Model Structure of Decorin and Implications for Collagen Fibrillogenesis*

Irene T. Weber‡§, Robert W. Harrison‡§, and Renato V. Iozzo¶‡¶

From the ‡Department of Biochemistry and Molecular Pharmacology, the §Kimmel Cancer Center, and the ¶Department of Pathology, Anatomy and Cell Biology, Jefferson Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania 19107

The three-dimensional structure of human decorin, a secreted proteoglycan involved in the regulation of collagen fibrillogenesis and cellular growth, has been modeled based on the crystal structure of the porcine ribonuclease inhibitor. Both proteins contain leucine-rich repeats and share 18% identical residues. This model structure of decorin has an arch shape with the single glycosaminoglycan chain and the three N-linked oligosaccharides located on the same side of the molecule. Decorin was modeled as binding to a polar sequence of collagen type I found in the d band. The inner concave surface is the appropriate size and shape to accommodate only one collagen triple helix of ~3 nm in length. The binding of one collagen triple helix to decorin is proposed to play a major role in the formation of the staggered arrangement of collagen molecules within the microfibrils by preventing lateral fusion of collagen molecules.

Decorin belongs to a growing family of structurally related proteoglycans, grouped as the small leucine-rich proteoglycans (SLRPs),1 whose major functional roles include regulation of collagen fibrillogenesis, modulation of growth factor activity, and regulation of cellular growth (1). Comparison of the human (2), bovine (3), avian (4), and murine (5) decorin has allowed a better understanding of its structure. The mature protein is highly conserved across species and consists of a central domain harboring ten leucine-rich repeats (LRR) flanked by disulfide-bonded terminal sequences. The amino terminus contains a single attachment site for either chondroitin or dermatan sulfate, whereas the central domain harbors three attachment sites for N-linked oligosaccharides. The sequence homologies are most pronounced in the central domain, and it is most likely that the high affinity binding site(s) (Kd in the nanomolar range) of decorin for various fibrillar collagens, including types I and II (6), III (7) and VI (8), is located to this central region. Indeed, results from three independent laboratories that have used recombinant decorin (9), decorin peptides (10), or biglycan/decorin chimeric proteins (11) have demonstrated that the LRRs 4–6 contain the high affinity binding site(s) for collagen type I. The binding of decorin to fibrillar collagen has important biological consequences in vivo, as recently demonstrated by the phenotype of mice lacking the decorin gene.2 In these mutant animals, disruption of the decorin gene leads to skin fragility and abnormal collagen morphology, characterized by uncontrolled lateral fusion of fibrils. Additional evidence that LRRs are potent mediators of specific homophilic and heterophilic interactions is provided by the observation that a single LRR in the nerve growth factor receptor TrkA can bind its ligand with nanomolar affinity and can mediate neurotrophin selectivity (12).

The crystal structure of porcine ribonuclease inhibitor (RI), a leucine-rich protein with structural homology to the SLRPs, has recently been elucidated at 2.5 Å resolution (13) and also its complex with RNase (14). The entire RI is made up of 15 tandem LRRs with alternating short β-strands and α-helices. These units form a superbhelix and are arranged so that the β-strands and α-helices are parallel to a common axis, thereby forming a horseshoe structure with the β-sheets lining the concave surface of the molecule and the α-helices flanking the convex face (13). This unusual shape, together with the vast exposure of the concave face to the solvent and its conformational flexibility, may explain why LRRs are utilized to achieve strong protein-protein interactions (1, 14).

Comparative modeling approaches capable of predicting the three-dimensional structure of a given polypeptide from sequence similarities with known structures have been previously used to investigate ligand binding to specific proteins and receptors, to solve experimental structures in solution or in the crystal state, and to design protein engineering experiments (15). In the present investigation, we modeled the three-dimensional structure of human decorin based on the crystal structure of the porcine RI using the molecular mechanics method AMMP, which uses all atoms without a distance cut-off (16). The decorin model is arch-shaped, and the concave surface can accommodate one triple helix of collagen type I. Decorin proteoglycan is proposed to bind in the 0.6-D (1 D = 67 nm) gap occurring between staggered collagen molecules and could account for the proposed binding sites and functional role of decorin in regulating collagen fibrillogenesis.

