Biochemical analysis of collagen binding by invertebrate SPARC and engineered monomeric decorin

Mehwaesh Islam
Department of Life Sciences
Imperial College London

March 2014

Thesis Submitted For the Degree of Doctor of Philosophy
Abstract

Collagens are a family of extracellular matrix proteins that are critically important for providing structural support to tissues and for regulating cell behaviour. The thesis describes the biochemical analysis of collagen binding by two globular proteins that profoundly influence collagen fibril formation in vivo, decorin and SPARC.

Decorin is the archetypal small leucine-rich repeat proteoglycan and an important regulator of collagen fibrillogenesis. The crystal structure of decorin, published in 2004, revealed a dimeric structure in which the presumed collagen binding site was not accessible. Whether the dimer is functional has been controversial, however, and it has been claimed that biologically active decorin is monomeric. In order to resolve this controversy, I designed a number of decorin mutants to disrupt the crystallographic dimer, including two mutants which introduced glycosylation sites into the dimer interface. Size exclusion chromatography with multi-angle laser light scattering and analytical ultracentrifugation were used to determine the oligomeric states of decorin and the designed mutants. I found that wild-type decorin dimerises in solution with a dissociation constant of ~1 μM. The mutants with engineered glycosylation sites were pure monomers while other mutants remained dimeric. Thermal unfolding experiments showed that the engineered decorin monomers were as stable as wild-type decorin. Mutations on the concave face of decorin abolished collagen binding, regardless of whether the mutant proteins retained the ability to dimerise or not. Thus the concave face of decorin is involved in collagen binding and the dimer must dissociate in order to bind collagen.

The crystal structure of human SPARC bound to a collagen-like triple-helical peptide was determined in the Hohenester lab in 2008. The key collagen binding residues are conserved between human and invertebrate (Drosophila and C. elegans) SPARCs. There is a key difference between the orthologues, however: high-affinity collagen binding to human SPARC requires proteolytic cleavage of an inhibitory loop that is absent from invertebrate SPARCs. To investigate the functional consequences of this structural difference for collagen binding, I compared the interactions of the different SPARCs with
collagen I and IV, using solid-phase assays and surface plasmon resonance. I found that invertebrate SPARCs did not bind collagen more tightly than human SPARC, suggesting that the absence of the inhibitory loop does not confer a higher affinity for collagen in invertebrate SPARCs. I made several unsuccessful attempts to crystallise an invertebrate SPARC, including experiments in which the glycan was trimmed by endoglycosidase digestion or removed by mutagenesis of acceptor sites.
Declaration

I declare that the work presented in this thesis has been carried out by myself unless otherwise stated.

Signed

Mehwaesh Islam

Copyright declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.
Acknowledgements

First and foremost, I owe a debt of gratitude to my supervisor, Professor Erhard Hohenester, for the opportunity to pursue a PhD. It has been a pleasure working with you. I have benefitted greatly from your guidance, mentorship, and the many lessons you have taught me along the way. Your ability to balance work interests and personal pursuits has been an inspiration. I hope we stay in touch.

Special thanks to my advisors, Professor Stephen Curry and Dr Bernadette Byrne, for the constructive discussions that initially helped shape my projects.

Thanks to our collaborators at UCL, Professor Steve Perkins and Jayesh Gor, for help with analytical centrifugation, Dr Jon Taylor for help with SEC-MALS, Professor Kurt Drickamer for help with calorimetry, and Dr Huifang Xu for help with solid-phase assays. Many thanks to the members of the Hohenester group past and present: Federico and Saffi helped me learn the ropes of the lab, Alan was always full of of good advice, Jan’s dedication as a researcher was inspiring. Thanks to Eoin for the laughs, and to Amar, Trevor, Marta, Olga and Beto for being amazing office mates. Thanks to my students, Amanda and Raj, for contributing to the work presented in this thesis.

Beyond academia, I must thank my wonderful friends for all the fun times and celebration of moments big and small: Ali, Rahul, Yusuf, Tasneem and gang, Clare, Jenny, Zara, and my friends on social media. Thanks to my dear friend, Rohini, for the pep talks and laughter.

Big thanks to Saad and Tasweer for always believing in me. Thanks to Nanu for her love and prayers. I am grateful to my in-laws for welcoming me into the family with arms wide open. Paras, you dealt with my occasional absences and shared all the highs and lows these last few years. For all this and more, you have my love and gratitude.

Last but not the least, I would like to thank my parents for their unflinching support and encouragement through all the endeavours I have undertaken to date. Maa and papa, thank you for instilling in me a love of learning. Of all that I am, of all that I could ever be, I will never be prouder than to say that I am of you.
For my parents, Qamrul and Gulerana
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADAMTS</td>
<td>A disintegrin and metalloproteinase with thrombospondin motifs</td>
</tr>
<tr>
<td>Amp</td>
<td>Ampicillin</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>AUC</td>
<td>Analytical ultracentrifugation</td>
</tr>
<tr>
<td>BM</td>
<td>Basement membrane</td>
</tr>
<tr>
<td>BMP</td>
<td>Bone morphogenetic protein</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary deoxyribonucleic acid</td>
</tr>
<tr>
<td>C. elegans</td>
<td>Caenorhabditis elegans</td>
</tr>
<tr>
<td>CIP</td>
<td>Calf intestinal alkaline phosphatase</td>
</tr>
<tr>
<td>CMV</td>
<td>Cytomegalovirus</td>
</tr>
<tr>
<td>Cp</td>
<td>Heat capacity</td>
</tr>
<tr>
<td>CS</td>
<td>Chondroitin sulphate</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco’s modified Eagle’s medium</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxyribonucleotide triphosphate</td>
</tr>
<tr>
<td>DS</td>
<td>Dermatan sulphate</td>
</tr>
<tr>
<td>DSC</td>
<td>Differential scanning calorimetry</td>
</tr>
<tr>
<td>DSF</td>
<td>Differential scanning fluorimetry</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>EC</td>
<td>Extracellular Ca$^{2+}$-binding</td>
</tr>
<tr>
<td>ECM</td>
<td>Extracellular matrix</td>
</tr>
<tr>
<td>E. coli</td>
<td>Escherichia coli</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EDC</td>
<td>N-ethyl-N’-(diethylaminoethyl)carbodiimide</td>
</tr>
<tr>
<td>EGF</td>
<td>Epidermal growth factor</td>
</tr>
<tr>
<td>EGFR</td>
<td>Epidermal growth factor receptor</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic reticulum</td>
</tr>
<tr>
<td>FBS</td>
<td>Foetal bovine serum</td>
</tr>
<tr>
<td>FGF</td>
<td>Fibroblast growth factor</td>
</tr>
<tr>
<td>FS</td>
<td>Follistatin-like</td>
</tr>
<tr>
<td>GAG</td>
<td>Glycosaminoglycan</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-Acetylglucosamine</td>
</tr>
<tr>
<td>HEK</td>
<td>Human embryonic kidney</td>
</tr>
<tr>
<td>HEPES</td>
<td>4-(2-hydroxyethyl)-1-piperazine-1-ethanesulfonic acid</td>
</tr>
<tr>
<td>HRP</td>
<td>Horseradish peroxidase</td>
</tr>
<tr>
<td>IGF</td>
<td>Insulin-like growth factor</td>
</tr>
<tr>
<td>IGFR</td>
<td>Insulin-like growth factor receptor</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>kb</td>
<td>Kilobase</td>
</tr>
</tbody>
</table>
kDa  Kilodalton
KS   Keratan sulphate
LB   Luria-Bertani medium
LRR  Leucine-rich repeat
MALDI-MS Matrix-assisted laser desorption/ionisation–mass spectrometry
MMP  Matrix metalloproteinase
MW   Molecular weight
MWCO Molecular weight cut-off
NC   Non-collagenous
NEB  New England Biolabs
NHS  N-hydroxysuccinimide
OD$_{600}$ Optical density at 600nm
OPD  o-Phenylenediamine dihydrochloride
PBS  Phosphate-buffered saline
PBS-T Phosphate-buffered saline + Tween 20
PCR  Polymerase chain reaction
PDL  Periodontal ligament
PEG  Polyethylene glycol
PMSF Phenylmethanesulphonylfluoride
PNGase F-MBP Peptide N-glycosidase F- Maltose binding protein
Puro Puromycin
rmsd  Root-mean-square deviation
rpm  Revolutions per minute
RTK  Receptor tyrosine kinase
RU  Response unit
$s_{20,w}$  Sedimentation coefficient corrected for water
SDS-PAGE  Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC  Size exclusion chromatography
SEC-MALS  Size exclusion chromatography with multi-angle light scattering
SLRP  Small leucine-rich repeat proteoglycan
SOE  Splicing by overlap extension
SPARC  Secreted protein acidic and rich in cysteine
SPR  Surface plasmon resonance
SV  Sedimentation velocity
SV40  Simian vacuolating virus 40
TAE  Tris-acetate EDTA
TBS  Tris-buffered saline
TEMED  Tetraethylmethylenediamine
TGF-β  Transforming growth factor - β
$T_m$  Melting temperature
Tris  Tris(hydroxymethyl)aminomethane
UV  Ultra-violet
V Volume

VEGF Vascular endothelial growth factor

w Weight

WT Wild-type
# Table of Contents

## Chapter 1 Introduction

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>The extracellular matrix (ECM)</td>
<td>21</td>
</tr>
<tr>
<td>1.1.1</td>
<td>The basement membrane</td>
<td>22</td>
</tr>
<tr>
<td>1.2</td>
<td>Collagen</td>
<td>25</td>
</tr>
<tr>
<td>1.2.1</td>
<td>Fibril-forming collagens: collagen I</td>
<td>27</td>
</tr>
<tr>
<td>1.2.2</td>
<td>Network-forming collagens: collagen IV</td>
<td>30</td>
</tr>
<tr>
<td>1.3</td>
<td>Small leucine rich repeat proteoglycans (SLRPs)</td>
<td>33</td>
</tr>
<tr>
<td>1.3.1</td>
<td>SLRPs: characteristics and classification</td>
<td>33</td>
</tr>
<tr>
<td>1.3.2</td>
<td>Biological functions of SLRPs</td>
<td>38</td>
</tr>
<tr>
<td>1.4</td>
<td>Decorin</td>
<td>39</td>
</tr>
<tr>
<td>1.4.1</td>
<td>Decorin: structure</td>
<td>39</td>
</tr>
<tr>
<td>1.4.2</td>
<td>Decorin: function in extracellular assembly</td>
<td>42</td>
</tr>
<tr>
<td>1.4.3</td>
<td>Decorin: role in cell signalling and cancer</td>
<td>43</td>
</tr>
<tr>
<td>1.4.4</td>
<td>Decorin: monomer or dimer?</td>
<td>46</td>
</tr>
<tr>
<td>1.5</td>
<td>SPARC</td>
<td>47</td>
</tr>
<tr>
<td>1.5.1</td>
<td>SPARC: general structure</td>
<td>48</td>
</tr>
<tr>
<td>1.5.2</td>
<td>SPARC – growth factor interactions</td>
<td>50</td>
</tr>
<tr>
<td>1.5.3</td>
<td>SPARC – collagen interactions</td>
<td>51</td>
</tr>
<tr>
<td>1.5.4</td>
<td>SPARC in ECM assembly</td>
<td>53</td>
</tr>
<tr>
<td>1.5.5</td>
<td>Collage binding by vertebrate and invertebrate SPARC</td>
<td>55</td>
</tr>
<tr>
<td>1.6</td>
<td>Aims of the project</td>
<td>55</td>
</tr>
</tbody>
</table>

## Chapter 2 Materials and Methods

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Materials</td>
<td>58</td>
</tr>
<tr>
<td>2.1.1</td>
<td>Chemicals and reagents</td>
<td>58</td>
</tr>
</tbody>
</table>
2.1.2 Bacterial cells and plasmids
2.1.3 Mammalian cells and cell culture reagents
2.1.4 Equipment
2.2 Methods
2.2.1 General molecular biology methods
2.2.1.1 Polymerase chain reaction (PCR)
2.2.1.2 Splicing by overlap extension (SOE) PCR
2.2.1.3 Agarose gel electrophoresis
2.2.1.4 DNA purification by gel extraction
2.2.1.5 Restriction digests
2.2.1.6 DNA ligation
2.2.1.7 DNA transformation
2.2.1.8 Screening for transformed DNA
2.2.2 Expression vectors
2.2.3 Mammalian cell culture
2.2.3.1 Transfection
2.2.3.2 Growth and maintenance of cells
2.2.4 Protein purification
2.2.4.1 Immobilised metal ion chromatography
2.2.4.2 Size exclusion chromatography (SEC)
2.2.4.3 Bacterial expression and purification of PNGase F-MBP
2.2.5 Proteins methods
2.2.5.1 SDS-PAGE
2.2.5.2 NuSEP®
2.2.5.3 Deglycosylation
2.2.5.4 Crystallisation trials
2.2.6 Size exclusion chromatography – multi angle light scattering
2.2.7 Surface plasmon resonance (SPR)/Biacore
2.2.8 Solid phase assays
2.2.9 Biotinylation
2.2.10 Biotin-avidin assay 76
2.2.11 Competition assay 77
2.2.12 Collagen fibrillogenesis assay 77
2.2.13 Differential scanning calorimetry (DSC) 78
2.2.14 Differential scanning fluorimetry (DSF) 78
2.2.15 Analytical ultracentrifugation (AUC) 79

Chapter 3 Results: Analysis of decorin dimerisation

3.1 Mouse decorin construct design 83
3.2 Purification of wild-type decorin 85
3.3 Purification of engineered glycosylation mutants 86
3.4 Purification of other mutants 88
3.5 Oligomeric states of wild-type and mutant decorin 91
3.6 Study of monomer-dimer equilibrium 94
3.7 Thermal stability of WT and mutant decorin 101
3.8 Collagen binding by WT and mutant decorin 103
3.8.1 Collagen-decorin binding assays 103
3.8.2 Collagen-decorin competition assay 109
3.8.3 Collagen fibrillogenesis assay 114

Chapter 4 Discussion: Analysis of decorin dimerisation 117

Chapter 5 Results: Structural and functional studies of invertebrate SPARC binding to collagen

5.1 Purification of SPARC proteins 128
5.1.1 Purification of SPARC His.myc tagged constructs 128
5.2 Expression and purification of PNGase F-MBP 130
5.3 Purification of SPARC His tagged constructs and crystallisation trials

5.3.1 Purification of dSPARC EC constructs for crystallisation trials

5.3.2 Purification of dSPARC FS-EC constructs for crystallisation trials

5.3.2.1 Analytical deglycosylation of dSPARC FS-EC NHis

5.3.2.2 Lectin treatment

5.3.2.3 dSPARC glycosylation site mutant

5.3.3 Purification of cSPARC proteins for crystallisation trials

5.4 Collagen binding analyses of SPARC proteins by ELISA and SPR

5.4.1 Solid phase binding assays (ELISA)

5.4.2 SPR studies

Chapter 6 Discussion: Interaction of SPARC with collagen

6.1 Biological implications of SPARC – collagen interactions

6.2 Structural studies

6.3 Functional studies

Chapter 7 Conclusions and future directions

References

Appendix

Appendix I

Appendix II
## List of Figures

### Figures Description

#### Chapter 1

1.1 The basement membrane  
1.2 The collagen triple helix  
1.3 The collagen fibril  
1.4 Type IV collagen network formation  
1.5 Phylogenetic tree of SLRPs  
1.6 Schematic of Class I SLRPs, decorin and biglycan  
1.7 Schematic of bovine decorin core protein and its internal organisation  
1.8 Transmission electron micrographs of dermal collagen of the skin from the decorin (Dcn)-deficient mice  
1.9 Role of decorin in tumour progression  
1.10 Schematic of the modular domains of SPARC protein  
1.11 Crystal structure of SPARC FS-EC Δα-C (inhibitory loop removed) bound to a 33-residue collagen peptide

#### Chapter 2

2.1 Splicing by overlap extension PCR for site-directed mutagenesis  
2.2 Vector map indicating features of pCEP-Pu

#### Chapter 3

3.1 Amino acid sequence alignment of bovine, mouse and human decorin  
3.2 Purification of WT decorin  
3.3 SEC analysis of decorin glycosylation mutants  
3.4 SDS-PAGE of wild-type decorin and glycosylation site mutants under reducing conditions  
3.5 SDS-PAGE of mouse decorin mutants under reducing conditions  
3.6 Positions of the engineered mutations in mouse decorin mapped onto the crystal structure of bovine decorin  
3.7 SEC-MALS chromatograms of wild-type and mutant decorin  
3.8 SDS-PAGE of wild-type decorin before and after the AUC run
3.9 Sedimentation velocity analytical ultracentrifugation of WT decorin
3.10 Continuous size distributions at seven different concentrations
3.11 Determination of the dimer dissociation constant for WT decorin.
3.12 SDS-PAGE of mouse decorin mutants (R151E and Y51A/R52A/Q54A) before and after the AUC run
3.13 Continuous size distributions at three different concentrations of (A) Y51A/R52A/Q54A and (B) R151E mutants
3.14 Determination of the dimer dissociation constants for decorin mutants
3.15 Thermal denaturation studies of WT and monomeric decorin by DSC
3.16 Thermal denaturation studies of WT and monomeric decorin by DSF
3.17 Binding of dSPARC, WT and monomeric decorin to collagen I
3.18 Surface representation of bovine decorin crystal with location of mutations highlighted
3.19 Reducing SDS-PAGE analysis of WT decorin and monomeric decorin mutants
3.20 Collagen binding by WT decorin and monomeric mutants
3.21 Collagen binding by Q61N and other monomeric mutants
3.22 Decorin - collagen I competitive binding assay
3.23 Anti-mouse decorin antibody binding to WT and mutant mouse decorin
3.24 Binding of WT and mutant mouse decorin to immobilised collagen
3.25 Collagen binding by WT and mutant mouse decorin
3.26 Effect of decorin on collagen fibrillogenesis

Chapter 5

5.1 Purification of SPARC proteins
5.2 SDS-PAGE analysis of PNGaseF-MBP IMAC purification
5.3 Deglycosylation of dSPARC EC NHis and its purification by SEC
5.4 SDS-PAGE analysis showing degree of deglycosylation of N-linked glycans from dSPARC FS-EC NHis
5.5 Purification and deglycosylation of dSPARC FS-EC NHis
5.6 SDS-PAGE analysis showing degree of deglycosylation of N-linked glycans from dSPARC FSEC-NHis
5.7 Purification of cSPARC EC NHis
5.8 Purification of cSPARC FS-EC NHis
5.9 Solid phase binding assays with recombinant dSPARC, hSPARC, cSPARC and collagen I
5.10 Solid phase binding assays with recombinant dSPARC, hSPARC, cSPARC and collagen IV

5.11 SPR (Biacore) analysis of dSPARC FS-EC NHis-myc and collagen I interaction

5.12 SPR (Biacore) analysis of dSPARC EC NHis-myc and collagen I interaction

5.13 Kinetic SPR analysis of hSPARC FS-EC NHis-myc and collagen I interaction

5.14 SPR (Biacore) analysis of hSPARC FS-EC NHis-myc and collagen I interaction

5.15 SPR (Biacore) analysis of hSPARC EC NHis-myc and collagen I interaction

5.16 SPR (Biacore) analysis of cSPARC FS-EC NHis-myc and collagen I interaction

Chapter 6

6.1 Function of SPARC in procollagen deposition and assembly
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Classification and properties of SLRPs including size, type of GAG, binding partners and knockout-mice phenotypes (where available)</td>
</tr>
<tr>
<td>2.1</td>
<td>Primers used for generating decorin expression plasmids</td>
</tr>
<tr>
<td>2.2</td>
<td>Primers used for generating SPARC expression plasmids</td>
</tr>
<tr>
<td>3.1</td>
<td>Molecular masses of decorin and its mutants as determined by SEC-MALS</td>
</tr>
<tr>
<td>4.1</td>
<td>Summary of results obtained with mouse decorin and its mutants</td>
</tr>
<tr>
<td>5.1</td>
<td>Molecular masses of SPARC proteins, number of N-linked glycosylation sites and yields from 1.5 L of serum-free medium</td>
</tr>
<tr>
<td>5.2</td>
<td>Summary of crystallisation trials</td>
</tr>
<tr>
<td>5.3</td>
<td>Summary of crystallisation trials (continued)</td>
</tr>
<tr>
<td>5.4</td>
<td>Apparent dissociation constants (mean ± standard error) of SPARC FS-EC NHis-myc proteins binding to immobilised type I collagen</td>
</tr>
<tr>
<td>5.5</td>
<td>Apparent dissociation constants (mean ± standard error) of SPARC FS-EC NHis-myc proteins binding to immobilised type IV collagen</td>
</tr>
</tbody>
</table>
Chapter 1:
Introduction
Chapter 1. Introduction

Collagen-rich extracellular matrices are critically important not only for providing structural support to cells but also in cell adhesion, signalling, regulation of tissue remodelling during growth, differentiation and wound healing (Myllyharju and Kivirikko, 2004). While the basic structure of the collagen triple helix has been known for half a century, protein-collagen interactions remain poorly understood at the atomic level. In the projects outlined in this thesis, I have studied collagen binding by two secreted proteins, mouse decorin and invertebrate SPARC. This introduction will first provide a brief outline of the environment of these two proteins i.e. the extracellular matrix. The basic structure and composition of two subclasses of the collagen superfamily will be discussed. This will be followed by an overview of our current understanding of the structure and biological roles of the two proteins under study.

1.1 The extracellular matrix (ECM)

The ECM is a complex network of macromolecules (structural proteins, proteoglycans, growth factors and matricellular proteins) that provide structural support to tissues while regulating cell behaviour and function including cell migration and proliferation, homeostasis, and developmental patterning (Frantz et al., 2010). ECM proteins are large multidomain molecules - the domains are often repeated and in tandem within the same molecule such that adjacent domains can cooperate in biological activities (Hohenester and Engel, 2002). Different tissues have unique ECM compositions. The ECM is dynamic and continuously remodelled to adapt to the functional requirements of tissues (Bosman and Stamenkovic, 2003). Genetic abnormalities in ECM proteins can give rise to a wide range of diseases (Frantz et al., 2010, Bateman et al., 2009).

A list of core ECM proteins which form the “core matrisome” have been identified - consisting of almost 300 proteins, including 43 collagen subunits, 200 glycoproteins and three dozen proteoglycans (Hynes and Naba, 2012). In addition, there are various “ECM-
affiliated” proteins (Naba et al., 2012), ECM-modifying enzymes such as proteases and cross-linking enzymes and growth factors that are present in the ECM. The main molecular components can be divided into two groups: fibrous proteins such as collagen and fibronectin and b) proteoglycans (PGs) such as those with leucine rich repeats (LRRs), perlecan and agrin (Schaefer and Schaefer, 2010).

1.1.1 The basement membrane

Basement membranes (BM) or basal laminae are a specialised form of sheet-like ECM underlying epithelial cells and surrounding endothelial, muscle, fat and peripheral nerve cells (Rowe and Weiss, 2008). BM genes are very ancient: they originated early during metazoan evolution and are highly conserved (Hynes, 2012). The discovery of the basement membrane-rich mouse Engelbreth-Holm-Swarm (EHS) tumour in the late 1970s led to the identification of a set of proteins that form a typical 50-100 nm thick lamina: cross-linked type IV collagen which makes up 50% of all BM proteins; laminin, a heterotrimer made up of related α, β and γ subunits; nidogen, a laminin-binding glycoprotein and perlecan and agrin, heparan sulphate proteoglycans (Figure 1.1) (Yurchenco, 2011). These molecules are large, ranging from ~ 75 nm to ~400 nm in length. BM also often contains the ubiquitous extracellular matrix components SPARC, fibulins, and fibronectin (Erickson and Couchman, 2000). Some specialised BMs are thicker, e.g. those of the extra-embryonic Reichert’s membrane, the kidney glomerulus and the EHS tumour matrix. A clear understanding of BM architecture by imaging in situ is difficult as it involves harmful extraction from tissues (Hohenester and Yurchenco, 2013). All BMs appear similar under the electron microscope but their detailed molecular compositions are unique in each tissue (Kalluri, 2003).

Unlike other components of the BM, laminin and collagen IV can self-assemble to form polymers (Yurchenco et al., 1992). BM components are assembled into functional units and secreted by cells. The α, β and γ chains of laminin twist around each other to form a cruciform trimer with three short arms and one long arm stabilised by disulphide bonds (Beck et al., 1990). Laminin polymers are initially deposited and these are anchored to the cell surface by interactions via cell surface receptors such as integrins, α-
dystroglycan and sulphated glycolipids/sulphatides (Figure 1.1B). This is followed by accumulation of non-laminin components such as heparan sulphate proteoglycans, agrin and perlecan, which provide collateral linkages to tether laminins to the cell surface. Collagen IV polymers are subsequently deposited on the cell surface. The laminin and collagen networks are bridged by nidogens 1 and 2 (BM glycoproteins). Studies with nidogen-deficient mice have shown that these proteins are not crucial for bridging laminin and collagen so other proteins must also be involved in this process (Bader et al., 2005). Other components of the BM such as heparan sulphates act in concert with nidogen and interact with this scaffold to form a fully functional BM (Hohenester and Yurchenco, 2013).
Figure 1.1 The basement membrane. A. Scanning electron micrograph of the basal lamina of the cornea of the chick embryo. The upper surface of the sheet-like basal lamina is exposed by removal of some of the epithelial cells. A network of collagen fibrils in the underlying connective tissue interacts with the lower face of the lamina (Alberts, 2002). B. Schematic of binding interactions between basement membrane components. Laminins anchor to the cell surface by interacting with receptors such as integrins, α-dystroglycan and sulphatides, using their long arms. Agrin and perlecan provide collateral linkages to cell surfaces. Type IV collagen forms an independent network, through interactions of its N-terminal 7S and C-terminal NC1 domains, as well as through lateral associations of the triple helices. The laminin and collagen networks are linked by nidogen and heparan sulphates (black double-headed arrows). Taken from Hohenester and Yurchenco (2013).
BMs separate the epithelium from the stroma of tissues. They function to provide structural support to adherent epithelial cells, and also influence cell behaviour such as providing cues to influence cell polarity. BMs are important in cell adhesion, growth, differentiation, tissue repair and molecular ultrafiltration (Sundaramoorthy et al., 2002). With a pore size in the order of ~ 50 nm, BMs only allow small molecules to passively diffuse through them, thus acting as selective barriers (Rowe and Weiss, 2008).

Many genetic and vascular diseases are caused by defects in BM proteins. For example, mutations in type IV collagen α5 chain causes Alport syndrome, which is characterised by a defective glomerular BM. Goodpasture syndrome - an autoimmune disease, is caused by auto-antibodies formed against type IV collagen α3 chain (Kalluri, 2003).

1.2 Collagen

Collagens are the most abundant proteins in mammals where they account for ~30% of the total protein mass. This family of extracellular proteins is characterised by the presence of a triple-helix that can make up most of their structure (96% of collagen I) or just a fraction (less than 10% in collagen XII) (Ricard-Blum, 2011). All collagen molecules consist of three polypeptide (α) chains that coil around each other to form a right-handed triple helix with a one-residue stagger between adjacent α chains (Figure 1.2). Each polypeptide chain contains repeating peptide triplets of glycine-X-Y. X and Y can be any amino acid but are often proline and 4-hydroxyproline respectively. The presence of glycine, the smallest amino acid, as every third residue allows close packing of the chains within the triple helix. Hydroxyproline is important for stability of the triple helix, which is further stabilised by hydrogen bonds and water bridges and electrostatic interactions involving lysine and aspartic acid (Brodsky and Ramshaw, 1997). All three polypeptide chains in a triple helix can be identical, which is true for most collagens (homotrimers), or may differ, consisting of two or three different α chains (heterotrimer) (Kadler et al., 2007). For example, type I collagen is made up of two α1 chains and one α2 chain wound into a triple helix whereas collagen II consists of three identical α chains. The primary structure of α1 chain of one type of collagen
differs from that of the α1 chain of another type of collagen. All collagens also contain non-collagenous (NC) domains that are numbered from the C-terminus (Kadler et al., 2007).

**Figure 1.2 The collagen triple helix.** Collagen is initially synthesised as pro-collagen with the triple-helical domain flanked by N- and C- terminal propeptides. These propeptides are eventually cleaved by enzymes (N- and C- proteinases). The mature collagen still contains non-triple helical short peptides called N- and C-telopeptides, and are subsequently assemble into fibrils. Taken from Boudko et al. (2012).

Collagens are mostly synthesised by cells in the ECM: fibroblasts, myofibroblasts, osteoblasts and chondrocytes (Bosman and Stamenkovic, 2003). Vertebrates have at least 28 different collagen types but the existence of several isoforms of the same collagen type gives rise to further diversity (Ricard-Blum, 2011). In addition, there exists a large group of collagen-like proteins such as C1q, ficolin, surfactant protein and macrophage scavenger receptor (Myllyharju and Kivirikko, 2001). The function of the collagens depends on the proper supramolecular assembly of the triple helices (Gordon and Hahn, 2010). Collagens can be subdivided into subfamilies based on their supramolecular assemblies: fibril- forming collagens, network-forming collagens, fibril-associated collagens with interrupted triple helices (FACITs), anchoring fibrils and beaded filaments (Ricard-Blum, 2011).
1.2.1 Fibril-forming collagens: collagen I

Tissues such as tendon, bone, cartilage and skin, which have to resist tensile and shear forces, contain collagen that are arranged into fibrils (Bosman and Stamenkovic, 2003). Members of this family include collagen types I, II, III, V and XI, type I being the archetypal collagen and the most abundant. With the exception of collagens XXIV and XXVII, the Gly-X-Y triple helical domain of fibril-forming collagens contains ~1000 residues and is uninterrupted (Kadler et al., 2007). Figure 1.3 shows how trimeric collagen molecules assemble into fibrils. Each triple helix is staggered from its neighbour by 67 nm in the direction of the helix and laterally, the fibrils are arranged quasi-hexagonally with respect to each other within the fibril (Perumal et al., 2008) and are connected by intermolecular covalent cross-links. The quarter-stagger gives rise to the banded pattern of fibrils in electron micrographs. The D-period, which is ~67 nm long, forms the basic morphological structure of the fibril. It is composed of a gap zone and an overlap zone and each D-period contains the complete monomer sequence from elements of five monomers (Sweeney et al., 2008).
**Figure 1.3 The collagen fibril.** A. A single type I collagen triple helix (monomer). Type I collagen is secreted as a procollagen monomer 300 nm long, extracellularly, the N- and C-propeptides are cleaved by proteinases (dashed vertical lines). B) the tropocollagen monomers assemble in a staggered fashion into collagen fibrils (C) where one D-period repeat (67 nm segment of microfibril, box) contains the complete collagen sequence from elements of the five monomers and includes an overlap and gap zone. (D). Collagen fibrils appear as periodic banded structures by electron microscopy; arrow, left border of overlap zone; heparin-albumin gold conjugates, which appear as circular dark objects, were used to map the heparin-binding site (E). Taken from Sweeney et al. (2008).

