

**STRUCTURE AND FUNCTION OF THE  
IRON-REGULATORY PROTEIN,  
REPULSIVE GUIDANCE MOLECULE C /  
HEMOJUVELIN**

By

Mahta Nili

A DISSERTATION

Presented to the Department of Biochemistry and Molecular Biology

and Oregon Health & Science University

School of Medicine

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

September 2011

Department of Biochemistry & Molecular Biology

School of Medicine

OREGON HEALTH & SCIENCE UNIVERSITY

---

**CERTIFICATE OF APPROVAL**

---

This is to certify that the Ph.D. dissertation of

**Mahta Nili**

has been approved by the following:

---

Peter Rotwein, M.D., dissertation advisor

---

Linda Musil, Ph.D., committee chair

---

William Skach, M.D, committee member

---

Ujwal Shinde, Ph.D., committee member

---

David Dawson, Ph.D., committee member

## TABLE OF CONTENTS

	<i>Page</i>
List of Tables	iv
List of Figures	v
List of Abbreviations	viii
Acknowledgements	xiii
Abstract	xiv
<u>Chapter 1</u> Introduction	
1.1 Iron Regulation	3
1.2 Structure and Function of RGMc/HJV	9
1.3 Disulfide Mapping by Tandem MS Combining ETD and CID	19
1.4 Dissertation Overview	30
<u>Chapter 2</u> Soluble RGMc/Hemojuvelin is a Broad Spectrum BMP Antagonist and Inhibits both BMP2 and BMP6-mediated Signaling and Gene Expression	
2.1 Summary	48
2.2 Introduction	49

2.3	Materials and Methods	52
2.4	Results	59
2.5	Discussion	66
2.6	Acknowledgements	73
<u>Chapter 3</u>	<b>Defining the Disulfide Bonds of Insulin-Like Growth Factor Binding Protein-5 by LC-MS with Electron Transfer Dissociation</b>	
3.1	Summary	103
3.2	Introduction	104
3.3	Materials and Methods	108
3.4	Results	113
3.5	Discussion	119
3.6	Acknowledgements	123
<u>Chapter 4</u>	<b>Disulfide Mapping of RGMc/Hemojuvelin by LC-MS Reveals Two Oxidation States</b>	
4.1	Summary	153
4.2	Introduction	154
4.3	Materials and Methods	156
4.4	Results	158
4.5	Discussion	161
4.6	Acknowledgements	166

<u>Chapter 5</u>	Conclusions and Future Directions	
5.1	Conclusions	186
5.2	Future Directions	191
5.3	Dissertation Summary	197
	References	198
<u>Appendix</u>	Juvenile-hemochromatosis linked RGMc C-terminal truncation mutants are cleaved into smaller isoforms	
A.1	Summary	227
A.2	Introduction	228
A.3	Materials and Methods	230
A.4	Results	233
A.5	Discussion	236
A.6	Acknowledgements	239

## LIST OF TABLES

<i>Number</i>		<i>Page</i>
<b>1.1</b>	HJV mutations linked to juvenile hemochromatosis	32
<b>2.1</b>	Primers used for RT-PCR analysis of BMP-mediated signaling	74
<b>2.2</b>	Genes induced $\geq 2.0$ -fold by BMP2 or BMP6 in Hep3B cells	75
<b>2.3</b>	Effects of Noggin and RGMc40 on genes induced by BMP2	76
<b>2.4</b>	Effects of Noggin and RGMc40 on genes induced by BMP6	77
<b>3.1</b>	m/z values of IGFBP-5 cysteine-containing tryptic peptides	124
<b>3.2</b>	CID-MS3 y and b ions for IGFBP-5 C-terminal peptides	125
<b>3.3</b>	CID-MS2 y and b ions for IGFBP-5 C-terminal peptides	126
<b>3.4</b>	y and b ions for establishing Cys <sup>47</sup> -Cys <sup>60</sup> disulfide-linkage	127
<b>3.5</b>	y and b ions for establishing linkages between N-terminal chymotrypsin peptides	128
<b>3.6</b>	y and b ions for establishing Cys <sup>54</sup> -Cys <sup>80</sup> disulfide-linkage	129
<b>4.1</b>	m/z values of selected RGMc cysteine-containing tryptic peptides	167
<b>4.2</b>	y and b CID-MS3 ions for establishing Cys <sup>178</sup> -Cys <sup>314</sup> linkage	168
<b>4.3</b>	y and b CID-MS3 ions for establishing Cys <sup>160</sup> -Cys <sup>310</sup> linkage	169
<b>4.4</b>	CID-MS3 y and b ions for defining Cys <sup>160</sup> -Cys <sup>178</sup> disulfide-linkage	170

## LIST OF FIGURES

<i>Number</i>		<i>Page</i>
<b>1.1</b>	Systemic iron homeostasis	34
<b>1.2</b>	Normal iron homeostasis mediated by an iron-sensing feedback loop	36
<b>1.3</b>	Domain map and isoforms of RGMc	38
<b>1.4</b>	Opposing actions of cell-associated and soluble RGMc/HJV in BMP regulation	40
<b>1.5</b>	Generalized tandem MS protocol flowchart	42
<b>1.6</b>	The activation steps CID and ETD	44
<b>1.7</b>	Sequences of the cysteine-rich proteins IGFBP-5 and RGMc	46
<b>2.1</b>	BMP2 and BMP6 rapidly stimulate Smad phosphorylation in the Hep3B liver cell line	79
<b>2.2</b>	BMP2 and BMP6 up-regulate BMP-responsive genes in Hep3B cells	81
<b>2.3</b>	BMP6 does not induce hepcidin gene expression in the mouse AML12 liver cell line	83
<b>2.4</b>	Purification of soluble RGMc proteins from mammalian cells	85
<b>2.5</b>	Characterization of soluble RGMc proteins from mammalian cells	87
<b>2.6</b>	Purification and characterization of RGMc40	89
<b>2.7</b>	Dose-dependent inhibition of BMP2- and BMP6-mediated signaling by RGMc50 and RGMc40	91
<b>2.8</b>	Time course of RGMc40 inhibition of BMP2-mediated signaling	93

<b>2.9</b>	Dose-dependent inhibition of BMP2 and BMP6-mediated gene expression by RGMc50 and RGMc40	95
<b>2.10</b>	RGMc40 inhibits BMP2- and BMP6-mediated gene expression	97
<b>2.11</b>	Purification of mutant RGMc proteins	99
<b>2.12</b>	Selected mutant RGMc proteins inhibit BMP-mediated gene expression	101
<b>3.1</b>	Experimental plan for identifying the disulfide bonds of IGFBP-5	131
<b>3.2</b>	Purification and characterization of IGFBP-5	133
<b>3.3</b>	Expected cysteine-containing peptides of IGFBP-5	135
<b>3.4</b>	Analysis of IGFBP-5 tryptic peptides by MS	137
<b>3.5</b>	Defining disulfide linkages in the C-terminal segment of IGFBP-5 (I)	139
<b>3.6</b>	Defining disulfide linkages in the C-terminal segment of IGFBP-5 (II)	141
<b>3.7</b>	Cysteines in the N-terminal domain of IGFBP-5 are located in highly intertwined peptides	143
<b>3.8</b>	Identifying disulfide bonds within the N-terminal domain of IGFBP-5	145
<b>3.9</b>	Alternative steps to identify disulfide linkages in the N-terminal domain of IGFBP-5	147
<b>3.10</b>	Structural predictions and disulfide bond map of the N-terminal domain of IGFBP-5	149
<b>3.11</b>	Amino acid substitution mutations in the N-terminal domain of IGFBP-5 do not alter disulfide-linkages	151
<b>4.1</b>	Experimental model for identifying the disulfide bonds of RGMc	172
<b>4.2</b>	Selected RGMc cysteine-containing tryptic peptides	174

<b>4.3</b>	The N-terminus of RGMc contains one disulfide bond linking Cys <sup>39</sup> to Cys <sup>82</sup>	176
<b>4.4</b>	<i>Ab initio</i> model of the N-terminal domain of RGMc	178
<b>4.5</b>	The disulfide linkages, Cys <sup>160</sup> -Cys <sup>310</sup> and Cys <sup>178</sup> -Cys <sup>314</sup> , in RGMc	180
<b>4.6</b>	Cys <sup>160</sup> is linked to Cys <sup>178</sup> in a second RGMc oxidation state	182
<b>4.7</b>	<i>Ab initio</i> models of single- and two-chain RGMc	184
<b>A.1</b>	Endogenous RGMc exists as soluble 50 and 40 kDa isoforms	241
<b>A.2</b>	A PC inhibitor prevents accumulation of soluble 40 kDa RGMc	243
<b>A.3</b>	Map of mouse RGMc C-terminal truncation mutants	245
<b>A.4</b>	Juvenile Hemochromatosis-linked RGMc C-terminal truncation mutants are cleaved into smaller isoforms	247
<b>A.5</b>	A PC inhibitor affects processing of selected RGMc C-terminal truncation mutants	249

## LIST OF ABBREVIATIONS

$\alpha$ -	anti-; alpha-
A	absorbance
A (amino acid)	alanine
Ad	adenovirus
Ala	alanine
Arg	arginine
Asp	aspartic acid
$\beta$ -	beta-
BCA	bicinchoninic acid
BMP	bone morphogenetic protein
C (amino acid)	cysteine
$^{\circ}\text{C}$	degrees Celsius
C-	carboxyl-
CD	circular dichroism
cDNA	complementary DNA
CHAPS	3-((3-Cholamidopropyl)dimethylammonium)-1-propanesulfonate
CID	collision induced dissociation
Cys	cysteine
D (amino acid)	aspartic acid
DMEM	dulbecco's modified eagle medium

DMT1	divalent metal transporter 1
DNA	deoxyribonucleic acid
DTT	dithiothreitol
DcytB	duodenal cytochrome b
E (amino acid)	glutamic acid
EDTA	ethylenediaminetetraacetic acid
ER	endoplasmic reticulum
ESI	electrospray ionization
ETD	electron transfer dissociation
FBS	fetal bovine serum
f.s.	frameshift
G (amino acid)	glycine
Gln	glutamine
Glu	glutamic acid
Gly	glycine
GPI	glycosylphosphatidylinositol
HJV	hemojuvelin
HOX1	heme oxygenase 1
hr	hour
IGF	insulin-like growth factor
IGFBP	insulin-like growth factor binding protein
IgG	immunoglobulin G
IR	infrared

JH	juvenile hemochromatosis
K (amino acid)	lysine
kDa	kilodalton
LC	liquid chromatography
Lys	lysine
M	molar
MALDI	matrix assisted laser desorption/ionization
MEM	minimum essential medium
μg	microgram
μl	microliter
mg	milligram
min	minute
ml	milliliter
μM	micromolar
mM	millimolar
mRNA	messenger RNA
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MW	molecular weight
N (amino acid)	asparagine
N-	amino-
NaCl	sodium chloride
NaDOC	sodium deoxycholate

ng	nanogram
nm	nanometer
nM	nanomolar
NP-40	Tergitol-type NP-40
P (amino acid)	proline
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PC	pro-protein convertase
PCR	polymerase chain reaction
PDI	protein disulfide isomerase
Phe	phenylalanine
PNGaseF	N-Glycosidase F
Pro	proline
PTM	post-translational modification
R (amino acid)	arginine
RGD	Arg-Gly-Asp (motif)
RGM	repulsive guidance molecule
RGMa/b/c	RGM family members a, b, or c
RGMc40	RGMc soluble 40 kDa isoform
RGMc50	RGMc soluble 50 kDa isoform
RNA	ribonucleic acid
RT-PCR	reverse transcriptase-PCR

RVKR	furin convertase inhibitor decanoyl-RVKR-chloromethylketone
SDS	sodium dodecyl sulfate
sec	second
Ser	serine
T (amino acid)	threonine
TEV	tobacco etch virus
TGF- $\beta$	transforming growth factor- $\beta$
Trp	tryptophan
U	unit
Val	valine
vWD	von Willebrand type-D
WT	wild type

## ACKNOWLEDGEMENTS

The path to my dissertation was filled with challenges, determination, and moments of fear, all mixed in with fun. I am sincerely grateful to many people for supporting me throughout this process, both scientifically and personally. I would like to first thank my advisors, Dr. Peter Rotwein and Dr. Ujwal Shinde, who genuinely provided a yin yang style of mentorship. They constantly challenged me to think critically about my work and that of others, which led me to become the scientist I am today. I would also like to thank key members of the Rotwein lab – Dr. David Kuning, Lisa Wilson, Dr. Aditi Mukherjee, Christopher Severyn, Dr. Robin Kuns-Hashimoto, and Dr. Dennis Chia – for providing an enriching intellectual scientific environment to work in. I especially wish to acknowledge Dr. Larry David for his assistance and expertise on mass spectrometry. I thank the members of my research advisory committee – Dr. Linda Musil, Dr. William Skach, Dr. David Dawson, and Dr. Ujwal Shinde – for their helpful guidance.

Finally, I wish to thank my family and friends for making this dissertation an achievable accomplishment. I thank my parents, Nasrin and Ali, for their constant love, support, and encouragement and of course my sisters and new brother-in-law, Mahyar, Mahhaun, and Sebastian, for their love and friendship. I also thank many extended family members of both the Riazi and Nili clans, with a special thanks to my grandmother for always thinking about me. Several friends – Melanie Laederich, Eric Stoffregen, Aditi Mukherjee, Jonathan Fay, Tony Capps, and Eileen Burkart-Hartman – definitely made graduate school a fun and enjoyable place to be each and every day.

## ABSTRACT

Iron homeostasis is a tightly regulated process that is maintained by several key factors that ensure proper iron distribution and storage throughout the body. Disruption in homeostatic processes can lead to inadequate iron levels, which impairs hemoglobin synthesis and red cell production, while chronic excess of iron causes tissue damage. Excessive iron accumulation in organs such as the liver, heart and pancreas is a hallmark characteristic of hereditary hemochromatosis. Mutations in RGMc/HJV cause juvenile hemochromatosis (JH), an early onset aggressive form of this disorder. Patients with juvenile hemochromatosis, and RGMc knockout mice, have reduced expression of the critical iron-regulatory peptide, hepcidin. Although RGMc plays a key role in the regulation of iron homeostasis, many of the mechanisms of RGMc actions are unknown. This dissertation explores RGMc actions through a structure-function approach. More specifically, functional interactions with members of the bone morphogenetic family (BMP) are explored, and disulfide mapping studies provide insight into the three-dimensional structure of RGMc isoforms.

Several significant findings are described. First, RGMc is demonstrated to be a potent antagonist of BMP-mediated signaling and gene expression with a specificity that encompasses two distinct BMP sub-families. Full-length and PC-cleaved soluble RGMc isoforms, and selected mutants that retain the ability to bind BMPs, are shown to antagonize the actions of both BMP2 and BMP6. Secondly, by using a newly-described MS3 disulfide bond mapping method combining ETD and CID, a definitive map has

been established of several disulfide bonds of the growth factor binding protein, insulin-like growth factor binding protein-5 (IGFBP-5), representing the first use of this method to define disulfide bonds of a protein with unknown structure. Finally, through the same MS3-based investigative approach coupled with *ab initio* molecular modeling, two disulfide bonding patterns representing two distinct oxidation states of RGMc are uncovered. In summary, the structure-function studies reported in this dissertation provide major insight into the iron-regulatory mechanisms of action of RGMc.

