PROTEINS: Structure, Function, and Genetics 38:210–225 (2000)

Super-Motifs and Evolution of Tandem Leucine-Rich Repeats Within the Small Proteoglycans—Biglycan, Decorin, Lumican, Fibromodulin, PRELP, Keratocan, Osteoadherin, Epiphycan, and Osteoglycin

Norio Matsushima,[1]* Toshio Ohyanagi,[1] Takanori Tanaka,[2] and Robert H. Kretsinger[3]
[1]*School of Health Sciences, Sapporo Medical University, Sapporo, Hokkaido, Japan*
[2]*Division of Biological Science, Graduate School of Science, Hokkaido University, Sapporo, Hokkaido, Japan*
[3]*Department of Biology, University of Virginia, Charlottesville, Virginia*

ABSTRACT **Leucine-rich repeats (LRRs) with 20–30 amino acids in unit length are present in many proteins from prokaryotes to eukaryotes. The LRR-containing proteins include a family of nine small proteoglycans, forming three distinct subfamilies: class I contains biglycan/PG-I and decorin/PG-II; class II: lumican, fibromodulin, PRELP, keratocan, and osteoadherin; and class III: epiphycan/PG-Lb and osteoglycin or osteoinductive factor. Comparative sequence analysis of the 34 available protein sequences reveals that these proteoglycans have two types of LRRs, which we call *S* and *T*. The type *S* LRR is 21 residues long and has the consensus sequence of xxaPzxLPxx-LxxLxLxxNxI. The type *T* LRR has 26 residues; its consensus sequence is zzxxaxxxxFxxaxxLxxLxLxxNxL. In both "x" indicates variable residue; "z" is frequently a gap; "a" is Val, Leu, or Ile; and I is Ile or Leu. These type *S* and *T* LRRs are ordered into two super-motifs—*STT* with about 73 residues in classes I and II and *ST* with about 47 residues in class III. The 12 LRRs in the small proteoglycans of I and II are best represented as $(STT)_4$; the seven LRRs of class III as $(ST)T(ST)_2$. Our analyses indicate that classes I/II and III evolved along different paths after the establishment of the precursor *ST*, and classes I and II also diverged after the establishment of the precursor $(STT)_4$. Proteins 2000;38:210–225.** © 2000 Wiley-Liss, Inc.

Key words: tandem repeat; super-repeat; proteoglycan I; proteoglycan II; proline arginine-rich and leucine-rich protein; PG-Lb; chondroadherin; gene duplication

INTRODUCTION

Leucine-rich repeats (LRRs) were first discovered in α2-glycoprotein from human serum.[1] More than 100 different LRR proteins have been described. They are present in organisms that range from bacteria to man.[2,3] 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, which take part in signal transduction, cell adhesion, DNA repair, recombination, transcription, RNA processing, disease resistance, and ice nucleation.

The number of LRRs motifs ranges from 1 to 30/33, for example, in platelet GPIb and chaoptin/Cf-2, respectively; LRRs are usually present in tandem. The most common length of LRRs is 24 residues, but the lengths of LRR units range from 20 to 30.[4,5] Kobe and Deisenhofer[4] proposed an overall consensus sequence for the LRR unit—xLxxLxLxx-Nzxaxxzazzzzazzxzz—in which "x" indicates a variable residue; "z" is frequently a gap; and "a" is Val, Leu, or Ile. The Leus are often replaced by Val or Ile, and the consensus Asn is sometimes replaced by Cys or Thr. All LRRs have a highly conserved LxxLxLxxNxL.[5] Recently it was proposed that LRR proteins can be subdivided into at least six or seven subfamilies, characterized by different lengths and consensus sequences of the variable parts of the repeats.[6,7] Furthermore, the repeats from these different subfamilies were proposed to differ in three-dimensional structure.[7]

Structures of two LRR proteins are available. Ribonuclease inhibitor consists of 15 alternating LRRs of 28 or 29 residues each, the consensus sequence of which is xLExLx-LxxCxLTxxxCxxLxxaLxxxx or xLExLxLxxNxLGDxGa-xxxxxxLxxPxx, showing that it forms higher-order repeating units, each 57 residues long.[8–10] Individual repeats correspond to β-α units that have a short β-strand and α-helix approximately antiparallel to each other, in which the LxL sequence of highly conserved LxxLxLxxNxL forms the β-strand and the CxxLxxaLx or DxGaxxxxxxL sequences of the variable parts form α-helices. Tandem repeats are arranged consecutively and parallel to a common axis so that the structure adopts a curved shape resembling a horseshoe with α-helices lining its outer circumference and β-strands forming a parallel β-sheet along its inner circumference.[11–15] The U2A' protein contains five and a half LRR domains between residues 22 and 146.[16,17] The crystal structure of the splicesome U2B'-U2A' protein complex bound to a fragment of U2 small

Grant sponsor: Suhara Foundation; Grant sponsor: Grants-in-Aid for Scientific Research from the Ministry Education, Science, Sports, and Culture of Japan.

*Correspondence to: Norio Matsushima, School of Health Sciences, Sapporo Medical University, Sapporo, Hokkaido 060–8556, Japan. E-mail: matusima@shs.sapmed.ac.jp

Received 24 May 1999; Accepted 12 August 1999

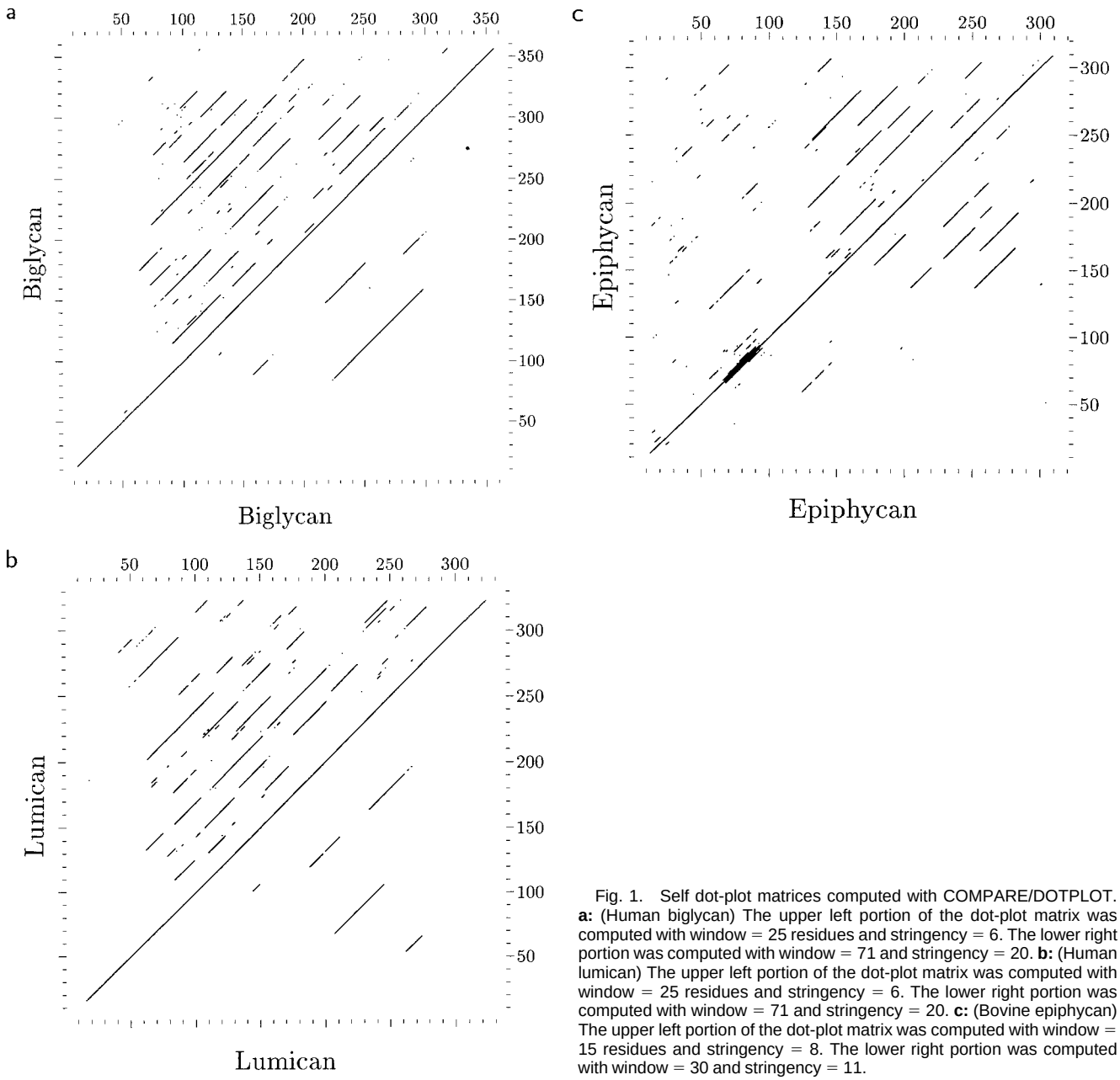


Fig. 1. Self dot-plot matrices computed with COMPARE/DOTPLOT. **a:** (Human biglycan) The upper left portion of the dot-plot matrix was computed with window = 25 residues and stringency = 6. The lower right portion was computed with window = 71 and stringency = 20. **b:** (Human lumican) The upper left portion of the dot-plot matrix was computed with window = 25 residues and stringency = 6. The lower right portion was computed with window = 71 and stringency = 20. **c:** (Bovine epiphycan) The upper left portion of the dot-plot matrix was computed with window = 15 residues and stringency = 8. The lower right portion was computed with window = 30 and stringency = 11.

nuclear RNA forms a solenoid similar to, but less regular than, that of ribonuclease inhibitor.[13] The parallel β-sheet within the LRR region is canonical; it contains all of the LxL sequences in the highly conserved segments, but the chains forming the outer circumference are either short α-helices, 3_{10} helices, or less regular structures.

Several small proteoglycans of the extracellular matrix contain LRRs, which carry chondroitin/dermatan sulfate or keratan sulfate glycosaminoglycan chains.[19–23] The sequences of biglycan/PG-I, decorin/PG-II, lumican, fibromodulin, PRELP (proline arginine-rich and leucine-rich protein), keratocan, osteoadherin, epiphycan/PG-Lb, and osteoglycin/osteoinductive factor consist of three main regions: an amino-terminal region with negatively charged glycosamino-glycans or tyrosine sulfate, a central domain with varying numbers LRRs, and a carboxyl-terminal region of poorly defined function. In all cases, the central domain is flanked by cysteine-rich clusters. These residues at the N-terminal region are similarly spaced in a stretch of about 20 amino acids with the consensus sequence: $Cx_{2-3}CxCx_{6-9}C$. These proteoglycans form three distinct subfamilies.[22,24] Class I consists of biglycan and decorin; they have the highest sequence similarity (≈57%). Class II has three subclasses; lumican and fibromodulin (IIA); PRELP and keratocan (IIB), and osteoadherin (IIC). Class III consists of epiphycan and osteoglycin. Class IV is more distantly related and consists of chondroadherin.[24]