EXPERIMENTAL PROCEDURES

The crystal structure of the RI (13) was used to build the model of decorin. The alignment of the amino acid sequences of human decorin and the RI was made by first aligning the LRRs. Then insertions and deletions were positioned on the basis of the secondary and tertiary structure of the ribonuclease inhibitor. The program AMMP (16) was run on a Dell PC (133 MHz Pentium) in Windows to build and minimize the model of decorin. The Windows 95 implementation of AMMP is comparable in speed to a Unix workstation. The procedure has been optimized so that minimization of the crystal structure of a protein results in atomic positions that are within the experimental errors in protein crystal structures (17, 18). The molecular mechanics calculations used the UFF potential set with modified atomic parameters as listed previously (18). The force constants for the planarity terms of the

*This work was supported by National Institutes of Health Grants RO1 CA39481 and RO1 CA47282 (to R. V. I.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed: Dept. of Pathology, Anatomy and Cell Biology, Rm. 249, Jefferson Alumni Hall, Thomas Jefferson University, Philadelphia, PA 19107. Tel.: 215-923-7969; Fax: 215-923-7969; E-mail: Iozzo@lac.jci.tju.edu.

‡ The abbreviations used are: SLRP, small leucine-rich proteoglycan; LRR, leucine-rich repeat; RI, ribonuclease inhibitor; D, the length of 1 cross-striation period of 67 nm.

© 1996 by The American Society for Biochemistry and Molecular Biology, Inc.

Printed in U.S.A.
to build a model of the LRRs of the thyrotropin receptor (20).

The model structure of decorin consists of an arch shape with an inner concave surface formed by the curved β-sheet and the outer convex surface formed by α-helices (Fig. 2, A and B) as in the crystal structure of the RI (13). The overall dimensions are 6.5 nm (the distance between the two arms) × 4.5 nm (the distance between the base of the arch and the apex) × 3 nm (overall thickness). In support of our model, a preparation of mixed SLRPs from bovine sclera, containing decorin and fibromodulin/lumican proteoglycans, when examined by rotary shadowing electron microscopy, revealed horseshoe-shaped images (21), resembling our arch-shaped model. As expected for dry preparations of proteoglycans, the dimensions were somewhat smaller (6 × 3.8 × 2 nm) than those predicted by our model. The RI dimensions are larger (7 × 6.2 × 3.2 nm) than those of decorin, as predicted from the greater number of LRRs.

The α-helical structure in the decorin model is not fully conserved in the LRR3 helical region, because decorin has two proline residues. However, the RI also has proline in the middle of two helices. The conformation of regions with insertions or deletions is very difficult to predict accurately (17, 22, 23), but the overall architecture of proteins and common ligand binding sites can often be predicted with relatively high accuracy (17, 22). Fortunately, the conserved residues of the LRR are located in the β-strands and amino terminus of the helix, and their conformation is essentially unchanged in this model structure of decorin. These results support the view that other SLRPs would fold in a manner similar to porcine RI and human decorin. This is likely due to the relatively large number of tandem repeats of leucine-rich modules, which would force the two parallel sets of α-helices and β-strands into a curved shape. Our model of decorin forms a structure somewhat more open than the horseshoe of the RI, which would allow a greater access of ligand proteins to the inner concave surface and thus a greater capacity of forming favorable contact points with other proteins.

**Interaction of a Collagen Triple Helix with Decorin**—The overall shape of the decorin molecule immediately suggested that collagen molecules would bind to the inner concave surface. It was clear that only a single collagen triple helix would be able to fit well inside the curved surface of decorin (Fig. 2). The inner diameter of the decorin model is about 2.5 nm, which provides a good complementary surface for the 1.5 nm diameter rod formed by the collagen triple helix (24). Interestingly, there are several polar and charged side chains exposed on the concave surface of decorin (Fig. 2C), and therefore, decorin would be expected to bind preferentially to a more polar region of the collagen molecule. Although the binding of decorin to the d band of collagen has been proposed to occur near its carboxyl end, at residues 896–891 of collagen (21, 25), a recent study using isolated collagen molecules and purified decorin has shown convincingly that the binding occurs preferentially near the amino-terminal end of collagen (26). The precise mapping of the decorin-binding site was possible because procollagen molecules harboring a globular C-propeptide were also tested and revealed the molecular polarity. Specifically, the major binding site, which accounts for ~60% of decorin/collagen associations, occurs between collagen residues 188 and 198 and corresponds exactly to the d band of collagen quarter stagger (26). This region of collagen has many charged residues and the central stretch of 13 amino acid residues (171-KGEAGPEGARGPE-186), which harbors 2 basic and 3 acidic residues, was selected for modeling. Thus, the collagen triple helix will have 6 basic and 9 acidic residues in this region (Fig. 2D). Different models of collagen suggested that different combinations of these charged residues would form salt bridges that intercon-
nect the three strands in the triple helix. This region of collagen is predicted to form ionic and polar interactions with the residues on the concave surface of decorin. Decorin has 10 basic and 11 acidic residues facing the inner cavity of the C-shaped molecule (Fig. 2 C). The concave inner surface is lined with charged residues. D is a space-filling model of a collagen triple helix showing the distribution of charged residues based on a polar sequence in the d band of collagen α1(I), 174KGEAPGAPGE186 (26). Basic residues are in blue, acidic residues are in red, and the other residues are in yellow. E is a model of the decorin complexed with a triple helix of collagen in a space-filling representation with decorin in green and collagen in yellow. The glycosaminoglycan attachment site on Ser7 is in orange, and the N-linked oligosaccharide attachment sites are in purple. F is a model of the decorin/collagen complex in a view perpendicular to that of E. About 10 residues of collagen fit inside the cavity of the decorin molecule.

