**Structural and functional characterization of the extracellular matrix proteins COMPcc and NtA**

By

Ainsley Alana Carole McFarlane

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry

University of Manitoba

Winnipeg

## Abstract

The extracellular matrix (ECM) is a complex of proteins and carbohydrates that not only provides a structural support between tissues but also functions in a wide variety of cellular activities. ECM proteins are large, complex proteins with many repeating domains. Individual domains can be analyzed for the investigation of structure-function relationships. This thesis focuses on two ECM proteins of interest: N-terminal agrin (NtA) and the coiled-coil domain of cartilage oligomeric matrix protein (COMPcc). Agrin is an important ECM protein involved in postsynaptic differentiation at the neuromuscular junction, mediated by binding in the NtA domain. In agrin, the NtA domain is followed by nine follistatin-like (FS) domains. Structural studies showed a novel interdomain disulfide bridge between the NtA and first FS domain in agrin. This disulfide bridge compensates for a seven residue splice insert in the C-terminus of NtA, suggesting that the interdomain disulfide bond may be necessary for the proper folding of agrin.

COMP is another important ECM protein that is found in cartilage, tendon, and ligament. It is a homopentamer held together by disulfide bonds in the central coiled-coil oligomerization domain. Previous structural studies demonstrated that COMPcc forms a pentameric α-helical coiled-coil structure containing a 73Å-long cavity with a diameter of 2–6Å that is capable of binding hydrophobic compounds. This binding capacity of COMPcc was investigated and the high-resolution crystal structures of COMPcc in complex with five naturally-occurring fatty acids were solved. Additionally, the binding properties of COMPcc in solution were investigated through the use of fluorescence spectroscopy. Both the x-ray crystallographic and solution data reveal that binding favourability of fatty acids to COMPcc is driven by length of

the methylene tail and degree of unsaturation. These results suggest the possibility of COMPcc to be used in targeted drug delivery systems.

## Acknowledgements

Financial support for this research project was provided by the Faculty of Science Studentship and the University of Manitoba Graduate Fellowship.

I would like to thank my supervisor, Dr. Jörg Stetefeld, without whom this would not have been possible. I appreciate all the support you have given me over the past four years and I value all that you have taught me. Thank you for being a wonderful teacher, mentor, advisor, and friend.

Thank you to my advisory committee Drs. Deborah Court, Sean McKenna, and John Sorensen, and to my external examiner Dr. Heiko Heerklotz.

I would also like to acknowledge the help and support from the other members of the Stetefeld research group. A heartfelt thanks to Drs. George Orriss, Markus Meier, Trushar Patel, and Simone Karrasch. I have learned so much from each of you, thank you for sharing your knowledge and experience with me. Thank you as well to the other students of our research group, past and present, that I have had the pleasure of working with: Kelly, Alex, Mina, Laura, and Natalie. I had a lot of fun working with you all and you helped make this a very enjoyable experience for me.

Finally, I would like to express my gratitude to my family for all of their love and support through the years. To my daughter, Rachel, who came into this world during my PhD program, you are the light of my life and I am grateful everyday that I get to be your Mom. And last but not least, my husband Michael, I am so glad we were able to go on this journey together. Thank you for being with me every step of the way.

## Dedication

This thesis is dedicated to my husband, Michael. Thank you for being my partner in life, love, and chemistry.

## Contents of Thesis

This dissertation comprises of the following manuscripts published/submitted for publication in peer-reviewed scientific journals. The list below is provided as it is arranged in this dissertation:

Chapter 2:

1. McFarlane, A. A., Orriss, G. L., and Stetefeld, J. (2009) The use of coiled-coil proteins in drug delivery systems, *Eur. J. Pharmacol. 625*, 101-107.

Chapter 3:

2. McFarlane, A.A., Khajehpour, M., and Stetefeld, J. (2011) The coiled-coil domain of Cartilage Oligomeric Matrix Protein binds fatty acids via a ring of conserved glutamines. (In preparation for submission to *The Journal of Biological Chemistry* – 2011).

Chapter 4:

3. McFarlane, A.A., and Stetefeld, J. (2009) An interdomain disulfide bridge links the NtA and first FS domain in agrin. *Protein Sci. 18*, 2421-2428.

**Contributions of the Authors**

All experimental works were performed by the author, under the supervision of Dr. J. Stetefeld. The manuscript for Chapter 2 was prepared by the author, with contributions and revisions from Drs. G.L. Orriss and J. Stetefeld. The fluorescence spectroscopy experiments performed in Chapter 3 were conducted under the supervision of Dr. M. Khajehpour. The manuscript for this chapter was prepared by the author, with contributions and revisions from

Drs. M. Khajehpour and J. Stetefeld. The manuscript for Chapter 4 was prepared by the author with contributions and revisions from Dr. J. Stetefeld.

## Table of Contents

## List of Tables

## List of Figures

## List of Abbreviations

| | |
|---|---|
| © | Copyright sign |
| $^{o}$ | Degrees |
| $^{o}$C | Degrees Celsius |
| 3D | 3-Dimensional |
| Å | Angstrom |
| AChR | Acetylcholine Receptor |
| Ala (A) | Alanine |
| Arg (R) | Arginine |
| Asn (N) | Asparagine |
| Asp (D) | Aspartic Acid |
| BM-40 | 40 kDa Calcium-binding Basement Membrane Protein |
| Ca | Calcium |
| CAD | Cadherin Domain |
| CAPB | Cocoamidoprylbetaine |
| CNBr | Cyanogen Bromide |
| COMP | Cartilage Oligomeric Matrix Protein |
| COMPcc | Coiled-coil domain of Cartilage Oligomeric Matrix Protein |
| CPA | Cis-Parinaric Acid |
| Cys (C) | Cysteine |
| Da | Dalton |
| DNA | Deoxyribonucleic Acid |

| | |
|---|---|
| DTT | Dithiothreitol |
| ECM | Extracellular Matrix |
| EF | Epidermal Growth Factor-like domain |
| EGF | Epidermal Growth Factor |
| FA | Fatty Acid |
| FABP | Fatty Acid-Binding Protein |
| Fig. | Figure |
| FN | Fibronectin Domain |
| FOLN | Follistatin N-terminal Domain-like |
| FS | Follistatin |
| Gln (Q) | Glutamine |
| Glu (E) | Glutamic Acid |
| Gly (G) | Glycine |
| h | Hour |
| HCl | Hydrochloric Acid |
| HDL | High-Density Lipoprotein |
| His (H) | Histidine |
| ICP-OES | Inductively-Coupled Plasma Optical Emission Spectroscopy |
| Ig | Immunoglobulin |
| Ile | Isoleucine |
| kcal | Kilocalorie |
| $K_d$ | Dissociation Constant |

| | |
|---|---|
| kDa | Kilodalton |
| kEv | Kiloelectron volt |
| LamG | Laminin G-like domain |
| LDL | Low-Density Lipoprotein |
| LE | Laminin Epidermal Growth Factor-like domain |
| Leu | Leucine |
| Lys (K) | Lysine |
| MALDI-TOF | Matrix-Assisted Laser Desorption Ionization-Time of Flight |
| Met (M) | Methionine |
| mg | Milligram |
| min | Minute |
| mL | Millilitre |
| MMP | Matrix Metalloproteinase |
| mol | Mole |
| mRNA | Messenger Ribonucleic Acid |
| MS | Mass Spectrometry |
| nL | Nanolitre |
| nm | Nanometre |
| NMJ | Neuromuscular Junction |
| NtA | N-terminal Agrin |
| N-TIMP-1 | N-terminal Proteinase Inhibition domain of the Tissue Inhibitor of Metalloproteinases-1 |

| | |
|---|---|
| OB | Oligonucleotide/oligosaccharide Binding |
| PBS | Phosphate Buffered Saline |
| PDB | Protein Database |
| Phe (F) | Phenylalanine |
| Pro (P) | Proline |
| RHCC | Right-handed coiled coil |
| rms | Root Mean Square |
| RNA | Ribonucleic Acid |
| RP-HPLC | Reverse-phase High Performance Liquid Chromatrography |
| r.p.t. | Residues per turn |
| SARS | Severe Acute Respiratory Syndrome |
| SDS-PAGE | Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis |
| Ser (S) | Serine |
| SFA | Saturated fatty acid |
| T3 | Calcium-binding domain |
| TC | C-terminal domain |
| TFA | Trifluoroacetic Acid |
| Thr (T) | Threonine |
| TIMP | Tissue Inhibitor of Metalloproteinases |
| $T_m$ | Midpoint Transition Temperature for Unfolding |
| Trp (W) | Tryptophan |
| TSP | Thrombospondin |

| | |
|---|---|
| Tyr (Y) | Tyrosine |
| Val (V) | Valine |
| VASP | Vasodilator-stimulated Phosphoprotein |
| VLP | Virus-like Particle |
| wt | Wild-type |

**List of Copyrighted Material for which Permission has been Obtained**

**Figure 1.1** – Reproduced with permission from Nature Publishing Group.

Chen, F. H., Rousche, K. T., and Tuan, R. S. (2006) Technology Insight: adult stem cells in cartilage regeneration and tissue engineering, *Nat. Clin. Pract. Rheumatol. 2*, 373-382.

http://s100.copyright.com/CustomerAdmin/PLF.jsp?lID=2011081_1313857734969

**Figure 1.2** – Reproduced with permission from The American Society for Biochemistry and Molecular Biology, Inc.

Morgelin, M., Heinegard, D., Engel, J., and Paulsson, M. (1992) Electron microscopy of native cartilage oligomeric matrix protein purified from the Swarm rat chondrosarcoma reveals a five-armed structure, *J. Biol. Chem. 267*, 6137-6141.

**Chapter 2** – Reproduced with permission from Elsevier.

McFarlane, A. A., Orriss, G. L., and Stetefeld, J. (2009) The use of coiled-coil proteins in drug delivery systems, *Eur. J. Pharmacol. 625*, 101-107.

https://s100.copyright.com/CustomerAdmin/PLF.jsp?lID=2011050_1305248713538

**Chapter 4** – Reproduced with permission from John Wiley and Sons.

McFarlane, A. A., and Stetefeld, J. (2009) An interdomain disulfide bridge links the NtA and first FS domain in agrin, *Protein Sci. 18*, 2421-2428.

http://s100.copyright.com/CustomerAdmin/PLF.jsp?lID=2011050_1305247901210

# Chapter 1: Introduction

## 1.1. Extracellular matrix proteins

The extracellular matrix (ECM) is the structural scaffold found in all tissues and organs. It consists of a network of proteins and carbohydrates that are secreted by cells in the matrix. Along with its structural functions, it also plays a role in tissue morphogenesis and migration, cell differentiation, signal transduction, and homeostasis (*1, 2*). Tissue specificity arises from variations in the topology of the ECM and its components, such as post-translational modifications and protein isoforms (*3*). Genetic defects affecting ECM proteins lead to a wide variety of diseases, including osteogenesis imperfecta, Stickler syndrome, Marfan syndrome, and muscular dystrophy (*4, 5*). The significance of this complex molecular network is evident, and increasing the understanding of its intricate functions is of great importance.

The ECM components can be divided into two main classes: fibrous proteins and proteoglycans. The major fibrous proteins found in the ECM include the collagens, fibronectins, elastins, and laminins. Proteoglycans are divided into three main families: the small leucine-rich proteoglycans, modular proteoglycans (such as perlecan and agrin), and cell surface proteoglycans (such as syndecans and glypicans) (*1*). The composition of the components in the ECM and their architecture is tissue-specific. For example, in articular cartilage, the dry weight of the ECM is composed of 60 – 86% collagens, 15 – 40% proteoglycans, and several other non-collagenous proteins such as fibronectin, fibromodulin, and cartilage oligomeric matrix protein (COMP) (Fig. 1.1). The biomechanical properties of articular cartilage such as its tensile strength and compression resistance is dependent upon the elaborate network of ECM constituents (*6*).

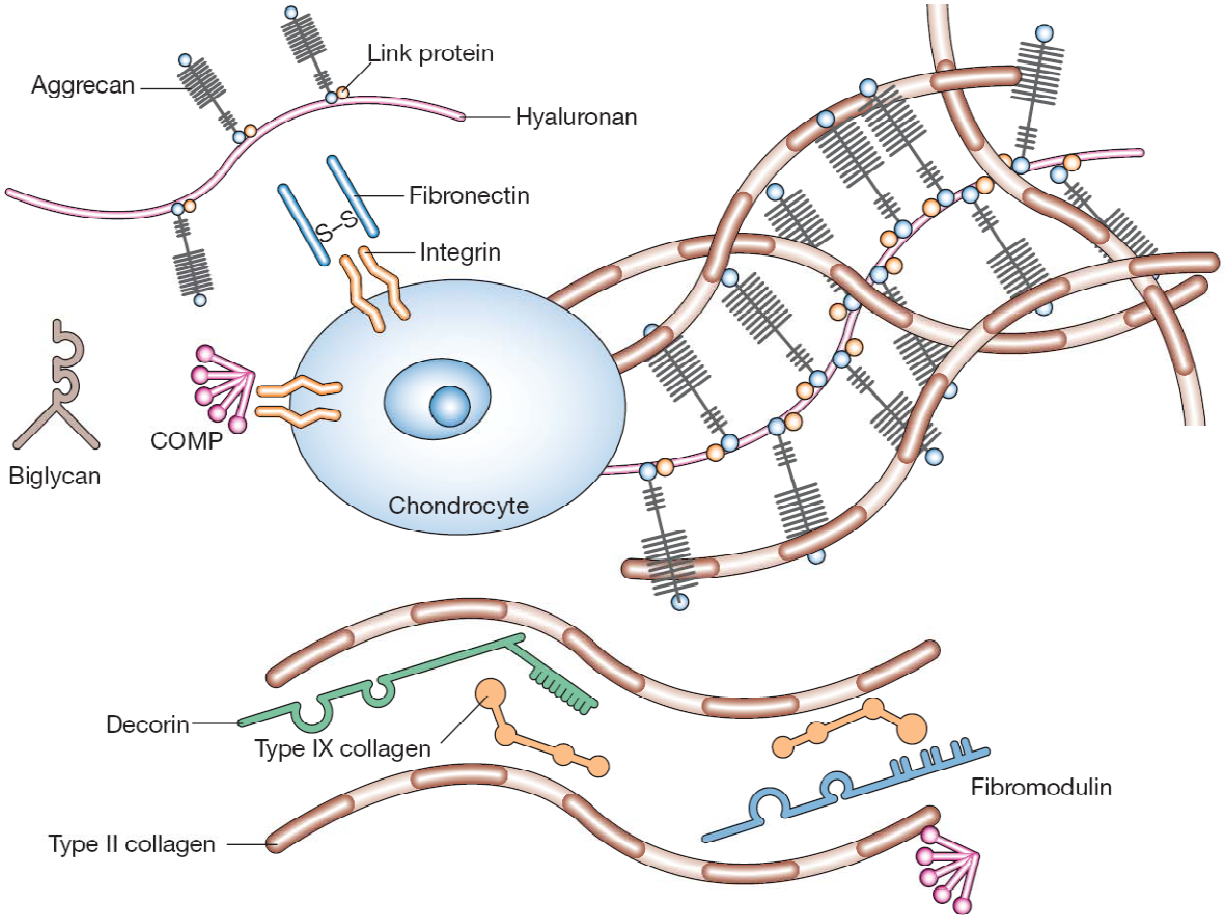


**Fig. 1.1. The extracellular matrix components of articular cartilage.** The main components of the articular cartilage ECM are the collagens, proteoglycans (mainly aggrecan but also smaller proteoglycans including biglycan, decorin, and fibromodulin), and the non-collagenous proteins such as link protein, fibronectin, and COMP (*6*). Genetic defects in the ECM proteins of cartilage lead to diseases such as early-onset osteoarthritis, Stickler syndrome, pseudoachondroplasia, and multiple epiphysial dysplasia (*4*). For copyright permission please see: http://s100.copyright.com/CustomerAdmin/PLF.jsp?lID=2011081_1313857734969

ECM proteins are mostly large, multidomain proteins that are composed of frequently repeated modules. They are also often difficult to solubilise, have abundant post-translational modifications, and are frequently regulated by alternative splice events, all of which makes them difficult to study. Little is known about the function of the multiple domains of ECM proteins and their biological interactions. Due to their modular nature, however, individual domains of ECM proteins can be investigated to characterize their structural and functional properties (*7*). Clearly these domains play an important role in the overall biochemistry of the ECM and this is an exciting field in which there is a great deal of knowledge that is lacking. This thesis aims to elucidate some of these knowledge gaps by focusing on two important ECM proteins: cartilage oligomeric matrix protein (COMP) and agrin. Domains of interest of these two proteins were studied using a multi-disciplinary approach combining various structural biology techniques. The data gathered in this thesis contributes to the overall understanding of the ECM and adds a few more pieces to this scientific puzzle.

### 1.2. Cartilage oligomeric matrix protein (COMP)

Cartilage oligomeric matrix protein (COMP) is a non-collagenous glycoprotein that is found in cartilage (*8*), tendon (*9, 10*), and ligament (*11*). It belongs to the thrombospondin (TSP) gene family of glycoproteins and has a bouquet-like shape with five 87 kDa arms extending out from a cylindrical assembly domain, as revealed by electron microscopy (Fig. 1.2) (*12*). There are four known TSP's (TSP1-4) with TSP3 and TSP4 forming pentameric structures similar to COMP (*13-15*) , while TSP1 and TSP2 form trimeric structures (*16*). COMP is a multidomain protein containing an N-terminal coiled-coil oligomerization motif (cc), followed by four epidermal growth factor (EGF)-like domains (EF), seven calcium binding domains (T3) and a C-

terminal globular domain (TC) (Fig. 1.3). Hereditary diseases of bone development such as pseudoachondroplasia and multiple epiphysial dysplasia are known to be caused by mutations in the COMP gene affecting the conformation of the T3 domains (*17-20*).

The crystal structure of the recombinant coiled-coil domain of COMP (COMPcc) comprising residues 27–72 revealed a 73Å-long hydrophobic pore with a diameter of 2–6Å (*21, 22*) within the pentameric α-helical bundle (Fig. 1.4). This pore is able to bind many biologically relevant hydrophobic compounds, and the crystal structures of the COMPcc-vitamin $D_3$ (*23*), COMPcc-all-*trans* retinol, and COMPcc-benzene complexes have been solved (*24*). The binding properties of the hydrophobic channel suggest the potential of COMPcc to be used as a storage and delivery system for hydrophobic compounds (*25*).

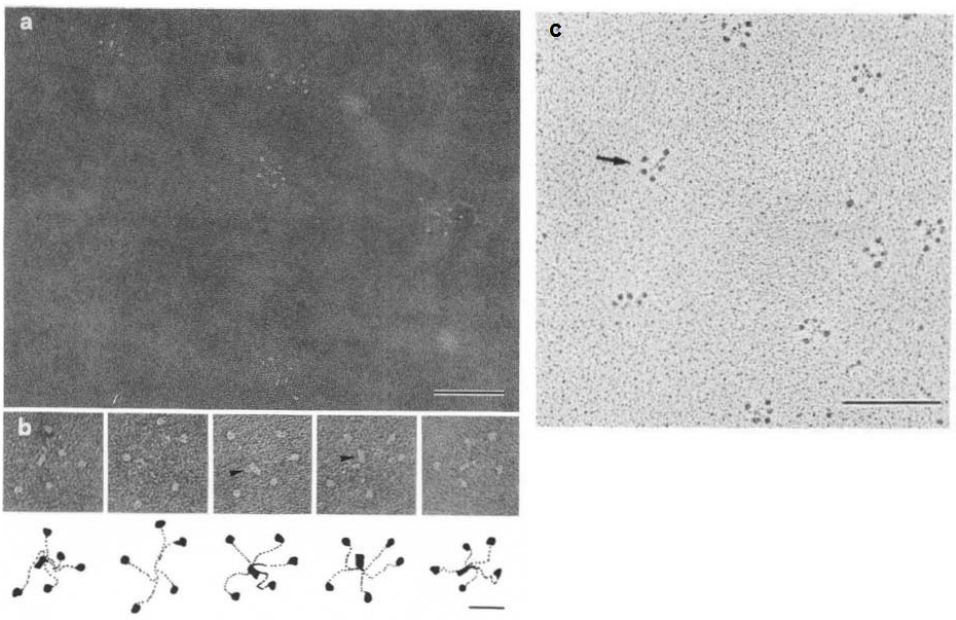


**Fig. 1.2. The structure of COMP.** (a) Electron micrograph of COMP molecules obtained by negative staining with uranyl formate. The bar represents 100 nm. (b) Higher magnification of select COMP molecules and their corresponding schematic drawings. The bar represents 25 nm.

(c) Glycerol sprayed/rotary shadowed electron micrograph of COMP. The arrow indicates a representative molecule of COMP with the characteristic five-armed structure. The bar represents 200 nm (*12*).


**Fig. 1.3. Domain architecture of COMP.** Schematic showing the domains of COMP. The N-terminal oligomerization domain (cc) is shown in blue, the EGF-like domains (EG) in yellow, the calcium-binding domains in purple (T3), and the C-terminal globular domain in green (TC).

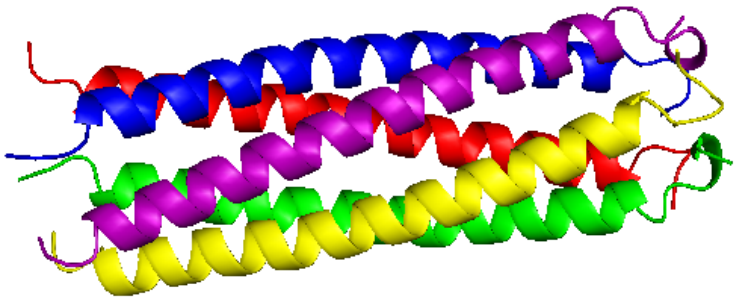

**Fig. 1.4. X-ray crystal structure of COMPcc.** Five α-helices form a pentameric coiled-coil with a central cavity that has the ability to bind hydrophobic compounds. The five α-helices are shown in different colours for contrast. PDB ID: 1VDF (*22*).

### 1.3. Agrin

Agrin is a multidomain heparan sulfate proteoglycan with an apparent molecular weight of 400–600 kDa on SDS-PAGE. It mediates the induction of postsynaptic specializations at the neuromuscular junction (NMJ), including the aggregation of acetylcholine receptors (AChR) (*26*). At the N-terminus of agrin, the N-terminal agrin (NtA) domain is found. The NtA domain comprises the first 135 amino acid residues and is required for binding to the basal membrane component laminin. The NtA domain is followed by nine follistatin-like (FS) domains. The C-terminal half of agrin contains three laminin G-like domains (LamG), which contain the α-dystroglycan binding site and the AChR clustering activity (Fig. 1.5) (*27, 28*).