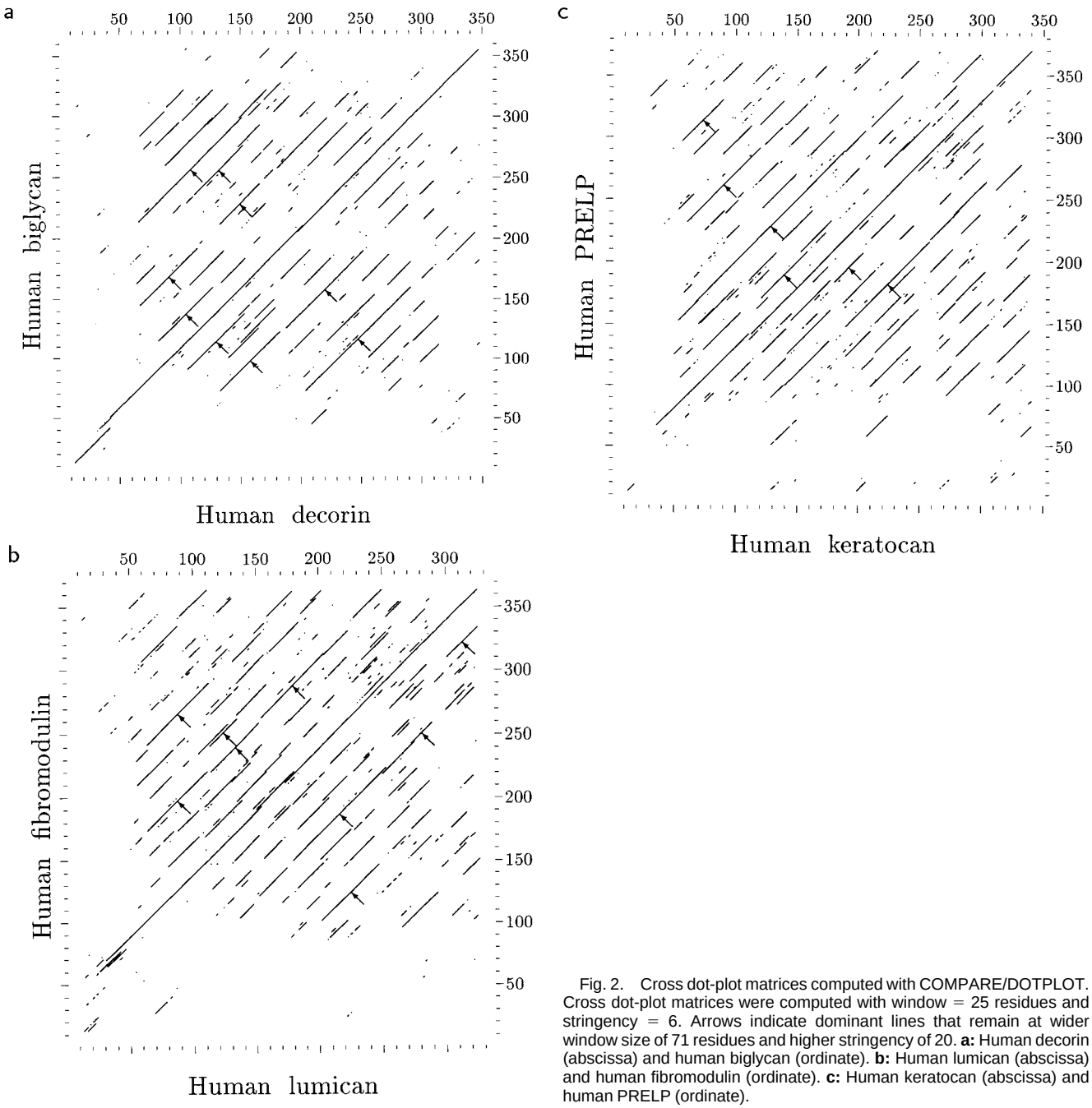


Fig. 2. Cross dot-plot matrices computed with COMPARE/DOTPLOT. Cross dot-plot matrices were computed with window = 25 residues and stringency = 6. Arrows indicate dominant lines that remain at wider window size of 71 residues and higher stringency of 20. **a:** Human decorin (abscissa) and human biglycan (ordinate). **b:** Human lumican (abscissa) and human fibromodulin (ordinate). **c:** Human keratocan (abscissa) and human PRELP (ordinate).

The assignments of number and "phasing (that is, which segment or residue is at the beginning of the first repeating unit) of the LRRs within these proteoglycans have varied among researchers. Nine to twelve LRRs were assigned within class I and II: biglycan,[25–29] decorin,[30–37] lumican,[38–41] fibromodulin,[42–44] PRELP,[45,46] keratocan,[47–49] and osteoadherin.[24] Six to eight LRRs were assigned within class III: epiphycan,[50,51] and osteoglycin[21–23]; however, the researchers who determined the osteoglycin sequence did not discuss repeat number.[52,53] Ten or 11 repeats were proposed for chondroadherin.[54–57]

The evolution of LRRs is not well understood. It is not even known whether all LRRs share a common ancestor. A probable pattern of evolution of LRR proteins would entail unequal crossing-over and duplications of gene fragments corresponding to prototypic LRR building blocks.[4] On the basis of the variation of lengths and consensus sequences of LRRs in different proteins, Kobe and Deisenhofer,[2,4] and Kajava[7] suggested that LRRs may have arisen independently several times.

In addition to ribonuclease inhibitor,[8–10] it was argued that higher-order LRRs repeats occur in *Trypansoma* VSG[58–60] and in tomato Cf-2.[61] Many proteins other than those with LRRs have super-repeats of tandem repeat

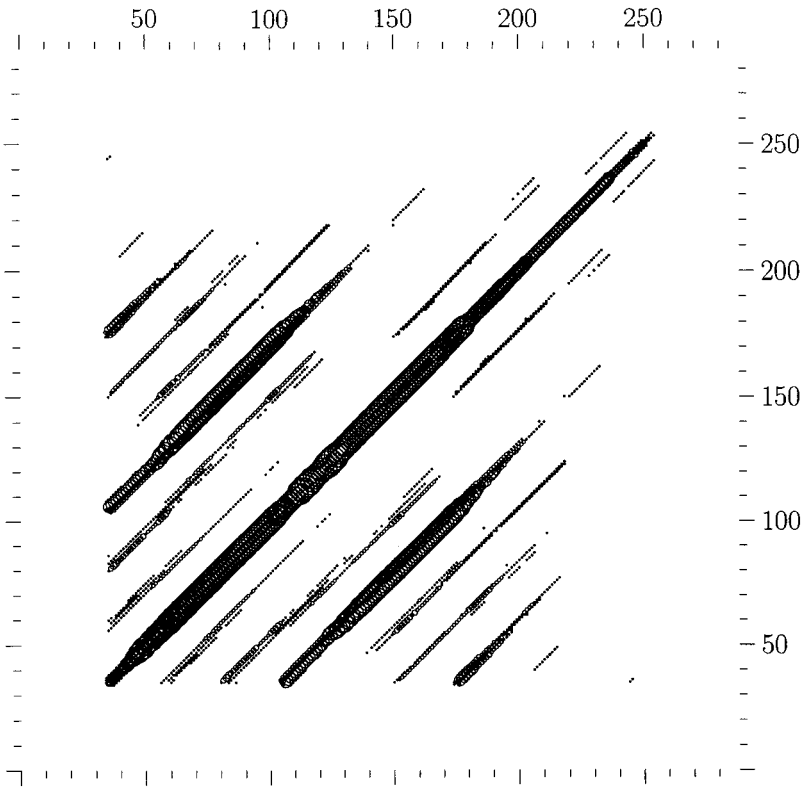


Fig. 3. Forty-two superimposed, cross dot-plot matrices from biglycan, lumican, decorin, fibromodulin, PRELP, keratocan (all from human), and bovine osteoadherin. The summed scores for the 42 (7×6) comparisons are represented by larger circles for higher scores. Before use in the COMPARE/DOTPLOT program, the non-LRR residues 1–81, 1–72, 1–58, 1–97, 1–94, 1–63, and 1–83 of biglycan, decorin, lumican and fibromodulin, PRELP, keratocan, and osteoadherin, respectively, were deleted. The window size is 71 residues and the stringency is 30.

units. Examples include nebulin,[62] titin/connectin,[63] mucin,[64] spectrin,[65] adhesive protein of mussles,[66] kininogens,[67] ice nucleation proteins,[68,69] and certain zinc finger proteins.[70]

We analyzed the LRR regions of these small proteoglycans in greater detail first to understand the character, number, and phasing of the LRRs. This analysis also provides insight into the evolutionary relationships of these proteins and of the hundreds of others that contain LRRs. Furthermore, these insights should help us understand the different structures and functions of these LRR proteins[21–23] and the other proteins that contain super-repeats based on tandem repeats.

MATERIALS AND METHODS

Amino Acid Sequences

The LRRs alignments were made for biglycan from seven species (human [Genbank accession number J04599], bovine [S82652], mouse [X53928], rat [U17834], dog [U83140], sheep [AF034842], and horse [AF035934]); for decorin from seven species (human [M14219], bovine [Y00712], mouse [X53929], rat [Z12298], rabbit [S76584], dog [U83141], and chicken [S76797]); lumican from five species (human [U18728], bovine [L11063], mouse [X84039], rat [S79461/AF013264], and chicken [M80584]); fibromodulin from five species (human [X72913], bovine [X16485], mouse [X94998], rat [X82152], and chicken [U34977]); PRELP from human [U29089/U41344]; keratocan from two species (human [AF065988/AF057301] and mouse [AF022256]); osteoadherin from human [U67279]; epiphycan from three species (mouse [U77127], bovine [D78274], and chicken [D10485]); osteoglycin from three species (human [M20774], bovine [M37974], and mouse [D31951]); and chondroadherin from four species (human [U96769], bovine [U08018], mouse [U96626], and rat [AF004953]).

Dot-Plots Analyses

Dot-matrix comparisons were made with COMPARE/DOTPLOT in the Genetics Computer Group software using the Blosum 62 scoring matrix. Window sizes and stringencies are indicated in figure legends.

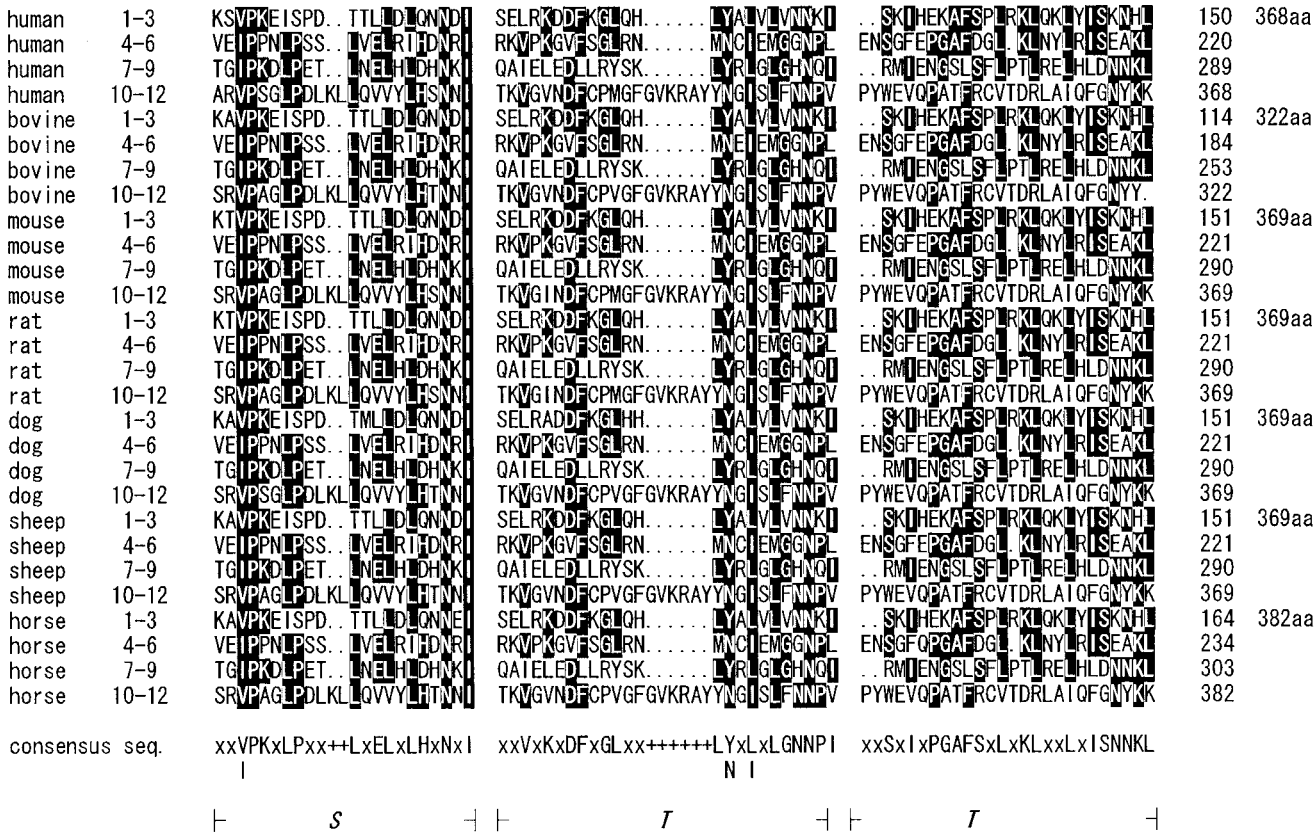
Sequence Analysis and Multiple Alignments

To recognize more divergent LRR repeat units, we relaxed the stringencies as much as possible. The lengths and the phasings of repeats were best recognized in multiple sequence alignments of individual LRR regions. On the basis of the results of these multiple sequence alignments of S and of T units, we aligned the super-repeats of LRRs. The final determinations of the locations of repeat sequences and their alignments with each other were refined by eye.