# **Chapter 1**

## **Introduction**

Hemochromatosis is a hereditary iron overload disorder in which the excessive accumulation of iron in organs such as the liver, heart and pancreas leads to their damage and dysfunction. Mutations in repulsive guidance molecule c (RGMc) / hemojuvelin (HJV) cause juvenile hemochromatosis, an aggravated form of this disorder that presents earlier in life. Patients with juvenile hemochromatosis, and RGMc knockout mice, have diminished expression of the key iron-regulatory peptide, hepcidin. Thus RGMc plays a critical role in the regulation of iron homeostasis; however the mechanisms of RGMc actions under physiological conditions and during the progression of disease are unknown. Although it is known that RGMc binds to both members of the bone morphogenetic protein (BMP) family and the transmembrane receptor neogenin, limited structural analyses of the protein has made it difficult to fully appreciate the different properties of various RGMc isoforms. Recent advances in MS have provided new disulfide mapping methodologies, which have become instrumental in obtaining tertiary structural data for cysteine-rich proteins such as RGMc.

This introductory chapter will provide background information and a brief review of the literature pertinent to the experimental results in this dissertation. A dissertation overview is presented at the end of this chapter.

## **1.1 IRON REGULATION**

### **1.1.1 Iron Absorption and Recycling**

Proper regulation of iron levels is critical for many fundamental cellular and metabolic processes. Organisms must acquire sufficient iron for maintaining homeostasis; however, a threshold must not be surpassed as excess free iron is toxic (1). Iron is a key component of several proteins, including oxygen-carrying globins, cytochromes that are essential for oxidative phosphorylation, and enzymes that transfer electrons (2).

Iron is almost exclusively taken up from the diet at the brush border of duodenal enterocytes via the divalent metal transporter 1 (DMT1) (3). Non-heme iron is insoluble and bio-unavailable, thus before absorption it must be reduced by a ferrireductase such as duodenal cytochrome b (DcytB) (4). To date, the mechanism by which heme iron is absorbed is unresolved; however, it is likely that once it is disassembled by heme oxygenase 1 (HOX1), the free iron enters the same pathways as non-heme inorganic iron (5). Internalized cytosolic iron can be released into circulation by the dual actions of hephaestin and ferroportin, which oxidize and export iron, respectively (6, 7). Circulating iron is rapidly bound to transferrin, a high-affinity iron-binding protein that increases the solubility and dampens the reactivity of iron (1).

The human body stores between 3 to 5 g of iron (8). Only 1-2 mg of iron is absorbed daily to compensate for loosely regulated losses, such as shedding of skin and urinary cells, loss of blood or sweat, or routine sloughing of intestinal epithelial cells (2). The majority of iron routinely used by an organism is recycled by macrophages located

primarily in the spleen (1). Macrophages take up damaged or aged erythrocytes by phagocytosis, wherein the erythrocyte membrane is degraded by the lysosome (1). HOX1 catalyzes the oxidation of heme to free iron which can then be pumped into the cytoplasm by DMT1 and further released by the macrophage via ferroportin (6). Figure 1.1 depicts the distribution flux of iron throughout the human body.

### **1.1.2 Heparidin, the master iron regulator**

The iron-regulatory hormone hepcidin plays the most crucial role in countering toxic accumulation of iron by regulating its storage, concentration, and distribution in tissues (9). Heparidin was discovered ten years ago as a gene that was up-regulated approximately 10-fold in mice overloaded with iron (10). Subsequently it was shown that mice with a disruption in the hepcidin gene experienced severe iron overload (11). Heparidin is a small peptide hormone that consists of 25 amino acids and contains 4 disulfide bonds (9). Mice administered synthetically-derived hepcidin display a severe anemic phenotype (12).

Near the same time that hepcidin was discovered, 3 groups using distinct approaches independently identified the iron exporter ferroportin in macrophages and duodenal enterocytes (13-15). Subsequent to its discovery, Nemeth et al. showed that in cells expressing ferroportin-GFP, hepcidin treatment caused the rapid internalization and degradation of ferroportin within the lysosome (16). This study also demonstrated that ferroportin was the direct receptor for hepcidin through radiolabeled binding studies (16), adding to its role as the hepcidin-regulated iron exporter. The interactions between hepcidin and ferroportin are crucial for proper maintenance of systemic iron homeostasis

(5). When hepcidin concentrations are low, iron is released through ferroportin from primary iron storage cells, including duodenal enterocytes, macrophages, and hepatocytes, and conversely, when hepcidin levels are high, ferroportin is not present on the surface of these cells, thereby holding iron within them (5) (Figure 1.2).

It has been demonstrated through mouse models, that hepcidin mRNA is induced by iron loading (10, 17), while suppressed by anemia (17). Determining how the hepcidin protein in circulation was regulated by iron proved to be difficult for many years due to the lack of hepcidin antibodies for immunoassays, however recent advances in mass spectrometric assays have allowed for hepcidin detection in serum (9). These methods have allowed for detection of an approximate 5-fold increase in hepcidin levels in healthy volunteers within a day after loading with 65 mg of iron (18). This rise in circulating hepcidin occurs ~4-8 hr after oral iron loading (19). Recapitulating these findings in a cell culture model system were not initially successful, as many hepatocyte-derived cell lines did not respond to iron treatment (9). However, mouse hepatocytes treated with physiological concentrations of holo-transferrin increased their hepcidin mRNA (20, 21).

Hepcidin is transcriptionally regulated by several iron-regulatory proteins. Much of the evidence for this has been presented through transgenic mouse models in combination with biochemical studies. The iron carrier transferrin presents circulating iron to transferrin receptors 1 and 2 which sense and uptake iron (22-24), and it is thought that HFE is the signaling component of a HFE-transferrin receptor complex that regulates hepcidin levels (23, 25). Increased hepatic iron concentrations enhance the expression of

BMP6 (26), which in combination with hemojuvelin (27) and neogenin (28, 29), increase BMP signaling and hepcidin expression, while TMPRSS6 acts as a negative regulator of this hepcidin up-regulation by cleaving hemojuvelin (30). Although these key players in iron regulation have been identified, a complete model of how these components physically interact, and all the intracellular signaling components involved that regulate hepcidin levels, remain to be determined (9).

### **1.1.3 BMP Signaling**

BMPs are members of the transforming growth factor- $\beta$  (TGF- $\beta$ ) superfamily of ligands that regulate a wide range of developmental and cell fate decisions (31). BMPs bind as dimers to specific type I and type II serine/threonine kinase receptors, and initiate a protein kinase cascade which culminates in the activation by serine phosphorylation of Smads 1, 5, and 8 (31). Activated Smads bind to the co-Smad, Smad4, and this heterodimeric signal transducer complex regulates the expression of many BMP-dependent target genes (31, 32). To finely tune these signals and generate specificity, regulation of this pathway occurs at multiple levels (33).

Binding of ligands or co-receptors can modulate BMP binding to cell surface receptors resulting in either the enhancement or repression of downstream signaling (31). The first reported co-receptors for the BMP subfamily of ligands are members of the RGM family, RGMa, b, and c (34-36), which have all been described as enhancers of BMP signaling. A general mode of inhibition of BMPs is through binding of antagonists to them and blocking receptor binding sites (37). Various antagonists have different

efficacies for different BMP family members (38). For example, Chordin is a relatively potent inhibitor of BMP2 and 4 (39) and follistatin is fairly effective against BMP7 (40), while DAN and follistatin-related protein are fairly weak toward BMP2, 4, and 7 (41-43). Chordin-like and Sclerostin can block the actions of BMP6, but not BMP2 (44, 45), and CTGF and Nov family members primarily inhibit BMP2 and 4 (46, 47).

#### **1.1.4 BMPs and Iron Regulation**

In 2006, two laboratories demonstrated in *in vitro* cell culture systems that hepcidin expression was induced by the BMP signaling pathway thereby providing the first link between BMPs and iron regulation (34, 48). Truksa et al. demonstrated that BMP2, 4 and 9 up-regulated hepcidin expression (48), while Babitt et al. supported this finding and additionally demonstrated that RGMc acts as a co-receptor that enhances BMP2-mediated hepcidin expression (34). Further evidence of a link between BMPs and iron regulation come from mice with a conditional liver-specific knockout of Smad4 that exhibit low levels of hepcidin and develop iron overload (49).

Kautz et al. demonstrated through gene array analyses that BMP6 was the only BMP family member that was increased in the liver of mice, at the mRNA level, after being fed high iron diets (26). Subsequent to this finding, two pivotal papers demonstrated that BMP6 played a crucial role *in vivo* in the regulation of iron (27, 50). Meynard et al. reported that mice with a liver-specific knockout of BMP6 experienced a massive accumulation of iron in the liver and several other tissues (50). In addition, it was reported that these knockout mice had dramatically reduced expression of hepcidin,

and lower levels of phosphorylated Smads1, 5, and 8 (50). In separate studies, Andriopoulos et al. demonstrated that mice treated with BMP6 exhibited higher levels of hepcidin expression and had lower levels of circulating iron in their serum (27). Thus, although several BMPs can stimulate the up-regulation of hepcidin in cell culture model systems, BMP6 appears to be the prime regulator of hepcidin *in vivo*.

## **1.2 STRUCTURE AND FUNCTION OF RGMc/HJV**

### **1.2.1 RGM Family**

The three members of the repulsive guidance molecule (RGM) family, RGMa, RGMb, and RGMc (51) were named based on the axonal guidance properties of the first discovered member, RGMa (52). RGM family members exhibit unique tissue expression patterns (53, 54), with RGMa and RGMb in distinct regions of the developing and adult nervous systems (52, 55, 56) and RGMc residing in skeletal muscle, heart, and liver (54, 57). More specifically, RGMa localizes to regions in the brain, including the hippocampus, midbrain, brainstem and the ventricular zone of the cortex, while RGMb has been localized to the dorsal root ganglia of the sympathetic nervous system, in the developing neural tube, and in regions of the spinal cord (53, 54).

RGM family members, expressed as primary translation products ranging from 420 to 438 amino acids (54-56, 58), are approximately 45% similar in sequence (53, 54, 59) and share several motifs and structural features (51). The three glycoproteins contain a N-terminal signal peptide, a C-terminal glycosylphosphatidylinositol (GPI) signal sequence, and three asparagine-linked glycosylation motifs (51). Each protein also possesses a RGD motif, which may serve a role in integrin binding (60), and a partial von Willebrand type-D domain (vWD) which has been shown to play roles in clotting factors (61). The partial vWD domain contains a site of internal cleavage which can give rise to two-chain species, shown to exist in RGMa and RGMc (52, 58).

Although RGMa, b, and c share several domain features, these three proteins possess divergent functions that appear to be mediated either by the cell surface receptor

neogenin (29, 62-66) or members of the BMP family (20, 27, 34-36), or a combination of both (67). Through its interaction with neogenin, RGMa has been shown to regulate the repulsive guidance of retinal axons (63, 64) and promote neuronal survival (63). To date, RGMb has not been shown to bind to neogenin, and its only known function is a possible role in promoting cell-cell adhesion by homophilic interactions (56). RGMc was linked to systemic iron metabolism when mutations in the gene were found in patients with the severe iron overload disorder, juvenile hemochromatosis (59). The two-chain RGMc isoform preferentially binds to neogenin (68), however a definitive role of this interaction in iron regulation has not been established. It has been postulated that neogenin is required for RGMc to be released from the cell surface (65, 69) and is an essential component of a complex including RGMc, BMPs and BMP receptors (67).

All members of the RGM family have been shown to bind members of the BMP family (34-36, 68, 70), while they do not bind to TGF- $\beta$  (70), proteins that reside in a larger superfamily with BMPs. Many of these studies, including solution-based pull-downs and cross-linking assays, have been performed with RGM-Fc fusion proteins wherein the Fc domain of IgG is linked to the C-termini of RGM proteins lacking GPI-anchor signal sequences (27, 68). Through over-expression studies it has been shown that RGMa, b, and c increase gene activity of a promoter-reporter construct containing a BMP response element (34-36), while for RGMa it has been shown that endogenous knockdown of this protein results in reduced promoter-reporter gene activity (71). The prevailing accepted role of cell surface RGMs in BMP biology is that they serve as BMP co-receptors that enhance BMP-mediated signaling (27, 35, 36). For example, RGMc has

been shown to augment BMP-mediated up-regulation of hepcidin gene expression (27, 34).

### **1.2.2 RGMc/HJV Structure**

Murine RGMc and human hemojuvelin, expressed as primary translation products of 420 and 426 amino acids, respectively, are 88% identical (53, 54, 58, 59). See Figure 1.3A for a RGMc domain map. The predicted length of the N-terminal signal sequence of RGMc/HJV is ~31 residues, while the C-terminal GPI-signal sequence is ~45 amino acids (59). The RGD domain is located at amino acids 91-93 in RGMc and although this motif has been shown to play a role in integrin binding in other proteins (72), these amino acids may be crucial for binding to BMPs as mutations in this region abolish this interaction (68). The amino acids GDPH, which reside within the N-terminal region of a partial vWF domain, are the site at which RGMc undergoes internal cleavage, between the Asp and Pro residues (73). It has still not been established if cleavage event at this site is enzyme-based or due to acid-hydrolysis. RGMc can also be cleaved at its C-terminus to generate smaller soluble isoforms (30, 58, 74-76). A defined PC-cleavage site exists at arginine 328 (75), and several groups have shown that furin can cleave RGMc at this position (74-76). More recently, the iron-regulatory protein TMPRSS6 has been shown to cleave membrane-bound RGMc from the cell surface (30), however attempts at defining the position of this cleavage have been unsuccessful (77).

RGMc contains three N-linked glycosylation sites (NXS/T), and it has been established through studies with the endo-glycopeptidases PNGaseF and Endo H, that RGMc is a glycoprotein (58, 78), however it has not yet been established which of the

three sites are linked to carbohydrate groups. Unpublished studies performed by our lab show markedly decreased levels of protein expression when these N-linked sites are mutated singly or in combination with one another. Therefore it is possible that these glycans are required for the proper folding of RGMc, and potentially have further roles in interactions with binding partners.

