Kuroki, Y. A new method for the identification of leucine-rich repeats by incorporating protein second structure prediction. In *Bioinformatics: Genome Bioinformatics and Computational Biology*; Nova Science Publishers: Hauppauge, NY, USA, **2011**; pp. 61–88.
24. Crooks, G.E.; Hon, G.; Chandonia, J.-M.; Brenner, S.E. WebLogo: A sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190.
25. Lee, J.K.; Belogrudov, G.I.; Stroud, R.M. Crystal structure of bovine mitochondrial factor B at 0.96-Å resolution. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 13379–13384.

26. Mena, E.L.; Jevtić, P.; Greber, B.J.; Gee, C.L.; Lew, B.G.; Akopian, D.; Nogales, E.; Kuriyan, J.; Rape, M. Structural basis for dimerization quality control. *Nature* **2020**, *586*, 452–456.
27. Deneka, D.; Sawicka, M.; Lam, A.K.; Paulino, C.; Dutzler, R. Structure of a volume-regulated anion channel of the LRRC8 family. *Nature* **2018**, *558*, 254–259.
28. Neves, D.; Job, V.; Dortet, L.; Cossart, P.; Dessen, A. Structure of internalin InlK from the human pathogen *Listeria monocytogenes*. *J. Mol. Biol.* **2013**, *425*, 4520–4529.
29. Faralla, C.; Bastounis, E.E.; Ortega, F.E.; Light, S.H.; Rizzuto, G.; Gao, L.; Marciano, D.K.; Nocadello, S.; Anderson, W.F.; Robbins, J.R. *Listeria monocytogenes* InlP interacts with afadin and facilitates basement membrane crossing. *PLoS Pathog.* **2018**, *14*, e1007094.
30. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589.
31. Varadi, M.; Bertoni, D.; Magana, P.; Paramval, U.; Pidruchna, I.; Radhakrishnan, M.; Tsenkov, M.; Nair, S.; Mirdita, M.; Yeo, J.; et al. AlphaFold protein structure database in 2024: Providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* **2024**, *52*, D368–D375.
32. Chebrek, R.; Leonard, S.; de Brevern, A.G.; Gelly, J.-C. PolyprOnline: Polyproline helix II and secondary structure assignment database. Database 2014, **2014**, bau102.
33. Hutchinson, E.G.; Thornton, J.M. PROMOTIF - a program to identify and analyze structural motifs in proteins. *Protein Sci.* **1996**, *5*, 212–220.
34. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637.
35. Horn, M.; Collingro, A.; Schmitz-Esser, S.; Beier, C.L.; Purkhold, U.; Fartmann, B.; Brandt, P.; Nyakatura, G.J.; Droege, M.; Frishman, D.; et al. Illuminating the evolutionary history of chlamydiae. *Science* **2004**, *304*, 728–730.
36. Domman, D.; Collingro, A.; Lagkouvardos, I.; Gehre, L.; Weinmaier, T.; Rattei, T.; Subtil, A.; Horn, M. Massive expansion of ubiquitination-related gene families within the Chlamydiae. *Mol. Biol. Evol.* **2014**, *31*, 2890–2904.
37. Choe, J.; Kelker, M.S.; Wilson, I.A. Crystal structure of human Toll-like receptor 3 (TLR3) ectodomain. *Science* **2005**, *309*, 581–585.
38. He, X.L.; Bazan, J.F.; McDermott, G.; et al. Structure of the Nogo receptor ectodomain: A recognition module implicated in myelin inhibition. *Neuron* **2003**, *38*, 177–185.
39. Yamagata, A.; Goto-Ito, S.; Sato, Y.; Shiroshima, T.; Maeda, A.; Watanabe, M.; Saitoh, T.; Maenaka, K.; Terada, T.; Yoshida, T.; et al. Structural insights into modulation and selectivity of transsynaptic neuroligin–LRRTM interaction. *Nat. Commun.* **2018**, *9*, 3964.
40. Ghoulane, A.; Joseph, A.P.; Bornot, A.; de Brevern, A.G. Analysis of protein chameleon sequence characteristics. *Bioinformatics* **2009**, *3*, 367–369.
41. Roterman, I.; Slupina, M.; Stapor, K.; Konieczny, L.; Gądek, K.; Nowakowski, P. Chameleon sequences - Structural effects in proteins characterized by hydrophobicity disorder. *ACS Omega* **2024**, *9*, 38506–38522.
42. Matsushima, N.; Miyashita, H.; Tamaki, S.; Kretsinger, R.H. Shrinking of repeating unit length in leucine-rich repeats from double-stranded DNA viruses. *Arch. Virol.* **2021**, *166*, 43–64.