The NtA domain is the most highly conserved domain in agrin among known species, with a >90% amino acid sequence similarity. However, there is a very small sequence similarity to other proteins (<15%), suggesting that the NtA domain is unique to agrin. The x-ray crystal structure of NtA revealed a β-barrel structure with α-helices at the N- and C-termini (Fig. 1.6). Based on structure similarity, it would appear that NtA belongs to the class of oligosaccharide/oligonucleotide-binding (OB) family of proteins. Also, NtA has a high structural similarity to the N-terminal proteinase inhibition domain of the tissue inhibitor of metalloproteinases-1 (N-TIMP-1). Matrix metalloproteinases (MMPs) are zinc-dependent endopeptidases that function in the degradation of ECM components, and their proteolytic activities are tightly regulated by TIMPs (*29*). The interesting structural features of NtA suggest the possibility of additional functions of NtA besides the known laminin-binding activity.

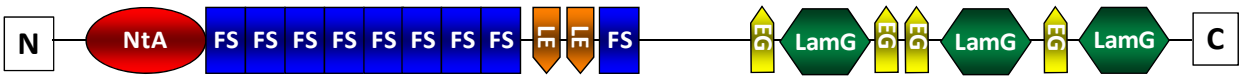


**Fig. 1.5. Domain organization of agrin.** Schematic showing the domains of agrin. The N-terminal agrin domain (NtA) is shown in red, the follistatin-like domains (FS) in blue, the laminin EGF-like domains (LE) in orange, the EGF-like domains (EG) in yellow, and the laminin G-like domains (LamG) in green.

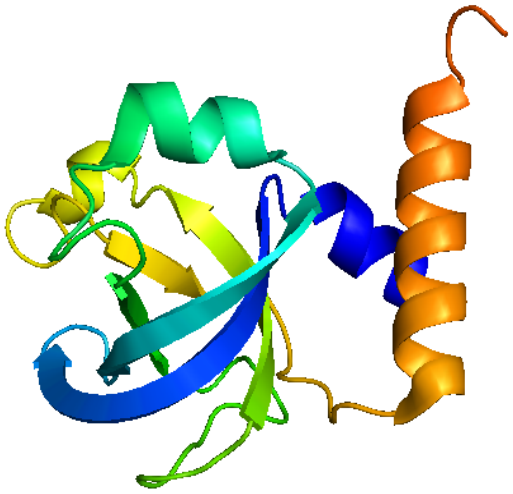

**Fig. 1.6. X-ray crystal structure of NtA.** Ribbon diagram of the NtA structure showing the overall β-barrel structure capped by an α-helix, and with the two α-helices at the N- and C-termini. PDB ID: 1JB3 (*30*).

### 1.4. Research objectives

The objective of this research project was to understand the structure-function relationships of two important ECM proteins, COMPcc and NtA. The review paper which forms

the basis of Chapter 2 opens the possibility of COMPcc to be used as a targeted drug delivery system. This led to the studies done in Chapter 3, where the binding properties of COMPcc were investigated by loading fatty acids into the hydrophobic channel of COMPcc. The structures of the COMPcc-fatty acid complexes were solved by x-ray crystallography, and the binding in solution was studied by fluorescence spectroscopy. In Chapter 4, the structure-function relationship of NtA was studied by using long-wavelength x-ray crystallography, mass spectrometry, and molecular modelling. The goal was to elucidate novel structural features of NtA that are important for the important functions of the protein.

Overall, the entire field of ECM biochemistry is significantly lacking in the understanding of the important functions of individual ECM components. The nature of the interactions between various ECM proteins, and the specific functions of individual domains are poorly understood. This thesis aims to further the knowledge in this area by studying domains of interest of two important ECM proteins. By examining the three-dimensional structures of these domains, insights into the overall function of these proteins can be gained.

### 1.5. Organization of the dissertation

This dissertation is prepared as a sandwich thesis, comprising three first-author papers published in peer-reviewed scientific journals. Each of these manuscripts contributes to the overall theme of the dissertation, which is the structural and functional characterization of ECM proteins. Throughout the thesis, a variety of techniques is used to add to the understanding of two important ECM proteins, COMPcc and NtA and to unveil novel functions and properties of these proteins (Fig. 1.7).

Chapter 1 provides a brief introduction to the research project. It introduces the general topic of ECM proteins, illustrating their major functions and characteristics. It then describes the two main proteins of interest, COMPcc and NtA, and outlines the objectives of the thesis project.

Chapter 2 is entitled *"The use of coiled-coil proteins in drug delivery systems."* It is an invited review article published in the special issue *"New vistas in anti-cancer therapy"* of the *"European Journal of Pharmacology."* The article introduces the topic of coiled-coil proteins and provides a comprehensive review of the significant features and functions of these proteins. It focuses on the ECM protein COMPcc and highlights the unique properties of this coiled-coil protein that make it an important protein for further research. This review also proposes the use of coiled-coil proteins as targeted drug delivery systems based on the evidence provided by the compiled research. This opens the possibility for investigating the storage functions of COMPcc and provides the background for the research done in Chapter 3.

Chapter 3 is entitled *"The coiled-coil domain of Cartilage Oligomeric Matrix Protein binds fatty acids via a ring of conserved glutamines."* This is an article in preparation for submission to the *"Journal of Biological Chemistry."* In this article, the binding characteristics of the COMPcc hydrophobic channel are investigated. The x-ray crystal structures of COMPcc in complex with five fatty acids are presented. These data are supported by solution studies using fluorescence spectroscopy.

Chapter 4 is entitled *"An interdomain disulfide bridge links the NtA and first FS domain in agrin."* It is an accelerated communication article published in the journal *"Protein Science."* The study published in this article revealed a novel interdomain disulfide bridge between the C-terminus of NtA and the FOLN subdomain of the first FS domain of agrin. This linkage causes a

rigid interface between the two domains that is independent of an alternative mRNA splice insert in NtA, indicating that the disulfide bond may be required for the proper folding of agrin.

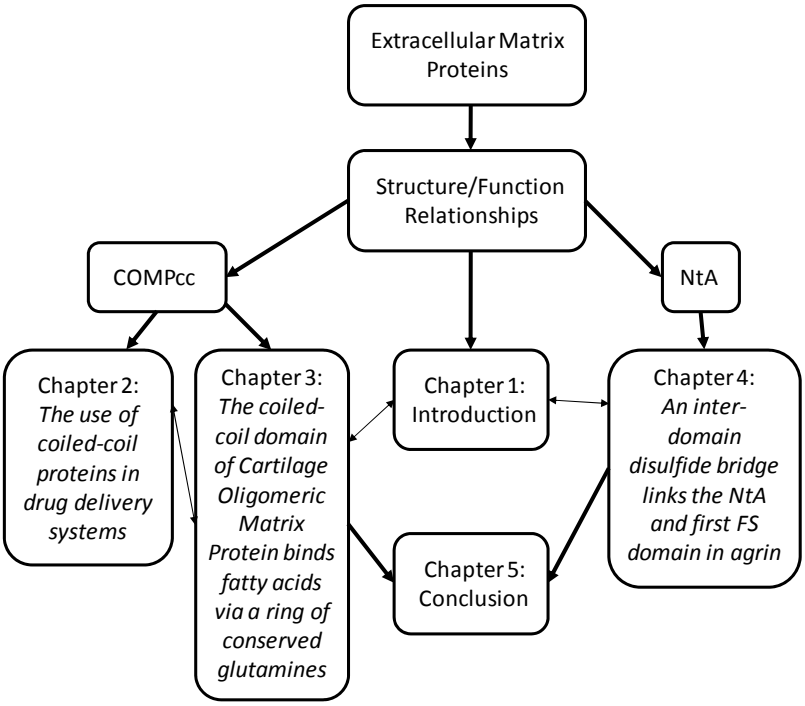


**Fig. 1.7. Dissertation organization.** Scheme showing the general organization, flow, and connecting themes of the thesis.

## References


1. Frantz, C., Stewart, K. M., and Weaver, V. M. (2010) The extracellular matrix at a glance, *J. Cell Sci. 123*, 4195-4200.
2. Hohenester, E., and Engel, J. (2002) Domain structure and organisation in extracellular matrix proteins, *Matrix Biol. 21*, 115-128.
3. Leeming, D. J., Bay-Jensen, A. C., Vassiliadis, E., Larsen, M. R., Henriksen, K., and Karsdal, M. A. (2011) Post-translational modifications of the extracellular matrix are key

events in cancer progression: opportunities for biochemical marker development, *Biomarkers 16*, 193-205.

4. Bruckner-Tuderman, L., and Bruckner, P. (1998) Genetic diseases of the extracellular matrix: more than just connective tissue disorders, *J Mol Med (Berl) 76*, 226-237.
5. Jarvelainen, H., Sainio, A., Koulu, M., Wight, T. N., and Penttinen, R. (2009) Extracellular matrix molecules: potential targets in pharmacotherapy, *Pharmacol. Rev. 61*, 198-223.
6. Chen, F. H., Rousche, K. T., and Tuan, R. S. (2006) Technology Insight: adult stem cells in cartilage regeneration and tissue engineering, *Nat. Clin. Pract. Rheumatol. 2*, 373-382.
7. Bork, P., Downing, A. K., Kieffer, B., and Campbell, I. D. (1996) Structure and distribution of modules in extracellular proteins, *Q. Rev. Biophys. 29*, 119-167.
8. Hedbom, E., Antonsson, P., Hjerpe, A., Aeschlimann, D., Paulsson, M., Rosa-Pimentel, E., Sommarin, Y., Wendel, M., Oldberg, A., and Heinegard, D. (1992) Cartilage matrix proteins. An acidic oligomeric protein (COMP) detected only in cartilage, *J. Biol. Chem. 267*, 6132-6136.
9. DiCesare, P., Hauser, N., Lehman, D., Pasumarti, S., and Paulsson, M. (1994) Cartilage oligomeric matrix protein (COMP) is an abundant component of tendon, *FEBS Lett. 354*, 237-240.
10. Smith, R. K., Zunino, L., Webbon, P. M., and Heinegard, D. (1997) The distribution of cartilage oligomeric matrix protein (COMP) in tendon and its variation with tendon site, age and load, *Matrix Biol. 16*, 255-271.

11. Muller, G., Michel, A., and Altenburg, E. (1998) COMP (cartilage oligomeric matrix protein) is synthesized in ligament, tendon, meniscus, and articular cartilage, *Connect. Tissue Res. 39*, 233-244.

12. Morgelin, M., Heinegard, D., Engel, J., and Paulsson, M. (1992) Electron microscopy of native cartilage oligomeric matrix protein purified from the Swarm rat chondrosarcoma reveals a five-armed structure, *J. Biol. Chem. 267*, 6137-6141.

13. Qabar, A., Derick, L., Lawler, J., and Dixit, V. (1995) Thrombospondin 3 is a pentameric molecule held together by interchain disulfide linkage involving two cysteine residues, *J. Biol. Chem. 270*, 12725-12729.

14. Shen, Z., Heinegard, D., and Sommarin, Y. (1995) Distribution and expression of cartilage oligomeric matrix protein and bone sialoprotein show marked changes during rat femoral head development, *Matrix Biol. 14*, 773-781.

15. Ekman, S., Reinholt, F. P., Hultenby, K., and Heinegard, D. (1997) Ultrastructural immunolocalization of cartilage oligomeric matrix protein (COMP) in porcine growth cartilage, *Calcif. Tissue Int. 60*, 547-553.

16. O'Rourke, K. M., Laherty, C. D., and Dixit, V. M. (1992) Thrombospondin 1 and thrombospondin 2 are expressed as both homo- and heterotrimers, *J. Biol. Chem. 267*, 24921-24924.

17. Hecht, J. T., Nelson, L. D., Crowder, E., Wang, Y., Elder, F. F., Harrison, W. R., Francomano, C. A., Prange, C. K., Lennon, G. G., Deere, M., and et al. (1995) Mutations in exon 17B of cartilage oligomeric matrix protein (COMP) cause pseudoachondroplasia, *Nat. Genet. 10*, 325-329.

18. Hecht, J. T., Montufar-Solis, D., Decker, G., Lawler, J., Daniels, K., and Duke, P. J. (1998) Retention of cartilage oligomeric matrix protein (COMP) and cell death in redifferentiated pseudoachondroplasia chondrocytes, *Matrix Biol. 17*, 625-633.

19. Briggs, M. D., Mortier, G. R., Cole, W. G., King, L. M., Golik, S. S., Bonaventure, J., Nuytinck, L., De Paepe, A., Leroy, J. G., Biesecker, L., Lipson, M., Wilcox, W. R., Lachman, R. S., Rimoin, D. L., Knowlton, R. G., and Cohn, D. H. (1998) Diverse mutations in the gene for cartilage oligomeric matrix protein in the pseudoachondroplasia-multiple epiphyseal dysplasia disease spectrum, *Am. J. Hum. Genet. 62*, 311-319.

20. Thur, J., Rosenberg, K., Nitsche, D. P., Pihlajamaa, T., Ala-Kokko, L., Heinegard, D., Paulsson, M., and Maurer, P. (2001) Mutations in cartilage oligomeric matrix protein causing pseudoachondroplasia and multiple epiphyseal dysplasia affect binding of calcium and collagen I, II, and IX, *J. Biol. Chem. 276*, 6083-6092.

21. Efimov, V. P., Engel, J., and Malashkevich, V. N. (1996) Crystallization and preliminary crystallographic study of the pentamerizing domain from cartilage oligomeric matrix protein: a five-stranded alpha-helical bundle, *Proteins 24*, 259-262.

22. Malashkevich, V. N., Kammerer, R. A., Efimov, V. P., Schulthess, T., and Engel, J. (1996) The crystal structure of a five-stranded coiled coil in COMP: a prototype ion channel?, *Science 274*, 761-765.

23. Ozbek, S., Engel, J., and Stetefeld, J. (2002) Storage function of cartilage oligomeric matrix protein: the crystal structure of the coiled-coil domain in complex with vitamin D(3), *EMBO J. 21*, 5960-5968.

24. Guo, Y., Bozic, D., Malashkevich, V. N., Kammerer, R. A., Schulthess, T., and Engel, J. (1998) All-trans retinol, vitamin D and other hydrophobic compounds bind in the axial pore of the five-stranded coiled-coil domain of cartilage oligomeric matrix protein, *EMBO J. 17*, 5265-5272.

25. McFarlane, A. A., Orriss, G. L., and Stetefeld, J. (2009) The use of coiled-coil proteins in drug delivery systems, *Eur. J. Pharmacol. 625*, 101-107.

26. Ruegg, M. A., and Bixby, J. L. (1998) Agrin orchestrates synaptic differentiation at the vertebrate neuromuscular junction, *Trends Neurosci. 21*, 22-27.

27. Gesemann, M., Denzer, A. J., and Ruegg, M. A. (1995) Acetylcholine receptor-aggregating activity of agrin isoforms and mapping of the active site, *J. Cell Biol. 128*, 625-636.

28. Gesemann, M., Brancaccio, A., Schumacher, B., and Ruegg, M. A. (1998) Agrin is a high-affinity binding protein of dystroglycan in non-muscle tissue, *J. Biol. Chem. 273*, 600-605.

29. Nagase, H., and Woessner, J. F., Jr. (1999) Matrix metalloproteinases, *J. Biol. Chem. 274*, 21491-21494.

30. Stetefeld, J., Jenny, M., Schulthess, T., Landwehr, R., Schumacher, B., Frank, S., Ruegg, M. A., Engel, J., and Kammerer, R. A. (2001) The laminin-binding domain of agrin is structurally related to N-TIMP-1, *Nat. Struct. Biol. 8*, 705-709.

## Chapter 2: The use of coiled-coil proteins in drug delivery systems

This chapter is based on the following manuscript published in the journal "*European Journal of Pharmacology*". Full citation of the paper is as follows:

McFarlane, A. A., Orriss, G. L., and Stetefeld, J. (2009) The use of coiled-coil proteins in drug delivery systems, *Eur. J. Pharmacol. 625*, 101-107.

For copyright permission please see:

https://s100.copyright.com/CustomerAdmin/PLF.jsp?lID=2011050_1305248713538

**Abstract**

The coiled-coil motif is found in approximately 10% of all protein sequences and is responsible for the oligomerization of proteins in a highly specific manner. Coiled-coil proteins exhibit a large diversity of function (eg. Gene regulation, cell division, membrane fusion, drug extrusion) thus demonstrating the significance of oligomerization in biological systems. The classical coiled-coil domain comprises a series of consecutive heptad repeats in the protein sequence that are readily identifiable by the location of hydrophobic residues at the '*a*' and '*d*' positions. This gives rise to an α-helical structure in which between 2 and 7 helices are wound around each other in the form of a left-handed supercoil. More recently, structures of coiled-coil domains have been solved that have an 11 residue (undecad) or a 15 residue (pentadecad) repeat, which show the formation of a right-handed coiled-coil structure.

The high stability of coiled coils, together with the presence of large internal cavities in the pentameric coiled-coil domain of cartilage oligomerization matrix protein (COMPcc) and the

tetrameric right-handed coiled coil of *Staphylothermus marinus* (RHCC) has led us and others to look for therapeutic applications. In this review, we present evidence in support of a vitamin A and vitamin $D_3$ binding activity for the pentameric COMPcc molecule. In addition, we will discuss exciting new developments which show that the RHCC tetramer is capable of binding the major anticancer drug cisplatin and the ability to fuse it to an antigenic epitope for the development of a new generation of vaccines.

**Key Words**: Coiled coils, drug delivery, COMPcc, RHCC, vitamin $D_3$, vitamin A, cisplatin

## 2.1. Introduction

The α-helical coiled coil is the most frequently encountered subunit oligomerization motif found in proteins (*1-3*), with an analysis of the protein database revealing that approximately 10% of all proteins contain the signature heptad repeat $(abcdefg)_n$ which is characteristic of coiled-coil proteins (*4*). Its existence was first predicted in a series of seminal papers by Crick (*5-7*) and supported by an observation by Perutz and others that the long fibrous protein tropomyosin and other members of the *k-m-e-f* proteins exhibited diffraction peaks with a repeated spacing of 5.15Å. Crick proposed that this could be explained by an arrangement in which two α-helices wrapped around each other at an angle of approximately 20° such that their side chains were interlocked together in a motif that repeated itself every 7 residues, or 2 turns of the helix. This arrangement would result in a helix with 3.5 residues per turn compared to the usual 3.6 for an undistorted α-helix, with the driving force for the distortion of the helices coming from the hydrophobic residues at the core '*a*' and '*d*' positions (Fig. 2.1). Crick referred to this packing as a "knobs into holes" arrangement in which one residue (*knob*) fits into a space

generated by four residues of the opposing helix (*hole*). The 4,3 spacing observed in coiled coils is the minimum spacing that allows for the formation of an amphipathic α-helix that has the core hydrophobic residues positioned on the same side of the helix, thereby facilitating the "knobs into holes" packing. Due to the supercoiling of the two helices around each other in a coiled coil, the handedness of the coiled coil is the opposite handedness to that of the helix which formed it. The simplest coiled-coil structure is the dimer, but trimers, tetramers, pentamers (*8, 9*), and heptamers (*10*) have been reported. In addition, if the helices come from the same protein chain, the resultant coiled coil is described as a homo-oligomer. Conversely, if the helices originate from different protein chains, the coiled coil is referred to as a hetero-oligomer.

Given that the α-helix is right-handed (*11*) and that the majority of coiled-coil proteins display a heptad repeat, it follows that the majority of coiled coils will adopt a left-handed supercoiled structure. However, more recent examples of right-handed coiled coils have been reported in the literature (*12-15*). This is not entirely unexpected as in addition to the classical 7/2 repeat suggested by Crick, Pauling proposed that α-helices could undergo deformations that resulted in repeats with an 11/3 (11 residues in 3 turns of the helix), 15/4, or 18/5 periodicity (*16*). More recently, a protein which displays an additional 25/7 periodicity has been reported (*17*). Of interest with respect to this review, was the prediction that the 7/2 and 15/4 repeats would have the opposite handedness (*18*). This has since been confirmed with the determination of coiled-coil structures that have a 15/4 (*13-15*) and an 11/3 periodicity (*19*), which revealed a right-handed coiled-coil structure (Fig. 2.1). The 18/5 and 25/7 repeats both form left-handed coiled coils (*16, 17*).

The prevalence of the coiled-coil motif across biological systems indicates its significance in many key biological processes. Indeed, coiled-coil motifs are found in proteins involved in transcription (*20*), cell/cell communication, membrane fusion, and proteins which act as molecular spacers and motors (see review Lupas and Gruber, 2005) (*21*). Such a diversity of function requires a simple design that is both robust and highly adaptable. Research is now being directed towards the use of these properties with applications in the field of nanotechnology. Due to the ability of certain coiled-coil structures to bind heavy metals they may also have an important environmental function (*19*). A promising new role relates to the ability of certain coiled-coil domains to bind biologically relevant molecules such as the major anticancer drug cisplatin, which raises the possibility that they could be used in the treatment of cancers. In addition, the use of coiled-coil domains for the production of a new generation of vaccines is another recent and exciting development in the field that will be watched with interest.

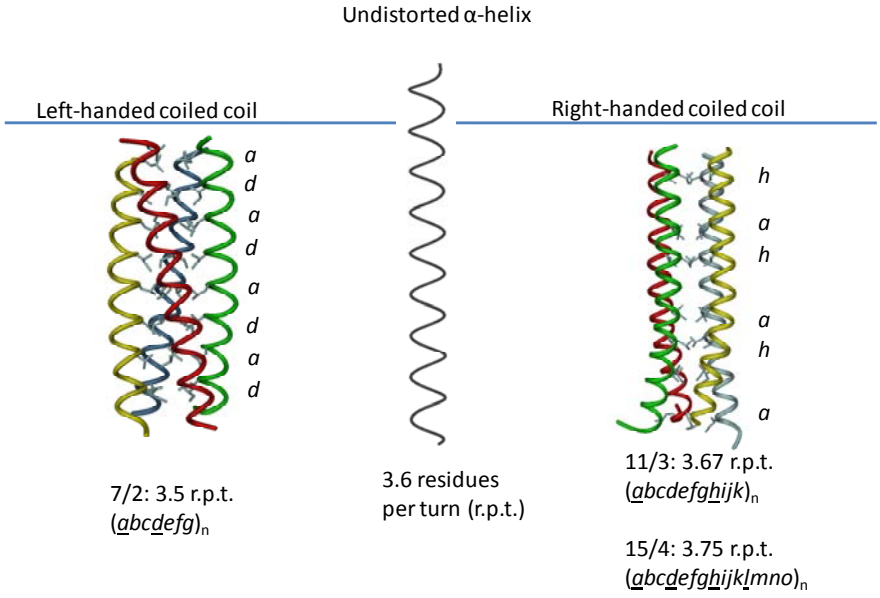


**Fig. 2.1: From α-helices to left- and right-handed coiled coils.**

In an undistorted α-helix one amino acid residue is rotated 100° along the screw axis and subsequently 3.6 residues are required for a full turn (**centre**). Coiled coils, which are pairs of α-helices winding around each other, show deviations from this rule. In left-handed coiled coils 7 residues are required for 2 turns and therefore are more tightly wound around each other in a left-handed supercoil to compensate for this missing 20° (**left**). In contrast, naturally occurring right-handed coiled coils reveal repeats of 11residues per 3 turns or 15 residues per 4 turns of the helix (**right**). As a consequence they are supercoiled either 20° or 60° to form a right-handed supercoil. Characteristic in all cases are highly repetitive sequence motifs of 7, 11, or 15 amino acid residues (in italics) with conserved aliphatic core residues (underlined) forming the hydrophobic core of coiled-coil channels.