RGMc was first cloned in 2004 and to date no structural data has been reported. Three-dimensional structural information can provide crucial insight into the function of RGMc. Fourteen evolutionarily conserved cysteines reside in RGMc (51), which can potentially give rise to 7 disulfide bonds. The disulfide bonding status of these cysteines remains to be elucidated. Unpublished circular dichroism studies performed by our lab indicate that loss of secondary structure is achieved upon exposure of RGMc to reducing agents. Additional studies performed with SDS-PAGE have indicated that at least one disulfide bond exists within the two-chain RGMc isoform, as two protein bands of 20 and 30 kDa appear under reducing conditions while only one protein band of 50 kDa is present when the protein is not reduced (58). Our lab has generated comprehensive three-dimensional *ab initio* models for the RGM family that predicted the placement of disulfide linkages through computational methods (51). Every model predicted RGMc to be a two domain protein with one or two disulfide linkages in the N-terminal domain, two disulfide bonds within the C-terminal domain, and one bond linking the N- and C-terminal domains together (51). Preliminary modeling has provided a framework for predicting disulfide linkages within RGMc; however, it is clear that experiments directed at mapping the disulfide bonds are necessary to provide definitive structural data.

### 1.2.3 Complex Biosynthesis and Processing of RGMc/HJV

RGMc undergoes a series of complex biosynthesis and processing steps to give rise to four distinct protein isoforms, including both GPI-linked and soluble protein species (Figure 1.3B) (20, 58, 68, 73, 75). These isoforms have been identified *in vivo* in human and mouse serum (73, 74), endogenously in muscle differentiation assays (58), and through over-expression studies analyzing the recombinant protein in a variety of mammalian cell-culture model systems (58, 73, 75). Recently a fifth putative RGMc soluble protein species was identified that may be specifically localized to the liver (77).

Two RGMc isoforms, a single-chain 50 kDa species and a two-chain 20 and 30 kDa protein, are linked to the plasma membrane (58). Immunocytochemistry experiments initially demonstrated that RGMc is present on the cell surface of differentiated C2 myotubes (58). Furthermore, incubation of differentiated myoblasts with bacterial PI-PLC, which cleaves GPI-linked proteins (79), led to the release of these two isoforms from the cell surface into the conditioned medium (58). Time course studies with an adenovirus encoding RGMc demonstrated that both cell surface isoforms rapidly appear on the cell surface within 4 hours post-infection (58). It is not yet known what cleavage event gives rise to the two-chain species, therefore it is possible that single-chain and two-chain cell surface RGMc traffic to the cell surface via distinct cellular pathways.

50 and 40 kDa soluble RGMc isoforms (RGMc50 and RGMc40, respectively) have been identified in both mouse and human serum, and are found in the conditioned medium of differentiating myoblasts and after over-expression of RGMc in a variety of

cell systems (20, 58, 73-75). Both RGMc50 and 40 are produced by differentiating C2 myoblasts as soon as 24 hr after the onset of differentiation (58). The mechanism by which RGMc50 is released by cells has not yet been determined. RGMc50 is cleaved by furin, and potentially other PCs, at a PC recognition site within the C-terminus of the protein to give rise to RGMc40 (74-76). Three consensus PC sites, RXXR, reside between residues 318 and 328, and furin can cleave RGMc after the arginine at position 328 (75). It has been shown that TMPRSS6, expressed in hepatocytes, can also release RGMc from the cell surface (30, 77). Initial over-expression studies by Silvestri et al. showed that RGMc was cleaved into multiple protein species (30), however a more recent study demonstrated that TMPRSS6 cleaves RGMc into a smaller soluble isoform than RGMc40 with an undefined cleavage site (77).

Mutational analyses of RGMc have revealed important findings about the processing of RGMc (58, 68, 80). Table 1.1 lists selected HJV disease-causing mutations. It has been suggested that RGMc mutations, in general, cause the protein to be retained within the endoplasmic reticulum (80), however several labs have studied mutant RGMc proteins that reside on the cell surface (34, 58, 68, 81). In our lab, we found that a mutation of Asp to Glu at position 165 (D165E) in the internal cleavage site abolished processing of RGMc to the two-chain species (68). A Gly to Val mutation at position 313 (G313V) also does not give rise to the two-chain species (68). However, both of these RGMc mutants were present on the cell surface (68).

#### **1.2.4 RGMc/HJV Interacts with BMPs and Neogenin**

Distinct RGMc isoforms display differential binding to neogenin and BMPs (68). Neogenin is a trans-membrane protein that is also a receptor for netrins, secreted molecules involved in cell survival and neuronal development (82). RGMa was initially discovered to mediate axonal guidance through interactions with neogenin (83). Subsequent to this it was shown that RGMc can bind to the extracellular domain of neogenin (84). It has been postulated that neogenin regulates release of RGMc from the cell surface (65, 69), perhaps playing a role in regulating soluble levels of the protein, and thereby modulating iron levels.

Through cell surface based binding studies it was demonstrated that two-chain RGMc preferentially binds to neogenin, and does not bind to BMPs (68). Although single-chain full-length 50 kDa RGMc isoforms bind to neogenin, binding is reduced compared to two-chain RGMc (68). Soluble RGMc40 does not bind to neogenin (68), implying that a region in the C-terminus along with cleavage in the N-terminal domain, produce an RGMc isoform that displays preferential binding regions for neogenin. Both soluble RGMc50 and 40 and cell surface single-chain RGMc bind to BMPs (27, 34, 68).

The prevailing model of the role of RGMc in hepcidin expression is that cell-associated RGMc enhances BMP signaling through its receptors to up-regulate hepcidin, while soluble RGMc has been proposed to act as a BMP inhibitor (9, 85) (Figure 1.4). A study by Babitt et al. first described RGMc as being a BMP co-receptor that signals via the Smad pathway to regulate expression levels of hepcidin (34). They reported that RGMc, in response to primarily BMP2 stimulation, was able to enhance the

phosphorylation of Smads1, 5, and 8, and further showed in a cell-based system that activity of a BMP response element was enhanced with co-transfection of RGMc (34). Subsequent to this finding the same group reported that BMP6 was the key regulator of hepcidin expression, and RGMc up-regulated BMP6 stimulatory effects (27). There are conflicting results as to whether neogenin plays a role in enhancement of BMP-mediated signaling (28, 67). Xia et al. showed that enhanced Smad signaling by BMPs and HJV were not altered by neogenin over-expression or knockdown (28), while Zhang et al. demonstrated that the interaction between RGMc and neogenin is required for BMP-induced hepcidin expression (67).

Lin et al. first proposed that soluble RGMc40 acted as an opposing regulator of hepcidin (73). In their initial studies recombinant RGMc40 was shown to suppress hepcidin mRNA in cultured primary human hepatocytes in a dose-dependent manner (73). Subsequent to these findings, the same group demonstrated that soluble RGMc40 inhibited BMP2, 4, and 9-mediated hepcidin expression in primary mouse hepatocytes (20). Babitt et al, showed that administration of soluble RGMc50, as a IgG-Fc fusion protein, to mice was able to decrease hepcidin levels (86).

Selected RGMc mutants display differential binding to BMPs and neogenin (68). Neither RGMc D165E or G313V, mutants that do not give rise to the internally cleaved form of the protein, are able to bind to neogenin (68). A Gly to Val mutation at position 92 (G92V) in the RGD domain does give rise to both single- and two-chain species on the cell surface (68), that are able to bind neogenin (68). However RGMc G92V cannot bind to BMPs (68), indicating that either the RGD domain plays a role in binding or that

the valine substitution at that position alters the structure of RGMc in a manner that precludes binding. In activity assays with a hepcidin promoter reporter construct it was shown that upon co-transfection of RGMc G313V or G92V, the G313V mutant was a more effective enhancer of BMP-mediated actions than the G92V mutant, however it was not as effective of an enhancer as wild-type RGMc (34).

### **1.2.5 RGMc/HJV and Iron Regulation**

Most of the mutations responsible for the iron overload disorder juvenile hemochromatosis, an aggravated form of hereditary hemochromatosis, have been linked to HJV, a gene that was identified through a positional cloning approach by Papanikolaou et al. in 2003 (59). The most common mutation is G320V (G313V in mice), which accounts for nearly two-thirds of patients with this disorder (59). Mouse RGMc knockout models have confirmed that this protein plays a crucial role in the development of juvenile hemochromatosis, by demonstrating that these mice accumulated extremely high levels of iron in several tissues (87, 88). Humans with mutations in HJV and mice lacking RGMc have low levels of serum and urinary hepcidin (59, 87), which results in increased intestinal iron absorption (87), suggesting that RGMc/HJV is a positive regulator of hepcidin.

To date, a clear link between iron levels and expression of RGMc/HJV has not been established *in vivo*. Several studies have indicated that enhanced processing of RGMc occurs in cell culture model systems upon iron loading, in that cell surface RGMc is cleaved to soluble RGMc (65, 74), however no changes in RGMc mRNA or protein levels have been detected. Humans with hereditary hemochromatosis showed no

statistically significant change in HJV levels, when compared to control patients (89), and furthermore mice lacking HFE, and therefore displaying iron overload, did not experience changes in RGMc mRNA or protein levels (90). Of note, these studies were in models of hereditary hemochromatosis and not in patients or mice with a juvenile hemochromatosis background. However, mean HJV expression levels were found to be significantly higher in physiological or secondary iron overload (91), iron loading models that may be more closely related to pathways activated in juvenile hemochromatosis. Contrary to this study, Krijt et al. reported that after iron overload in mice, hepcidin expression increased, while RGMc expression remained unchanged, as determined by real-time PCR (92). The conclusion reached in this study was that the hepcidin response was unlikely to be mediated by RGMc (92).

It is known that RGMc is a critical player in iron homeostasis, however many details of its mechanism of action, and regulation of its expression, have not been fully elucidated. Many more intricate details about the interactions between iron-regulatory players, predominantly RGMc/HJV, HFE, TfR1 and 2, and hepcidin need to be uncovered to understand how cell surface interactions and signal transduction pathways intersect to regulate systemic iron homeostasis.

## **1.3 DISULFIDE MAPPING BY TANDEM MS COMBINING CID AND ETD**

### **1.3.1 Formation of Disulfide Bonds**

One of the few post-translational modifications (PTMs) that takes place during the process of protein folding is the formation of the disulfide bond (93). Most secretory proteins and the extracellular domains of membrane proteins contain disulfide bonds in intra- or inter-molecular linkages, as this key modification is required not only for proper folding but also stability and function (94). If correct disulfide linkages are not formed, final native conformations are not achieved, oftentimes leading to protein aggregation (93).

Most disulfide bonds are stabilizing factors for protein structure, as it is generally accepted that they destabilize denatured protein states (95). Theoretical studies have indicated that as the distance (in primary amino acid sequence) between cysteine residues involved in a linkage increases, the stabilization provided to the final native conformation also increases (96). Although most disulfide bonds affect the rate of protein folding and provide stability, some serve functional roles (93, 94, 97). Cytosolic proteins, such as ribonucleotide reductase and oxio-reductases, must be oxidized as part of their catalytic cycles (93, 94). Disulfide bonds that regulate function have also been identified in photosynthetic enzymes and several transcription factors (94).

Formation of the disulfide bond is a reversible reaction that can be carried out by sulfhydryl-containing substrates, including small thiol-containing compounds such as glutathione (98), or thiol-containing enzymes such as protein disulfide isomerase (PDI) (94, 99). In a disulfide-exchange reaction, a free thiol is deprotonated to form a thiolate

anion, which displaces one sulfur of the disulfide bond of the oxidized species (94). In a second exchange, the next thiolate anion attacks the mixed-disulfide, whereby the originally oxidized protein or substrate is reduced and the originally reduced protein is oxidized (94). In addition to this, de novo disulfide bond formation can occur via the transfer of electrons to molecular oxygen (93, 94).

Oxidative folding requires specialized compartments where disulfide bond formation is favored. The two main sites for forming disulfide bonds in eukaryotes are the endoplasmic reticulum (ER) and the mitochondrial intermembrane space (IMS), while prokaryotic disulfide bonds are formed in the periplasm (97). In the cytosol, proteins are synthesized in the presence of high concentrations of reduced glutathione (~10 mM), thereby counteracting the formation of disulfide bonds (97, 98). Secretory proteins which traverse through the ER generally experience coupled translocation into the ER and disulfide bond formation (97). Electrons are transferred from reduced proteins to oxidized PDI family members causing reduction of conserved C-X-X-C motifs (94, 97, 100). Ero1 re-oxidizes PDIs by the transfer of electrons via FAD to oxygen (100). To ensure that the proper native conformation is achieved, cycles of oxidation and reduction occur with the ER to break and re-form disulfide bonds with proteins (94).

### **1.3.2 Tandem Mass Spectrometry**

Tandem mass spectrometry (MS/MS) is routinely used to provide structural information about biomolecules. Figure 1.5 shows a general MS/MS experiment. MS/MS consists of two consecutive mass analysis steps used to identify fragment ions

formed from a precursor ion (101, 102). In the first step, a precursor ion is isolated based on mass-to-charge ( $m/z$ ), while in the second step product ions formed from either spontaneous or induced fragmentation are analyzed based on  $m/z$  values (101). Generally fragmentation is induced through collision or electron transfer (103-105). Interpretation of product ions in the second step provides information regarding sequence, types of PTMs, and sites of attachment (106).

Prior to analysis by MS/MS, proteins are routinely subjected to digestion by selected proteases (101). The most commonly used protease, trypsin, cleaves proteins after the basic residues arginine and lysine (106). Due to the abundance of basic residues within proteins, trypsin generally produces small peptides, providing two ionization sites, one each at the N- and C-termini (101, 107). Histidines within peptides can provide extra ionization sites. In addition to this, missed tryptic cleavage sites, due to sequential basic residues, blockage by PTMs, or prolines after the basic residues which renders sites inaccessible to trypsin, allow for higher ionization states (106).

The two main ionization methods, electrospray ionization (ESI) and matrix assisted laser desorption/ionization (MALDI), are primarily used to generate molecular ions of peptides (102, 105, 106). ESI produces both single and multiple charged ions based on the number of basic residues within a peptide, while MALDI only produces singly charged species, charge that is passed from the matrix employed (102, 106). Higher charged species generally provide more fragmentation information, as more sites within a peptide are subject to dissociation (106).

During ESI, ionized peptides are formed by spraying a dilute solution of sample from a fine metal capillary at atmospheric pressure in a high electric field (101, 105). The peptide solvent evaporates rapidly and the peptides pick up protons based on possible sites of proton attachment (101). Usually ESI does not result in peptide fragmentation (101). ESI is easily coupled to chromatographic separation, thereby liquid chromatography (LC) is commonly used to separate complex mixtures before delivery for mass analysis (106). LC is generally performed with a mixture of highly pure water and volatile organic solvents, which is also ideal for ESI (101). Currently the preferred sample delivery method is online reverse phase nano-flow LC that provides a built in fractionation method that is high speed and requires low sample volume (106).

As components are eluted after separation, mass spectra are continuously recorded (101). Total ion traces, or the total number of ions recorded plotted against elution time, aid in identification of regions of data likely to contain useful spectra (101). Linear ion traps provide MS/MS capabilities for multiple steps in which fragment ions are isolated and further fragmented prior to a second stage of mass analysis (108). These mass analyzers allow for high ion trapping capacity and have a slow scanning function to provide high resolution (108).