Fibril-forming collagens are synthesised as pro-collagens in the endoplasmic reticulum (ER) having N- and C-propeptides, short non-helical N- and C-telopeptides, and a central triple helix (Ricard-Blum, 2011). The intracellular steps in collagen biosynthesis involve a number of post-translational modifications including hydroxylation of certain proline and lysine residues to 4-hydroxyproline, 3-hydroxyproline and hydroxylysine, glycosylation of some of the hydroxylysine residues and formation of some intra-chain
and inter-chain disulphide bonds catalysed by protein disulphide isomerase (PDI) (Myllyharju and Kivirikko, 2004). These processes are stopped by the formation of the triple helix. This happens when the C-propeptides associate and the triple helix is propagated in a zipper-like fashion towards the N-terminus (Myllyharju and Kivirikko, 2001). The chaperone, Hsp47, binds to procollagen the ER and it has been suggested that intracellular secreted protein acidic and rich in cysteine (SPARC) also acts as a chaperone in concert with Hsp47 to ensure that correctly folded procollagen exits the ER (Martinek et al., 2007). The secretion of procollagen fibres requires formation of large COPII vesicle coats which can accommodate these macromolecules (Jin et al., 2012). Once outside the cell, the propeptides of procollagen are cleaved. The N-propeptides are cleaved by procollagen N-proteinases which are disintegrin and metalloproteinase with thrombospondin motifs (ADAMTS)-2, -3 and -14 proteinases. The C-propeptides are cleaved by procollagen C-proteinases which are bone morphogenetic protein-1 (BMP-1) proteinases (Kadler et al., 2007). Meprin α and meprin β also act as both C- and N-proteinases for type I procollagen (Broder et al., 2013). The cleavage of the N- and C-propeptides exposes the telopeptides which are non-helical regions containing binding sites for fibrillogenesis – they form non-reducible covalent cross-links with neighbouring triple helices so molecules can form fibrils in a quarter-staggered arrangement (Eyre et al., 1984). This process is catalysed by lysyl oxidase and gives collagen its tensile strength. The inhibition of this enzyme does not affect fibril formation. However, cross-link-deficient collagen is more likely to be degraded than cross-linked collagen. Thus lysyl oxidase inhibitors have potential to be used as anti-fibrotic drugs (Myllyharju and Kivirikko, 2001).

Collagen fibrils often contain more than one collagen type. For e.g. type I collagen may contain small amounts type III, V and XII. Collagen II, the principal component of cartilage, contains small amounts of collagen XI in its core and collagen type IX on its surface (Myllyharju and Kivirikko, 2004). In vitro, collagen fibrillogenesis is a self-assembly process that does not require the presence of cells. In vivo, noncollagenous molecules are required to initiate fibrillogenesis (Kadler et al., 2008). Nearly 50 molecules have been found to interact with type I collagen in vivo (Di Lullo et al., 2002) which may generate a variety of fibril patterns from parallel bundles in tendon and
ligament to orthogonal arrangements in cornea. The initial site of fibril formation has not been defined so far (Ricard-Blum, 2011). In embryonic development, cells may control fibril formation extracellularly by localising this process to cell membrane through fusion of collagen fibril intermediates (Birk and Trelstad, 1986, Zhang et al., 2005) or they may start intracellularly in Golgi-to-membrane carriers that transport procollagen to cell membrane protrusions called fibripositors (Canty et al., 2004). There are speculations that fibronectin and integrins may bind collagen and induce a conformational change to accelerate fibrillogenesis (Kadler et al., 2008). Collagen type V and type XI (which is homologous to collagen V) are used to nucleate collagen fibrils. Using mouse models of Ehlers-Danlos syndrome, a hereditary connective tissue disorder resulting from mutations in COL5A1 and COL5A2 genes, it has been shown that collagen V controls initiation of collagen fibril assembly in vivo (Wenstrup et al., 2004). Mice lacking collagen V do not contain collagen fibrils and die at embryonic day 10. Mice deficient in small leucine-rich repeat proteoglycans (SLRPs) that bind collagen including decorin, biglycan, lumican and fibromodulin, also exhibit defects in collagen fibril size and morphology (Schaefer and lozzo, 2008) (more on this in later sections).

The collagen triple helix is resistant to proteolytic cleavage by pepsin, trypsin and papain (Kadler et al., 2007). Matrix metalloproteinases MMP-1, -8 and -13 cleave fibril-forming collagens I, II and III, generating three-quarter and one-quarter fragments (Ricard-Blum, 2011) as part of the ECM degradation and remodelling process.

### 1.2.2 Network-forming collagens: Collagen IV

The second class of collagens are the network-forming collagens which include collagen types IV, VI, VIII, and X. The ability to self-assemble into organised networks distinguishes these collagens from fibrillar ones. Type IV collagen is unique among the collagen superfamily since it is only found in BMs. It is the ancestral type of collagen from which the other 27 vertebrate collagen types have evolved (Pastor-Pareja and Xu, 2011). Six genes encoding six polypeptide α chains, α1(IV) – α6(IV) have been identified
Khoshnoodi et al., 2008). These chains are closely related (50 – 70% identity at the amino acid level) and can assemble into unique heterotrimers in specific stoichiometries to make three different types of network-forming building blocks called protomers. These include the (α1)2α2, α3α4α5 and (α5)2α6 protomers (Miner, 2012). The most common variant is (α1)2α2, found nearly in all BMs, whereas α3α4α5 and (α5)2α6 are found in alveolar and glomerular BMs (Yurchenco, 2011). The existence of different α chains for collagen IV and their restricted tissue distribution determine the structural and functional specificity of BMs (Ortega and Werb, 2002). Each α chain is ~400 nm long and contains three domains: a middle triple helical domain flanked by a cysteine-rich N-terminal 7S domain and a C-terminal globular NC1 domain which is about 200 amino acids in length (Kalluri, 2003). The triple-helical region is ~1,400 amino acids in length, but unlike most other collagen types, the collagenous Gly-X-Y repeats have 21-26 interruptions in which glycine does not occur in every third position, presumably to make a more flexible trimer and to serve as cell-binding sites (Khoshnoodi et al., 2008). The interruptions do not disturb the folding of the triple helix but give rise to kinks in the rod-like helix (Boudko et al., 2012). As demonstrated by electron microscopy, trimer assembly begins at the C-terminus when NC1 domains from three α chains interact and trimerisation proceeds from C- to N-terminus in a zipper-like fashion to form a protomer (Figure 1.4). This is followed by two trimers interacting via the C-terminal NC1 domains to form an NC1 hexamer. The crystal structure of the NC1 domain reveals that it exposes a flat face to its opposite trimer for dimerisation and becomes stabilised through an unusual covalent linkage between the side chains of a Met residue on one chain and a Lys residue on the opposite chain (Sundaramoorthy et al., 2002, Than et al., 2002). (Bhave et al., 2012) have shown that the BM enzyme, peroxidasin, catalyses the formation of this bond and Drosophila peroxidasin mutants show BM defects due to fewer collagen IV crosslinks. Next, two protomers associate through the N-terminal 7S domains to form a tetramer. In this way, a collagen scaffold is formed which has been compared with a chicken-wire fence, although the molecules further interact to form supramolecular assemblies by lateral association of collagenous domains (Boudko et al., 2012).
Type IV collagens are present in mammalian BMs throughout development and adulthood. However, in invertebrate BMs, collagen IV is deposited much later in development, around the time of early muscle contractions, suggesting that collagen IV is not required for initial BM assembly but BMs need collagen IV to withstand mechanical stress (Yurchenco et al., 2004). In *Drosophila*, dorsal-ventral patterning is regulated by a host of secreted proteins such as Dpp that form a concentration gradient. There are two type IV collagen proteins in *Drosophila*: Viking (Vkg) and Cg25C. It has
been shown that collagen IV promotes gradient formation by binding Dpp in early *Drosophila* embryo (Wang et al., 2008).

### 1.3 Small leucine rich repeat proteoglycans (SLRPs)

#### 1.3.1 SLRPs: characteristics and classification

Proteoglycans are glycoproteins composed of covalently linked diverse carbohydrate chains called glycosaminoglycans (GAGs) attached to serine residues of a protein core (Schaefer and Schaefer, 2010). GAGs are linear, consisting of repeating units of disaccharides with carboxyl and sulphate groups attached. They consist of an amino sugar, N-acetylglucosamine or N-acetylgalactosamine and a second sugar which is usually a uronic acid, iduronic or glucuronic acid (Kjellen and Lindahl, 1991). GAGs can vary in size, number, sulphation and epimerisation in various tissues (Chen and Birk, 2013). Owing to their negative charge, GAGs cause proteoglycans to be extended in conformation and enable them to bind water and divalent cations such as calcium, thus conferring space-filling and hydration functions (Hynes and Naba, 2012). In addition to the GAG chains, most core proteins of proteoglycans carry *N*- and/or *O*-linked oligosaccharides (Kjellen and Lindahl, 1991).

Mammalian genomes encode roughly three dozen ECM proteoglycans. They can be broadly divided into two families: modular proteoglycans and small leucine-rich repeat proteoglycans (SLRPs). Modular proteoglycans, such as aggrecan, perlecan and agrin, are multidomain assemblies which are relatively elongated and often highly glycosylated (Iozzo and Murdoch, 1996). Unlike modular proteoglycans, SLRPs have relatively smaller protein cores (36-42 kDa) (Schaefer and Iozzo, 2008).

SLRPs belong to the leucine-rich repeat (LRR) superfamily of proteins (Hocking et al., 1998). The SLRP gene family encodes 18 genes mapped onto relatively few (7) chromosomes, suggesting that some degree of functional redundancy has been generated during evolution (Chen and Birk, 2013, Iozzo, 1998). At the protein level, SLRPs are characterised by a central LRR region flanked by cysteine-rich clusters on the N- and C- termini and at least one GAG chain. The LRRs are 20-29 amino acids long, the
most common length being 24 residues, with a conserved motif of LxxLxLxxNxL, where L is leucine, isoleucine or valine and x is any amino acid (McEwan et al., 2006). The LRR motif has been found in >100 intracellular, extracellular, and cell surface proteins (Scott et al., 2006). The LRRs adopt a curved solenoid structure with convex and concave faces that allow protein-protein interactions (Kresse and Schonherr, 2001). All SLRPs have four cysteines with class-conserved internal spacing at the N-terminus and two cysteine residues at the C-terminus (Chen and Birk, 2013). Another hallmark of SLRPs is the presence of the C-terminal ‘ear repeat’, a long repeat that extends laterally away from the protein core. LRR-XI contains one of the cysteine residues which forms a disulphide bond with the other cysteine residue on LRR-XII (McEwan et al., 2006).

The SLRP family can be divided into five classes (Classes I-V) based on conservation and homology at the genomic and protein levels and the spacing of cysteine residues at the N-terminus (Table 1) (Schaefer and Iozzo, 2008). SLRPs are closely related but are differentially expressed and have acquired special functions (Kalamajski and Oldberg, 2010). Figure 1.5 is a sequence-based evolutionary tree (dendogram) of the SLRP family.

![Figure 1.5 Phylogenetic tree of SLRPs.](image-url) Evolutionary tree based on a multiple sequence alignment by ClustalW. Adapted from McEwan et al. (2006).
Table 1 Classification and properties of SLRPs including size, type of GAG, binding partners and knockout-mice phenotypes (where available). Adapted from (Kalamajski and Oldberg, 2010, Chen and Birk, 2013)

<table>
<thead>
<tr>
<th>SLRPs</th>
<th>Size (kDa)</th>
<th>Type of GAG</th>
<th>Abundant in</th>
<th>Binding partners</th>
<th>Knock-out mice phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Class I</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decorin</td>
<td>36</td>
<td>CS/DS</td>
<td>Skin, tendon, cartilage, kidney, muscle, predentin and PDL</td>
<td>Collagen I; collagen II and III; collagen V; collagen VI; collagen XII; collagen XIV; fibronectin; thrombospondin-1; microfibril-associated glycoprotein-1 and fibrillin-1; tenascin-X</td>
<td>Fragile skin, weak tendons, lower airway resistance, slow wound healing and angiogenesis</td>
</tr>
<tr>
<td>Biglycan</td>
<td>38</td>
<td>CS/DS</td>
<td>Skin, tendon, bone, cartilage, kidney, muscle, predentin and PDL</td>
<td>Collagen I; collagen II; collagen III; collagen VI; collagen IX and biglycan; collagen II and VI complex; tropoelastin and microfibril-associated glycoprotein-1</td>
<td>Bone mass reduction, aortic rupture, and weak tendons</td>
</tr>
<tr>
<td>Asporin</td>
<td>42</td>
<td>*</td>
<td>Liver, heart, aorta, uterus, perichondrium, periosteum, predentin and PDL</td>
<td>Collagen I</td>
<td></td>
</tr>
<tr>
<td><strong>Class II</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fibromodulin</td>
<td>42</td>
<td>KS</td>
<td>Tendon, ligament, cartilage, predentin and PDL</td>
<td>Collagen I; collagen II; collagen VI; collagen IX; collagen XII</td>
<td>Weak tendons and osteoarthritis</td>
</tr>
<tr>
<td>Lumican</td>
<td>38</td>
<td>KS</td>
<td>Widespread, predentin, and PDL</td>
<td>Collagen I; aggregcan; β1 integrin; β2 integrin; α2β1 integrin</td>
<td>Fragile skin and opaque cornea</td>
</tr>
<tr>
<td>Osteoadherin</td>
<td>42</td>
<td>KS</td>
<td>Mineralised bone</td>
<td>αβ3 integrin; Non-collagenous domain 4 of collagen IX</td>
<td></td>
</tr>
<tr>
<td>PRELP</td>
<td>44</td>
<td>KS</td>
<td>Cartilage, lung kidney, and skin</td>
<td>Perlecan and collagen</td>
<td></td>
</tr>
<tr>
<td><strong>Class III</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osteoglycin</td>
<td>35</td>
<td>KS</td>
<td>Cornea and skin</td>
<td>Collagen I</td>
<td>Slight skin fragility, but clear (normal) cornea</td>
</tr>
<tr>
<td>Opticin</td>
<td>35-45</td>
<td>*</td>
<td>Vitreous (eye)</td>
<td>Heparan and chondroitin sulphate proteoglycans, collagen</td>
<td></td>
</tr>
<tr>
<td><strong>Class IV</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chondroadherin</td>
<td>36</td>
<td>KS</td>
<td>Cartilage, cornea, lens, and retina</td>
<td>α2β1 integrin; collagen II</td>
<td></td>
</tr>
<tr>
<td><strong>Class V</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Podocan</td>
<td>70</td>
<td></td>
<td></td>
<td>Collagen I</td>
<td></td>
</tr>
</tbody>
</table>

*Approximate molecular mass of the protein core without post-translational modifications  

*Type of glycosaminoglycan - (CS chondroitin sulphate, DS dermatan sulphate, KS keratan sulphate)  

*These SLRPs are non-glycanated but may have acidic regions (e.g. poly Asp in asporin) or sulphated tyrosine residues
Class I SLRPs – This family includes decorin, biglycan and asporin (Table 1). These are encoded by eight exons with highly conserved intron/exon junctions (Schaefer and Iozzo, 2008). They contain 10-12 LRRs and a defined N-terminal cysteine sequence (CX₃CXCX₆C) which forms two disulphide bonds (Ameye and Young, 2002). Biglycan and decorin share 57% sequence identities (Iozzo, 1998). They contain one or two dermatan or chondroitin sulphate chains, respectively; the choice of chains is tissue-specific (Hocking et al., 1998). Asporin, on the other hand, does not contain GAG chains, so is not a classical proteoglycan but has been placed in this class based on sequence homology (Schaefer and Iozzo, 2008). Biglycan and decorin are secreted containing N-propeptides of 21 and 14 residues respectively, which can be cleaved by BMP-1 to release the mature forms in certain tissues (Scott et al., 2000). Figure 1.6 is a schematic of the structures of class I SLRP members, decorin and biglycan.

![Figure 1.6 Schematic of Class I SLRPs, decorin and biglycan.](image)

Multiple LRRs (shown as green boxes) in the central domain make up about 70% of each core protein. The LRRs are flanked by cysteine-rich clusters on two sides that can form disulphide bonds (S-S) (arrows). Decorin carries one and biglycan carries two dermatan sulphate or chondroitin sulphate glycosaminoglycans in the N-terminal region (red wavy lines). Potential N-linked glycosylation sites are shown as red dotted lines.
Class II SLRPs – This group comprises of fibromodulin, lumican, osteoadherin and PRELP. These are encoded by three exons, with a large central exon encoding almost all LRRs (Ameye and Young, 2002). They have a conserved N-terminal cysteine sequence (CX\textsubscript{3}CXC\textsubscript{9}C) and contain sulphated tyrosine residues at the N-terminal end making them polyanionic. They primarily carry keratan sulphate proteoglycans (Schaefer and Iozzo, 2008).

Class III SLRPs – This class includes osteoglycin, epiphycan and opticin. These are encoded by seven exons and consist of seven LRRs which are encoded by the last 3 exons. The N-terminal cysteine sequence is CX\textsubscript{2}CXC\textsubscript{6}C. They also contain N-terminal sulphated tyrosine residues (Ameye and Young, 2002).

Class IV SLRPs – This family contains two members, chondroadherin and nyctalopin, which are non-canonical SLRPs. They do not have ‘ear repeats’ but contain four C-terminal cysteines forming two disulphide bonds (McEwan et al., 2006). Interestingly, nyctalopin is the first described glycosylphosphatidylinositol-anchored SLRP (Schaefer and Iozzo, 2008).

Class V SLRPs – This class of non-canonical SLRPs includes podocan. It contains a relatively large number of LRRs (22) and is the only SLRP not to have a disulphide-bonded C-terminal cap (McEwan et al., 2006). Hence, it is not a ‘classical’ SLRP but is classed as one due to functional commonality with other SLRP members such as binding to collagen I (Schaefer and Iozzo, 2008).

In the ECM, SLRPs are found ubiquitously but their distribution is regulated temporally and spatially during development. For example, the corneal stroma contains decorin and biglycan but biglycan expression decreases significantly postnatally whereas decorin expression remains stable. Lumican is expressed throughout the corneal stroma during development but restricted to the posterior stroma in the adult (Chakravarti et al., 2006). Fibromodulin, which is not considered to be a corneal component, is expressed briefly into the central cornea during early postnatal development (Chen and Birk, 2013).
1.3.2 Biological functions of SLRPs

One of the major functions of SLRPs involves regulating collagen fibril assembly as evidenced by the presence of disorganised fibrils leading to defective connective tissue phenotypes in SLRP knockout mice (Table 1) (Kalamajski and Oldberg, 2010). Various SLRPs cooperate to modulate the process of fibrillogenesis. SLRPs belonging to the same class have the same binding site on collagen and therefore compete with one another for collagen binding (for e.g. lumican competes with fibromodulin) (Chen and Birk, 2013). This gives rise to some level of functional redundancy within SLRP members. For example, mice deficient in either decorin or biglycan show a mild corneal stromal phenotype, but decorin-biglycan double knockout mice have a severe skin phenotype, indicating that the two SLRPs are sufficiently similar to compensate for the loss of either of the two (Zhang et al., 2009). It is generally accepted that SLRPs bind collagen fibrils via their concave faces. For example, lumican and fibromodulin binding to collagen type I is mediated by a homologous sequence in LRR 5-7 (Kalamajski and Oldberg, 2009). The sulphated GAG chains extend outwards into the inter-fibril space and are involved in hydration of the matrix (Chen and Birk, 2013). However, SLRP core proteins can regulate fibrillogenesis in vitro in the absence of GAGs (Rada et al., 1993).

How exactly the SLRPs regulate collagen fibrillogenesis is not known. (Kalamajski and Oldberg, 2010) suggest three possible mechanisms: a) they bind to collagen and prevent uncontrolled fibril assembly by sterical hindrance, b) SLRPs with more than one binding site can add collagens to the multimers and bridge fibrillar and FACIT collagens or, c) they may hinder access to potential cross-linking lysine residues on collagen monomers.

Apart from their role in collagen fibril assembly, SLRPs are involved in modulating signalling pathways including those of the BMP/TGFβ superfamily (Schaefer and Iozzo, 2008). At least three SLRPs (decorin, biglycan and fibromodulin) bind TGFβ, a cytokine involved in inflammation, control of cell proliferation, and apoptosis, via their protein cores (Hildebrand et al., 1994). Mice deficient in biglycan show age-dependent osteopenia as a result of defective osteoblast differentiation due to reduced BMP-4 binding, indicating a modulating role of biglycan on BMP-4 binding (Chen et al., 2004).
Decorin binds to four different receptor tyrosine kinases (RTKs) (Iozzo and Schaefer, 2010): it activates the epidermal growth factor receptor (EGFR) leading to growth suppression (Iozzo et al., 1999); it binds to insulin-like growth factor 1 receptor (IGF-IR) on endothelial cells (Schonherr et al., 2005); it binds to vascular endothelial growth factor receptor 2 (VEGFR2) and antagonises proliferation and migration of extravillous trophoblast (EVT) cells (Khan et al., 2011); and it acts as a novel antagonistic ligand of the Met receptor (Goldoni et al., 2009). These interactions have potential to be exploited as protein-based therapies for various cancers and inflammatory disorders (Schaefer and Iozzo, 2008).

1.4 Decorin

1.4.1 Structure

Decorin is the archetypal SLRP, so named because it “decorates” collagen fibrils. It is almost ubiquitously expressed in connective tissues and is conserved across species. It is synthesised mainly by fibroblasts, stressed vascular endothelial cells, and smooth muscle cells (Neill et al., 2012). It consists of a signal sequence, a short cleavable propeptide, a single N-terminal dermatan/chondroitin sulphate GAG chain attached to a serine residue, and a core protein of ∼45 kDa containing 12 LRRs (numbered 1-XII), flanked by disulphide-bonded sequences (Figure 1.7). The crystal structure of the decorin core protein, determined by Scott et al., 2004, revealed that the twelve LRRs form a right-handed curved solenoid which is “banana-shaped”, in contrast to the more closed “horseshoe-shaped” porcine ribonuclease inhibitor crystal structure (Kobe and Deisenhofer, 1993), which had previously been used to generate homology models for decorin (Weber et al., 1996). The 12 LRRs are in tandem array and vary in length from 21 to 30 amino acids, following a short-long-long regular pattern throughout the molecule (Scott et al., 2004). The first nine LRRs have a 21-24-24 repeating structure, except LRR-VI which is 26 residues long. The last three LRRs show a 23-30-27 pattern (McEwan et al., 2006). The inner concave face is made up of parallel β-sheets and the convex back consists of irregular loops and single helical turns. Differences in the length of the LRRs and secondary structure elements forming the convex face gives rise to the
variation in molecular curvature. The core proteins of bovine and mouse decorin have 3 and 4 consensus sites for N-linked glycosylation, respectively. The presence of decorin core protein without a GAG chain has been reported in some tissues but its relevance is unknown (Seo et al., 2005). Two distinctive features of SLRPs also present in decorin include a) the N-terminal capping motif which forms where four cysteine residues form a disulphide knot which buries the hydrophobic core of the first LRR and, b) the C-terminal “ear-repeat” where LRR-XII extends laterally and a conserved cysteine residue forms a disulphide bond with another one on the last LRR (Scott et al., 2004). The decorin crystal structure revealed that two monomers associate through their concave faces, burying a large surface area (~2,300 Å²) to form an antiparallel dimer.
**Figure 1.7** Schematic of bovine decorin core protein and its internal organisation. Green arrows represent β-strands, red ribbons show α-helical loops and glycosylation sites are shown in cyan. The N- and C- termini are labelled. The figure was generated using PyMOL (http://www.pymol.org), from the PDB entry: 1xku. Adapted from Scott et al. (2004). Amino acid sequences for the 12 LRRs are tabulated. LRR consensus sequences are highlighted in green.

<table>
<thead>
<tr>
<th>LRR</th>
<th>Length</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>G P V C P F R C Q C H</td>
<td>22–32</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>21</td>
<td>P R V V Q C S D G E K V P K D L P P D</td>
</tr>
<tr>
<td>II</td>
<td>24</td>
<td>A L D D Q N K T E I K D G D F K N L K N</td>
</tr>
<tr>
<td>III</td>
<td>24</td>
<td>H T L I L I N K I S K I S P G A F A P L V K</td>
</tr>
<tr>
<td>IV</td>
<td>21</td>
<td>E R Y L S K Q K E L P E K M P K T</td>
</tr>
<tr>
<td>V</td>
<td>24</td>
<td>Q E R V H E E H T K V R K S V F N G L N Q</td>
</tr>
<tr>
<td>VI</td>
<td>26</td>
<td>I V V E G T S P K S G I E N G A F Q G M K K</td>
</tr>
<tr>
<td>VII</td>
<td>21</td>
<td>S Y E R A D A N E T T I P Q G L P P S</td>
</tr>
<tr>
<td>VIII</td>
<td>24</td>
<td>T E H E D G N K T K V D A A S L K G L N N</td>
</tr>
<tr>
<td>IX</td>
<td>24</td>
<td>A K E G S F S F S A V D N S L A N T P H</td>
</tr>
<tr>
<td>X</td>
<td>23</td>
<td>R E L E H N N K V K P G G L A D H K Y</td>
</tr>
<tr>
<td>XI</td>
<td>30</td>
<td>Q V Y E H N N N S A I G S N D F C P P G Y N T K K A S</td>
</tr>
<tr>
<td>XII</td>
<td>27</td>
<td>S G V S C F S P V Q W E I Q P S T F R C V Y V R</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A A V Q</td>
</tr>
</tbody>
</table>

Amino acid sequences for the 12 LRRs are tabulated. LRR consensus sequences are highlighted in green.
1.4.2 Function in extracellular assembly

Several important functions have been attributed to decorin, based on the interaction of the core protein with other proteins, including regulation of collagen fibrillogenesis, corneal transparency, tensile strength of skin and tendon, and tumour repression (Keene et al., 2000). The main driving force behind fibril formation is the collagen structure itself but SLRPs seem to be important in the process as well. One of the main roles of decorin is to regulate collagen fibril formation; it does so by delaying fibril assembly and reducing fibril diameter (Figure 1.8) (Chen and Birk, 2013, Kalamajski and Oldberg, 2010). Decorin has been shown to bind collagen types I and II, III, VI and XIV (Vogel et al., 1984, Bidanset et al., 1992, Douglas et al., 2006). The concave face is thought to be involved in protein-protein interactions (Kobe and Kajava, 2001).

---

**Figure 1.8** Transmission electron micrographs of dermal collagen of the skin from the decorin (Dcn)-deficient mice. (A and B) Dcn -/- and (C) Dcn+/- mice. Dcn -/- shows larger (>200 nm) and irregular fibrils (A and B, asterisks) in addition to smaller (30–40 nm) fibrils (A, circles). Arrowheads in B indicate coarser fibrils showing lateral fusion to an adjacent tapered segment. C. Collagen from the wild-type mouse shows a more compact and uniform pattern of fibril diameter and distribution. Adapted from Danielson et al. (1997).
Decorin knock-out mice are viable and grossly normal but have fragile skin, with reduced tensile strength and a thinner than normal dermis. The collagen defects are very similar to the cutaneous defects observed in human Ehlers-Danlos syndrome (EDS) (Ameye and Young, 2002). Compared to wild-type mice, decorin-deficient mice have significantly delayed dermal wound healing, indicating that decorin has a role in tissue repair (Jarvelainen et al., 2006). Targeted disruption of the decorin gene in the periodontal ligament (PDL) of mice shows collagen fibrils with abnormal morphology arranged in a random orientation instead of the normal parallel orientation (Hakkinen et al., 2000). Mice lacking decorin or biglycan show changes in collagen fibril size and shape in bone. However, there are no changes in bone mass or other skeletal phenotypes in decorin-deficient mice, unlike biglycan-deficient animals, indicating that despite their high level of homology, the two SLRPs have different functions in vivo (Corsi et al., 2002). Decorin-deficient mice have higher biglycan expression, but the reverse is not the case. Decorin-biglycan double-KOs show severely disrupted collagen fibril structures in the cornea, indicating that biglycan compensates for the loss of decorin (Zhang et al., 2009). Three frameshift mutations at the C-terminus of the decorin gene have been associated with a human disease called congenital stromal corneal dystrophy (Bredrup et al., 2005) that leads to corneal opacity, which has also been observed in mouse models of this disease (Chen et al., 2011).

Decorin can act as an adhesion site for *Borrelia burgdorferi*, the causative agent of Lyme disease, which resides in the ECM and expresses two surface-exposed decorin-binding adhesins (Guo et al., 1995). Decorin-deficient mice are more resistant to the disease (Brown et al., 2001).

### 1.4.3 Decorin: role in cell signalling and cancer

In addition to being a structural protein, decorin is recognised as having key roles in cell signalling with its ability to regulate downstream signalling indirectly via sequestration of growth factors, and directly by binding and down-regulating several RTKs, which are often overexpressed in cancer cells (Neill et al., 2012). When expressed in tumour xenograft-bearing mice or injected systemically, decorin inhibits both primary tumour
growth and metastatic spreading (Goldoni and Iozzo, 2008). The potential of decorin as a novel biological target for the treatment of tumours is being investigated these days by many researchers (Sofeu Feugaing et al., 2013).