Phylogenic Analyses

Dendrograms of repeats and of super-repeats were computed with ProtML in the MOLPHY program.[71] The ProtML program infers evolutionary trees from protein sequence using the maximum likelihood method.[72] First, the NJdist program was used for inferring a tree from a

a. Biglycan



b. Decorin

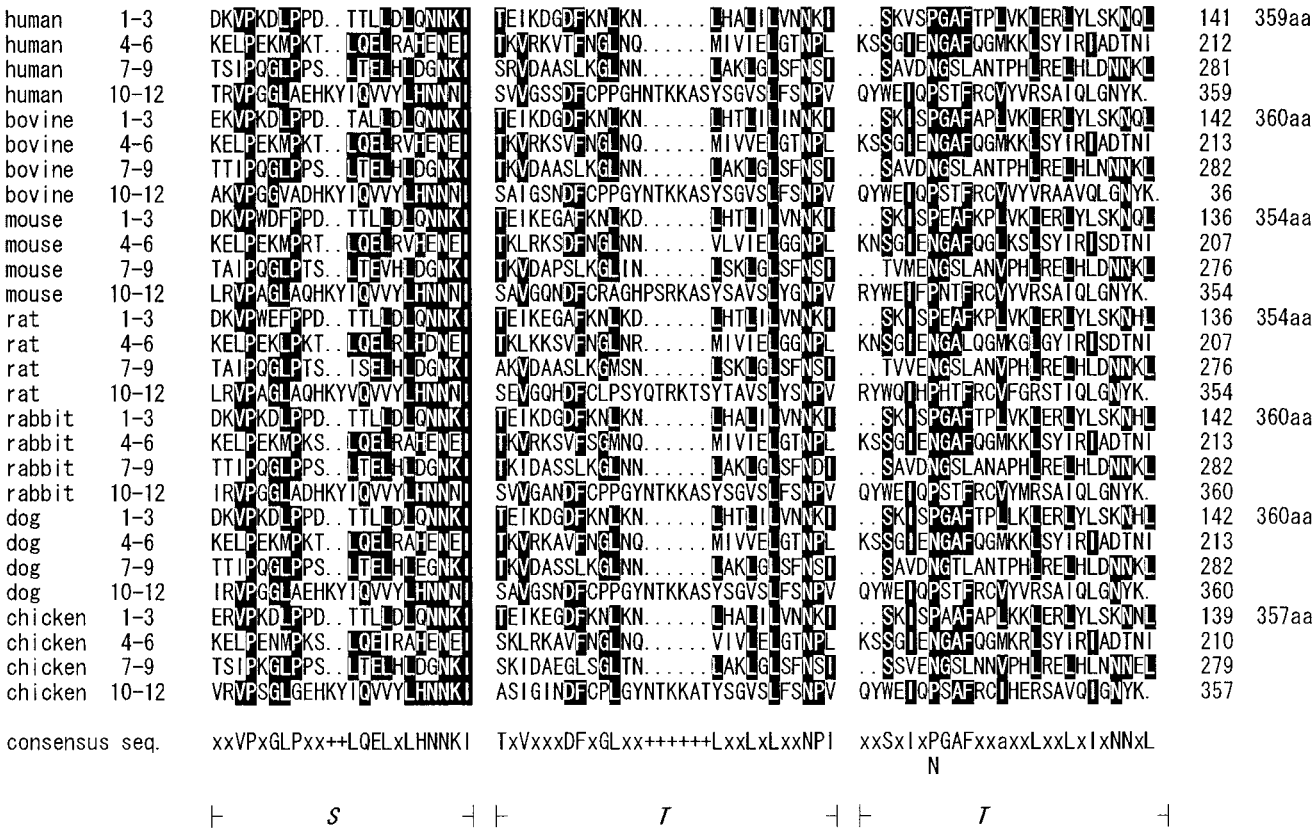


Figure 4.

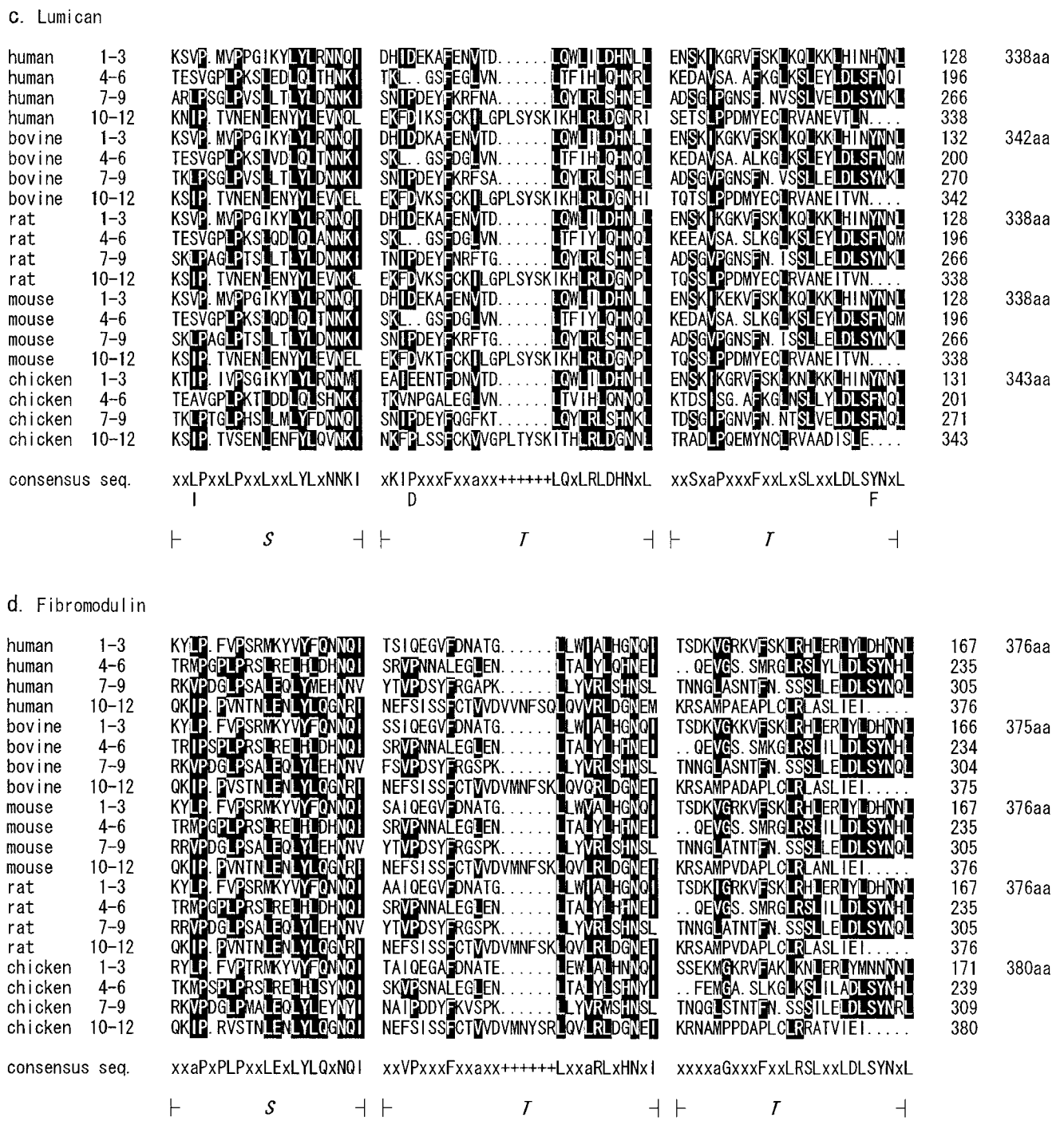


Fig. 4. Sequence alignments of super-repeats in nine subfamilies of small proteoglycans. **a:** biglycan; **b:** decorin; **c:** lumican; **d:** fibromodulin; **e:** PRELP; **f:** keratocan; **g:** osteoadherin; **h:** epiphycan; **i:** osteoglycin. For class I (biglycan, decorin, lumican, fibromodulin, and PRELP) and class II (keratocan and osteoadherin) alignments are made for each of four super-repeats: ***STT***. For class III (epiphycan and osteoglycin) alignments are made for ***ST***, ***T***, ***ST***, and ***ST***. For the nine alignments the left column indicates species; next repeat numbers shown in that row; sequence; and finally the number in the right-hand column is the residue number at the end of that super-repeat unit. Consensus residues that are based on 50% identity are highlighted with reverse-contrast. Gaps imposed by sequence alignment are indicated by (.). The consensus sequence is indicated at the bottom for each molecule. Possible insertions indicated by "+" is based on 25% occupancy. "a" indicates Leu, Ile, or Val. "x" represents nonconserved residues (<50% identity).

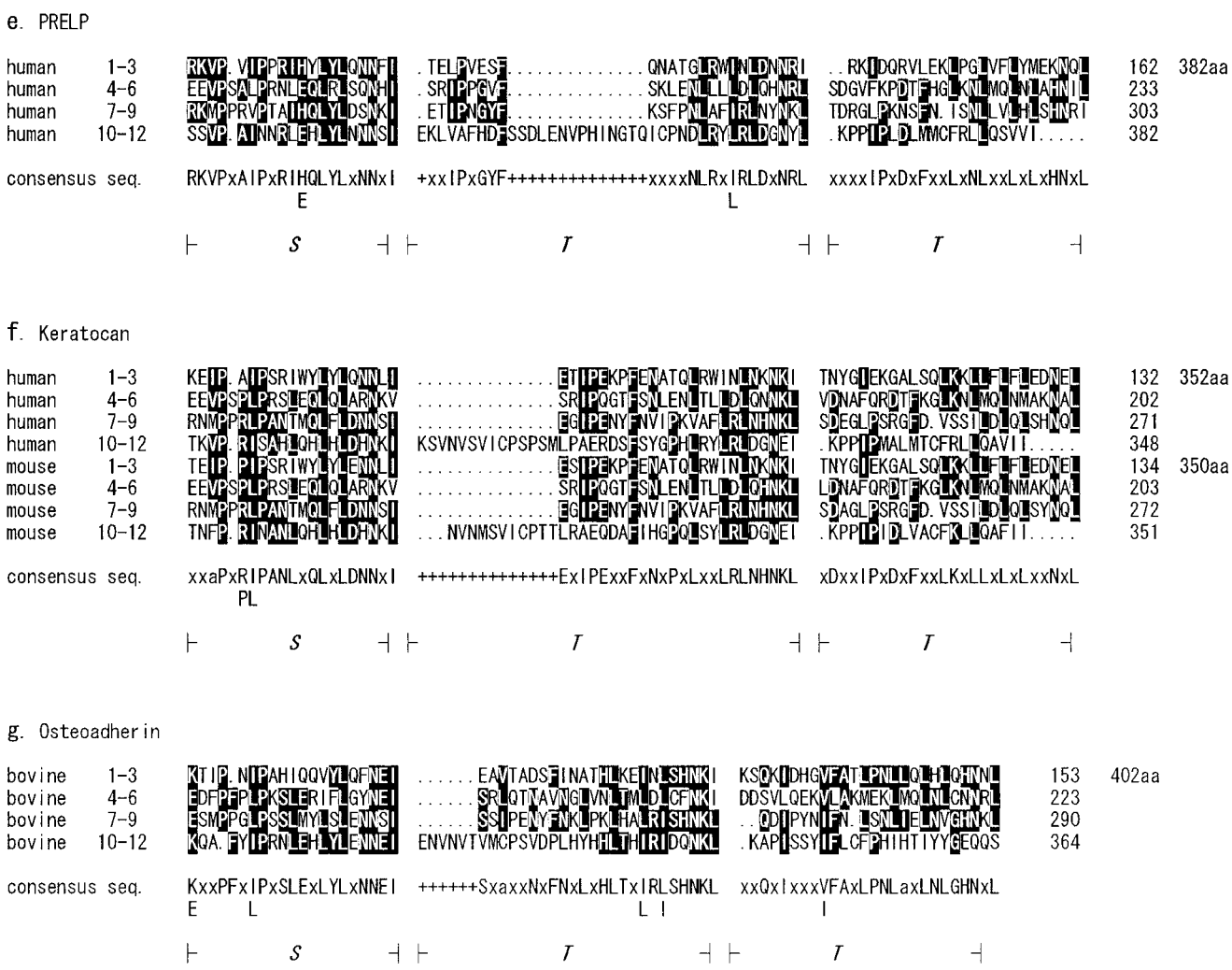


Figure 4. (Continued.)

distance matrix by the neighboring-joining method.[73] Next, the ProtML tree was obtained by repeated local rearrangement.