### **1.3.3 Collision-Induced Dissociation and Electron Transfer Dissociation**

The most commonly used activation step in MS is collision-induced dissociation (CID), a low energy method to fragment peptide ions through many collisions with rare gas atoms (105). CID supplies sufficient energy to induce covalent bond cleavage, and within peptides the preferred cleavage sites are the backbone amide bonds (105, 107).

Upon dissociation a homologous series of y and b ions are generated, creating a collection of peptide ions that differ in mass by a single amino acid (Figure 1.6) (104, 105). Preferred target peptides for CID are short to medium length (5 – 20 amino acid) sequences (103).

The distribution of oscillation energy along the peptide backbone from CID occurs on a slow time scale (picoseconds), ultimately cleaving the weakest bonds (105). PTMs, such as phosphorylation, often attached through labile bonds, can cause peptides to fragment in a different manner, in that cleavage of the PTM becomes a preferred dissociation site (105, 107). Due to this, peptides with PTMs rarely give rise to a homologous series of y and b ions, and therefore do not give rise to sequence identifications (107). In addition to this, very little information about the PTM itself is acquired after CID activation (107).

In 2004, a new activation step termed electron transfer dissociation (ETD) was introduced that provided an alternate fragmentation method to CID that could allow for thorough analysis of PTMs (109, 110). ETD is based on ion-ion chemistry that fragments peptides via electron transfer from a radical anion to a protonated peptide (103, 111). ETD primarily cleaves the C $\alpha$ -N bonds of peptides producing c and z ions (Figure 1.6) (103, 107). Upon ETD activation, most PTMs (i.e. phosphorylation, glycosylation, and sulfonation) remain intact, while backbone fragmentation is observed. This allows for the identification of the site of attachment and mass of the PTM (105). CID causes the dissociation of many labile PTMs; however, disulfide linkages are not typically broken

by CID. On the other hand, ETD, in addition to dissociating C $\alpha$ -N bonds, preferentially cleaves disulfide bonds (111, 112).

CID and ETD cleave peptide backbones by distinct chemical mechanisms that give rise to y and b ions and c and z ions, respectively. CID is based on the transfer of kinetic energy from a neutral gas to the internal energy of a precursor peptide ion, that is distributed across the entire peptide as oscillation energy (103). This energy transfer occurs within a collision cell or ion trap containing a high pressure of the inert gas (commonly argon, helium, or nitrogen) (103). The amount of energy exchange typically dictates fragmentation. At lower energies, neutral losses such as water or methyl groups occur, while at higher energies uncontrolled fragmentation between carbon atoms can occur (107). To obtain optimal backbone amide bond cleavages collision energies at the fragmentation threshold are employed and typically, in a gas-phase peptide ion, the amide bonds are the preferred sites of cleavage (103).

In ETD, an amide carbonyl group that is hydrogen-bonded to a protonated basic amino acid side chain captures an electron (107). The most commonly used electron donor is fluoranthene (103). The resulting radical anion abstracts a proton and generates an amino-ketal radical site which is further triggered to dissociate into complementary pairs of c and z ions (107). Most often even electron c ions ( $c^+$ ) and odd electron z ions ( $z^{\dot{-}}$ ) are generated, however odd electron c ions and even electron z ions have been observed (107). Disulfide linkages have a higher radical affinity and tend to be the preferred cleavage site over peptide backbones (111). In the presence of disulfide bonds,

initial electron capture occurs at a carbonyl group, however the electron is subsequently transferred to the anti-bonding orbital of the disulfide bond, de-stabilizing this site (107).

The combination of ETD and CID allows for a powerful MS methodology for mapping disulfide bonds. ETD and CID can be employed in tandem, such that in an initial ETD step a disulfide linkage between peptides can be cleaved. In a subsequent CID step, the liberated peptides can be further dissociated and identified.

### **1.3.4 Disulfide Mapping by MS**

MS is an essential analytical tool for establishing disulfide linkages, which provide crucial structural information for proteins (113, 114). The task of identifying disulfide bonds in a protein can be quite challenging, because as the number of cysteines increases, the possible combination of disulfide linkages increases to an even greater extent (115). Strategies used to map disulfide bonds via MS have dramatically improved throughout the years (113, 116). A traditional MS approach relying on the comparison of reduced and non-reduced proteins will be discussed below, followed by a novel MS3 methodology combining ETD and CID.

Standard approaches for mapping disulfide linkages in proteins involve LC-MS analysis of peptides generated by enzymatic digestion before and after reduction (114). Optimization of many factors, including proteases, reducing agents, alkylating agents, and de-glycosylating enzymes contribute to successful mapping in these experiments (117). Generally, the first step in these experiments is to determine the number of free cysteines by alkylating non-reduced protein samples and establishing which cysteine

residues are modified (115). In certain instances disulfide-linked peptides may be glycosylated, therefore prior incubation with de-glycosylating enzymes is necessary for peptide identification (115). To determine the disulfide linkages within a protein, the protein must be digested while the disulfide bonds remain intact (115). Comparison of peptide map assignments for the reduced and non-reduced proteins generally allows for identification of cysteine residues involved in disulfide bonding (117). Major problems that arise from this strategy are the inability to make assignments when more than a single disulfide bond links two peptides or when three or more peptides are disulfide-linked together (114, 116). Therefore it is very difficult to map the disulfide bonds of proteins with an intricate network of linkages between numerous closely spaced cysteines.

Since the advent of ETD, problems encountered with traditional disulfide mapping are being resolved (103, 107). A major advantage of using ETD is that protein samples no longer need to be reduced prior to MS analysis, as disulfide bonds are dissociated during the ETD activation step (103, 107, 111), as noted in the previous section. Karger and colleagues have used online LC-MS with ETD and CID to map the disulfide bonds of several recombinant proteins, including an immunoglobulin light chain (116, 118), human growth hormone (116), and tissue plasminogen activator (116, 119). All of the proteins that were analyzed by this method had known disulfide maps, and their experiments provided proof-of-principle that this method was effective (116, 119). The MS3 protocol they employed consisted of a full scan MS, followed by a MS2 ETD step and a MS3 CID step (116, 118, 119). In essence the MS2 ETD cleaved disulfide-

linkages, while the final MS3 CID step fragmented released peptides, allowing for their identification (116, 118, 119).

Disulfide-linkages between highly interlinked peptides can now be assigned by use of MS3 combining ETD and CID (119). In the case of 3 peptides that are linked (i.e. Peptides A and C are both linked to Peptide B), a single ETD activation step can release either Peptide A or C, which can then be identified by CID fragmentation (116, 119). To assign the exact site of the disulfide-linkage, CID fragmentation of the peptide backbones of the remaining disulfide bonded pair of peptides can be followed (116, 119). The same holds true for two peptides that are linked together by two disulfide-linkages, in that an ETD step can dissociate one of the bonds and then the cysteines linked can be established through CID fragmentation data (119). Further advances in MS methodologies may allow for direct identification of released peptides as it is possible that two sequential ETD steps can cleave two disulfide bonds within highly intertwined peptides.

### **1.3.5 *Ab Initio* Modeling and Disulfide Mapping**

Computational modeling methods such as homology modeling, protein threading, and *ab initio* modeling provide critical insights into protein structures (120-123).

Homology modeling relies on three-dimensional template structures that share greater than 30% sequence identity with the amino acid sequence of the protein of interest (121).

When structures of homologous proteins are unavailable, protein threading (also known as fold recognition) can be employed (122). Threading relies on optimal fitting of amino acid sequences onto resolved protein structures in databases (122). *Ab initio* modeling

relies solely on the physio-chemical properties of the amino acids in the protein sequences (120, 123, 124).

Based on critical assessment of techniques for protein structure predictions (CASP) the most accurate and consistent *ab initio* modeling software is Rosetta (120, 123). Structure prediction in Rosetta is based on obtaining the lowest free energy states of protein conformations, which relies on fairly accurate energy functions (120). Several hallmark features of the predicted folded structures are necessary including the burial of nonpolar groups away from water, the close packing of buried groups, and the formation of intramolecular hydrogen bonds by buried polar atoms (120).

The Rosetta method and homology modeling share some of the same limitations (124, 125). The issue of false positives and negatives in Rosetta-based *ab initio* modeling is significant (125). The failure to generate a model by this method does not mean that one cannot be generated and it does not necessarily indicate that the structure is a novel one (125). Additionally, the generation of a model does not guarantee it is correct (124, 125). There are also limitations with regards to protein size (124). Most of the successful *de novo* prediction methods that are applicable to larger protein segments (up to ~150 residues) use information from known protein structures (124). Although the program can be modified, by default Rosetta employs an ensemble of short structural fragments extracted from the PDB (124).

In Chapters 3 and 4 of this dissertation we use *ab initio* modeling in combination with tandem MS with ETD and CID to predict complete structures of IGFBP-5 and RGMc based on identified disulfide linkages. To perform *ab initio* modeling for these

studies, segments were generated using the Rosetta fragment server. Five thousand independent simulations were generated and organized into five clusters. The centers of the five largest clusters were chosen as the best models, defined as having the lowest standard deviation of the mean among positions of  $\alpha$ -carbon atoms of all residues when compared with all other simulations in a cluster.

### **1.3.6 The Cysteine-Rich Proteins, IGFBP-5 and RGMc**

Two chapters of this dissertation are dedicated to establishing the disulfide-linkages of the secreted proteins IGFBP-5 and RGMc by MS3 combining ETD and CID. Limited structural data is available for either of these proteins. This brief sub-section of the introduction provides sequences for the cysteine-rich mouse proteins IGFBP-5 and RGMc (Figure 1.7) that can be used as a reference throughout this dissertation.

IGFBP-5 contains 18 evolutionarily conserved cysteines, 12 of them residing in the N-terminal domain (~ residues 1-85) and 6 in the C-terminal domain (~residues 170-252), which can give rise to 9 disulfide bonds (Figure 1.7A) (126, 127). Mouse and human IGFBP-5 are 97% identical (128). Full-length mouse RGMc50 contains 14 cysteines (Figure 1.7B) and is 88% identical to its human homologue, hemojuvelin (57).

## 1.4 DISSERTATION OVERVIEW

The principal theme of this dissertation is to enhance understanding of soluble RGMc isoforms through a structure-function approach. The primary focus of this dissertation will be (i) to elucidate the role of soluble RGMc isoforms in regulating BMP-mediated actions, (ii) to establish novel methodologies of disulfide mapping by tandem MS with ETD and CID, and (iii) to provide structural data for RGMc via disulfide mapping by MS in combination with *ab initio* modeling.

The previous sections of this introduction provide an overview of mammalian iron homeostasis, insight into the biosynthesis, function and structure of RGMc, and methodologies for exploring the tertiary structure of proteins, by disulfide mapping via tandem MS combining CID and ETD.

The aim of Chapter 2 is to elucidate the functional relationship between the soluble RGMc isoforms and two members of the BMP family, BMP2 and BMP6. Data in this chapter demonstrate that (i) soluble RGMc is a potent inhibitor of BMP2- and BMP6-mediated signaling and gene expression, (ii) BMP2 and BMP6 stimulate to a similar extent the same cohort of genes in a cultured liver cell line, and (iii) selected RGMc mutants that can bind BMPs retain the ability to inhibit BMP-mediated actions.

Chapter 3 presents novel tandem MS methodologies combining CID and ETD aimed at determining the disulfide maps of proteins, using IGFBP-5 as a model. The major findings of this chapter establish (i) five definitive disulfide linkages within IGFBP-5, (ii) that MS and *ab initio* modeling provide in combination a powerful

approach to mapping disulfide bonds, and (iii) that an N-terminal IGFBP-5 mutant that does not bind IGFs does not have an altered disulfide mapping pattern.

The primary focus of Chapter 4 is to provide the first insights into the structure of distinct RGMc isoforms via a bimodal approach that combines disulfide mapping of the proteins by tandem MS methods using CID and ETD with *ab initio* modeling. Data in this chapter demonstrate (i) that *de novo* protein disulfide mapping can be achieved via MS3 combining ETD and CID, (ii) that RGMc is a two domain protein with cysteine linkages that dictate the structure of each domain and (ii) that distinct RGMc isoforms possess unique disulfide bonding patterns.

The appendix presents additional results related to the dissertation. By analysis of disease-associated truncation mutants, the processing of RGMc by PCs is discussed. Chapter 5 summarizes the major findings in this dissertation and presents future areas of study that can be pursued. Collectively, this dissertation provides a detailed understanding of the structure and function of soluble RGMc.

**Table 1.1 HJV Mutations Linked to Juvenile Hemochromatosis**

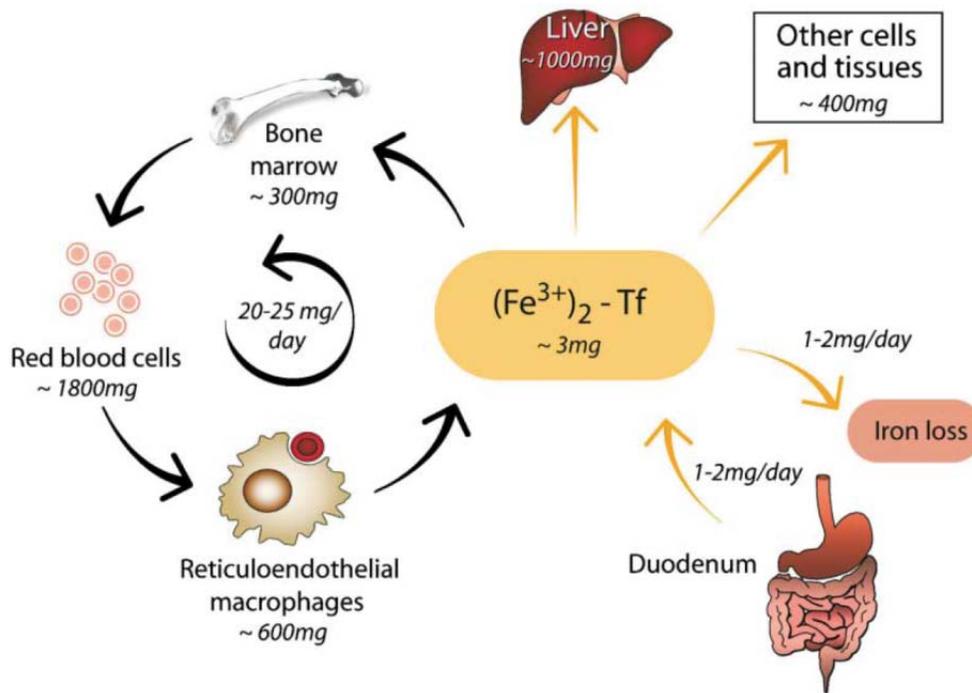
<b>Wild-Type</b>	<b>A.A.</b>	<b>Mutation</b>	<b>Possible Structure-Function Disruption</b>
Cys	80	Arg	Altered disulfide bonds
Gly	99	Val	Disrupts RGD motif
Gly	99	Arg	Disrupts RGD motif
Cys	119	Phe	Altered disulfide bonds
Asp	149	f.s.	Only soluble isoforms synthesized?
Asp	172	Glu	Interrupts internal cleavage site
Arg	176	Cys	Altered disulfide bonds
Trp	191	Cys	Altered disulfide bonds
Gln	312	f.s.	Only soluble isoforms synthesized?
Gly	320	Val	unknown, most prevalent mutation
Cys	321	f.s.	Only soluble isoforms synthesized?
Cys	321	Trp	Altered disulfide bonds
Ser	328	f.s.	Only soluble isoforms synthesized?
Cys	361	f.s.	Only soluble isoforms synthesized?
Arg	385	f.s.	Only soluble isoforms synthesized?

f.s. denotes frameshift mutation; A.A. is amino acid

Mutations from references (59, 129-134)

**Figure 1.1 Systemic iron homeostasis.** Major pathways of iron flow in the human body. Approximate daily amounts of iron flux and normal values of iron in human tissues and organs are displayed. Of note, an average of 1 – 2 mg of iron is absorbed from the diet, while the same amount is lost by un-regulated events such as blood loss and sloughing of intestinal and skin cells. Figure taken from Hentze et al. (135).

