Structural Principles of Leucine-Rich Repeat (LRR) Proteins

Purevjav Enkhbayar,[1,3] Masakatsu Kamiya,[2,3] Mitsuru Osaki,[1] Takeshi Matsumoto,[3] and Norio Matsushima[3]*

[1]*Division of Biological Resources and Production, Graduate School of Agriculture, Hokkaido University, Sapporo, Hokkaido, Japan*

[2]*Division of Biological Science, Graduate School of Science, Hokkaido University, Sapporo, Hokkaido, Japan*

[3]*School of Health Sciences, Sapporo Medical University, Sapporo, Hokkaido, Japan*

ABSTRACT **LRR-containing proteins are present in over 2000 proteins from viruses to eukaryotes. Most LRRs are 20–30 amino acids long, and the repeat number ranges from 2 to 42. The known structures of 14 LRR proteins, each containing 4–17 repeats, have revealed that the LRR domains fold into a horseshoe (or arc) shape with a parallel β-sheet on the concave face and with various secondary structures, including α-helix, 3_{10}-helix, and pII helix on the convex face. We developed simple methods to characterize quantitatively the arc shape of LRR and then applied them to all known LRR proteins. A quantity of $2R\sin(\bar{\varphi}/2)$, in which R and $\bar{\varphi}$ are the radii of the LRR arc and the rotation angle about the central axis per repeating unit, respectively, is highly conserved in all the LRR proteins regardless of a large variety of repeat number and the radius of the LRR arc. The radii of the LRR arc with β-α structural units are smaller than those with β-3_{10} or β-pII units. The concave face of the LRR β-sheet forms a surface analogous to a part of a Möbius strip. Proteins 2003;54:394–403.**

Key words: geometric analysis; three-dimensional circle fitting; curvature; twist; Möbius strip; tilt angle; β-sheet direction vector

INTRODUCTION

LRRs are present in over 2000 proteins. LRR proteins have been identified in viruses, bacteria, arcaea, and eukaryotes. They include hormone receptors, tyrosine kinase receptors, cell-adhesion molecules, bacterial virulence factors, enzymes, and extracellular matrix-binding glycoproteins. All of the LRR proteins appear to be involved in protein–protein interactions; these include signal transduction, cell adhesion, DNA repair, recombination, transcription, RNA processing, disease resistance, ice nucleation, apoptosis, the plant immune response, and the mammalian innate immune response.

LRRs vary in their lengths and pattern of conserved residues and are usually present in tandem.[1,2] The number of LRR repeats ranges from 2 to 45, and their length ranges from 20 to 30. The LRRs are highly variable in the consensus sequence and the length of repeats. All LRR repeats are divided into a highly conserved segment and a variable segment. The highly conserved segment consists of an 11-residue stretch, LxxLxLxxNxL, or a 12-residue stretch, LxxLxLxxCxxL, in which "L" is Val, Leu, or Ile, "N" is Asn, Thr, Ser, or Cys, and "C" is Cys or Ser.[3–5] Seven classes of the LRRs had been proposed, characterized by different lengths and consensus sequences of the variable segments of repeats.[3,4,6] They are "typical," "RI-like," "CC," "PS," "SDS22-like," "bacterial," and "TpLRR."[3,7]

The structures of 14 different proteins that contain LRRs are available. They are RI[2,8,9]; RanGAP[10,11]; Skp2[12,13]; three internalins: Inl-B,[14,15] Inl-H,[15] and Inl-A,[16]; RabGGTα[17]; TAP[18,19]; U2A′[20]; YopM,[21] Chlamydomonus DLC-1[22,23]; GPIbα[24,25]; Tmod[26]; and nogo receptor.[27] All the LRR domains in these proteins adopt an arc shape or a horseshoe shape, a bowed tube on the concave (inner) face, with a parallel β-sheet and the convex (outer) face with a variety of secondary structures, such as α-helix, 3_{10}-helix, pII, and β-turn (Fig. 1).

Solenoid structures including the LRR structures that contain repeating structural units have been regarded to form a continuous superhelix.[28] The solenoid structures were qualitatively characterized by curvature, handedness, and twist. For the quantitative characterization, it is necessary to determine the curved axis of the superhelix.

In this study, we developed two simple methods to characterize quantitatively the LRR arc and applied them to all the known structures. One of the two methods is a 3D circle fitting using atomic coordinates of individual LRR β-sheets for the known structures of LRR proteins. The 3D circle fitting yields two geometric parameters: the radius

Abbreviations: 3D, three-dimensional; α, α-helix; β, β-strand; DLC-1, outer arm dynein light chain 1; GPIbα, glycoprotein Ibα; Inl-A, internalin A; Inl-B, internalin B; Inl-H, internalin H; LRR, leucine-rich repeat; pII, polyproline II; RabGGTα, Rab Geranylgeranyltransferase α-subunit; RanGAP, GTPase-activating protein; RI, ribonuclease inhibitor Skp2, Skp2 protein; TAP, mRNA export factor or tip associating protein; Tmod, tropomodulin; U2A′, spliceosomal U2A′ protein; VWF-A1, von Willebrand factor A1 domain; YopM, outer protein YopM.

Grant sponsor: Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture of Japan (to N. Matsushima).

*Correspondence to: Norio Matsushima, School of Health Sciences, Sapporo Medical University, Sapporo 060-8589, Japan. E-mail: matusima@sapmed.ac.jp

Received 26 May 2003; Accepted 29 July 2003

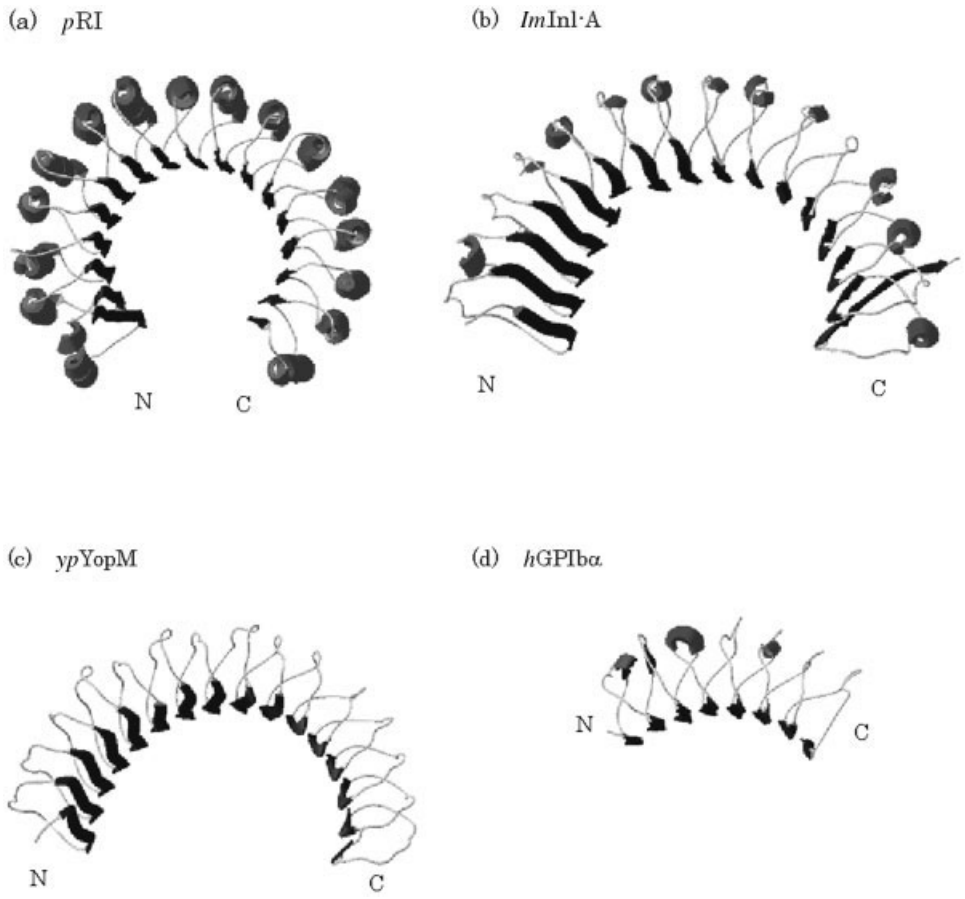

Fig. 1. Ribbon presentation of LRR domains within four representative LRR proteins: (**a**) *p*RI, (**b**) *Im*Inl-A, (**c**) *yp*YopM, and (**d**) *h*GPIbα.