RESULTS

Self Dot-Plot Analyses

Self dot-matrix plots were computed for all 10 subfamilies as illustrated in Figure 1a–c for biglycan (class I), lumican (II), and epiphycan (III). The entire sequences were used in these calculations. The upper left halves were computed with smaller windows and lower stringencies; the lower right with larger windows and higher stringencies. For all subfamilies the upper left plots show multiple LRR repeats. For classes I and II the lower right plots strongly indicate super-repeats and more weakly hint at super-repeats for epiphycan and osteoglycin (class III). However, there is no trace of super-repeats for chondroadherin (results not shown). There are significant regions at the N-terminus and shorter regions at the C-terminus that have no diagonal lines; they precede or follow the regions of repeated LRRs.

Cross Dot-Plot Analyses

Cross dot-plots were computed for all pairs of the 10 subfamilies, as illustrated in Figure 2a–c for biglycan (class I) versus decorin (I), for lumican (IIA) versus fibromodulin (IIA), and for PRELP (IIB) versus keratocan (IIC), indicated by arrows, so that superimposed lines reinforce each other. The super-repeats seen in the self dot-plots (Fig. 1) are confirmed in the cross dot-plots. More important the super-repeat is about 70 residues for classes I. The cross dot-plot of epiphycan and osteoglycin (class III) and their self dot-plots appear not to reveal the existence of the super-repeat (results not shown). Furthermore, there is no super-repeat for chondroadherin (class IV) (results not shown). It is a fortuitous control of our analyses.

Superposition of 42 (7×6) cross dot-plots for the seven proteins of classes I and II emphasize their common super-repeat of 73 residues (Fig. 3), in which human small proteoglycans were used except for bovine osteoadherin. The non-LRR residues 1–81, 1–72, 1–58, 1–97, 1–94, 1–63,

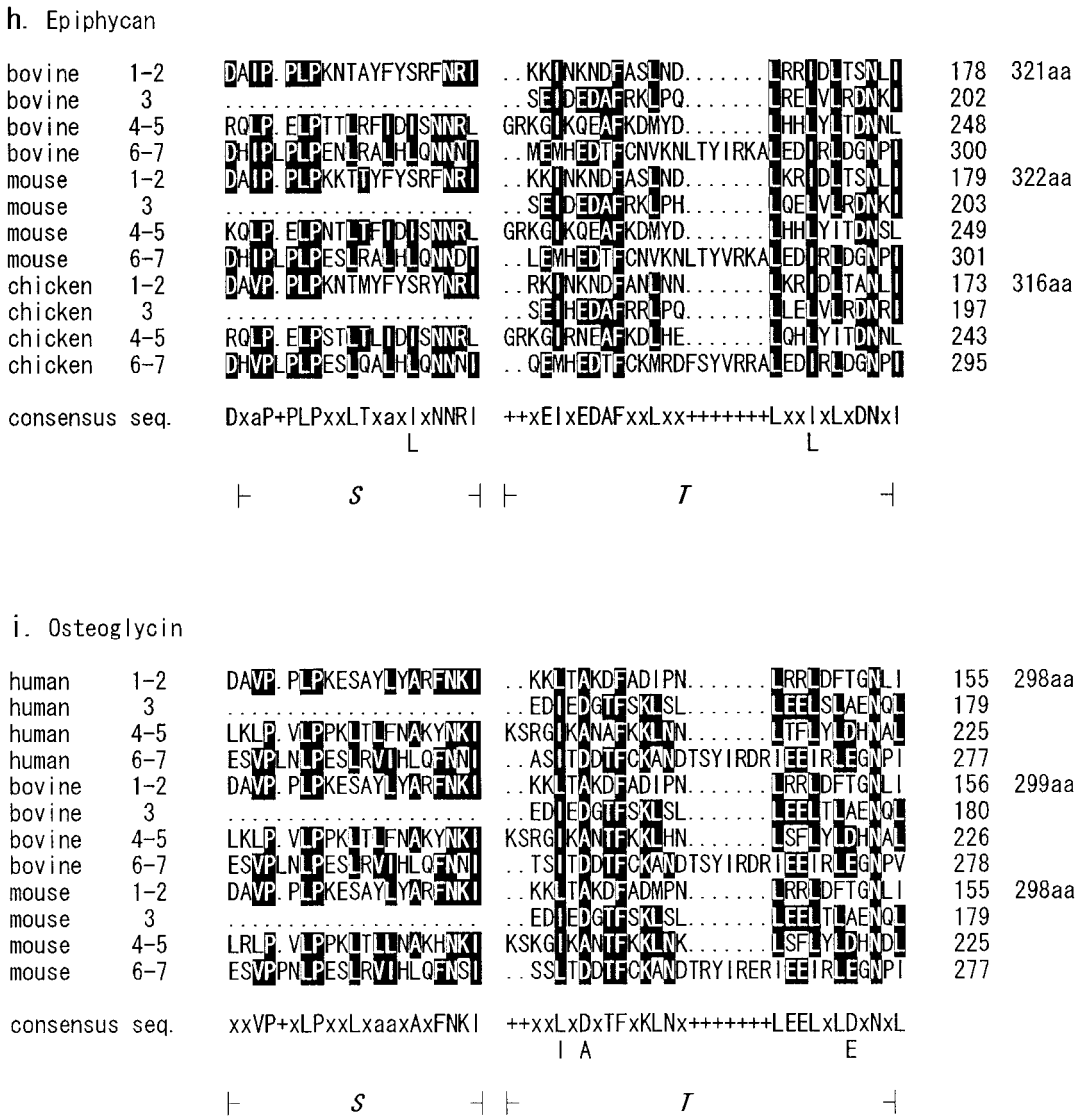


Figure 4. (Continued.)

and 1–83 from biglycan, decorin, lumican, and fibromodulin, PRELP, keratocan, and osteoadherin, respectively, were deleted before computing the final cross dot-plots shown in Figure 3.

Number and "Phasing" of LRRs

We made multiple sequence alignments among the super-repeats to determine the number and phasing of LRR units. We used the highly conserved segment, LxxLxLxxNxL, to identify LRR units and took account of the existence of super-motifs with about 70 residues in classes I and II. Classes I and II both contain 12; class III contains 7 LRRs. Chondroadherin has 10 LRRs but no super-repeats.

All LRRs regions within classes I, II, and III start with the variable part of type ***S*** LRR (Fig. 4). In contrast, LRRs within chondoroadherin forming class IV start with the highly conserved parts.

Two Types of LRRs

These sequence alignments reveal two types of LRR units (Fig. 4). The 10 ***T*** repeats of chondroadherin were not included because we focused on super-repeats. Type ***S*** has about 21; type ***T*** has about 26 residues. Figure 5 shows their consequence sequences based on all 34 sequences available for classes I, II, and III.

The type ***S*** LRR is observed in the first, fourth, seventh, and tenth units of the 12 tandem LRRs in classes I (biglycan and decorin) and II (lumican, fibromodulin, PRELP, keratocan, and osteoadherin). Type ***S*** was also observed in the first, fourth, and sixth LRRs of the super-repeats in class III (epiphycan and osteoglycin). All other LRR repeats are type ***T***.

Super-Repeat Units of LRRs

There are four super-repeats, ***STTSTTSTTSTT***, in classes I and II. There are also four super-repeats,

a.

Protein	LRR repeat unit number	Consensus sequence of *S*
Biglycan	1, 4, 7, 10	xxVPKxLPxxLxELxLHxNxI I
Decorin	1, 4, 7, 10	xxVPxGLPxxLQELxLHNNKI
Lumican	1, 4, 7, 10	KxLPxxLPxxLxxLYLxNNKI I
Fibromodulin	1, 4, 7, 10	xxaPxPLPxxLExLYLQxNQI
PRELP	1, 4, 7, 10	RKVPxAIPxRIHQLYLxNNxI E
Keratocan	1, 4, 7, 10	xxaPxRIPANLxQLxLDNNxI PL
Osteoadherin	1, 4, 7, 10	KxxPxxIPxSLExLYLxNNEI E L
Epiphycan	1, 4, 6	DxaP+xLPxxLTxaxIxNNRI L
Osteoglycin	1, 4, 6	xxVP+xLPxxLxaaxAxFNKI
Con. Con.		xxaPzxLPxxLxxLxLxxNxI

b.

protein	LRR repeat unit number	Consensus sequence of *T*
Biglycan	2, 5, 8, 11	xxVxKxDFxGLxxLYxLxLGNNPI N I
Decorin	2, 5, 8, 11	TxVxxxDFxGLxxLxxLxLxxNPI
Lumican	2, 5, 8, 11	xKIPxxxFxxaxxLQxLRLDHNxL D
Fibromodulin	2, 5, 8, 11	xxVPxxxFxxaxxLxxaRLxHNxI
PRELP	2, 5, 8, 11	xxIPNGxFxxxxNLRxIRLDxNRL L
Keratocan	2, 5, 8, 11	ExIPExxFxNaDxLxxLRLNHNKL
Osteoadherin	2, 5, 8, 11	SxaxxNxFNxLxxLxxIRLSHNKL L I
Biglycan	3. 6. 9, 12	xxSxIxPGAFSxLxKLxxLxISNNKL
Decorin	3. 6. 9, 12	xxSxIxPGAFxxaxxLxxLxIxNNxL N
Lumican	3. 6. 9, 12	xxSxaPGxxFxxLxSLxxLDLSYNxL F
Fibromodulin	3. 6. 9, 12	xxxxaGxxxFxxLRSLxxLDLSYNxL
PRELP	3. 6. 9, 12	xxxxIPxDxFxxLxNLxxLxLxHNxL
Keratocan	3. 6. 9, 12	xDxxIPxDxFxxLKxLLxLxLxxNxL
Osteoadherin	3. 6. 9, 12	xxQxIxxxVFAxLPNLaxLNLGHNxL I
Epiphycan	2, 3, 5, 7	++xEIxEDAFxxLxxLxxIxLxDNxI L
Osteoglycin	2, 3, 5, 7	++xxLxDxTFxKLNxLEELxLxDNxL I A E
Chondroadherin	1–10	xxLxxxAFxxLxxLxxLxLxNNxL H
Con. Con.		zzxxaxxxxFxxaxxLxxLxLxxNxL

Fig. 5. Consensus of consensus sequences for ***S*** and ***T*** domains in ten small proteoglycans. The consensus of consensus sequences for ***S*** (**a**) and ***T*** (**b**) domains was derived by selecting residues with 80% identity from the consensus sequences of the nine subfamilies whose sequences are shown in Figure 4. "x" represents nonconserved residues. "z" indicates frequent deletions.

ST_TSTST, in class III; however, the second super-repeat lacks one type ***S*** LRR. It is probable that this lack obscured the detection of the super-repeats by the dot-plot analysis. The consensus sequence of each super-repeat unit, based on a minimum of 50% identity or higher, is annotated at the bottom of each super-repeat unit (Fig. 4).