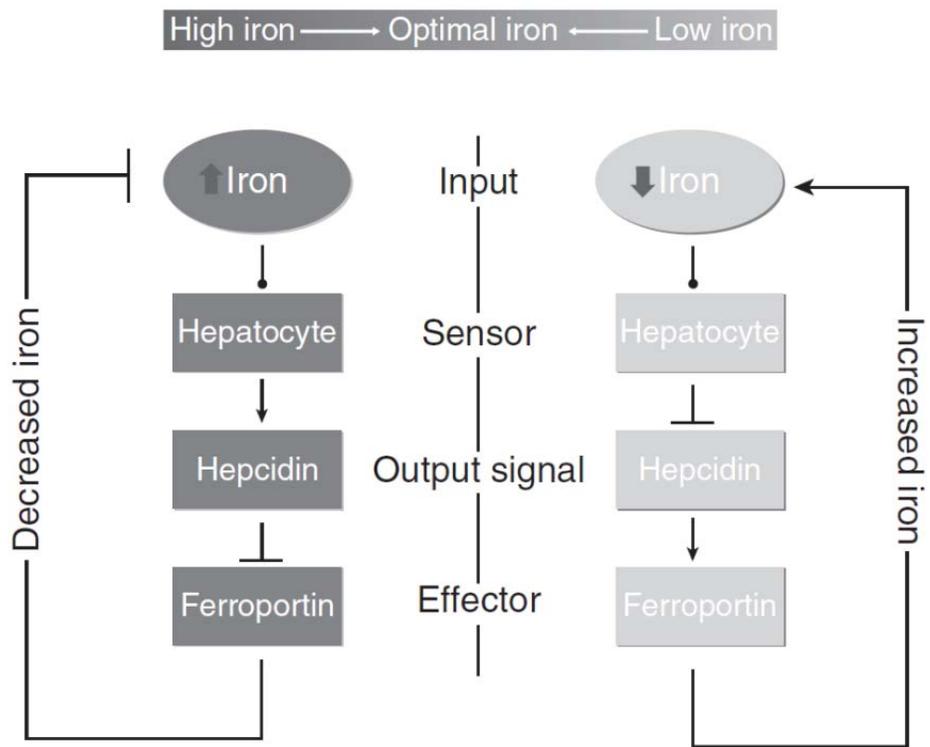
Figure 1.1



**Figure 1.2 Normal iron homeostasis mediated by an iron-sensing feedback loop.**

Iron levels in circulation direct the amount of iron released from the primary storage sites, the hepatocytes, macrophages, and duodenal enterocytes. When iron levels are high, sensors on the hepatocytes up-regulate hepcidin gene expression through identified and unknown signaling mechanisms. Binding of hepcidin to ferroportin causes its internalization and degradation, resulting in lower circulating levels of iron. When iron levels are low, hepcidin expression is decreased, and thus iron is exported through the iron exporter ferroportin. This feedback mechanism allows for the adjustment of circulating iron levels to meet the demands of the body. Figure taken from Wrighting and Andrews (5).

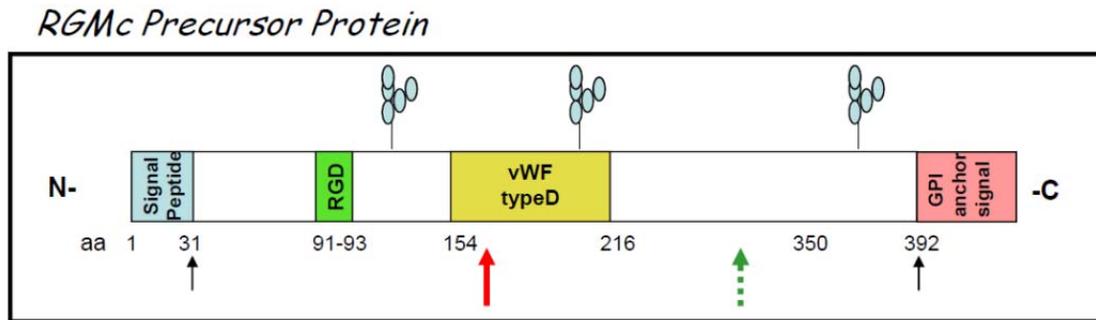
Figure 1.2



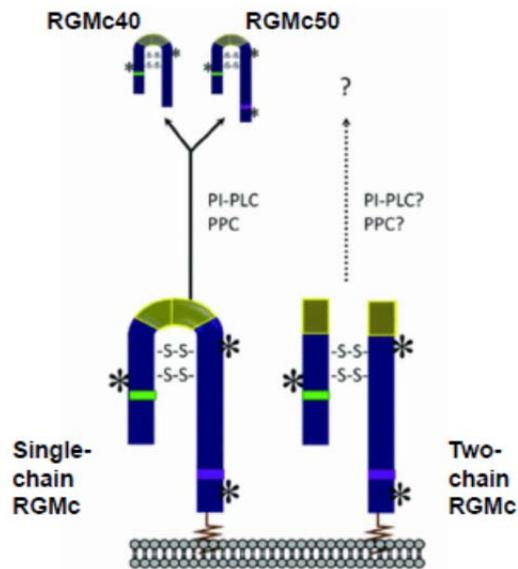
**Figure 1.3 Domain map and isoforms of RGMc.** **A.** The primary RGMc translation product is 420 amino acids and contains a N-terminal signal sequence and C-terminal GPI anchor sequence. RGMc contains several recognizable features: a RGD domain (a.a. 91-93), a partial vWF type D domain (a.a. 154-216), and three N-linked glycosylation motifs (NXS/T). Red arrow indicates internal cleavage site. Green broken arrow points to PC-cleavage site; black arrows indicate sites at which RGMc signal sequences are cleaved. **B.** RGMc exists as four distinct isoforms. Two-chain and single-chain species are linked to the cell surface via GPI anchors. Soluble full-length (RGMc50) and PC-cleaved (RGMc40) isoforms arise from single-chain cell-associated RGMc. Figure **B.** taken from Severyn et al. (51).

Figure 1.3

A.



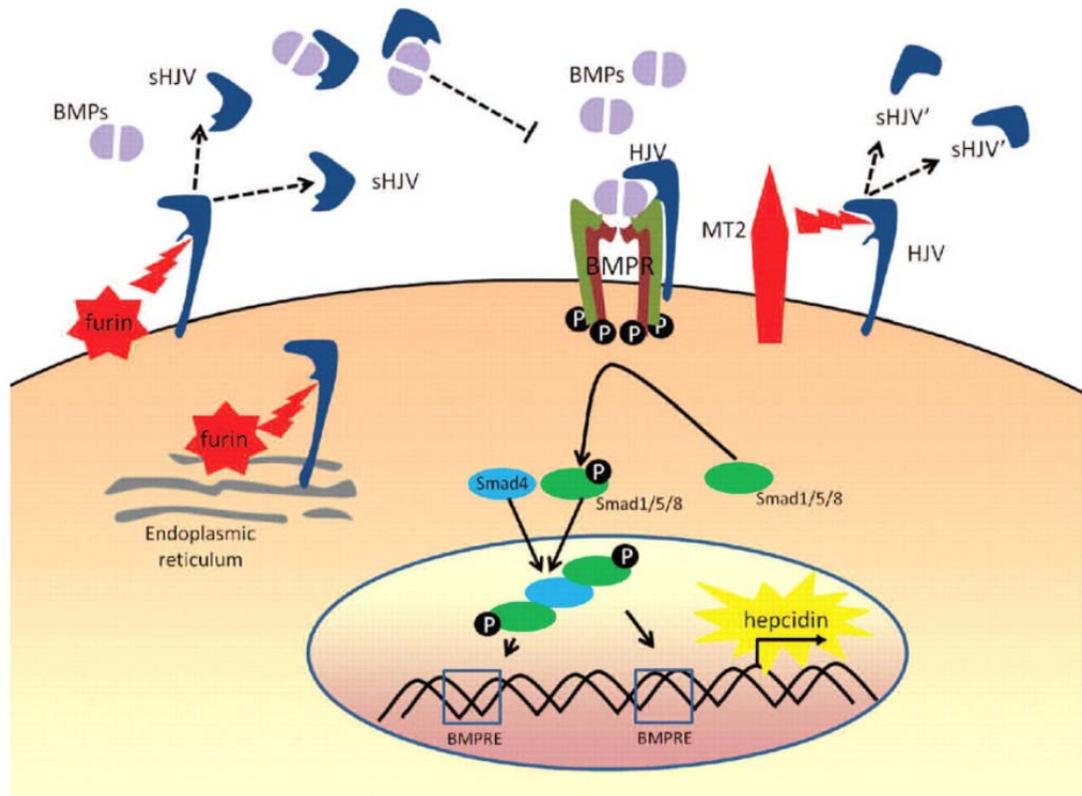
B.



**Figure 1.4 Opposing actions of cell-associated and soluble RGMc/HJV in BMP**

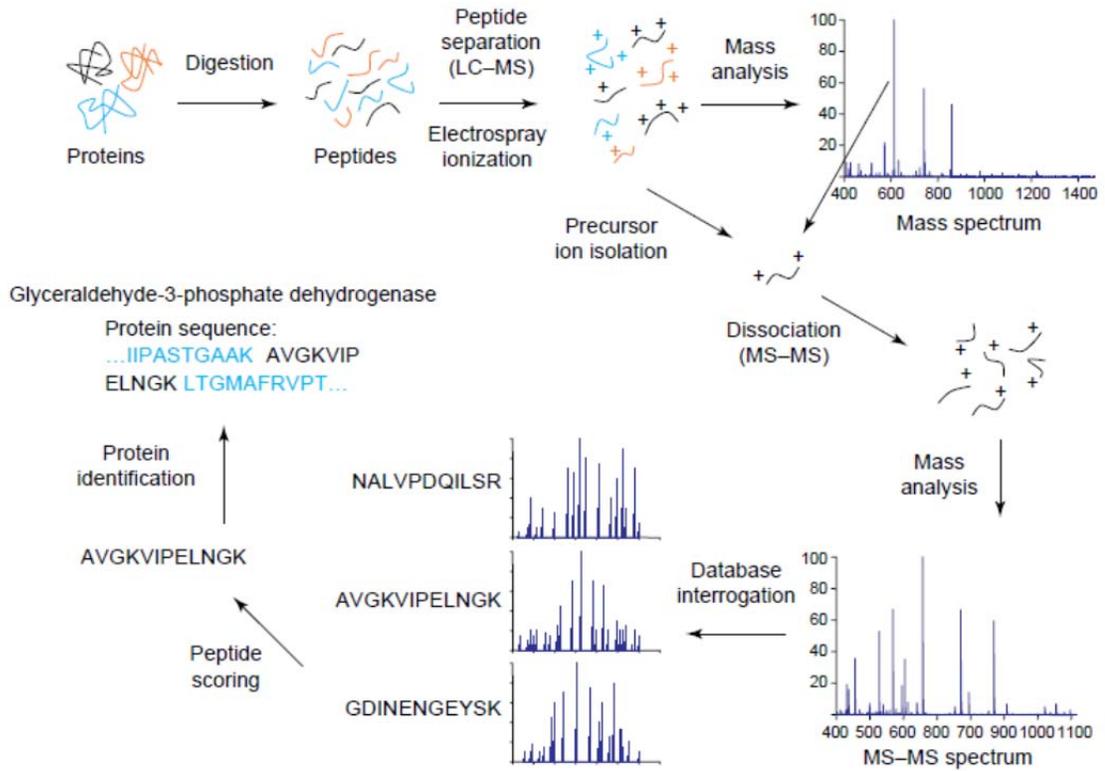
**regulation.** The prevailing view of RGMc/HJV as a BMP co-receptor. Cell-associated RGMc/HJV binds to BMPs and up-regulates Smad phosphorylation and BMP-responsive genes, including hepcidin. RGMc/HJV on the cell surface can be cleaved to soluble isoforms by PCs, such as furin, or Matriptase-2 (MT2, also known as TMPRRS6) that can bind and sequester BMPs, thus functioning as BMP inhibitors. Figure taken from Kautz and Nemeth (85).

Figure 1.4



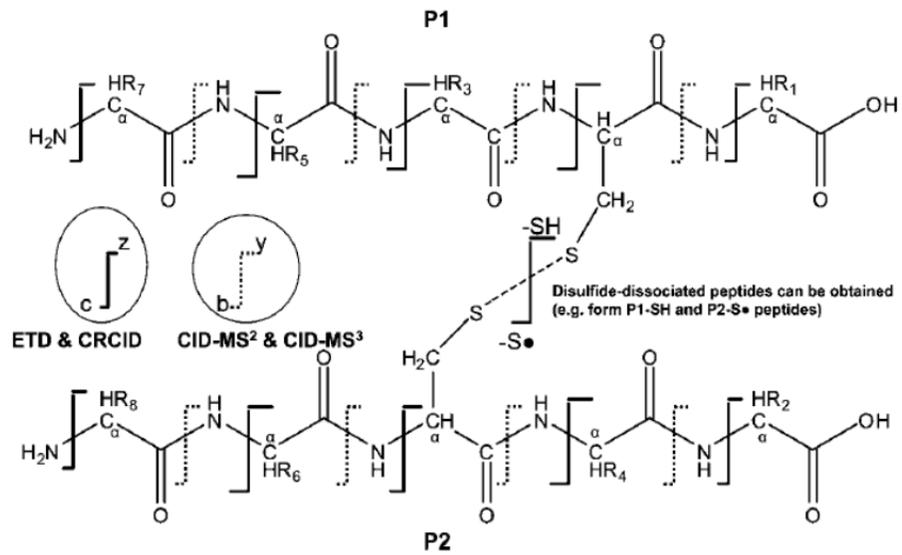
**Figure 1.5 Generalized tandem MS protocol flowchart.** Purified proteins or complex mixtures of proteins are digested with proteases and analyzed by LC-MS. In the first MS scan peptide masses are determined. Individual precursor peptides are selected, isolated, and subjected to a second analysis scan that results in fragment ions. The combination of the original peptide mass and fragment ions allow for peptide identification through database searches, which then allows for protein identifications. Figure taken from Kolker et al. (106)

**Figure 1.5**



**Figure 1.6 The activation steps CID and ETD.** CID and ETD induce dissociation preferentially at different sites within a disulfide-linked peptide. In the diagram two peptides, P1 and P2, are linked by a disulfide bond. CID at either the MS2 (CID-MS2) or at the MS3 (CID-MS3) analysis step induces cleavage of the peptide backbone amides to generate y and b ions. Fragmentation by ETD results in backbone C $\alpha$ -N cleavages, giving rise to c and z ions or more preferentially dissociation of the disulfide-linkage. Figure taken from Wu et al. (116)

Figure 1.6



**Figure 1.7 Sequences of the cysteine-rich proteins, IGFBP-5 and RGMc. A.**

Sequence of mouse IGFBP-5 used for disulfide mapping studies in Chapter 3 of this dissertation. The leucine (L) at position 1 is the first residue of the mature protein. The 18 cysteines are underlined. **B.** Sequence of RGMc used for disulfide mapping studies in Chapter 4 of this dissertation. The first residue of the sequence displayed, glutamine (Q) at position 33, is the predicted start site of mature RGMc. The 14 cysteines are underlined.