### 2.2. Structure of coiled coils

The determination of the protein sequence for tropomyosin in 1975 showed that a heptad repeat $(\textit{abcdefg})_n$ could be identified from analysis of the protein sequence alone (*22*). This spawned a whole field in its own right aimed at correctly predicting coiled-coil motifs from an analysis of any given protein sequence (*23-29*). However, it was not until the publication of the influenza haemagglutinin structure (*30*) that the coiled coil was first visualized showing the core hydrophobic residues adopting the "knobs into holes" packing as predicted by Crick at positions '*a*' and '*d*' in the heptad repeat. Since then much effort has been directed towards a better description of the coiled coil both in terms of the driving forces that determine the oligomeric state and also what additional forces stabilize the coiled-coil oligomer.

The 33 residue leucine zipper region of the yeast transcription factor GCN4 has proven a useful experimental tool to address the issue of oligomerization state. GCN4 has the standard

heptad repeat with a 7/2 periodicity and has Ile predominantly at the '*a*' and Leu at the '*d*' positions, respectively. In this configuration the crystal structure shows that the protein adopts a dimeric coiled coil (*31*). However, if the positions of the Ile and Leu are reversed such that Leu is found at the '*a*' and Ile at the '*d*' position then the protein becomes a tetramer (*32*). Any other combination of Ile, Leu, or Val at the '*a*' and '*d*' positions in the same study was shown to result in a trimeric state (*33*).

The classical left-handed coiled coil displays a heptad repeat with a so-called 7/2 periodicity, which places the core hydrophobic residues at an average spacing of 3.5 residues per turn of the helix. The next nearest number larger than 3.6 which allows residues to assume quasi-equivalent positions after a small number of turns is 3.67. This is achieved with an undecad repeat $(abcdefghijk)_n$ that has a periodicity of 11/3. In this motif, hydrophobic residues are present at the '*a*' and '*d*' positions as for the heptad repeat but also additionally at the '*e*' and '*h*' positions. Such a coiled-coil motif was identified from a naturally occurring archeal protein (RHCC) from *Staphylothermus marinus* (*12*) and its structure solved in 2000 (*19*). Of interest, is the fact that the undecad motif gives rise to a coiled coil that adopts a right-handed configuration. In this structure, the hydrophobic residues at the '*a*' and '*h*' positions are equivalent to the '*a*' and '*d*' positions of left-handed coiled coils and form the expected "knobs into holes" packing, whereas the hydrophobic residues at the '*d*' and '*e*' positions pack in a "knobs into knobs" arrangement. Given that the number of residues per turn for the 11/3 right-handed coiled coil at 3.67 is only just above the 3.6-3.64 expected for an undistorted α-helix then it is perhaps not surprising to find that the RHCC is only weakly supercoiled to the right.

A final coiled-coil motif discussed here is the pentadecad repeat $(abcdefghijklmno)_n$ which is obtained by having fifteen residues over four turns of the helix and has an average

spacing between hydrophobic residues of 3.75. The tetramerization domain of VASP adopts such a 15/4 motif and its structure has been solved to atomic resolution (*13*). It forms a right-handed coiled coil, with the hydrophobic core residues located at positions '*a*', '*d*', '*e*', '*h*', and '*l*' of the repeat. In the VASP structure the residues at positions '*a*' and '*h*' form the expected "knobs into holes" packing as in the RHCC structure, whereas the residues at positions '*d*', '*e*', and '*l*' pack in a "knobs into knobs" arrangement. The larger spacing between hydrophobic residues is suggestive of a higher degree of supercoiling of the VASP right-handed coiled coil and this is reflected in the structure. Thus right-handed coiled coils with a 15/4 periodicity form coiled-coil structures that supercoil as much to the right as the more numerous left-handed coiled coils with a 7/2 periodicity coil to the left.

### 2.3. Stability, helix packing, and cavities

A key feature for the stability and oligomerization states of coiled coils is the hydrophobic core formed by the "knobs into holes" interactions between residues at the '*a*' and '*d*' positions of the heptad repeat (Fig. 2.1). In left-handed coiled coils, the β-branched side chains Leu, Ile, and Val together with Ala are found almost exclusively at the '*a*' and '*d*' positions within the heptad repeat and polar helix favouring residues are generally found elsewhere (*24, 34*). In a 4,3-hydrophobic pattern they form either parallel or perpendicular "knobs into holes" with amino acid side chains of the adjacent helix (*35*). In the RHCC from *Staphylothermus marinus*, the "knobs into holes" interactions between helices are formed by the side chains at positions '*a*' and '*h*' of the undecad repeat, which is consistent with a 7,4-hydrophobic repeat (Fig. 2.1). This was an unexpected result, as formation of an artificial right-handed coiled coil (*36*) was achieved using a 4,4,3-repeat. A further feature that stabilizes the

RHCC tetramer is provided by the "knobs into knobs" packing between the hydrophobic residues at the '*d*' position of one helix and the '*e*' position of the neighbouring helix to form the *de*-layer (Fig. 2.2A). The additional hydrophobic interactions present in the RHCC result in a large buried surface of ~9.500 $Å^2$ which accounts for ~50% of the total solvent-accessible area of the four isolated helices which compares with an expected value of 20-25% for a non coiled-coil protein. As a consequence, the RHCC is extremely thermostable and resists heating to temperatures >90°C (*37*). In contrast, the hydrophobic pattern of the right-handed VASP tetramer revealed a 4,3,4,4-hydrophobic pattern of "knobs into holes" that is equivalent to a heptad repeat and two 4-residue insertions (*13*).

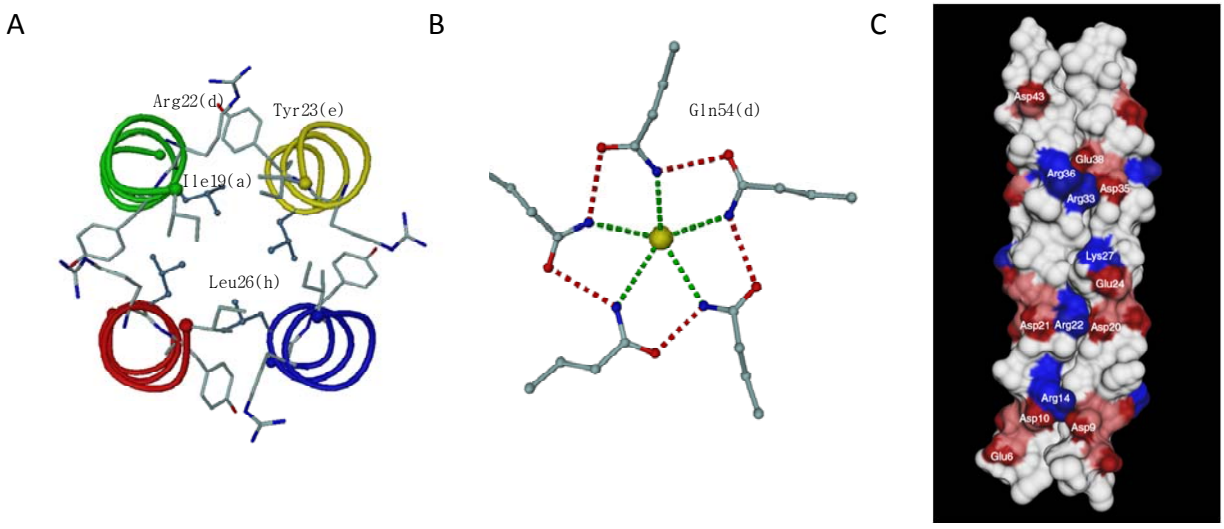


**Fig. 2.2. Inside and outside forces for coiled coil stabilization.**

(**A**) Inside forces are classical knobs-into-holes formations of aliphatic side chains oriented inside the coiled-coil channel. In addition to *a* and *h* positioned knobs-into-holes the right-handed coiled coil from *Staphylothermus marinus* reveals a so far unknown inter-helical hydrophobic core composed of residues in *d* and *e* positions (*de*-layer). (**B**) In the case of the pentameric COMPcc (PBD-code 1VDF), a ring of five $Gln^{54}$ residues in position *d* is arranged to form an inter-helical network of ionic interactions (red dotted lines) and contributes to the

binding (green dotted lines) of a chloride ion (highlighted in yellow). (**C**) Molecular surface model of the tetrameric RHCC structure. Only the front two helices of the RHCC tetramer are shown. Hydrophobic residues are depicted in grey, whereas positively- and negatively-charged residues, which form the major i-i'+2 (inter-helical) and i-i+3 (intra-helical) salt bridges are depicted in blue and red, respectively.

Of particular interest with respect to the topic of this review is that the left-handed pentameric coiled-coil domain of cartilage oligomerization matrix protein (COMPcc) and the right-handed tetrameric RHCC from *Staphylothermus marinus* both form large internal cavities. To date, no other examples exist of coiled coils that form such large cavities and so these two domains are unique. COMPcc forms a pentameric coiled-coil structure (*8, 9*) that contains a 73Å-long axial pore that has a diameter of between 2-6Å along its length, with the exception of the constriction formed by a ring of five conserved glutamine residues at position 54 which divides the channel into two (Fig. 2.2B, 2.3A). The positioning of the glutamine residue at a '*d*' position of the heptad repeat is somewhat surprising, but further analysis shows that this is related to its ligand binding activities which are discussed in detail in section 5. By contrast, the RHCC molecule contains four large cavities on the inside of the RHCC tetramer (Fig. 2.3B), a feature which is absent from all known left-handed tetrameric coiled-coil structures deposited in the protein database (PDB). The cavities are connected to a continuous central channel which is exclusively lined with aliphatic side chains. The diameter of the channel as defined by the van der Waals radii varies between 2.0 and 8.4Å, which exceeds the 6Å maximum diameter of the pentameric COMPcc pore. The cavities are located in the gap between the '*a*' and '*h*' layers. In the native structure, the cavities are occupied with water molecules that form water clusters due to the lack of polar groups. Analysis of the heavy atom derivatives used for phasing the structure

showed that the major heavy atom binding sites are located within these cavities. The presence of the heavy atoms in the cavities is almost certainly responsible for the excellent quality of the derivative parameters. The narrow 2Å-wide constriction formed at the '*a*' and '*h*' layers suggest that the derivatives do not enter the cavities via the termini. Instead it would appear likely that the heavy atoms penetrate directly into the cavities through gates formed by a complex network of surface salt bridges and ionic interactions (*19*).

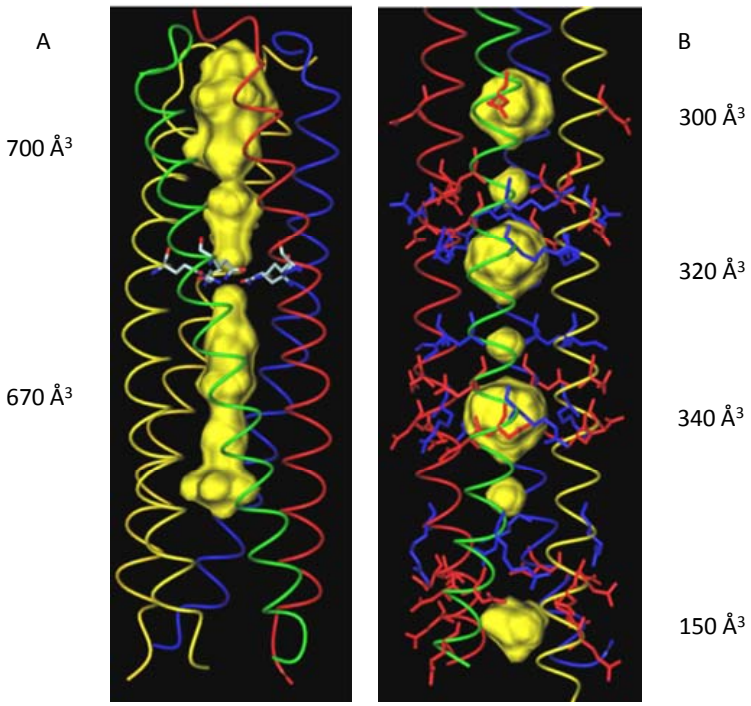


**Fig. 2.3. Cavities inside tetrameric and pentameric coiled-coil channels.**

Extraordinary features of COMPcc (**A**) and RHCC (**B**) are large cavities along the aliphatic channel core. These cavities (highlighted as yellow van der Waals spheres) can function as storage spaces for a great variety of different cargo systems with large biomedical importance. The volumes of individual cavities are shown in cubic Angstroms ($Å^3$). Salt bridges surrounding individual cavities in RHCC are highlighted in red (acidic resdiues) and blue (basic residues).

The $Gln^{54}$ ring in COMPcc is drawn in colour atom type. Remarkably, both channels allow for different shapes of interior spaces. Whereas RHCC cavities are regularly-arranged separated balls, both cavities in COMPcc are elongated and separated by the $Gln^{54}$ ring in the *d*-position.

## 2.4. Electrostatics

After the hydrophobic forces, the electrostatic interactions that are formed within (*intra*-) or between (*inter*-) the helices of the coiled coil are the next most significant factors involved in stabilization of coiled-coil domains. This is facilitated by an exceptionally high content (up to ~27 %) of charged residues. Such a high frequency of favourable electrostatic interactions is not observed in non coiled-coil protein structures and may explain in part the high thermostability of many coiled-coil peptides. The inter-helical salt bridges are typically formed between the '*e*' position of one helix and the '*g*' position of the adjacent helix (i-i'+2) with the residues involved having the opposite charges (*24*). With regard to the intra-helical salt bridges, these are typically between residues in an i-i+3 or an i-i+4 arrangement (*38*). In the RHCC structure, besides the expected i-i'+2 inter-helical salt bridges, a unique i-i'+5 interaction is also observed. Furthermore, in the RHCC tetramer the salt bridges are organized into three networks of complex salt bridges involving residues that are flanked by polar interactions (Fig. 2.2C).

## 2.5. Coiled coils and drug delivery

### *2.5.1. COMPcc*

The design of efficient drug delivery systems, especially for anticancer therapies, is of great interest as many of these drugs cause toxic side effects due to their lack of specificity for cancerous versus healthy cells. Coiled coils are attractive candidates for drug delivery systems

due to their ability to bind a number of biologically relevant molecules. The five-stranded coiled-coil domain of cartilage oligomeric matrix protein (COMPcc) is one such example, where the storage function of the protein has been studied. Through the use of x-ray crystallography and in-solution assays, the ability of COMPcc to reversibly bind various cargo molecules has been thoroughly characterized (*39, 40*).

COMPcc is a homopentamer consisting of five α-helices that are joined by interchain disulfide bridges. It forms an internal, hydrophobic pore that is 73Å-long and has a diameter of 2–6Å. This pore is divided into two by a ring of conserved glutamines at position 54, which via an intricate network of hydrogen bonds act as a biological magnet to attract electronegative ligands (*8, 9*). The ability of the pore to bind a wide spectrum of ligands is directly related to its physicochemical properties. For example, small electronegative ions such as chloride are held in the channel by virtue of an electrostatic interaction with the glutamine 54 ring, whereas vitamins A and $D_3$ are bound by a combination of hydrophobic forces along the length of the pore and an electrostatic interaction with the glutamine 54 ring (*8, 39, 40*). In addition, small hydrophobic compounds such as benzene can also bind the channel due to its hydrophobicity (*39*). Thus the COMPcc channel has the potential to act as a "Trojan horse" to deliver therapeutic compounds in the treatment of disease.

A possible storage and delivery function of COMPcc was first suggested by Guo et al. with the observation that the coiled-coil protein was able to bind a number of important cell signalling molecules, including vitamin A (all-*trans* retinol) and vitamin $D_3$. They were able to determine a dissociation constant of $K_D = 0.6\mu M$ for all-*trans* retinol using fluorescence titration. Furthermore, they solved the crystal structure of the COMPcc-all-*trans* retinol complex to validate the ligand binding within the hydrophobic channel. A single molecule of retinol is

bound in the "lower" N-terminal cavity of COMPcc, with its β-ionone ring oriented towards the N-terminus and the isoprene tail oriented towards the C-terminus (Fig. 2.4A). The hydroxyl group of retinol forms hydrogen bonds with two of the five glutamine residues (*39*). A subsequent study by Özbek et al. confirmed the binding of vitamin $D_3$ to COMPcc by x-ray crystallography of the COMPcc-vitamin $D_3$ complex (*40*). An earlier study by Guo et al. had been suggestive of vitamin $D_3$ binding as the midpoint transition temperature $T_m$ for unfolding increased by $8^oC$ in the presence of the molecule (*39*). Interestingly, the crystal structure by Özbek et al. shows that two molecules of vitamin $D_3$ are bound to COMPcc. One molecule of vitamin $D_3$ is bound in each cavity, with the molecules oriented in a head-to-head fashion in the hydrophobic channel (Fig. 2.4B). Of significance is that upon binding vitamin $D_3$, the hydrophobic cavities increase in volume by ~30%. This is achieved by a reorientation of the β-branched side chains at positions *'a'* and *'d'* of the coiled coil, which are responsible for forming the "knobs into holes" (*40*) and shows the dynamic nature of the channel when binding ligands. Both the studies by Guo et al. and Özbek et al. are supportive of a storage and delivery function for COMPcc.

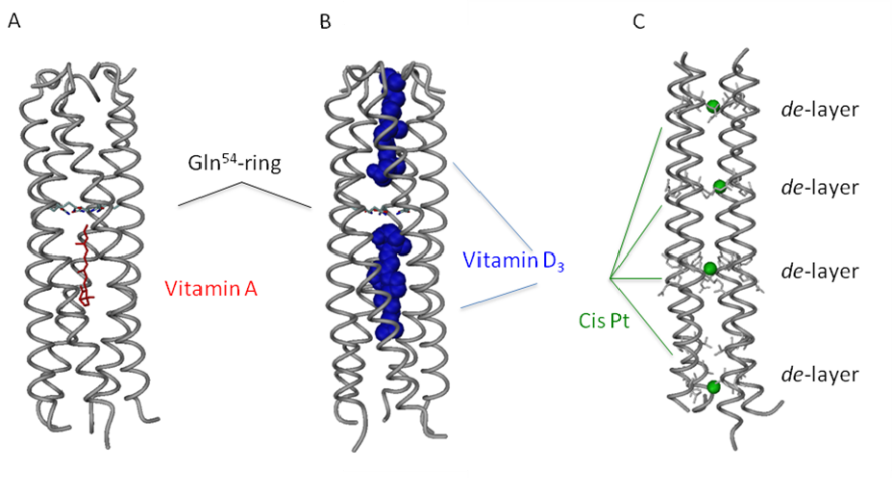


**Fig. 2.4. Storage properties of coiled-coil domains.**

COMPcc has been shown to store vitamin A (**A**) and vitamin $D_3$ (**B**). In both cases the elongated cargo is oriented such that its hyroxyl group can form an electrostatic interaction with the $Gln^{54}$ ring system. In contrast to vitamin A where only molecule is bound, COMPcc can store two vitamin $D_3$ molecules, one in each cavity. (**C**) The larger cavities in RHCC can store one molecule of cisplatin each. Whereas aliphatic, elongated cargos in COMPcc are diffusing via both N- and C-terminus into the hydrophobic channel, the ionic network in RHCC allows for a diffusion of cisplatin via the inter-helical interface.

*2.5.2. RHCC*

Another coiled-coil domain of interest is the right-handed coiled-coil domain (RHCC) from *Staphylothermus marinus*. This domain forms a tetrameric right-handed coiled coil, which contains four large internal cavities that are capable of binding large metals such as gold and platinum (*19*). Of relevance to this review, a recent study by Eriksson et al. demonstrated that the anticancer drug cisplatin could bind to RHCC. Furthermore, the complex was delivered to tumor cell lines *in vitro* (*41*). Cisplatin is a widely-used chemotherapeutic drug that is used to treat a number of cancers including those of the reproductive system, as well as stomach and

bladder (*42*). Unfortunately, treatment with cisplatin leads to numerous toxic side effects, thus the development of an effective drug delivery system aimed at selectively targeting cancerous cells is highly desirable.

Eriksson et al. incubated RHCC with cisplatin and the excess was removed by gel filtration. The amount of bound cisplatin was determined by inductively-coupled plasma optical emission spectroscopy (ICP-OES), and it was found that on average one molecule of cisplatin is bound per cavity in each RHCC tetramer (Fig. 2.4C). They also demonstrated that cisplatin-bound RHCC was stable for up to 12h in solution. Fluorescently-labelled RHCC was incubated with FADu carcinoma cells and the uptake of the protein into the cells was monitored by fluorescence microscopy. It was determined that the cell lines were able to take up the protein into intracellular vesicles. The cytotoxicity of RHCC and cisplatin-bound RHCC was analyzed by incubating the proteins with various carcinoma cell lines. RHCC exerted no cytotoxic effect on cells, while cisplatin-bound RHCC was found to either show an equivalent effect or a significantly higher effect than cisplatin alone. Perhaps the most encouraging findings of the study by Eriksson *et al.* were in the trials of cisplatin-bound RHCC *in vivo*. Mice treated with the drug-loaded protein exhibited no significant immune response (*41*). These findings indicate the great potential of RHCC as a carrier system for cisplatin.

### *2.5.3. Future prospects*

One of the main challenges of drug delivery is selectively targeting toxic compounds to their desired locations. In particular with chemotherapeutic drugs, the cytotoxicity of these compounds causes many harsh side effects, thus the development of a targeted system to bypass healthy cells and deliver therapies directly to cancer cells is of great interest. The unique

properties of coiled coils make them highly suitable for use in these systems. Since the major function of coiled coils is to act as oligomerization domains, the N- and C-terminals of the proteins can be easily conjugated to numerous epitopes that can bind specific cell surface markers, thereby creating a powerful targeted drug delivery system. The ability to fuse coiled coil domains to other protein domains was first shown by Tomschy et al. who conjugated five extracellular CAD domains of E-cadherin to the N-terminus of COMPcc (*43*). In addition, Ahrens et al. fused five CAD domains of P-cadherin via a linker region to the N-terminus of COMPcc (*44*). The classical cadherin domains (E-, N-, P-, and C-cadherin) function in $Ca^{2+}$-dependent selective cell recognition and adhesion (*45, 46*). Therefore, by attaching COMPcc to molecules that are involved in cell recognition and adhesion, this raises the possibility that the complex can be directed to a target cell where it would release its cargo. These studies illustrate the potential of coiled coils in the development of new targeted drug delivery systems.