Phylogenetic Analysis of Individual Super-Repeat Units

To establish the possible origin of classes I, II, III, and IV of the small proteoglycans, three phylogenetic trees were constructed (Fig. 6a–c). Figure 6a shows the dendrogram for the four super-repeat units in classes I and II. Class I contains seven biglycans and seven decorins. Class II contains five lumicans, five fibromodulins, one PRELP, two keratocans, and one osteoadherin. All of the 28 first ***STT*** units (***STT1***) of classes I and II cluster together. Correspondingly, the 28 second ***STT*** units (***STT2***) of classes I and II themselves cluster together. This congruence also holds for the 28 third ***STT*** units (***STT3***) and for the 28 fourth ***STT*** units (***STT4***). In each of the four major branches biglycans and decorins are closely related; lumicans, fibromodulins, PRELP, keratocans, and osteoadherin are closely related. Thus, classes I and II form distinct but congruent subfamilies. Super-repeat ***STT1*** is

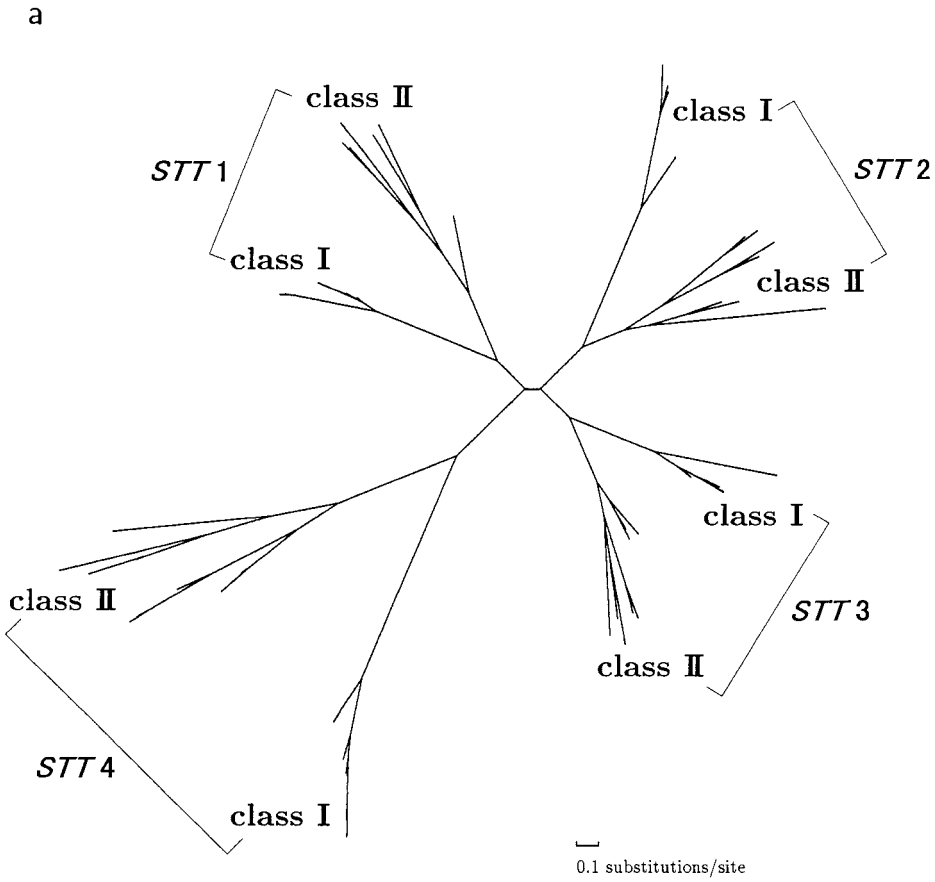


Fig. 6. Dendrograms of individual super-repeats within each of classes I/II and III, and of individual LRRs within class IV. (**a**) Class I: ***STT*** from seven biglycans (human, bovine, mouse, rat, dog, sheep, and horse), seven decorins (human, bovine, mouse, rat, rabbit, dog, and chicken); class II: five lumicans (human, bovine, mouse, rat, and chicken, five fibromodulins (human, bovine, mouse, rat, and chicken), human PRELP, two keratocans (human and mouse), and human osteoadherin. (**b**) Class III: ***T*** (only) from three epiphycans (mouse, bovine, and chicken) and three osteoglycins (human, bovine, and mouse). (**c**) Class IV: ***T*** from four chondroadherins (human, bovine, mouse, and rat). Branch lengths were computed by using the JTT-F model of sequence evolution.[96] The bar indicates 0.1 substitutions per position. The first, second, third, and fourth units of four super-repeats ***STTSTTSTTSTT***, in classes I and II are indicated by ***STT*****1**, ***STT*****2**, ***STT*****3**, and ***STT*****4**, respectively. The type ***T***-LRR of the first, second, third, and fourth units of four super-repeats, ***ST_TSTST***, in class III are indicated by (***S***)***T*****1, *T*****2,** (***S***)***T*****3**, (***S***)***T*****4**, respectively. The first to tenth units in LRRs, $(\boldsymbol{T})_{10}$, in chondroadherin (class IV) are indicated by ***T*****1**, ***T*****2**, ***T*****3**, ***T*****4**, ***T*****5**, ***T*****6**, ***T*****7**, ***T*****8**, ***T*****9**, and ***T*****10**, respectively. The following sequences are completely identical to each other; ***STT*****1** between mouse and rat biglycans, and between bovine and sheep biglycans, ***STT*****2** among human, mouse, rat, dog, and sheep biglycans, and between mouse and rat fibromodulins; ***STT*****3** among human, bovine, horse, mouse, rat, dog, and sheep biglycans, and between mouse and rat fibromodulins; ***STT*****4** between horse and sheep biglycans, and between mouse and rat biglycans; (***S***)***T*****1** between mouse and bovine osteoglycins; (***S***)***T*****3** between mouse and bovine osteoglycins; ***T*****2** between human and bovine chondroadherins, and between rat and mouse chondroadherins; ***T*****3** between bovine and rat chondroadherins; ***T*****4** between human and bovine chondroadherins, and between rat and mouse chondroadherins; ***T*****5** between rat and mouse chondroadherins; ***T*****8** among human, bovine, rat, and mouse chondroadherins.

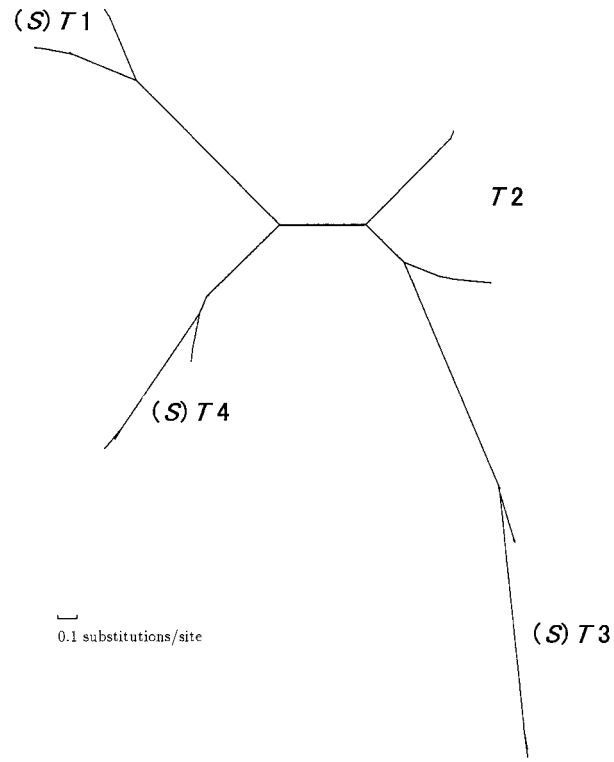
most closely related to ***STT4*** and that ***STT2*** is most closely related to ***STT3*** (Fig. 6a).

As noted, the second super-repeats of epiphycan and osteoglycin lack an ***S*** unit. We, therefore, constructed a dendrogram for only the four ***T*** units of individual super-repeat units in three epiphycans and three osteoglycins (Fig. 6b). The six species of the epiphycan and osteoglycin show good, but not perfect congruence.

Four chondroadherins contain 10 LRRs consisting of only type ***T***. The sequences from the four species are similar and for several domains identical among two to four species. Figure 6c shows the dendogram for the 10 ***T*** units in four chondroadherins. All of the four first **T** units cluster together as is also the case for domains 2–10. The four second ***T*** units (**T2**) and the four fifth **T** units (**T5**) are more closely related; the four fourth **T** units (**T4**) and the four seventh **T** units (***T*****7**) are closely related. The eighth and ninth units, ***T***8 and ***T***9, are also closely related.

The most parsimonious interpretation of the evolution of the nine small proteoglycans (classes I, II, and III) is shown diagrammatically in Figure 7. The evolution involves duplication and fusion of a UR/LRR domain with subsequent divergence, in the ancestor of small proteoglycans to form the UR/***ST*** super-repeat unit (**1** and **2** in Fig. 7). It duplicated without fusion to form the precursor ***ST*** (classes I and II) and the precursor ***ST*** (III) super-repeats (**3** in Fig. 7). Duplication and fusion of the two LRR domains form the four LRR domains, ur***STST***, and then

b



c

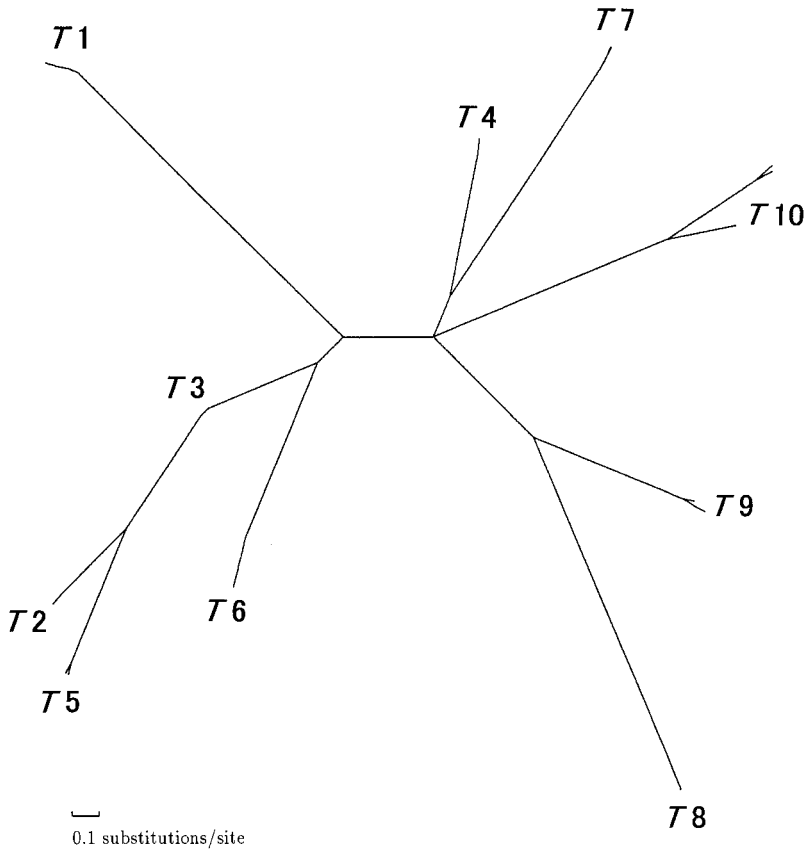


Figure 6. (Continued.)