Figure 1.7

A. IGFBP-5

LGSFVHCEPC DEKALSMCPP SPLGCELVKE PGCGCCMTCA LAEGQSCGVY 50  
TERCAQGLRC LPRQDEEKPL HALLHGRGVC LNEKSYGEQT KIERDSREHE 100  
EPTTSEMAEE TYSPKVFRPK HTRISELKAE AVKKDRRKKL TQSKFVGGAE 150  
NTAHPRVIPA PEMRQESEQG PCRRHMEASL QEFKASPRMV PRAVYLPNCD 200  
RKGFYKRKQC KPSRGRKRG I CWCVDKYGMK LPGMEYVDGD FQCHAFDSSN 250  
VE

B. RGMc

QCKILRCNAE YVSTLRRLRG GGSPTPRGG GRGGLASGGL CRALRSYALC 82  
TRRTARTCRG DLAFHSAVHG IEDLMIQHNC SRQGPTAPPP ARGPALPGAG 132  
PAPLTPDPCD YEARFSRLHG RAPGFLHCAS FGDPHVRSFH NQFHTCRVQG 182  
AWPLLDNDFL FVQATSSPVS SGANATTIRK ITIIFKNMQE CIDQKVYQAE 232  
VDNLPAAFED GSINGGDRPG GSSLSIQTAN LGSHVEIRAA YIGTTIIIRQ 282  
TAGQLSFSIR VAEDVARAFS AEQDLQLCVG GCPPSQRLSR SERNRRGAIA 332  
IDTARRLCKE GLPVEDAYFQ SCVFDVSVSG DPNFTVAAQT ALDDARVFLT 382  
DLENLHLFPS DA

## **Chapter 2**

### **Soluble RGMc/Hemojuvelin is a Broad Spectrum BMP Antagonist and Inhibits both BMP2 and BMP6-mediated Signaling and Gene Expression**

Mahta Nili, Ujwal Shinde, and Peter Rotwein

Department of Biochemistry and Molecular Biology, Oregon Health & Science  
University, Portland, OR 97239-3098, USA

All experiments and data analysis reported in this chapter were performed by the author of this dissertation. This work has been previously published (Nili, M., Shinde U., Rotwein, P. (2010) *J. Biological Chemistry* 285: 24783-92; (136)) and presented (poster at the 92<sup>nd</sup> Annual Endocrine Society Meeting, San Diego, CA, 2010).

## 2.1 SUMMARY

Inactivating mutations in hemojuvelin/repulsive guidance molecule c (HJV/RGMc) cause juvenile hemochromatosis (JH), a rapidly progressive iron overload disorder in which expression of hepcidin, a key liver-derived iron-regulatory hormone, is severely diminished. Several growth factors in the bone morphogenetic protein (BMP) family, including BMP2 and BMP6, can stimulate production of hepcidin, a biological effect that may be modified by RGMc. Here we demonstrate that soluble RGMc proteins are potent BMP inhibitors. We find that 50 and 40 kDa RGMc isoforms, when added to cells as highly purified IgG Fc fusion proteins, are able to block the acute effects of both BMP2 and BMP6 at the levels of Smad induction and gene activation, and thus represent a potentially unique class of broad-spectrum BMP antagonists. Whole transcript microarray analysis revealed that BMP2 and BMP6 each stimulated expression of a nearly identical cohort of ~40 mRNAs in Hep3B cells and demonstrated that 40 kDa RGMc was an effective inhibitor of both growth factors, although its potency was less than that of the known BMP2-selective antagonist, Noggin. We additionally show that JH-linked RGMc mutant proteins that retain the ability to bind BMPs also are able to function as BMP inhibitors, and like wild type soluble RGMc species, can block BMP-activated hepcidin gene expression. The latter results raise the question of whether disease severity in JH will vary depending on the ability of a given mutant RGMc protein to interact with BMPs.

## 2.2 INTRODUCTION

Iron is an essential co-factor for many cellular processes, and plays a vital role in regulating respiration, energy metabolism, and oxygen transport (1, 135). Iron homeostasis is tightly controlled, and detrimental consequences arise from both its deficiency and excess (1, 135). In hemochromatosis, chronic accumulation of excess iron in multiple tissues leads to organ damage and dysfunction (1, 135, 137). Juvenile hemochromatosis (JH) is a rare and rapidly progressive form of this disorder, and has been linked to alterations in the HJV/HFE2 gene, which encodes for hemojuvelin (HJV) (59, 137). Over two-dozen HJV gene mutations have been identified in patients with JH, and include single nucleotide changes that lead to amino acid substitutions and DNA frame-shifts that could give rise to truncated proteins (59, 137). In all patients with JH, levels of the key liver-derived iron-regulatory hormone, hepcidin, are reduced, leading to inappropriately elevated iron absorption from the duodenum (137). A similar pattern of excessive iron uptake has been seen in mice engineered to lack HJV (87, 88).

HJV is identical to RGMc, which with RGMa and RGMb comprise the repulsive guidance molecule (RGM) family (59). RGMc is produced in the liver and striated muscle (54, 57, 59) while RGMa and RGMb are primarily synthesized in the central nervous system (54), where they are involved in regulating neuronal survival and axonal patterning during development (55, 56). RGM family members are glycoproteins that share several structural motifs (54, 59), and all three can undergo a series of similar biosynthetic and processing steps leading to cell-associated GPI-linked and soluble protein species (35, 36, 58). RGMc is composed of two cell membrane associated forms,

a ~50 kDa single-chain species and a two-chain isoform with ~30 and ~20 kDa disulfide-bonded subunits, and two soluble single-chain members of ~50 and ~40 kDa (58, 73). To date, the biochemical mechanisms responsible for production of multiple RGMc protein species have not been elucidated fully.

The precise biological roles for each RGMc species in systemic iron balance also have not been defined yet. Recent studies have hypothesized that cell-associated RGMc functions as a co-receptor for selected bone morphogenetic proteins (BMPs), and can facilitate the ability of BMPs to stimulate hepcidin gene expression in the liver (34), although the evidence supporting this co-receptor concept is incomplete. BMPs are members of the transforming growth factor- $\beta$  (TGF- $\beta$ ) super-family, and play crucial roles in a variety of developmental and cell fate decisions (31, 32). BMPs bind as dimers to specific type I and type II serine/threonine kinase receptors, and initiate a protein kinase cascade which culminates in the activation by serine phosphorylation of Smads 1, 5, and 8, signal transducers that regulate the expression of many BMP-dependent target genes (31, 32). Several sub-classes of BMPs exist that bind to distinct receptors (138, 139). For example, BMP2 and BMP6 are 61% identical in amino acid sequence and belong to the BMP2/4 and BMP5/6/7/8 sub-classes, respectively (140). BMP2 preferentially binds to the type II receptor BMPRIIA, and the type I receptor ALK3, while BMP6 interacts with ACVRIIa and ALK2 (141). BMP2 also binds with high affinity to the soluble inhibitor, Noggin (142, 143), while BMP6 binds poorly (144). Despite these differences, all BMPs promote receptor-mediated activation of Smads 1, 5, and 8, and thus may control the expression of similar groups of genes.

Here we demonstrate that soluble RGMc proteins are potent BMP inhibitors. We find that 50 and 40 kDa RGMc isoforms are able to block the acute effects of both BMP2 and BMP6 at the level of Smad activation, and thus represent a potentially unique class of broad-spectrum BMP modifiers. We additionally show that JH-linked RGMc mutant proteins that retain the ability to bind BMPs also are able to function as BMP inhibitors, and like wild type soluble RGMc species, can block BMP-activated hepcidin gene expression.

## **2.3 MATERIALS and METHODS**

### **2.3.1 Materials**

Fetal bovine serum (FBS), Ultralow IgG FBS, Dulbecco's modified Eagle's medium (DMEM), Modified Eagle's Medium (MEM), insulin-transferrin-selenium, phosphate-buffered saline (PBS), trypsin/EDTA, AcTEV Protease, Protein A Sepharose 4B, TRIzol Reagent, and Superscript III first-strand cDNA synthesis kit were purchased from Invitrogen (Carlsbad, CA). N-Glycosidase F (PNGaseF) was from New England Biolabs (Beverly, MA). Ham's F12 medium was from Thermo Scientific Hyclone (Waltham, MA). Recombinant mouse Noggin-Fc, and recombinant human BMP2 and BMP6 were purchased from R&D Systems (Minneapolis, MN). NiSepharose 6 Fast Flow was from GE Healthcare (Waukesha, WI). The furin convertase inhibitor decanoyl-RVKR-CMK and okadaic acid were purchased from Alexis Biochemicals (San Diego, CA); protease inhibitor tablets were from Roche Applied Sciences (Indianapolis, IN). Restriction enzymes, buffers, ligases, and polymerases were purchased from Roche Applied Sciences, BD Biosciences-Clontech (Palo Alto, CA), and Fermentas (Hanover, MD). Sodium orthovanadate and dexamethasone were from Sigma-Aldrich (St. Louis, MO). AquaBlock EIA/WIB solution was from East Coast Biologicals (North Berwick, ME). The BCA protein assay kit and GelCode Blue Stain Reagent were purchased from Pierce Biotechnologies (Rockford, IL). TransIT-LT1 was from Mirus Bio (Madison, WI). NitroBind nitrocellulose was from GE Water & Process Technologies (Trevose, PA). Polyclonal anti- $\alpha$ -tubulin antibody was purchased from Sigma Aldrich (St. Louis, MO). Monoclonal anti-Smad5 and anti-phospho-Smad5 antibodies were from Abcam

(Cambridge, MA). Secondary antibodies, AlexaFluor 680-conjugated goat-anti-rabbit IgG and AlexaFluor 680-conjugated goat-anti-mouse IgG, were purchased from Invitrogen. IR800-conjugated goat-anti-mouse IgG and IR800-conjugated goat-anti-human IgG were from Rockland Immunochemical (Gilbertsville, PA). Other chemicals and reagents were purchased from commercial suppliers.

### **2.3.2 Cell Culture and Transient Transfections**

The following cell lines were purchased from American Type Culture Collection: HEK293 (CRL-1573), Hep3B (HB-8064), and AML12 (CRL-2254). HEK293 cells were used for expression of all Fc and RGMc-IgG1-Fc fusion proteins. Hep3B and AML12 cells, human and mouse liver cells, respectively, were used to assess the effects of RGMc proteins on BMP-mediated signaling and gene expression. All cells were grown at 37°C in humidified air and 5% CO<sub>2</sub>. HEK293 cells were maintained in DMEM plus 10% FCS and transfected at ~70% confluent density with TransIT-LT1. AML12 cells were maintained in a 1:1 mixture of DMEM and Ham's F12 medium with 10% FCS, 5 µg/ml insulin, 5 µg/ml transferrin, 5 ng/ml selenium, and 40 ng/ml dexamethasone. Hep3B cells were maintained in MEM plus 10% FCS.

### **2.3.3 Generation and Purification of RGMc-IgG1-Fc Fusion Proteins**

For generation of full-length RGMc-Fc fusion proteins, codons 1-391 of mouse RGMc were sub-cloned via 5' EcoRI and 3' NotI sites into pcDNA3 (Invitrogen) in frame with the Fc region of human IgG1 (from pFuse-hFc1, Invivogen; San Diego, CA). Codons comprising a TEV protease recognition site (glutamate-asparagine-lysine-

tyrosine-phenylalanine-glutamine), plus three additional glycine residues, were added to the fusion plasmid between RGMc and IgG1-Fc segments, just 5' to a NotI restriction site. RGMc codon substitutions G92V and G313V, described previously (68), were sub-cloned into the full-length RGMc-IgG1-Fc backbone. RGMc truncation mutants were generated by PCR to replace codons after Q318 and C141 with the TEV recognition site and IgG1-Fc. DNA sequencing was used to confirm all nucleotide changes. For generation of IgG1-Fc, codons 1-120 of neogenin (comprising the signal peptide) were sub-cloned via 5' HindIII and 3' NotI sites into pcDNA3 in frame with the Fc region of IgG1. All RGMc-IgG1-Fc fusion proteins and IgG1-Fc were purified from conditioned medium of HEK293 cells that were transfected at ~70% confluent density with 10 µg of DNA per 100-mm diameter culture dish. At 5 hr post-transfection medium was changed to DMEM with 2% ultralow IgG FBS, and conditioned medium was collected 48 hr later. To minimize proteolytic cleavage of RGMc-IgG1-Fc fusion proteins, the furin convertase inhibitor decanoyl-RVKR-CMK was added to collection medium at a final concentration of 5 µM (68). Batch protein purifications were performed by IgG affinity chromatography. Conditioned medium was incubated with protein A Sepharose 4B for 16 hr at 4°C. Protein A beads were pelleted and washed three times with PBS containing 0.1% Tween. Bound RGMc-IgG1-Fc or IgG1-Fc proteins were eluted with 100 mM glycine, 1mM EDTA, pH 3.0 followed immediately by neutralization with 1M Tris-Cl, pH 8.0 to achieve a final Tris concentration of 100 mM. Purification was assessed after SDS-PAGE by staining of protein bands with GelCode Blue. Protein concentrations

were estimated by comparison with bovine serum albumin standards ranging from 100 ng to 1 µg electrophoresed in adjacent lanes.

#### **2.3.4 Purification of 40 kDa RGMc (RGMc40)**

RGMc40-Fc purified from conditioned medium was treated with TEV protease (5 U/10 µg RGMc40-Fc) for 3 hr at 20°C in buffer (50 mM Tris-Cl, 500 nM EDTA, pH 8.0) supplied by manufacturer. TEV protease was removed from the sample by batch affinity chromatography using the polyhistidine tag at the N-terminus of the protein. Samples were incubated with NiSepharose for 2 hr at 20°C followed by centrifugation to pellet the beads. Samples were then incubated with Protein A Sepharose for 2 hr at 20°C to remove cleaved off IgG1-Fc or un-cleaved RGMc-IgG1-Fc fusion proteins. Purification was assessed and protein concentrations were estimated as described in section 2.3.3 above.