(R) for the LRR arc and its rotation angle (ϕ) about the central axis. The other method is a 3D representation of the β-sheet direction vectors of individual β-sheets of LRRs, and its numerical analysis yields the tilt angle between the direction vector and the central axis.

The radii of the LRR arc with β-α structural units are smaller than those with β-3_{10} or β-pII units. We found a relationship between the R and ϕ parameters on the LRR arc regardless of the repeat number. The concave face of the LRR β-sheet forms a surface analogous to a part of a Möbius strip.

METHODS

LRR Proteins

The known structures of the LRR domains are summarized in Table I. The data were taken from 14 different LRR proteins (Table I). Only the data for DLC-1 were taken from a solution structure determined by NMR technique, while all the others were taken from crystal structures.

The 3D Circle Fitting

The known LRR structures reveal that 3 residues at positions 3–5 in the highly conserved segments, LxxLxLxx-NxL or LxxLxLxxCxxL, form a short β-strand, which is completely, or almost completely, conserved in all structures of the LRR proteins studied here (Table II). As noted, the β-strands stack parallel and then form an arc. Therefore, the atomic coordinates of the α-carbon ($^{\alpha}$C) of the consensus leucine residue at position 4 (corresponding to the middle of each β-strand) in individual LRR repeats were used as reference points for the 3D circle fitting. In the case of RI, RabGGTα, TAP, and Tmod, putative LRR motifs at the N- or C-terminal repeat display a weaker homology to the internal repeat motif. However, they form parallel β-strands. Thus, the putative LRR motifs were added as an LRR. Inl-A has complete, or almost complete, 15 LRRs, while the 16th repeat is incomplete; the LRR adopts a β-strand that combines to form a parallel β-sheet. Thus, the 16th repeat was also added as an LRR.

The 3D circle fitting is the method for directly fitting a circle to a set of 3D data points. We used a multistep process to accomplish the 3D circle fitting:

Step 1: Compute the least-squares plane of the data.
Step 2: Project the data points onto the plane.
Step 3: Compute the 2D circle fit in the plane.
Step 4: Perform a full, 3D minimization search over all the parameters.

TABLE I. Fourteen Different LRR Proteins of Known Structures

Name and origin of the protein	Short name	PDB code	Resolution Å	Comments
Porcine ribonuclease inhibitor	*p*RI	2nbh	2.30	Monomer
	*p*RI	1dfj	2.50	RI-RNase A complex
Human ribonuclease inhibitor	*h*RI	1a4y	2.00	RI-angiogenin complex
Chicken tropomodulin	*c*Tmod	1io0	1.45	Monomer; C-terminal half
GTPase-activating protein RNA1 from *Schizosaccharomyces pombe*	*sp*RanGAP	1yrg	2.66	Homodimer
RanGTPase-activating protein 1 from *Schizosaccharomyces pombe*	*sp*RanGAP	1k5d	2.70	Ran-Gppnhp-Ranbp1-RanGAP complex
	*sp*RanGAP	1k5g	3.10	Ran-Gdp-Alfx-Ranbp1-RanGAP complex
Human Skp2 protein	*h*Skp2	1fqv	2.80	Skp1-Skp2 complex
	*h*Skp2	1fs2	2.90	Skp1-Skp2 complex
Internalin A from *Listeria monocytogenes*	*lm*Inl-A	1o6t	1.60	Monomer
	*lm*Inl-A	1o6v	1.50	Homodimer
	*lm*Inl-A	1o6s	1.80	Inl-A-human E-cadherin complex
Internalin B from *Listeria monocytogenes*	*lm*Inl-B	1d0b	1.86	Monomer
	*lm*Inl-B	1h6t	1.60	Monomer; C-terminal Ig-like domain
Internalin H from *Listeria monocytogenes*	*lm*Inl-H	1h6u	1.80	Monomer; C-terminal Ig-like domain
Rat rab geranylgeranyltransferase α subunit	*r*RabGGTα	1dce	2.00	RabGGTα-RabGGTβ complex
Human U2 A′ protein	*h*U2A′	1a9n	2.38	U2B′-U2A′ RNA ternary complex
Human tip associating protein	*h*TAP	1ft8	3.15	Homotetramer
Human nuclear RNA export factor 1	*h*TAP	1fo1	2.90	Homodimer
Outer arm dynein from *Chlamydomonas reinhardtii*	*cr*DLC1	1ds9	NMR	Solution structure
Outer protein YopM from *Yersinia pestis*	*yp*YopM	1g9u	2.35	Homotetramer
	*yp*YopM	1jl5	2.10	Homotetramer
Human Glycoprotein lbα	*h*GPIbα	1m0z	1.85	Homodimer
	*h*GPIbα	1m10	3.10	GPIBα-VWF A1 complex
Human nogo receptor	*h*Nogo	1ozn	1.52	Monomer

Structures are listed here in the same order as they appear in Table III.

TABLE II. Repeat Number, Unit Length, Consensus Sequence, Type, and Structural Unit of the LRRs of Known Structures

Protein name	Repeat number (*N*)[a]	Unit length	Consensus sequence[b]	Type[c]	Structural unit[d]
RI	17	28–29	LxxLxLxx(N/C)xxLxxxgoxxLxxoLxxxxxx	RI-like	β-α (16)
Tmod	5	28–30	LxxLxLxxNxLxxxxxxxxxxxxxxxxx	?	β-α (5)
RanGAP	11	28–37	LxxLxLxxNxLxxxxxxxxxxxxxxxxxx	RI-like	β-α (7)
Skp2	11	23–27	LxxLxLxxNxLxxxxxxxxxxxx	CC	β-α (7)
Inl-A	16	21–22	LxxLxLxxNxIxDIxxLxxLxx	SDS22-like	β-3_{10} (15)
Inl-B	8	22	LxxLxLxxNxIxDIxxLxxLxx	"	β-3_{10} (7)
Inl-H	8	22	LxxLxLxxNxIxDIxxLxxLxx	"	β-3_{10} (7)
RabGGTα	6	22–27	LxxLxLxxNxLxxxxxxxxxx	"	β-3_{10} (4); β-α (1)
U2A′	5	22–27	LxxLxLxxNxLxxxxxxxxxx	"	β-3_{10} (2); β-β-α (1)
TAP	4	24–43	LxxLxLxxNxLxxxxxxxxxxxxxxx	"	β-3_{10} (1); β-3_{10}/α (3)
DLC1	6	22–25	LxxLxLxxNxLxxxxxxxxxx	"	β-β-α (4); β-α (2)
YopM	16	20–22	LxxLx(A/V)xxNxLxxLP(D/E)LPPx	bacterial	β-pII (15)
GPIbα	8	22–24	LxxLxLxxNxLxxxxxxxxxxxx	?	β-t (3); β-β (2)
Nogo	9	22–24	LxxLxLxxNxLxxLxxxAFxxLxx	Typical	β-α (1)

[a]The repeat number of LRRs is the number of β-strands on the concave face in the highly conserved segment within respective LRR proteins.
[b]The length of the consensus sequence is that with the most highest frequency.
[c]The type of LRR motif was identified by Kajava.[3,7]
[d]The number in parenthesis is the number of secondary structures on the convex face of the arc shape of the LRR domains. The secondary structure of β-strand, 3_{10}-helix and polyproline II helix was identified by the dihedral angles of each residue. The β-turn in GPIbα was identified after Uff et al.[24] Abbreviations: β, β-strand; 3_{10}, 3_{10}-helix; t, β-turn; pII, polyproline II helix.