An interesting feature of decorin model is that all the carbohydrate moieties are positioned on one side of the arch-shaped molecule (Fig. 2 E). Accordingly, complementarity of charge would enhance binding between collagen and decorin. The model shows collagen interacting most closely with LRRs 3 and 4 (Fig. 2, E and F), in agreement with experimental data based on direct interaction between decorin peptides or decorin/biglycan chimeric proteins and collagen type I (10, 11). This predicted decorin-binding sequence located in d band of collagen type I is also present with somewhat lower fidelity in collagen types II, III, and V and is highly conserved in murine, avian, bovine, and human collagens, an important point insofar as decorin association with the d band of fibrillar collagens occurs in all species so far examined (27).

An interesting feature of decorin model is that all the carbohydrate moieties are positioned on one side of the arch-shaped molecule (Fig. 2 E). The single dermatan/chondroitin sulfate chain located on Ser7 is at one edge of the decorin molecule, so that the glycosylation will increase the thickness of the arch-shaped decorin molecule.

Decorin Could Act as a Spacer During Lateral Assembly of Collagen Molecules—The collagen molecule consists of a triple helical arrangement that forms a 300-nm-long flexible rod, 1.5 nm in diameter (24, 29). There is ample evidence that three α-chains with a repeating Gly-X-Y sequence staggered by one residue relative to one another are required to make a collagen molecule (29). An individual α-chain is not stable as a helix. These collagen molecules aggregate in parallel into a fibril, a D-staggered array (1 D = 567 nm, or 234-residue spacings) that overlaps about three-quarters of the length of the molecule (Fig. 3). This aggregation gives rise to the characteristic D periodicity of fibrils in which each D period is subdivisible into an overlap zone and a more loosely packed gap zone (26). The triple helix molecules are in contact with each other for about three-quarters of their length with a staggered spacing (gap zone) of 0.6 D, 40 nm, between molecules. Type I collagen triple helices are further stabilized by cross-linking of their lysine residues to produce the four-dimensional stagger (29). The decorin molecule is proposed to enclose one triple helix of collagen in the gap between molecules at the d band, located at −0.8 D from the amino terminus. Decorin binding to a fibril of five collagen triple helices is shown in Fig. 3. A secondary binding site, located at −1.6 D from the amino ter-

FIG. 2. Model structure of human decorin with a single triple helix of collagen type I. A is a decorin model in a ribbon representation showing the LRRs forming a central cavity with the β-sheet on the inner concave surface and the α-helices on the outer convex surface. The decorin model has the external dimensions of about 6.5 × 4.5 × 3.0 nm. B is the same in a ribbon representation in a view perpendicular to that of A. C is a space-filling representation of decorin showing the distribution of charged residues with basic or positively charged residues in blue and acidic or negatively charged residues in red; other residues are in green. The concave inner surface is lined with charged residues. D is a space-filling model of a collagen triple helix showing the distribution of charged residues based on a polar sequence in the d band of collagen α1(I), 174KGEAPGAPGE186 (26). Basic residues are in blue, acidic residues are in red, and the other residues are in yellow. E is a model of the decorin complexed with a triple helix of collagen in a space-filling representation with decorin in green and collagen in yellow. The glycosaminoglycan attachment site on Ser7 is in orange, and the N-linked oligosaccharide attachment sites are in purple. F is a model of the decorin/collagen complex in a view perpendicular to that of E. About 10 residues of collagen fit inside the cavity of the decorin molecule.
Model of Decorin

concave surface and contact the secondary site on an adjacent collagen molecule with one of the terminal ends, perhaps via its carboxyl end as proposed before (9, 10). In any event, the presence of decorin at these strategic sites would assist in the correct positioning of the collagen molecules within the staggered conformation of the fibril. Interactions with other SLRPs and enzymatic cross-linking of triple helices would represent additional aids for proper collagen fibrillogenesis in vivo.