A final therapeutic application discussed here is the fusion of coiled-coil domains to epitopes for the production of therapeutic vaccines. As coiled coils form oligomers, such fusions should result in the repetitive display of antigens thereby generating a strong immune response. This property was examined by Schroeder et al. and Pimentel et al., who generated a peptide nanoparticle for the presentation of actin antigens and severe acute respiratory syndrome (SARS) virus epitopes, respectively (*47, 48*). The core polypeptide of the nanoparticle contains a modified sequence of the pentameric COMPcc connected to a *de novo* designed trimeric coiled coil via two glycine residues. The desired epitope can then be linked to the C-terminus of the core polypeptide for antigenic display. Schroeder et al. linked a short sequence corresponding to the hydrophobic loop of actin and Pimentel et al. linked the HRC1 epitope from the SARS coronavirus spike protein to the trimeric coiled coil, respectively (*47, 48*). The resultant

polypeptide self-assembles into an icosahedron and resembles a virus-like particle (VLP) (*49*). Both studies demonstrated that this repetitive arrangement of epitopes on the surface of the peptide nanoparticle was able to elicit a specific immunogenic response to the projected epitopes, indicating the potential of this system in the design of vaccines (*47, 48*).

## 2.6. Conclusion

Coiled coils are primarily oligomerization domains that are capable of forming well defined higher order multimers, which are extremely stable. They were initially identified in proteins of the *k-m-e-f* series, which includes the keratins, tropomyosin, and fibrinogen; therefore it was believed that they would simply provide rigidity to proteins. Despite the fact that these proteins led to the proposal of a coiled-coil structure, they are only poorly characterized at the structural level due to their biochemical characteristics. For example, the full length tropomyosin has only been solved to 7Å resolution (*50*) although the N-terminal (*51*) and C-terminal fragments (*52*) have been solved separately to 2Å and 2.7Å, respectively. Of all the *k-m-e-f* proteins, the only one whose structure has been completely solved is fibrinogen (*53-55*). However, since their initial discovery they have been shown to be involved in key biological processes that include gene transcription (*20*), regulation (*15*), cytokinesis (*56*), molecular motors (*57*), muscle fibres (*58*), multidrug resistance (*59, 60*), membrane fusion (*61*), protein stability in harsh environments (*12, 37*), and more recently as potential delivery systems either for the treatment of cancer (*41*) or in the development of novel vaccines (*47, 48*). Despite this diversity of function, the one common characteristic is the requirement for a stable oligomerization domain and this is provided by the coiled-coil domain.

It is perhaps not surprising to find that the α-helical coiled coil is derived from one of the two stable secondary structure elements found in proteins, namely the α-helix. Thus, nature has taken an already stable structural element and improved its stability by virtue of a simple repeating motif that takes advantage of the inherent properties of the α-helix. This has the major advantage that such domains can be readily incorporated into existing structural motifs. Furthermore, all that is required to adjust the oligomeric state of the coiled coil is a substitution of the core residues at the '*a*' and '*d*' positions in the heptad repeat (*32, 33*). Therefore, the coiled coil can be easily adapted to suit the end function, which may explain why this domain is so prevalent in biological systems. Interestingly, all the right-handed coiled coils identified to date (*12, 15, 19*) are tetrameric, so it remains to be seen if other oligomeric states of such domains are identified in the future or if the tetrameric state is preferred above all other combinations. As a simple rule of thumb, it would appear that when the average spacing between hydrophobic residues is <3.6Å the coiled coil is left-handed, whereas when the spacing is >3.6Å it adopts a right-handed configuration. Thus it seems that biological systems have devised an easy way in which to differentiate between these two coiled-coil forms.

With regard to this review, the two coiled-coil domains of the most interest are the left-handed COMPcc from *Homo sapiens* and the right-handed RHCC from *Staphylothermus marinus*. The shared feature of these two proteins is the presence of large internal cavities that are absent from all other coiled-coil structures solved to date. COMPcc contains two internal cavities that have been shown to bind the signalling molecules vitamins A and $D_3$. These molecules are largely hydrophobic with polar groups at one end and bind readily inside the hydrophobic cavity of COMPcc. Many drugs used in the treatment of cancers are hydrophobic, which makes drug delivery by conventional means problematic due to solubility issues.

Therefore, this problem could be overcome by loading the therapeutic drugs inside the cavity. This is exactly the approach that Eriksson et al. used when they demonstrated that the cavities of RHCC could be loaded with the anticancer drug cisplatin (*41*). There are two main advantages to using coiled-coil domains as "Trojan horses" for the delivery of therapeutic compounds. Firstly, by shielding the therapeutic drug inside the cavity it reduces the possibility of the drug being metabolized before reaching the cancerous site. Secondly, it avoids the recognition of the drug by one of the many membrane-bound drug exporters, which actively pump the drug back out of the cell before it reaches a high enough concentration to kill the cancer cell.

An additional property of coiled coils that is beneficial for drug delivery relates to its oligomeric state. Studies in the late 1990s and early 2000s showed that it was possible to fuse coiled coils to other protein domains involved in cell recognition and adhesion (*43, 44*). These proof of principle experiments are important as they raise the possibility that a targeting molecule can be attached to a coiled coil in order to direct it to a specific location in the cell for delivery of a target cargo. Following on from this, more recent experiments have shown that coiled-coil domains can be fused to various epitopes for the production of therapeutic antibodies. This is possible as a result of the experiments by Eriksson et al., which showed that the coiled coil itself did not elicit an immune response (*41*), and also because any immune response generated by fusion to an antigenic epitope will be amplified due to the multiple copies of protein chains coming from the coiled coil. Preliminary results (*47, 48*) are encouraging and further developments in this area are eagerly awaited.

**Acknowledgements**

This work was supported by a grant from the University of Manitoba Research and IPM Program. Dr. Jörg Stetefeld is supported by the Canada Research Chair Program. We thank Dr. Simone Karrasch and Mr. Michael McFarlane for their editorial assistance.

**References**

1. Cohen, C., and Parry, D. A. (1990) Alpha-helical coiled coils and bundles: how to design an alpha-helical protein, *Proteins 7*, 1-15.
2. Lupas, A. (1996) Coiled coils: new structures and new functions, *Trends Biochem. Sci. 21*, 375-382.
3. Kohn, W. D., Mant, C. T., and Hodges, R. S. (1997) Alpha-helical protein assembly motifs, *J. Biol. Chem. 272*, 2583-2586.
4. Walshaw, J., and Woolfson, D. N. (2001) Socket: a program for identifying and analysing coiled-coil motifs within protein structures, *J. Mol. Biol. 307*, 1427-1450.
5. Crick, F. H. C. (1953) The Fourier transform of a coiled-coil, *Acta Crystallogr. 6*, 685-689.
6. Crick, F. H. C. (1952) Is alpha-keratin a coiled coil?, *Nature 170*, 882-883.
7. Crick, F. H. C. (1953) The packing of alpha-helices: simple coiled-coils, *Acta Crystallogr. 6*, 689-697.
8. Malashkevich, V. N., Kammerer, R. A., Efimov, V. P., Schulthess, T., and Engel, J. (1996) The crystal structure of a five-stranded coiled coil in COMP: a prototype ion channel?, *Science 274*, 761-765.

9. Efimov, V. P., Engel, J., and Malashkevich, V. N. (1996) Crystallization and preliminary crystallographic study of the pentamerizing domain from cartilage oligomeric matrix protein: a five-stranded alpha-helical bundle, *Proteins 24*, 259-262.

10. Liu, J., Zheng, Q., Deng, Y., Cheng, C. S., Kallenbach, N. R., and Lu, M. (2006) A seven-helix coiled coil, *Proc. Natl. Acad. Sci. U. S. A. 103*, 15457-15462.

11. Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963) Stereochemistry of polypeptide chain configurations, *J. Mol. Biol. 7*, 95-99.

12. Peters, J., Baumeister, W., and Lupas, A. (1996) Hyperthermostable surface layer protein tetrabrachion from the archaebacterium Staphylothermus marinus: evidence for the presence of a right-handed coiled coil derived from the primary structure, *J. Mol. Biol. 257*, 1031-1041.

13. Kuhnel, K., Jarchau, T., Wolf, E., Schlichting, I., Walter, U., Wittinghofer, A., and Strelkov, S. V. (2004) The VASP tetramerization domain is a right-handed coiled coil based on a 15-residue repeat, *Proc. Natl. Acad. Sci. U. S. A. 101*, 17027-17032.

14. Hoiczyk, E., Roggenkamp, A., Reichenbecher, M., Lupas, A., and Heesemann, J. (2000) Structure and sequence analysis of Yersinia YadA and Moraxella UspAs reveal a novel class of adhesins, *EMBO J. 19*, 5989-5999.

15. Nooren, I. M. A., Kaptein, R., Sauer, R. T., and Boelens, R. (1999) The tetramerization a domain of the Mnt repressor consists of two right-handed coiled coils, *Nat. Struct. Biol. 6*, 755-759.

16. Pauling, L., Corey, R. B., and Branson, H. R. (1951) The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain, *Proc. Natl. Acad. Sci. U. S. A. 37*, 205-211.

17. Tarbouriech, N., Curran, J., Ruigrok, R. W. H., and Burmeister, W. P. (2000) Tetrameric coiled coil domain of Sendai virus phosphoprotein, *Nat. Struct. Biol. 7*, 777-781.

18. Pauling, L., and Corey, R. B. (1953) Compound helical configurations of polypeptide chains; structure of proteins of the alpha -keratin type, *Nature (London, United Kingdom) 171*, 59-61.

19. Stetefeld, J., Jenny, M., Schulthess, T., Landwehr, R., Engel, J., and Kammerer, R. A. (2000) Crystal structure of a naturally occurring parallel right-handed coiled coil tetramer, *Nat. Struct. Biol. 7*, 772-776.

20. Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins, *Science 240*, 1759-1764.

21. Lupas, A. N., and Gruber, M. (2005) The structure of alpha-helical coiled coils, *Adv. Protein Chem. 70*, 37-78.

22. Stone, D., Sodek, J., Johnson, P., and Smillie, L. B. (1975) Tropomyosin: correlation of amino acid sequence and structure, *Proc. IX FEBS Meeting 31*, 125-136.

23. Parry, D. A. D. (1975) Analysis of the primary sequence of alpha -tropomyosin from rabbit skeletal muscle, *J. Mol. Biol. 98*, 519-535.

24. McLachlan, A. D., and Stewart, M. (1975) Tropomyosin coiled-coil interactions. Evidence for an unstaggered structure, *J. Mol. Biol. 98*, 293-304.

25. McLachlan, A. D., and Karn, J. (1983) Periodic features in the amino acid sequence of nematode myosin rod, *J. Mol. Biol. 164*, 605-626.

26. Lupas, A., Van Dyke, M., and Stock, J. (1991) Predicting coiled coils from protein sequences, *Science 252*, 1162-1164.

27. Berger, B., Wilson, D. B., Wolf, E., Tonchev, T., Milla, M., and Kim, P. S. (1995) Predicting coiled coils by use of pairwise residue correlations, *Proc. Natl. Acad. Sci. U. S. A. 92*, 8259-8263.

28. Wolf, E., Kim, P. S., and Berger, B. (1997) MultiCoil: a program for predicting two- and three-stranded coiled coils, *Protein Sci. 6*, 1179-1189.

29. Delorenzi, M., and Speed, T. (2002) An HMM model for coiled-coil domains and a comparison with PSSM-based predictions, *Bioinformatics 18*, 617-625.

30. Wilson, I. A., Skehel, J. J., and Wiley, D. C. (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution, *Nature 289*, 366-373.

31. O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil, *Science 254*, 539-544.

32. Harbury, P. B., Kim, P. S., and Alber, T. (1994) Crystal structure of an isoleucine-zipper trimer, *Nature 371*, 80-83.

33. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants, *Science 262*, 1401-1407.

34. Sodek, J., Hodges, R. S., Smillie, L. B., and Jurasek, L. (1972) Amino-acid sequence of rabbit skeletal tropomyosin and its coiled-coil structure, *Proc. Natl. Acad. Sci. U. S. A. 69*, 3800-3804.

35. Burkhard, P., Stetefeld, J., and Strelkov, S. V. (2001) Coiled coils: a highly versatile protein folding motif, *Trends Cell Biol. 11*, 82-88.

36. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T., and Kimt, P. S. (1998) High-resolution protein design with backbone freedom, *Science 282*, 1462-1467.

37. Peters, J., Nitsch, M., Kuhlmorgen, B., Golbik, R., Lupas, A., Kellermann, J., Engelhardt, H., Pfander, J. P., Muller, S., Goldie, K., and et al. (1995) Tetrabrachion: a filamentous archaebacterial surface protein assembly of unusual structure and extreme stability, *J. Mol. Biol. 245*, 385-401.

38. Meier, M., and Burkhard, P. (2006) Statistical analysis of intrahelical ionic interactions in alpha -helices and coiled coils, *J. Struct. Biol. 155*, 116-129.

39. Guo, Y., Bozic, D., Malashkevich, V. N., Kammerer, R. A., Schulthess, T., and Engel, J. (1998) All-trans retinol, vitamin D and other hydrophobic compounds bind in the axial pore of the five-stranded coiled-coil domain of cartilage oligomeric matrix protein, *EMBO J. 17*, 5265-5272.

40. Ozbek, S., Engel, J., and Stetefeld, J. (2002) Storage function of cartilage oligomeric matrix protein: the crystal structure of the coiled-coil domain in complex with vitamin D(3), *EMBO J. 21*, 5960-5968.

41. Eriksson, M., Hassan, S., Larsson, R., Linder, S., Ramqvist, T., Lovborg, H., Vikinge, T., Figgemeier, E., Muller, J., Stetefeld, J., Dalianis, T., and Ozbek, S. (2009) Utilization of a right-handed coiled-coil protein from archaebacterium Staphylothermus marinus as a carrier for cisplatin, *Anticancer Res. 29*, 11-18.

42. Boulikas, T., and Vougiouka, M. (2004) Recent clinical trials using cisplatin, carboplatin and their combination chemotherapy drugs (review), *Oncol. Rep. 11*, 559-595.

43. Tomschy, A., Fauser, C., Landwehr, R., and Engel, J. (1996) Homophilic adhesion of E-cadherin occurs by a co-operative two-step interaction of N-terminal domains, *EMBO J. 15*, 3507-3514.

44. Ahrens, T., Pertz, O., Haussinger, D., Fauser, C., Schulthess, T., and Engel, J. (2002) Analysis of heterophilic and homophilic interactions of cadherins using the c-Jun/c-Fos dimerization domains, *J. Biol. Chem. 277*, 19455-19460.

45. Takeichi, M. (1988) The cadherins: cell-cell adhesion molecules controlling animal morphogenesis, *Development (Cambridge, England) 102*, 639-655.

46. Rimm, D. L., and Morrow, J. S. (1994) Molecular cloning of human E-cadherin suggests a novel subdivision of the cadherin superfamily, *Biochem. Biophys. Res. Commun. 200*, 1754-1761.

47. Schroeder, U., Graff, A., Buchmeier, S., Rigler, P., Silvan, U., Tropel, D., Jockusch, B. M., Aebi, U., Burkhard, P., and Schoenenberger, C.-A. (2009) Peptide Nanoparticles Serve as a Powerful Platform for the Immunogenic Display of Poorly Antigenic Actin Determinants, *J. Mol. Biol. 386*, 1368-1381.

48. Pimentel, T. A. P. F., Yan, Z., Jeffers, S. A., Holmes, K. V., Hodges, R. S., and Burkhard, P. (2009) Peptide nanoparticles as novel immunogens: design and analysis of a prototypic severe acute respiratory syndrome vaccine, *Chem. Biol. Drug Des. 73*, 53-61.

49. Raman, S., Machaidze, G., Lustig, A., Aebi, U., and Burkhard, P. (2006) Structure-based design of peptides that self-assemble into regular polyhedral nanoparticles, *Nanomed. 2*, 95-102.

50. Whitby, F. G., and Phillips, G. N., Jr. (2000) Crystal structure of tropomyosin at 7 Angstroms resolution, *Proteins 38*, 49-59.

51. Brown, J. H., Kim, K.-H., Jun, G., Greenfield, N. J., Dominguez, R., Volkmann, N., Hitchcock-DeGregori, S. E., and Cohen, C. (2001) Deciphering the design of the tropomyosin molecule, *Proc. Natl. Acad. Sci. U. S. A. 98*, 8496-8501.

52. Li, Y., Mui, S., Brown, J. H., Strand, J., Reshetnikova, L., Tobacman, L. S., and Cohen, C. (2002) The crystal structure of the C-terminal fragment of striated-muscle alpha - tropomyosin reveals a key troponin T recognition site, *Proc. Natl. Acad. Sci. U. S. A. 99*, 7378-7383.

53. Brown, J. H., Volkmann, N., Jun, G., Henschen-Edman, A. H., and Cohen, C. (2000) The crystal structure of modified bovine fibrinogen, *Proc. Natl. Acad. Sci. U. S. A. 97*, 85-90.

54. Yang, Z., Mochalkin, I., Veerapandian, L., Riley, M., and Doolittle, R. F. (2000) Crystal structure of native chicken fibrinogen at 5.5-A resolution, *Proc. Natl. Acad. Sci. U. S. A. 97*, 3907-3912.

55. Yang, Z., Kollman, J. M., Pandi, L., and Doolittle, R. F. (2001) Crystal structure of native chicken fibrinogen at 2.7 A resolution, *Biochemistry (Mosc.) 40*, 12515-12523.

56. Lee, H. H., Elia, N., Ghirlando, R., Lippincott-Schwartz, J., and Hurley, J. H. (2008) Midbody Targeting of the ESCRT Machinery by a Noncanonical Coiled Coil in CEP55, *Science 322*, 576-580.

57. Yun, M., Bronner, C. E., Park, C. G., Cha, S. S., Park, H. W., and Endow, S. A. (2003) Rotation of the stalk/neck and one head in a new crystal structure of the kinesin motor protein, Ncd, *EMBO J. 22*, 5382-5389.

58. Strelkov, S. V., Herrmann, H., Geisler, N., Wedig, T., Zimbelmann, R., Aebi, U., and Burkhard, P. (2002) Conserved segments 1A and 2B of the intermediate filament dimer: their atomic structures and role in filament assembly, *EMBO J. 21*, 1255-1266.

59. Higgins, M. K., Bokma, E., Koronakis, E., Hughes, C., and Koronakis, V. (2004) Structure of the periplasmic component of a bacterial drug efflux pump, *Proc. Natl. Acad. Sci. U. S. A. 101*, 9994-9999.

60. Koronakis, V., Sharff, A., Koronakis, E., Luisi, B., and Hughes, C. (2000) Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export, *Nature 405*, 914-919.

61. Deng, Y., Liu, J., Zheng, Q., Yong, W., and Lu, M. (2006) Structures and Polymorphic Interactions of Two Heptad-Repeat Regions of the SARS Virus S2 Protein, *Structure 14*, 889-899.

## Chapter 3: The coiled-coil domain of Cartilage Oligomeric Matrix Protein binds fatty acids via a ring of conserved glutamines

This chapter is based on the following manuscript in preparation for submission to the journal "*The Journal of Biological Chemistry*".

McFarlane, A. A., Khajepour, M., and Stetefeld, J. (2011). The coiled-coil domain of Cartilage Oligomeric Matrix Protein binds fatty acids via a ring of conserved glutamines.


### Abstract

Here we present high-resolution crystal structures of five naturally-occurring fatty acids in complex with the coiled-coil domain of cartilage oligomeric matrix protein (COMPcc). Binding is mediated by a ring of glutamines at position 54, which form a network of hydrogen bonds assembled into a funnel-like arrangement. In addition, we show by fluorescence spectroscopy the binding behaviour in solution. Our findings reveal a binding trend within the hydrophobic channel of COMPcc that is driven by length of the methylene tail and incorporation of unsaturation. This opens the exciting possibility that COMPcc could be developed as a targeted drug delivery system.



**Key Words**: COMPcc, saturated fatty acids, unsaturated fatty acids, pentamer, coiled coil, drug delivery

## Introduction

Cartilage oligomeric matrix protein (COMP) is a non-collagenous glycoprotein of the thrombospondin family that is found in cartilage (*1*), tendon (*2, 3*), and ligament (*4*). It is a homopentamer consisting of five subunits held together by interchain disulfide bridges in the N-terminal coiled-coil domain (residues 27-72). The crystal structure of the recombinant coiled-coil domain of COMP (COMPcc) revealed a 73Å-long axial pore with a diameter of 2–6Å. This pore is divided into two hydrophobic cavities by a ring of conserved glutamines at residue 54, which form a network of hydrogen bonds (*5, 6*). The binding of numerous biologically relevant hydrophobic compounds to recombinantly expressed COMPcc has been shown, and the crystal structures of the COMPcc-vitamin $D_3$ (*7*), COMPcc-all-*trans* retinol, and COMPcc-benzene complexes have been solved (*8*). The binding properties of the hydrophobic channel suggest the potential of COMPcc to be used as a storage and delivery system for hydrophobic compounds (*9*).

Saturated fatty acids (SFAs) have diverse and important biological functions in cells. They are involved in protein acylation, transcription regulation, apoptosis, energy production and storage, and membrane synthesis (*10, 11*). SFAs comprise approximately 30–40% of total fatty acids in animal tissues, with the majority being palmitic acid (15–25%), followed by stearic acid (10–20%), myristic acid (0.5–1%), and lauric acid (<0.5%) (*10, 12*). Myristic acid and palmitic acid are important cell signaling molecules as they are involved in the two types of protein fatty acid acylation: N-terminal myristoylation and S-palmitoylation. N-terminal myristoylation is the attachment of a modified myristic acid to the $NH_2$-terminal glycine of a protein via an amide linkage. It is estimated that ~0.5% of all eukaryotic proteins are N-myristoylated (*13*). Proteins with these modifications have important functions in cell signaling pathways, oncogenes,

membrane anchoring, and virus structural proteins (*14*). S-palmitoylation is the attachment of palmitic acid to a cysteine residue of a protein via a thioester linkage. It is a reversible association and is primarily involved in regulating protein-membrane interactions and trafficking of proteins. Many cell signaling proteins undergo both N-myristoylation and S-palmitoylation (*15*).

It has long been thought that a high dietary intake of SFAs is associated with elevated cholesterol levels and increased risk of coronary heart disease and cardiovascular disease. However, more recent studies have suggested that this may not be the case (*10, 12, 16*). An intake of stearic acid has been shown to have no adverse effect on cardiovascular disease and to decrease low-density lipoprotein (LDL) cholesterol (*17*). It has also been reported that a moderate intake of myristic acid (between 1.2 and 1.8% of total energy) has a favourable effect on the activity of lecithin-cholesterol acyltransferase, a key enzyme in cholesterol removal from tissues (*18*).