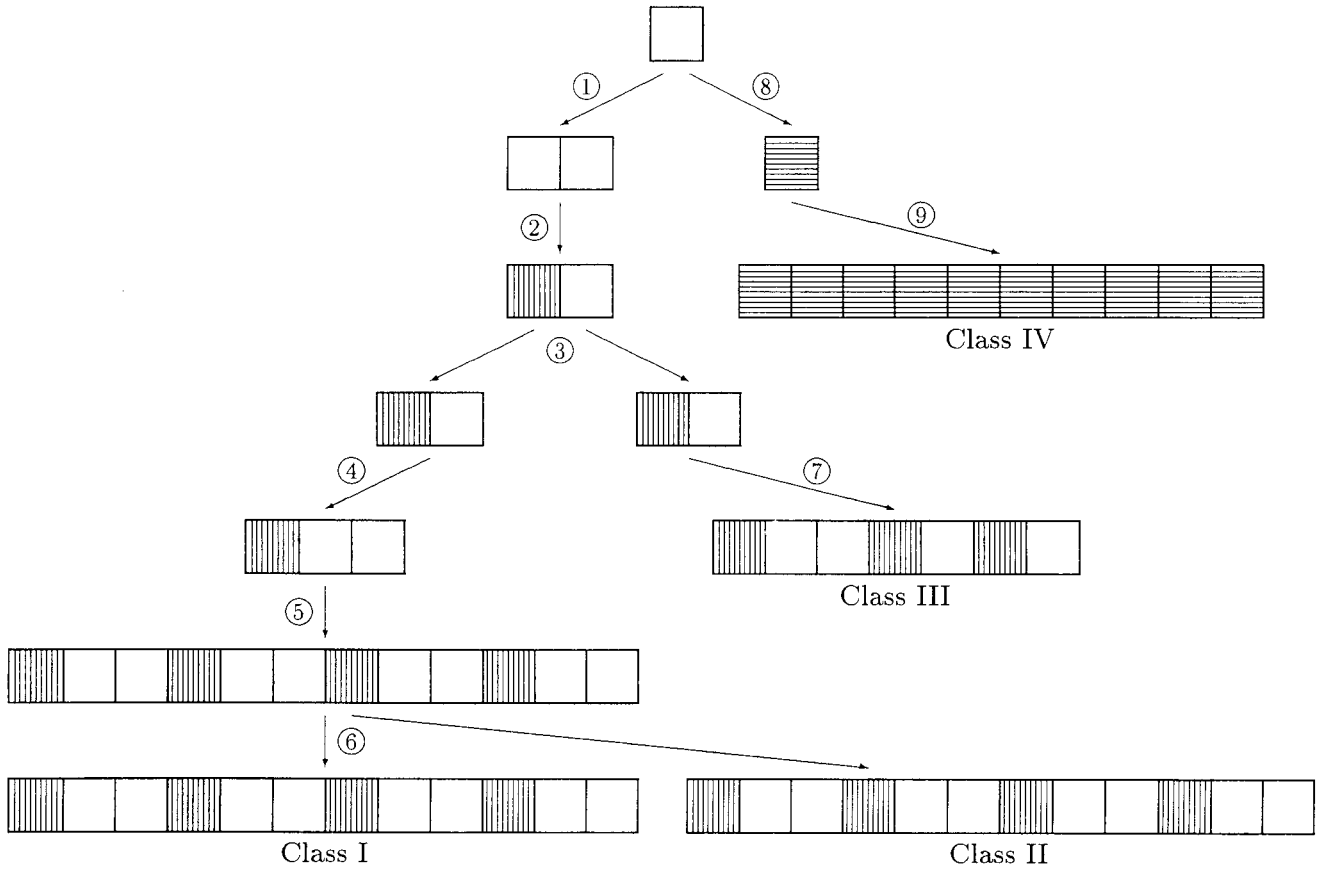


Fig. 7. Evolution of the super-repeats of LRRs of the small glycoproteins. The most parsimonious interpretation of the evolution of the super-repeats of LRRs of the small glycoproteins involves: 1. Duplication and fusion of the ***T***-like UR/LRR gene in the ancestor of all small proteoglycans. 2. Divergence of the two LRR domains to form ur***ST***. 3. Duplication without fusion to form the precursor ***ST*** (classes I and II) and the precursor ***ST*** (III) super-repeats. 4. Duplication and fusion of the two LRR domains to form the four LRR domains, ur***STST*** and then unequal crossing over of the four LRR domains with ur***STST*** to form ur***STT*** (classes I and II). Alternative gene conversion of the second unit of the ur***ST*** to form ur***STT*** (classes I and II). 5. Duplication and fusion, and second duplication and fusion of ur***STT*** to form the precursor of 12 LRRs domains. 6. Duplication without fusion to form the precursor $(\textbf{\textit{STT}})_4$ super-repeats of classes I and II. 7. Duplication, fusion, and deletion of ur***ST*** to form the precursor of seven LRR domains of class III. 8. Divergence of the UR/LRR gene in the ancestor of all small proteoglycans to form the precursor ***T*** of 10 LRR domains within chondroadherin. 9. Several duplications and fusions, and/or deletions of the precursor ***T*** to form the 10 LRRs within chondroadherin.

unequal crossing over of the four LRR domains with ur***STST*** form ur***STT*** (classes I and II). Alternatively, gene conversion of the second unit of the ur***ST*** forms ur***STT*** (**4** in Fig. 7). Subsequent duplications, fusions, and a second duplication and fusion formed the precursor of 12 LRR domains (**5** in Fig. 7). It duplicated without fusion to form classes I and II (**6** in Fig. 7). Subsequent duplications of the ur***ST***, fusions, and deletions formed class III (**7** in Fig. 7). It appears that the precursor ***S*** subunit that gave rise to chondroadherin (class IV) diverged before the establishment of UR/***ST*** (**8** in Fig. 7), because the phasing of LRRs in class IV differs from that of LRRs in classes I, II, and III. Several duplications and fusions, and/or deletions of the precursor ***T*** occurred to form the 10 LRRs within chondroadherin (**9** in Fig. 7).

DISCUSSION

Number of LRRs Within Small Proteoglycans

The comparative sequence analysis indicates that classes I and II have 12 LRRs. This is consistent with the repeat number proposed for human biglycan and decorin by Kresse et al.,[20] although their and our phasings differ. In addition to cysteine-rich clusters ($Cx_{2-3}CxCx_{6-9}C$) at the N-termini, there are conserved cysteine-clusters ($Cx_{31-39}C$) at the C-termini. The present results mean that the $Cx_{31-39}C$ sequence is contained in the LRR region.

Dot-plots indicate that there are significant regions at the N-terminus and shorter regions at the C-terminus that have no diagonal lines (Figs. 1 and 2). No diagonal lines at the C-terminus is due mainly to weak homology of the eleventh and twelfth of the 12 tandem LRRs. The eleventh unit has additional insertions of 6–15 residues, and the twelfth unit shows deletions of 1–5 at the C-termini (Fig. 4). The absence of diagonal lines at the N-terminus is due to non-LRR domains, including the cysteine-rich clusters in addition to the window size used.

The sequence alignment indicates that class III has four super-repeats, *ST_TSTST*, the second of which is incom-

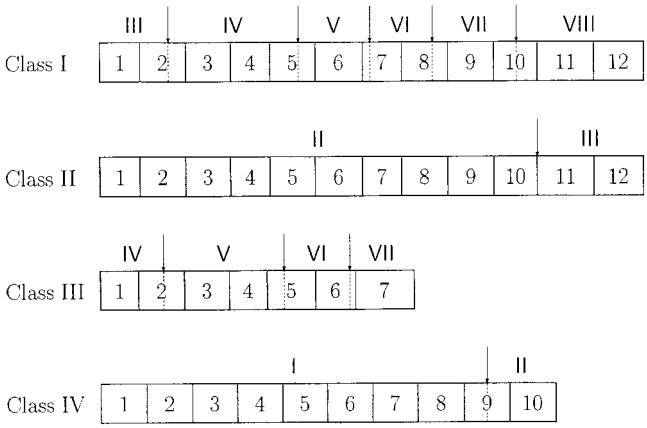


Fig. 8. Exon distribution and the positions of exon/intron boundaries in LRRs domains of four classes of the small glycoproteins. Class I (human biglycan and human decorin); class II (human lumican, human fibromodulin, mouse keratocan, and human PRELP); class III (mouse epiphycan); class IV (human and mouse chondroadherins). Individual LRRs are indicated by box in which an Arabic numeral shows each unit number of LRRs. Exons for each class are indicated by a Roman numeral. The arrowheads indicate the positions of exon/intron boundaries in the LRRs domains.

plete. This "incompleteness" prevents the detection of super-repeats within class III by the dot-plot analysis.

Features of Two Types of LRRs Within Small Proteoglycans

Type **S** LRR with 21 residues is characterized by the consensus Pro at positions 4 and 8, and consensus aliphatic residues at positions 3 and 7 (Figs. 4 and 5). Position 5 is a gap. Position 8 favors Asn. Interestingly, position 21 is Ile instead of Leu seen in other LRR proteins. Such a short repeating unit has been observed in most of LRRs in Ipa4.5, and Ipa7.8 from *Shigella flexneri*[74] and yopM from *Yersinia pestis*.[75] The repeating number of LRRs is 8 for Ipa4.5, 7 for Ipa7.8, and 13.5 for yopM. The consensus sequence of LRRs within Ipa4.5, Ipa7.8, and yopM are represented by xxIPxLPxxLxxLxLxxNxL, which is almost identical to type ***S*** LRR found here. Human and mouse CD14[76,77] and mouse RP105[78] contain at least one ***S***-like LRR. This type of short LRR is characteristic of six subfamilies proposed by Kajava.[7]

Type ***T*** LRR with 26 residues is characterized by the consensus "a" at positions 5 and 13 and the consensus Phe at position 10 (Figs. 4 and 5). Positions 1 and 2 frequently have gap and, thus, this type is 24 residues in length. Positions 6 and 23 favor Pro and Asn or His, respectively. Repeating units similar to type ***T*** LRR have been observed in a large number of LRR proteins such as chaoptin,[79] connectin,[80,81] Toll,[82] tartan,[83] NLRR-1,[84] ALS,[85,86] and 18 wheeler.[87,88] The type ***T*** LRR belongs to a family of the typical 24-residue LRR that is the most populated unit.[5,7] The length of the typical 24-residue LRR ranges from 20 to 27.[7] This allows insertions and deletions at some positions as seen in positions 12 and 14 of type ***T*** LRR.

Predicted Structures of Two Types of LRRs Within Small Proteoglycans

Two available structures of LRR proteins reveal that the LxL sequences of highly conserved LxxLxLxxNxL forms a β-strand.[11–15,18] It is likely that the β-strand appears in the corresponding sequence of the LRRs units within the small proteoglycans. However, it appears that the structure of variable parts of type ***S*** and ***T*** LRR units differ from each other. Kajava et al.[5] proposed that the variable part of the typical 24-residue LRR that is similar to type ***T*** contains an α-helix, although the α-helix is shorter than those seen in the ribonuclease inhibitor.

By contrast, it was proposed that short and proline-rich LRRs that are almost identical to type ***S*** contain polyproline II (*p*) instead of α-helix.[7] Thus, it is reasonable to predict that type ***S*** and ***T*** LRRs have β–*p* and β–α structural units, respectively.

Rotary shadowed electron micrographs of the small proteoglycans (decorin, lumican, and fibromodulin)[89,90] reveal the unique horseshoe-shaped structure seen in the structure of ribonuclease inhibitor, as expected from the predicted structural units. The overall width of the horseshoes of these three small proteoglycans (5–6 nm for decorin and lumican, and >6 nm for fibromodulin) is comparable with 7 nm of ribonuclease inhibitor. In fact, the three-dimensional model of human decorin[91] was constructed on the basis of the crystal structure of the ribonuclease inhibitor and, moreover, predicted binding of collagen triple helix to the inner surface of the decorin arch.

Evolution of Super-Repeats of LRRs Within Small Proteoglycans

As noted, super-repeat ***STT*1** is most closely related to ***STT*4** and ***STT*2** is most closely related to ***STT*3** (Fig. 6a). This branch order, which has strong statistical significance, implies an evolution more complex than two gene duplication, fusion events—ur***STT***->***STT*od-*STT*ev->*STT*1-*STT*2-*STT*3-*STT*4** (**5** in Fig. 7). At least one more gene rearrangement was involved, e.g., ur***STT***->***STT*od-*STT*ev->*STT*od–*STT*ev-*STT*od–*STT*ev-switch->*STT*od–*STT*ev-*STT*ev-*STT*od->*STT*1-*STT*2-*STT*3-*STT*4.**

The branch order of domain ***T*** in class III is illustrated in Figure 6b. As is the case for classes I/II the 1,4/2,3 similarity implies a more complex evolution than two gene duplication/fusion events. As indicated in Figure 7, the paths leading to subfamilies I/II and to III diverged early (**3** in Fig. 7). It is especially interesting that the rather unusual 1,4/2,3 pattern of relatedness occurs independently in both lineages (**7** in Fig. 7). Whether this reflects structural constraints or simply the coincidence of two unusual events we do not know. The branch order of domain B in calss IV also indicates a complex sequence of gene duplications, fusions, and deletions (**9** in Fig. 7).