We are fitting a set of data points so $\{\boldsymbol{x_i}\}$, $i = 1, 2, \ldots, N$. Scalar quantities are represented in plain type and vector quantities with boldface type. Other notations are

$\boldsymbol{c} = (c_x, c_y, c_z)$ The center of the circle
$\boldsymbol{x_i} = (x_i, y_i, z_i)$ The ith data point
$\boldsymbol{n} = (n_x, n_y, n_z)$ The direction vector normal to the circle's plane
R The radius of the circle.

The least-squares plane is determined by the minimization of the sum of the squares of the distance from the data points to a plane (*Step 1*). The 2D circle fit is performed using the data points projected onto the least-squares plane (*Step 3*).

The orthogonal distance (d_i) from a single point of the ith data to the circle is given by

$$d_i = d(x_i) = [\{g_i^2 + (f_i^c - R)^2\}]^{1/2}, \quad (1)$$

where $g_i = \boldsymbol{n} \cdot (\boldsymbol{x_i} - \boldsymbol{c})$ and f_i^c is the distance from the center of the circle to the data point projected onto the least-squares plane.

The quantity J^c that is the sum of the squares of the orthogonal distance (d_i) is given by

$$J^c(R, \boldsymbol{c}, \boldsymbol{n}) = \Sigma d_i^2 = \Sigma\{g_i^2 + (f_i^c - R)^2\}. \quad (2)$$

The quantity J^c is minimized, subject to the constraint:

$$|\boldsymbol{n}| = 1 \quad (3)$$

This procedure optimizes all the parameters simultaneously (*Step 4*).[29]

The initial parameters of the least-squares plane in *Step 1* were determined by the atomic coordinates of the 3 $^{\alpha}C_i$ atoms at the first, the middle, and the last repeats. The initial $\boldsymbol{c}$ of the circle's center in *Step 3* was calculated by averaging the coordinates of all $^{\alpha}C_i$ atoms projected onto the least-squares plane.

In addition to the radius R, the rotation angle (ϕ) about the central axis was estimated. The R and ϕ gives a good approximation of the radius and the rotation angle of the LRR arc with N repeats, as illustrated in Figure 2(a). The average rotation angle ($\bar{\varphi}$) about the central axis per repeating unit was also calculated by $\phi/(N - 1)$. Furthermore, the repeat number per turn (N_t) was predicted by $360/\bar{\varphi}$.

The Analysis of the β-Sheet Direction Vector

The β-sheet direction vector ($\boldsymbol{b}$) in individual repeats was defined as the vector joining the atomic coordinates of the α-carbon ($^{\alpha}C$) of the residue at two positions, 3 and 5, which correspond to the first and the last residues of each β-strand, respectively. The β-sheet direction vectors in individual repeats were represented three-dimensionally.

Two parameters were calculated using the β-sheet direction vector ($\boldsymbol{b}$) in individual repeats. One is the β-sheet twist (τ) between adjacent repeats defined as the acute angles between the β-sheet direction vectors, $\boldsymbol{b_i}$ of the i-th repeat and $\boldsymbol{b_{i+1}}$ of the (i+1)-th repeat [Fig. 2(b)]. If the interstrand pair is right-hand twisted, then the cross-product $\boldsymbol{b_i} \times \boldsymbol{b_{i+1}}$ will be pointing roughly in the direction of $\boldsymbol{d_{i,i+1}}$: the vector from $^{\alpha}C$ (i) to $^{\alpha}C$ (i+1). Otherwise, the interstrand pair is left-hand twisted. Consequently, we defined the sign of the sheet twist as the sign of the scalar product between $\boldsymbol{d_{i,i+1}} \times \boldsymbol{b_i}$ and $\boldsymbol{b_{i+1}}$; $(\boldsymbol{d_{i,i+1}} \times \boldsymbol{b_i}) \cdot \boldsymbol{b_{i+1}}$.

The other parameter is the tilt angle (θ) between the β-sheet direction vector $\boldsymbol{b}$ and the central axis of the best-fit circle in individual repeats [Fig. 2(c)]. The vector $\boldsymbol{r_i}$ of the i-th repeat is the position vector projected to the circle plane, which is made by joining the center of the circle and the $^{\alpha}C$ coordinate of residue at position 4 in the highly conserved segment of the LRRs. The position vector $\boldsymbol{r_i}$ and the unit vector $\boldsymbol{n}$ define a plane. The tilt angle (θ_i) of the i-th repeat is the angle between a component of the β-sheet direction vector $\boldsymbol{b_i}$ parallel to this plane and the vector $\boldsymbol{n}$, as is illustrated in Figure 2(c).

To calculate the parallel component of the $\boldsymbol{b_i}$ vector, we define two unit vectors, $\rho_i = \boldsymbol{r_i}/|\boldsymbol{r_i}|$ and $\tau = \boldsymbol{n} \times \rho_i$ [Fig. 2(c)]. The three vectors, $\boldsymbol{n}$, ρ_i, and τ, are perpendicular to each other. The vector $\boldsymbol{b_i}$ is given by

$$\boldsymbol{b_i} = \boldsymbol{b_i^{\tau}} + \boldsymbol{b_i^{n\rho}}, \quad (4)$$

where $\boldsymbol{b_i^{\tau}}$ is a component parallel to τ and $\boldsymbol{b_i^{n\rho}}$ is a component parallel to the plane defined by $\boldsymbol{n} \times \rho_i$. The vector $\boldsymbol{b_i^{\tau}}$ is also given by $\boldsymbol{b_i}$ and τ:

$$\boldsymbol{b_i^{\tau}} = (\boldsymbol{b_i} \cdot \tau)\tau. \quad (5)$$

Thus, $\boldsymbol{b_i^{n\rho}}$ is calculated using Eqs. (4) and (5), and the angle between the two vectors, $\boldsymbol{b_i^{n\rho}}$ and $\boldsymbol{n}$ gives the tilt angle θ_i. The sign of the tilt angle is also defined as the sign of the scalar product between $\boldsymbol{b_i}$ and ρ_i; $(\boldsymbol{b_i} \cdot \rho_i)$.

The average change in the tilt angle (θ) per repeating unit ($\Delta\theta/\Delta N$) was calculated. The change in the tilt angle per turn (θ_t) was also predicted by the product of $\Delta\theta/\Delta N$ and N_t.

RESULTS AND DISCUSSION

A Relationship Between the Radius R and the Rotation Angle $\bar{\varphi}$

The best-fit circles obtained by the 3D circle fitting were closely fitted to the data points with small errors (Table III). The reliability of the 3D circle fitting was also confirmed in the case of three internalins (Inl-A, Inl-B, and Inl-H). Thsee three internalins have the same consensus sequence of LxxLxLxxNxLxxLxxLxxLxx and all individual repeats adopt β-3_{10} units. However, the repeat number differs; those of Inl-A, Inl-B, and Inl-H are 18, 8, and 8, respectively (Table I). Irrespective of such differences in the repeat number, the radii of these internalins were very similar to each other.

As noted, the 3D circle fitting yields two geometric parameters, the radius of the best-fit circle (R) and the rotation angle (ϕ) of the arc of LRRs having N repeats. In addition to R and ϕ, the average rotation angle about the central axis per repeating unit ($\bar{\varphi}$) and the repeat number per turn (N_t) are shown in Table III.