Yamaguchi et al. (1990) first discovered that the stable transfection of decorin causes growth arrest in Chinese hamster ovary cells and this was due to the anti-proliferative effect of decorin, i.e. its ability to bind and block TGF-β (Neill et al., 2012). Decorin directly interacts with EGFR, causing rapid phosphorylation of the EGFR in squamous carcinoma cells, leading to activation of the MAP kinase signal pathway, induction of endogenous p21 and growth suppression (Iozzo et al., 1999). This signal transduction process leads to down-regulation of the receptor (Csordas et al., 2000), endocytosis of decorin-bound EGFR and ultimately to lysosomal degradation of the internalised decorin (Zhu et al., 2005). Decorin interaction with the EGFR is mediated by LRR-7 of the decorin protein core (Santra et al., 2002). A stoichiometry of 1:1 for the decorin protein core and EGFR is likely, due to structural constraints of the EGFR binding region, suggesting that biologically active decorin is a monomer (Iozzo and Schaefer, 2010, Goldoni et al., 2004). Decorin also binds Met, the receptor for hepatocyte growth factor, with high affinity (Goldoni et al., 2009). This interaction leads to transient activation of the receptor, recruitment of the E3 ubiquitin ligase c-Cbl, followed by rapid intracellular degradation of Met. Signalling through Met leads to suppression of β-catenin, a downstream Met effector, leading to inhibition of tumour growth and metastasis (Goldoni et al., 2009). It has been shown that ~30% of decorin-null mice develop intestinal tumours, associated with up-regulation of β-catenin signalling (Bi et al., 2008). In normal cells, decorin exerts proliferative effects and a prosurvival response, favouring cell growth, via its effects on IGF-IR (Iozzo and Schaefer, 2010). Decorin binds to IGF-IR in endothelial cells and also binds to and sequesters IGF, the natural ligand for IGF-IR. Decorin addition causes IGF-IR phosphorylation and activation, which is followed by receptor down-regulation. These effects are caused by the core protein of decorin, and the binding region could be mapped to the N terminus of the molecule. Binding to IGF-IR leads to phosphorylation of protein kinase B, induction of p21, resulting in inhibition of apoptosis in endothelial cells (Schonherr et al., 2005). Figure 1.9 summarises the roles of decorin in tumour progression.
Figure 1.9 Role of decorin in tumour progression. Decorin is produced and secreted by fibroblasts of the tumour stroma (light yellow spindle-shaped cells) and affects tumour cells (centre) via multiple pathways: (1) decorin affects signal transduction via multiple RTKs, thus modulating cell proliferation, cell cycle progression and apoptosis. For example, interaction of the decorin with erbB family receptors leads to signal transduction characterised by stimulation of MAP kinases, up-regulation of p21, an inhibitor of cyclin-dependent kinases, and ultimately to growth suppression. (2) Decorin induces endocytosis of the EGFR, leading to receptor downregulation and a reduction in extracellular decorin. (3) Decorin promotes matrix organisation (collagen fibrillogenesis) which constitutes a physical barrier against migration/motility of cancerous cells, thus preventing metastasis (4) Decorin can negatively regulate signal transduction by interfering with growth factor binding, such as TGF-β. As a result, synthesis of matrix molecules and growth factor-dependent modulation of cell proliferation and migration are inhibited. (5) Decorin affects tumour angiogenesis via the VEGF-R, PDGF-R and FGF-R pathways. Taken from Sofeu Feugaing et al. (2013).
1.4.4 Decorin: monomer or dimer?

Ultrastructural studies of tissue-derived collagen fibrils using stains such as Cupromeronic blue (Scott and Orford, 1981) and antibodies against the decorin core protein (Pringle and Dodd, 1990) have shown that decorin binds in the gap zone of the D-period. Keene et al. (2000) mapped the binding site of decorin core protein to a narrow region in the C-terminus of the collagen type I triple helix, very close to one of the major intermolecular cross-linking sites, using rotary shadowing electron microscopy and photoaffinity labelling. Interaction studies using peptides derived by cyanogen bromide (CNBr) cleavage of type I and type II collagen with decorin extracted from bovine tendon showed that decorin makes contacts with multiple sites in type I collagen and probably also in type II collagen and that some collagen Lys/Hyl residues are essential for the binding (Tenni et al., 2002).

Several structural and biophysical studies have indicated that decorin and other SLRPs form stable dimers, that monomeric decorin cannot exist in solution and that dimerisation is mediated by the concave face, potentially making this surface unavailable for ligand binding (Scott et al., 2003, Scott et al., 2004, Scott et al., 2006). In contrast to this view, Goldoni et al. (2004) have claimed that biologically active decorin is monomeric and dimerisation is an artefact of lyophilisation. Other researchers have found that like most SLRPs, decorin-collagen interactions appear to be mediated via the concave face of the core protein, buried in the dimer interface, which is also the most conserved region. Using chimeras of recombinant decorin and biglycan, it has been shown that the decorin binding sites for collagen I are located in LRRs 5-6 (Svensson et al., 1995). A critical role of Glu-180 residue in LRR-6 has been suggested by studying interactions of recombinant decorin, expressed in mammalian cells, with collagen type I (Kresse et al., 1997). Site-directed mutagenesis studies have further localised the binding site to the sequence, SYIRIADTNIT, in LRR-6 (Kalamajski et al., 2007). It is to be noted that these studies have a slightly different numbering system for the LRRs than that used in Figure 1.7 since the presence of the first LRR only became evident after the elucidation of the crystal structure. Molecular modelling studies of the interactions between decorin core protein and type I collagen fibrils were investigated using their respective X-ray diffraction structures (Orgel et al., 2009, Weber et al., 1996). The
monomeric form of decorin core protein showed the most appropriate shape complementarity with the collagen fibril surface and favourable calculated energies of interaction (Orgel et al., 2009).

These data raise questions about the oligomeric state of decorin and its impact on collagen binding. Hence, the aim of my project was to resolve this controversy by determining the oligomeric state of decorin and studying its interactions with collagen using various biophysical and biochemical techniques.

1.5 SPARC

Matricellular proteins are ECM molecules that have no primary structural roles in the ECM but are involved in mediating cell-matrix interactions (Brekken and Sage, 2000). Members of this family of extracellular proteins include thrombospondin (TSP) 1 and 2, SC1/hevin, tenascin C and X and SPARC (also known as osteonectin or BM-40). Some distinguishing characteristics of matricellular proteins are: a) they are expressed at high levels during embryogenesis and in response to injury, declining postnatally; b) they bind to many cell-surface receptors, the ECM, growth factors, cytokines and proteases; c) they generally induce de-adhesion of anchorage-dependent cells; and d) in most cases, targeted gene disruption in mice produces either a grossly normal or a subtle phenotype that is exacerbated upon injury (Bornstein and Sage, 2002, Aszodi et al., 2006).

SPARC is a matricellular glycoprotein that is conserved in all animals from C. elegans and Drosophila, to mammals (Bradshaw, 2009). It was first described as a major component of human and bovine bone and as a protein secreted by proliferating cells in vitro (Lane and Sage, 1994). SPARC is expressed by many different cell types including endothelial cells, fibroblasts, osteoblasts and macrophages (Trombetta-Esilva and Bradshaw, 2012). It has been shown to induce cell rounding in endothelial cells, fibroblasts and smooth muscle cells and functions as a counter-adhesive protein (Sage et al., 1989). A third function of SPARC is in the regulation of the ECM production and
turnover through its effects on collagen and extracellular proteases (Bradshaw and Sage, 2001).

1.5.1 SPARC: general structure

In both vertebrates and invertebrates, SPARC is a single copy gene that is highly conserved (Table 2). The murine SPARC gene has 10 exons whereas the *C. elegans* homologue only contains 6 exons, as a result of which the nematode protein is truncated in its first acidic module. However, this truncation does not affect the protein’s ability to bind Ca$^{2+}$ and collagen (Brekken and Sage, 2000). In contrast to *C. elegans*, *Drosophila* SPARC is longer than vertebrate SPARCs, mainly due to additional amino acid residues in the N-terminal domain. Whether these additional residues confer unique functions to *Drosophila* SPARC remains to be determined (Martinek et al., 2002).

Human SPARC protein is 32 kDa in size and has 303 amino acids, with the initial 17 amino acids containing the signal sequence, which is removed upon secretion. The mature SPARC protein has 286 amino acids with three distinct modular domains: i) an N-terminal acidic domain (residues 1-52) which binds several Ca$^{2+}$ ions with low affinity and also to hydroxyapatite, ii) a central follistatin-like (FS) domain (residues 53-137) and iii) an $\alpha$-helical extracellular high affinity Ca$^{2+}$-binding (EC) domain (residues 138-286) containing a pair of EF-hands and the collagen binding site (Figure 1.10) (Bradshaw and Sage, 2001). This domain organisation is phylogenetically conserved (Martinek et al., 2002).
Table 2: Amino acid sequence identities of SPARC from different organisms; Adapted from Brekken and Sage (2000)

<table>
<thead>
<tr>
<th>Accession no. from NCBI databank</th>
<th>SPARC % amino acid sequence identity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bovine</strong></td>
<td>AAA30678</td>
</tr>
<tr>
<td><strong>Mouse</strong></td>
<td>CAA27642</td>
</tr>
<tr>
<td><strong>Drosophila</strong></td>
<td>CAB39319</td>
</tr>
<tr>
<td><strong>C. elegans</strong></td>
<td>AAA16827</td>
</tr>
</tbody>
</table>

![Figure 1.10 Schematic of the modular domains of SPARC protein](image)

The follistatin-like domain is shown in red and the EC module (aa 138-286) is shown in blue. The FS domain contains the peptide 2.1 shown in green; The (K)GHK angiogenic peptide shown in black; The EC domain contains peptide 4.2 (aa 255-274) shown in yellow. Taken from (Brekken and Sage, 2000), originally drawn from PDB entry 1BMO (Hohenester et al., 1997).
The FS domain contains 10 cysteines and an N-linked glycan at Asn99. It has an elongated structure consisting of an N-terminal β-hairpin and a small core of mixed α/β structure. The structure is homologous to a repeated domain in follistatin, a protein that binds to members of the TGF-β family, and to domains in agrin. The N-terminal β-hairpin is very similar to that of EGF and the core resembles a serine protease inhibitor of the Kazal family (Hohenester et al., 1997). The FS domain also contains peptides that exert different effects on endothelial cells. Peptide 2.1 inhibits the proliferation of endothelial cells whereas peptide 2.3 stimulates endothelial cell proliferation (Figure 1.10) (Funk and Sage, 1993).

The EC domain is a compact, highly α-helical structure that contains two EF-hand motifs, which bind Ca\(^{2+}\) with high affinity (Hohenester et al., 1996). Binding of Ca\(^{2+}\) to the EF-hands increases the α-helical secondary structure content in SPARC, which is dependent on cooperative interactions between the EF-hands and FS domain (Maurer et al., 1995). The fibril-forming collagen types I, III, and V, and the basement membrane collagen type IV, bind this domain in a Ca\(^{2+}\)-dependent fashion (Termine et al., 1981, Sage et al., 1989, Mayer et al., 1991, Maurer et al., 1995). This domain contains peptide 4.2 which has been shown to inhibit endothelial cell proliferation (Brekken and Sage, 2000).

### 1.5.2 SPARC – growth factor interactions

SPARC modulates the activity of at least three growth factor/cytokine signalling pathways, mediated by cell surface receptors including vascular endothelial growth factor (VEGF) receptor, basic fibroblast growth factor (bFGF, FGF2), and TGF-β1 (Bradshaw, 2012). SPARC directly interacts with VEGF-A and prevents it from inducing VEGFR1 activity in cultured endothelial cells and in vivo. SPARC inhibits the FGF2-induced activation of endothelial cells in vitro. Another growth factor crucial for vascular homeostasis is TGF-β1. Genetic ablation of TGF-β1 results in abnormal vessel formation and embryonic lethality in mice (Rivera et al., 2011). Whether TGF-β1 and SPARC interact directly is unknown but TGF-β1 induces expression of SPARC in a
variety of cell types including fibroblasts and mesangial cells (Francki et al., 1999). Conversely, SPARC also regulates expression of TGF-β1.

In Drosophila, SPARC is upregulated by "loser" cells – cells proximal to "killer" or "winner" cells, during cell competition (Portela et al., 2010). The loser cells are induced to undergo apoptosis by a secreted factor from winner cells. SPARC expression by loser cells prolongs the life of these cells by inhibiting an unidentified secreted killing signal and preventing, at least temporarily, apoptosis of the loser cell (Bradshaw, 2012).

1.5.3 SPARC – collagen interactions

The binding of SPARC to collagen is specific to the collagen triple helix as it is abolished by heat denaturation of collagen (Giudici et al., 2008). Proteolytic cleavage of full-length SPARC or the EC module by several MMPs at the peptide bond between Leu197-Leu198 in helix-αC results in a 10-fold increase in collagen affinity. Cleavage of helix-αC exposes the collagen binding site on helix-αA, thereby increasing binding affinity (Sasaki et al., 1997). Deletion of helix-αC in recombinant SPARC gives a similar increase in binding affinity (Sasaki et al., 1998). Differentially glycosylated forms of SPARC may have different functional roles in vivo. Bone SPARC carrying high-mannose structures binds collagen I with higher affinity than platelet SPARC carrying bi-antennary and tri-antennary glycans (Kaufmann et al., 2004). On the other hand, a single N99Q mutation to prevent N-glycosylation causes no significant change in the binding of SPARC to collagens I and IV (Sasaki et al., 1997).

The SPARC binding site on collagen I has been mapped in two separate studies. Wang et al. (2005) identified a broad distribution of SPARC binding sites, with a primary binding site ~1/3 (87.5 – 125nm) from the C-terminus of procollagen I, using tapping-mode atomic force microscopy. Another study by Giudici et al., using rotary shadowing followed by electron microscopy, found a major site located ~180nm from the C-terminus and a less preferred site 60-100nm from the C-terminus. The discrepancy between the two studies could be due to the different sources of SPARC and procollagen, and the different experimental methods used (Giudici et al., 2008).
The crystal structure of human SPARC FS-EC bound to a 33-residue collagen-like triple helical peptide has been recently determined (Figure 1.11) (Hohenester et al., 2008). The peptide represents residues 564-590, the major binding site identified by Giudici et al. SPARC has been shown to bind the GVMGFO (O stands for hydroxyproline) motif of the middle and trailing collagen chains, burying a total of 720 Å² of solvent-accessible collagen surface. This motif also binds two other unrelated proteins: von Willebrand factor and discoidin domain receptor 2. The structure revealed that the collagen triple helix is not distorted upon SPARC binding. In contrast, the conformation of SPARC is significantly altered, resulting in a deep pocket that accommodates the phenylalanine residue of the trailing collagen chain (Hohenester et al., 2008).
1.5.4 SPARC in ECM assembly

The phenotype of SPARC-null mice is grossly normal except that they display some abnormalities in the ECM: a) early onset cataract formation and rupture of the lens capsule (Gilmour et al., 1998); b) osteopenia, with decreased bone formation and
decreased osteoblast and osteoclast surface and number (Delany et al., 2000); and c) accelerated cutaneous wound closure (Bradshaw et al., 2002). These phenotypes can be attributed to significant changes in collagen fibril morphology and decreases in collagen concentrations in SPARC-null mice compared with wild-type animals (Bradshaw, 2009). The skin of SPARC-null mice has roughly half the amount of collagen in comparison with wild-type, as measured by hydroxyproline analysis (Bradshaw et al., 2003). SPARC and collagen I may be involved in ECM remodelling. For example, Mov-13 mice carry a mutation in the collagen α1(I) gene and do not express collagen I in the ECM. These mice have impaired deposition of SPARC within the ECM, suggesting that collagen I influences the spatial distribution of SPARC (Iruela-Arispe et al., 1996).

SPARC also interacts with collagen IV and influences basal lamina assembly. Studies with SPARC-GFP fusion protein show that the protein is localised to most BMs along body wall and sex muscles and overlaps with the distribution of collagen IV in C. elegans. Elimination of SPARC by RNA interference results in embryonic or larval lethality (Fitzgerald and Schwarzbauer, 1998). This may be due to the fact that unlike vertebrates, C. elegans does not contain any SPARC homologues so there is no functional compensation. Collagen IV is required during development to provide mechanical strength in order to withstand contractile forces during embryonic movements (Yurchenco et al., 2004, Martinek et al., 2008). SPARC and other BM components such as network-forming collagen IV are synthesised and secreted by haemocytes and assembled in BMs mid-embryogenesis in Drosophila (Martinek et al., 2002). Inhibition of SPARC expression in Drosophila leads to the absence of collagen IV in the BM of certain structures such as the ventral nerve chord, resulting in embryonic lethality (Martinek et al., 2008). SPARC mutants and collagen IV mutants also show striking phenotypic similarity in flies suggesting a mutual dependence of the two proteins (Martinek et al., 2008). It has been proposed that intracellular SPARC acts as the principal collagen IV chaperone in invertebrates, which do not have other molecular chaperones such as Hsp47, thus ensuring stability of the triple helix prior to export from the endoplasmic reticulum (Martinek et al., 2002). Martinek et al. (2002) have further hypothesised that SPARC promotes collagen fibril assembly by modulating the binding of molecules such as decorin.
1.5.5 Collagen binding by vertebrate and invertebrate SPARC

Although the collagen binding residues between human and invertebrate SPARC (Drosophila (dSPARC) and C. elegans (cSPARC)) are highly conserved (Appendix 1), there is one difference between the orthologues: collagen binds to human SPARC following proteolytic nicking of an inhibitory loop which is not present in invertebrate SPARC. Lacking helix αB, invertebrate SPARCs have a much shorter connection between helices αA and αD (Hohenester et al., 2008). Therefore, it can be hypothesised that collagen binding to invertebrate SPARC does not require proteolytic activation and is of high affinity constitutively. To test this hypothesis, I aimed to study binding interactions between human and invertebrate SPARC and collagen I and IV and attempted to determine the crystal structure of dSPARC and cSPARC.

1.6 Aims of the Project

The broad aim of the projects outlined in this thesis was to study the interaction of the proteins, decorin and SPARC, with collagen.

More specifically, the aims were as follows:

1) The crystal structure of decorin revealed a dimeric structure in which the presumed collagen binding site is not accessible. Whether the dimer is functional remains controversial and it has been claimed that biologically active decorin is a monomer. The main aim of the study was to resolve the controversy about the oligomeric state of decorin by designing a number of decorin mutants to disrupt the crystallographic dimer. Size exclusion chromatography with multi-angle laser light scattering and analytical ultracentrifugation was used to determine the oligomeric state of decorin. A further aim was to identify whether decorin binds collagen as a monomer or a dimer, indirectly by measuring the inhibition of collagen fibrillogenesis and directly using solid-phase assay.

2) The crystal structure of human SPARC (hSPARC) bound to a collagen-like triple-helical peptide was determined in the Hohenester lab. The key collagen binding
residues are conserved between human and invertebrate SPARC (in this study, dSPARC and cSPARC). However, collagen binds to hSPARC following proteolytic cleavage of an inhibitory loop that is not present in invertebrate SPARC. The aim of this research was to investigate the role of the inhibitory loop on collagen binding by studying binding interactions between human and invertebrate SPARC and collagen I and IV, using solid phase assays and surface plasmon resonance. Another aim was to attempt to determine the crystal structure of dSPARC and cSPARC, which involved enzymatic deglycosylation or mutagenesis of the constructs to aid crystallisation.
Chapter 2:
Materials and Methods
Chapter 2. Materials and Methods

2.1 Materials

2.1.1 Chemicals and Reagents

All chemicals used in this study were of analytical grade and all solutions were prepared with ultrapure deionised water. Agarose, Luria agar, calcium chloride (CaCl₂), sodium hydrogen carbonate (NaHCO₃), 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES), Tris(hydroxymethyl)aminomethane (Tris base), sodium dodecyl sulphate (SDS), polyoxyethylenesorbitan monolaurate (Tween - 20), ethylenediaminetetraacetic acid (EDTA), ampicillin, avidin peroxidase, K-casein from bovine milk, rat tail type I collagen, collagen IV from human placenta and bovine serum albumin (BSA) were purchased from Sigma Aldrich (Gillingham, UK). Luria Bertani broth was purchased from Melford (Suffolk, UK). Imidazole buffer substance was from MERCK (Nottingham, UK). Gibco® phosphate-buffered saline (PBS) tablets were purchased from Life technologies (Paisley, UK). Ammonium persulphate (APS) was obtained from Fisons Scientific (Loughborough, UK). Sodium chloride (NaCl), tetramethyleneethylenediamine (TEMED), Acrylamide:Bis-Acrylamide 37.5:1, ethanol and isopropanol were purchased from Fisher Scientific (Loughborough, UK). Quick Coomasie stain and InstantBlue™ gel stain were obtained from Generon (Berkshire, UK). Protein Ladder (10-250 kDa), restriction enzymes and reaction buffers were purchased from New England Biolabs (NEB) (Ipswich, UK). HyperLadder™ I was from Bioline Reagents Ltd (London, UK). Platinum® Pfx Polymerase Polymerase Chain Reaction (PCR) kit was purchased from Invitrogen (Paisley, UK). Plasmid DNA purification kits were purchased from QIAGEN (Crawley, UK). Mouse anti c-myc antibody was purchased from AbD Serotec (Oxford, UK). Polyclonal goat anti-mouse immunoglobulins/HRP was purchased from DAKO (High Wycombe, UK). Goat anti-mouse decorin polyclonal antibody and rabbit anti-goat IgG HRP affinity purified polyclonal antibody were from R&D Systems Europe Ltd (Abingdon, UK). SYPRO orange dye was from Invitrogen (Paisley, UK). EZ-Link® Sulfo-NHS-LC-Biotin was purchased from Thermo Scientific (Illinois, USA). PCR primers were ordered from Sigma Aldrich (Gillingham, UK) and are listed in Table 2.1. Agarose bound lectins were purchased from Vector Laboratories (Peterborough, UK).
2.1.2 Bacterial Cells and Plasmids

Subcloning Efficiency™ DH5α™ Competent Cells were purchased from Invitrogen (Paisley, UK). *Drosophila* SPARC DNA (AJ_133736) was a gift from Takako Sasaki (University of Erlangen, Germany). Plasmid cDNA containing human SPARC (hSPARC) DNA was provided by Erhard Hohenester (Imperial College London, UK). Bacterial strain transformed with *C. elegans* SPARC cDNA, a Bluescript plasmid isolated from a λ-phage cDNA library, called F4-15, was kindly provided by Dr Jean Schwarzbauer (Princeton University, USA). Plasmid pCMV6 containing mouse decorin cDNA was purchased from Origene (Maryland, USA).

2.1.3 Mammalian Cells and Cell Culture Reagents

Human embryonic kidney (HEK)293 c18 cells were purchased from American Type Culture Collection (Teddington, UK). Dulbecco's modified Eagle's medium (D-MEM) F12 (1:1), foetal bovine serum (FBS), penicillin/streptomycin (Pen/Strep), L-glutamine, trypsin-EDTA solution and Geneticin® selective antibiotic (G418 sulphate) were purchased from Invitrogen (Paisley, UK). Puromycin dihydrochloride was obtained from Sigma Aldrich (Gillingham, UK). Fugene® 6 transfection reagents were purchased from Roche Diagnostics (Burgess Hill, UK) and Promega Corporation (Madison, USA).

2.1.4 Equipment

A NanoDrop ND-1000 UV spectrophotometer from Labtech International (Ringmer, UK) was used to quantify DNA and protein concentrations. VisionWorks®LS Image Acquisition and Analysis Software was used to visualise DNA and protein gels. Mastercycler® Personal by Eppendorf (Stevenage, UK) was used for PCR cycling. Routine centrifugation was carried out using a ALC PK121R multispeed centrifuge supplied by Jencon-PLS (Bedfordshire, UK) or Eppendorf bench-top centrifuge. Slide-A-Lyser® dialysis cassettes and SnakeSkin™ dialysis tubing (10kDa molecular weight cut-off) were from Thermo Scientific (Illinois, USA). Nunc tissue culture flasks (T25, T75
and T175) were purchased from Scientific Laboratory Supplies (Yorkshire, UK). Corning HYPERFlask M cell culture vessels were purchased from Fisher Scientific (Loughborough, UK). 5 ml HisTrap™ High performance (HP) columns were obtained from GE Healthcare (Buckinghamshire, UK) for use in immobilised metal affinity chromatography. Superdex™ 200 HR 10/30 and HiLoad™ 16/60 Superdex™ 200 prep grade columns from GE Healthcare (Buckinghamshire, UK) were used for gel filtration of proteins. An ÄKTApurifier liquid chromatography system and an ÄKTAFPLC from GE Healthcare were used for protein purification. Nunc Maxisorp® flat-bottom 96 well plates were obtained from VWR (Leicestershire, UK). A Sunrise™ microplate absorbance reader from Tecan Group Ltd. (Reading, UK) was used to measure absorbance in solid phase assays. Size exclusion chromatography - multi angle light scattering (SEC-MALS) was carried out using 1260 Infinity high pressure liquid chromatography (HPLC) from Agilent Technologies (Berkshire, UK), in-line with a light scattering detector, Wyatt Mini Dawn, and a refractive index detector, Optilab T-rEX. Biacore experiments were carried out on a Biacore® 3000 instrument from GE Healthcare (Buckinghamshire, UK). Crystallisation trials were set up using a Mosquito robot from TTP Labtech. A Shimadzu UV-2501PC spectrophotometer was used for collagen fibrillogenesis assay. Differential scanning calorimetry (DSC) experiments were performed using an N-DSCIII Differential Scanning Calorimeter from TA Instruments (Herts, UK). Differential scanning fluorometry (DSF) were performed using a Stratagene Mx3005p system from Agilent Technologies. Analytical ultracentrifugation (AUC) experiment was performed using a Beckman XL-1 analytical centrifuge from Beckman Coulter (UK) Ltd (High Wycombe, UK).
2.2 Methods

2.2.1 General molecular biology methods

2.2.1.1 Polymerase Chain Reaction (PCR)

PCR was performed using Platinum Pfx polymerase in 50 µl reaction mixtures. The reaction mixtures consisted of 5 µl of 10x amplification buffer, 3 µl of 10 mM dNTP, 5 µl of 10x enhancer, 3 µl of 50 mM MgSO₄, 1 µl of template DNA (up between 50 ng and 200 ng), 1 µl of each 100 µM primer, 1 µl (1.5 U) of Pfx DNA polymerase and distilled water. The PCR cycle was initiated by denaturation at 94 °C for 1 min, followed by 29 cycles of denaturation at 94 °C for 30 sec, annealing at 55 °C for 45 sec and extension at 72 °C for 2 min. The reaction was completed by a final extension at 72 °C for 6 min. The primer sequences are tabulated in Table 2.1.

2.2.1.2 Splicing by Overlap Extension (SOE) PCR

Overlap extension PCR (Heckman and Pease, 2007) (Figure 2.1) was used to generate decorin mutant cDNA constructs. Briefly, complementary oligodeoxyribonucleotide (oligo) primers containing the nucleotide changes and PCR was used to generate two mutated DNA fragments having overlapping ends. These fragments were gel purified and subsequently combined in a 'fusion' PCR reaction in which the overlapping, complementary regions containing the desired mutation anneal.
Figure 2.1 Splicing by overlap extension PCR for site-directed mutagenesis. The first step involves two independent PCR reactions using external primers (a and d) and mutagenic primers (b and c), yielding two intermediate products AB and CD, each containing the desired mutation indicated by the asterisk. Products AB and CD hybridise via their overlapping complementary region and form the template for the second PCR reaction which uses external primers a and d to make the full-length double-stranded product AD.
Table 2.1 Primers used for generating decorin expression plasmids. Underlined sequences GCTAGC and GCGGCCGC correspond to restriction sites, \textit{NheI} and \textit{NotI}, incorporated into the forward and reverse primers respectively. The coloured triplets correspond to the mutated nucleotides.

<table>
<thead>
<tr>
<th>DNA Construct</th>
<th>Primer Sequence* (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dec CHis</td>
<td>F: TCGCTAGCCCTGATATCTATGTGCCCCTACCGATG</td>
</tr>
<tr>
<td></td>
<td>R: ATCCGCCGCGCCGCTTGTAGTTTCCAAGTTGAATGGCAGACG</td>
</tr>
<tr>
<td>Decorin mutants external primers</td>
<td>F (182bp_up): GCGCCAGCTGATCAAGCTCTGCTGCC or F: TCGCTAGCCCTGATATCTATGTGCCCCTACCGATG</td>
</tr>
<tr>
<td></td>
<td>R: ATCCGCCGCGCCGCTTGTAGTTTCCAAGTTGAATGGCAGACG</td>
</tr>
<tr>
<td>Dec Q61N internal primers</td>
<td>F: CTT CGA GTG GTG AAC TGT TCT GAT CTG GGT</td>
</tr>
<tr>
<td></td>
<td>R: CAG ATG AGA ACA GTT CAC CAC TCG AAG ATG</td>
</tr>
<tr>
<td>Dec Y130N internal primers</td>
<td>F: TTG GAA AGG CTT AAC CTG TCT AAG AAC CAA</td>
</tr>
<tr>
<td></td>
<td>R: GTT CTT AGA CAG GTT AAG CCT TTC CAA CTT CAC</td>
</tr>
<tr>
<td>Dec YRQ-A internal primers</td>
<td>F: TCT ATG TGC CCC GCA GCA TGC GCA TGT CAT CTT CGA GTG</td>
</tr>
<tr>
<td></td>
<td>R: CAC TCG AAG ATG ACA TGC GCA TGC TGC GGG GCA CAT AGA</td>
</tr>
<tr>
<td>Dec R121E internal primers</td>
<td>F: CTC CAG GAA CTT GAA GT C AT GAG AAT GAG</td>
</tr>
<tr>
<td></td>
<td>R: ATT CTC ATG GAC TTC AAG TCT GTG AGT TCT</td>
</tr>
<tr>
<td>Dec E156K internal primers</td>
<td>F: CCC AGA ACT CTC CAG AAG CTT CTG GTG CAT GAG</td>
</tr>
<tr>
<td></td>
<td>R: CTC ATG GAC ACG AAG CTT CTG GAG AGT TCT GGG</td>
</tr>
<tr>
<td>Dec K159E internal primers</td>
<td>F: GAG AAT GAG ATC ACC GAG CTG CCG AAA TCC</td>
</tr>
<tr>
<td></td>
<td>R: GGA TTT CCG CAG CTC GGT GAT CTC ATT CTC</td>
</tr>
<tr>
<td>Dec K111S internal primers</td>
<td>F: TTG ATC CTT GTC AAC AAG AGT ATC AGC AAA ATC AGT CCA GAG</td>
</tr>
<tr>
<td></td>
<td>R: CTC TGG ACT GAT TTT GCT GAT ACT GTT GAC AAG GAT CAA</td>
</tr>
<tr>
<td>Dec RE267A internal primers</td>
<td>F: GTT CCT CAT CTG GCT GCT CAC TCG TAC GAC AAC</td>
</tr>
<tr>
<td></td>
<td>R: GTT GTC CAA GTG GAG AGC AGC CAG ATG AGG AAC</td>
</tr>
<tr>
<td>Dec D204R internal primers</td>
<td>F: CGC ATC TCA GCA ACC AAT ACC AAT AAT GCG</td>
</tr>
<tr>
<td></td>
<td>R: CGC AGT TAT GTT GGT TCG TGA GAT GCG</td>
</tr>
<tr>
<td>Dec G246N internal primers</td>
<td>F: TTG TCT AAA CTG AAT TTG AGC TCC AAC AGC ATC</td>
</tr>
</tbody>
</table>
**Table 2.2 Primers used for generating SPARC expression plasmids.** Underlined sequences GCTAGC and GCGGCCGC correspond to restriction sites, *NheI* and *NotI*, incorporated into the forward and reverse primers respectively. The coloured triplets correspond to the mutated nucleotides.