Natural receptors for SFAs include members of the albumin and fatty acid-binding protein (FABP) families. These proteins serve to increase the solubility of fatty acids and mediate their transport within cells. While there are many members of the FABP family with a great deal of variance in protein sequence, all members share a common β-barrel structural motif. The 10-stranded antiparallel β-barrel contains a hydrophobic core to which fatty acids bind. The core is capped on one end by an N-terminal helix-turn-helix motif. Inside the binding pocket, the carboxyl group is coordinated through electrostatic interactions with tyrosine and 2 arginine residues and the hydrocarbon tail is oriented with hydrophobic residues on one side and ordered water molecules on the other side (*19, 20*).

In the current work, we have solved the crystal structures of COMPcc in complex with myristic acid (C14:0), palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1) and arachadic acid (C20:0). In agreement with fluorescence spectroscopy studies in solution, we can decipher a trend in binding favourability that is determined by length of the aliphatic tail and geometry altered by the introduction of a *cis*-configured double bond.

## Results

### *X-Ray structures of the individual COMPcc-fatty acid complexes*

The coiled-coil domain of COMP comprising residues 20-72 was obtained by recombinant expression in *E. coli* as described previously (see also Materials and Methods) (*5*). The crystal structure of the COMPcc/fatty acid complexes were solved by molecular replacement using the apo-COMPcc version (PDB code 1MZ9) (*21*) as a search template (Fig. 3.1; see also Table 3.1). The COMPcc chain fragment forms a parallel left-handed coiled-coil pentamer with an average length of 70 Å and an average outer diameter of about 30 Å. The axial pore of the pentamer is divided by the hydrophilic Gln54 ring system into two hydrophobic cavities exclusively lined with aliphatic side chains (*6*). The electron density for the protein is well defined with exceptions in chain E showing disordering of the last seven C-terminal residues (66-72). Accordingly, the ring of interchain disulfide bridges between cysteines 71 and 68 of two neighbour chains is disrupted at this position leaving chain D and E disconnected in the otherwise fully oxidized pentamer.

In the individual COMPcc-fatty acid complex structures, one molecule of the respective fatty acid is bound inside of the N-terminal hydrophobic compartment in a linear, elongated conformation, with the length axis of the ligands paralleling the channel symmetry (Fig. 3.1).

The fatty acids are retained in the binding pocket due to an interaction between the highly electronegative carboxylate head group and the elaborate hydrogen bonding network formed by the Gln54 ring. The hydrogen bonds of the Gln54 ring are assembled into a funnel-like arrangement, such that partial positive charges on the amide nitrogens are oriented towards the bottom of the funnel and partial negative charges on the carbonyl oxygens are oriented towards the top. This creates a dipole, which is parallel to the dipole moment of the α-helices. The positively-charged bottom of the funnel can act as a trap for negatively-charged ions, as is seen in the native structure of COMPcc where a chloride ion is bound (*6*).

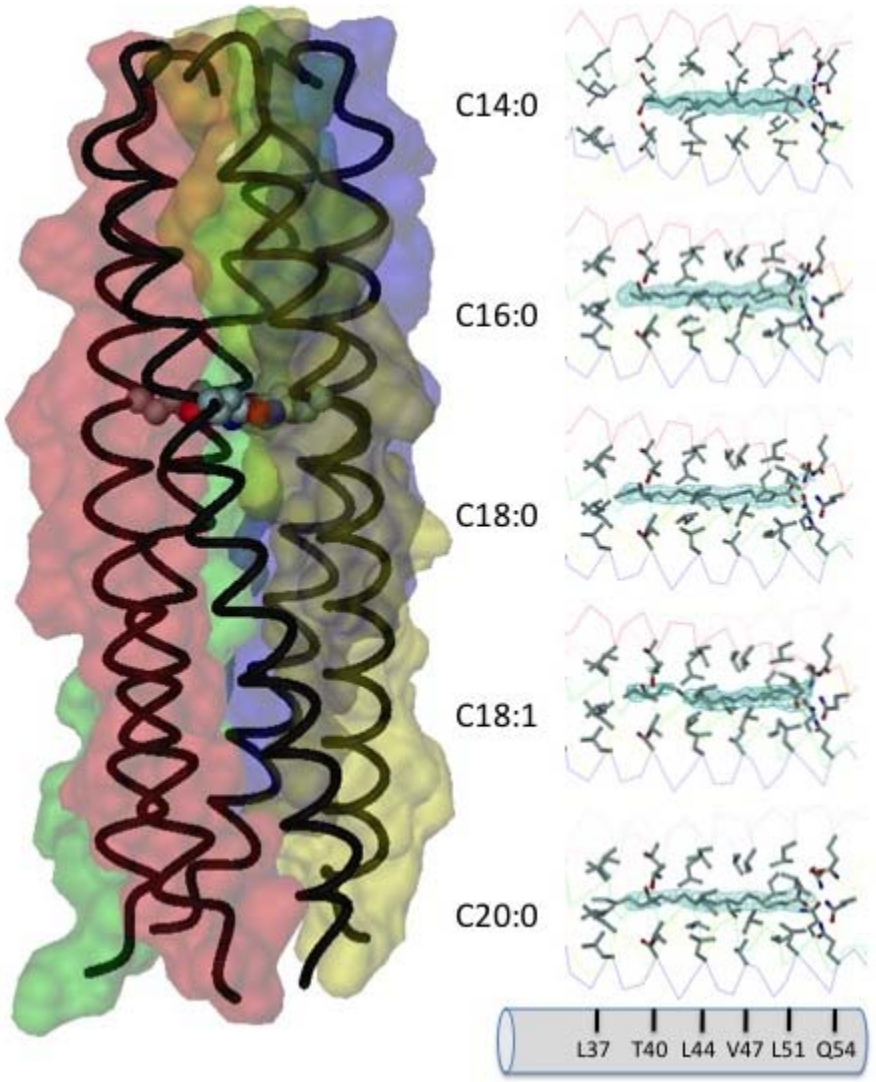


**Fig. 3.1. Crystal structures of COMPcc in complex with fatty acids.**

Overall view of the individual 2Fo-Fc-omit maps (1.2σ contour level) superimposed on: myristic acid (C14:0), palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), and arachadic acid (C20:0). The five water molecules surrounding the methylene tail are shown as red balls and green electron density maps. Amino acid residues forming knobs-into-holes according the heptad repeat pattern are shown in atom color type with *a* and *d* positions indicated. A superimposition of all five complexes is shown on the left.

**Table 3.1: Data collection and refinement data.**

| **Data collection** | | | | | |
|---|---|---|---|---|---|
| Fatty acid | ***C14*** | ***C16*** | ***C18*** | ***C18 :1*** | ***C20*** |
| Resolution range (Å) | 20.6-1.8 (1.9-1.8) | 24-2.2 (2.32-2.2) | 29.4-1.87 (1.99-1.87) | 29.4-2.75 (2.92-2.75) | 27.9-3.0 (3.19-3.0) |
| Unique reflections | 17542 | 9799 | 15795 | 4761 | 3580 |
| Completeness % | 94.9 (94.3) | 99.1 (98.7) | 99.3 (95.7) | 92.6 (93.0) | 94.9 (93) |
| Redundancy | 2.0 | 4.4 | 4.0 | 2.9 | 2.9 |
| $R_{sym}$ [1] | 10.1 (31.8) | 8.1 (29.6) | 7.4 (28.9) | 8.3 (32.4) | 14.5 (29.1) |
| Cell dimensions (Å,°) | | | | | |
| A | 38.35 | 38.7 | 37.9 | 37.86 | 37.8 |
| B | 49.39 | 48.5 | 48.9 | 49.11 | 48.9 |
| C | 54.82 | 53.3 | 53.9 | 54.25 | 53.1 |
| β | 103.86 | 103.8 | 104.1 | 104.14 | 103.5 |
| **Refinement** | | | | | |
| Rfactor[2] (%), no sigma cutoff | 21.2 (26) | 20.8 (23.0) | 20.7 (28.5) | 20.1 (31.3) | 20.3 (23.7) |
| $R_{free}$ (%), no sigma cutoff | 27.6 (33) | 27.5 (31.1) | 25.5 (35.7) | 27.5 (36.3) | 23.2 (28.5) |
| Average B factors ($Å^2$) | | | | | |
| Protein/ Water/ Ligands | 26.3/44.3/35.2 | 48.9/66.1/45.8 | 29.6/39.3/38.5 | 41.6/38.5/37.1 | 32.4/34.7/37.9 |
| Bond length (Å)[3] | 0.005 | 0.006 | 0.005 | 0.008 | 0.008 |
| Bond angles (°)[3] | 0.9 | 1.0 | 0.9 | 1.1 | 1.2 |
| Ramachandran plot[4] | 96.7/2.4/0.5/0.5 | 95.7/3.3/0/1 | 97.6/1.5/0/1.0 | 93.8/5.2/0.5/0.5 | 93.8/4.3/1/1 |

Values for the highest resolution shells are in parentheses.

[1] $R_{sym} = 100 * \Sigma I - \langle I \rangle | / \Sigma I$.

[2] $R_{factor} = \Sigma||F_{obs}|-|F_{calc}|/\Sigma|F_{obs}|$.

[3] Root-mean-square error.

[4] Percentage of residues in most favoured/ additional allowed/ generously allowed and disallowed regions. All datasets were measured at a wavelength of 1.5418Å.

The aliphatic side chains inside of the coiled coil cause 6 regular constrictions to the diameter of the pore, varying in size between 2 and 6 Å as defined by the van der Waals radii (see also Fig. 3.2). The aliphatic side chains are located in the hydrophobic cavities between Thr40 and Leu51 to accommodate different variations of the long fatty acid tail. The terminal methylene groups are fixed by Thr40 (C14:0), Leu37-Thr40 (C16:0) and Leu37 (C18:0). This coupling is originated by the length extension of fully saturated elongated chains. Remarkable differences can be detected for C18:1, which is fixed to this *cis*-double bond kink at Leu44, the remaining portion of the aliphatic tail remains highly disordered. Also the C20:0 fatty acid complex structure shows a significantly disordered tail for the methylene group C18-C20.

The opening of the unliganded channel had been described to be maximal at positions Val47 (*6*), which here serve to accommodate the *cis*-configured double bond in the C18:1 systems of the oleic acid molecule. The sp2 hybridized double bond is tightly fitted into a

hydrophobic ring of β-branched residues in heptad position d, Val47. This interaction imposes a planar 9-s-*trans* conformation on the oleic molecules in the COMPcc. The C9-C10 bond exhibits a torsion angle of 162°. The methyl C11 points towards Leu44 in close van der Waals contacts to the side chain atoms of three of the five chains (distances between 2.5Å and 3.5Å).

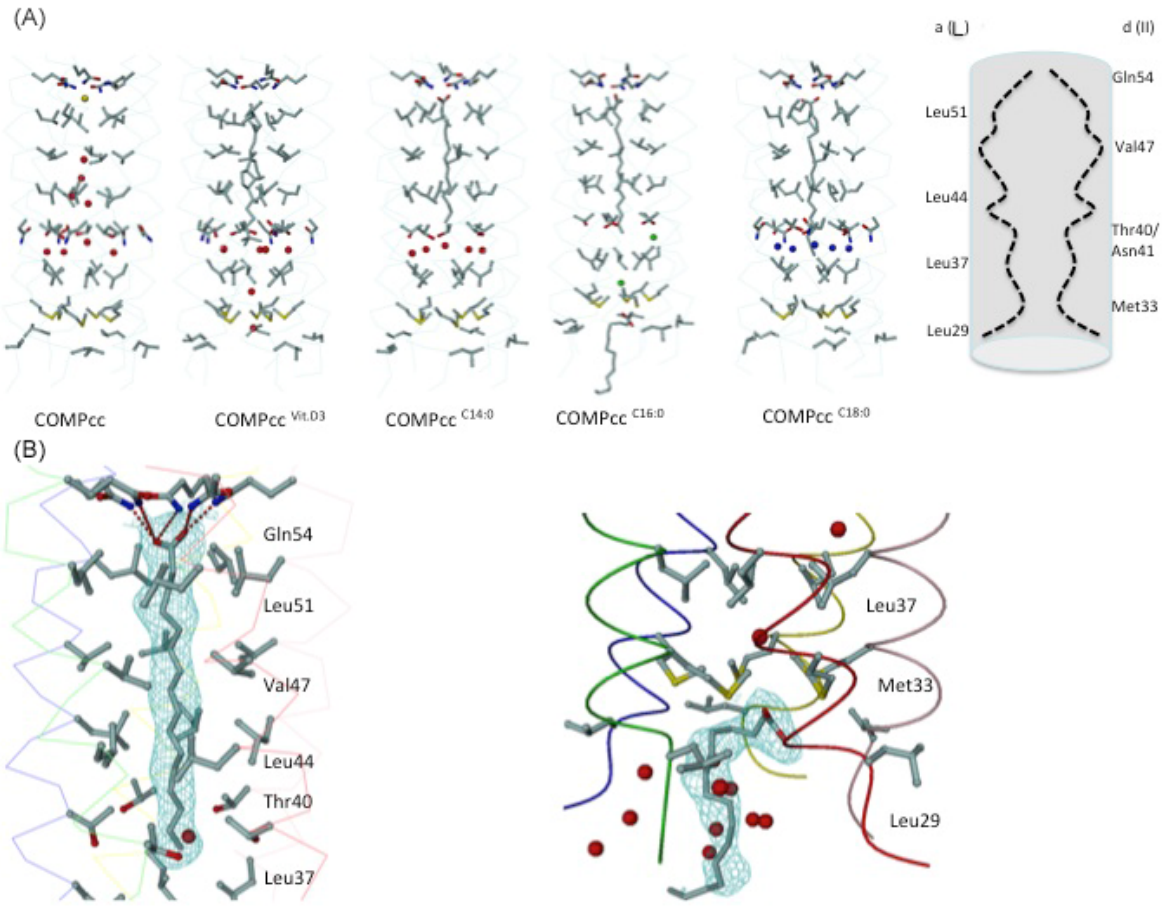


**Fig. 3.2. Binding mechanism.**

(A) Comparison of wild type COMPcc (left) with the pentameric channel complexed with vitamin $D_3$, myristic acid (C14:0), palmitic acid (C16:0), and stearic acid (C18:0). The inner radius properties of COMPcc are shown at the very right. (B) High resolution electron density map of a palmitic acid molecule inside the hydrophobic channel (left) and outside the channel

(right). Hydrogen bond distances are drawn as dotted lines and aliphatic knobs-into-holes residues are labelled. Water molecules are shown as red spheres.

***Fluorescence spectroscopy***

The probe cis-parinaric acid (CPA) is a fluorescence probe often used to investigate the fatty acid binding sites of proteins. The free probe has low fluorescence in aqueous solution, when the probe binds proteins its fluorescence is increased approximately five-fold (Fig. 3.3a). This increase can be used to determine the binding constant of CPA to the protein. From the fluorescence enhancement profile, the fraction of ligand bound protein can be calculated via (eq-a):

$$\text{Fraction of ligand bound protein} = \frac{F - F_0}{F_{420} - F_0} = f \qquad \text{(a)}$$

Where $F_0$ is the fluorescence of protein sample when no CPA has been added, F is the protein fluorescence at any given CPA concentration and $F_{420}$ is the protein fluorescence in the presence of 3 μM of CPA. In the case of one ligand binding site, *f* follows a hyperbolic dependence upon ligand concentration given by (eq-b):

$$f = \frac{B[L]_{free}}{K_d + [L]_{free}} \qquad \text{(b)}$$

Where *B* is a constant close to one, $K_d$ is the binding constant and $[L]_{free}$ is the concentration of free ligand (in this case CPA). The data in Fig. 3.3b show a good hyperbolic correlation; therefore, the binding of CPA to COMPcc is a hyperbolic one site binding with a binding constant of 0.7 ± 0.1 μM.

The probe CPA can also be used to characterize the binding of other fatty acids to COMPcc. The addition of fatty acids (FA) to the CPA-COMPcc complex will displace CPA

leading to a decrease in fluorescence. This effect is shown in Fig. 3.3c. This decrease in fluorescence follows a hyperbolic profile as shown in the correlation line and from the correlation line we can obtain $[FA]_{1/2}$ the concentration of fatty acid at which the fluorescence is half that of the original. Values of $[FA]_{1/2}$ for a variety of fatty acids are given in Table 3.2.

If the amounts of COMPcc and CPA are kept significantly lower than the $K_d$, the following equilibria exist in solution:

$$COMPcc + CPA \rightleftarrows COMPcc - CPA \qquad K_d^{CPA} = \frac{[CPA][COMPcc]}{[CPA - COMPcc]}$$

$$COMPcc + FA \rightleftarrows COMPcc - FA \qquad K_d^{FA} = \frac{[FA][COMPcc]}{[FA - COMPcc]}$$

Let us now consider the case in which the concentration of COMPcc is kept low both relative to $K_d^{CPA}$ and relative to CPA and the concentration of CPA is also significantly lower than $K_d^{CPA}$. In these experiments the initial concentration of COMPcc is ~ 0.1μM and CPA is equal to 0.4 μM. In the case where there is no FA added the concentrations are $[COMPcc] = [COMPcc]_0$ and $[CPA] \approx [CPA]_0 = 0.4$ μM. In this case the concentration of the CPA-COMPcc complex will be equal to $[CPA - COMPcc]_0 = \frac{[CPA]_0[COMPcc]_0}{K_d^{CPA}}$. At $[FA]_{1/2}$, the concentration of the COMPcc-CPA complex is equal to $[CPA - COMPcc] = \frac{1}{2}\frac{[CPA]_0[COMPcc]_0}{K_d^{CPA}}$. Writing a mass balance for COMPcc we obtain:

$$[COMPcc]_0 = [COMPcc] + [COMPcc - CPA] + [COMPcc - FA]$$

when $[FA] = [FA]_{1/2}$

$$[COMPcc]_0 = [COMPcc] + \frac{1}{2}\frac{[CPA]_0[COMPcc]_0}{K_d^{CPA}} + \frac{[FA]_{1/2}[COMPcc]}{K_d^{FA}}$$

$$[COMPcc]_0\left(1 - \frac{1}{2}\frac{[CPA]_0}{K_d^{CPA}}\right) = [COMPcc]\left(1 + \frac{[FA]_{1/2}}{K_d^{FA}}\right) \quad \text{(e)}$$

Therefore by plugging equation e in the CPA equilibrium constant we obtain the following relationship at fatty acid concentrations equal to $[FA]_{1/2}$:

$$K_d^{CPA} = \frac{[CPA][COMPcc]}{[CPA - COMPcc]} = \frac{[CPA]_0[COMPcc]_0\left(\frac{1 - \frac{[CPA]_0}{K_d^{CPA}}}{1 + \frac{[FA]_{1/2}}{K_d^{FA}}}\right)}{\frac{1}{2}\frac{[CPA]_0[COMPcc]_0}{K_d^{CPA}}}$$

$$\frac{1}{2} = \frac{1 - \frac{[CPA]_0}{K_d^{CPA}}}{1 + \frac{[FA]_{1/2}}{K_d^{FA}}} \quad \text{(f)}$$

Therefore the binding constant $K_d^{FA}$ can be calculated using $[FA]_{1/2}$. These results are tabulated in Table 3.2.

**Table 3.2: Fluorescence spectroscopy data.**

| Cargo | $[FA]_{1/2}$ (μM) | $K_d$ (μM) |
|---|---|---|
| C14:0 [Myristic acid] | 0.3 | 0.19 ± 0.08 |
| C16:0 [Palmitic acid] | 0.22 | 0.14 ± 0.06 |
| C18:0 [Stearic acid] | 0.195 | 0.12 ± 0.05 |
| C18:1 [Oleic acid] | 0.39 | 0.25 ± 0.09 |
| C20:0 [Arachadic acid] | 0.28 | 0.18 ± 0.07 |
| cis-Parinaric acid | N/A | 0.7 ± 0.1 |
| Wt vitamin A (*22*) | N/A | 0.6 |
| Q54I vitamin A (*22*) | N/A | 0.8 |

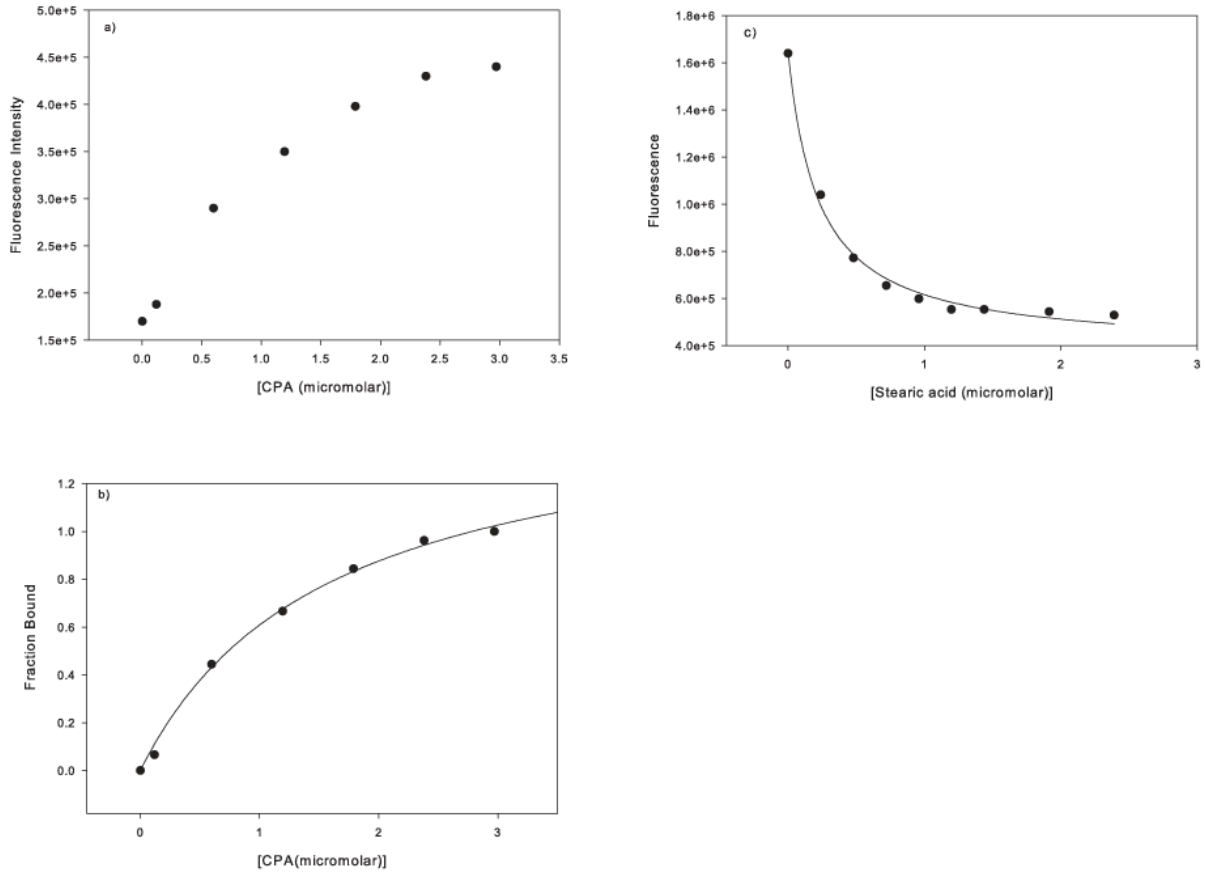


**Fig. 3.3. Fluorescence spectra.**

(a) Fluorescence spectrum of CPA bound to COMPcc in solution, measured at 420nm. (b) The fraction of CPA bound to COMPcc follows a hyperbolic one-site binding with a binding constant of 0.7 ± 0.1 μM. The curve is obtained by fitting the data to eq-b in the text. (c) The effect of fatty acids on the fluorescence of the COMPcc-CPA complex. In this case stearic acid is shown. The decrease in fluorescence follows a hyperbolic trend, with the correlation line used to find $[FA]_{1/2}$ (see text).