The radius R values ranged from 15 to 24 Å, and the ϕ values from 11 to 16°, except for RabGGTα and DLC-1. The values of the R and ϕ for RabGGTα and DLC-1 were 45–46 Å and 5.9–6.4°, respectively. The numerical quantity, D =

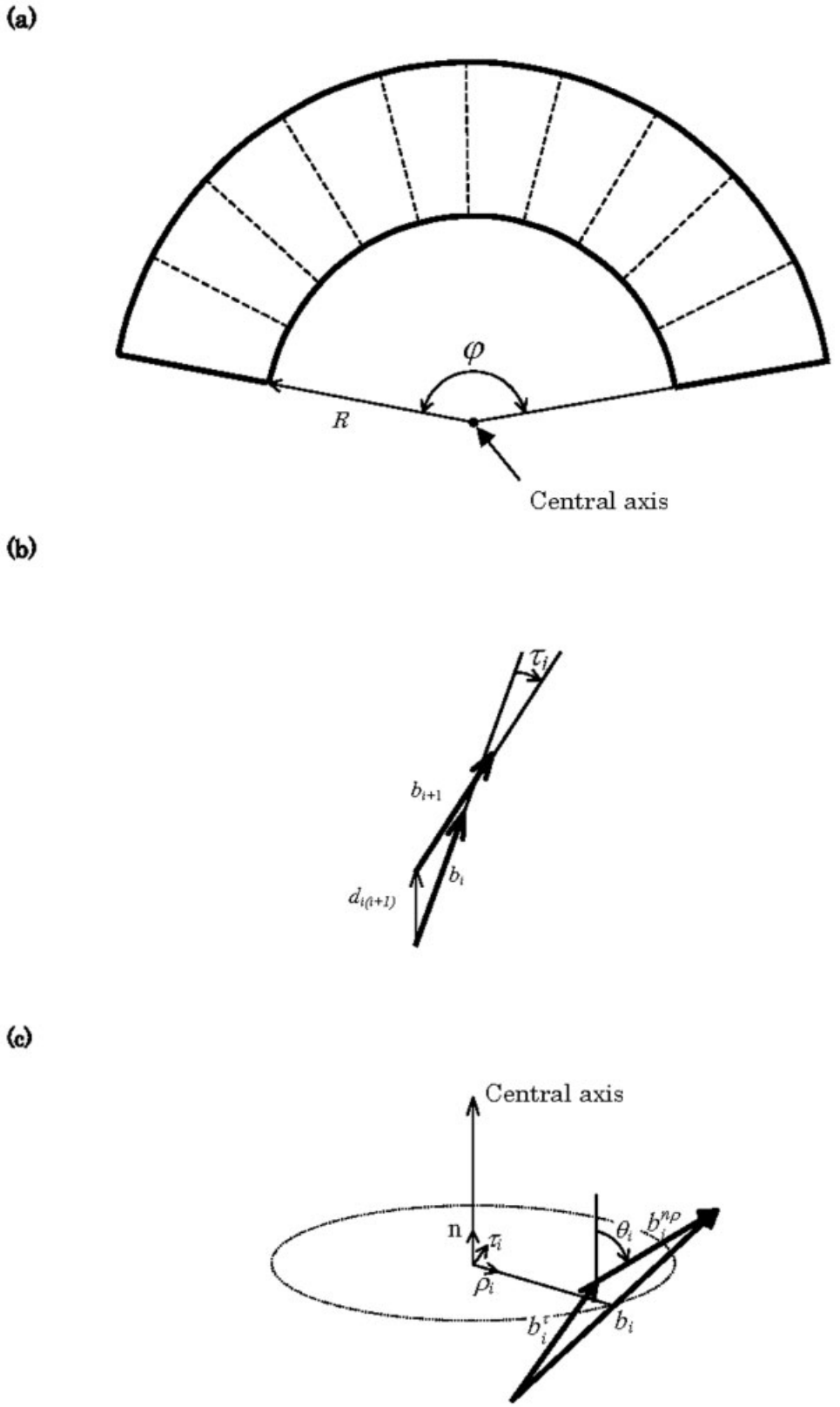

Fig. 2. Schematic illustration of the arc shape of LRR domain having $N (= 10)$ repeats and an illustration of the parameters used for the geometric analysis of the LRRs. (**a**) Two geometric parameters, R and ϕ indicate the radius and the rotation angle of the arc shape, respectively; (**b**) the parameter τ indicates the β-sheet twist between adjacent repeats; (**c**) the parameter θ indicates the tilt angle between the β-sheet direction vector (b) and the central axis of the best-fit circle.