<table>
<thead>
<tr>
<th>DNA Construct</th>
<th>Primer Sequence* (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>dSPARC FS-EC NHis</td>
<td>F: CGGAATTCGCTAGCCGATCTGTGCGAAACGATGAGC&lt;br&gt;R: ATCCCGGGCCGCTAACCAGGAGATGGGGCTG</td>
</tr>
<tr>
<td>dSPARC FS-EC NHis.myc</td>
<td>F: CGGAATTCGCTAGCCGATCTGTGCGAAACGATGAGC&lt;br&gt;R: ATCCCGGGCCGCTAACCAGGAGATGGGGCTG</td>
</tr>
<tr>
<td>dSPARC EC NHis</td>
<td>F: CGGAATTCGCTAGCCGAGAGCTGCGAGGGCGAG&lt;br&gt;R: ATCCCGGGCCGCTAACCAGGAGATGGGGCTG</td>
</tr>
<tr>
<td>dSPARC EC NHis.myc</td>
<td>F: CGGAATTCGCTAGCCGAGAGCTGCGAGGGCGAG&lt;br&gt;R: ATCCCGGGCCGCTAACCAGGAGATGGGGCTG</td>
</tr>
<tr>
<td>dSPARC FS-EC NHis glycosylation double mutant internal primers</td>
<td>F: ACCACCACCAGGAGACCTGG (Mutation in FS domain)&lt;br&gt;R: CCAGGTCTCCAGGAGGAGCATCAGGAGTGGGGCTG&lt;br&gt;R: ACCCGAGTTTCGCTCCAGGAGATGTCAGTC</td>
</tr>
<tr>
<td>hSPARC FS-EC His.myc</td>
<td>F: TTCGCTAGCC AATCCCTGCCGAAGCCAC CACTGC&lt;br&gt;R: ATCCCGGGCCGCTAACCAGGAGATGGGGCTG</td>
</tr>
<tr>
<td>hSPARC EC His.myc</td>
<td>F: TTCGCTAGCCCCCCCTTGCCGACTCTGAG&lt;br&gt;R: ATCCCGGGCCGCTTACCAAGATCCTCTGCGAT</td>
</tr>
<tr>
<td>cSPARC FS-EC NHis</td>
<td>F: CGGAATTCGCTAGCCGAGAGCTGCGAGGGCGAG&lt;br&gt;R: ATCCCGGGCCGCTAACCAGGAGATGGGGCTG</td>
</tr>
<tr>
<td>cSPARC EC NHis</td>
<td>F: CGGAATTCGCTAGCCGAGAGCTGCGAGGGCGAG&lt;br&gt;R: ATCCCGGGCCGCTAACCAGGAGATGGGGCTG</td>
</tr>
<tr>
<td>cSPARC FS-EC NHis.myc</td>
<td>F: CGGAATTCGCTAGCCGAGAGCTGCGAGGGCGAG&lt;br&gt;R: ATCCCGGGCCGCTAACCAGGAGATGGGGCTG</td>
</tr>
<tr>
<td>cSPARC EC NHis.myc</td>
<td>F: CGGAATTCGCTAGCCGAGAGCTGCGAGGGCGAG&lt;br&gt;R: ATCCCGGGCCGCTAACCAGGAGATGGGGCTG</td>
</tr>
</tbody>
</table>
2.2.1.3 Agarose Gel Electrophoresis

1% (w/v) agarose was dissolved in 1x Tris-acetate EDTA (TAE) buffer by boiling. Cooled, molten agarose was supplemented with 1x SYBR® Safe DNA gel stain and poured into a gel chamber. DNA samples containing 1x gel loading dye (2.5% (w/v) Ficoll®-400, 3.3 mM Tris-HCl, pH 8.0, 11 mM EDTA, 0.017% (w/v) SDS, 0.015% (w/v) bromophenol blue) were loaded onto the gel and run at 100 V for 50 minutes. All samples were compared to 100bp or 1 kb DNA ladder and/or Hyperladder I™, depending on the estimated size of the products. SYBR® Safe-stained DNA was visualised using a DNA gel documentation system.

2.2.1.4 DNA Purification by Gel Extraction

The amplified PCR products were purified to remove excess primers and nucleotides. This was done by separating the products on 1% (w/v) agarose gels. Bands of interest were excised and the DNA purified using a QIAquick gel extraction kit according to the manufacturer's protocol. The DNA was quantified using a Nanodrop ND-1000 spectrophotometer.

2.2.1.5 Restriction Digests

Double digests of the amplicons and the vectors were performed using NheI/NotI at 37 °C for 1-1.5 hours. The digestion mixtures contained DNA, NEB reaction buffer 4, 0.1 μg/ml BSA and 10 U restriction enzymes in a 50 μl reaction volume. 1 μl of calf intestinal alkaline phosphatase (CIP) was added directly to the reaction mixture containing the cloning vector 30 minutes prior to the end of the reaction. The digested products were purified by gel extraction using a gel extraction kit (Qiagen) following the manufacturer's protocol.
2.2.1.6 DNA Ligation

The purified digestion products were ligated using a Rapid T4 DNA ligation kit (Roche) according to manufacturer's protocol. 75 ng vector was mixed with a 3-fold molar excess of digested insert, in addition to 1x T4 DNA ligase buffer and 1 U T4 DNA ligase in a 10 µl reaction volume. The mixtures were incubated at 16 °C for one hour.

2.2.1.7 DNA Transformation

Frozen Subcloning Efficiency™ DH5α™ Competent Cells were thawed on ice and 6 µl of each ligation mixture was added to 50 µl cells and left on ice for 30 minutes. The cells were heat-shocked at 42 °C for 45 seconds, then returned to ice for 2 minutes before being plated out on LB agar containing 100 µg/ml ampicillin and incubated overnight at 37 °C.

2.2.1.8 Screening for transformed DNA

Colonies were picked and grown overnight at 37 °C in 5 ml LB broth containing 100 µg/ml ampicillin. Plasmid DNA was extracted using Qiagen DNA Miniprep kit according to manufacturer's protocol. The correct ligation products were ascertained by performing 'diagnostic' NotI/NheI double digests. Insert-containing vectors were subsequently sequenced (Eurofins MWG).

2.2.2 Expression vectors

The eukaryotic episomal expression vector pCEP-Pu, with an N-terminal His6-tag or an N-terminal His.myc tag or a C-terminal His6-tag, was used for expression in mammalian 293 c18 cells (Figure 2.2). The 9.6 kb vector is based on Invitrogen's pCEP-Sh3 vector, with modifications made by Rupert Timpl's lab in München, Patrik Maurer's lab in Köln,
and in our lab (Kohfeldt et al., 1997). These modifications include introduction of puromycin resistance gene, BM-40 secretion sequence cloned into the multiple cloning site and addition of fused and cleavable tags at the N- and C- termini. The vector also contains an EBNA-1 gene and a replication origin (oriP) site which mediate extrachromosomal replication and partitioning of the episome during host cell division. Figure 2.2 shows a vector map indicating the features of pCEP-Pu.

The sequence-verified inserts were cloned into one of the modified pCEP-Pu vectors using NheI and NotI. For C-terminal tagged proteins, a vector-derived APLA sequence remains at the N terminus of the secreted recombinant protein following cleavage of the BM-40 sequence signal, and an AAAHHHHHHH sequence is added at the C terminus. For N-terminal tagged proteins, the N-terminus of the mature protein is APLV-His6-ALA followed by the insert.

Figure 2.2 Vector map indicating features of pCEP-Pu. This plasmid (~ 9.6 kb) is used for episomal expression in HEK293-EBNA cells. It contains the BM-40 secretion signal which allows secretion of the protein from the cell, a multiple cloning site (MCS), cytomegalovirus promoter (pCMV) for transcription of recombinant genes inserted into the multiple cloning site, an EBNA-1 gene and a replication origin (OriP) site which facilitate extrachromosomal replication of the plasmid during host cell division. Marker for puromycin resistance gene allows puromycin selection in transfected mammalian cells.
2.2.3 Mammalian Cell Culture

2.2.3.1 Transfection
Expression vectors were transfected into HEK 293-EBNA cells cultured in DMEM supplemented with 10% (v/v) foetal bovine serum (FBS), penicillin/streptomycin, 2 mM L-glutamine and 25 µg/ml geneticin. Transfection was performed in T25 flasks using Fugene reagent according to the manufacturer's protocol (Roche). Cells containing the plasmid were selected 24 hours later using 1 µg/ml puromycin. Transfectants were maintained under continuous selection with puromycin. All solutions were warmed to 37 °C prior to use.

2.2.3.2 Growth and Maintenance of Cells
Cells were grown at 37 °C in a humid atmosphere of 5 % carbon dioxide. The adherent cells were passaged by decanting the growth medium, then trypsinising with 2 ml of trypsin:EDTA solution for 2-3 minutes at 37 °C to detach them from the flask. Resistant cells, when greater than 80% confluent, were expanded progressively into T75, T175 and finally into Corning HYPERFlasks. Once the cells in HYPERflasks had reached confluence, the selection medium was discarded; the cells were washed twice with PBS and incubated with serum-free medium to eliminate serum proteins from large-scale protein preparations. Serum-free medium containing soluble proteins (conditioned medium) was collected after one week. Flasks were refilled with serum-free medium and up to four exchanges were carried out. The media were centrifuged at 3,500 rpm for 10 minutes, filtered to remove cell debris and stored at 4 °C.

2.2.4 Protein Purification
Milligram (mg) quantities of pure and stable protein are required for structural and functional studies. This section describes the purification strategies used to obtain such
quantities of various constructs of decorin and SPARC proteins, and the bacterial expression and purification of PNGaseF - MBP enzyme.

2.2.4.1 Immobilised Metal Ion Chromatography

Conditioned serum-free medium (pH adjusted to 7.5) was loaded onto a 5 ml HisTrap™ column equilibrated with PBS (10 mM sodium phosphate pH 7.4, 2.68 mM KCl, 140 mM NaCl), using an ÄKTAPurifier. Proteins were eluted using 300 mM imidazole in PBS. The absorbance at 280 nm was monitored constantly. Selected 1 ml fractions were analysed by SDS-PAGE and pooled as appropriate.

All SPARC protein samples were dialysed overnight at 4 °C against 2 L 20 mM HEPES pH 7.5, 150 mM NaCl, 2 mM CaCl₂. The protein concentration was then determined using a spectrophotometer.

2.2.4.2 Size Exclusion Chromatography (SEC)

SEC was carried out in order to further purify proteins. A Superdex 200 10/300 column or a Superdex 200 16/60 column (GE Healthcare) was equilibrated with running buffer: 20 mM HEPES pH 7.5, 150 mM NaCl, 2 mM CaCl₂ for SPARC proteins and Tris-buffered saline (TBS) (20 mM Tris pH 7.5, 150 mM NaCl) or PBS for decorin proteins. 530 μl or 5 ml of each protein sample was injected into the ÄKTA system per run, depending on the loop used (500 μl for the smaller 10/30 column and 5 ml for the larger 16/60 column). Elution was monitored by UV absorbance at 280 nm. 1 or 2 ml fractions were collected. 10 μl of each fraction were run on SDS-PAGE to identify the eluted proteins. Protein concentrations were quantified using a spectrophotometer at 280 nm using extinction coefficients calculated from the primary amino acid sequence.

2.2.4.3 Bacterial Expression and Purification of PNGase F-MBP

E. coli BL21 cells transformed with PNGase F-MBP (peptide-N-glycosidase F-maltose binding protein) construct (glycerol stock provided by Prof Erhard Hohenester) were
streaked on to an LB and Agar plate containing 100 µg/ml ampicillin and incubated overnight at 37 °C. A colony was picked and used to inoculate 100 ml LB culture containing 100 µg/ml ampicillin and grown overnight at 250 rpm at 37 °C. 50 ml of this starter culture was added to 1 L of LB broth and the culture incubated again at 37 °C. At OD$_{600nm}$ = 0.6, as the bacteria were in logarithmic growth phase, the culture was induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and left overnight at 30 °C in a shaker at 250 rpm. The next day, cells were harvested by centrifugation at 5000 $g$ for 15 minutes at 4 °C. The pellet was stored at -80 °C until purification.

Frozen cell pellets were slowly thawed on ice and resuspended in 30 ml of PBS (Buffer A) with 1 mg/ml lysozyme, 0.1% Triton X-100 and 0.1 mM PMSF (phenylmethanesulphonylfluoride). The sample was incubated on ice for 10 minutes. The lysate was sonicated on ice for 10 cycles of alternating 30 seconds ‘on’ followed by 30 second rest to shear the DNA. The sample was centrifuged at 30,000 $g$ for 20 minutes at 4 °C to pellet the cell debris and remove unbroken cells. 1 ml of 1 mg/ml protamine sulphate was added to the supernatant to precipitate the DNA and the sample was centrifuged again at 30,000 $g$ for 20 minutes at 4 °C.

The cleared sample was loaded onto a 5 ml amylose resin column (NEB) washed with column buffer (20 mM Tris pH 7.5, 200 mM NaCl, 1 mM EDTA). The protein was eluted using 10 mM maltose, 20 mM Tris-HCl pH 7.5, 200 mM NaCl, 1 mM EDTA. Selected fractions were analysed by SDS-PAGE and pooled as appropriate.

The protein was further purified by size exclusion chromatography using a Superdex S200 column on an Äkta FPLC equilibrated with running buffer (40 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM EDTA). The volume of the protein sample was reduced to ~500 µl using a 10 kDa molecular weight cut-off (MWCO) Vivaspin tube and then injected into the ÄKTA system. Elution was monitored by UV absorbance at 280 nm. Fractions of 1 ml were collected. Peak fractions were pooled and concentrated to 2 mg/ml. Glycerol was added in a 1:1 ratio. The samples were then aliquoted, flash frozen in liquid nitrogen and stored at -80 °C.
2.2.5 Protein Methods

2.2.5.1 SDS-PAGE

SDS-PAGE (Laemmli, 1970) was carried out using a Hoefer™ SE-250 gel tank. Separating gels were either 10% (w/v) or 15% (w/v) acrylamide:bis-acrylamide, in addition to 0.375 M Tris, pH 8.8, 0.1% (w/v) SDS, 0.04 % (w/v) APS and 0.2% (w/v) TEMED. Stacking gels were 4% acrylamide: bis-acrylamide plus 0.125 M Tris, pH 6.8, 0.1% (w/v) SDS, 0.04% APS and 0.2% (w/v) TEMED. 10 µl samples containing 1x SDS-PAGE loading buffer (60 mM Tris pH 6.8, 25% glycerol, 2% SDS, 0.7 M β-mercaptothanol, 0.1% bromphenol blue) were boiled at 95 °C for 7 minutes and electrophoresed at 180 V for 1 hour in 1x SDS-PAGE running buffer (125 mM Tris base pH 8.3, 0.96 M glycine, 0.1% SDS). 10 µl of pre-stained broad range protein ladder (10-250 kDa) was used.

After separation, gels were stained with either Quick Coomasie stain or InstantBlue™ gel stain for at least 60 minutes at room temperature with gentle shaking. This was followed by de-staining with deionised water overnight at room temperature.

2.2.5.2 NuSep®

NuSep® Tris-HEPES 4 - 12% gels (Generon) were used for SDS-PAGE analysis of protein samples. Samples containing 1 x sample buffer (60 mM Tris pH 6.8, 25% glycerol, 2% SDS, 0.7 M β-mercaptothanol, 0.1% bromphenol blue) were boiled at 95 °C for 7 minutes. The gel cassettes were loaded onto the Hoefer SE-250 gel tank and immersed in 1 x NuSep Tris-HEPES-SDS running buffer (100 mM Tris pH 8.0, 100 mM HEPES, 0.1% SDS). The boiled samples were loaded and electrophoresed at 120 V for 1 hour. 10 µl of pre-stained broad range protein ladder (10-250 kDa) was used.

2.2.5.3 Deglycosylation

For diagnostic purposes, the N-linked glycan in SPARC proteins was removed by treatment with New England Biolabs (NEB) PNGase F treatment according to the manufacturer’s protocol. Samples were heat-denatured by adding 1 x glycoprotein
Denaturing buffer (0.5% SDS, 40 mM DTT) for 10 minutes at 100 °C prior to addition of PNGase F diluted in 1 x G7 buffer (50 mM sodium phosphate pH 7.5) and 1 x NP-40.

For large scale deglycosylation, the protein samples were incubated with PNGaseF-MBP (produced as described in section 2.2.4.3) overnight at room temperature in an enzyme : protein ratio of 1:5.

*Drosophila* SPARC fragments were treated with lectin in an attempt to remove remaining glycosylated forms following PNGaseF treatment. Lectin bead slurry, consisting of three types of lectin beads (agarose *Lens culinaris* agglutinin, agarose concanavalin A and wheat germ agglutinin) was used for this purpose. 2 ml of each type of beads were spun down at 800 rpm for 5 minutes and the supernatant discarded. The beads were then washed with double-distilled water followed by wash buffer (10 mM HEPES pH 7.5, 0.15 M NaCl, 0.1 mM CaCl₂, 0.01 mM MnCl₂) to remove sugar added to stabilise the lectin. The PNGaseF - treated protein sample was incubated with the slurry overnight on a bench-top rotator at room temperature. The supernatant from the beads was collected, which should contain deglycosylated protein. The lectin beads were loaded onto a gravity flow column. The remaining glycosylated protein bound to the beads were eluted using 10 ml elution buffer (10 mM HEPES pH 7.5, 0.15 M NaCl, 0.1 mM CaCl₂, 0.5 M N-acetyl-D-glucosamine, 0.2 M α-methyl mannoside, 0.2 M α-methyl glucoside).

### 2.2.5.4 Crystallisation Trials

Crystallisation trials were set up using various screens (see Chapter 5, tables 5 and 6). The protein was mixed 1:1 with well solution (100 nl + 100 nl) in 96-well sitting drop plates using a Mosquito robot and incubated at 20 °C.
2.2.6 Size Exclusion Chromatography- Multi Angle Light Scattering (SEC - MALS)

400 μl of wild-type decorin and each mutant sample, at a concentration of 3 mg/ml, were injected and separated on a Superdex 200 10/30 column which had been pre-equilibrated in 1 x TBS (20 mM Tris pH 7.5, 150 mM NaCl) at room temperature at a flow rate of 0.2 ml/min. Light scattering data were recorded on an in-line Wyatt miniDAWN light scattering (LS) detector which detects scattered light at three different angles. The light scattering detector was in series with an Optilab rEX refractive index detector. UV absorbance was recorded at 280 nm. A specific refractive index increment (dn/dc) value of 0.185 ml/g was used for protein component and 0.145 ml/g for the carbohydrate component of each glycoprotein (Wen et al., 1996). Each putative site for N-linked glycosylation was assumed to add a molecular mass of 2 kDa. The dn/dc value was calculated using the equation: (((0.185/100) x protein % + (0.145/100) x sugar %).

The mass of the polypeptide fraction of the glycoproteins was determined by the three-detector method described by (Wen et al., 1996) using an extinction coefficient of 24961 M⁻¹cm⁻¹ for the decorin protein. The equation used to calculate the polypeptide mass for a protein, Mp, containing carbohydrates is described by Wen et. al as:

\[ M_p = \frac{K_{RI}^2}{K_{LS} K_{UV}} \left( \frac{LS}{UV} \right) \frac{(RI)}{\varepsilon_p} \]

Where \((K_{RI}^2/K_{LS}K_{UV})\) is an instrument calibration constant, subscript \(p\) refers to the polypeptide component, \((LS)\) is the intensity of light-scattering signal, \((UV)\) is the signal from the UV absorbance detector, \((RI)\) is the refractive index signal and \(\varepsilon\) is the molar extinction coefficient i.e. the absorbance of 1 mg/ml glycoprotein at a 1 cm pathlength. All contributions from the carbohydrate are eliminated by selecting a wavelength where the carbohydrate does not absorb.

The Wyatt ASTRA (version 5.3.4.20) software was used for all MW calculations. The laser signals were aligned, and peaks selected to determine delays to correct for the
difference in flow path between the UV and LS measurements. The despiking level was set to ‘normal’. Appropriate pre-peak baselines were selected for the UV and the three LS signals. The peak boundaries were defined before checking the results section of the software.

2.2.7 Surface Plasmon Resonance (SPR)/ BIACORE

All SPR measurements were performed on a BIACORE 3000 instrument. The carboxymethylated dextran surface of a CM5 chip was activated by the injection of a mixture of equal volumes of 0.2 M N-ethyl-N-(diethylaminopropyl)carbodiimide (EDC) and 0.05 M N-hydroxysuccinimide (NHS). Collagen I or collagen IV (both at 2 mg/ml in 0.1 M acetic acid) was diluted to 100 µg/ml in 10 mM sodium citrate pH 3.2 and immobilised via its primary amine groups at 25 °C at a flow rate of 10 µl/min. An immobilisation level of ~2000 resonance units was obtained. Unreacted NHS groups were deactivated and non-covalently bound proteins removed using 1 M ethanolamine pH 8.5. A control flow cell, without any coupled protein, was treated identically.

The binding assay was performed by using recombinant proteins at different concentrations, ranging from 0 µM to 40 µM, diluted in running buffer (10 mM HEPES pH 7.4, 0.15 M NaCl, 0.05% Tween-20) using a two-fold dilution series. The running buffer was flowed at a rate of 20 µl/min. The surface was regenerated using 10 mM HEPES pH 7.4, 0.15 M NaCl, followed by 0.1 M NaHCO₃ pH 9.2, 2 M NaCl for 30 seconds each at a flow rate of 30 µl/minute. Some interactions were measured without regeneration of the sensor surfaces between runs. A titration series with long injection phases was used for the steady-state analysis. The sensorgrams were then evaluated with the BIAevaluation 3.0 software. Control sensorgrams were subtracted from the sensorgrams obtained on immobilised collagen to account for changes in refractive index and nonspecific binding. The response to the running buffer was defined as the baseline level, and all responses were expressed relative to this baseline.
2.2.8 Solid Phase Assays

Solid phase assays were performed in enzyme-linked immunosorbent assay (ELISA) format and used to test the interaction of proteins with collagen I and IV.

5 mg of rat tail type I collagen was dissolved in 0.1 M acetic acid. 10 µg/ml of this collagen solution (25 µg/ml for decorin assays) diluted in 50 mM Tris pH 8.5, 100 mM NaCl was immobilised onto Nunc Maxisorp 96-well immunoplates by incubation overnight at 4 °C. Wells were washed once with 1x PBS, then blocked with 150 µl of 0.04 mg/ml casein in 1x PBS containing 0.05% Tween-20 for two hours at room temperature. Proteins were diluted in incubation buffer (0.04 mg/ml κ-casein in 1x PBS with 0.05% Tween-20). 50 µl of proteins ranging from 0 µM to 40 µM were added to each well and incubated at room temperature for 3 hours. The wells were washed six times with 150 µl/well of incubation buffer. His.myc-tagged proteins were detected first by adding mouse anti-myc antibody (1:1000 dilution) for 1 hour at room temperature followed by sheep-anti-mouse-HRP antibody (1:1000 dilution) again for an 1 hour. Decorin proteins were incubated for 1 hour with 50 µl/well of goat anti-mouse decorin antibody (1:500 dilution in incubation buffer). After six washes with incubation buffer, the wells were incubated for 1 hour with 50 µl/well of horseradish peroxidase-conjugated rabbit anti-goat IgG antibody (1:1000 dilution in incubation buffer). In another set of experiments, only anti-His-tag HRP conjugated antibody was used. Following three washes with incubation buffer and three washes with PBS (150 µl/well), the assay was developed using 75 µl/well o-phenylenediamine dihydrochloride (OPD) substrate for 5 minutes and the reaction stopped using 50 µl/well of 3 M H2SO4. Absorbance was measured at 492 nm with an ELISA microplate reader.

Data were fitted using SigmaPlot to an equation for one-site saturable binding plus a linearly increasing background of nonspecific binding as shown in the equation below:

\[
y = \frac{B_{\text{max}} \cdot x}{K_{0.5} + x} + a
\]
Where y is the OD\textsubscript{492} reading, x is the ligand concentration, B\textsubscript{max} is maximum saturable binding, a is the background binding of antibodies to collagen and the blocked plastic surface and K\textsubscript{0.5} is the ligand concentration for half-maximal binding.

2.2.9 Biotinylation

Wild-type and mutant decorin and dSPARC (control) were labelled with biotin as follows. The samples were first dialysed against PBS, then incubated on ice with a 20-fold molar excess of EZ-Link Sulfo-NHS-LC biotin for 2 h. Concentrated Tris/HCl buffer, pH 7.5, was then added to 50 mM final concentration and the samples incubated for 1 h. The samples were then dialysed overnight against PBS to remove excess non-reacted biotin and stored at 4 °C.

2.2.10 Biotin - Avidin Assays

Nunc Maxisorp 96-well microtitre plates were coated with 50 μl of 25 μg/ml collagen I in PBS, overnight at 4 °C. The wells were then washed once with PBS and blocked with 150 μl/well of 10 mg/ml casein in 1x PBS containing 0.05% Tween-20 for two hours at room temperature. Proteins were serially diluted in incubation buffer (0.04 mg/ml casein in 1x PBS with 0.05% Tween-20). The wells were washed six times with incubation buffer (150 μl/well). Different concentrations of wild-type and decorin and glycosylation mutants ranging from 0 μM to 10 μM were added to each well (50 μl/well) and incubated at room temperature for 3 hours. Following three washes with incubation buffer and three with PBS, biotinylated proteins were detected using 50 μl/well avidin peroxidase (1:1000 dilution). The assay was developed using 75 μl/well o-phenylenediamine dihydrochloride (OPD) substrate for 5 minutes and the reaction stopped using 50 μl/well of 3 M H\textsubscript{2}SO\textsubscript{4}. Absorbance was measured at 492 nm with an ELISA microplate reader. The apparent dissociation constants were obtained by fitting the data with Equation 2 using SigmaPlot 12.0,
\[ y = \frac{B_{\text{max}} \cdot x}{K_{0.5} + x} + nx + a \]

where \( y \) is the OD\textsubscript{492} reading, \( x \) is the protein concentration, \( K_{0.5} \) is the dissociation constant, \( B_{\text{max}} \) is the maximum of specific binding, and \( nx + a \) is a linear term to account for non-specific binding to collagen and the blocked plastic surface.

### 2.2.11 Competition assays

Biotinylated wild-type decorin (1 \( \mu \)M) was pre-incubated with increasing concentrations of non-biotinylated wild-type decorin (up to 20 \( \mu \)M) at room temperature and the mixture added to collagen I-coated plates. The assay was developed as described above.

### 2.2.12 Collagen fibrillogenesis assay

All samples were kept on ice to avoid fibril formation prior to assaying. A 1.05 mg/ml solution of mouse tendon collagen I isolated from tendon, treated with pepsin and precipitated using NaCl (provided by Prof Hans Peter Bächinger), in 50 mM acetic acid was neutralised by diluting it 33 times with 150 mM sodium phosphate, 150 mM NaCl, pH 7.8 and immediately placed in a plastic cuvette in a Shimadzu UV-2501PC spectrophotometer that was pre-warmed to 37 °C. Collagen fibril formation was monitored as absorbance at 400 nm at 3 minute intervals over 100 minutes. The inhibition assays contained wild-type and decorin mutants at a concentration of 50 \( \mu \)g/ml (1.38 \( \mu \)M). Turbidity curves were analysed in terms of the maximum turbidity and the rate of turbidity change observed.
2.2.13 Differential Scanning Calorimetry (DSC)

DSC experiments were performed to assess the stability of the different decorin mutants, using an N-DSCIII Differential Scanning Calorimeter. Decorin samples at 3 mg/ml were dialysed against PBS. 1 ml aliquots of sample and buffer were de-gassed under vacuum for 15 minutes. 300 µl each of sample and buffer were loaded onto the sample and reference cells respectively. Following a pre-scan equilibration period of 10 minutes, the sample was heated linearly from 5 - 20 °C at 1 °C per minute followed by cooling over the same temperature range. The scans were monitored and increased as necessary until a stable baseline was obtained. For the actual denaturation runs, the temperature scan ranged from 5 - 65 °C at a heating rate of 1 °C per minute. Four heating scans between 5 - 65 °C were recorded for each sample. A small excess constant pressure of 3 atm was applied to the sample and reference cell to allow scanning to high temperatures without the sample boiling or forming bubbles. The temperature corresponding to maximum Cp was determined as the apparent melting temperature (Tm) i.e. the temperature at which 50% of the protein is unfolded.

Data are quoted per mole of monomeric protein (36.6 kDa). Data analysis was performed by working out a conversion factor for power/heating rate to express the results in terms of heat capacity, Cp, having the units, kJ/ (K · mole).

2.2.14 Differential Scanning Fluorimetry (DSF)

DSF experiments were performed to assess the stability of WT and decorin mutants, using a Stratagene Mx3005P qPCR System with an excitation of 492 nm and emission of 610 nm. Each sample was prepared in a total volume of 20 µl containing protein solutions at 5 µM final concentration and 1:500 SYPRO orange. The PCR plates were sealed with the optical foil seal. The samples were heated at a rate of 1 °C/min, from 25 °C to 95°C, and the Tm values were calculated from the inflection points of the transition curves.
2.2.15 Analytical Ultracentrifugation (AUC)

Sedimentation velocity (SV) experiment was performed in Professor Stephen Perkins' lab at University College London, using a Beckman XL-1 analytical centrifuge at a rotor speed of 40,000 rpm and a temperature of 20 °C. Wild-type decorin was studied in a concentration series ranging from 0.075 mg/ml to 3.6 mg/ml. An eight-hole AnTi50 rotor was used containing double-sector cells with 12 mm column heights. Wild-type decorin was extensively dialysed against PBS. The sample cells were filled with 400 µl decorin and the reference cells with 450 µl sample buffer (PBS).

**Wild-type decorin composition:** Decorin has an absorption coefficient at 280 nm of 7.15 cm⁻¹ (1%, 1 cm path length). Partial specific volume \( \bar{\rho} \) for wild-type decorin was taken to be 0.7289 ml/g and MW of 44,310 Da, calculated using a programme called SLUV which calculates the \( \bar{\rho} \)-bar and MW from the amino acid and carbohydrate composition of the protein. The N-linked oligosaccharides on decorin were assumed to be bi-antennary, consisting of a core of GlcNAc₂Man₃ and two NeuNAc.Gal.GlcNAc antennae.