#### **2.3.4 Analysis of N-linked Glycosylation**

To detect the presence of N-linked sugars, purified RGMc-IgG1-Fc fusion proteins, IgG1-Fc and RGMc40 were incubated with PNGaseF (500 U) for 16 hr at 37°C according to supplier's instructions, and products were resolved after SDS-PAGE by staining of proteins with GelCode Blue, and by immunoblotting as described in section 2.3.7 below.

### **2.3.5 Circular Dichroism Spectroscopy**

All proteins analyzed by CD spectroscopy were dialyzed with PBS containing 1mM EDTA for 16 hours at 4°C to remove glycine from elution buffer. To assess secondary structure of RGMc40 CD spectroscopy was performed on an AVIV model 215 CD spectrometer maintained at 4°C. For each protein 3 to 5 spectra were measured from 190 to 260 nm at 0.5-nm intervals for 3 sec at each wavelength, using a protein concentration of 0.2 mg/ml and a path length of 0.1 cm, and results were averaged. To establish a minimal baseline starting value before and after each sample PBS containing 1mM EDTA was measured. To determine the secondary structure of RGMc-IgG1-Fc proteins, data was collected as for RGMc40; however, an averaged spectrum for IgG1-Fc alone was subtracted from that of the RGMc-IgG1-Fc protein.

### **2.3.6 Analysis of BMP-mediated Signaling.**

Confluent Hep3B or AML12 cells were serum starved in MEM for 16 hr, followed by addition of BMP2 or BMP6 (0 - 200 ng/ml) for up to 8 hr in the absence or presence of various concentrations of purified IgG1-Fc, Noggin-Fc, RGMc40, or RGMc-Fc fusion proteins. Whole cell protein extracts and RNA were isolated and analyzed as described below in sections 2.3.7 and 2.3.8, respectively.

### **2.3.7 Protein Extraction and Immunoblotting**

Whole cell protein lysates were prepared by washing the cells with PBS, scraping them into 200 µl RIPA buffer (150 mM NaCl, 50 mM Tris-Cl, 1% NP-40, 0.5% NaDOC, 0.1% SDS, pH 7.5) and drawing them through a 21 gauge needle 5 times such

that foaming was minimized. Samples were centrifuged at 14K RPM for 10 minutes at 4°C to remove insoluble material. Protein concentrations of lysates were determined using the Pierce BCA protein assay kit, and aliquots were stored at -80°C until use. Protein samples (10-30 µg/lane) were resolved by SDS-PAGE and transferred to nitrocellulose membranes. After blocking with 50% AquaBlock solution for 1 hr at 20°C, membranes were incubated sequentially with primary and secondary antibodies. Primary antibodies were used at the following dilutions: anti-phospho-Smad5, 1:1000; anti-Smad5, 1:500; anti- $\alpha$ -tubulin, 1:30,000. Secondary antibodies were used at 1:10,000. Results were visualized and images captured using the LiCOR Odyssey and version 3.0 analysis software.

### **2.3.8 RNA Isolation and Analysis**

Total cellular RNA was extracted by suspending them in 1 mL of TRIzol reagent and following instructions supplied by manufacturer. RNA concentrations were determined spectrophotometrically [absorbance at a ratio of 260 to 280 nm ( $A_{260/280}$ )  $\geq$  1.9], and RNA quality was assessed by agarose gel electrophoresis. RNA (2 µg) was reverse transcribed with Superscript III first-strand cDNA synthesis kit using oligo (dT) primers in a final volume of 25 µL. PCR reactions were performed with 0.5 µL of cDNA according to a protocol supplied with Advantage2 GC kit (Clontech), using primer pairs listed in Table 2.1. PCR assays were performed within the linear range of cycle numbers (18-35) for each primer pair. PCR products were separated on 1.2% agarose gels and images were captured and quantified with a GelDoc imager and Quantity One® Software (BioRad; Hercules, CA).

### **2.3.9 Microarray Analysis of Gene Expression**

Hep3B cells were treated with 100 ng/ml BMP2 or BMP6 alone or in combination with a 10-fold molar excess of Noggin-Fc or a 20-fold molar excess of RGMc40-Fc for 4 hr. Un-treated Hep3B cells were used as controls. Total cellular RNA was extracted with TRIzol, followed by an additional sodium acetate-ethanol precipitation. Preparation of cDNAs, labeling, hybridization, quality control, and data acquisition were performed at the OHSU Gene Microarray Shared Resource using the Affymetrix GeneChip® Human Gene 1.0 ST Array. Two biological replicates were obtained for each treatment group. Results were normalized using the robust multichip average algorithm and analyzed with web-based GeneSifter® software. Transcripts from  $\log_2$ -transformed data were filtered according to the following criteria: for pairwise comparisons (t-test,  $p < 0.05$ ); and for multi-group comparison analysis (one-way ANOVA,  $p < 0.02$ ). For all comparisons the Benjamini and Hochberg correction was applied. The complete microarray data set (series record GSE20671) may be found at the National Center for Biotechnology Information Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>).

## **2.4 RESULTS**

### **2.4.1 Rational of Experimental Approach**

Several studies suggest that soluble RGMc is an inhibitor of BMP-mediated signaling and gene expression while cell-associated RGMc acts as a BMP co-receptor that up-regulates BMP-mediated processes (20, 27, 34, 86). Experiments to test these hypotheses were based on long duration luciferase reporter assays of BMP response elements (27, 34, 86) which can have confounding effects due to cross-talk from other signaling pathways. Here we chose to focus directly on the actions of RGMc on BMP signaling mediators, Smads, and known BMP gene targets. We also wanted to look at very early events in signaling such that effects from other signaling pathways were negligible. We have previously shown that RGMc exists as two isoforms in serum from humans and mice (74), and in several cell-culture model systems using various gene delivery methods (58, 68). In our studies here we set out to examine the effects of both of these soluble RGMc isoforms on BMP-mediated processes. Several BMP inhibitors only antagonize the effects of distinct classes of BMPs. Here we tested the effects of RGMc on both BMP2 and BMP6, members of different BMP sub-classes, to assess whether RGMc represents a new member of a more broad-based class of BMP inhibitor. Finally, from previous studies we it was demonstrated that selected RGMc disease-associated mutants displayed differential binding to BMPs (68). In our studies presented here, we further assessed several mutant RGMc species to determine if they are able to antagonize BMP-mediated gene expression.

#### **2.4.2 Hep3B Liver Cells are Robustly Responsive to BMP2 and BMP6**

Several BMPs have been linked to regulation of systemic iron metabolism through effects on hepcidin gene expression in the liver (27, 34, 48, 50), and it has been proposed that RGMc modifies these actions of BMPs (20, 27, 34, 86), although the biochemical and molecular mechanisms have not been elucidated. Incubation of either BMP2 or BMP6 with confluent Hep3B cells led to rapid stimulation of intracellular signaling via BMP receptors, as seen by inducible serine phosphorylation of Smad5 within 15 min of growth factor addition that was sustained for up to 240 min (Figure 2.1). Both BMPs also rapidly induced hepcidin gene expression, with mRNA being detectable within 1 hr after BMP2 treatment, and within 2 hr after addition of BMP6 (Figure 2.2 A). In contrast, only BMP2 could promote hepcidin mRNA accumulation in the mouse AML12 liver cell line (Figure 2.3), making this not a useful model for examining the actions of different BMPs. Dose-response curves showed that BMP2 was slightly more effective than BMP6 in inducing gene expression in Hep3B cells, as indicated by ~2-fold greater increases in hepcidin, Smad7, and Id1 mRNAs at any growth factor dose up to 200 ng/ml (Figure 2.2 B). Taken together, these results show that both BMP2 and BMP6 rapidly activate Smads and stimulate gene expression in Hep3B cells, and indicate that this cell line would be a good model for examining the effects of different RGMc protein species on BMP-mediated signaling and biological effects.

#### **2.4.3 RGMc Expressed and Purified as an IgG1-Fc Fusion Protein**

RGMc undergoes a complex series of biochemical and processing steps leading to membrane-bound and extracellular forms of the protein, including ~50 and ~40 kDa

soluble single-chain species (58, 73-75) here termed RGMc50 and RGMc40. We expressed both of these isoforms as IgG1-Fc fusion proteins (see Figure 2.4 A for domain maps) and purified them from HEK293 cell conditioned culture medium by protein A affinity chromatography (Figure 2.4 B). RGMc40-Fc and IgG1-Fc are predominantly single species of ~73 and ~33 kDa respectively on reducing SDS PAGE, as evidenced by staining of the purified proteins with Coomassie blue, and by detection with antibodies to human IgG and mouse RGMc (Figure 2.4 B). The expected full-length RGMc50-Fc fusion protein of ~83 kDa is seen as the major purified band, but two smaller and less abundant immunoreactive species of ~68 and ~45 kDa also are detected, and represent cleavage products of the full-length protein.

#### **2.4.4 Biochemical Characterization of RGMc-IgG1-Fc Fusion Proteins**

We performed a series of analytical experiments to evaluate each purified soluble RGMc species. Since IgG proteins naturally form dimers via disulfide linkages through their Fc region (145) we showed that RGMc50-Fc, RGMc40-Fc, and IgG1-Fc exist as dimers by demonstrating that their relative electrophoretic mobility was diminished on non-reducing compared with reducing SDS-PAGE gels (Figure 2.5 A). RGMc50, RGMc40, and IgG1-Fc each are predicted to contain N-linked sugars (Figure 2.4 A), and incubation with PNGaseF resulted in enhanced mobility on reducing SDS-PAGE (Figure 2.5 B).

#### **2.4.5 Purification and Analysis of RGMc40**

By nature of the design of the RGMc-IgG1-Fc fusion proteins, RGMc can be cleaved from the IgG1-Fc domain with TEV protease (Figure 2.4A). After removal of

the IgG1-Fc domain by incubation of RGMc40-Fc with TEV protease (Figure 2.6 A), the resultant purified recombinant RGMc40 was still a substrate for PNGaseF (Figure 2.6 B), indicating that both the RGMc and IgG1-Fc moieties are N-linked glycoproteins (Figures 2.5B and 2.6B). We analyzed secondary structural characteristics of RGMc40-Fc and RGMc40 by CD spectroscopy, and found that both proteins were very similar to each other in their  $\alpha$ -helical and  $\beta$ -sheet content (Figure 2.6C). Our analysis of RGMc50-Fc and RGMc40-Fc demonstrate that these proteins maintain the characteristics of native RGMc50 and RGMc40 species, and indicate that these recombinant soluble RGMc isoforms will be useful tools to probe the functional consequences of interactions with BMPs.

#### **2.4.6 Soluble RGMc Inhibits BMP2 and BMP6-mediated Signaling**

We assessed the effects of RGMc fusion proteins on BMP receptor activation in Hep3B cells by monitoring changes in BMP-stimulated phosphorylation of Smad5. Both RGMc50-Fc and RGMc40-Fc caused a dose-dependent decline in the extent of BMP-mediated Smad5 phosphorylation, with a consistently greater inhibitory effect seen on BMP6 than on BMP2 (Figure 2.7 A, B). BMP2 stimulated Smad5 phosphorylation more rapidly than BMP6, and at these earlier time points RGMc40-Fc also effectively inhibited these actions of BMP2 (Figure 2.8). Noggin is a well-described inhibitor of BMP2 but is not very effective against BMP6 (142, 143), and Noggin-Fc was more potent than RGMc50-Fc or RGMc40-Fc in blocking the effects of BMP2, as evidenced by a complete inhibition of Smad5 phosphorylation at 4-fold molar excess, but did not reduce the effects of BMP6 at 20-fold molar excess (Figure 2.7). In contrast, addition of IgG1-

Fc at a dose as high as 100-fold molar excess did not interfere with the actions of either growth factor (Figure 2.7). Moreover, as RGMc40-Fc and RGMc40 were equipotent in blocking the effects of BMP2 and BMP6 on Smad5 phosphorylation (Figure 2.7 C), we conclude that the IgG1-Fc domain does not influence the actions of the RGMc moiety in the recombinant fusion protein.

#### **2.4.7 BMP2 and BMP6-mediated Gene Expression Inhibited by RGMc50 and RGMc40**

We next examined the impact of RGMc40-Fc and RGMc50-Fc on BMP-activated gene expression. Addition of each fusion protein caused a dose-dependent decrease in the extent of BMP-stimulated accumulation of hepcidin, Smad7, and Id1 mRNAs, as measured by semi-quantitative RT-PCR, with greater inhibitory effects being observed on BMP6 than on BMP2 (Figure 2.9). Of note, Id1 gene expression was reduced less by RGMc-Fc than was hepcidin or Smad7. As expected, Noggin-Fc was ~5 times more potent than RGMc-Fc in blocking the actions of BMP2 on gene expression, but had little effect on BMP6, and the IgG1-Fc domain on its own did not interfere with BMP-regulated gene activity at 100-fold molar excess (Figure 2.9). In addition, RGMc40 and RGMc40-Fc were equivalently inhibitory (Figure 2.9 C).

#### **2.4.8 Microarray Analysis of Gene Expression**

To extend the scope of our experiments on the actions of BMPs and RGMc in Hep3B cells, we examined changes in gene expression using microarray assays. Whole transcript analysis with Human Gene 1.0 ST GeneChip® arrays revealed that BMP2 and BMP6 stimulated the accumulation of a nearly identical cohort of mRNAs, with ~40

transcripts being induced by at least 2-fold (Table 2.2), and 25 being increased by > 2.5-fold (Figure 2.10 A) within 4 hrs of BMP treatment. Additional pair-wise analyses demonstrated that Noggin completely inhibited all mRNAs whose expression was stimulated by BMP2, but had no effect on gene expression by BMP6 (Figure 2.10 A, Tables 2.3 and 2.4). In contrast, RGMc40 equivalently inhibited all transcripts induced by BMP2 or BMP6, but was generally less potent than Noggin in counteracting the actions of BMP2 (Figure 2.10 A, Tables 2.3 and 2.4). Of note, expression of Id1 and Id3 were reduced below basal levels by Noggin in the presence of BMP2, but were decreased substantially less than other transcripts by RGMc40. We cannot explain the enhanced potency of Noggin to block expression of Id1 or Id3, or the reduced efficacy of RGMc40. Except for these two outliers, the inhibitory potency of RGMc40 was generally about half of that of Noggin for genes whose expression was induced by BMP2 (Figure 2.10 A, Tables 2.3 and 2.4). To verify and extend the results of microarray studies, we tested the effects of Noggin and RGMc40-Fc on BMP-stimulated gene expression for ATOH8, SLC6A19, DUSPI, and GATA2 by semi-quantitative PCR. As also shown by microarray results, Noggin was more effective than RGMc40 in blocking the actions of BMP2, but was less effective in inhibiting BMP6-induced genes (Figure 2.10 B).