TABLE III. Geometric Parameters of 14 Different LRR Proteins

Protein name	Chain ID	N	R (Å)	φ (°)	$\bar{\varphi}$ (°)	D (Å)	N_t	$\lvert\bar{\tau}\rvert$ (°)	$\frac{\Delta\theta}{\Delta N}$ (°)	θ_t (°)
*p*RI	A2nbh	17	17.68 ± 0.05	269.9	16.87	5.19	21.3	4.27 ± 0.64	0.90 ± 0.31	19.3
	Ildfj	17	18.68 ± 0.08	252.4	15.78	5.13	22.8	3.90 ± 0.65	1.21 ± 0.39	27.7
*h*RI	Ala4y	17	17.56 ± 0.08	269.7	16.86	5.15	21.4	4.26 ± 0.71	1.73 ± 0.39	37.9
	D1a4y	17	17.54 ± 0.07	270.3	16.89	5.15	21.3	4.43 ± 0.75	1.82 ± 0.39	39.7
*c*Tmod	A1io0	5	17.51 ± 0.05	65.6	16.41	5.00	21.9	4.05 ± 1.45	0.68 ± 0.79	14.9
*sp*RanGAP	A1yrg	11	19.57 ± 0.08	148.8	14.88	5.07	24.2	5.01 ± 0.77	3.09 ± 0.58	74.8
	B1yrg	11	19.56 ± 0.08	149.1	14.91	5.08	24.1	5.02 ± 0.84	3.21 ± 0.58	77.4
	C1k5d	11	19.98 ± 0.10	147.5	14.75	5.13	24.4	4.94 ± 1.08	3.65 ± 0.62	89.1
	F1k5d	11	19.94 ± 0.10	147.7	14.77	5.13	24.4	4.91 ± 1.08	3.66 ± 0.62	89.3
	I1k5d	11	19.99 ± 0.10	147.4	14.74	5.13	24.4	5.02 ± 1.06	3.65 ± 0.62	89.1
	L1k5d	11	19.91 ± 0.10	148.0	14.80	5.13	24.3	5.59 ± 1.01	4.22 ± 0.55	102.5
	C1k5g	11	19.51 ± 0.08	149.2	14.92	5.07	24.1	5.88 ± 1.14	3.41 ± 0.73	82.2
	F1k5g	11	19.51 ± 0.08	149.5	14.95	5.08	24.1	5.63 ± 1.34	4.66 ± 0.92	112.3
	I1k5g	11	19.54 ± 0.08	149.2	14.92	5.07	24.1	5.95 ± 1.19	3.43 ± 0.73	82.7
	L1k5g	11	19.50 ± 0.08	149.4	14.94	5.07	24.1	5.90 ± 1.14	3.42 ± 0.74	82.4
*h*Skp2	A1fqv	11	18.91 ± 0.19	151.3	15.13	4.98	23.8	6.89 ± 1.72	3.43 ± 0.65	85.4
	C1fqv	11	18.90 ± 0.19	151.5	15.15	4.98	23.8	6.97 ± 1.65	3.45 ± 0.68	85.9
	E1fqv	11	18.96 ± 0.20	151.0	15.10	4.98	23.8	6.76 ± 1.70	3.47 ± 0.67	87.1
	G1fqv	11	18.95 ± 0.19	151.1	15.11	4.98	23.8	6.90 ± 1.66	3.61 ± 0.69	90.3
	I1fqv	11	19.03 ± 0.20	150.3	15.03	4.98	23.9	6.97 ± 1.71	3.54 ± 0.68	88.9
	K1fqv	11	18.90 ± 0.19	152.0	15.20	5.00	23.7	6.98 ± 1.69	3.38 ± 0.68	83.8
	M1fqv	11	18.88 ± 0.20	152.1	15.21	5.00	23.7	6.82 ± 1.62	3.47 ± 0.69	86.4
	O1fqv	11	18.96 ± 0.19	151.0	15.10	4.98	23.8	6.71 ± 1.65	3.54 ± 0.69	88.9
	A1fs2	9	19.04 ± 0.08	120.9	15.11	5.01	23.8	7.67 ± 1.80	4.97 ± 0.60	118.3
	C1fs2	9	19.35 ± 0.10	119.7	14.96	5.04	24.1	7.92 ± 1.90	5.05 ± 0.68	121.7
*Im*1nl-A	A1o6t	16	22.66 ± 0.09	186.0	12.40	4.89	29.0	7.03 ± 1.00	6.62 ± 0.26	192.0
	A1o6v	16	23.07 ± 0.08	183.5	12.23	4.92	29.4	7.01 ± 1.04	6.37 ± 0.22	187.3
	B1o6v	16	23.00 ± 0.07	183.7	12.25	4.91	29.4	7.10 ± 0.96	6.39 ± 0.21	187.9
	A1o6s	16	23.57 ± 0.06	179.0	11.93	4.90	30.2	7.30 ± 0.92	6.82 ± 0.24	206.0
*lm*Inl-B	A1d0b	8	21.33 ± 0.06	94.0	13.43	4.99	26.8	7.00 ± 1.19	6.18 ± 0.36	165.6
	A1h6t	8	21.66 ± 0.05	92.5	13.21	4.98	27.2	6.45 ± 0.94	6.09 ± 0.23	165.6
*lm*Inl-H	A1h6u	8	24.02 ± 0.04	83.4	11.92	4.99	30.2	5.06 ± 1.02	4.82 ± 0.34	145.6
*r*RabGGTα	A1dce	6	44.85 ± 0.03	31.7	6.35	4.97	56.7	3.21 ± 0.67	0.14 ± 0.58	7.9
	C1dce	6	44.67 ± 0.05	31.9	6.39	4.98	56.3	3.22 ± 0.64	0.03 ± 0.54	1.7
*h*U2A′	A1a9n	5	24.21 ± 0.04	48.3	12.07	5.09	29.8	8.42 ± 2.54	−2.63 ± 2.68	−76.8
	C1a9n	5	25.44 ± 0.04	46.3	11.57	5.13	31.1	7.79 ± 2.47	−2.10 ± 2.57	−63.0
*h*TAP	A1ft8	4	15.53 ± 0.02	55.8	18.61	5.02	19.3	2.86 ± 1.34	−1.15 ± 1.09	−22.2
	B1ft8	4	15.13 ± 0.02	57.5	19.16	5.04	18.8	3.32 ± 1.35	−0.92 ± 1.19	−17.3
	C1ft8	4	15.08 ± 0.01	57.6	19.21	5.03	18.7	3.61 ± 1.30	−0.98 ± 1.51	−18.3
	D1ft8	4	15.65 ± 0.02	55.5	18.48	5.03	19.5	3.14 ± 1.20	−0.83 ± 1.30	−16.2
	A1fo1	4	14.37 ± 0.04	60.9	20.31	5.07	17.7	4.46 ± 0.23	−1.56 ± 1.55	−27.6
	B1fo1	4	14.20 ± 0.02	60.8	20.25	4.99	17.8	5.16 ± 0.94	−1.65 ± 1.73	−29.4
*cr*DLC1	A1ds9	6	46.76 ± 0.15	29.4	5.87	4.79	61.3	5.60 ± 1.85	−0.40 ± 0.98	−24.5
*yp*YopM	A1g9u	16	24.66 ± 0.11	174.0	11.60	4.98	31.0	5.77 ± 0.75	4.87 ± 0.25	151.0
	A1j15	16	23.88 ± 0.07	172.5	11.50	4.78	31.3	5.90 ± 0.85	5.21 ± 0.24	163.1
*h*GPIbα	A1m0z	8	23.75 ± 0.02	85.9	12.28	5.08	29.3	3.20 ± 0.82	3.05 ± 0.43	89.4
	B1m0z	8	21.74 ± 0.02	94.0	13.43	5.08	26.8	3.16 ± 0.70	2.90 ± 0.42	77.7
	B1m10	8	23.30 ± 0.03	87.0	12.43	5.04	29.0	3.68 ± 0.73	3.40 ± 0.33	98.6
*h*Nogo	A1ozn	9	33.70 ± 0.04	69.5	8.69	5.11	41.4	3.47 ± 0.44	3.18 ± 0.10	131.7

N is the repeat number of LRRs. φ is the rotation angle of the LRR arc shape. $\bar{\varphi}$ indicates the average rotation angle about the central axis per repeating unit that is given by $\bar{\varphi} = \varphi/(N-1)$. N_t indicates the repeat number per turn that is given by $360/\bar{\varphi}$. D is given by $2R\sin(\bar{\varphi}/2)$. θ_t is the change in θ per turn that is given by $N_t(\Delta\theta/\Delta N)$ where θ is the angle between the β-sheet direction vector and the central axis. The slope $(\Delta\theta/\Delta N)$ of *h*Skp2 with 11 and 9 LRRs was calculated for the 10 LRRs that removed the last repeat and for the 7 LRRs that removed the last two repeats, respectively, because the tilt angle of these removed repeats differed drastically from the other repeats.

$2R\ \sin(\bar{\varphi}/2)$, was calculated for individual LRR proteins. We found that the D values are highly conserved in all the known LRR structures, regardless of repeat number (Table III). This clearly indicates that $D \approx$ constant is an important parameter restraining the curvature of the LRR arc.

The D gives a good approximate value of the interstrand distance, which remains in range of 4.5 ± 1 Å to allow for H bonds to form between strands, because the quantity D is the average $^{\alpha}C(i) - {}^{\alpha}C(i+1)$ distance between adjacent repeats, in which the $^{\alpha}C$ atoms are at position 4 in the

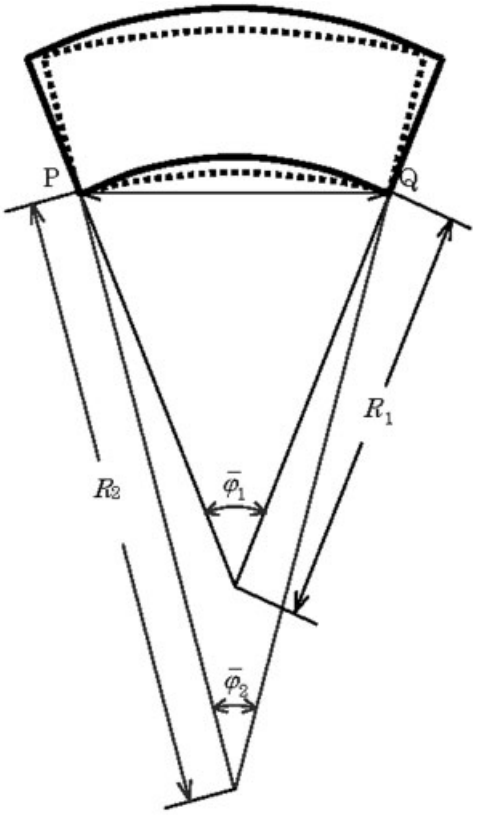

Fig. 3. The relationship between the radius (R) and the average rotation angle about the central axis per repeating unit ($\bar{\varphi}$) of the LRR arc shape in the 14 LRR proteins. P and Q are the reference points located on the β-sheets between adjacent repeats. The distance between P and Q (= D) is given by $2R_1 \sin(\bar{\varphi}_1/2) = 2R_2 \sin(\bar{\varphi}_2/2)$.

β-strand, as noted. In fact, the D values were consistent with those calculated directly by the atomic coordinates of the $^{\alpha}$C's at the i and $i + 1$ repeats. It is clear that the D constancy is the results of the interstrand distance = constant; that is, if $R_1 < R_2$, then $\bar{\varphi}_1 > \bar{\varphi}_2$ is true for all the LRR proteins, because $\bar{\varphi} < 180°$. This relationship is illustrated in Figure 3.