Genomic Organization of Small Proteoglycans

The dendrogram of the LRRs and the patterns of the super-repeats based on amino acid sequences indicate that

the 10 small proteoglycans form four distinct subfamilies. The organization and distribution of exons reinforce this classification (Fig. 8). The human[26] and murine biglycan[92] and human decorin[32] genes (class I) are all encoded by eight distinct exons of which exons III–VIII encode the LRRs. The human,[38] murine,[93] and chicken lumican,[94] human fibromodulin,[44] human PRELP,[45] and mouse keratocan[48] genes (class II) are all encoded by three exons of which II and III encode the LRRs. The mouse PG-Lb/epiphycan gene (class III) has seven exons; the LRRs are encoded by exons IV– VII.[95] The human and mouse chondroadherin gene (class IV) has three exons; I and II encode the LRRs.[56,57]

The exon/intron boundaries within regions encoding LRRs are shown in Figure 8. In human biglycan[26] and human decorin[32] introns split the highly conserved 11-residue segments of the second, fifth, eighth, and tenth LRRs and the variable part of the seventh LRR without any recognizable pattern. The exon/intron boundaries in the class II LRRs are located at the N-termini of the variable part of the eleventh LRRs. In mouse epiphycan introns split the variable parts of the second and fifth units and the highly conserved 11-residue segment of the sixth LRR.[91] Human and mouse chondroadherin genes[56,57] (class IV) have one intron at the N-terminus of the ninth LRR.

The patterns of cysteine-rich clusters at the N-termini also differ in the four classes: Cx_3CxCx_6C in class I, Cx_3CxCx_9C in II, Cx_2CxCx_6C in III, and Cx_3CxCx_8C in class IV.

Evolutionary Implication of Super-Motifs in LRRs

Most of the tandem LRRs in ribonuclease inhibitor,[8–10] *Trypanosoma* VSG,[58–60] and tomato Cf-2[61] and the nine small proteoglycans (classes I, II, and III) contain super-motifs. The super-motifs of LRRs within ribonuclease inhibitor, VSG, and Cf-2 consist of two alternating repeats. The present analyses indicate that, rather than a simple duplication in an LRR unit, the LRR regions within classes I, II, and III were derived from tandemly repeated super-motifs that encode two LRR units, ***ST***.

Super-repeats of tandem repeat units also exist in the sequences within many proteins other than those with LRRs.[58–66] It has been proposed that α- and β-spectrins arose after the establishment of a large super-motif composed of eight of the 106-residue motifs.[65] We suggest that such super-motifs arose in many other proteins that have tandem repeat units.

CONCLUSIONS

1. There are two types of LRRs in the 10 subfamilies of small proteoglycans. Their consensus sequences are: ***S*** xxaPzxLPxxLxxLxLxxNxI and ***T*** zzxxaxxxxFxxaxxLxx-LxLxxNxL.
2. The super-repeat pattern of classes I and II is: ***STTSTTSTTSTT***. The super-repeat pattern of class III is: ***ST_TSTST***.
3. The division of the small proteoglycans into classes I (biglycan and decorin), II (lumican, fibromodulin, PRELP, keratocan, and oseoadherin), and III (epiphycan and osteoglycin) is consistent with the dendrogram of their LRRs, the patterns of their super-repeats, the distributions of introns in their encoding genes, and the patterns of cysteine-rich clusters. Chondroadherin forms a distinct class IV. It has 10 ***T*** domains with no super-repeats.
4. The most parsimonious interpretation of the evolution of these 10 LRR proteins involves duplication and fusion of a UR/LRR domain with subsequent divergence, in the ancestor of the four classes of small proteoglycans to form the UR/***ST*** super-repeat unit. It duplicated without fusion to form the precursor ***ST*** (classes I and II) and the precursor ***ST*** (III) super-repeats. The subsequent duplications, fusions, and/or deletions to form classes I–III are diagrammed in Figure 7. It appears that the precursor ***T*** subunit that gave rise to chondroadherin diverged from the UR/LRR domain.

The present analysis provides a reference for evaluation of the super-repeats observed in other LRR-containing proteins and in other proteins that have super-repeats based on tandem repeats of homologous domains.

ACKNOWLEDGMENTS

The authors thank Prof. Hideo Kano (Sapporo Medical University) and Prof. Katsutoshi Nitta (Hokkaido University) for their helpful support and Dr. Susumu Nakayama (Nagasaki University) for assistance with the MOLPHY program. Dot-plot analyses were performed by the Genetics Computer Group software at Genome Center of Tokyo University. This work was supported in part by the Suhara Foundation (to N.M.) and by Grants-in-Aid for Scientific Research from the Ministry Education, Science, Sports and Culture of Japan (to N.M.). We also thank Hiromi Konno for her help in preparing the manuscript.

REFERENCES

1. Takahashi N, Takahashi Y, Putnam FW. Periodicity of leucine and tandem repetition of a 24-amino acid segment in the primary structure of leucine-rich α 2-glycoprotein of human serum. Proc Natl Acad Sci USA 1985; 82:1906–1920.
2. Kobe B, Deisenhofer J. Proteins with leucine-rich repeats. Curr Opin Struct Biol 1995;5:409–416.
3. Buchanan SG, Gay NJ. Structural and functional diversity in the leucine-rich repeat family of proteins. Prog Biophys Mol Biol 1996; 65:1–44.
4. Kobe B, Deisenhofer J. The leucine-rich repeat: a versatile binding motif. Trends Biochem Sci 1994;19:415–421.
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. Ohyanagi T, Matsushima N. Classification of tandem leucine-rich repeats within a great variety of proteins. FASEB J 1997;11:A949.
7. Kajava AV. Structural diversity of leucine-rich repeat proteins. J Mol Biol 1998;277:519– 527.
8. Lee FS, Fox EA, Zhou HM, Strydom DJ, Vallee BL. Primary structure of human placental ribonuclease inhibitor [erratum appears in Biochemistry 28:7138, 1989]. Biochemistry 1988;27: 8545–8553.
9. Schneider R, Schneider-Scherzer E, Thurnher M, Auer B, Schweiger M. The primary structure of human ribonuclease/angiogenin inhibitor (RAI) discloses a novel highly diversified protein superfamily with a common repetitive module. EMBO J 1988;7:4151–4156.

10. Hofsteenge J, Kieffer B, Matthies R, Hemmings BA, Stone SR. Amino acid sequence of the ribonuclease inhibitor from porcine liver reveals the presence of leucine-rich repeats. Biochemistry 1988;27:8537–8544.
11. Kobe B, Deisenhofer J. Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. Nature 1993;366: 751–756.
12. Kobe B, Deisenhofer J. A structural basis of the interactions between leucine-rich repeats and protein ligands. Nature 1995;374: 183–186.
13. Kobe B, Deisenhofer J. Mechanism of ribonuclease inhibition by ribonuclease inhibitor protein based on the crystal structure of its complex with ribonuclease A. J Mol Biol 1996;264:1028–1043.
14. Kobe B, Ma Z, 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;24:288–291.
15. Papageorgiou AC, Shapiro R, Acharya KR. Molecular recognition of human angiogenin by placental ribonuclease inhibitor-an X-ray crystallographic study 2.0 Å resolution. EMBO J 1997;16:5162–5177.
16. Sillekens PT, Beijer RP, Habets WJ, van Verooij WJ. Molecular cloning of the cDNA for the human U2 snRNA-specific A' protein. Nucleic Acids Res 1989;17:1893–1906.
17. Cross M, Wieland B, Palfi Z, et al. The trans-splicesomal U2 snRNP protein 40K of *Trypanosoma bruccei*: cloning and analysis of functional domains reveals homology to a mammalian snRNP protein. EMBO J 1993;12:1239–1248.
18. 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.
19. Kresse H, Hausser H, Schonherr E. Small proteoglycans. Experientia 1993;49:403–416.
20. Kresse H, Hausser H, Schonherr E. Small proteoglycans. EXS 1994;70:73–100.
21. Iozzo RV, Murdoch AD. Proteoglycans of the extracellular environment: clues from the gene and protein side offer novel perspectives in molecular diversity and function. FASEB J 1996;10:598–614.
22. Iozzo RV. The family of the small leucine-rich proteoglycans: key regulators of matrix assembly and cellular growth. Crit Rev Biochem Mol Biol 1997;32:141–174.
23. Hocking AM, Shinomura T, McQuillan DJ. Leucine-rich repeat glycoproteins of the extracellular matrix. Matrix Biol 1998;17:1–19.
24. Sommarin Y, Wendel M, Shen Z, Hellman U, Heinegard D. Osteoadherin, a cell-binding keratan sulfate proteoglycan in bone, belongs to the family of leucine-rich repeat proteins of the extracellular matrix. J Biol Chem 1998;273:16723–16729.
25. Fisher LW, Termine JD, Young MF. Deduced protein sequence of bone small proteoglycan I (biglycan) shows homology with proteoglycan II (decorin) and several nonconnective tissue proteins in a variety of species. J Biol Chem 1989;264:4571–4576.
26. Fisher LW, Heegaard, AM, Vetter U, et al. Human biglycan gene: putative promoter, intron-exon junctions, and chromosomal localization. J Biol Chem 1991;266:14371–14377.
27. Dreher KL, Asundi V, Matzura D, Cowan K. Vascular smooth muscle biglycan represents a highly conserved proteoglycan within the arterial wall. Eur J Cell Biol 1990;53:296–304.
28. Neame PJ, Choi HU, Rosenberg LC. The primary structure of the core protein of the small leucine-rich proteglycan (PG-I) from bovine articular cartilage. J Biol Chem 1989;264:8653–8661.
29. Xu JH, Radhakrishnamurthy B, Srinivasan, SR, Berenson, GS. Primary structure of bovine aorta biglycan core protein deduced from cloned cDNA. Biochem Mol Biol Int 1995;37:263–272.
30. Li W, Vergnes JP, Cornuet PK, Hassell JR. cDNA clone to chick corneal chondroitin/dermatan sulfate proteoglycan reveals identity to decorin. Arch Biochem Biophys 1992;296:190–197.
31. Asundi VK, Dreher KL. Molecular characterization of vascular smooth muscle decorin: deduced core protein structure and regulation of gene expression. Eur J Cell Biol 1992;59:314–321.
32. Danielson KG, Fazzio A, Cohen I, Cannizzaro LA, Eichstetter I, Iozzo RV. The human decorin gene: intron-exon organization, discovery of two alternatively spliced exons in the 5' untranslated region, and mapping of the gene to chromosome 12q23. Genomics 1993;15:146–160.
33. Day AA, McQuillan CI, Termine JD, Young MR. Molecular cloning and sequence analysis of the cDNA for small proteoglycan II of bovine bone. Biochem J 1987;248:801–805.
34. Abramson SR, Woessner JF Jr. cDNA sequence for rat dermatan sulfate proteoglycan-II (decorin). Biochim Biophys Acta 1992;1132: 225–227.
35. Zhan Q, Burrows R, Cintron,C. Cloning and in situ hybridization of rabbit decorin in corneal tissues. Invest Ophthalmol Visual Sci 1995;36:206–215.
36. Kokenyesi R, Woessner JF Jr. Purification and characterization of a small dermatan sulphate proteoglycan implicated in the dilation of the rat uterine cervix. Biochem J 1989;260:413–419.
37. Scholzen T, Solursh M, Suzuki S, et al. The murine decorin: complete cDNA cloning, genomic organization, chromosomal assignment, and expression during organogenesis and tissue differentiation. J Biol Chem 1994;269:28270–28281.
38. Grover J, Chen XN, Korenberg JR, Roughley PJ. The human lumican gene: organization, chromosomal location, and expression in articular cartilage. J Biol Chem 1995;270:21942–21949.
39. Blochberger TC, Vergen JP, Hempel J, Hassell JR. cDNA to chick lumican (corneal keratan sulfate proteoglycan) reveals homology to the small interstitial proteoglycan gene family and expression in muscle and intestine. J Biol Chem 1992;267:347–352.
40. Funderburgh JL, Funderburgh ML, Brown SJ, et al. Sequence and structural implications of a bovine corneal keratan sulfate proteoglycan core protein: protein 37B represents bovine lumican and proteins 37A and 25 are unique. J Biol Chem 1993;268:11874–11880.
41. Funderburgh JL, Funderburgh ML, Hevelone ND, et al. Sequence, molecular properties, and chromosomal mapping of mouse lumican. Invest Ophthalmol Visual Sci 1995;36:2296–2303.
42. Oldberg A, Antonsson P, Lindblom K, Heinegard D. A collagen-binding 59-kd protein (fibromodulin) is structurally related to the small interstitial proteoglycans PG-S1 and PG-S2 (decorin). EMBO J 1989;8:2601–2604.
43. Nurminskaya MV, Birk DE. Differential expression of fibromodulin mRNA associated with tendon fibril growth: isolation and characterization of a chicken fibromodulin cDNA. Biochem J 1996;317:785–789.
44. Antonsson P, Heinegard D, Oldberg A. Structure and deduced amino acid sequence of the human fibromodulin gene. Biochim Biophys Acta 1993;1174:204–206.
45. Grover J, Chen XN, Korenberg JR, Recklies AD, Roughley PJ. The gene organization, chromosome location, and expression of a 55-kDa matrix protein (PRELP) of human articular cartilage. Genomics 1996;38:109–117.
46. Bengtsson E, Neame PJ, Heinegard D, Sommarin Y. The primary structure of a basic leucine-rich repeat protein, PRELP, found in connective tissues. J Biol Chem 1995;270:25639–25644.
47. Dunlevy JR, Chakravarti S, Gyalzen P, Vergnes JP, Hassell JR. Cloning and chromosomal localization of mouse keratocan, a corneal keratan sulfate proteoglycan. Mamm Genome 1998;9:316–319.
48. Liu CY, Shiraishi A, Kao CWC, et al. The cloning of mouse keratocan cDNA and genomic DNA and the characterization of its expression during eye development. J Biol Chem 1998;273: 22584–22588.
49. Corpuz LM, Funderburgh JL, Funderburgh ML, Bottomley GS, Prakash S, Conrad GW. Molecular cloning and tissue distribution of keratocan: bovine corneal keratan sulfate proteoglycan 37A. J Biol Chem 1996;27:9759–9763.
50. Kurita K, Shinomura T, Ujita M, et al. Occurrence of PG-Lb, a leucine-rich small chondroitin/dermatan sulphate proteoglycan in mammalian epiphyseal cartilage: molecular cloning and sequence analysis of the mouse cDNA. Biochem J 1996;18:909–914.
51. Shinomura T, Kimata K. Proteoglycan-Lb, a small dermatan sulfate proteoglycan expressed in embryonic chick epiphyseal cartilage, is structurally related to steoinductive factor. J Biol Chem 1992;267:1265–1270.
52. Madisen L, Neubauer M, Plowman G, et al. Molecular cloning of a novel bone-forming compound: osteoinductive factor. DNA Cell Biol 1990;9:303–309.
53. Ujita M, Shinomura T, Kimata K. Molecular cloning of the mouse osteoglycin-encoding gene. Gene 1995;158:237–240.
54. Shen Z, Gantcheva S, Mansson B, Heinegard D, Sommarin Y.