**Data analysis:** The SV data were analysed with SEDFIT software (Version 14.0) using the continuous c(s) analysis method to determine the sedimentation coefficients \( s_{20,w} \) (\( s_{20,w} \) refers to sedimentation coefficient corrected for water). The buffer density and viscosity values were taken to be the theoretical values calculated by SEDNTERP - in this case, for PBS (10 mM Na₂PO₄, 140mM NaCl, 2.68mM KCl), 1.00543 and 0.01020 respectively. SEDFIT fits the sedimentation boundaries using the Lamm equation (Schuck, 2000) which assumes that all species in solution have the same frictional ratio, \( f/f_0 \) in each fit. The final c(s) distribution plot used a fixed resolution of 200 sedimentation coefficients between 0.5 and 15 S. All size distributions were determined with a confidence level of \( p = 0.95 \). Each fit was optimised by floating \( f/f_0 \), baseline and meniscus and keeping the partial specific volume, solvent density and the cell volume fixed until the lowest root-mean-square deviation (rmsd) was found and the fit looked good on visual inspection. The c(s) integration function (area under the curve) was used to find the percentage of different species in the total loading concentration.
Theoretical sedimentation coefficients $s_{20,w}$ for decorin monomer and dimer were calculated using the HYDROPRO programme (Garcia De La Torre et al., 2000a). This programme uses the atomic coordinates of the protein to calculate $s_{20,w}$ by converting the protein to a hollow shell model consisting of spheres only. Mouse decorin monomer and dimer were modelled based on the crystal structure of bovine decorin (PDB code 1xku) with bi-antennary carbohydrates added perpendicular to the protein domain at all four of the predicted glycosylation sites. The hydration shell was represented using the default atomic-element radius (AER) value of 0.31 nm for all atoms.

**Estimation of $K_d$ for protein in monomer-dimer equilibrium:**

An equation that relates the dimer fraction to the protein concentration was derived as described by Benfield et al. (2011) and modified as follows:

For a system in monomer-dimer equilibrium,

$$M + M \leftrightarrow D \quad \text{Eq.}(1)$$

where $M$ and $D$ are monomers and dimers respectively. The equilibrium dissociation constant $K_d$ is defined by:

$$K_d = \frac{[M]^2}{[D]} \quad \text{Eq.}(2)$$

where $[M]$ and $[D]$ represent molar concentrations (mol/L) of monomer and dimer respectively. The total protein concentration $[M]_T$ in equivalent monomers is:

$$[M]_T = [M] + 2[D] \quad \text{Eq.}(3)$$

This can be re-arranged as:

$$[D] = \frac{[M]_T - [M]}{2} \quad \text{Eq.}(4)$$
Substituting (4) into (2):

\[ K_d = \frac{2[M]^2}{[M]_T - [M]} \quad \text{Eq. (5)} \]

Solving Eq. (5) For \([M]\)

\[ [M] = \frac{-K_d + \sqrt{K_d^2 + 8K_d[M]_T}}{4} \quad \text{Eq. (6)} \]

% monomer = \(\frac{[M]}{[M]_T} \cdot 100\)

% dimer = \(\frac{2[D]}{[M]_T} \cdot 100\)

It then follows that,

% monomer = \(\frac{[M]}{[M]_T} = \frac{\sqrt{K_d^2 + 8K_d[M]_T} - K_d}{4[M]_T} \quad \text{Eq. (7)} \)

Hence,

% dimer = \(\left(1 - \frac{\sqrt{K_d^2 + 8K_d[M]_T} - K_d}{4[M]_T}\right) \cdot 100 \quad \text{Eq. (8)} \)

Plotting % dimer against \([M]_T\) gives us the \(K_d\) which is the only fittable parameter in the above equation.
Chapter 3: Results

Analysis of decorin dimerisation
3. Results

The crystal structure of decorin revealed a seemingly tight dimeric structure in which the concave faces of two monomers form the dimer interface, thus making the presumed collagen binding site inaccessible. The physiological relevance of the decorin dimer remains controversial and it has been claimed that biologically active decorin is a monomer.

The main aim of the study was to resolve the controversy about the oligomeric state of decorin by designing a number of decorin mutants to disrupt the crystallographic dimer. I used size exclusion chromatography with multi-angle laser light scattering and analytical ultracentrifugation was used to determine the oligomeric states of wild-type and mutant decorin. The stabilities of wild-type and mutant decorin were quantified by differential scanning calorimetry and differential scanning fluorimetry. A further aim was to identify whether decorin binds collagen as a monomer or a dimer, directly using solid-phase assay and indirectly by measuring the inhibition of collagen fibrillogenesis.

3.1 Mouse decorin construct design

The amino acid identity of decorin across species is about 80% (Figure 3.1). A wild-type (WT) mouse decorin expression construct (referred to as decorin) corresponding to the ordered residues in the crystal structure of bovine decorin core protein was created. This construct consists of residues 45-354 of the protein (UniProt P28654, residue numbering scheme includes the signal peptide). A His$_6$-tag was added at the C-terminus. Since the C-terminal cap is not involved in the dimer interface, the tag is not expected to affect dimerisation.
Figure 3.1 Amino acid sequence alignment of bovine, mouse and human decorin. Conserved residues are shaded in grey. The engineered glycosylation sites are shaded pink. Putative N-glycosylation sites are in pink. Various other independent mutations are coloured cyan. Putative collagen binding residues (Kalamajski et al., 2007, Kresse et al., 1997) are shaded yellow. Cysteines are in red. The residue numbering above the alignment corresponds to that of bovine decorin. Sequences from the N-terminus to LRR12 are shown. The LRR consensus sequences (L..L..L..N.L) are indicated below the alignment.
3.2 Purification of wild-type decorin

The WT decorin construct was used for the episomal transfection of HEK293 cells since the post-translational modifications produced by these cells are similar to those of native mammalian glycoproteins. WT decorin protein was purified to homogeneity from serum-free medium by nickel affinity chromatography, followed by SEC (Figure 3.2A). The yield from 2 L of medium was 15.5 mg of protein. The SEC fractions were analysed by SDS-PAGE (Figure 3.2B). All purified decorin gave two bands on reducing SDS-PAGE - the lower band corresponding to glycosylated decorin at ~45 kDa (the predicted molecular mass from the amino acid sequence, 36.2 kDa). Due to the presence of N-linked glycosylation sites in decorin, the purified proteins migrate as diffuse bands at higher molecular mass on SDS-PAGE. The band at ~ 80 kDa, is unlikely to be a disulphide-linked dimer as the gel has been run under reducing conditions. It is possibly a gel artefact since the protein elutes as a single species in SEC and it is not seen on all gels (e.g. Figure 3.19).

Figure 3.2 Purification of WT decorin. A. The elution profile of purified WT Dec CHis on a Superdex S200 16/60 column is shown. The arrows show the peak elution volumes of molecular mass standards in kDa. Protein samples were eluted in TBS (20 mM Tris pH 7.5, 150 mM NaCl). B. Reducing SDS-PAGE analysis of Dec CHis purification (Coomassie Blue stain). Peak fractions correspond to the SEC peak in A. The positions of selected molecular mass markers are shown on the left.
3.3 Purification of engineered glycosylation mutants

In order to design mouse decorin constructs that are incompatible with the dimer structure presented by Scott et. al (2004), I designed a number of mutants including two which had glycosylation sites engineered into the dimer interface (Figure 3.1). Glu61 (Gln37 in bovine decorin) in the N-terminal cap (Q61N mutant) and Tyr130 (Tyr106 in bovine decorin) in LRR4 (Y130N mutant) were mutated to Asn, which created consensus sites for glycosylation at these locations. Based on the NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/), the asparagines at positions 61 and 130 have predicted glycosylation potentials of 0.61 and 0.72 respectively (threshold = 0.5), suggesting that these engineered glycosylation sites have a high probability of being modified.

These mutants were secreted from HEK293 cells as efficiently as wild-type decorin, indicating that the proteins were folded correctly. The yields from 2 L of medium were 33.4 mg of Q61N CHis and 23.5 mg of Y130N CHis. The SEC profiles of the two mutants overlaid with that of wild-type decorin are shown in Figure 3.3. The two mutants with engineered glycosylation sites elute later compared with the wild-type protein, suggesting that they are smaller in size, i.e. possibly monomeric in solution. The presence of extra peaks in the Y130N mutant SEC profile suggests protein aggregation.
Figure 3.3 SEC analysis of decorin glycosylation mutants. The elution profiles of purified WT Dec CHis, Q61N and Y130N mutants on a Superdex S200 16/60 column are shown. The arrows show the peak elution volumes of molecular weight standards in kDa. Protein samples were eluted in TBS (20 mM Tris pH 7.5, 150 mM NaCl).

SDS-PAGE analysis of wild-type decorin under reducing conditions shows a band at ~45 kDa, consistent with a calculated molecular mass, 36.2 kDa plus four N-linked glycosylation sites in the mouse decorin sequence (Figure 3.4). The two glycosylation mutants (Q61N and Y130N) appear fuzzier and migrate at a higher molecular mass, most likely due to the addition of the extra glycan. The shift in molecular mass of the Y130N is less clear, either due to the absence of an extra glycan or because the glycan is not detectable by SDS-PAGE. Since I already had one mutant with a clearly visible extra glycan (Q61N mutant), I did not further investigate the glycosylation profile of the Y130N mutant. Removal of the N-linked glycans by PNGase F digestion following denaturation of the samples generated identical sharp bands of lower molecular weight (~30 kDa) corresponding to the core deglycosylated proteins (Figure 3.4).
The purified proteins (WT Dec CHis, Q61N and Y130N mutants) were left undigested or treated with PNGase F following denaturation and then analysed by SDS-PAGE. The positions of selected molecular weight markers are indicated on the left.

3.4 Purification of other mutants

Two other constructs were made to introduce mutations into the decorin dimer interface: a charge reversal mutation, R151E in LRR5 (Arg127 in bovine decorin), which reverses a charge in the extensive hydrogen-bonding network at the dimer interface; and a triple mutation in the N-terminal cap, Y51A/R52A/Q54A (Phe27, Arg28 and Gln30 in bovine decorin), which removes three side-chains involved in the dimer interface. As controls, I designed two mutants on the outer surface of decorin (E156K and K159E) which should leave the dimer interface unaffected. All mutants were highly expressed and the yields following affinity purification and SEC from 2 L of medium were as follows: 11.8 mg of R151E, 5.5 mg of Y51A/R52A/Q54A, 11.9 mg of E156K and 15.1 mg of K159E (Figure 3.5).
Figure 3.5 SDS-PAGE of mouse decorin mutants under reducing conditions. The purified proteins (R151E, Y51A/R52A/Q54A, E156K and K159E mutants) were analysed by SDS-PAGE. The positions of selected molecular weight markers are indicated on the left.

Figure 3.6 shows the positions of the engineered mutations in mouse decorin mapped onto the crystal structure of bovine decorin.
Figure 3.6 Positions of the engineered mutations in mouse decorin mapped onto the crystal structure of bovine decorin. The independently mutated residues on the dimer interface are each colour-coded and the control mutants located outside the dimer interface (E156K and K159E) are coloured orange. The N- and C- termini are labelled on one subunit. Drawn from PDB entry 1XKU (Scott et al., 2004).
3.5 Oligomeric states of wild-type and mutant decorin

Size exclusion chromatography with multi-angle laser light scattering (SEC-MALS) was used to determine oligomeric states of wild-type mouse decorin and its dimer interface mutants. This technique allows the determination of absolute molecular masses of proteins in a reference-free manner. In conventional SEC, the elution position is affected by the shape (hydrodynamic radius) of the protein. So if the protein is elongated it will elute earlier from the column, hence appearing larger than its true molecular weight or if it is ‘sticky’ i.e. it interacts with the column matrix, it will elute later. However, molecular masses determined by light scattering are independent of the shape of the protein and the elution volume. Another advantage of SEC - MALS is that in the case of glycoproteins, it allows the de-convolution of the effect of the carbohydrates from that of the core protein (Wen et al., 1996).

Wild-type mouse decorin injected at a concentration of 3 mg/ml (83 μM) eluted as an asymmetric peak with a pronounced tail (Figure. 3.7). The molecular mass of the polypeptide fraction of the protein was determined to be 64.4 kDa (Table 3.1). This value is much closer to the calculated mass of a dimer (72.4 kDa) than that of a monomer (36.2 kDa). The average molecular mass of the dimer is slightly lower than expected because there is some dissociation into monomers, giving rise to an asymmetric peak. The molecular mass of the glycoconjugate (i.e. protein plus carbohydrate modifications) was determined to be 83.5 kDa, closely matching the reported mass of 84.6 kDa for dimeric bovine decorin core glycoprotein (Scott et al., 2003).

As expected, the control mutants, E156K AND K159E, of mouse decorin were dimeric with molecular weights of 65 and 60 kDa respectively for the core proteins without modifications (Figure 3.7, Table 3.1). The Y51A/R52A/Q54A and R151E mutants were also dimeric, indicating that fairly dramatic amino acid substitutions do not disrupt the dimer. For some of the samples, Y51A/R52A/Q54A mutant in particular, peak broadening was observed due to polydispersity, resulting in the molecular masses not being constant across the peaks. The elution profiles of Q61N and Y130N mutants indicate that these are pure monomers as symmetric peaks were detected by SEC-MALS,
with molecular masses consistent with those calculated from their primary sequences (Fig. 3.7, Table 3.1). It can be inferred that the addition of the extra glycan in the Q61N mutant resulted in the disruption of the decorin dimer. The same cannot be inferred categorically for the Y130 mutant, since the additional glycan modification is not evident in SDS-PAGE (Figure 3.4). However, our aim was only to obtain monomeric decorin mutants, hence we did not investigate the modifications in the Y130N mutant further. Thus, we were able to disrupt the decorin dimer by engineering glycosylation sites into the dimer interface.

Table 3.1 Molecular masses of decorin and its mutants as determined by SEC-MALS. The relative errors of the experimentally determined masses are <5%.

<table>
<thead>
<tr>
<th>Protein(s)</th>
<th>Calculated molecular mass (kDa)</th>
<th>N-linked glycosylation sites</th>
<th>Peak elution volume (ml)</th>
<th>Experimental mass of glycoprotein (kDa)a</th>
<th>Experimental mass of polypeptide fraction (kDa)b</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>36.2</td>
<td>4</td>
<td>14.5</td>
<td>83.5</td>
<td>64.4</td>
</tr>
<tr>
<td>Y51A/R52A/Q54A</td>
<td>36.0</td>
<td>4</td>
<td>14.1</td>
<td>85.3</td>
<td>65.3</td>
</tr>
<tr>
<td>R151E</td>
<td>36.2</td>
<td>4</td>
<td>14.1</td>
<td>88.4</td>
<td>68.4</td>
</tr>
<tr>
<td>E156K</td>
<td>36.2</td>
<td>4</td>
<td>14.5</td>
<td>79.7</td>
<td>65.0</td>
</tr>
<tr>
<td>K159E</td>
<td>36.2</td>
<td>4</td>
<td>14.4</td>
<td>78.9</td>
<td>60.3</td>
</tr>
<tr>
<td>Q61N</td>
<td>36.2</td>
<td>5</td>
<td>15.5</td>
<td>52.1</td>
<td>37.6</td>
</tr>
<tr>
<td>Y130N</td>
<td>36.2</td>
<td>5</td>
<td>15.1</td>
<td>51.5</td>
<td>36.7</td>
</tr>
</tbody>
</table>

aDerived from the refractive index and light scattering signals.

bDerived from the absorbance, refractive index and light scattering signals (Wen et al., 1996).
Figure 3.7 SEC-MALS chromatograms of wild-type and mutant decorin. Proteins were injected at 3 mg/ml (83 µM) onto a Superdex S200 column. Differential refractive index (ΔRI) is a measure of protein concentration (solid lines; left y-axis). Molecular masses (dashed lines) of the polypeptide fractions are plotted across the peaks. The theoretical molecular mass of decorin core protein (including purification tag) is 36.2 kDa.
3.6 Study of monomer-dimer equilibrium

The initial SEC-MALS data gave the first indication that wild-type decorin exists in a monomer-dimer equilibrium in solution. To investigate the monomer-dimer equilibrium quantitatively, we used sedimentation velocity analytical ultracentrifugation (AUC). AUC is a powerful technique to quantitatively study the interaction of proteins in solution. In a sedimentation velocity (SV) experiment, the rates of movement of samples are observed as they sediment at very high rotor speed due to differences in density, mass and shape. Two optical systems are used to monitor sedimentation: absorbance at 280 nm and interference optics, as a function of radial distance. The sedimentation coefficient, $S_{20,w}$, of the protein refers to the rate at which the sedimentation boundary moves. It gives information about both the mass and the shape of the protein. The units of $S_{20,w}$ are in Svedbergs (S) where 1 S = $10^{-13}$ seconds.

Sedimentation velocity data were collected at seven concentrations of wild-type decorin, ranging from 0.028 to 3.4 mg/ml (0.77 to 94 μM) in PBS, at a rotor speed of 40,000rpm. SDS-PAGE before and after the AUC run revealed no aggregation (Figure 3.8).

Figure 3.8 SDS-PAGE of wild-type decorin before and after the AUC run. Analysis by SDS-PAGE prior to and following the AUC runs shows no apparent formation of aggregates. The positions of selected molecular weight markers are shown on the left.
HYDROPRO (Garcia de la Torre et al., 2000b) was used to calculate the theoretical $S_{20,w}$ based on an atomic shell model of bovine decorin with appropriate bi-antennary oligosaccharide additions, which yielded $S_{20,w}$ values of 3.0 S and 4.7 S for monomeric and dimeric mouse decorin respectively. SEDFIT (Brown and Schuck, 2006) was used to perform continuous size distribution c(s) analysis of the interference optics data (Figure 3.9). Two distinct peaks were observed, one at 3.6 S and one at 4.7-5.3 S, which were interpreted to be monomer and dimer respectively (Figure 3.10). The agreement between the calculated and experimental $S_{20,w}$ values confirms the existence of mouse decorin in these oligomeric states. The relative proportion of these species varied with protein concentration. At 0.028 mg/ml (0.77 µM), the distribution shows predominantly monomer and a small amount of dimer. The ratio is reversed at 3.4 mg/ml (94 µM), where mostly dimer is present (Figure 3.10). The dissociation constant $K_d$ for the decorin dimer was calculated to be 1.37 ± 0.30 µM (Figure 3.11), indicating that at submicromolar concentrations, decorin is predominantly monomeric.
Figure 3.9 Sedimentation velocity analytical ultracentrifugation of WT decorin.
Representative boundary fits of fringes collected by interference measurements, showing every third scan (from 0.43 to 3.4 mg/ml) and every fifth scan at the 3 lowest concentrations (upper panel) for reason of clarity. Scan boundaries are in black and their fits in red. Experiments were performed at 20 °C with WT decorin in PBS. Scans were fitted to the Lamm equation using SEDFIT.
Figure 3.10 Continuous size distributions at seven different concentrations. The peaks for monomer (m) and dimer (d) are indicated. At 0.028 mg/ml, the peaks in the c(s) distributions show predominantly monomer and a small amount of dimer. The ratio is reversed at 3.4 mg/ml, where mostly dimer is present. Experiments were performed at 20 °C with WT decorin in PBS.
Figure 3.11 Determination of the dimer dissociation constant for WT decorin. Fraction dimer plotted as a function of protein concentration. The relative amounts of monomer and dimer were determined from integration of the peaks in the c(s) distribution. The data were fitted to a model of monomer-dimer equilibrium (Eq. 8 in section 2.2.15).

Similar AUC experiments were carried out with the Y51A/R52A/Q54A and R151E mutants to determine their dissociation constants. SDS-PAGE before and after the AUC runs revealed no aggregation (Figure 3.12). Continuous size distribution c(s) analysis of selected interference optics data of the mutants are shown (Figure 3.13).
Figure 3.12 SDS-PAGE of mouse decorin mutants (R151E and Y51A/R52A/Q54A) before and after the AUC run. Analysis by SDS-PAGE prior to and following the AUC runs shows no apparent formation of aggregates. The positions of selected molecular weight markers are shown on the left.

Figure 3.13 Continuous size distributions at three different concentrations of (A) Y51A/R52A/Q54A and (B) R151E mutants. The peaks for monomer (m) and dimer (d) are indicated. Experiments were performed at 20 °C with decorin mutants in PBS.
The dissociation constant of the Y51A/R52A/Q54A mutant was derived to be $2.3 \pm 0.8 \, \mu M$ (Figure 3.14A) and that of R151E was $0.47 \pm 0.09 \, \mu M$ (Figure 3.14B), i.e. both were found to be similar to that of WT decorin.

**Figure 3.14 Determination of the dimer dissociation constants for decorin mutants.** A) Y51A/R52A/Q54A and B) R151E. Fraction dimer plotted as a function of protein concentration. The relative amounts of monomer and dimer were determined from integration of the peaks in the $c(s)$ distribution. The data were fitted to a model of monomer-dimer equilibrium (Eq. 8 in section 2.2.15).
3.7 Thermal stability of WT and mutant decorin

So far I have shown that WT decorin and dimeric mutants, R151E and Y51A/R52A/Q54A, exist in concentration-dependent monomer-dimer equilibria with comparable dimer dissociation constants, and that stable monomeric mutants (Q61N and Y130N) can be obtained. I then sought to analyse the thermal stabilities of wild-type decorin and the monomeric Q61N mutant using differential scanning calorimetry (DSC). DSC is a thermal analysis technique that can be used to study how the heat capacity of a protein (Cp) changes with temperature. A DSC instrument measures the amount of heat absorbed or released by a sample if heated or cooled. Proteins unfold at a characteristic temperature as the heating progresses, giving rise to an endothermic peak.

WT decorin denatures at a melting temperature of 50 °C and the Q61N mutant protein at 52 °C at pH 7.4 (Figure 3.15). This compares with the reported T_m of 46 °C for bovine decorin (Scott et al., 2006). The denaturation was not completely reversible as the signal following the second up-scan was roughly 10% that of the first. This is in contradiction to reports by Scott et al. (2006) that heat denaturation of decorin is completely reversible but is in agreement with the findings of Krishnan et al. (1999) that recombinant decorin does not recover after heat denaturation up to 60 °C.

The protein samples showed signs of precipitation when taken out of the DSC cell. Due to the irreversibility of the thermal denaturation process, enthalpies could not be calculated.

These results show that engineered monomeric mouse decorin is as stable as wild-type decorin.
Figure 3.15 Thermal denaturation studies of WT and monomeric decorin by DSC. Differential scanning calorimetry unfolding curves of WT (black line) and monomeric (red line) decorin as a function of temperature at pH 7.4. Proteins were at a concentration of 3 mg/ml in PBS. The melting temperatures are 50 °C for WT decorin and 52 °C for the Q61N mutant.

We also tested the thermal stabilities of the remaining mutants using differential scanning fluorimetry (DSF). DSF is a rapid technique for the assessment of protein stability. It monitors thermal unfolding of proteins in the presence of a dye that fluoresces in non-polar environments, such as the hydrophobic sites on unfolded proteins (Niesen et al., 2007). It is superior to other thermal scanning methods such as DSC in being relatively high-throughput and requiring small amounts of protein.

The melting curves for WT decorin and mutants indicate that they unfold within a narrow temperature range of 49 – 51 °C (Figure 3.16). The monomeric mutants (Q61N and Y130N) were slightly more stable than the wild-type protein.
3.16 Thermal denaturation studies of WT and monomeric decorin by DSF.
Differential scanning fluorimetry denaturation curves are shown for WT and mutant decorin as a function of temperature. Proteins were at a concentration of 0.181 mg/ml in PBS. The melting temperatures derived from the inflection points of the curves are: 49 °C for R151E, 50 °C for WT, Y51A/R52A/Q54A, and 51 °C for the Q61N and Y130 mutants. Shown is a representative of two independent experiments performed in triplicate.

3.8 Collagen binding by WT and mutant decorin

3.8.1 Collagen - decorin binding assays

The ability of WT and mutant decorin to bind collagen was tested in a solid phase binding assay previously described by Tenni et al. (2002). Immobilised collagen I was incubated with serial dilutions of biotinylated mouse decorin proteins. Bound decorin was detected with avidin coupled to peroxidase from horseradish. As a positive control, I tested the interaction of biotinylated Drosophila SPARC with collagen I, which showed saturable binding (Figure 3.17A). Wild-type decorin and the monomeric Q61N and Y130N mutants bound collagen in a very similar concentration-dependent manner,
indicating that the mutations might not have an effect on collagen binding (Figure 3.17B). There was, however, significant binding of decorin to plates not coated with collagen (Figure 3.17B, open circle) despite the presence of high concentrations of casein in the blocking and incubation buffers.

Figure 3.17 Binding of dSPARC, WT and monomeric decorin to collagen I. A. Collagen binding by dSPARC FS-EC His-myc as a positive control. B. Collagen binding by WT and mutant decorin (Q61N and Y130N). Collagen I was immobilised onto plastic and increasing amounts of biotinylated decorin added. The binding was detected using avidin peroxidise from horseradish. Data shown are representative of multiple independent experiments. Data were fitted by a single-site saturation model with a linearly increasing nonspecific background.
Since the additional glycan modifications at positions 61 and 130 did not inhibit collagen binding, we inferred that the engineered single mutations were not enough to affect binding of decorin to collagen. However, we were still keen to know which residues in the decorin dimer interface partake in its interaction with collagen. Based on the bovine decorin crystal structure, we selected amino acid residues located on the concave face and produced five mutants built onto the Q61N mutant that we expected to have an effect on collagen binding (Figure 3.18):

1) Q61N/K111S: a mutation which introduces another consensus site for N-linked glycosylation with a glycosylation potential of 0.59

2) Q61N/R267A/E268A: a double alanine substitution that removes two side chains that are involved in the interface

3) Q61N/D204R: a charge reversal mutation in LRR7 - the Asp residue has been previously shown to be important in collagen binding (Kalamajski et al., 2007)

4) Q61N/G246N: a mutation which introduces another consensus site for N-linked glycosylation with a glycosylation potential of 0.54

5) Q61N/G246R: a Gly to Arg substitution to introduce a charged residue to plug a ‘hole’ in the interface.

We hoped that at least one of these mutants would have a profound impact on collagen binding so we identity the collagen binding site in decorin.
Figure 3.18 Surface representation of bovine decorin crystal with location of mutations highlighted. Orange indicates conserved region based on the sequence conservation of decorin from various organisms (cow, mouse, human, rabbit, pig, dog, horse, chick, zebrafish and xenopus). The region that is not conserved is in grey. The location of the Q61N mutation is shown in magenta. The five independent mutations built onto the Q61N mutant are coloured green. The residue numbers correspond to those in mouse decorin. Two out of three naturally occurring N-glycosylation sites are shown in red.

All five mutants were highly expressed from episomally transfected HEK293 cells and purified to homogeneity by nickel affinity chromatography from serum-free medium, followed by SEC. The purified samples were analysed by SDS-PAGE (Figure 3.19).

Figure 3.19 Reducing SDS-PAGE analysis of WT decorin and monomeric decorin mutants (Coomassie Blue stain). The mutants have higher electrophoretic mobility according to their glycosylation status. The glycosylated mutant Q61N runs at higher molecular mass compared to the WT protein. Q61N/K111S and Q61N/G246N have additional glycosylation sites introduced and hence appear slightly bigger than Q61N mutant. The positions of selected molecular weight markers are indicated on the left.

We tested the interaction of these mutants (Q61N/K111S, Q61N/R267A/E268A, Q61N/D204R, Q61N/G246N and Q61N/G246R) with collagen I using a solid phase assay with biotinylated decorin proteins (Figure 3.20). Surprisingly, all five monomeric mutants bound collagen in a similar manner. However, the unacceptably high levels of non-specific binding despite using high concentrations of casein (1%) in the blocking and incubation buffers, made us suspicious of the validity of the results of this assay.
Figure 3.20 Collagen binding by WT decorin and monomeric mutants. Solid-phase binding assay was carried out with increasing amounts of biotinylated WT decorin and four double mutants (K111S, D204R, G246N, G246R) on immobilised rat tail collagen I. Data shown are representative of at least three independent experiments.

Minimal biotinylation using only 2-fold molar excess biotin was tried to reduce the amount of non-specific binding but this did not show saturable binding (Figure 3.21).
Figure 3.21 Collagen binding by Q61N and other monomeric mutants. Solid-phase binding assay was carried out with increasing amounts of minimally biotinylated Q61N and five double mutants (Q61N/K111S, Q61N/R267A/E268A, Q61N/D204R, Q61N/G246N, Q61N/G246R) added to immobilised rat tail collagen I.

3.8.2 Collagen - decorin competition assay

My results so far showed that the engineered glycans (Q61N and Y130N) and also the double mutations failed to obstruct collagen binding in a solid phase assay using biotinylated decorin proteins. However, I was wary of the high levels of non-specific binding and decided to check the validity of my biotin-avidin assay. The specificity of decorin - collagen I interaction was tested in a competition binding assay, in which increasing concentrations of non-biotinylated WT decorin (up to 20 µM) compete with biotinylated WT decorin (1 µM) for binding to collagen. This experiment suggests that at least ~50% of the binding of biotinylated decorin was non-specific as it could not be effectively competed with 20-fold excess of unlabelled decorin (Figure 3.22).
Figure 3.22 Decorin - collagen I competitive binding assay. Solid-phase binding assay with recombinant biotinylated decorin and increasing concentrations of non-biotinylated decorin, added for three hours at room temperature to 96 wells coated with collagen at 25 μg/ml. The addition of a 20-fold excess of non-biotinylated wild-type decorin only partially inhibits the interaction between biotinylated wild-type decorin and collagen I.

Since I could not be confident about the results of my biotin-avidin solid phase assay, I decided to measure decorin binding using an anti-mouse decorin primary antibody.

By carrying out a titration experiment with various dilutions of the antibodies, I established that the optimum antibody concentrations for use in my solid phase assay i.e. the concentration(s) which give the best colorimetric reading with minimum background, were those quoted by the manufacturers (data not shown).

I also confirmed that the mutations are not within the epitopes of the anti-decorin antibody (Figure 3.23). The primary antibody exhibits equal or better binding to all decorin proteins, except Q61N/K111S and Q61N/D204R.
I proceeded to carry out solid phase assays with WT, Q61N, Y130N and R151E mutants. WT decorin showed dose-dependent, saturable binding to collagen with an apparent dissociation constant of $K_{0.5} = 0.31 \pm 0.14 \, \mu \text{M}$ (Figure 3.24). The $K_{0.5}$ value represents the mean ± SEM from three independent experiments. This is consistent with previous observations by Kalamajski et al. (2007) and Tenni et al. (2002). In contrast to the wild-type protein, the Q61N, Y130N and R151E mutants showed greatly reduced binding affinities but were not completely inactive. These results show that the concave face of decorin is involved in collagen binding.