43. Hao, B.; Zheng, N.; Schulman, B.A.; Wu, G.; Miller, J.J.; Pagano, M.; Pavletich, N.P. Structural basis of the Cks1-dependent recognition of p27Kip1 by the SCFSkp2 ubiquitin ligase. *Mol. Cell* **2005**, *20*, 9–19.
44. Kuchay, S.; Wang, H.; Marzio, A.; Jain, K.; Homer, H.; Fehrenbacher, N.; Philips, M.R.; Zheng, N.; Pagano, M. GGTase3 is a newly identified geranylgeranyltransferase targeting a ubiquitin ligase. *Nat. Struct. Mol. Biol.* **2019**, *26*, 628–636.
45. Xing, W.; Busino, L.; Hinds, T.R.; Marionni, S.T.; Saifee, N.H.; Bush, M.F.; Pagano, M.; Zheng, N. SCFFBXL3 ubiquitin ligase targets cryptochromes at their cofactor pocket. *Nature* **2013**, *496*, 64–68.
46. Wang, H.; Shi, H.; Rajan, M.; Canarie, E.R.; Hong, S.; Simoneschi, D.; Pagano, M.; Bush, M.F.; Stoll, S.; Leibold, E.A. FBXL5 regulates IRP2 stability in iron homeostasis via an oxygen-responsive [2Fe-2S] cluster. *Mol. Cell* **2020**, *78*, 31–41.

47. Cao, S.; Garcia, S.F.; Shi, H.; James, E.I.; Kito, Y.; Shi, H.; Mao, H.; Kaisari, S.; Rona, G.; Deng, S.; et al. Recognition of BACH1 quaternary structure degrons by two F-box proteins under oxidative stress. *Cell* **2024**, *187*, 7568–7584.
48. Buckley, D.L.; Van Molle, I.; Gareiss, P.C.; Tae, H.S.; Michel, J.; Noblin, D.J.; Jorgensen, W.L.; Ciulli, A.; Crews, C.M. Targeting the von Hippel–Lindau E3 ubiquitin ligase using small molecules to disrupt the VHL/HIF-1 α interaction. *J. Am. Chem. Soc.* **2012**, *134*, 4465–4468.
49. Chakraborty, S.; Nguyen, B.; Wasti, S.D.; Xu, G. Plant leucine-rich repeat receptor kinase (LRR-RK): Structure, ligand perception, and activation mechanism. *Molecules* **2019**, *24*, 3081.
50. Matsushima, N.; Miyashita, H.; Enkhbayar, P.; Kretsinger, R.H. Comparative geometrical analysis of leucine-rich repeat structures in the nod-like and toll-like receptors in vertebrate innate immunity. *Biomolecules* **2015**, *5*, 1955–1978.
51. Kibby, E.M.; Conte, A.N.; Burroughs, A.M.; Nagy, T.A.; Vargas, J.A.; Whalen, L.A.; Aravind, L.; Whiteley, A.T. Bacterial NLR-related proteins protect against phage. *Cell* **2023**, *186*, 2410–2424.
52. Barrangou, R.; Marraffini, L.A. CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Mol. Cell* **2014**, *54*, 234–244.
53. Duine, J.A.; Jongejan, J.A. Quinoproteins, enzymes with pyrrolo-quinoline quinone as cofactor. *Annu. Rev. Biochem.* **1989**, *58*, 403–426.
54. Schubert, W.-D.; Göbel, G.; Diepholz, M.; Darji, A.; Kloer, D.; Hain, T.; Chakraborty, T.; Wehland, J.; Domann, E.; Heinz, D.W. Internalins from the human pathogen *Listeria monocytogenes* combine three distinct folds into a contiguous internalin domain. *J. Mol. Biol.* **2001**, *312*, 783–794.
55. Helmann, J.D.; Chamberlin, M.J. Structure and function of bacterial sigma factors. *Annu. Rev. Biochem.* **1988**, *57*, 839–872.
56. Sanders, C.; Turkarslan, S.; Lee, D.-W.; Daldal, F. Cytochrome c biogenesis: The Ccm system. *Trends Microbiol.* **2010**, *18*, 266–274.
57. Wolanin, P.M.; Webre, D.J.; Stock, J.B. Mechanism of phosphatase activity in the chemotaxis response regulator CheY. *Biochemistry* **2003**, *42*, 14075–14082.
58. Soanes, D.M.; Talbot, N.J. Comparative genome analysis reveals an absence of leucine-rich repeat pattern-recognition receptor proteins in the kingdom Fungi. *PLoS ONE* **2010**, *5*, e12725.
59. Rivas-Marin, E.; Canosa, I.; Devos, D.P. Evolutionary cell biology of division mode in the bacterial Planctomycetes-Verrucomicrobia-Chlamydiae superphylum. *Front. Microbiol.* **2016**, *7*, 1964.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.