Chondroadherin expression changes in skeletal development. Biochem J 1998;330:549–557.
55. Neame PJ, Sommarin Y, Boynton RE, Heinegard D. The structure of a 38kDa leucine-rich protein isolated from bovine cartilage. J Biol Chem 1994;269:21547–21554.
56. Grover J, Chen XN, Korenberg JR, Roughley PJ. The structure and chromosome location of the human chondroadherin gene (CHAD). Genomics 1997;45:379–385.
57. Landgren C, Beier DR, Fassler R, Heinegard D, Sommarin Y. The mouse chondroadherin gene: characterization and chromosomal localization. Genomics 1998;47:84–91.
58. Revelard P, Lips S, Pays E. A gene from the VSG expression site of *Trypanosoma brucei* encodes a protein with both leucine-rich repeats and a putative zinc finger. Nucleic Acids Res 1990;18:7299–7303.
59. Ross DT, Raibaud A, Florent IC, et al. The *trypanosome* VSG expression site encodes adenylate cyclase and a leucine-rich putative regulatory gene. EMBO J 1991;10:2047–2053.
60. Smiley BL, Stadnyk AW, Myler PJ, Stuart K. The *trypanosoma* leucine repeat gene in the variant surface glycoprotein expression site encodes a putative metal-binding domain and a region resembling protein-binding domains of yeast, *Drosophila*, and mammalian proteins. Mol Cell Biol 1990;10:6436–6444.
61. Dixon MS, Jones DA, Keddie JS, Thomas CM, Harrison K, Jones JD. The tomato Cf-2 disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. Cell 1996; 84:451–459.
62. Wang K, Knipfer M, Huang QQ, et al. Human skeletal muscle nebulin sequence encodes a blueprint for thin filament architecture: sequence motifs and affinity profiles of tandem repeats and terminal SH3. J Biol Chem 1996;271:4304–4314.
63. Labeit S, Barlow DP, Gautel M, et al. A regular pattern of two types of 100- residue motif in the sequence of titin. Nature 1990;345:273–276.
64. Desseyn JL, Guyonnet-Duperat V, Porchet N, Aubert JP, Laine A. Human mucin gene MUC5B, the 10.7-kb large central exon encodes various alternate subdomains resulting in a super-repeat: structural evidence for a 11p15.5 gene family. J Biol Chem 1987;272:3168–3178.
65. Byers TJ, Brandin E, Lue RA, Winograd E, Branton D. The complete sequence of *Drosophila* to β-spectrin reveals supramotifs comprising eight 106-residue segments. Proc Natl Acad Sci USA 1992;89:6187–6191.
66. Inoue K, Takeuchi Y, Takeyama S, et al. Adhesive protein cDNA sequence of the mussel *Mytilus coruscus* and its evolutionary implications. J Mol Evol 1996;43:348–356.
67. Kitamura N, Kitagawa H, Fukushima D, Takagaki Y, Miyata T, Nakanishi S. Structural organization of the human kininogen gene and a model for its evolution. J Biol Chem 1985;260:8610–8617.
68. Wolber P, Warren G. Bacterial ice-nucleation proteins. Trends Biochem Sci 1989;14:179–182.
69. Warren G, Wolber P. Molecular aspects of microbial ice nucleation. Mol Microbiol 1991;5:239–243.
70. Nietfeld W, El-Baradi T, Mentzel H, et al. Second-order repeats in *Xenopus laevis* finger proteins. J Mol Biol 1989;208:639–659.
71. Adachi J, Hasegawa M. Computer science monography, No. 28, MOLPHY: Programs for molecular phylogenetics based on maximum likelihood. Tokyo: Institute of Statistical Mathematics; 1996.
72. Kishino H, Miyata T, Hasegawa M. Maximum likelihood inference of protein phylogeny, and the origin of chloroplasts. J Mol Evol 1990;31:150–160.
73. Saitou N, Nei M. The neighboring-joining method: a new method of reconstructing phylogenetic trees. Mol Biol Evol 1987;4:406–425.
74. Venkatesan MM, Buysse JM, Hartman AB. Sequence variation in two ipaH genes of *Shigella flexneri* 5 and homology to the LRG-like family of proteins. Mol Microbiol 1991;5:2435–2445.
75. Leung KY, Straley SC. The yopM gene of *Yersinia pestis* encodes a released protein having homology with the human platelet surface protein GPIb alpha. J Bacteriol 1989;171:4623–4632.
76. Setoguchi M, Nasu N, Yoshida S, Higuchi Y, Akizuki S, Yamamoto S. Mouse and human CD14 (myeloid cell-specific leucine-rich glycoprotein) primary structure deduced from cDNA clones. Biochim Biophys Acta 1989;1008:213–222.
77. Ferreo E, Hsieh CL, Francke U, Goyert SM. CD14 is a member of the family of leucine-rich proteins and is encoded by a gene syntenic with multiple receptor genes. J Immunol 1990;145:331–336.
76. Miyake K, Yamashita Y, Ogata M, Sudo T, Kimoto M. RP105, a novel B cell surface molecule implicated in B cell activation, is a member of the leucine-rich repeat protein. J Immunol 1995;154: 3333–3340.
77. Reinke R, Krantz DE, Yen D, Zipursky SL. Chaoptin, a cell surface glycoprotein required for *Drosophila* photoreceptor cell morphogenesis, contains a repeat motif found in yeast and human. Cell 1988;52:291–301.
79. Nose A, Mahajan VB, Goodman CS. Connectin: a homophlic cell adhesion molecule expressed on a subset of muscles and the motoneurons that innervate them in *Drosophila*. Cell 1992;70:553–567.
81. Gould AP, White RA. Connectin, a target of homeotic gene control in *Drosophila*. Development 1992;116:1163–1174.
82. Hashimoto C, Hudson KL, Anderson KV. The Toll gene of *Drosophila*, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein. Cell 1988;52:269–279.
83. Chang Z, Price BD, Bockheim S, Boedigheimer MJ, Smith R, Laughon A. Molecular and genetic characterization of the *Drosophila* tartan gene. Dev Biol 1993;160:315–332.
84. Taniguchi H, Tohyama M, Takagi T. Cloning and expression of a novel gene for a protein with leucine-rich repeats in the developing mouse nervous system. Brain Res Mol Brain Res 1996;36:45–52.
85. Leong SR, Baxter RC, Camerato T, Dai J, Wood WI. Structure and functional expression of the acid-labile subunit of the insulin-like growth factor-binding protein complex. Mol Endocrinol 1992;6: 870–876.
86. Dai J, Baxter RC. Molecular cloning of the acid-labile subunit of the rat insulin-like growth factor binding protein complex. Biochem Biophys Res Commun 1992;188:304–309.
87. Eldon E, Kooyer S, D'Evelyn D, et al. The *Drosophila* 18 wheeler is required for morphogenesis and has striking similarities to Toll. Development 1994;120:885–899.
88. Chiang C, Beachy PA. Expression of a novel Toll-like gene spans the parasegment boundary and contributes to hedgehog function in the adult eye of *Drosophila*. Mech Dev 1994;47:225–239.
89. Scott JE, Cummings C. The proteins of proteodermatan and proteokeratan sulphates (decoron and fibromodulon/lumicon) are horseshoe shaped, resembling ribonuclease inhibitor. Biochem Soc Transact 1995;23:515S.
90. Scott JE. Proteodermatan and proteokeratan sulfate (decorin, lumican/fibromodulin) proteins are horseshoe shaped: implications for their interactions with collagen. Biochemistry 1996;35: 8795–8799.
91. Weber IT, Harrison RW, Iozzo RV. Model structure of decorin and implications for collagen fibrillogenesis. J Biol Chem 1996;271: 31767–31770.
92. Wegrowski Y, Pillarisetti J, Danielson KG, Suzuki S, Iozzo RV. The murine biglycan: complete cDNA cloning, genomic organization, promoter function, and expression. Genomics 1995;30:8–17.
93. Ying S, Shiraishi A, Kao CW, et al. Characterization and expression of the mouse lumican gene. J Biol Chem 1997;272:30306–30313.
94. Hassell JR, Rada J, Cornuet P, Vergnes JP, Kinchington PR. Gene structure of chick lumican and identification of the first exon. Biochim Biophys Acta 1998;1397:119–125.
95. Iwata Y, Shinomura T, Kurita K, Zako M, Kimata K. The gene structure and organization of mouse PG-Lb, a small chondroitin/dermatan sulphate proteoglycan. Biochem J 1998;331:959–964.
96. Cao Y, Adachi J, Janke A, Paabo S, Hasegawa M. Phylogenetic relationship among Eutherian orders estimated from inferred sequence mitochondrial proteins: instability of a tree based a single gene. J Mol Evol 1994;39:519–527.