Diversity in the Radius of the LRR Arc

The supersecondary structure observed in the known LRR structures may be classified into at least three groups: β-α, β-pII, and β-3_{10} structural units (Table II). All or most of the individual repeats within RI, RanGAP, Skp2, and Tmod adopt β-α units. The radius for the LRR arc of RI, RanGAP, Skp2, and Tmod ranges from 17 to 20 Å (Table III). All repeats within YopM and the internalins (Inl-A, Inl-B, and Inl-H) adopt β-pII and β-3_{10} units, respectively. YopM has a value of 24 Å for the radius and 21–24 Å for the internalins. The radius of U2A′, in which 3 out of the 5 repeats within U2A′ adopt β-3_{10} units, is 23–24 Å. Thus, it can be concluded that the LRR arc shape having β-α units has a smaller radius than the arc having β-pII units or β-3_{10} units.

The pronounced curvature of the LRR arc observed in RI with the β-α units was indicated to be the result of different diameters of the α-helices and β-strands when packed in a parallel arrangement.[8,30] It was also suggested that the curvature is caused by the first nonpolar residue (usually) in the highly conserved 11-residue stretch with the consensus sequence LxxLxLxxNxL.[31] Bearing this in mind, YopM LRR was compared and contrasted to repeats within Inl-B, yeast RanGAP, U2A′, and RabGGTα.[21] For most LRR types, the right-hand side β-sheet portion, the lower-left corner turn, and the top of the backbone "box" overlap almost exactly with corresponding parts of RI LRR. However, the left sides are quite variable and do not align well.[21] Therefore, it is suggested that the different diameters among α-helix, 3_{10}-helix, and pII helix affect the radius of the LRR arc. The diameter of α-helix is bigger than those of the 3_{10}-helix and pII helix. Thus, it is inferred that the radius of the LRR arc with β-α units is smaller than those with β-3_{10} and β-pII units. In fact, this was observed, as noted.

Similar consideration is applicable to the radii of LRR arcs with β-pII and β-3_{10} units. The 3_{10}-helix of the Inl-B LRR occupies almost exactly the same space as the pII helix of the YopM LRR.[21] As expected, the radius of YopM with β-pII units is comparable with that of the three internalins (Inl-A, Inl-B, and Inl-H) with β-3_{10} units.

The above observations indicate that the radius is closely related to the different diameters among α-, 3_{10}-, and pII-helices on the convex face.

Moreover, it may be of interest to note that the radius of GPIbα is also comparable with those of the LRR proteins with β-pII or β-3_{10}. Two out of 8 LRR repeats in GPIbα adopt β-β structural units, although the secondary structures of the other repeats are not specialized.[24] Furthermore, TAP, containing 4 LRR repeats that adopt 3 β-3_{10}/α units and 1 β-3_{10} unit, has the smallest radius among the known LRRs. A long 3_{10}/α-helix on the convex face at the first repeat with 43 residues appears to cause a small radius.

The radius of RabGGTα and DLC-1, with 6 LRRs, is the largest among the known LRR structures (Table III). As noted, most LRRs within RabGGTα adopt β-3_{10} structural units, while DLC-1 contains 4 β-β-α units and 2 β-α units (Table I). The LRRs in these two proteins are significantly similar to those in U2A′. However, U2A′ has a radius of 24 Å, which is half that of RabGGTα and DLC-1. At present, this large difference in the radius between RabGGTα or DLC-1 and U2A′ cannot be adequately explained.

The β-Sheet Concave Face as a Möbius Strip

The β-sheet direction vectors in individual repeats were represented in 3D, as shown in Figure 4. This 3D representation appears to suggest that the concave face of the LRR β-sheet is approximately confined to a Möbius strip. An ideal Möbius strip is a continuous one-sided surface that can be formed from a rectangular strip by rotating one end 180° and attaching it to the other end (Fig. 5).

If the concave face of LRRs is regarded as a part of a Möbius strip, it is predicted that the change in the tilt angle per turn (θ_t) will reach 180°. As expected, the θ values of most LRR proteins increase with increasing N; some proteins are exceptions (Fig. 6 and Table III). The

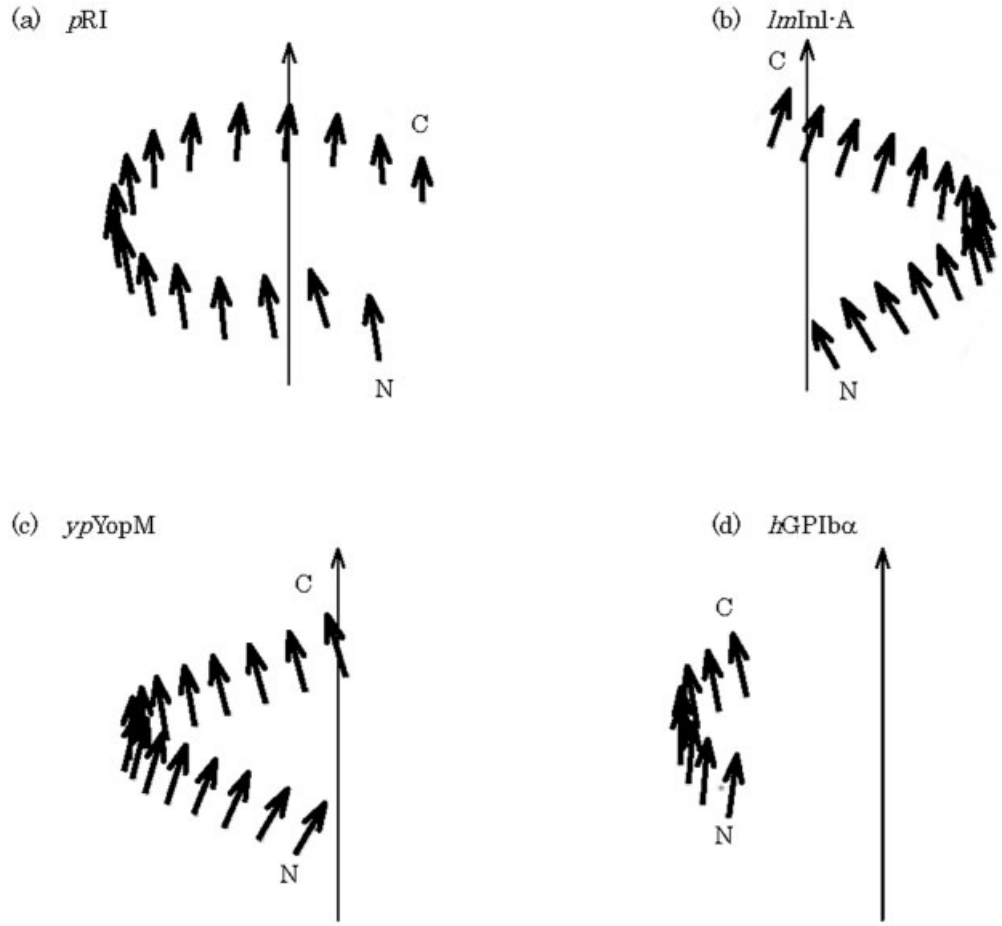

Fig. 4. The direction vector of parallel β-sheets of individual LRRs within four representative LRR proteins in 3D space: (**a**) *p*RI, (**b**) *lm*Inl-A, (**c**) *yp*YopM, and (**d**) *h*GPIbα.