## Discussion

### *Binding mechanism*

The individual COMPcc-fatty acid complex structures show that one molecule of each fatty acid is bound in the "lower" N-terminal cavity of the hydrophobic channel (Fig. 3.1). A diffusion via the N-terminus for the lipophilic cargos ligand seems to be plausible. Additional electron density in the cocrystal structure of palmitic acid (C16:0) underlines this fact (Fig. 3.2B). This assumption is underlined by changes within the helical backbone at the very N-terminus and reorientations at the Thr40 ring to accommodate the methylene tail fixation. The pentameric water ring between Leu37 and Thr40 is altering significantly (Fig. 3.2). These facts taken together suggest that the N-terminal position is the primary binding site in COMPcc for hydrophobic compounds, although data concerning the oxidative state of native COMP in cartilage are still lacking.

The Gln54 residue belongs to a four amino acid motif (QVKE) that is conserved among the pentameric thrombospondins (TSP-3, TSP-4, and COMP) (*23*). Gln54 is in position *d* of the characteristic heptad repeat $(a\text{-}g)_n$, which is unusual as the *a* and *d* positions are normally comprised of hydrophobic residues. It adopts a counterclockwise knobs-into-holes configuration with its Cα-Cβ vector (knob) oriented against the Cα-Cα vector of Q54-V55 (hole) of the neighbouring chain. Interestingly, it was shown that the thermal stability of the COMPcc was increased from 73°C to 104°C when Gln54 was mutated to a Leu residue (*24*). This implies an evolutionary advantage of the less thermostable wild type COMPcc over the Q54L mutant and suggests an additional function of the glutamine inside the pentameric channel. This decrease in thermal stability can be compensated by the addition of ligands, as the midpoint transition

temperature ($T_m$) of unfolding was increased by 2°C when benzene and cyclohexane were bound, and increased by 8 and 10°C when vitamin $D_3$ and elaidic acid were bound, respectively (*22*).

A comparison of the various energy-minimized COMPcc structures shows that the most energetically favoured state is when a cargo is loaded into the system. For example, the uncomplexed glutamine ring has a ΔG of –6.02 kcal/mol in comparison to –9.8 kcal/mol when C16:0 is bound. This data provides further support for ligand binding role of the Gln54 ring.

***Biological Implications***

Coiled-coil proteins such as COMPcc are attractive candidates for the design of drug delivery systems (*9*). The "breathing effect" demonstrated by the dynamic opening of the COMPcc channel to accommodate spacious molecules clearly illustrates the remarkable capability of this protein to adapt to its bound cargo. This effect is especially dramatic in the COMPcc-vitamin $D_3$ complex, in which the volume of the cavities increases by ~30% upon binding of the ligand (*21*). Drug design strategies involving COMPcc can be aimed at utilizing the affinity of the hydrophobic channel for diverse chemical compounds to load potential therapeutics into the binding pocket. The protein termini can be exploited for targeted drug delivery approaches as potential attachment sites for binding partners of cell surface markers of interest (*9*).

The x-ray crystal data along with the solution data reveal an overall binding trend for the fatty acids to COMPcc. As the length of the methylene tail is increased from C14 to C18, the $K_d$ values decrease, indicating a tighter binding within the COMPcc channel. Introduction of a double bond in the case of oleic acid causes an increase in the $K_d$, as the protein must accommodate the *cis*-double bond of the fatty acid. Interestingly, the most favourable fatty acid

ligand for COMPcc is the C18:0 stearic acid. A diet containing high amounts of stearic acid has been reported in some cases to lower both LDL and HDL cholesterol compared with other SFA and unsaturated fatty acid diets (*17*). Taken together, our data suggest a putative therapeutic role for a COMPcc carrier-cargo system.

## Conclusion

In the present work we have demonstrated the remarkable ability of COMPcc to bind biologically significant hydrophobic molecules. The x-ray crystal structure of COMPcc in complex with five naturally-occurring fatty acids reveals a potential target for the design of drug delivery systems. In the crystal structure, ligands are tightly bound within the inner pore of COMPcc, with binding mediated by a ring of glutamines at position 54. We also confirmed the ligand binding in solution through fluorescence spectroscopy where the $K_d$ values are in the micromolar range. Our data suggest an exciting prospective for this new and emerging field.

## Materials and Methods

### *Expression and purification of recombinant COMPcc*

The coiled-coil domain of rat COMPcc comprising residues 27–72 was prepared as described previously (*5, 8*). Purified COMPcc was dialyzed against PBS pH 7.4 and concentrated to 10mg/mL with a 10kDa Amicon membrane (Millipore).

### *Crystallization and Data Collection*

Crystallization experiments were performed at room temperature employing the vapour diffusion technique. Hanging droplets were made by mixing 2µl protein solution (10 mg/ml)

with 0.2 M sodium acetate, 0.1 M Tris-HCl, pH 8.5 and 30% PEG 4000. Individual fatty acids obtained from Sigma were soaked into the crystals overnight. The crystals belong to spacegroup $P2_1$ and contain one molecule of the pentameric COMPcc within the asymmetric unit. To analyze the influence of different effectors (pH, ions and organic solvents) four crystal structures performing different crystallization conditions were determined (data not shown). The high resolution data set were collected at synchrotron CLS (PX-Beamline) on a MAR research imaging plate detector. Diffraction images were processed using program suite MOSFLM (*25*) and the structure factors were scaled and reduced using SCALA from the CCP4 package (*26*). Statistics of the merged data are given in Table 3.1.

***Structure determination and refinement***

Molecular replacement was performed using the AMORE program of the CCP4 package (*26*). A Poly-serine model of native COMPcc structure (PDB-code:1MZ9) was used as search template. Positional refinement was performed with CNS using the maximum likelihood method (*27*). 10% of the reflections were excluded for use in a cross validation set. Refinement with CNS was alternated with manual electron density refitting of side-chains and terminal regions using MAIN (*28*). At this stage the individual fatty acid molecules have been fitted into a 3.0σ contoured Fo-Fc difference map. To determine the favoured axial orientation of the ligands within the pentameric channel a 2° stepwise refinement (conjugated gradient minimization together with individual B-factor refinement) along the five-fold local symmetry axis was performed. Interpretation of the electron density maps for each solution together with monitoring of the Rfree/Rvalue ratio revealed that one orientation is preferred which is shown in Fig. 3.1. In further refinement overall anisotropic B-factor and bulk solvent corrections were utilized.

Simulated annealing omit maps confirmed the correctness of the protein and ligand structures. Water molecules were added chosen by distance criteria and hydrogen bonding geometry and were tested for position in spherical density, reasonable temperature factors, real space R-values, and improvement of the R-factors.

***Fluorescence spectroscopy of COMPcc-fatty acid complexes***

Steady-state fluorescence spectra were measured on a Fluorolog-3 Horiba Jobin Yvon spectrofluorometer (Edison, NJ). The sample was held in a 10 × 10mm quartz cuvette equipped with a continuous stirrer. The data was analyzed with Sigmaplot (Point Richmond, CA) software. The reaction was thermostatically controlled at 25$^{o}$C by a Jeio-Tech refrigerating bath circulator (Des Plaines, IL). All fluorescence data were collected in 1X PBS buffer, pH 7.4.

**References**


1. Hedbom, E., Antonsson, P., Hjerpe, A., Aeschlimann, D., Paulsson, M., Rosa-Pimentel, E., Sommarin, Y., Wendel, M., Oldberg, A., and Heinegard, D. (1992) Cartilage matrix proteins. An acidic oligomeric protein (COMP) detected only in cartilage, *J. Biol. Chem. 267*, 6132-6136.
2. DiCesare, P., Hauser, N., Lehman, D., Pasumarti, S., and Paulsson, M. (1994) Cartilage oligomeric matrix protein (COMP) is an abundant component of tendon, *FEBS Lett. 354*, 237-240.
3. Smith, R. K., Zunino, L., Webbon, P. M., and Heinegard, D. (1997) The distribution of cartilage oligomeric matrix protein (COMP) in tendon and its variation with tendon site, age and load, *Matrix Biol. 16*, 255-271.

4. Muller, G., Michel, A., and Altenburg, E. (1998) COMP (cartilage oligomeric matrix protein) is synthesized in ligament, tendon, meniscus, and articular cartilage, *Connect. Tissue Res. 39*, 233-244.

5. Efimov, V. P., Engel, J., and Malashkevich, V. N. (1996) Crystallization and preliminary crystallographic study of the pentamerizing domain from cartilage oligomeric matrix protein: a five-stranded alpha-helical bundle, *Proteins 24*, 259-262.

6. Malashkevich, V. N., Kammerer, R. A., Efimov, V. P., Schulthess, T., and Engel, J. (1996) The crystal structure of a five-stranded coiled coil in COMP: a prototype ion channel?, *Science 274*, 761-765.

7. Ozbek, S., Engel, J., and Stetefeld, J. (2002) Storage function of cartilage oligomeric matrix protein: the crystal structure of the coiled-coil domain in complex with vitamin D(3), *EMBO J. 21*, 5960-5968.

8. Guo, Y., Bozic, D., Malashkevich, V. N., Kammerer, R. A., Schulthess, T., and Engel, J. (1998) All-trans retinol, vitamin D and other hydrophobic compounds bind in the axial pore of the five-stranded coiled-coil domain of cartilage oligomeric matrix protein, *EMBO J. 17*, 5265-5272.

9. McFarlane, A. A., Orriss, G. L., and Stetefeld, J. (2009) The use of coiled-coil proteins in drug delivery systems, *Eur J Pharmacol 625*, 101-107.

10. Legrand, P., and Rioux, V. (2010) The complex and important cellular and metabolic functions of saturated fatty acids, *Lipids 45*, 941-946.

11. Spector, A. A. (1999) Essentiality of fatty acids, *Lipids 34 Suppl*, S1-3.

12. Rioux, V., and Legrand, P. (2007) Saturated fatty acids: simple molecular structures with complex cellular functions, *Curr. Opin. Clin. Nutr. Metab. Care 10*, 752-758.

13. Resh, M. D. (2006) Trafficking and signaling by fatty-acylated and prenylated proteins, *Nat. Chem. Biol. 2*, 584-590.

14. Hayashi, N., and Titani, K. (2010) N-myristoylated proteins, key components in intracellular signal transduction systems enabling rapid and flexible cell responses, *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci. 86*, 494-508.

15. Salaun, C., Greaves, J., and Chamberlain, L. H. (2010) The intracellular dynamic of protein palmitoylation, *J. Cell Biol. 191*, 1229-1238.

16. Siri-Tarino, P. W., Sun, Q., Hu, F. B., and Krauss, R. M. Meta-analysis of prospective cohort studies evaluating the association of saturated fat with cardiovascular disease, *Am. J. Clin. Nutr. 91*, 535-546.

17. Hunter, J. E., Zhang, J., and Kris-Etherton, P. M. (2010) Cardiovascular disease risk of dietary stearic acid compared with trans, other saturated, and unsaturated fatty acids: a systematic review, *Am. J. Clin. Nutr. 91*, 46-63.

18. Vaysse-Boue, C., Dabadie, H., Peuchant, E., Le Ruyet, P., Mendy, F., Gin, H., and Combe, N. (2007) Moderate dietary intake of myristic and alpha-linolenic acids increases lecithin-cholesterol acyltransferase activity in humans, *Lipids 42*, 717-722.

19. Chmurzynska, A. (2006) The multigene family of fatty acid-binding proteins (FABPs): function, structure and polymorphism, *J. Appl. Gen. 47*, 39-48.

20. Furuhashi, M., and Hotamisligil, G. S. (2008) Fatty acid-binding proteins: role in metabolic diseases and potential as drug targets, *Nat. Rev. Drug Discov. 7*, 489-503.

21. Ozbek, S., Engel, J., and Stetefeld, J. (2002) Storage function of cartilage oligomeric matrix protein: the crystal structure of the coiled-coil domain in complex with vitamin D(3), *EMBO J. 21*, 5960-5968.

22. Guo, Y., Bozic, D., Malashkevich, V. N., Kammerer, R. A., Schulthess, T., and Engel, J. (1998) All-trans retinol, vitamin D and other hydrophobic compounds bind in the axial pore of the five-stranded coiled-coil domain of cartilage oligomeric matrix protein, *EMBO J. 17*, 5265-5272.

23. Efimov, V. P., Lustig, A., and Engel, J. (1994) The thrombospondin-like chains of cartilage oligomeric matrix protein are assembled by a five-stranded alpha-helical bundle between residues 20 and 83, *FEBS Lett. 341*, 54-58.

24. Terskikh, A. V., Potekhin, S. A., Melnik, T. N., and Kajava, A. V. (1997) Mutation Gln54Leu of the conserved polar residue inthe interfacial coiled coil position (d) results in significant stabilization of the original structure of the COMP pentamerization domain, *Lett. Pept. Sci.*, 297-304.

25. Leslie, A. G. W. (1992) Recent changes to the MOSFLM package for processing film and image plate data, *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*.

26. CCP4. (1994) The CCP4 suite: programs for protein crystallography, *Acta Crystallogr. D Biol. Crystallogr. 50*, 760-763.

27. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D Biol. Crystallogr. 54*, 905-921.

28. Turk, D. (1992) Weiterentwicklung eines programmes für molekülgraphik und electronendichte- manipulation und seine anwendungen auf verschiedene protein-strukturaufklärungen, Technische Universität München, Germany.

## Chapter 4: An interdomain disulfide bridge links the NtA and first FS domain in agrin

This chapter is based on the following manuscript published in the journal "*Protein Science*". Full citation of the paper is as follows:

McFarlane, A. A., and Stetefeld, J. (2009) An interdomain disulfide bridge links the NtA and first FS domain in agrin, *Protein Sci. 18*, 2421-2428.



**Abstract**

Agrin is a multidomain heparan sulfate proteoglycan involved in postsynaptic differentiation at the neuromuscular junction. Binding of agrin to synaptic basal lamina is mediated by the N-terminal agrin (NtA) domain. The NtA domain of agrin is followed by a tandem of nine follistatin-like (FS) domains forming a rod-like spacer to the laminin G-like domains of the molecule. Here we report that the most C-terminal cysteine residue of NtA (Cys 123) forms an interdomain disulfide bond with the FOLN subdomain of the FS module. Remarkably, this single cysteine is flanked by Leu117 and Val124, which are two essential β-branched amino acids forming the heterocomplex of NtA with the γ1 chain of laminin. Moreover, we show that this covalent linkage compensates for the seven amino acid residue splice insert at the very C-terminal helix H3 and causes a rigid interface between NtA and FS independent of the alternative mRNA splice event. These results suggest that the interdomain

disulfide bond between the NtA and the first FS domain might be important for the proper folding of agrin.

**Keywords**

Agrin, N-terminal agrin, follistatin, interdomain disulfide, FOLN subdomain, Kazal subdomain

## Introduction

The extracellular matrix (ECM) is a network of extracellular macromolecules that provides a structural support between tissues in eukaryotic organisms. The ECM plays an important role in tissue development, cell differentiation, and signal transduction. It is made up of a complex variety of polysaccharide and protein molecules, which are typically produced by the cells in the matrix (*1*). ECM proteins are mostly large, modular proteins consisting of sets of frequently repeated domains. This allows for the acquisition of structural information from individual modules (*2*).

The neuromuscular junction (NMJ) mediates the efficient communication between neurons and muscle fibers. It is composed of pre- and post-synaptic cells that are separated by a synaptic cleft containing the synaptic basal lamina. The ECM protein agrin is released by motor neurons and induces postsynaptic specializations at the NMJ, including clustering of acetylcholine receptors (*3*). Agrin is a multi-domain heparan sulfate proteoglycan with an apparent molecular weight of 400-600 kDa on SDS-PAGE. The N-terminus of agrin is composed of the N-terminal agrin (NtA) domain, which is required for binding of laminin and localization of agrin at basal lamina. Remarkably, NtA is a target of alternative mRNA splicing causing an insert of seven amino acid residues (*4, 5*). The NtA domain is followed by nine

repeats of follistatin-like (FS) domains (*6, 7*). Mapping of the laminin-binding site of NtA revealed that two amino acid residues are critical for binding (*8*). Leu117 and Val124 form a heterotetrameric interface between the triple helical coiled coil of laminin and the C-terminal helix H3 of NtA. The C-terminus of agrin is composed of three laminin G-like (LG) domains, which function in the clustering of acetylcholine receptors (*9*).

FS domains are found in proteins such as agrin, BM-40, follistatin, and ovomucoid-like chicken protein and are predominantly arranged as repetitive modules in tandem arrays forming rod-like structures (*1, 2*). The structure of the FS domain of mouse BM-40 revealed an N-terminal β-hairpin with four cysteines linked 1-3 and 2-4 (FOLN subdomain), directly followed by a larger globular α/β fold with six cysteines linked 5-9, 6-8, and 7-10 (Kazal subdomain) (*10*). The separation of cysteines and their linkage pattern within the FOLN subdomain are reminiscent of epidermal growth factor (*2*). The KAZAL domain is known as a member of the pancreatic secretory trypsin inhibitor family. A double-headed inhibitor of trypsin (on one site) and of chymotrypsin, subtilisin, and elastase (on the other site) is present in large amounts in submandibular glands and consists of two tandem KAZAL type domains (*11*).

In this work, we show in an integrated approach combining long-wavelength X-ray crystallography and mass spectrometry that there is an interdomain disulfide bridge between Cys123 of NtA and the following first FS domain in agrin. Intradomain disulfide bridges are commonly seen in ECM proteins and serve to stabilize the structural conformation of these molecules. Interdomain disulfide bridges are rare, and to date, only a few examples are known (*12, 13*). We propose that this interdomain disulfide linkage in agrin plays an important role in stabilizing the overall structure as it is independent of an alternative mRNA splice event.

## Results

### *Long-wavelength structure of NtA*

The crystal structure of chicken NtA was determined to 2.3Å resolution by molecular replacement using the previously solved chicken NtA core domain as search template (PDB code: 1PXU) (*14*). The structure is composed of five β-strands forming a canonical shaped β-barrel domain flanked by α-helices at both termini [H1 and H3; Fig. 4.1(A)]. The most extraordinary feature of the crystal structure is the high flexibility of Cys123 located in the C-terminal α-helix H3 implying a potential disulfide linkage [Fig. 4.1(B)]. The use of soft X-rays with a wavelength of λ=1.9995Å (6.213 kEv) and the possibility to collect highly redundant diffraction data were crucial to study the conformational flexibility of Cys123 of the C-terminal Helix H3 in NtA. In contrast to data collection protocols at λ=1.54Å, we could show for the first time an inherent structural flexibility of this free accessible thiol group. This finding is particularly striking based on the nonreducing milieu within the ECM and the close spatial neighborhood of the proceeding first FS domain.

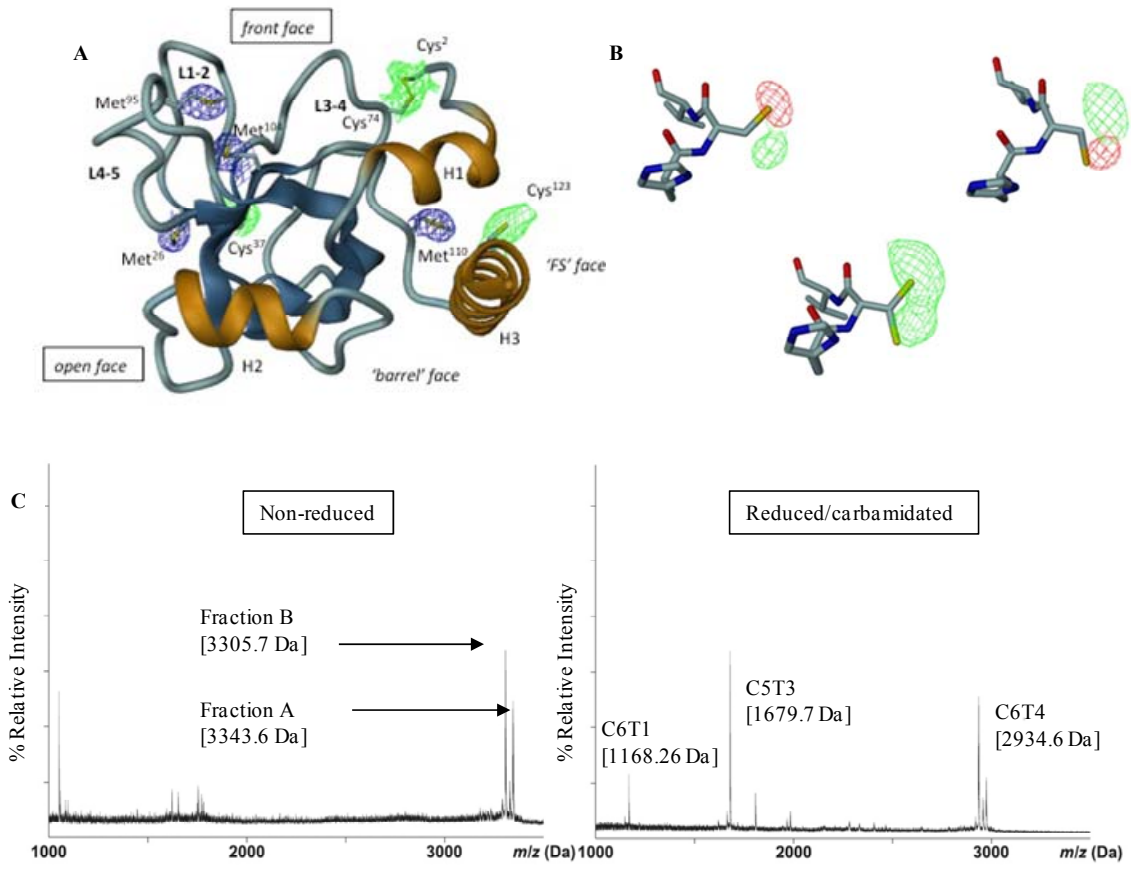


**Fig. 4.1. Long-wavelength X-Ray structure of NtA and MS mapping of the interdomain disulfide linkage.** (A) Ribbon diagram of the NtA structure. The β-strands are in blue and α-Helices (H1-H3) are in gold; loops L1-2, L3-4, and L4-5 connecting individual β-strands are oriented to the same surface of the protein. The 4σ Fo-Fc electron density maps for individual sulfur atoms in methionine thioether groups (in blue) and cysteine thiol moieties (in blue) are highlighted. (B) The alternate conformations adopted by Cys123. Upper row: calculated Fo-Fc electron density map (4σ contour level) around Cys123 is indicated each by means of the positive (green) and negative (red) electron density peak. The diffraction data were measured at λ=1.54Å. Lower row: a 4σ contour level Fo-Fc difference map (diffraction data collected at

1.9995Å) indicating the free degree of rotation at κ angle at the thiol group of Cys123. (C) Left panel: MALDI-TOF spectrum of nonreduced fractions A (3339.6 Da) and B (3305.7 Da) obtained by sequential cleavage of NtA-FS by CNBr and trypsin. Right panel: the corresponding MALDI-TOF spectra of reduced and carbamidated trypsin digests of fraction A (C5T3 and C6T1) and B (C6T4).