Figure 3.23 Anti-mouse decorin antibody binding to WT and mutant mouse decorin. Solid-phase binding assay was carried out with 96-well plates coated with decorin proteins. Data are expressed as the average of duplicate data points.
Figure 3.24 Binding of WT and mutant mouse decorin to immobilised collagen. Solid-phase binding assay was carried out with increasing amounts of WT or mutant decorin added to 96-well plates coated with rat tail collagen I. Bound decorin was detected by anti-mouse decorin antibody followed by HRP-conjugated rabbit anti-goat IgG antibody, and measured as absorbance at 492 nm. Shown is a representative of three independent experiments, each performed in duplicate. The curves were generated by non-linear least squares fitting of the data by an equation describing single-site binding with linear background.

However, the strong binding seen in this set of experiments did not prove to be very reproducible when repeated with a different batch of proteins, regardless of the amount of blocking agent used (up to 10 mg/ml casein) and/or the use of different antibodies (anti-mouse decorin antibody, anti-His-tag antibody). The data obtained using anti-His-tag antibody for detection are shown in Figure 3.25. I also included the Y51A/R52A/Q54A mutant in this assay.
Figure 3.25 Collagen binding by WT and mutant mouse decorin. Solid-phase binding assay was carried out with increasing amounts of WT or mutant decorin added to 96-well plates coated with rat tail collagen I. Bound decorin was detected by anti-His-tag HRP-conjugated antibody and measured as absorbance at 492 nm. Shown is a representative of three independent experiments, each performed in duplicate. The curves were generated by non-linear least squares fitting of the data by an equation describing single-site binding with linear background.

Even though I am not confident about reporting $K_{0.5}$ values for these interactions as they varied from experiment to experiment, I consistently observe stronger collagen binding by wild-type decorin and the Y51A/R52A/Q54A mutant than any of the single point mutants (Figure 3.25).
3.8.3 Collagen fibrillogenesis assay

Since the results of the solid phase assays were not very robust, I decided to carry out a collagen fibrillogenesis assay which is the classic assay for testing decorin function.

Wild-type decorin retarded fibrillogenesis of type I collagen as has been reported previously (Douglas et al., 2006, Zhang et al., 2009, Raspanti et al., 2007, Kalamajski et al., 2007). The Y51A/R52A/Q54A mutant delayed fibrillogenesis similar to the WT protein. The mutants Q61N, Y130N and R151E did not. This assay was repeated six times using different batches of collagen with similar results (Figure 3.26). These findings indicate that mutations in the decorin dimer interface interfere with collagen binding.

![Figure 3.26 Effect of decorin on collagen fibrillogenesis. Mouse type I collagen (32 μg/ml) was incubated at 37 °C and pH 7.8. Fibril formation was monitored by turbidity readings at 400 nm at 0.5 minute intervals. WT and mutant mouse decorin proteins were added at a concentration of 50 μg/ml. A and B show representative experiments using two different batches of collagen. For each data set, data shown are representative of three independent experiments](image-url)
Thus the combined results of the solid phase and fibrillogenesis assays indicate that the concave face of decorin (residues 61, 130 and 151 in particular) is involved in collagen binding.
Chapter 4: Discussion

Analysis of decorin dimerisation
4. Discussion

SLRPs such as decorin and biglycan are known to bind various types of collagens, thereby regulating the assembly of fibrils in skin, tendons, and cornea (Kalamajski and Oldberg, 2010, Chen and Birk, 2013). The biological significance of these interactions *in vivo* can be deduced from the phenotype of decorin-deficient mice. Mice lacking decorin are characterised by fragile skin, and their collagen fibrils have irregular diameters due to uncontrolled lateral fusion of fibrils (Danielson et al., 1997). Molecular modelling studies of the interactions between decorin core protein and type I collagen fibrils have suggested that the concave face of decorin may bind to one or more triple helices (Weber et al., 1996, Orgel et al., 2009). Consistent with this notion, the concave face is the most conserved region (Scott et al., 2004). However, the crystal structures of decorin and biglycan revealed that the concave face is involved in forming a tight dimer, making it potentially unavailable for ligand binding (Scott et al., 2004, Scott et al., 2006). Hence, there has been controversy about the oligomeric status of decorin (Goldoni et al., 2004) and how it binds collagen was unclear. In this thesis, I have attempted to resolve this controversy by defining the oligomeric state of decorin and studying its interactions with collagen.

The key findings from my thesis are:

1) Wild-type decorin exists in a monomer-dimer equilibrium with a dimer dissociation constant of $1.37 \pm 0.30 \mu M$.

2) Engineering glycosylation sites into the dimer interface abolished dimerisation; other interface mutants remained dimeric.

3) Both wild-type and mutant decorin are equally stable in thermal denaturation studies.

4) Mutants in the dimer interface abolished collagen binding, regardless of the oligomeric status of the mutants.
Goldoni et al. (2004) provided some lines of evidence to suggest that decorin is monomeric in solution and that the oligomerisation of decorin observed by Scott et al. (2004) is a result of extensive dialysis of the samples against water followed by freeze-drying, leading to non-specific association. However, each of the techniques used by Goldoni et al. to determine the oligomeric state of decorin had their disadvantages. Unlike Scott et al. (2004), they used recombinant decorin under non-denaturing conditions. Conventional SEC was used to show that decorin behaved as a monomer since it nearly coelutes with BSA (MW \( \sim 66 \) kDa). But SEC is not as suitable as SEC-MALS in determining molecular masses due to reasons discussed in chapter 3. Chemical crosslinking was used to show that very high concentrations of cross-linking agents are required for preferential dimer formation in decorin, compared with other dimeric proteins such as the EGFR. But crosslinking depends more on the availability of lysine residues than the concentration of cross-linking agents, and only two of the fifty-four lysines available on decorin are close enough to be crosslinked for dimer formation (Scott et al., 2006). The third line of evidence was from mass-spectrometry (MALDI-MS) data which gave a mass of \( \sim 44 \) kDa for the decorin core protein, but this technique is performed \textit{in vacuo}, not in solution. In addition, Goldoni et al. did not provide any indication of the stability of their decorin preparation. Hence, the controversy still prevailed and a comprehensive study using robust biophysical techniques to determine the oligomeric status of decorin was missing.

I attempted to analyse the role of the concave face of decorin by mutating solvent-exposed residues in the LRRs. Since the LRR fold is very resilient to non-conservative replacements of residues (Howitt et al., 2004), dramatic amino acid substitutions were chosen, including the introduction of two new glycosylation sites. Wild-type decorin and interface mutants were expressed in mammalian cells to ensure appropriate post-translational modifications. Previous studies have reported that removal of N-linked oligosaccharides leads to aggregation and reduced solubility of decorin core protein (Scott and Dodd, 1990). All proteins were highly expressed and soluble, indicating correct folding. The yield of Y51A/R52A/Q54A was relatively poor compared to WT and other mutants and this could be due to the triple mutation being more dramatic compared with point mutations.
The oligomeric states of WT and mutant decorin at 3 mg/ml (83 µM) were determined using SEC-MALS, which allows continuous monitoring of light scattering and refractive index as the samples elute from a gel filtration column. These experiments suggested that WT decorin exists in a monomer-dimer equilibrium. The elution profiles of Y51A/R52A/Q54A and R151E mutants, and the negative control mutants located outside the interface, E156K and K159E, all resembled that of WT decorin i.e. they were dimeric. However, the mutants with engineered glycosylation sites are purely monomeric (MW~ 36.2 kDa), even at such high concentrations, indicating that the decorin dimer had been disrupted.

I investigated the monomer-dimer equilibrium quantitatively using analytical ultracentrifugation and determined a dimer dissociation constant of 1.37 ± 0.30 µM for WT decorin. This is similar to the dimer dissociation constant of 4.5 µM obtained for biglycan (calculated from a free energy of association of -7.3 kcal/mol), which has been shown to dimerise reversibly in solution (Liu et al., 1994). Two distinct peaks in the c(s) distribution (corresponding to monomers and dimers) can be seen in sedimentation velocity studies, indicating that the rate of interconversion of species is slow on the time-scale of sedimentation. This is in contrast to the SEC-MALS experiment where a dimeric peak with a pronounced tail and a slightly lower than expected molecular mass is observed, suggesting a more rapidly equilibrating system. The two results are not necessarily inconsistent and are likely due to differences in the two methods. In SEC, the resolution of the technique may not be high enough for the two peaks to be separated chromatographically. In AUC, the dimer sediments faster than the monomer and is therefore, always in the presence of the monomer. The longer running time of this technique (~15 hours) means that both monomers and dimers sediment, giving rise to two peaks in the c(s) distribution.

Sedimentation velocity studies on the dimeric R151E and Y51A/R52A/Q54A mutants yielded dimer dissociation constants similar to that of WT decorin. Thus, at high concentrations typically used in solution scattering (Scott et al., 2006, Scott et al., 2003) and crystallisation experiments (Scott et al., 2004, Scott et al., 2006), decorin and biglycan are dimeric. At likely physiological concentrations, dimers will dissociate into monomers.
All forces that stabilise the native structures of proteins are sensitive to changes in temperature. Hence, the susceptibility of proteins to thermal denaturation is an indicator of their stabilities (Sola and Griebenow, 2009). The stabilities of WT and mutant decorin were assessed by DSC and DSF. All mutants showed similar denaturation profiles and unfolded within a narrow temperature range of 49 – 51 °C. Thus, the mutations did not have an adverse effect on the structure and stability of the decorin core. Engineered monomeric mouse decorin was found to be marginally more stable than wild-type decorin. This is not unexpected since the stabilising effect of glycans is well-known and glycoengineering is regularly used by the pharmaceutical industry to improve the stability of protein pharmaceuticals (Sola and Griebenow, 2009). Stability can be increased in two different ways: a) lowering the energy level of the native folded state, and b) increasing the energy level of the unfolded state. Experiments by Kwon et al. to study the effects of glycosylation on the unfolding and refolding rates of human alpha 1-antitrypsin showed that glycosylation slows down the unfolding process without affecting the refolding rates significantly (Kwon and Yu, 1997). It was proposed that the increase in thermodynamic stability caused by glycosylation could be due to stabilisation of the native state and not due to destabilisation of the unfolded state (Sola and Griebenow, 2009). Each glycan unit attached to the protein surface provides ~ 1-2 kcal/mol of thermodynamic stabilisation (Sola et al., 2007). Conversely, another study using Kazal-type serine protease inhibitors hypothesised that carbohydrates sterically restrict peptide backbone mobility, thus destabilising the unfolded state by reducing the entropy of unfolding (DeKoster and Robertson, 1997). Our results contradict the conclusion of Scott et al. that monomeric decorin cannot be stable in solution (Scott et al., 2003). Unfolding experiments with decorin and biglycan revealed transitions from folded (dimer) to unfolded (monomer) (Scott et al., 2003, Scott et al., 2006). This has been taken as evidence that dimerisation is required for stabilisation of decorin and biglycan, but I have managed to obtain monomeric decorin by engineering glycosylation sites into the dimer interface (Q61N and Y130N). Scott et al. also showed that the denaturation of decorin was completely reversible, in contrast to our findings and those of Krishnan et al. (1999). Scott et al. attributed this to the lack of disulphides due to the vaccinia virus/T7 bacteriophage expression system used by Krishnan et al. (1999). I have used a
mammalian expression system to produce recombinant decorin, so it is unlikely that the inability of decorin to refold following heat denaturation is due to the absence of disulphides.

LRR proteins such as internalinA (Schubert et al., 2002), the Nogo receptor (He et al., 2003) and domain 3 of Slit (Howitt et al., 2004) exist as stable monomers. In contrast, many crystal structures of LRR proteins reveal putative dimerisation interfaces. Proteins such as AMIGO-1 (Kajander et al., 2011) and LRR domain 4 of Slit2 (Seiradake et al., 2009) dimerise similarly to decorin and do not dissociate into monomers. Mutations in the dimer interface of AMIGO-1 affect folding and stability (Kajander et al., 2011). Some LRR proteins such as the ectodomain of the human Toll-like receptor 3 have glycosylation sites on both the convex and concave faces and one of the two major side faces, leaving only one of the side faces for ligand binding or formation of homodimer (Choe et al., 2005). The TLR3 ectodomain is monomeric in solution but the full-length transmembrane protein may be able to form stable or transient dimers. However, for most of the crystallographic dimers of LRR proteins elucidated so far, there is no biophysical and/or functional evidence to suggest that these dimers are stable in solution and not a consequence of associations formed in the crystal lattice (Bella et al., 2008). In addition, the functional relevance of oligomerisation is not clear. For example, the crystal structure of *Yersinia pestis* cytotoxin, YopM, an LRR protein, revealed a tetramer wherein the four monomers bury a large amount of the hydrophobic surfaces. However, SEC-MALS shows that the protein is monomeric, with or without the presence of calcium and cross-linking studies show oligomerisation only in the presence of calcium. In addition, the biological significance of oligomerisation is unclear, hence it is assumed that tetramerisation is an artefact resulting from the high concentration of calcium in crystallisation solutions (Evdokimov et al., 2001). Thus, in the absence of further evidence, it should be assumed that at physiological concentrations, the quaternary structures of these proteins are predominantly monomeric.

The total buried surface area between LRR-domain dimer interfaces or between LRR domains and their ligands, are often taken as indicators of affinity of interactions i.e. in principle, large buried surfaces suggest tight complexes. For LRR – ligand complexes,
the total buried surfaces are typically in the 2500-3400 Å² range (i.e. the size of the interface is half of this value) (Bella et al., 2008). For ribonuclease inhibitor bound to one of its ligands, angiogenin, such high contact surfaces (2680 Å²) are associated with equilibrium dissociation constants in the femtomolar range (Papageorgiou et al., 1997). For the decorin crystallographic dimer, a similar contact area of 2300 Å² has been used to explain the observed tight interaction between the two monomers (Scott et al., 2004). The crystal structure revealed a banana-shaped molecule that maximise the surface area, unlike the more rounded, horseshoe-shaped ribonuclease inhibitor. However, there are instances when the total buried surface does not relate to the binding affinity. Interaction between internalinA and its ligand, human E-cadherin1 domain, buries a total area of 2400 Å² but the affinity of the interaction is remarkably weak, in the micromolar range (~ 50 µM) (Schubert et al., 2002). Thus, the total contact area alone is not a reliable indicator of the strength of interactions between the LRR domain and its ligands, and specific interactions need to be taken into account, together with other evidence from biophysical studies. In the case of decorin, molecular interactions at the decorin dimer interface are made up of a few hydrophobic and aromatic residues which are sandwiched, and hydrogen bonds or salt bridges. The latter are intrinsically weak in an aqueous environment as polar/charged residues buried in the interface form weak hydrogen bonds with protein atoms, rather than strong hydrogen bonding to the solvent (Xu et al., 1997). The geometries of polar and charged residues are also important in determining their quality since owing to constraints imposed by bond lengths and bond angles, the geometries of hydrogen bonds are unlikely to be optimised across protein interfaces (Xu et al., 1997). I determined the shape complementarity between the two decorin monomers using a shape correlation statistic, S_c, which depends both on the relative shapes of the surfaces with respect to each other and on the extent to which the interaction brings individual elements of the opposing surfaces into proximity (Lawrence and Colman, 1993). Interfaces with perfect fits have an S_c value of one, whereas uncorrelated interfaces have S_c values nearing zero. An S_c value of 0.73 was obtained for the decorin dimer which falls within the typical range of 0.70 to 0.76 found for oligomeric proteins, indicating good shape complementarity. This value is a measure of the ‘global’ shape correlation but does not account for areas such as
solvent-filled pockets seen in the concave face of decorin, which have poorer shape correlation.

To summarise, here, I have provided compelling biophysical evidence that in the case with decorin, dimerisation is weak and reversible.

The decorin protein core binds collagen while the GAGs extend outward into the inter-fibril space. In vitro fibrillogenesis assays and solid phase assays indicate the involvement of the core protein in collagen binding, independent of the GAG chains (Rada et al., 1993, Kalamajski et al., 2007, Vogel et al., 1987). The GAG chain on decorin has no detectable effect of decorin structure and conformational stability when examined by circular dichroism spectroscopy (Krishnan et al., 1999). GAGs bound to decorin act like bridges and transfer forces between contiguous collagen fibrils, thus providing mechanical integrity to tissues (Redaelli et al., 2003). The concave face of decorin, LRR6 and LRR7 in particular, has been implicated in collagen binding (Kalamajski et al., 2007, Kresse et al., 1997). The convex face of decorin is made of residues with a higher degree of evolutionary sequence variability than the concave face. In addition, the presence of endogenous N-oligosaccharides on one of the side faces of the dimer indicates that this side is not involved in decorin-collagen interactions. McEwan et al. have suggested hypothetical ligand-binding scenarios for dimeric decorin including a) dimeric decorin binding collagen with a 2:2 stoichiometry, and b) the dimer dissociating to interact with collagen as monomers (McEwan et al., 2006).

In order to characterise decorin binding to collagen type I, I carried out solid phase binding assays with immobilised collagen I. This technique proved to be less robust than expected due to high levels of non-specific binding and issues with reproducibility using different batches of proteins. Solid phase decorin-collagen binding data in the literature also show weak and noisy signals and high levels of non-specific binding (Kalamajski et al., 2007, Kresse et al., 1997). Tenni et al. (2002) have performed solid phase assays to test the interaction of biotinylated decorin with methylated type I collagen peptide fragments but I have not been able to reproduce this using biotinylated decorin and immobilised rat tail collagen I. The problem is likely to be in the nature of decorin rather than the assay format itself since I encountered no such issues when
performing the same assay with other proteins such as SPARC (chapter 4 of this thesis). Even though I am not confident in quoting $K_{0.5}$ values for the interactions, the behaviour of the mutants was consistent i.e. WT decorin and Y51A/R52A/Q54A mutant bind collagen more strongly than any of the dimer-interface single mutants (Q61N, Y130N and R151E). The minimal binding observed with the single mutants is likely to be non-specific.

The collagen fibrillogenesis assay, on the other hand, is very robust and was reproducible with different batches of decorin and collagen. Collagen incubated at 37 °C in buffers with a neutral pH assembles into insoluble fibrils; the kinetics of this process can be monitored by turbidimetry (Vogel et al., 1984). Collagen fibrillogenesis is a multi-step process which begins with the assembly of collagen to form fibril intermediates. These immature preformed intermediates undergo linear and lateral growth to form mature fibrils. Genetic studies of fibrillogenesis in the cornea of decorin- and biglycan-deficient mice have shown that these SLRPs, positioned on fibrils in an orthogonal array, do not affect the initial formation of fibril intermediates but regulate linear growth and lateral association (Zhang et al., 2009).

Several studies have used turbidity assays to examine the effects of decorin on collagen fibrillogenesis (Vogel et al., 1984, Uldbjerg and Danielsen, 1988, Kalamajski et al., 2007). The lag phase of the turbidity curve is mainly characterised by longitudinal growth; the collagen appears as thin filaments. The increase in turbidity is due to lateral association of these fibrils (Hedbom and Heinegard, 1989). Mutations in the dimer interface abolished the ability of decorin to inhibit fibrillogenesis by hindering lateral growth of fibrils. In contrast, a triple mutant (Y51A/R52A/Q54A), with mutations in the N-terminal cap region, significantly inhibited fibrillogenesis although not as efficiently as WT decorin. Previous studies have reported the importance of LRR6 (Kresse et al., 1997) and LRR7 (Kalamajski et al., 2007) in collagen binding. Hence, a large part of the concave face of decorin seems to be involved in collagen binding. This view is supported by molecular modelling studies using the experimentally derived structure of fibrillar type I collagen (Orgel et al., 2006, Perumal et al., 2008) which shows that monomeric decorin shows the most appropriate shape complementarity with collagen and can interact with up to six triple helices on the fibril surface (Orgel et al., 2009). Rather than
any small collection of amino acid interactions being specifically responsible for the decorin-collagen association, there seems to be an array of potential hydrogen bonds and electrostatic interactions (Orgel et al., 2009).

Another interpretation of my findings would be that collagen binding requires an intact decorin dimer and that the monomeric mutants, Q61N and Y130N, are inactive due to disruption of the dimer. In that case, the dimeric R151E mutant should have been active as well but this is not the case. At the concentrations used in the fibrillogenesis assay, WT and R151E mutant are 50% and 34% monomer respectively, calculated from their dissociation constants determined by AUC. Such a small difference cannot explain the complete loss of binding by the R151E mutant. Hence, our model that the concave face of decorin mediates collagen binding by dissociation of dimer to monomers, still stands.

To summarise, I have shown that decorin dimerisation is reversible and that the concave face of decorin is alternately involved in dimerisation and collagen binding (Table 4.1).

Table 4.1 Summary of results obtained with mouse decorin and its mutants.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Location of mutated residue(s)</th>
<th>Oligomeric state(^a)</th>
<th>Collagen binding(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td></td>
<td>Monomer-dimer equilibrium</td>
<td>Yes</td>
</tr>
<tr>
<td>Y51A/R52A/Q54A</td>
<td>N-terminal cap</td>
<td>Monomer-dimer equilibrium</td>
<td>Yes</td>
</tr>
<tr>
<td>Q61N</td>
<td>LRR1</td>
<td>Monomer</td>
<td>No</td>
</tr>
<tr>
<td>Y130N</td>
<td>LRR4</td>
<td>Monomer</td>
<td>No</td>
</tr>
<tr>
<td>R151E</td>
<td>LRR5</td>
<td>Monomer-dimer equilibrium</td>
<td>No</td>
</tr>
</tbody>
</table>

\(^a\) Determined by SEC-MALS and AUC

\(^b\) Determined by collagen fibrillogenesis and solid-phase binding assays
**Future work**

The work outlined in this thesis has made significant progress in resolving the controversy about the oligomeric status of decorin and how this affects collagen binding. Identification of a stable monomeric mutant that retains the ability to bind collagen would be definitive proof of our model but this may be difficult to obtain since the dimer interface is the same region that is involved in collagen binding. It would be useful to identify collagen peptides that bind decorin since these can be used to carry out solution scattering and/or co-crystallisation experiments to show that the decorin monomer mediates collagen binding. Residues in the core region of decorin and other class I SLRPs such as biglycan and asporin have high sequence identity so it would be interesting to study how these proteins interact with collagen as well.
Chapter 5: Results
Structural and functional studies of invertebrate SPARC binding to collagen
5. Results

The crystal structure of human SPARC (hSPARC) bound to a collagen-like triple-helical peptide was determined in the Hohenester lab (Hohenester et al., 2008). The key collagen binding residues are conserved between human and invertebrate SPARC (Drosophila and C. elegans). However, collagen binds to hSPARC following proteolytic cleavage of an inhibitory loop that is not present in invertebrate SPARC (Sasaki et al., 1997). The aim of this study was to determine the crystal structures of dSPARC and cSPARC. This involved enzymatic deglycosylation or mutagenesis of the constructs to aid crystallisation. A further aim was to investigate the role of the inhibitory loop on collagen binding by studying binding interactions between human and invertebrate SPARC and collagen I and IV, using solid phase assays and surface plasmon resonance.

5.1 Purification of SPARC proteins

5.1.1 Purification of SPARC His.myc tagged constructs

SPARC FS-EC and EC constructs, containing N-terminal His-myc tags, were produced in episomally transfected HEK293 cells. These constructs were generated for use in biochemical assays such as ELISA and BIAcore. All 6 constructs (from hSPARC, dSPARC, cSPARC) were purified to homogeneity by nickel affinity chromatography from serum-free medium, followed by dialysis against appropriate Ca\(^{2+}\)-containing buffer. The yields for SPARC His-myc proteins from 1.5 L of medium are tabulated in Table 5.1. The purified samples were analysed by SDS-PAGE (Figure 5.1). Due to the presence of N-linked glycosylation sites in SPARC (Table 5.1), the purified proteins migrate as diffuse bands at higher molecular mass on SDS-PAGE (Figure 5.1).
Table 5.1 Molecular masses of SPARC proteins, number of N-linked glycosylation sites and yields from 1.5 L of serum-free medium

<table>
<thead>
<tr>
<th>Protein(s)</th>
<th>Residues</th>
<th>Calculated molecular mass (kDa)</th>
<th>N-linked glycosylation sites</th>
<th>Yield per 1.5 L of serum-free medium (mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>dSPARC FS-EC NHis-myc</td>
<td>83 - 304</td>
<td>31.3</td>
<td>2</td>
<td>19.5</td>
</tr>
<tr>
<td>dSPARC EC NHis-myc</td>
<td>169 - 304</td>
<td>21.6</td>
<td>1</td>
<td>15.8</td>
</tr>
<tr>
<td>hSPARC FS-EC NHis-myc</td>
<td>70 - 303</td>
<td>32.7</td>
<td>2</td>
<td>18.4</td>
</tr>
<tr>
<td>hSPARC EC NHis-myc</td>
<td>153 - 303</td>
<td>23.4</td>
<td>1</td>
<td>13.3</td>
</tr>
<tr>
<td>cSPARC FS-EC NHis-myc</td>
<td>52 - 264</td>
<td>30.1</td>
<td>2</td>
<td>10.9</td>
</tr>
<tr>
<td>cSPARC EC NHis-myc</td>
<td>139 - 264</td>
<td>20.3</td>
<td>1</td>
<td>9.4</td>
</tr>
</tbody>
</table>

Figure 5.1 Purification of SPARC proteins. A. SDS-PAGE analysis of purified dSPARC & hSPARC NHis-myc tagged constructs EC. B. SDS-PAGE analysis of cSPARC NHis-myc tagged constructs. Protein samples were dialysed against 20 mM HEPES pH 7.5, 150 mM NaCl, 2 mM CaCl₂. The positions of selected molecular mass markers are shown on the left.
5.2 Expression and Purification of PNGase F-MBP

Peptide-N-glycosidase F (PNGase F) is an enzyme that cleaves between the innermost GlcNAc and asparagine residues of N-linked glycan modifications. Purified active PNGase F-MBP protein was required in order to deglycosylate SPARC constructs for crystallisation trials. The PNGase F-MBP protein was expressed in *E. coli* BL21 cells. Purification was performed by affinity chromatography on an amylose resin column and the purified protein was eluted in 10 mM maltose, 20 mM Tris-HCl pH 7.5, 200 mM NaCl, 1 mM EDTA (Figure 5.2) and gel filtered in 20 mM Tris-HCl pH 7.5, 50 mM NaCl, 5 mM Na₂EDTA. A 1 L culture yielded 13 mg of protein.

![Figure 5.2 SDS-PAGE analysis of PNGaseF-MBP IMAC purification.](image)

Figure 5.2 SDS-PAGE analysis of PNGaseF-MBP IMAC purification. SDS PAGE analysis shows: Son, whole cell lysate after sonication; SnP, supernatant after centrifugation with protamine sulphate; new, purified PNGase F-MBP eluted in 10 mM maltose, 20 mM Tris-HCl pH 7.5, 200 mM NaCl, 1 mM EDTA; old, PNGase F-MBP sample from previous stock. The band at 80 kDa corresponds to the intact fusion enzyme. The lower band(s) most likely corresponds to a cleaved product containing the MBP tag. The positions of selected molecular mass markers are shown on the right.
5.3 Purification of SPARC His tagged constructs and crystallisation trials

5.3.1 Purification of dSPARC EC constructs for crystallisation trials

A dSPARC EC domain construct containing an N-terminal His-tag was produced in episomally transfected HEK293 cells and purified by nickel affinity chromatography from serum-free medium.

The EC NHis construct was deglycosylated using PNGase F. Figure 4.3A shows the SDS-PAGE gel of the deglycosylated form along with the undigested protein. The deglycosylated protein band migrated at its predicted molecular mass of \(~19\) kDa. SEC was carried out in order to further purify the samples for crystallisation trials. The SEC elution profile for SPARC EC NHis is shown in Figure 5.3B. A small shoulder in the eluted peak (at 14.9 ml) suggests that the protein can form higher oligomers in solution. When analysed by SDS-PAGE all the fractions have the same band pattern suggesting that the association may be non-covalent (data not shown). The protein was concentrated to 10.3 mg/ml and 17.8 mg/ml and crystallisation screens were set up and kept at 21°C. These trials failed to yield crystals (Table 5.2).
5.3.3 Deglycosylation of dSPARC EC NHis and its purification by SEC

A. Reducing SDS-PAGE analysis showing removal of N-linked glycans from SPARC EC-NHis. The band corresponding to a possible SPARC EC-NHis dimer disappears upon prolonged boiling of the protein sample. The positions of selected molecular mass markers are shown on the left. B. The elution profile of purified SPARC EC-NHis on a Superdex S200 column is shown. Protein samples were eluted in 20 mM HEPES pH 7.5, 150 mM NaCl, 2 mM CaCl$_2$. The arrows show the peak elution volumes of molecular mass standards in kDa.

5.3.2 Purification of dSPARC FS-EC constructs for crystallisation trials

A dSPARC construct containing the FS-EC domain with an N-terminal His tag was produced in episomally transfected HEK293 cells and purified by nickel affinity chromatography from serum-free medium. The yield from 1.5 L of cell culture medium was 32.4 mg of protein.

5.3.2.1 Analytical deglycosylation of dSPARC FS-EC NHis

Enzymatic methods of deglycosylation are required to remove heterogeneity in glycoproteins for subsequent structural analysis. A small-scale deglycosylation
experiment was carried out using PNGaseF-MBP to test out various conditions to find the gentlest method of removing N-linked glycans, without denaturing the protein. A 1:25 enzyme : substrate ratio seems to partially deglycosylate the sample, producing an intermediate product in addition to the deglycosylated protein (Figure 5.4). Using 1:5 enzyme : protein molar ratio resulted in a high level of deglycosylation. Since the reactions seem to be independent of incubation temperature, subsequent large-scale deglycosylation experiments were performed overnight at room temperature using a 1:5 enzyme : protein molar ratio.

![Figure 5.4 SDS-PAGE analysis showing degree of deglycosylation of N-linked glycans from dSPARC FS-EC NHis. The deglycosylation efficiency was determined by having two different enzyme to protein molar ratios (1:5 and 1:25) at two different temperatures (4 °C and 35 °C). Molecular weight of dSPARC FS-EC NHis, based on amino acid composition, is 27 kDa. The positions of selected molecular mass markers are indicated on the left.](image-url)
The FS-EC NHis construct was deglycosylated using PNGase F-MBP. Another round of affinity chromatography was performed to remove the enzyme. The protein was then further purified by SEC (Figure 5.5A). The elution profile shows the presence of mixed species in solution – presumed monomeric and dimeric. SDS-PAGE analysis following deglycosylation shows that the glycans were not removed completely by PNGase F-MBP treatment as two bands are still present – one above 30 kDa (the deglycosylated protein is 29 kDa) (Figure 5.5B). The sample was concentrated to 11.4 mg/ml and crystallisation screens were set up and kept at 21 °C. This trial failed to yield crystals most likely due to the presence of mixed species of different oligomeric states and also due to partial deglycosylation.