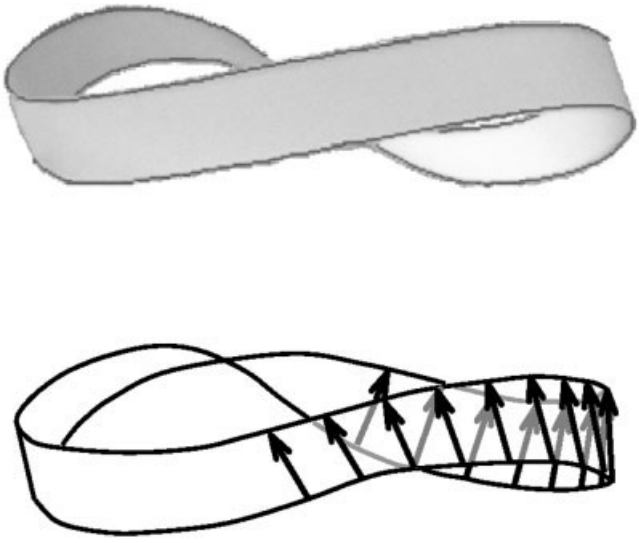

Fig. 5. A Möbius strip (**a**) and a schematic diagram of the concave face developed by stacking of the parallel β-sheets in LRRs (**b**).

estimated θ_t are 165–175° for YopM, 170° for Inl-B, 200–220° for Inl-A, and 146° for Inl-H, while the $\theta_t$of RanGAP, Skp2, and GPIbα are 75–120° (Table III). The θ_t values of the former group are very comparable with 180° in a Möbius strip.

As noted, the concave surfaces of RanGAP and Skp2, which have β-α structural units, have θ_t values of 75–120°. In contrast, RI, which has β-α structural units, has a much smaller θ_t (19–40°) than these two LRR proteins. The small θ_t observed in RI might be attributable to alternating LRRs of 28 or 29 residues each, which are not seen in RanGAP and Skp2.

It can be concluded that the tilt angle per turn (θ_t) satisfactorily characterizes the concave face of the LRR β-sheet in individual LRR structures. The concave face of the LRR β-sheet can be regarded as a part of a Möbius strip.

The Twist of β-Sheets

The β-sheets in protein structures strongly favor a right-hand twist. Intrinsic flexibility of the β-sheet twist (and coiling) allows two main types of structures to be formed by association of β-sheets. In one, twisted β-sheets pack face to face (aligned packing) to form sandwich or propeller structures. In the other, β-sheets twist and coil to form a closed barrel, in which the first strand is hydrogen bonded to the last.

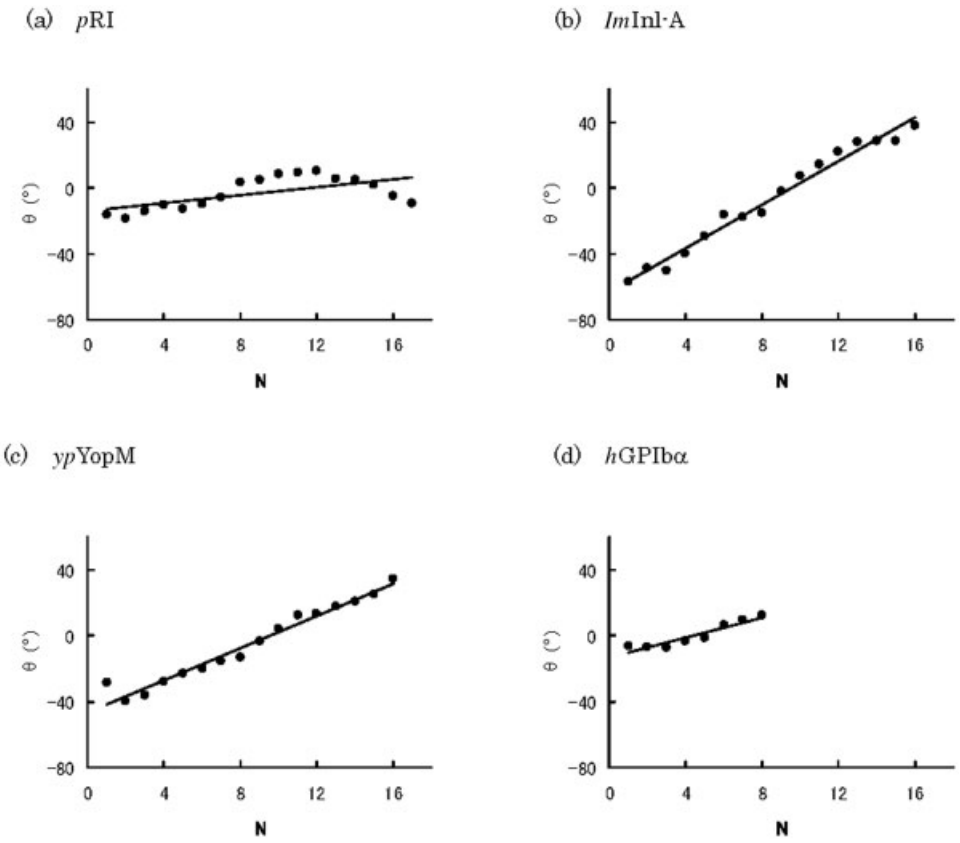

Fig. 6. The direction angle (θ) of individual repeats as a function of the repeat numbers (*N*) within LRR proteins: (**a**) *p*RI, (**b**) *yp*YopM, (**c**) *Im*Inl-A, and (**d**) *h*GPIbα.

The mean twist angle (τ) of β-sheets in LRR proteins was 3–8° (Table III). This value is much smaller than the 32° observed in β-barrels,[32–34] the 19° in α/β structures,[35] and 17° in the aligned class.[36] The small twists observed indicate that the LRR arc is grouped in a new class of β structures.

As noted, the LRR β-sheet concave face forms a surface analogous to a part of a Möbius strip. We think that the Möbius strip is the result of a combination of the small twist of the LRR β-sheets and the bending due to different diameters between β-strands on the concave face and the secondary structure, such as α-helix and 3_{10}-helix on the convex face when packed in a parallel arrangement. It may be significant to also note that the curved surface of the β-sheet barrels has been shown to be fitted satisfactorily by hyperboloids.[34,37]

Overall Structural Change in the Complexes of Some LRR Proteins

The crystal structure of three LRR proteins have been determined in the free state and their complexes: RI and its complex with RNase A,[8,38,39] Inl-A and its complex with the human receptor E-cadherin,[16] and GPIbα and its complex with VWF-A1.[13] In these RI, Inl-A, and GPIbα complexes, the concave face of the LRR β-sheet is involved in protein–protein interactions.

The radius of the LRR arc in the RI-RNase A complex is 1 Å larger than that of RI in the free form (Table III). This increase in the radius of the RI-RNase A complex corresponds to the increase in the shortest distance between the N- and C-terminal repeats.[2] Similarly, the complex between Inl-A and E-cadherin has a 0.5–0.6 Å larger radius than the protein in the free form (Table III). The radius in the complex GPIbα-VWF A1 is only slightly larger than that of the one chain of the homodimers in the free state (Table III). It is likely that their complexes show a tendency toward an increase in the radius of the arc of the LRR domains. Significantly, the formation of these complexes leads to an increase in θ_t, which has a range of 8–21° (Table III).

It can be concluded that the radius and θ_t are sensitive parameters to overall structural changes of the LRR domains.

CONCLUSIONS

In this study, we developed two new methods for characterizing quantitatively the LRR arc shape and then applied them to all known LRR proteins. The two methods deepened an understanding of the structural determinants and allowed a more in-depth analysis of the structural diversity and the overall structural change observed in the LRR proteins. The three geometric parameters, R, $\bar{\varphi}$, and θ_t, proposed here, satisfactorily characterized all the LRR structures.

We found the consistency of $2R \sin(\bar{\varphi}/2)$ that is a strong structural determinant that restrains the curvature of the

arc shape of the LRR domains. The radii of the LRR arcs having β-α structural units were smaller than those having β-3_{10} or β-pII units. The concave face of the LRR β-sheet can be regarded as a part of a Möbius strip.

Besides being of academic interest, these methods could be useful in a top-down approach for the de novo design of an LRR protein.[40]

ACKNOWLEDGMENTS

We thank Prof. Robert H. Kretsinger (University of Virginia) for his helpful suggestion and comments.