**Table 4.1. Data collection and refinement statistics**

| **Data collection at 1.9995 Å** | |
|---|---|
| Resolution (Å) (last shell) | 28–2.3 (2.44–2.3) |
| Observed reflections | 1.130368 |
| Unique reflections | 8736 |
| Completeness | 97.3 (97.3) |
| Redundancy | 31.1 |
| $R_{sym}$[a] | 8.6 (33.9) |
| **Unit cell dimensions (Å, °)** | |
| a/ b and c | 80.23/50.71 and 54.19 |
| β | 117.094 |
| **Refinement statistics** | |
| *R*-factor[b] (%) | 22.1 (42.7) |
| $R_{free}$ (%) | 27.6 (45.8) |
| Bond length[c] (Å) | 0.006 |
| Bond angles[c] (°) | 1.3 |
| Ramachandran plot[d] | 87.5/12.5/0/0 |

[a]$R_{sym} = 100* \Sigma|I - (I)| / \Sigma I$. Values for the final resolution shells are in parenthesis.
[b]$R_{factor} = 100* \Sigma ||F_{obs}| - |F_{calc}|| / \Sigma |F_{obs}|$.
[c]Root mean square error.
[d]Percentage of residues in most favoured/additional allowed/generously allowed, and disallowed regions.

***NtA is disulfide bridged with the FOLN subdomain of FS via C-terminal Cys123***

To elucidate a possible disulfide linkage status of Cys123, proteolytic fragments were generated by treatment with cyanogen bromide (CNBr), followed by reverse-phased high performance liquid chromatography (RP-HPLC) separation. The fragments obtained after

separation on RP-HPLC were subjected to MALDI-TOF mass spectroscopy (MS; for nomenclature and molecular masses of the fragments see Materials and Methods and Tables 4.2 and 4.3). Six fragments (C1-C6) are predicted from CNBr digestion of NtA-FS from which four contain cysteine residues (Table 4.2). The disulfide bridge between Cys2 and Cys74 in the crystal structure [Fig. 4.1(A)] suggests that CNBr fragments C1 and C2 are disulfide linked. Although C1 contains a single cysteine residue (Cys2), fragment C2 contains two cysteine residues (Cys37 and Cys74). The crystal structure of recombinant NtA expressed in eukaryotic cells revealed that Cys37 is free, oriented inside the β-barrel core domain (*4*). MS confirmed the presence of corresponding fragments and the observed data agree with the predicted masses (data not shown).

Further analysis of the remaining fragments (C3-C6) revealed a rather broad peak in the mass spectrum. Reduction with dithiothreitol (DTT) showed that the observed mass corresponds to the calculated mass of C6, however, no precise measurement for fragment C5 could be obtained, which is attributable to incomplete cleavage by CNBr that resulted in a fragment of C3-C4-C5. Heterogeneous N-glycosylation on C3 might explain the lack of a well-defined mass spectrum of the C3-C4-C5 fragment (*6, 7*). However, cleavage of NtA-FS by CNBr followed by reduction demonstrates that C5 is linked via disulfide bridges to C6. The connection can only be mediated by cysteines present on C5 (Cys123 and/or Cys150) since fragments C3 and C4 do not possess any cysteine residues. Therefore, a disulfide linkage between NtA (Cys123) and the first cysteine residue of FS (Cys150) can be excluded.

The next strategy to resolve disulfide bridging was to digest the CNBr-fragments C5-C6 with trypsin, followed by MALDI-TOF analysis before and after reduction with DTT and alkylation with iodacetoamide. In the nonreduced state, two peaks were identified as peptides A

and B with observed molecular masses of 3339.6 Da (fraction A) and 3305.7 Da (fraction B), respectively [Fig. 4.1(C), left panel].

The observed mass of 3339.6 Da (fraction A) corresponds to the combined theoretical mass of C5T3 (1622.72 Da) and C5T6 (667.27 Da) disulfide bonded to C6T1 (1053.48 Da; Table 4.3). A difference of 4 Da in molecular mass between the nonreduced (3339.6 Da) and sum of reduced (3343.47 Da) fragments is indicative of the formation of two disulfide bridges. The peptide was resolved by reduction with DTT and carbamidation with iodoacetamide into two peptides with masses of 1679.7 Da and 1168.3 Da, respectively [Fig. 4.1(C), right panel]. The mass of 1679.4 Da corresponds to C5T3 with one carbamidated cysteine (Cys123), whereas 1168.3 Da correspond to C6T1 with two carbamidated cysteine residues (Cys155 and Cys161). However, the peptide fragment corresponding to C5T6 (Cys150) could not be identified on reduction and alkylation because of its small size.

**Table 4.2. Calculated masses of the predicted CNBr fragments of NtA-FS**

| Fragment | Sequence | Mass (Da) |
|---|---|---|
| C1 | NC$^{2}$PERELQRREEEANVVLTGTVEEIM | 3044.48 |
| C2 | NVDPVHHTYSC$^{37}$KVRVWRYLKGKDIVTHEILLDGGNKVVIGGFGDPLIC$^{74}$DNQVS TGDTRIFFVNPAPQYM | 7741.93 |
| C3 | WPAHRNELM | 1153.56 |
| C4 | LNSSLM | 664.33 |
| C5 | R*ITLR*N LEEVEHC$^{123}$VEEHR*K*LLADKPNSYFTQTPPTPR*DAC$^{150}$RGM | 5050.52 |
| C6 | LC$^{155}$GFGAVC$^{161}$ER*SPTDPSQASC$^{173}$VC$^{175}$K*K*TAC$^{180}$PVVVAPVC$^{188}$GSDYSTYSN EC$^{199}$ELEK*AQC$^{206}$ NQQR*R*IK*VISK*GPC$^{220}$GSK*DPC$^{226}$AEVTR*SHHHHHH | 9194.25 |

The peptide nomenclature is described in Experimental Procedures. The molecular weights of the predicted CNBr fragments are given as the monoisotopic masses of the singly protonated peptides.

Cysteines are highlighted as superscripts, while $Ala^{132}$ and $Cys^{150}$ which mark the last and the first residue of the NtA and the FS domain, respectively, are underlined. Stars indicate putative trypsin cleavage sites in C5 and C6.

In summary, the data presented show that intrafragment disulfide linkages within C5 (between Cys123 and Cys150) and within C6T1 (between Cys155 and Cys161) are not formed. Furthermore, they show that C5T3 and C5T6 are associated to C6T1 via two disulfide bonds and provide evidence for disulfide bridges between Cys123 and Cys150 with Cys155 or Cys161. Therefore, it can be concluded that NtA is disulfide linked via Cys123, located at the C-terminal α-helix H3, with the FOLN subdomain of FS. However, the order of linkage between the fragments remains elusive (see Discussion).

The observed mass of 3305.7 Da (fraction B) corresponds to the theoretical mass of C6T4 (2764.22 Da) disulfide linked to C6T9 (548.25) with a mass difference of 7 Da between nonreduced and reduced fragments [Fig. 4.1(C), left panel and Table 4.2). The peptide was resolved on DTT reduction and alkylation with iodoacetamide into a peptide (C6T4) with a mass of 2934.6 Da that corresponds to three carbamidated cysteine residues [Cys180, Cys188 and Cys199; Fig. 4.1(C), right]. The smaller peptide corresponding to C6T9 (Cys220) could not be identified following reduction and alkylation. These data imply disulfide bridge formation between Cys180, Cys188 or Cys199 with Cys220 including an intrafragment disulfide bond within C6T4. All attempts to isolate fragments containing Cys173, Cys175 (C6T2) and Cys206 (C6T5) failed, most probably because they do not bind to the reverse-phase support during HPLC purification.

**Table 4.3. Trypsin cleavage products of the CNBr fragments C5 and C6.**

| Peptide Fraction | Peptide mass (non-reduced)[a] | Peptide mass (reduced)[a] | Δ Mass | Trypsin cleavage fragment | Sequence | Peptide mass (reduced)[a] | Peptide mass (reduced/ alkylated)[a] | Δ Mass |
|---|---|---|---|---|---|---|---|---|
| A | 3339.6 | 3343.5 | 4 | C5T3 | NLEEVEH $C^{123}$VEEHR | 1622.72 | 1679.72 | 57.0 |
| | | | | C5T6 | DA$C^{150}$RG M | 667.2[b] | ND | |
| | | | | C6T1 | L$C^{155}$GFG A$VC^{161}$ER | 1053.5 | 1168.26 | 115.1 |
| B | 3305.7 | 3312.5 | 7 | C6T4 | TA$C^{180}$PV VVAPV$C^{18}$ $^{8}$GSDYT YSNE$C^{199}$E LK | 2763.2 | 2934.55 | 171.34 |
| | | | | C6T9 | GP$C^{220}$GS K | 547.2 | ND | |
| | | | | C6T2[c] | SPTDPSQ AS$C^{173}$V $C^{175}$K | 1321.6 | | |
| | | | | C6T5[c] | AQ$C^{206}$NQ QR | 846.4 | | |

All Peptide masses are given in Da. Individual fragments are labeled sequentially. Fraction A consists of C5T3, C5T6, and C6T1, whereas fraction B comprises C6T4 and C6T9, respectively. The peptides C6T2 and C6T5 containing cysteine residues $Cys^{173}$, $Cys^{175}$, and $Cys^{206}$ could not be isolated. The calculated masses are given as monoisotopic values of the singly protonated peptides.

[a] Molecular mass was calculated based on combined theoretical mass of individual trypsin fragments.

[b] The oxidation and subsequent hydroxymethylation of the C-terminal Met adds +16Da to the peptide mass of 651.25 Da.

[c] CNBr-Trypsin fragments that could not be isolated.

**Discussion**

Many ECM modules contain disulfide bridges, most probably to increase stability of relatively small domains and to protect against proteolysis. Cysteines are the most important identifiers of a particular module by sequence analysis (*1*). Almost all conserved cysteine residues in various ECM proteins are involved in formation of disulfide bridges, mostly intramodular ones. Exceptions tp this rule are rare and have so far only been observed for Ig-like and FN-III domains. The 3D structures of several cell surface proteins containing Ig-like domains clearly demonstrate the switch of disulfide bridges between strands (*15*). Domains of the FN-III type usually do not contain disulfide bridges; however, the structure of neuroglian reveals a disulfide linkage (*16*). Cole et al. (*12*) demonstrated an intermodular disulfide linkage in the soluble Interleukin-6R domain, revealing a bridge between Cys6 of the N-terminal Ig-domain to Cys174 of the following first FN-III domain.

Crystal structures of FS domains within BM-40 and follistatin reveal a nonglobular structure that is stabilized by a small hydrophobic core in its larger, C-terminal portion (Kazal subdomain) and by a total of five disulfide bonds (*10, 17*). Remarkably, in both X-ray structures the N-terminal FOLN subdomain shows the largest degree of mobility. The subdivision of FS is reflected in a disulfide linkage pattern of two nonoverlapping sets (1-3 and 2-4) and (5-9, 6-8, and 7-10; Fig. 4.2).

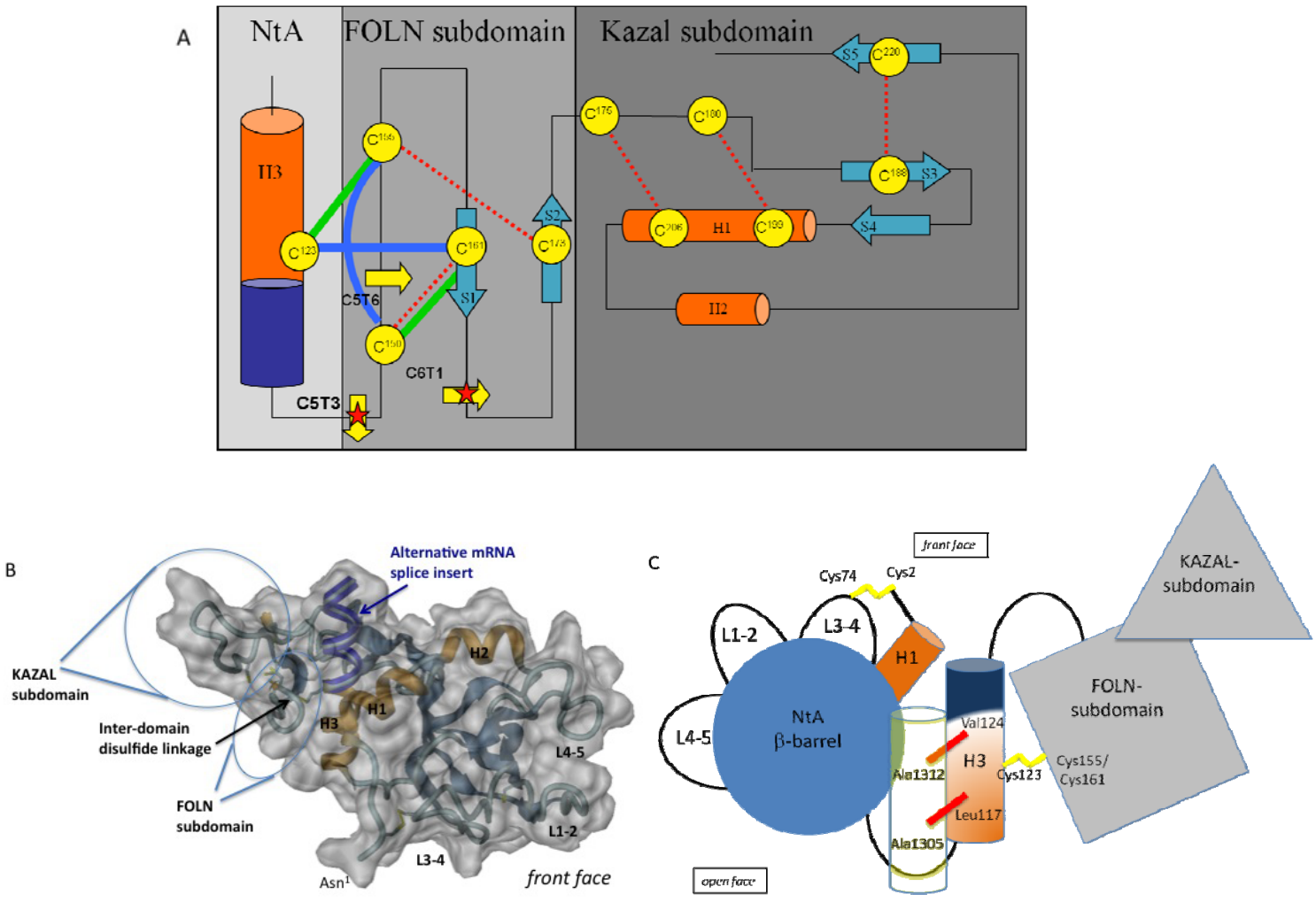


**Fig. 4.2. The proposed NtA-FS tandem.** (A) Topology model of interdomain disulfide bond formation between NtA and FS domain. Secondary structural elements (α-helices in orange and β-strands in blue) and conserved disulfide bond pattern within the FS domain (thin red dotted lines) are highlighted. Cysteine residues are shown in circles labeled sequentially. The FS domain is assembled from two rather weakly interacting subunits, a highly twisted N-terminal β-hairpin (FOLN subdomain) and a pair of antiparallel α-helices connected to a small three-stranded antiparallel β-sheet (Kazal subdomain). Arrows mark trypsin cleavage sites, and stars represent isolated, reduced, and alkylated fragments (see also Table 4.3). Possible new

interdomain disulfide bonds between Cys123 and Cys155 or Cys161, as well as the new intradomain disulfide bond between Cys150 and Cys155 are highlighted as thick green (model 1) and blue (model 2) lines. In any case, the disulfide bridge between Cys155 and Cys173 has to be reshuffled. A theoretical third model, comprising a free Cys150 within C5 has to be excluded, based on the fact that a separate CNBr-trypsin fraction before reduction via DTT containing Cys150 only, as well as a consequent disulfide bridge between 2 and 4 could not be detected. (B) Surface representation of the modeled NtA-FS tandem based on a FTDOCK approach. Color scheme and labels as in Fig. 4.1. Individual cysteine residues are shown in render mode. The alternative mRNA splice insert in H3 is highlighted in dark blue. (C) Topology diagram of the NtA-FS tandem. Individual domains of the tandem are highlighted in different color schemes according to previous diagrams. Loop regions extending the β-barrel (L1-2, L3-4, and L4-5) and the 9-residue linker between the barrel and α-helix H3, as well as the 15-residue long connection between the C-terminus of NtA and the N-terminus of FS, are shown as black lines. Disulfide bridges are highlighted in yellow. The alternative mRNA splice insert at H3 is shown in blue. For simplicity reasons, only the γ1-chain of laminin is shown in yellow transparent with key interface residues marked with red bars.

By sequential cleavage of NtA-FS using CNBr and trypsin coupled to RP-HPLC purification we have isolated fragments C5T3, C5T6, C6T1, C6T4, and C6T9. MALDI-TOF analysis of the reduced and carbamidated peptide fragments revealed that Cys123 from NtA is linked via a disulfide bridge with the FOLN subdomain of the following FS. This connection can be obtained either via its linkage to Cys155 or to Cys161, which leaves two possibilities for the formation of the disulfide bridge between NtA and FS [Fig. 4.2(A)]. The first possibility

requires only one new disulfide bond between Cys123 and Cys155 while the conserved 1-3 bridge (Cys150-Cys161) remains intact. This model is supported by a sequence alignment of all nine FS domains of chicken agrin showing that five disulfide bonds are highly conserved, with the exception of the third FS domain, where two disulfide bonds, 2-4 and 5-9, are replaced (Fig. 4.3). The first disulfide bridge (1-3) seems to be strictly required, however, a second connection within FOLN (2-4) and link to the Kazal subdomain (5-9) are not. In contrast, model 2 would require the formation of two new disulfide bridges between Cys123-Cys161 and Cys150-Cys155, respectively. Therefore, this possibility is less likely because (i) it would require reshuffling of two disulfide bridges within the FS domain (1-3 and 2-4) and (ii) the short distance of only 4 amino acids and the stereochemical constraints at the N-terminal tip of the β-hairpin make a disulfide bridge between Cys150 and Cys155 very unlikely.

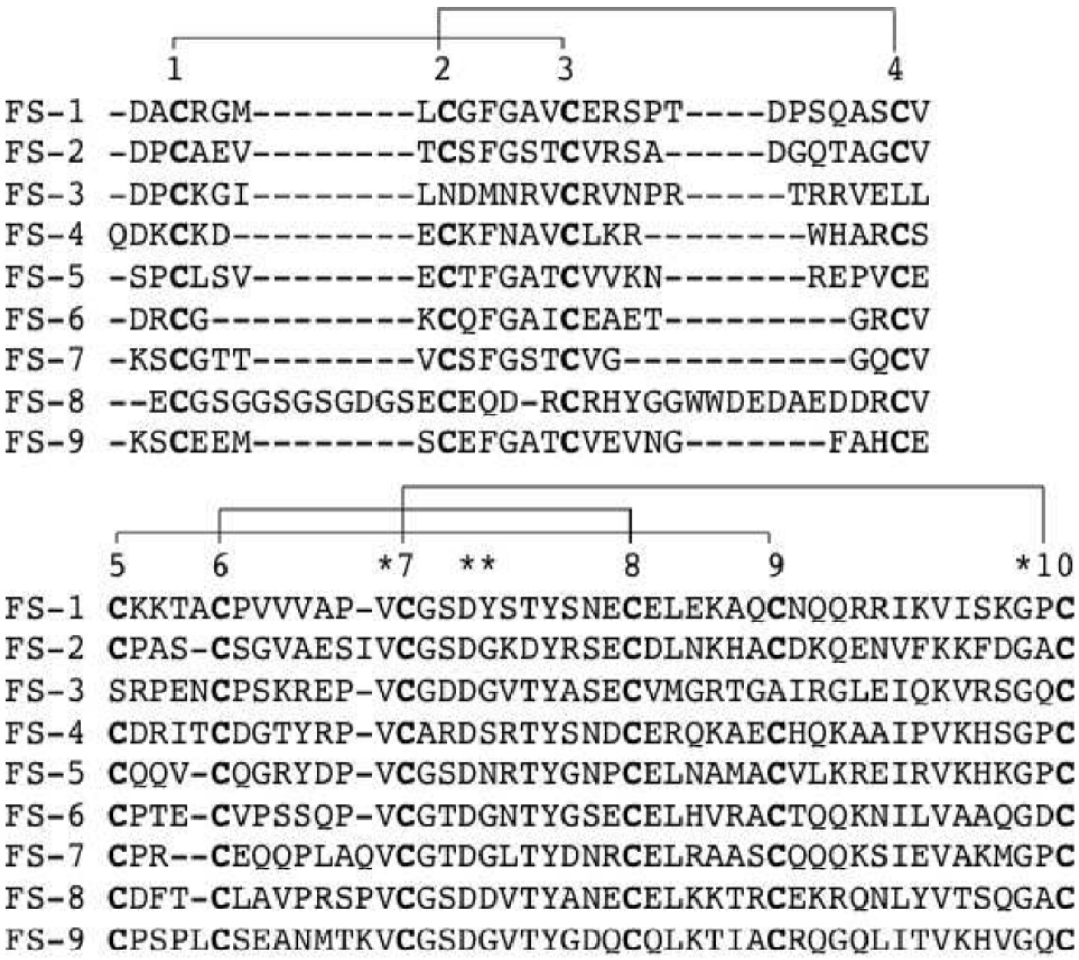


**Fig. 4.3. Sequence alignment of all nine FS domains in chicken agrin.**

The first FS domain starts with residue number 148. The upper scheme aligns the first four cysteine residues of the FOLN subdomain (disulfide linkage pattern between cysteine residues 1-3 and 2-4), whereas the Kazal subdomain with six cysteines is shown below (disulfide linkage pattern between cysteine residues 5-9, 6-8, and 7-10). Conserved residues within the Kazal subdomain are highlighted. Remarkably, the disulfide linkage between 2$^{nd}$ and 3$^{rd}$ as well as 5$^{th}$ and 9$^{th}$ cysteine residue is interrupted in the FS-3 domain.