Figure 5.5 Purification and deglycosylation of dSPARC FS-EC NHis A. The elution profile of purified dSPARC FS-EC NHis on a Superdex S200 column is shown. Protein samples were eluted in 20 mM HEPES pH 7.5, 150 mM NaCl, 2 mM CaCl₂. The profile shows presence of two different oligomeric states. The peak region below the black line corresponds to the pooled fractions in B. The arrows show the peak elution volumes of molecular mass standards in kDa. B. SDS-PAGE analysis of dSPARC FS-EC NHis deglycosylation. The gel shows that the protein has not been deglycosylated completely by PNGase F treatment as there are two bands present (shown by arrows). The positions of selected molecular mass markers are shown on the left.
5.3.2.2 Lectin treatment

Since enzymatic treatment did not result in complete deglycosylation of dSPARC constructs, the next approach was to treat the sample with lectins, carbohydrate-binding proteins, following PNGaseF-MBP digestion, to remove any residual glycosylated proteins. It was not possible to obtain fully deglycosylated protein using this method, as shown by the presence of two bands, one above 30 kDa (Figure 5.6).

![Figure 5.6 SDS-PAGE analysis showing degree of deglycosylation of N-linked glycans from dSPARC FSEC-NHis. The presence of two bands (shown by arrows), including one above 30 kDa, indicates that treatment with lectin did not remove all glycosylated species. The positions of selected molecular mass markers are shown on the left.](image)

5.3.2.3 dSPARC glycosylation site mutant

A dSPARC double mutant, with mutations in the 2 putative glycosylation sites (N126 and N214 to Q), was generated in the hope that removing carbohydrates might facilitate crystallisation. The yield was ~12 mg from 2 L of cell culture medium. Crystallisation
trials were set up at 10.2 mg/ml and at 19.4 mg/ml and no crystals were formed under these conditions (Table 5.2).

5.3.3 Purification of cSPARC proteins for crystallisation trials

At this point, attempts were made to crystallise SPARC from an alternative invertebrate species, namely C. elegans, in its native state. cSPARC FS-EC (residues 52-264) and EC (residues 139-264) constructs, containing N-terminal His-tags, were produced in episomally transfected HEK293 cells.

The yield from 2 L of cell culture medium was 3.7 mg of cSPARC EC NHis protein. Figure 5.7A shows the SEC profile of this protein. The first peak is within the void volume of the column. The second peak (starred) could be a dimer. The major peak, eluting at ~12 ml, contains monomeric SPARC EC NHis (~ 16 kDa). The two bands on SDS-PAGE (Figure 5.7B) correspond to different glycoforms of the protein.

Figure 5.7 Purification of cSPARC EC NHis A. The elution profile of purified cSPARC EC NHis on a Superdex S75 column is shown. Protein samples were eluted in 20 mM HEPES pH 7.5, 150 mM NaCl, 2 mM CaCl₂. The profile shows presence of cSPARC EC NHis monomer and a tiny proportion of dimer. The peak region below the black line corresponds to the pooled fractions in B. The arrows show the peak elution volumes of molecular mass standards in kDa. B. SDS-PAGE analysis of cSPARC EC NHis purification. The lane marked with an asterisk and the underlined fractions refer to the corresponding SEC peak(s). The positions of selected molecular mass markers are shown on the left.
For cSPARC FS-EC NHis protein, the yield from 2 L of cell culture medium was 13.4 mg. **Figure 5.8A** shows the SEC profile. The first peak is within the void volume of the column. The second peak (starred) could be a dimer. The major peak, eluting at \( \sim 10.8 \) ml, contains monomeric cSPARC FS-EC NHis (\( \sim 26 \) kDa).

**Figure 5.8 Purification of cSPARC FS-EC NHis.** A. The elution profile of purified cSPARC FS-EC NHis on a Superdex S75 column is shown. Protein samples were eluted in 20 mM HEPES pH 7.5, 150 mM NaCl, 2 mM CaCl\(_2\). The profile shows presence of cSPARC FS-EC NHis monomer and a small proportion of dimer. The peak region below the black line corresponds to the pooled fractions in B. The arrows show the peak elution volumes of molecular mass standards in kDa.

B. SDS-PAGE analysis of cSPARC FS-EC NHis purification. The lane marked with an asterisk and the underlined fractions refer to the corresponding SEC peak(s). The positions of known molecular mass markers are shown on the left.

Crystallisation trials were set up with cSPARC EC NHis at 7.9 mg/ml and cSPARC FS-EC NHis at 18.9 mg/ml. Both proteins failed to crystallise in their fully glycosylated states (*Table 5.2*).
Table 5.2 Summary of crystallisation trials. Concentrated proteins were screened against various pre-filled 96-well plate screens. ‘x’ indicates the screen has been used to set up trials. A list of the composition of the crystallisation screens can be found at http://www3.imperial.ac.uk/xraycrystallography/crystn/xtalscreens. Refer to table 5.3 for the names of the screens and their manufacturers.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Concentration in mg/ml</th>
<th>Plate ID/Screen</th>
</tr>
</thead>
<tbody>
<tr>
<td>dSPARC EC NHis</td>
<td>10.3/17.8</td>
<td>x/x</td>
</tr>
<tr>
<td>dSPARC FS-EC NHis, deglyc</td>
<td>11.4</td>
<td>x</td>
</tr>
<tr>
<td>dSPARC FS-EC NHis glyc mut</td>
<td>10.2/19.4</td>
<td>x/x</td>
</tr>
<tr>
<td>cSPARC FS-EC NHis</td>
<td>18.9</td>
<td>x</td>
</tr>
<tr>
<td>cSPARC EC NHis</td>
<td>7.9</td>
<td>x</td>
</tr>
</tbody>
</table>

Table 5.3 Summary of crystallisation trials (continued). List of pre-filled 96-well plate screens and their manufacturers.

<table>
<thead>
<tr>
<th>Plate ID</th>
<th>Screen</th>
<th>Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICL1</td>
<td>Crystal Screen 1 and 2</td>
<td>Hampton Research</td>
</tr>
<tr>
<td>ICL2</td>
<td>Wizard 1 and 2</td>
<td>Emerald Biosystems</td>
</tr>
<tr>
<td>ICL3</td>
<td>PEG/ion and Natrix</td>
<td>Hampton Research</td>
</tr>
<tr>
<td>ICL4</td>
<td>Index</td>
<td>Hampton Research</td>
</tr>
<tr>
<td>ICL5</td>
<td>SaltRx</td>
<td>Hampton Research</td>
</tr>
<tr>
<td>ICL6</td>
<td>Memstart/Memsys</td>
<td>Molecular Dimensions</td>
</tr>
<tr>
<td>ICL7</td>
<td>PACT Premier</td>
<td>Molecular Dimensions</td>
</tr>
<tr>
<td>ICL8</td>
<td>JCSG+</td>
<td>Molecular Dimensions</td>
</tr>
<tr>
<td>ICL9</td>
<td>MemGold</td>
<td>Molecular Dimensions</td>
</tr>
<tr>
<td>ICL10</td>
<td>PEG/ion 2 and Wizard 3</td>
<td>Hampton Research/Emerald Biosystems</td>
</tr>
<tr>
<td>ICL11</td>
<td>JBScreen Cryo</td>
<td>Jena Biosciences</td>
</tr>
<tr>
<td>ICL12</td>
<td>Proplex</td>
<td>Molecular Dimensions</td>
</tr>
<tr>
<td>ICL13</td>
<td>Morpheus</td>
<td>Molecular Dimensions</td>
</tr>
<tr>
<td>ICL14</td>
<td>PGAScreen</td>
<td>Molecular Dimensions</td>
</tr>
</tbody>
</table>
5.4 Collagen binding analyses of SPARC proteins by ELISA and SPR

5.4.1 Solid phase binding assays (ELISA)

Solid phase enzyme-linked immunosorbent assay (ELISA) have been used previously to test binding of nematode SPARC to collagens I and IV (Maurer et al., 1997). I carried out similar assays to measure the binding of soluble vertebrate and invertebrate SPARC proteins to immobilised collagens I and IV. These proteins were N-terminally tagged with a His_6 tag and a myc epitope (Section 5.1.1). The EC constructs did not show any appreciable interactions in this assay format (data not shown).

_Drosophila_ SPARC FS-EC showed high affinity, dose-dependent binding to immobilised rat tail collagen I with a half-maximal binding (K_{0.5}) of 0.8 ± 0.2 µM (Figure 5.9, Table 5.4). The K_{0.5} value represents the mean ± SEM from four independent experiments. Human SPARC FS-EC showed comparatively lower collagen binding (K_{0.5} = 7.1 ± 1.4 µM). cSPARC exhibited very little binding to collagen I (Figure 5.9, Table 5.4), with K_{0.5} about an order of magnitude higher than previously reported by Maurer et al. (1997). Binding of the SPARC proteins to the blocking protein, κ-casein, was negligible (Figure 5.9).
Figure 5.9 Solid phase binding assays with recombinant dSPARC, hSPARC, cSPARC and collagen I. Recombinant SPARC FS-EC proteins were added for 3 h at room temperature to 96-well plates coated with collagen I at 10 µg/ml. Shown is a representative of at least two independent experiments. The curves were generated by non-linear least squares fitting of the data by an equation describing single-site binding. Blank: binding of cSPARC to non collagen-coated wells.

Table 5.4 Apparent dissociation constants (mean ± standard error) of SPARC FS-EC NHis-myc proteins binding to immobilised type I collagen (n = 2-4 independent experiments).

<table>
<thead>
<tr>
<th>Protein</th>
<th>$K_{0.5}$ (µM)$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>dSPARC</td>
<td>0.8 ± 0.2</td>
</tr>
<tr>
<td>hSPARC</td>
<td>7.1 ± 1.4</td>
</tr>
<tr>
<td>cSPARC</td>
<td>20, 35</td>
</tr>
</tbody>
</table>

$^a$Concentration of SPARC proteins at 50% of saturable binding
In solid phase assays with immobilised collagen IV, human SPARC FS-EC showed high affinity, dose-dependent binding with a half-maximal binding ($K_{0.5}$) of $2.6 \pm 0.3 \, \mu M$ (Figure 5.10, Table 5.8). This is in agreement with previously reported values of $2 \, \mu M$ (Sasaki et al., 1997). Drosophila SPARC FS-EC showed comparatively lower collagen binding ($K_{0.5} = 12.9 \pm 2.6 \, \mu M$). cSPARC FS-EC exhibited very little binding to collagen IV (Figure 5.10, Table 5.5), with $K_{0.5}$ about an 10-20 fold higher than previously reported by Maurer et al. (1997).

![Graph showing binding curves for different proteins](image)

**Figure 5.10 Solid phase binding assays with recombinant dSPARC, hSPARC, cSPARC and collagen IV.** Recombinant SPARC FS-EC proteins were added for 3 h at room temperature to 96-well plates coated with collagen IV at 10 $\mu$g/ml. Shown is a representative of at least two independent experiments. The curves were generated by non-linear least squares fitting of the data by an equation describing single-site binding. Blank: binding of cSPARC to non-collagen-coated wells.

**Table 5.5 Apparent dissociation constants (mean ± standard error) of SPARC FS-EC NHis-myc proteins binding to immobilised type IV collagen (n = 2-4 independent experiments).**

<table>
<thead>
<tr>
<th>Protein</th>
<th>$K_{0.5}$ (µM)$^{a}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>dSPARC</td>
<td>12.9 ± 2.6</td>
</tr>
<tr>
<td>hSPARC</td>
<td>2.6 ± 0.3</td>
</tr>
<tr>
<td>cSPARC</td>
<td>46, 29</td>
</tr>
</tbody>
</table>

$^{a}$Concentration of SPARC proteins at 50% of saturable binding
5.4.2 SPR studies

Solid phase assays are not very reliable for low affinity interactions with fast off-rates, due to the numerous wash steps involved. In our hands, such experiments with SPARC constructs containing only the EC domain did not show much binding above background levels, most likely due to very weak binding. Therefore, I carried out surface plasmon resonance (SPR)/ Biacore assays to measure dissociation constants between SPARC proteins and human collagen I.

BIACORE is a label free method used to study macromolecular interactions in real time. It exploits the phenomenon of SPR which occurs when polarised light undergoes total internal reflection by striking a gold-coated glass prism. One binding partner (ligand) is immobilised on a chip surface while the other (analyte) is flowed across the surface. As the analyte binds to the ligand, an increase in the refractive index occurs due to the accumulation of protein on the chip surface. This change is measured and the result is displayed in a graph of response or resonance units (RU) versus time.

Collagen I was immobilised on the surface of CM5 sensor chips (2331 RU of collagen I). SPARC FS-EC and EC proteins were then passed over the sensor chip surface at concentrations ranging of 0 to 40 µM. The qualitatively observed rapid rates of association and dissociation indicate weak interactions (Figure 5.11A), hence interactions were measured under steady-state conditions and analysed by fitting the plateau values at equilibrium to a simple 1:1 Langmuir binding model (Figure 5.11B). dSPARC FS-EC bound collagen I with a half-maximal binding ($K_{0.5}$) of 14.6 ± 1.4 µM (Figure 5.11). As expected, dSPARC EC showed comparatively lower collagen binding, $K_{0.5} = 39.5 ± 7.0$ µM (Figure 5.12).
Figure 5.11 SPR (Biacore) analysis of dSPARC FS-EC NHis-myc and collagen I interaction. 
A. The sensorgrams were obtained by dSPARC concentrations ranging from 0-40 µM. The y-axis shows the response difference i.e. the difference between experimental and control flow cells, in resonance units (RU). 
B. Plot of the equilibrium binding response versus concentration. Data were fitted by a simple 1:1 Langmuir binding model to derive dissociation constants (K_{0.5}).
Figure 5.12 SPR (Biacore) analysis of dSPARC EC NHis-myc and collagen I interaction. A. The sensorgrams were obtained by dSPARC concentrations ranging from 0-40 µM. The y-axis shows the response difference i.e. the difference between experimental and control flow cells, in resonance units (RU). B. Plot of the equilibrium binding response versus concentration. Data were fitted by a simple 1:1 Langmuir binding model to derive dissociation constants ($K_{d}$).
It is possible to carry out kinetic analysis of SPR data, when the association and dissociation rates are small. However, erroneous interpretation of this type of data can be made due to factors such as neglect of mass transport, steric hindrance and/or the possibility of more-complex binding schemes (Schuck and Minton, 1996). Therefore, comparing data from equilibrium and kinetic analyses is a good way to test the validity of the experimental data.

The kinetic evaluation procedure determines association and dissociation constants by fitting the experimental data to a 1:1 interaction model. At equilibrium, association balances dissociation. For a simple bimolecular interaction, \( A + B \leftrightarrow AB \), the equilibrium dissociation constant, \( K_{0.5} = [A][B]/[AB] \), is measured by plotting equilibrium binding level against concentration. Both kinetic and equilibrium analyses were carried out on the dataset derived for hSPARC FS-EC binding to collagen I. For kinetic analysis, the experimental data were fitted by a 1:1 Langmuir binding isotherm \( (A + B \leftrightarrow AB) \) (Figure 5.14). Visual inspection of the fitted curves overlaid on the experimental data indicates good fit. A dissociation constant, \( K_{0.5} \), of 22.7 \( \mu \)M was obtained \( (\chi^2 = 1.9) \). This value is in excellent agreement with that obtained from equilibrium analysis of the data (Figure 5.13), where a \( K_{0.5} \) of 23.9 ± 0.9 \( \mu \)M was obtained. hSPARC EC binding to collagen I yielded a similar \( K_{0.5} \) value of 26.2 ± 0.8 \( \mu \)M (Figure 5.15).
Figure 5.13 Kinetic SPR analysis of hSPARC FS-EC NHis-myc and collagen I interaction. The sensorgrams were obtained with hSPARC concentrations ranging from 0-40 µM. The y-axis shows the response difference i.e. the difference between experimental and control flow cells, in resonance units (RU). The coloured traces are the experimental data and the grey traces are best fits of the kinetic model. Data analysis performed using the BIAevaluation 4.1.1 software.
Figure 5.14 SPR (Biacore) analysis of hSPARC FS-EC NHis-myc and collagen I interaction. 
A. The sensorgrams were obtained with hSPARC concentrations ranging from 0-40 µM. The y-axis shows the response difference i.e. the difference between experimental and control flow cells, in resonance units (RU). B. Plot of the equilibrium binding response versus concentration. Data were fitted by a simple 1:1 Langmuir binding model to derive dissociation constants (K_0.5).
Figure 5.15. SPR (Biacore) analysis of hSPARC EC NHis-myc and collagen I interaction. A. The sensorgrams were obtained with hSPARC concentrations ranging from 0-40 µM. The y-axis shows the response difference i.e. the difference between experimental and control flow cells, in resonance units (RU). B. Plot of the equilibrium binding response versus concentration. Data were fitted by a simple 1:1 Langmuir binding model to derive dissociation constants ($K_{d}$).
cSPARC FS-EC showed very weak binding to collagen I (compared to other invertebrate SPARCs) with an estimated $K_{0.5}$ value of $109 \pm 37 \, \mu M$ (Figure 5.16). These data do not allow reliable analysis since the curve does not plateau at the highest concentration used (20 µM). The interaction of collagen I with cSPARC EC would be weaker still so this was not tested.

**Figure 5.16** SPR (Biacore) analysis of cSPARC FS-EC NHis-myc and collagen I interaction. A. The sensorgrams were obtained with cSPARC concentrations ranging from 0-40 µM. The y-axis shows the response difference i.e. the difference between experimental and control flow cells, in resonance units (RU). B. Plot of the equilibrium binding response versus concentration. Data were fitted by a simple 1:1 Langmuir binding model to derive dissociation constants ($K_{0.5}$).
I was unable to study the binding of SPARC proteins to collagen IV using SPR as the CM5 sensor surface immobilised with collagen IV could not be regenerated.
Chapter 6: Discussion

Interaction of SPARC with collagen
6. Discussion

6.1 Biological implications of SPARC – collagen interactions

SPARC is an important matricellular glycoprotein that is highly expressed in tissues undergoing development and remodelling. Several studies have shown SPARC binds to a number of ECM proteins including collagen types I, III, IV and V (Termine et al., 1981, Sage et al., 1989, Mayer et al., 1991, Maurer et al., 1995). The biological significance of SPARC – collagen interactions is poorly understood but can be deduced from knockout mouse models, which exhibit grossly normal phenotype but have fragile skin and develop cataracts (Norose et al., 1998). Tissues from SPARC-null mice contain less collagen than those from WT mice (Bradshaw et al., 2003) in part due to abnormal ECM assembly, indicating that SPARC is involved in the supramolecular assembly and stability of both fibrillar and network-forming collagen. Studies in invertebrates have showed that SPARC is required for collagen IV deposition in the BM during embryonic development. Defects in collagen deposition have also been observed in mammals, in the absence of SPARC expression. Based on these observations, it has been proposed that SPARC acts in concert with molecular chaperones such as Hsp47 to stabilise the collagen triple helix and restrict the lateral aggregation of procollagen molecules prior to export from the ER (Martinek et al., 2007). Trombetta-Esilva and Bradshaw (2012) have suggested a hypothetical model of cellular mechanisms of SPARC which is shown in Figure 6.1.
Figure 6.1 Function of SPARC in procollagen deposition and assembly. SPARC is secreted from the cell bound to procollagen, thus preventing procollagen binding to cell-surface receptors such as DDR2. In the absence of SPARC, procollagen molecules fail to accumulate into fibrils as they get tethered to the cell surface, resulting in less total collagen and thinner collagen fibrils. From Trombeta-Esilva and Bradshaw (2012).

It has been hypothesised that SPARC and related proteins such as SC1/hevin can compensate functionally for each other in vertebrates (Brekken and Sage, 2000). SC1 has been shown to bind collagen (Hambrock et al., 2003) and the collagen binding residues are conserved in both SPARC and SC1. However, SPARC/SC1 double knockout mice show similar phenotype to SPARC-null mice (Barker et al., 2005), indicating that SPARC is more important than SC1 in collagen assembly and maturation.

Invertebrate SPARCs also bind collagens through their EC domain but lacking helix αB, they have a shorter connection between helices αA and αD (Appendix 1). Therefore, it can be hypothesised that invertebrate SPARCs may exist in an activated form i.e. they do not require proteolytic cleavage of helix αC to make the collagen binding site on helix αA accessible (Sasaki et al., 1998). To test this hypothesis, I expressed vertebrate and
invertebrate SPARC to be able to perform structural studies and investigate SPARC binding to collagen.

6.2 Structural studies

Structural characterisation of proteins by X-ray crystallography may be highly informative but presents two major bottlenecks: production of soluble proteins and growth of diffraction-quality crystals (Chang et al., 2007). SPARC proteins were recombinantly expressed in embryonic kidney HEK-293 c18 cells to ensure correct folding and post-translational modifications. The first construct considered for structural studies was the dSPARC EC domain, which contains a potential N-linked glycosylation site at Asn214. Glycosylated proteins are often difficult to crystallise due to heterogeneity and conformational flexibility of the carbohydrate moiety, leading to inhibition of formation of an ordered crystal lattice (Baker et al., 1994). Hence, enzymatic deglycosylation was achieved by PNGase F digestion, which cleaves between the Asn-oligosaccharide bond, and deglycosylation was confirmed by SDS-PAGE. However, the SEC profile showed the sample was not monodisperse (presence of monomer and dimer), and this might explain why the protein failed to yield crystals.

I attempted to crystallise dSPARC FS-EC, also treated with PNGase F to facilitate crystallography. The SEC profile showed the presence of mixed species – presumed monomeric and dimeric. Although deglycosylation was easily achieved on a small-scale analytical preparation of this protein, it did not work equally well for a large scale preparation, even after overnight incubation with PNGase F at relatively high concentration. SDS-PAGE showed the presence of two bands with reduced molecular weight compared with the native protein, indicating that some but not all carbohydrates had been removed. Following PNGase F digestion, the protein was incubated with lectins, but this treatment also failed to remove residual carbohydrates. The difference in level of deglycosylation between EC and FS-EC proteins could be due to accessibility of the glycosylation sites. PNGase F cleaves between the innermost GlcNAc and Asn residues, and the cleavage site on the glycan in the FS domain (Asn126) might be less accessible to PNGase F than the one in the EC domain (Asn214).
A dSPARC double mutant, with mutations in the 2 putative glycosylation sites (Asn126 and Asn214 to Gln), also failed to yield crystals.

At this stage, a SPARC homologue from another invertebrate species, *C. elegans*, was considered for structural studies. Some proteins do crystallise with full N-linked glycans (Mesters and Hilgenfeld, 2007), hence crystallisation trials of cSPARC FS-EC and EC were set up in their native states. Both proteins failed to crystallise in their fully glycosylated states.

It may be possible and worthwhile to deglycosylate invertebrate SPARC proteins using other approaches. One strategy is to express the proteins in the presence of the α-mannosidase inhibitor, kifunensine, purify them and then treat them with enzymes such as endoglycosidase F (Endo F). The glycoform profiles of proteins depend on the cell lines used to express these proteins. For example, glycoproteins expressed in insect cells are mainly oligomannans such as Man₉GlcNAc₂ whereas mammalian cell lines produce mainly complex bi-antennary N-glycans, built on a core of Man₃GlcNAc₂. Kifunensine acts downstream of the N-glycosylation biosynthetic pathway, i.e. it allows initial glycosylation and folding but inhibits mannosidase I enzyme in the ER (Elbein, 1991), thus leaving the proteins with high-mannose Man₉GlcNAc₂ oligosaccharides, which are more homogeneous than hybrid glycan structures and susceptible to cleavage by Endo F. To ensure complete deglycosylation, lectin treatment can be performed following enzymatic digestion of SPARC proteins, since commonly used lectins such as concanavalin A (ConA) bind oligomannose-type N-glycans with much higher affinity than complex-type bi-antennary N-glycans (Cummings and Etzler, 2009). These treatments could potentially lead to successful crystallisation of invertebrate SPARC proteins due to removal of heterogeneity.

### 6.3 Functional studies

We felt it was important to compare how vertebrate and invertebrate SPARCs differ in their ability to bind collagens type I and IV. In solid phase and surface plasmon resonance assays, dSPARC binds most strongly to collagen I, followed by hSPARC and
cSPARC. The solid assay yielded dissociation constants for dSPARC and hSPARC FS-EC that are an order of magnitude lower (i.e. tighter binding) than those obtained from SPR studies. This discrepancy may be due to differences in assay formats. Solid phase assays are not ideal for determination of dissociation constants due to surface effects and non-linearity of detection signals. In such assays, high surface density means several adsorbed molecules may be recognised by a multivalent ligand leading to apparent higher affinities. Very early saturation is observed at low ligand concentrations; at higher concentrations, steric hindrances between large antibodies result in non-linearity of detection signals, yielding affinities that can be up to 1000 fold too high (Tangemann and Engel, 1995). Our data also show that some of the dissociation constants are an order of magnitude higher than previously published binding data (Sasaki et al., 1997, Maurer et al., 1997). SPARC EC proteins could not be studied using solid phase assays possibly because in many cases, low affinity binding fails to be detected in these assays due to the long incubation times and several wash steps. In SPR studies, hSPARC FS-EC and EC proteins showed similar binding affinities to collagen I indicating a single binding site, as reported previously (Sasaki et al., 1997). SPR data for dSPARC FS-EC and EC and collagen interactions showed rapid rates of association and dissociation and could not be analysed kinetically. The binding profiles for dSPARC – collagen interactions indicate that the dissociation constants derived might not be best explained by a single-site model; the dissociation phase does not follow a 1:1 interaction model, therefore, the calculated $K_{0.5}$ value might be a combination of $K_{o.5}$ values corresponding to high- and low-affinity binding sites for dSPARC on collagen I.

In solid phase assays with immobilised collagen IV, hSPARC binds tightly with a $K_{0.5}$ of 2.6 µM, which is in agreement with previously reported values (Maurer et al., 1997). dSPARC shows relatively lower binding and cSPARC exhibits very little binding to collagen IV. One of the drawbacks of our study of invertebrate SPARC – collagen interactions is that it involves the use of mammalian collagens I and IV (from rat tail and human placenta, respectively). Unfortunately, collagen from invertebrates is not available commercially. The Collagen Toolkit, developed in the Farndale lab, consists of sets of overlapping collagen-derived-triple-helical peptides (Raynal et al., 2006). This has been used in several studies for mapping ligand binding sites on fibrillar collagens II
and III. No such toolkit exists for heterotrimeric collagens (e.g. type I and IV) due to technical challenges involved in the synthesis of these peptides.

Alanine mutagenesis studies of the collagen binding epitope of hSPARC have previously indicated the presence of five crucial amino acid residues for collagen binding: R149 and N156 in helix αA, L242, M245 and E246 in a loop region connecting the two EF-hands of SPARC (Sasaki et al., 1998). These essential residues are conserved in vertebrates. However, small differences are seen in invertebrates; an N156Q substitution is found in *C. elegans* and a M245L substitution found in *Drosophila*. These minor substitutions should not have a significant effect on the structure and function of the EC domain. However, despite the conservation of 3 of the 5 key collagen binding amino acids, our data shows that *C. elegans* SPARC has a lower collagen binding affinity compared with other SPARCs. Further studies are needed to understand invertebrate SPARC – collagen IV interactions in more detail and it may be possible do so in future if technological advances lead to the development of peptide libraries from heterotrimeric collagens.
Conclusions
And
Future Directions
7. Conclusions and Future Directions

Decorin

Decorin, the prototypical small leucine-rich repeat proteoglycan, has a broad binding repertoire ranging from collagens to growth factors and RTKs. Consequently, it regulates a host of cellular processes in the ECM such as fibrillogenesis and the control of cell proliferation.

Decorin is a major component of connective tissues and has been shown to bind various collagens including types I, II, II and VI (Bidanset et al., 1992, Scott and Orford, 1981, Vogel et al., 1984). Decorin is also known to bind and inhibit the activity of TGF-β, a powerful profibrotic cytokine (Yamaguchi et al., 1990). Decorin deficient mice are viable and grossly normal but have fragile skin with reduced skin tensile strength as a result of abnormal collagen fibrils of irregular diameter (Danielson et al., 1997).

Much work remains to be done to better understand the biological role of decorin in collagen fibrillogenesis and to exploit its potential as a therapeutic agent for various connective tissue and fibrous disorders involving collagen deposition, control and repair such as wound healing, tendon and ligament repair, etc. (Reed and Iozzo, 2002). Future therapies will rely on the Dcn -/- mouse model and double knockout studies to understand the molecular mechanisms through which decorin and other SLRPs operate in maintaining the integrity of connective tissues (Reed and Iozzo, 2002). My work has resolved the controversy about the physiological relevance of the decorin dimer by showing that dimerisation is reversible and that the concave face of decorin is involved alternatively in dimerisation or collagen binding. In addition, I have shown that wild-type and monomeric decorin have similar thermal stabilities. Most studies so far have focussed on only a few SLRPs, especially decorin, biglycan, lumican, and fibromodulin, with research concerning the other SLRPs trailing far behind (Nastase et al., 2014). Biophysical studies including characterisation of the functional oligomeric statuses and
stabilities of these other SLRPs warrant further studies as they help us better understand the roles these molecules play in several disease states.

Among SLRPs, the biological role of decorin in cancer progression has been extensively studied. Numerous studies have shown that decorin acts as a natural anticancer agent so enhancement of decorin gene expression should halt tumour growth (Bi and Yang, 2013). It is anticipated that decorin will garner more interest in the future due to its potential therapeutic and prognostic value. The concave face of decorin seems to be able to bind a myriad of ECM proteins with no obvious structural similarities, but having versatile effects in health and disease. Therefore, my findings detailed in this thesis, that the functional oligomeric state of decorin is monomeric and the concave face of decorin is available for ligand binding, are of importance in understanding core-protein mediated interactions of decorin and can be extended to study other binding partners, such as growth factors and RTKs in the ECM. One of the challenges of future research would be to identify the leucine-rich repeats of decorin that harbour specific bioactivities since these can be used in adjuvant peptide therapies or can be engineered to bind to specific receptors in the treatment of cancer (Neill et al., 2012).