REFERENCES

1. Kobe B, Deisenhofer J. The leucine-rich repeat: a versatile binding motif. Trends Biochem Sci 1994;19:415–421.
2. Kobe B, Deisenhofer J. Proteins with leucine-rich repeats. Curr Opin Struct Biol 1995;5:409–416.
3. Kajava AV. Structural diversity of leucine-rich repeat proteins. J Mol Biol 1998;277:519–527.
4. Ohyanagi T, Matsushima N. Classification of tandem leucine-rich repeats within a great variety of proteins. FASEB J 1997;11:A949.
5. Kajava AV, Vassart G, Wodak SJ. Modeling of the three-dimensional structure of proteins with the typical leucine-rich repeats. Structure 1995;3:867–877.
6. Matsushima N, Ohyanagi T, Tanaka T, Kretsinger RH. Super-motifs and evolution of tandem leucine-rich repeats within the small proteoglycans—biglycan, decorin, lumican, fibromodulin, PRELP, keratocan, osteoadherin, epiphycan, and osteoglycin. Proteins 2000;38:210–225.
7. Kajava AV. Review: proteins with repeated sequence—structural prediction and modeling. J Struct Biol 2001;134:132–144.
8. Kobe B, Deisenhofer J. Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. Nature 1993;366:751–756.
9. Papageorgiou AC, Shapiro R, Acharya KR. Molecular recognition of human angiogenin by placental ribonuclease inhibitor—an X-ray crystallographic study at 2.0 Å resolution. EMBO J 1997;16:5162–5177.
10. Hillig RC, Renault L, Vetter IR, Drell T IV, Wittinghofer A, Becker J. The crystal structure of RNA1p: a new fold for a GTPase-activating protein. Mol Cell 1999;3:781–791.
11. Seewald MJ, Korner C, Wittinghofer A, Vetter IR. RanGAP mediates GTP hydrolysis without an arginine finger. Nature 2002;415:662–666.
12. Schulman BA, Carrano AC, Jeffrey PD, Bowen Z, Kinnucan ER, Finnin MS, Elledge SJ, Harper JW, Pagano M, Pavletich NP. Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. Nature 2000;408:381–386.
13. Zheng N, Schulman BA, Song L, Miller JJ, Jeffrey PD, Wang P, Chu C, Koepp DM, Elledge SJ, Pagano M, Conaway RC, Conaway JW, Harper JW, Pavletich NP. Structure of the Cul1-Rbx1-Skp1-F box Skp2 SCF ubiquitin ligase complex. Nature 2002;416:703–709.
14. Marino M, Braun L, Cossart P, Ghosh P. Structure of the lnlB leucine-rich repeats, a domain that triggers host cell invasion by the bacterial pathogen *L. monocytogenes*. Mol Cell 1999;4:1063–1072.
15. Schubert WD, Gobel G, Diepholz M, Darji A, Kloer D, Hain T, Chakraborty T, Wehland J, Domann E, Heinz DW. Internalins from the human pathogen *Listeria monocytogenes* combine three distinct folds into a contiguous internalin domain. J Mol Biol 2001;312:783–794.
16. Schubert WD, Urbanke C, Ziehm T, Beier V, Machner MP, Domann E, Wehland J, Chakraborty T, Heinz DW. Structure of internalin, a major invasion protein of *Listeria monocytogenes*, in complex with its human receptor E-cadherin. Cell 2002;111:825–836.
17. Zhang H, Seabra MC, Deisenhofer J. Crystal structure of Rab geranylgeranyltransferase at 2.0 Å resolution. Structure 2000;8:241–251.
18. Liker E, Fernandez E, Izaurralde E, Conti E. The structure of the mRNA export factor TAP reveals a cis arrangement of a non-canonical RNP domain and an LRR domain. EMBO J 2000;19:5587–5598.
19. Ho DN, Coburn GA, Kang Y, Cullen BR, Georgiadis MM. The crystal structure and mutational analysis of a novel RNA-binding domain found in the human Tap nuclear mRNA export factor. Proc Natl Acad Sci USA 2002;99:1888–1893.
20. Price SR, Evans PP, Nagai K. Crystal structure of the spliceosomal U2B-U2A′ protein complex bound to a fragment of U2 small nuclear RNA. Nature 1998;394:645–650.
21. Evdokimov AG, Anderson DE, Routzahn KM, Waugh DS. Unusual molecular arcitecture of the *Yersinia pestis* cytotoxin YopM: a leucine-rich repeat protein with the shortest repeating unit. J Mol Biol 2001;312:807–821.
22. Wu H, Maciejewski MW, Marintchev A, Benashski SE, Mullen GP, King SM. Solution structure of a dynein motor domain associated light chain. Nat Struct Biol 2000;7:575–579.
23. Wu H, Blackledge M, Maciejewski MW, Mullen GP, King SM. Relaxation-based structure refinement and backbone molecular dynamics of the dynein motor domain-associated light chain. Biochemistry 2003;42:57–71.
24. Uff S, Clemetson JM, Harrison T, Clemetson KJ, Emsley J. Crystal structure of the platelet glycoprotein Ibalpha N-terminal domain reveals an unmasking mechanism for receptor activation. J Biol Chem 2002;277:35657–35663.
25. Huizinga EG, Tsuji S, Romijn RA, Schiphorst ME, de Groot PG, Sixma JJ, Gros P. Structures of glycoprotein Ibα and its complex with von Willebrand factor A1 domain. Science 2002;297:1176–1179.
26. Krieger I, Kostyukova A, Yamashita A, Nitanai Y, Maeda Y. Crystal structure of the C-terminal half of tropomodulin and structural basis of actin filament pointed-end capping. Biophys J 2002;83:2716–2725.
27. He XL, Bazan JF, McDermott G, Park JB, Wang K, Tessier-Lavigne M, He Z, Garcia KC. Structure of the nogo receptor ectodomain: a recognition module implicated in myelin inhibition. Neuron 2003;38:177–185.
28. Kobe B, Kajava AV. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. Trends Biochem Sci 2000;25:509–515.
29. Shakarji CM. Least-squares fitting algorithms of the NIST algorithm testing system. J Res Natl Inst Stand 1998;103:633–641.
30. Kobe B, Kajava AV. The leucine-rich repeat as a protein recognition motif. Curr Opin Struct Biol 2001;11:725–732.
31. Kajava AV, Kobe B. Assessment of the ability to model proteins with leucine-rich repeats in light of the latest structural information. Protein Sci 2002;1:1082–1090.
32. Murzin AG, Lesk AM, Chothia C. Principles determining the structure of beta-sheet barrels in proteins: I. A theoretical analysis. J Mol Biol 1994;236:1369–1381.
33. Murzin AG, Lesk AM, Chothia C. Principles determining the structure of beta-sheet barrels in proteins: II. The observed structures. J Mol Biol 1994;236:1382–1400.
34. Lasters I, Wodak SJ, Alard P, van Cutsem E. Structural principles of parallel beta-barrels in proteins. Proc Natl Acad Sci USA 1988;85:3338–3342.
35. Janin J, Chothia C. Packing of alpha-helices onto beta-pleated sheets and the anatomy of alpha/beta proteins. J Mol Biol 1980;143:95–128.
36. Chothia C, Janin J. Relative orientation of close packed β-sheets in proteins. Proc Natl Acad Sci USA 1981;74:4130–4134.
37. Novotny J, Bruccoleri RE, Newell J. Twisted hyperboloid (Strophoid) as a model of β-barrels in proteins. J Mol Biol 1984;177:567–573.
38. Kobe B, Deisenhofer J. A structural basis of the interactions between leucine-rich repeats and protein ligands. Nature 1995;374:183–186.
39. Kobe B, Deisenhofer J. Complex between bovine ribonuclease A and porcine ribonuclease inhibitor crystallizes in a similar unit cell as free ribonuclease inhibitor. J Mol Biol 1994;241:288–291.
40. Forrer P, Stumpp MT, Binz HK, Pluckthun A. A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. FEBS Lett 2003;539:2–6.