Taken together, only models having a free cysteine at position four (Cys173) are possible, and a disulfide reshuffling within the FS domain is still required. Based on the limited set of data concerning disulfide bridging within the Kazal subdomain, a possible role of a free Cys173 remains elusive. However, superposition of both FS domains from BM-40 and follistatin revealed a hingelike motion between FOLN and Kazal subdomains (*17*). If both Kazal-like moieties are superimposed, the respective FOLN subdomains are bent and axially rotated by 45° relative to one another. Together with the observed variability in the positioning of cysteines 2, 4, 5, and 9, the interface between both subdomains is likely to be flexible. This flexibility might be important for the different spatial orientations in the long-range tandem arrays of nine FS domains in agrin [Fig. 4.2(C)].

A particularly striking feature is the fact that the NtA domain has significant structural similarity with TIMP-1, an endogenous inhibitor of matrix metalloproteinases (MMPs). These enzymes are secreted as inactive zymogens and have to be activated by selective promatrixin proteinases (*18*). A possible interference of the KAZAL inhibitor domain with this activation step might explain the stabilization effect of agrin within NMJ and the diminished MMP-activities if agrin is released inside the synaptic cleft (*19*). Agrin remains associated for weeks with the synaptic basal lamina *in vivo*, and mice that overexpress an NtA-containing agrin construct at the NMJ show increased stability of postsynaptic structures. These authors (*19*) showed that regulated inhibition of MMP3-activity results in changes in structure and function of the NMJ. Consistent with its modular organization, different regions of agrin have been associated with distinct functional properties (*20*). An alternate function for agrin's FS domains is suggested by their homology with Kazal-type proteinase inhibitors. Studies *in vitro* have shown that agrin inhibits several proteinases, including a membrane-bound plasmin-like activity

(*21*). In addition to their structural role, the FS domains may also act as protease inhibitors. Interestingly, a proposed intermodular disulfide linkage between NtA and the FOLN-subdomain of the following FS module underlines this hypothesis.

The interdomain disulfide bridging between Cys123 and the FOLN subdomain of FS confirms the previously proposed model of NtA-laminin interaction, which clearly indicates a predominant role of residues at the "barrel face" of helix H3 (interaction interface) and a "FS face" at the opposite side. Cys123 on the FS face marks the linker for the interaction between NtA and the following first FS domain (see above). Disulfide bridges are well known to stabilize noncovalent conformations and are thus important and, usually, conserved structural features. The transition from an intradomain to an interdomain cysteine bridge may offer the advantage of a defined spatial orientation of the NtA and the first FS domain in agrin.

The C-terminal α-helix (H3) of chicken NtA contains a seven-residue splice insert comprising residues Glu126-Ala132. Motor neurons in developing spinal cord contain agrin transcripts that include the splice insert. However, the majority of agrin mRNA in non-neuronal tissues is characterized by the absence of the seven-residue insert (*6, 22*). We suggest that the additional disulfide bridge between NtA and FS performs a further stabilization effect, which is required to accommodate the loss of seven amino acid residues in the C-terminal α-helix H3 [Fig. 4.2(B)]. These findings are supported by the fact that alternative mRNA splice events are species specific. Only chicken NtA shows the insert of seven amino acid residues, whereas mouse, human, and rat NtA domains do not. However, all species reveal the same pattern of synapse formation and stabilization.

Despite the fact that the linker-region between NtA and FS (residues Asp133-Ala149) remains the same, a loss of the seven-residue insert within α-helix H3 would cause a remarkable

effect on the spatial orientation of both domains. The intermodular disulfide bond would be able to stabilize a required relative orientation to each other and preserve functional activity in the protein.

To solve this question, the 3D structure of the NtA-FS tandem would be needed. Work in this direction is in progress.

## Materials and Methods

### *Protein purification and CNBr cleavage*

The NtA-FS was dialyzed overnight against 1% acetic acid, followed by speed vac drying. The resulting pellet was resuspended in 70% formic acid and digested overnight with CNBr. The digest was again speed vac dried, the pellet was taken up in 0.1% trifluoroacetic acid (TFA), and the resulting peptide fragments were purified by RP-HPLC on a Vydac TP52 C18 (2.1 x 250mm) column.

### *Trypsin cleavage*

Fractions obtained after RP-HPLC were subjected to speed vac drying to evaporate acetonitrile. The pH was adjusted to 8.0, and digestion with trypsin (Promega) was performed at an enzyme/substrate ratio of 1:25 at 37°C for 2 hours. The reaction was terminated by the addition of TFA to a final concentration of 0.1%. The samples were desalted by absorption on C18 Zip Tips (P10 size, Millipore) washed with 0.1% TFA and desorbed with 2 ml of 50% acetonitrile/0.1% TFA. The samples were analyzed by MALDI-TOF as described below.

***Reduction and alkylation***

Reduction was carried out by 10 mM DTT in 100 mM Tris-HCl pH 8.0 for 1 h. Following reduction, cysteines were alkylated using 50 mM iodoacetamide and incubated for a further 15 min at room temperature. The reaction was stopped by the addition of 1/10 volume of 10% TFA. The samples were desalted as described above and analyzed on MALDI-TOF.

***MALDI-TOF mass spectrometry***

300nL of peptide solution were deposited onto anchor spots of a 36-sample support target (Bruker Daltonik), and the droplet was allowed to dry at room temperature. Mass spectra were recorded on a Bruker Reflex III instrument (Bruker Daltonik). Peptides were analyzed in linear mode. Assignment of the CNBr or tryptic peptides of NtA-FS was done by comparing the measured masses to those predicted by the protein sequence of NtA-FS.

***Peptide Nomenclature***

The first letter and number of the fragment name indicates the cyanogen cleavage products (C1 to C6). Peptides of CNBr fragments C5 (C5T3 and C5T6) and C6 (C6T4 and C6T9) generated by trypsin cleavage are labeled with an additional T. The peptides are numbered sequentially according to their position based on the amino acid sequence (Tables 4.1 and 4.2).

***Structure determination of NtA***

The NtA domain was produced in HEK 293 cells and purified from conditioned media as described previously (*8, 14*). To obtain crystals in a quality to perform long-wavelength X-ray

diffraction experiments cocoamidoprylbetaine (CAPB) was added to the original crystallization condition (*4, 5, 23*).

The long-wavelength data set was collected at a wavelength λ = 1.9995Å at the Canadian Light Source (CLS). Diffraction images were processed using program suite MOSFLM, and the structure factors were scaled and reduced using SCALA from the CCP4 package (*24*). The statistics of the merged data and the refined structure are given in Table 4.1. Molecular replacement was performed using the AMORE program of the CCP4 package. The core of the five stranded β-barrel OB-fold of NtA was used as search model comprising amino acid residues 18-110. Loop regions connecting β-strands S1-S2 (L1-2), S3-S4 (L3-4), S4-5 (L4-5) and all three helical segments (H1-H3) were truncated. Refinement with CNS was alternated with manual electron density refitting of side-chains and terminal regions (*25*). The target parameters of Engh and Huber, overall anisotropic B-factor scaling, and bulk solvent corrections were used. Finally, water molecules were added chosen by distance criteria and hydrogen bonding geometry and were tested for position in spherical density, reasonable temperature factors, real space R-values, and improvement of the R-factors. The coordinates have been deposited at Brookhaven database with the PDB-code 3I7O.

***Molecular modeling***

Docking of NtA (PDB code:1PXU) with the FS domain (PDB code:1BMO) was performed using FTDOCK (*26*). The homology model of the FS domain was generated using SWISS-Model. The lowest energy docked structures were subject to 100 cycles of unrestrained Powell minimization using CNS (*25*). Harmonic restraints were imposed on the protein atoms (3 kcal/mol $Å^2$) with increased weight (20 kcal/mol $Å^2$).

**References**


1. Hohenester, E., and Engel, J. (2002) Domain structure and organisation in extracellular matrix proteins, *Matrix Biol. 21*, 115-128.

2. Bork, P., Downing, A. K., Kieffer, B., and Campbell, I. D. (1996) Structure and distribution of modules in extracellular proteins, *Q. Rev. Biophys. 29*, 119-167.

3. Ruegg, M. A., and Bixby, J. L. (1998) Agrin orchestrates synaptic differentiation at the vertebrate neuromuscular junction, *Trends Neurosci. 21*, 22-27.

4. Stetefeld, J., Jenny, M., Schulthess, T., Landwehr, R., Schumacher, B., Frank, S., Ruegg, M. A., Engel, J., and Kammerer, R. A. (2001) The laminin-binding domain of agrin is structurally related to N-TIMP-1, *Nat. Struct. Biol. 8*, 705-709.

5. Stetefeld, J., and Ruegg, M. A. (2005) Structural and functional diversity generated by alternative mRNA splicing, *Trends Biochem. Sci. 30*, 515-521.

6. Denzer, A. J., Gesemann, M., Schumacher, B., and Ruegg, M. A. (1995) An amino-terminal extension is required for the secretion of chick agrin and its binding to extracellular matrix, *J. Cell Biol. 131*, 1547-1560.

7. Denzer, A. J., Hauser, D. M., Gesemann, M., and Ruegg, M. A. (1997) Synaptic differentiation: the role of agrin in the formation and maintenance of the neuromuscular junction, *Cell Tissue Res. 290*, 357-365.

8. Mascarenhas, J. B., Ruegg, M. A., Winzen, U., Halfter, W., Engel, J., and Stetefeld, J. (2003) Mapping of the laminin-binding site of the N-terminal agrin domain (NtA), *EMBO J. 22*, 529-536.

9. Gesemann, M., Cavalli, V., Denzer, A. J., Brancaccio, A., Schumacher, B., and Ruegg, M. A. (1996) Alternative splicing of agrin alters its binding to heparin, dystroglycan, and the putative agrin receptor, *Neuron 16*, 755-767.

10. Hohenester, E., Maurer, P., and Timpl, R. (1997) Crystal structure of a pair of follistatin-like and EF-hand calcium-binding domains in BM-40, *EMBO J. 16*, 3778-3786.

11. van de Locht, A., Lamba, D., Bauer, M., Huber, R., Friedrich, T., Kroger, B., Hoffken, W., and Bode, W. (1995) Two heads are better than one: crystal structure of the insect derived double domain Kazal inhibitor rhodniin in complex with thrombin, *EMBO J. 14*, 5149-5157.

12. Cole, A. R., Hall, N. E., Treutlein, H. R., Eddes, J. S., Reid, G. E., Moritz, R. L., and Simpson, R. J. (1999) Disulfide bond structure and N-glycosylation sites of the extracellular domain of the human interleukin-6 receptor, *J. Biol. Chem. 274*, 7207-7215.

13. Muralidhara, B. K., and Hirose, M. (2000) Structural and functional consequences of removal of the interdomain disulfide bridge from the isolated C-lobe of ovotransferrin, *Protein Sci. 9*, 1567-1575.

14. Mascarenhas, J. B., Ruegg, M. A., Sasaki, T., Eble, J. A., Engel, J., and Stetefeld, J. (2005) Structure and laminin-binding specificity of the NtA domain expressed in eukaryotic cells, *Matrix Biol. 23*, 507-513.

15. Bork, P., Holm, L., and Sander, C. (1994) The immunoglobulin fold. Structural classification, sequence patterns and common core, *J. Mol. Biol. 242*, 309-320.

16. Huber, A. H., Wang, Y. M., Bieber, A. J., and Bjorkman, P. J. (1994) Crystal structure of tandem type III fibronectin domains from Drosophila neuroglian at 2.0 A, *Neuron 12*, 717-731.

17. Innis, C. A., and Hyvonen, M. (2003) Crystal structures of the heparan sulfate-binding domain of follistatin. Insights into ligand binding, *J. Biol. Chem. 278*, 39969-39977.

18. Nagase, H., and Woessner, J. F., Jr. (1999) Matrix metalloproteinases, *J. Biol. Chem. 274*, 21491-21494.

19. Werle, M. J., and VanSaun, M. (2003) Activity dependent removal of agrin from synaptic basal lamina by matrix metalloproteinase 3, *J. Neurocytol. 32*, 905-913.

20. Smith, M. A., and Hilgenberg, L. G. (2002) Agrin in the CNS: a protein in search of a function?, *Neuroreport 13*, 1485-1495.

21. Biroc, S. L., Payan, D. G., and Fisher, J. M. (1993) Isoforms of agrin are widely expressed in the developing rat and may function as protease inhibitors, *Brain Res. Dev. Brain Res. 75*, 119-129.

22. Tsen, G., Napier, A., Halfter, W., and Cole, G. J. (1995) Identification of a novel alternatively spliced agrin mRNA that is preferentially expressed in non-neuronal cells, *J. Biol. Chem. 270*, 15934-15937.

23. Ozbek, S., Engel, J., and Stetefeld, J. (2002) Storage function of cartilage oligomeric matrix protein: the crystal structure of the coiled-coil domain in complex with vitamin D(3), *EMBO J. 21*, 5960-5968.

24. CCP4. (1994) The CCP4 suite: programs for protein crystallography, *Acta Crystallogr. D Biol. Crystallogr. 50*, 760-763.

25. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D Biol. Crystallogr. 54*, 905-921.

26. Gabb, H. A., Jackson, R. M., and Sternberg, M. J. (1997) Modelling protein docking using shape complementarity, electrostatics and biochemical information, *J. Mol. Biol. 272*, 106-120.

# Chapter 5: Conclusion

## 5.1. Thesis summary

ECM proteins play a vital role in many cellular processes. They are often difficult to assay due to their large size and high amount of post-translational modifications. The attainment of structural information of individual domains can aid immensely in the study of ECM proteins. Understanding the important structure-function relationships of ECM proteins enriches our knowledge of the cellular activities these proteins take part in and can lead to potential therapies for diseases concerning the ECM.

This thesis was aimed at investigating two ECM proteins: COMPcc and NtA. COMPcc is a key structural protein found in cartilage (*1*), tendon (*2, 3*), and ligament (*4*). The x-ray crystal structure of recombinant COMPcc revealed that the protein is a pentamer of five α-helices that form a left-handed supercoil containing a 73Å-long axial pore (*5*). Previous studies demonstrated the remarkable ability of the COMPcc channel to bind hydrophobic compounds. The crystal structures of COMPcc-vitamin $D_3$ (*6*), COMPcc-all-*trans* retinol, and COMPcc-benzene complexes have been solved (*7*). COMPcc belongs to the family of coiled-coil proteins, which are extremely stable and function in nature as oligomerization domains. These characteristics, along with the unique storage properties of COMPcc make it an attractive candidate for the design of drug delivery systems (*8*). I explored the use of COMPcc and other coiled-coil proteins as potential drug delivery vehicles in Chapter 2 of this thesis. This then led to the studies which were done in Chapter 3.

In Chapter 3, I investigated the ability of COMPcc to store hydrophobic compounds. I chose to study five naturally-occurring fatty acids: myristic acid (C14:0), palmitic acid (C16:0),

stearic acid (C18:0), arachadic acid (C20:0), and oleic acid (C18:1). I solved the x-ray crystal structure of COMPcc in complex with these fatty acids. Additionally, I performed fluorescence spectroscopy measurements on the COMPcc-fatty acid complexes to determine their binding behaviours in solution. Remarkably, COMPcc was able to bind all five fatty acids with varying affinities. An overall trend was established in which binding favourability is a function of the length of the methylene tail and degree of unsaturation of the fatty acids. This study was significant as it demonstrates a putative therapeutic role for COMPcc in a drug delivery system. Also, the data suggest a possible binding mechanism for the ligand molecules to COMPcc.

Chapter 4 explores the structural properties of NtA. NtA is the N-terminal domain of agrin, which is a multidomain heparan sulfate proteoglycan involved in post-synaptic differentiations at the NMJ. NtA binds laminin and localizes agrin at basal lamina. In agrin, NtA is followed by nine repeats of follistatin-like domains, which form a rod-like spacer to the laminin G-like domains at the C-terminal half of the protein (*9-11*). In this study, I examined the structure-function relationship of NtA and an NtA-FS tandem using a combination of x-ray crystallography, mass spectrometry, and molecular modeling. Mass spectrometry analysis revealed a novel interdomain disulfide bridge between the most C-terminal cysteine residue of NtA (Cys 123) and the FOLN subdomain of the FS domain in the NtA-FS tandem protein. These findings were supported by the long-wavelength crystal structure of NtA, which showed the structural flexibility of the Cys123 residue that had been lacking in the previously-published structures. This work is significant in that it illustrates a rare interdomain disulfide bridge. Moreover, this disulfide bond is independent of an alternative mRNA splice event that occurs in the C-terminal helix H3 of NtA. These results suggest that the interdomain disulfide linkage between NtA and FS has a functional role in stabilizing the overall structure of agrin.

COMP and agrin are two important ECM proteins whose biochemistry is not well characterized. Each are composed of a number of repeating domains, and the individual functions of these domains as well as their contributions to the overall functionality of their respective proteins is also poorly understood. By studying the three-dimensional structure of distinct domains, the intricate structure-function relationships can begin to be elucidated.

COMP plays a vital role in cartilage, tendon, and ligament and defects in the COMP gene can lead to severe afflictions such as pseudoachondroplasia and multiple epiphysial dysplasia. The oligomerization domain of COMP, COMPcc, is required for the formation of the pentameric structure of the protein (*5*). Its unique properties as a storage and delivery system for hydrophobic compounds make it a domain of great interest for investigation, which led in part to the research done in this dissertation. The data presented indicate a potential storage function of COMPcc *in vivo* and open the possibility for COMPcc to be developed as a targeted drug delivery system.

Agrin is also a key protein found in vertebrates. Synapse formation at the NMJ relies on specific interactions between agrin and other ECM components such as laminin and α-dystroglycan (*12*). Understanding the roles of the individual domains in agrin and how these domains function together is essential to elucidating the mechanism of postsynaptic differentiation at the NMJ. The discovery of a novel interdomain disulfide bridge between NtA and the first FS domain in agrin sheds new light on the understanding of agrin as a whole. In all, this thesis makes a distinct contribution to the comprehension of the structural biology of the ECM. The work presented brings new insights into this diverse field and opens the door for new and exciting research.

## 5.2. Future perspectives

The study of ECM proteins through an integrated approach utilizing various structural biology techniques is a powerful method for understanding the important structure-function relationships of these proteins. This thesis lays the groundwork for developing COMPcc as a targeted drug delivery carrier system. The extraordinary storage functions of COMPcc make it ideally suited for such an application. Future studies should be aimed at further investigating this type of system. The ability of COMPcc to store and deliver cargo molecules *in vivo* should be studied via cellular and pharmacological assays.

The Gln54 ring of COMPcc is a desirable target for future experimentation. Binding of hydrophobic compounds to COMPcc is achieved through the network of hydrogen bonds formed by the ring of five glutamines at amino acid position 54. It would be interesting to mutate this residue and examine the change in binding ability of COMPcc with various hydrophobic compounds. Additionally, larger and more complex molecules can be tested within the protein's channel to determine its size constraints.

Agrin is a complex ECM protein with important cellular functions. Clearly, the interplay between the multiple domains of agrin is essential to its function. Future work should be aimed at determining the structures of agrin domains in tandem to elucidate their structure-function relationships. Small angle x-ray scattering could be used to reveal the overall structure of the multidomain agrin protein. NtA itself is an interesting protein for future considerations. It is structurally similar to N-TIMP-1, suggesting that it may have an additional functional role as an MMP inhibitor (*13*). Research can be aimed at analyzing the potential binding and inhibitory properties of NtA on MMP's. This can be done through kinetic experiments and MMP enzyme

assays. Finally, x-ray crystallography can be used to determine the structure of agrin-MMP complexes in order to guide potential structure-based inhibitor design.

**References**


1. Hedbom, E., Antonsson, P., Hjerpe, A., Aeschlimann, D., Paulsson, M., Rosa-Pimentel, E., Sommarin, Y., Wendel, M., Oldberg, A., and Heinegard, D. (1992) Cartilage matrix proteins. An acidic oligomeric protein (COMP) detected only in cartilage, *J. Biol. Chem. 267*, 6132-6136.
2. DiCesare, P., Hauser, N., Lehman, D., Pasumarti, S., and Paulsson, M. (1994) Cartilage oligomeric matrix protein (COMP) is an abundant component of tendon, *FEBS Lett. 354*, 237-240.
3. Smith, R. K., Zunino, L., Webbon, P. M., and Heinegard, D. (1997) The distribution of cartilage oligomeric matrix protein (COMP) in tendon and its variation with tendon site, age and load, *Matrix Biol. 16*, 255-271.
4. Muller, G., Michel, A., and Altenburg, E. (1998) COMP (cartilage oligomeric matrix protein) is synthesized in ligament, tendon, meniscus, and articular cartilage, *Connect. Tissue Res. 39*, 233-244.
5. Malashkevich, V. N., Kammerer, R. A., Efimov, V. P., Schulthess, T., and Engel, J. (1996) The crystal structure of a five-stranded coiled coil in COMP: a prototype ion channel?, *Science 274*, 761-765.
6. Ozbek, S., Engel, J., and Stetefeld, J. (2002) Storage function of cartilage oligomeric matrix protein: the crystal structure of the coiled-coil domain in complex with vitamin D(3), *EMBO J. 21*, 5960-5968.

7. Guo, Y., Bozic, D., Malashkevich, V. N., Kammerer, R. A., Schulthess, T., and Engel, J. (1998) All-trans retinol, vitamin D and other hydrophobic compounds bind in the axial pore of the five-stranded coiled-coil domain of cartilage oligomeric matrix protein, *EMBO J. 17*, 5265-5272.

8. McFarlane, A. A., Orriss, G. L., and Stetefeld, J. (2009) The use of coiled-coil proteins in drug delivery systems, *Eur. J. Pharmacol. 625*, 101-107.

9. Ruegg, M. A., and Bixby, J. L. (1998) Agrin orchestrates synaptic differentiation at the vertebrate neuromuscular junction, *Trends Neurosci. 21*, 22-27.

10. Denzer, A. J., Gesemann, M., Schumacher, B., and Ruegg, M. A. (1995) An amino-terminal extension is required for the secretion of chick agrin and its binding to extracellular matrix, *J. Cell Biol. 131*, 1547-1560.

11. Denzer, A. J., Hauser, D. M., Gesemann, M., and Ruegg, M. A. (1997) Synaptic differentiation: the role of agrin in the formation and maintenance of the neuromuscular junction, *Cell Tissue Res. 290*, 357-365.

12. Mascarenhas, J. B., Ruegg, M. A., Winzen, U., Halfter, W., Engel, J., and Stetefeld, J. (2003) Mapping of the laminin-binding site of the N-terminal agrin domain (NtA), *EMBO J. 22*, 529-536.

13. Stetefeld, J., Jenny, M., Schulthess, T., Landwehr, R., Schumacher, B., Frank, S., Ruegg, M. A., Engel, J., and Kammerer, R. A. (2001) The laminin-binding domain of agrin is structurally related to N-TIMP-1, *Nat. Struct. Biol. 8*, 705-709.