In conclusion, the model of decorin/collagen complex predicts a tight fit between these two matrix proteins at the molecular level and provides a molecular explanation and a refined topology for the observed decorin-binding sites on collagen. The binding of decorin to one collagen triple helix within the gap zone and a second collagen molecule in the staggered arrangement will promote formation of the correct fibril and prevent incorrect addition/fusion of collagen in the gap.

Acknowledgments—We thank C. Reed and I. Eichstetter for help with the illustrations.

REFERENCES

1. Izzo, R. V., and Murdoch, A. D. (1996) J. Biol. Chem. 271, 598–614
2. Krusius, T., and Ruoslahti, E. (1986) Proc. Natl. Acad. Sci. U.S.A. 83, 7683–7687
3. Day, A. A., McQuillan, C. I., Termine, J. D., and Young, M. R. (1987) Biochem. Biophys. Res. Commun. 156, 190–197
4. Scholzen, T., Solursh, M., Suzuki, S., Reiter, R., Morgan, J. L., Buchberg, A. M., Siracusa, L. D., and Iozzo, R. V. (1994) J. Biol. Chem. 68, 28270–28281
5. Kobe, B., and Deisenhofer, J. (1995) Biochem. Biophys. Res. Commun. 213, 573–582
6. Watabe, K., and Yamada, K. M. (1995) J. Biol. Chem. 270, 28270–28281
7. Thiesen, S. L., and Rosenquist, T. H. (1994) Matrix Biol. 14, 573–582
8. Bidanset, D. J., Guidry, C., Rosenberg, L. C., Choi, H. U., Timpl, R., and Hoo, M. (1992) J. Biol. Chem. 267, 5250–5256
9. Spiro, R. C., Countaway, J. L., Gaaide, W. A., Garcia, J. A., Leisten, J., O’Neill, J. J., Smiley, W. R., Stecker, K., and Harper, J. R. (1995) Mol. Biol. Cell 55, 303 (abstr.)
10. Schönberg, E., Haussler, H., Beaven, L., and Kresse, H. (1995) J. Biol. Chem. 270, 8877–8883
11. Svensson, L., Heinegård, D., and oldberg, A. (1995) J. Biol. Chem. 270, 20712–20716
12. Windisch, J. M., Marksteiner, R., and Schneider, R. (1995) J. Biol. Chem. 270, 28133–28138
13. Kobe, B., and Deisenhofer, J. (1995) Nature 366, 751–756
14. Kobe, B., and Deisenhofer, J. (1995) Nature 374, 183–186
15. Ring, C. S., and Cohen, F. E. (1993) J. Biol. Chem. 268, 783–790
16. Harrison, R. W. (1993) J. Exp. Med. 114, 1112–1122
17. Harrison, R. W., Chatterjee, D., and Weber, I. T. (1995) Proteins 23, 463–471
18. Weber, I. T., and Harrison, R. W. (1996) Protein Eng. 9, 679–690
19. Nemethy, G., Gibson, K. D., Palmer, K. A., Yoon, C. N., Patil, G., Zargari, A., Rumsey, S., and Scheraga, H. A. (1992) J. Phys. Chem. 96, 6472–6484
20. Kajava, A. V., Vassart, G., and Wodak, S. J. (1996) Structure 4, 967–977
21. Scott, J. E. (1996) Biochemistry 35, 8795–8799
22. Weber, I. T. (1990) Proteins 7, 172–184
23. Greer, J. (1990) Proteins 7, 317–334
24. Prockop, D. J., and Kivirikko, K. I. (1995) Annu. Rev. Biochem. 64, 403–434
25. Scott, J. E., and Glanville, R. W. (1993) Biochem. Soc. Trans. 21, 123
26. Lan, Y., Cummings, C., Sheehan, J. K., Kadler, K. E., Holmes, D. F., and Chapman, J. A. (1993) in Dermatan Sulphate Proteoglycans (Scott, J. E., ed) pp. 183–192, Portland Press, London
27. Scott, J. E. (1988) Biochem. J. 252, 313–323
28. Scott, J. E., and Orford, C. R. (1981) Biochem. J. 197, 213–216
29. Piez, K. A. (1984) in Extracellular Matrix Biochemistry (Piez, K. A., and Reddi, A. H., eds) pp. 1–39, Elsevier Science Publishing Co., Inc., New York