**SPARC**

SPARC is multifunctional matricellular glycoprotein that is able to bind several proteins in the ECM, modulate the activity of growth factors, and alter cell shape due to its counteradhesive properties. High levels of SPARC are observed during development. In adult tissues, SPARC expression is usually restricted to sites of ECM turnover. The SPARC sequence is highly conserved among species, as diverse as *C. elegans* and humans, indicating the importance of this protein in multicellular life (Bradshaw, 2009).

It has been shown that SPARC is required for normal embryonic development in invertebrates. Inactivation of SPARC in *C. elegans* and *Xenopus* leads to embryonic lethality (Fitzgerald and Schwarzbauer, 1998, Purcell et al., 1993). In *Drosophila* embryos, haemocyte-derived collagen IV is not observed in the basal laminae in the
absence of SPARC, leading to decreased basal lamina stability (Martinek et al., 2008). Hence, the role of SPARC – collagen interactions has important consequences for our understanding of development in invertebrates.

The high conservation of collagen binding residues between vertebrate and invertebrate SPARC sequences suggests that all SPARCs bind collagen in a similar manner. However, invertebrate SPARCs do not undergo proteolytic activation similar to human SPARC since they do not contain an inhibitory loop which is present in mammalian SPARC. The studies outlined in this thesis have made some progress in understanding the differences in collagen binding between vertebrate and invertebrate SPARC in terms of binding affinities. Multiple SPARC binding sites along the collagen IV triple helix have been observed under electron micrographs (Mayer et al., 1991) but their precise locations remain to be elucidated. This would require the development of heterotrimeric peptide libraries the synthesis of which has so far proven to be technically challenging.
References
References


Appendix I

Sequence alignments of selected SPARC EC domains. The sequences are of human SPARC (SwissProt P09486), human SPARC-like/hevin (Q14515), Xenopus laevis SPARC (P36378), Drosophila melanogaster SPARC (O97365) and Caenorhabditis elegans SPARC (P34714). Conserved residues are highlighted in yellow and cysteines are in red. Residues involved in calcium and collagen binding to human SPARC are indicated by pink circles and black inverted triangles, respectively. The sequence numbering and α-helices A-G of human SPARC are indicated above the alignment. From (Hohenester et al., 2008)
Appendix II

Part of the work presented in this thesis is published in:


The concave face of decorin mediates reversible dimerisation and collagen binding

Mehwaesh Islam, Jayesh Gor, Stephen J. Perkins, Yoshihiro Ishikawa, Hans Peter Bachinger and Erhard Hohenester
The Concave Face of Decorin Mediates Reversible Dimerization and Collagen Binding*

Mehwaesh Islam1, Jayesh Gor5, Stephen J. Perkins6, Yoshihiro Ishikawa6, Hans Peter Bächinger6, and Erhard Hohenester3

From the 1Department of Life Sciences, Imperial College London, London SW7 2AZ, United Kingdom, 6Department of Structural and Molecular Biology, University College London, London WC1E 6BT, United Kingdom, 6Research Department, Shriners Hospital for Children, Portland, Oregon 97239, and 3Department of Biochemistry and Molecular Biology, Oregon Health and Science University, Portland, Oregon 97239

Received for publication, July 23, 2013, and in revised form, October 11, 2013 Published, JBC Papers in Press, October 29, 2013, DOI 10.1074/jbc.M113.504530

Significance:
The crystal structure of decorin, the concave faces of two monomers interact to form a tight dimer.

Conclusion:
Decorin binds collagen as a monomer.

Results:
The decorin dimer in solution is in equilibrium with stable monomers, and mutations on the concave face abolish collagen binding.

Significance:
These findings help resolve the controversy about the functional oligomeric state of decorin.

Decorin, the prototypical small leucine-rich proteoglycan, binds to collagen and thereby regulates collagen assembly into fibrils. The crystal structure of the decorin core protein revealed a tight dimer formed by the association of two monomers via their concave faces (Scott, P. G., McEwan, P. A., Dodd, C. M., Bergmann, E. M., Bishop, P. N., and Bella, J. (2004) Proc. Natl. Acad. Sci. U.S.A. 101, 15633–15638). Whether decorin binds collagen as a dimer has been controversial. Using analytical ultracentrifugation, we determined a dissociation constant of 1.37 ± 0.30 μM for the mouse decorin dimer. Dimerization could be abolished by engineering glycosylation sites into the dimer interface; other interface mutants remained dimeric. The monomeric mutants were as stable as wild-type decorin in thermal unfolding experiments. Mutations on the concave face of decorin abolished collagen binding regardless of whether the mutant proteins retained the ability to dimerize or not. We conclude that the concave face of decorin mediates collagen binding and that the dimer therefore must dissociate to bind collagen.

The small leucine-rich proteoglycans (SLRPs)2 comprise a diverse family of secreted glycoproteins that have in common a core protein consisting of multiple leucine-rich repeats (LRRs) flanked by cysteine-rich cap regions. One or several glycosaminoglycan chains are attached to the canonical LRRs; other family members have acidic regions or are modified by tyrosine sulfation (1, 2). Decorin is the prototypical SLRP. It has a well characterized role in regulating collagen fibrillogenesis (2, 3) and additionally modulates the activity of various growth factors and receptor tyrosine kinases (4). Ultrastructural studies of tissue-derived collagen fibrils have revealed decorin binding sites within the gap region of the D-period (5–7). A unique decorin binding site near the C terminus of the triple helix has been identified using type I procollagen produced in cell culture (8). Decorin inhibits collagen fibrillogenesis in vitro (9) and has a profound effect on the ultrastructure of the resulting fibrils (10). Decorin-deficient mice are viable and grossly normal but have fragile skin due to abnormal collagen fibrils (11). Mice lacking decorin and the related SLRP biglycan have a much more severe skin phenotype (12) and a severely disrupted collagen fibril architecture in the cornea (13).

The crystal structure of the decorin core protein revealed that the 12 LRRs form a curved solenoid; the concave face of the solenoid is a parallel β-sheet, and the convex back consists of irregular loops and single helical turns (14). In this crystal structure, two decorin monomers were found to interact through their concave faces, burying a large amount of decorin surface (see Fig. 1A). A strong tendency of decorin and other SLRPs to form dimers in solution was observed in several biophysical studies (15–17). Some even claimed that folded monomeric decorin cannot exist in solution (18), whereas others concluded that the crystallographic decorin dimer is an artifact (19). To complicate matters further, mutagenesis (20–22) and molecular modeling studies (23, 24) have implicated the concave face, which is largely buried in the decorin dimer, in collagen binding.

We felt that it was important to resolve the controversy about the oligomeric state of decorin and how it relates to collagen binding. Here, we show that decorin dimerization is relatively weak and reversible and that mutants that are stable monomers in solution can be obtained. Mutations on the concave face of decorin abolished collagen binding regardless of whether they disrupted the dimer or not. Thus, the same region of decorin...
mediates dimerization and collagen binding, and the decorin dimer must dissociate to bind collagen.

**EXPERIMENTAL PROCEDURES**

**Expression Constructs**—DNA coding for residues 45–354 of mouse decorin (UniProt P28654) was amplified by PCR from a full-length cDNA clone (OriGene) and inserted into a modified pCEP-Pu vector (25). After cleavage of the vector-encoded BM-40 signal peptide, vector-encoded APLA and AAAHHHHHHH sequences are present at the N and C termini of the mature protein, respectively. The mutations were introduced by overlap extension PCR. All expression constructs were verified by sequencing.

**Protein Production**—The proteins were produced in human embryonic kidney HEK293 c18 cells (ATCC). The cells were grown at 37 °C with 5% CO2 in Dulbecco’s modified Eagle’s medium/F-12 (Invitrogen) containing 10% fetal bovine serum, 2 mM glutamine, 10 units/ml penicillin, 100 μg/ml streptomycin, and 250 μg/ml Geneticin. The cells were transfected with the pCEP-Pu expression plasmid using FuGENE (Roche Diagnostics) and selected with 1 μg/ml puromycin (Sigma). Confluent cells in a HYPERFlask (Corning) were washed twice with phosphate-buffered saline (PBS; Invitrogen) and incubated in serum-free medium for 3–4 weeks with weekly medium exchanges. The pooled and filtered conditioned medium was loaded onto a 5-ml HisTrap column (GE Healthcare) using an ÄKTA Purifier (GE Healthcare). The protein was eluted with 300 mM imidazole in PBS, concentrated using a Vivaspin centrifugal device (Sartorius), and further purified on a Superdex 200 16/60 size exclusion chromatography column (GE Healthcare) with Tris-buffered saline (TBS; 20 mM Tris, 150 mM NaCl, pH 7.5) as the running buffer. The fractions containing pure protein were pooled and concentrated to 2–3 mg/ml, and aliquots were flash frozen in liquid nitrogen. All experiments were performed with freshly thawed proteins. For analytical purposes, the N-linked glycan was removed by peptide N-glycosidase F treatment under denaturing conditions according to the manufacturer’s protocol (New England Biolabs).

**Analytical Size Exclusion Chromatography with Laser Light Scattering**—Wild-type and mutant decorin samples at a concentration of 3 mg/ml (83 μM) were injected onto a Superdex 200 10/30 column (GE Healthcare) connected to a 1260 Infinity HPLC (Agilent Technologies). The running buffer was TBS, and the flow rate was 0.2 ml/min. Light scattering and refractive index changes were monitored using in-line Wyatt Mini Dawn and Optilab T-rEX detectors (Wyatt Technology Corp.). The data were analyzed with the Wyatt ASTRA V software using dn/dc values of 0.185 and 0.145 ml/g for the polypeptide and polysaccharide fractions of the glycoproteins, respectively. Each consensus site for N-linked glycosylation was assumed to add 2 kDa of molecular mass. The mass of the polypeptide fraction of the glycoproteins was determined by the three-detector method (26) using an extinction coefficient of 24,961 M−1 cm−1 for the decorin protein.

**Analytical Ultracentrifugation**— Sedimentation velocity experiments were performed at 20 °C using a Beckman XL-I analytical ultracentrifuge at a rotor speed of 40,000 rpm. The instrument was equipped with an eight-hole AnTi50 rotor with double sector cells with column heights of 12 mm. Sedimentation was monitored using absorbance (280 nm) and interference optics. Decorin proteins were dialyzed extensively against PBS and studied at concentrations ranging from 0.028 (0.77 μM) to 3.4 mg/ml (94 μM). The sedimentation boundaries were analyzed using direct boundary Lamm fits using the program SEDFIT (version 14.1) (27). A partial specific volume of 0.7289 ml/g was calculated from the amino acid and carbohydrate content. The buffer density and viscosity were taken to be 1.00543 g/ml and 1.02 centipoises, respectively, based on theoretical values provided by the program SEDNTERP. The continuous c(s) size distribution algorithm assumes that all species have the same friction ratio f/f0 in each fit. The final SEDFIT analyses used a fixed resolution of 200 and optimized the c(s) fit by floating the meniscus, bottom of the cell, base line, and f/f0 ratio until the overall root mean square deviation and visual appearance of the fits were deemed satisfactory. The relative amounts of monomer and dimer were derived using the c(s) integration function. The dimer dissociation constants were obtained by fitting the ratio of monomer and dimer with Equation 1 using SigmaPlot 12.0 software (Systat Software Inc.).

\[
y = 1 - \frac{K_d^2 + 8K_dx - K_d}{4x} \quad \text{(Eq. 1)}
\]

where y is the dimer fraction, x is the total protein concentration, and Kd is the dissociation constant. For a derivation of this equation, see Benfield et al. (28). HYDROPRO (29) was used to calculate s20,w values for the decorin monomer and dimer. The models were based on the crystal structure of dimeric bovine decorin core (14). Biantennary oligosaccharide chains, each consisting of a GlcNAc2Man3 core and two GlcNAc-Gal NeuNAc antennae (30), were added at each of the four predicted glycosylation sites of mouse decorin, and the hydration shell was represented by an atomic element radius of 0.31 nm for all atoms (29).

**Differential Scanning Calorimetry**—The experiments were performed using a Calorimetry Systems Nano III calorimeter. Wild-type and mutant decorin samples at a concentration of 3 mg/ml (83 μM) were dialyzed extensively against PBS. 1-ml aliquots of sample and dialysis buffer were degassed for 15 min. Following an equilibration period of 10 min, initial scans from 5 to 20 °C were repeated until a stable base line was obtained. Scans were then performed from 5 to 65 °C at a heating rate of 1 °C/min.

**Differential Scanning Fluorimetry**—The experiments were performed using a Stratagene Mx3005p qPCR instrument essentially as described (31). 10-μl aliquots of wild-type and mutant decorin in PBS at a concentration of 0.362 mg/ml (10 μM) were mixed with 10 μl of SYPRO Orange solution (Invitrogen) diluted 1:250 and heated from 25 to 95 °C at a heating rate of 1 °C/min. The excitation wavelength was 492 nm, and fluorescence was monitored at 610 nm.

**Collagen Fibrillogenesis Assay**—A 1.05 mg/ml stock solution of mouse type I collagen (isolated from tendon, treated with pepsin, and precipitated using NaCl) in 50 mM acetic acid was neutralized at 4 °C by diluting it 33-fold with 150 mM sodium phosphate, 150 mM NaCl, pH 7.8 and immediately placed into a


The R151E mutation in LRR5 (Arg-127 in the bovine decorin structure) reverses a charge in the network of polar interactions that accounts for most of the interface, the Q61N mutation (Gln-37 in the bovine decorin structure) introduces a consensus site for N-linked glycosylation into LRR1, and the Y130N mutation (Tyr-106 in the bovine decorin structure) introduces a consensus site for N-linked glycosylation into LRR4. Analysis by the NetNGlyc server predicted glycosylation potentials of 0.60 and 0.72 for asparagines at positions 61 and 130, respectively, indicating a high probability that the engineered glycosylation sites would be modified.

Wild-type mouse decorin and all four mutants were obtained with good yields from episomally transfected HEK293 cells and purified to homogeneity (Fig. 1B and data not shown). The wild-type protein (calculated molecular mass, 36.2 kDa) ran as a single band of ~43 kDa on reducing SDS-PAGE, consistent with the presence of four consensus sites for N-linked glycosylation in the mouse decorin sequence. The Q61N mutant ran as a broader band at higher molecular mass, demonstrating that the engineered glycosylation site at position 61 is modified by a glycan. In contrast, the electrophoretic mobility of the Y130N mutant resembled more closely that of the wild-type protein, indicating that the engineered glycosylation site at position 130 is either unmodified or that the attached glycan is not detectable by SDS-PAGE. Removal of the N-linked glycans by peptide N-glycosidase F digestion resulted in identical sharp bands at ~30 kDa for the wild-type construct and the two mutants with engineered glycosylation sites (Fig. 1B).

**Oligomeric States of Wild-type and Mutant Decorin**—To determine the oligomeric states of mouse decorin and its dimer interface mutants, we first used size exclusion chromatography with multangle light scattering (SEC-MALS). Wild-type mouse decorin injected at 3 mg/ml concentration (83 μm) eluted in an asymmetric peak with a pronounced tail (Fig. 2). The molecular mass of the protein without modifications was determined to be 64.4 kDa (Table 1). This value is much closer to the calculated mass of a dimer (72.4 kDa) than that of a monomer (36.2 kDa). The dimer appears to dissociate quite readily, however, giving rise to an asymmetric peak and an average mass that is slightly lower than that of a dimer. The molecular mass of the glycoprotein (i.e. protein plus carbohydrate modifications) was determined to be 83.5 kDa (Table 1), which is in excellent agreement with the reported mass of 84.6 kDa for dimeric bovine decorin core glycoprotein (17). The elution profiles and molecular masses of the Y51A/R52A/Q54A and R151E mutants resembled those of the wild-type protein (Fig. 2 and Table 1), indicating that these mutations had not disrupted the mouse decorin dimer. In contrast, the Q61N and Y130N mutants eluted later than wild-type decorin and displayed symmetric peak profiles with molecular masses closely matching those of a monomer (Fig. 2 and Table 1). For the Q61N mutant, the disruption of the dimer could be attributed unequivocally to an engineered glycan as there is clear evidence for an additional modification in SDS-PAGE (Fig. 1B). For the Y130N mutant, the presence of a disruptive glycan could only be inferred from the observation that this mutant is monomeric. It is possible, although unlikely, that introduction of an unmodified asparagine at position 130 disrupts the decorin dimer. Because our
objective was to obtain monomeric decorin mutants, we did not further investigate the presumed modification in the Y130N mutant.

The SEC-MALS experiment suggested that wild-type mouse decorin might exist in a concentration-dependent monomer-dimer equilibrium. To investigate the monomer-dimer equilibrium quantitatively, we used analytical ultracentrifugation. Using atomic models based on the crystal structure of bovine decorin (14) with appropriate carbohydrate modifications, we calculated $s_{20,w}^0$ values of 3.0 and 4.7 S for monomeric and dimeric mouse decorin (for details, see “Experimental Procedures”). We collected sedimentation velocity data at seven concentrations of wild-type decorin ranging from 0.028 to 3.4 mg/ml (0.77–94 μM) (Fig. 3A). The $c(s)$ distributions derived from these data are characterized by two peaks, one at 3.6 S and one at 4.7–5.3 S, the relative proportions of which varied with the protein concentration (Fig. 3B). These peaks were interpreted to correspond to monomers and dimers, respectively, and their relative areas were used to derive a dissociation constant of $1.37 \pm 0.30 \mu M$ for the mouse decorin dimer (Fig. 3C). Analogous experiments with the Y51A/R52A/Q54A and R151E mutants yielded comparable dimer dissociation constants of 2.3 and 0.47 ± 0.09 μM, respectively (Fig. 4).

Stability of Wild-type and Mutant Decorin—The experiments described so far show that wild-type mouse decorin exists in a monomer-dimer equilibrium and that mutants can be obtained that behave as pure monomers in SEC-MALS. To quantify the thermal stabilities of selected proteins, we used differential scanning calorimetry. Wild-type mouse decorin and the monomeric Q61N mutant unfolded in single transitions with melting temperatures of 50 and 52 °C, respectively (Fig. 5A), which compare with a reported melting temperature of 46 °C for bovine decorin (18). Unfolding was only partially reversible as the signals on a second up-scan were reduced by 90% (data not shown). To extend the analysis to the remaining mutants, we used differential scanning fluorimetry, which monitors thermal unfolding using a hydrophobic dye and requires only small amounts of protein (31). The unfolding curves of all decorin proteins were very similar with inflection points ranging from 49 to 51 °C (Fig. 5B). As in the differential scanning calorimetry experiment, the monomeric mutants with engineered glycosylation sites (Q16N and Y130N) were marginally more stable than wild-type decorin. A positive effect of glycans on protein stability is well documented and frequently exploited in the pharmaceutical industry (32).

Collagen Binding by Wild-type and Mutant Decorin—An important aim of the present study was to resolve the controversy whether decorin binds to collagen as a monomer or as a dimer (14, 19, 24). We first assessed collagen binding indirectly by measuring the inhibition of collagen fibrillogenesis, which is the classic assay for decorin activity (9). Wild-type mouse decorin robustly delayed fibrillogenesis of type I collagen (Fig. 6).

### TABLE 1

Molecular masses of mouse decorin and its mutants as determined by SEC-MALS

<table>
<thead>
<tr>
<th>Protein</th>
<th>Calculated molecular mass</th>
<th>N-Linked glycosylation sites</th>
<th>Peak elution volume</th>
<th>Experimental mass of glycoprotein</th>
<th>Experimental mass of polypeptide fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>36.2 kDa</td>
<td>4</td>
<td>14.5 ml</td>
<td>83.5 kDa</td>
<td>64.4 kDa</td>
</tr>
<tr>
<td>Y51A/R52A/Q54A</td>
<td>36.2 kDa</td>
<td>4</td>
<td>14.1 ml</td>
<td>85.3 kDa</td>
<td>65.3 kDa</td>
</tr>
<tr>
<td>Q61N</td>
<td>36.2 kDa</td>
<td>5</td>
<td>15.5 ml</td>
<td>52.1 kDa</td>
<td>37.6 kDa</td>
</tr>
<tr>
<td>Y130N</td>
<td>36.2 kDa</td>
<td>4</td>
<td>15.1 ml</td>
<td>51.5 kDa</td>
<td>36.7 kDa</td>
</tr>
<tr>
<td>R151E</td>
<td>36.2 kDa</td>
<td>4</td>
<td>14.1 ml</td>
<td>88.4 kDa</td>
<td>68.4 kDa</td>
</tr>
</tbody>
</table>

* Derived from the refractive index and light scattering signals.

* Derived from the absorbance, refractive index, and light scattering signals (26).
A) as reported previously for bovine and human decorin (10, 13, 14, 19, 22). Of the four dimer interface mutants, only the Y51A/R52A/Q54A mutant delayed fibrillogenesis similarly to wild-type protein. The Q61N, Y130N, and R151E mutants were completely inactive (Fig. 6A). We also attempted to measure collagen binding directly using a solid-phase assay with immobilized type I collagen (22, 33) but were frustrated by high and variable levels of nonspecific binding regardless of the blocking agent (albumin and casein) or detection method used (anti-mouse decorin antibody, anti-His tag antibody, biotinylation, and detection by avidin). Despite these problems, we consistently observed stronger collagen binding by wild-type mouse decorin and the Y51A/R52A/Q54A mutant than by any of the single point mutants (Fig. 6B). These observations corroborate the findings obtained with the more robust fibrillogenesis assay and indicate that decorin residues 61, 130, and 151 are important for collagen binding.

**DISCUSSION**

Decorin and the related SLRPs biglycan, lumican, and fibromodulin play a major role in regulating collagen fibril formation in the extracellular matrix (2, 3), but how they bind to collagen has been unclear. A sterically plausible binding mode involves one or several collagen triple helices interacting with the concave face of the curved SLRP molecule (23, 24), which also happens to be the most highly conserved surface region (14). However, in crystal structures of decorin and biglycan, the concave face forms the interface of a seemingly tight dimer (14, 18), leading to controversy about the physiological relevance of the dimers (16–19). Here, we have resolved this controversy by

---

**FIGURE 3.** Sedimentation velocity analysis of wild-type mouse decorin. Seven concentrations from 0.028 to 3.4 mg/ml were analyzed in PBS at a rotor speed of 40,000 rpm. A, scan boundaries (black) and their fits (red) at the highest and lowest protein concentration. Only every third (3.4 mg/ml) or fifth (0.028 mg/ml) scan is shown for clarity. B, four continuous size distributions obtained from fitting the scan boundaries with the Lamm equation. The peaks assigned to monomeric (m) and dimeric (d) decorin are labeled. C, determination of the dimer dissociation constant. The dimer fractions were obtained by integration of the monomer and dimer peaks in the c(s) distributions. The solid line is a non-linear least square fit of the data by the equation describing a monomer-dimer equilibrium (see “Experimental Procedures”).

**FIGURE 4.** Determination of the dimer dissociation constants of decorin mutants Y51A/R52A/Q54A and R151E by sedimentation velocity analysis. The dimer fractions were obtained by integration of the monomer and dimer peaks in the c(s) distributions. The solid lines are non-linear least square fits of the data by the equation describing a monomer-dimer equilibrium (see “Experimental Procedures”).

---

6A) as reported previously for bovine and human decorin (10, 13, 14, 19, 22). Of the four dimer interface mutants, only the Y51A/R52A/Q54A mutant delayed fibrillogenesis similarly to wild-type protein. The Q61N, Y130N, and R151E mutants were completely inactive (Fig. 6A). We also attempted to measure collagen binding directly using a solid-phase assay with immobilized type I collagen (22, 33) but were frustrated by high and variable levels of nonspecific binding regardless of the blocking agent (albumin and casein) or detection method used (anti-mouse decorin antibody, anti-His tag antibody, biotinylation, and detection by avidin). Despite these problems, we consistently observed stronger collagen binding by wild-type mouse decorin and the Y51A/R52A/Q54A mutant than by any of the single point mutants (Fig. 6B). These observations corroborate the findings obtained with the more robust fibrillogenesis assay and indicate that decorin residues 61, 130, and 151 are important for collagen binding.

**DISCUSSION**

Decorin and the related SLRPs biglycan, lumican, and fibromodulin play a major role in regulating collagen fibril formation in the extracellular matrix (2, 3), but how they bind to collagen has been unclear. A sterically plausible binding mode involves one or several collagen triple helices interacting with the concave face of the curved SLRP molecule (23, 24), which also happens to be the most highly conserved surface region (14). However, in crystal structures of decorin and biglycan, the concave face forms the interface of a seemingly tight dimer (14, 18), leading to controversy about the physiological relevance of the dimers (16–19). Here, we have resolved this controversy by
showing that dimerization is reversible and that the concave face of decorin is involved alternatively in dimerization or collagen binding (Table 2).

Using analytical ultracentrifugation, we determined a dissociation constant of 1.37 μM for the mouse decorin dimer. A similar study of biglycan dimerization determined a dissociation constant of 4.5 μM (calculated from a free energy of association of −7.3 kcal/mol) (15). Thus, at the high concentrations typically used in solution scattering (17, 18) and crystallization experiments (14, 18), decorin and biglycan are dimers, but at plausible physiological concentrations, dimers will dissociate into monomers. In unfolding experiments with decorin and biglycan, denaturation coincides with a transition from (folded) dimer to (unfolded) monomer (17, 18). This finding has been interpreted as evidence that folded monomers cannot exist (18). By engineering glycosylation sites into the dimer interface, we have created decorin mutants (Q61N and Y130N) that remain monomeric at high concentration. The thermal stability of these mutants slightly exceeded that of wild-type decorin, likely due to a commonly observed stabilizing effect of engineered glycans (32). Thus, dimerization clearly is not essential to stabilize the decorin fold. Proteins that are structurally related to decorin and are stable monomers, such as Nogo receptor (34, 35) or LRR domain 3 of Slit (36), exist. Other proteins that dimerize similarly to decorin, such as LRR domain 4 of Slit (37) or AMIGO-1 (38), also exist. In contrast to our findings with decorin, mutation of interface residues in AMIGO-1 affected protein folding and secretion (38). Thus, the possibility remains that some SLRPs do not dissociate into stable monomers, but we believe that this is no longer a tenable scenario for decorin.

**FIGURE 5.** Thermal stabilities of WT and mutant mouse decorin. **A,** unfolding transitions obtained by differential scanning calorimetry at a protein concentration of 3 mg/ml in PBS. The melting temperatures derived from the peak maxima are 50 (WT) and 52 °C (Q61N), respectively. **B,** unfolding transitions obtained by differential scanning fluorimetry at a protein concentration of 0.181 mg/ml in PBS. The melting temperatures derived from the inflection points of the curves are 49 (R151E), 50 (WT and Y51A/R52A/Q54A), and 51 °C (Q61N and Y130N), respectively. Shown is a representative of two independent experiments carried out in triplicate. A.U., arbitrary units.

**FIGURE 6.** Collagen binding by WT and mutant mouse decorin. **A,** inhibition of collagen fibrillogenesis by WT and mutant mouse decorin. Type I collagen (32 μg/ml) was incubated at pH 7.8 and 37 °C, and the turbidity arising from fibril formation was recorded as absorbance at 400 nm. The decorin proteins were added at a concentration of 50 μg/ml. Shown is a representative of three independent experiments. **B,** collagen binding by WT and mutant mouse decorin. Type I collagen was immobilized on microtiter plates and incubated with varying amounts of decorin proteins. Bound decorin proteins were detected as absorbance at 492 nm using an antibody-linked color reaction. The solid lines are fits of the data by an equation describing single site binding. Shown is a representative of three independent experiments carried out in duplicate.

**TABLE 2**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Location of mutated residue(s)</th>
<th>Oligomeric state</th>
<th>Collagen binding</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td></td>
<td>Monomer-dimer</td>
<td>Yes</td>
</tr>
<tr>
<td>Y51A/R52A/Q54A</td>
<td>N-terminal cap</td>
<td>Equilibrium</td>
<td>Yes</td>
</tr>
<tr>
<td>Q61N</td>
<td>LRR1</td>
<td>Monomer</td>
<td>No</td>
</tr>
<tr>
<td>Y130N</td>
<td>LRR4</td>
<td>Monomer-dimer</td>
<td>No</td>
</tr>
<tr>
<td>R151E</td>
<td>LRR5</td>
<td>Equilibrium</td>
<td>No</td>
</tr>
</tbody>
</table>

* Determined by SEC-MALS (Fig. 2) and analytical ultracentrifugation (Figs. 3 and 4).

* Determined by inhibition of collagen fibrillogenesis and solid-phase binding (Fig. 6).
Engineering glycosylation sites into LRR1 (Q61N mutant) and LRR4 (Y130N mutant) or reversing the charge of a key residue in LRR5 (R151E mutant) abolished the ability of decorin to inhibit collagen fibrillogenesis and reduced collagen binding in a solid-phase binding assay. In contrast, a drastic triple mutation in the N-terminal cap region (Y51A/R52A/Q54A) had only a modest effect on collagen binding. Previous mutagenesis studies additionally implicate LRR6 (21) and LRR7 (22) in collagen binding. Thus, the picture that is emerging is that a large part of the concave surface of monomeric decorin may be involved in collagen binding. In agreement with this view, a recent modeling study using the experimentally derived structure of fibrillar type I collagen (39, 40) concluded that the decorin monomer could interact with up to six triple helices at the fibril surface (24).

An alternative interpretation of our results is that the monomeric decorin mutants Q61N and Y130N are inactive because collagen binding requires an intact decorin dimer (14). If this were the case, the R151E mutant, which dimerizes similarly to wild-type decorin, would be expected to bind collagen, but this is not the case. Using Equation 1 and the experimentally determined dimer dissociation constants, we estimate that 50% of wild-type decorin and 34% of the R151E mutant are available as monomers in the fibrillogenesis assay. This modest difference cannot explain the complete loss of collagen binding resulting from the R151E mutation. The simplest explanation is that Arg-151 (and the concave face as a whole) is directly involved in the binding of monomeric decorin to collagen fibrils. The ultimate proof for this model would require a mutant that is monomeric and binds collagen like wild-type decorin. Such mutants may well be elusive given that the dimer interface is formed precisely by the region most likely responsible for collagen binding.

Acknowledgments—We thank Amanda Simmonds and Raj Gandhi for preliminary experiments, Alan Purvis for help with SEC-MALS, Kurt Drickamer for help with calorimetry, Rhian Jones for help with fluorimetry, and Peter Nixon and Karim Magdhaouli for use of the temperature-controlled spectrophotometer.

REFERENCES
protein, a vaccinia virus inhibitor of IKKβ-mediated activation of nuclear factor κB. J. Biol. Chem. 286, 20727–20735