#### **2.4.9 Selected JH-associated RGMc Mutants Inhibit BMP2 and BMP6**

We next addressed the effects of selected disease-associated RGMc mutant proteins on BMP-activated gene expression. We expressed and purified as IgG1-Fc fusion proteins the mouse versions of human RGMc amino acid substitution mutants G99V and G320V (RGMc-G92V-Fc and RGMc-G313V-Fc, respectively), as well as the

truncation mutant 148X (mouse RGMc141X-Fc), as illustrated in Figure 2.11. Addition of a 10-fold molar excess of either RGMc50-Fc or RGMc40-Fc to Hep3B cells caused a ~60% decrease in the extent of BMP-stimulated accumulation of transcripts encoding hepcidin and Smad7, while a 50-fold molar excess of IgG1-Fc was ineffective, and a 4-fold excess of Noggin-Fc completely inhibited the actions of BMP2 but not BMP6 (Figure 2.12). RGMc-G92V-Fc did not inhibit the effects of either BMP, but surprisingly RGMc-G313V-Fc was as potent as wild-type RGMc50-Fc or RGMc40-Fc, particularly for BMP2, and RGMc141X-Fc was about half as effective on a molar basis (Figure 2.12). Since previous studies indicated that RGMc-G313V-Fc could bind BMP2 ~10% as effectively as RGMc50-Fc, and that RGMc-G92V-Fc could not bind BMP2 (68), these results show that RGMc mutants inhibit BMP actions by direct growth factor binding, and indicate that there is a rough correlation between binding strength and inhibitory activity.

## **2.5 DISCUSSION**

Soluble RGMc50 and RGMc40 can function as broad-based BMP inhibitors, and prevent both BMP2 and BMP6 from activating their receptors, as evidenced by the acute impairment of Smad phosphorylation, and the longer-term reduction in gene activation that we observed. In our present work, we have employed signaling studies, RT-PCR and microarray-based gene expression analyses to assay the effects of RGMc50 and 40 on BMP2 and BMP6. Our results are discussed below.

### **2.5.1 RGMc-Fc Fusion Proteins Retain Native Properties**

The use of an IgG1-Fc fusion tag has become quite prevalent (27, 34, 68, 86), as the addition of this domain allows for enhanced solubility, higher protein expression levels, and a one-step purification scheme involving binding of the IgG1-Fc domain to Protein A (68). Lin's group used a HJV-IgG1-Fc expression construct to purify soluble RGMc for both *in vitro* and *in vivo* mouse studies (27, 34, 86), however the native properties of the protein were not characterized until here. We have shown that RGMc-IgG1-Fc, along with the IgG1-Fc domain itself, is N-linked glycosylated through carbohydrate cleavage studies with PNGaseF. The fact that RGMc is N-linked glycosylated at multiple sites and that the full-length protein contains 14 cysteines that could give rise to 7 disulfide bonds could very well be the reason why many attempts to express this protein in bacterial systems were unsuccessful. By separating the fusion proteins by non-reducing SDS-PAGE we were able to establish that the IgG1-Fc domains retain their ability to dimerize.

Through structural studies with CD spectroscopy we determined that RGMc40 retains the same secondary structure irrespective of being bound to IgG1-Fc. Unfortunately expression levels from most cell culture model systems expressing wild-type RGMc did not lend to purification of high enough yields of soluble protein for structural comparison. However, through functional analyses we previously showed that several RGMc-IgG1-Fc fusion proteins can bind to BMP2 (68).

### **2.5.2 Soluble RGMc Inhibits BMP2 and BMP6-mediated Signaling and Gene Expression**

Our ability to generate highly purified RGMc-IgG Fc fusion proteins has further allowed us to determine that RGMc50 and RGMc40 are fairly equivalent BMP inhibitors on a molar basis, and to demonstrate that their effects are similarly antagonistic to both BMP2 and BMP6, even though these two BMPs are only ~61% identical in amino acid sequence (140), and preferentially activate different Type I and Type II receptors (141). Remarkably, in Hep3B cells both BMPs are nearly equivalently active, as measured by their ability to stimulate Smad phosphorylation and to promote the expression of an identical cohort of ~40 genes to the same extent in microarray profiling studies, indicating that several classes of receptors must be present in these cells.

The results obtained with RGMc50 or RGMc40 on BMP-mediated signaling and gene regulation contrast with the effects of Noggin, which in our hands was ~5-fold more potent on a molar basis in inhibiting BMP2 than were either RGMc protein, but was ineffective toward BMP6. Among other previously described BMP antagonists, Chordin

is relatively potent toward BMP2 and 4 (39) and follistatin is fairly effective against BMP7 (40), while DAN and follistatin-related protein are fairly weak toward BMP2, 4, and 7 (41-43). Chordin-like and Sclerostin can block the actions of BMP6, but not BMP2 (44, 45), and CTGF and Nov family members primarily inhibit BMP2 and 4 (46, 47). Thus, soluble RGMc proteins appear to represent a fairly unique class of broad-spectrum BMP antagonists that target two different BMP sub-families.

### **2.5.3 BMP2 and BMP6 are Inhibited by Selected JH-associated RGMc Mutants**

The discovery that mutations in the gene encoding HJV/RGMc resulted in the rapidly progressive iron overload disorder, JH (59), first implicated this protein in the regulation of whole-body iron metabolism (87, 88). Several single-nucleotide mutations in the human HJV/RGMc gene that cause JH are predicted to encode RGMc proteins with single amino acid substitutions (59). In one of the more prevalent disease-associated mutants, glycine residue 320 is changed to valine (G320V (59); G313V in mouse RGMc (59)), and this protein can bind BMP2, although less effectively than wild type RGMc (68). We now find that purified recombinant mouse RGMc G313V can inhibit BMP2- and BMP6-stimulated hepcidin gene expression almost as effectively as wild type RGMc50. In contrast, the G92V mutation (G99V in human RGMc (59)), which cannot bind BMP2 (68), does not interfere with BMP-mediated signaling. Although a clinical relationship has not been established yet between a specific HJV genotype and disease phenotype in JH, and the precise levels of expression or secretion of mutated HJV/RGMc proteins in individuals with JH are unknown, it is conceivable that mutations which lead

to soluble RGMc species that bind BMPs may be more severe than those that do not, because of a greater inhibition of hepcidin production.

Very few studies to date have examined the actions of truncated forms of RGMc, such as those predicted from JH-associated HJV/RGMc gene mutations that cause frame shifts within the protein coding region (59). We now find that the mutant 141X truncation, consisting of ~110 NH<sub>2</sub>-terminal residues of mature RGMc, when produced and purified as an IgG Fc fusion protein, retains the ability to inhibit the actions of both BMP2 and BMP6. These results allow us to hypothesize that a major domain for BMP binding is found within the NH<sub>2</sub>-terminal third of RGMc. Since the RGMc G92V mutant protein cannot not bind BMPs (68), and does not reduce their biological actions, we further postulate that a glycine or similarly sized aliphatic amino acid at this position is essential for interactions of RGMc with BMPs. Additional studies will be needed to define the critical determinants for BMP binding.

#### **2.5.4 Possible Implications of these Findings on the Role of RGMc in Iron Regulation**

RGMc is expressed early in development in skeletal muscle, heart and liver. Proper expression and localization may be most critical during early development as mutations in RGMc lead to the aggravated early on-set iron overload disorder, JH. As this may be the case, the two soluble HJV and RGMc 50 and 40 kDa isoforms have been detected in the serum of both adult humans and mice, respectively (74) and therefore may

play a role in iron regulation, or as a broad spectrum BMP inhibitor, throughout the lifespan of an organism.

The prevailing view of the role of RGMc in iron regulation is that cell-associated RGMc acts as a co-receptor that enhances BMP-mediated actions while soluble RGMc is a decoy receptor that has an opposing function (34). Although that is an attractive scenario that explains the disease phenotype, no one has ever successfully shown a significant increase in BMP-mediated Smad phosphorylation or up-regulation of gene expression in the presence of RGMc. Our results show that soluble RGMc is a potent inhibitor of both BMP2- and BMP6-mediated signaling in Hep3B liver cells. We have demonstrated through microarray analyses that many BMP-response genes, both involved in iron regulation and not, are repressed. Although hepcidin, a key player in iron regulation, is stimulated by BMP2 and BMP6 1.4- and 1.3-fold, respectively, its expression was not nearly as enhanced as other well-described BMP-responsive genes, such as Id1, Smad7 and the growth hormone receptor whose increases were all greater than 4-fold. However, it is possible that BMP-mediated hepcidin expression is enhanced to a greater extent *in vivo* or in primary hepatocytes which we did not use in our studies.

Through its inhibitory actions on BMPs, soluble RGMc suppresses the expression of hepcidin. JH-linked mutations in RGMc result in reduced hepcidin levels, likely indicating that the predominant player in the JH disease phenotype is the cell-associated RGMc isoform. However, through mutational analyses with soluble RGMc we were able to assess several important details about the functional interaction between RGMc and BMPs. Our results show that a glycine to valine mutation at position 92 within the RGD

domain results in no inhibition of BMP-mediated gene expression. We also show that a glycine to valine mutation at position 313 or truncation of the protein down to the first 110 N-terminal residues does not alter the protein's ability to inhibit BMP-mediated gene expression. Therefore if cell-associated RGMc binds to BMP in the same manner as soluble RGMc, and up-regulates BMP-mediated hepcidin expression as proposed, the G92V mutation should not be tolerated while the G313V mutation would be acceptable. However, both mutations have been linked to JH and low hepcidin levels. Few studies have critically examined the role that the receptor neogenin plays in RGMc iron regulation. The RGMc G92V mutant binds to neogenin, while the G313V mutant does not (68), which may indicate that two unrelated signaling pathways contribute to RGMc modulation of hepcidin expression. RGMc mutations that resulted in truncated protein species can only exist as soluble species as they lack their C-terminal GPI anchors. These mutant proteins cannot enhance BMP-mediated actions as proposed for cell-associated isoforms; however the soluble truncated proteins that retain the ability to bind BMPs can inhibit BMP-mediated responses.

### **2.5.5 Summary**

In summary, through the use of highly purified recombinant fusion proteins, we have shown that soluble RGMc species are effective BMP antagonists, and appear to act by preventing access of at least two classes of BMPs to cell surface receptors. Soluble RGMc proteins thus have a broader spectrum of inhibitory effects than Noggin, which primarily targets the BMP2/4 sub-class (143), and other previously-described BMP antagonists, which also have a limited range of activity (41-47). We additionally find

that selected JH-associated RGMc mutant proteins also can inhibit BMP actions, raising the hypothesis that disease severity in JH may vary depending on potential functions of individual mutants. Moreover, deletion mapping studies with a JH-linked truncation mutant have allowed us to identify a minimal BMP binding domain within the NH<sub>2</sub>-terminal ~110 amino acids of RGMc. Our results thus establish an experimental framework for discerning the biological roles of soluble RGMc proteins in BMP-mediated signaling, and for defining the critical amino acids and structural determinants responsible for binding these diverse BMPs.

## **2.6 ACKNOWLEDGEMENTS**

Microarray assays were performed in the Affymetrix Microarray Core of the OHSU Gene Microarray Shared Resource. We thank Eric Olson and Hugh Arnold for training in GeneSifter® software. These studies were supported in part by National Institutes of Health Research Grants R01 DK042748 (to P. Rotwein), F31 HL095271 (to M. Nili) and National Science Foundation Grant 0746589 (to U. Shinde).

**Table 2.1 Primers used for RT-PCR analysis of BMP-mediated signaling**

<b>Gene</b>	<b>Location</b>	<b>DNA Sequence (5' - 3')</b>	<b>cDNA (bp)</b>
ATOH8	Exon 1 Exon 2	CCTCCTCCGAGATCAAAGC CGGCACTGTAGTCAAGGTCA	219
DUSPI	Exon 3 Exon 4	CTGCCTTGATCAACGTCTCA ACCCTTCCTCCAGCATTCTT	160
GATA2	Exon 2 Exon 3	GTCACTGACGGAGAGCATGA GCCTTCTGAACAGGAACGAG	232
HAMP (human hepcidin)	Exon 1 Exon 3	TGGCACTGAGCTCCCAGATC CGCAGCAGAAAATGCAGATG	209
ID1	Exon 1 Exon 1	AAACGTGCTGCTCTACGACA GATTCCGAGTTCAGCTCCA	153
SLC6A19	Exon 11 Exon 12	ACCCTGGCTACGAGGAATTT GTACTTCAGGTCCCCGTTC	212
SMAD7	Exon 1 Exon 2/3	TCCTGCTGTGCAAAGTGTTT TCTGGACAGTCTGCAGTTGG	211
S17	Exon 2 Exon 3/4	CATTATCCCCAGCAAAAAGC AGGCTGAGACCTCAGGAACA	155
<i>Hamp1</i> (mouse hepcidin 1)	Exon 2 Exon 3	GAGACAGACTACAGAGCTGCAG GTCAGGATGTGGCTCTAGGCTA	182
<i>S17</i>	Exon 2 Exon 5	ATCCCCAGCAAGAAGCTTCGGAACA TATGGCATAACAGATTAACAGCTC	332

**Table 2.2 Genes induced  $\geq 2.0$ -fold by BMP2 or BMP6 in Hep3B cells**

Gene	Accession No.	Gene Name	Fold Change	
			BMP2	BMP6
ID1	NM_181353	Inhibitor of DNA binding 1	13.2	11.2
ID3	NM_002167	Inhibitor of DNA binding 3	11.1	9.7
SLC6A19	NM_001003841	Solute carrier family 6, member 19	7.2	6.3
SMAD6	NM_005585	SMAD family member 6	6.7	6.2
SMAD7	NM_005904	SMAD family member 7	4.3	4.5
GHR	NM_000163	Growth hormone receptor	4.1	4.0
BAMBI	NM_012342	BMP and activin membrane-bound inhibitor homolog	4.1	3.8
	NR_015379	urothelial cancer associated 1	4.1	3.3
HEY1	NM_012258	Hairy/enhancer-of-split related with YRPW motif 1	4.0	3.9
RASSF5	NM_182663	Ras association (RalGDS/AF-6) domain family member 5	3.9	3.5
SLC29A3	NM_018344	Solute carrier family 29, member 3	3.8	3.6
DSE	NM_013352	Dermatan sulfate epimerase	3.7	3.7
SMAD9	NM_001127217	SMAD family member 9	3.7	3.5
PFKFB3	NM_004566	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	3.6	3.3
GPD1L	NM_015141	Glycerol-3-phosphate dehydrogenase 1-like	3.5	3.4
ATOX1	NM_032827	Atonal homolog 8	3.4	3.9
ID4	NM_001546	Inhibitor of DNA binding 4	3.4	2.8
GATA2	NM_032638	GATA binding protein 2	3.2	3.1
GATA3	NM_001002295	GATA binding protein 3	3.1	3.1
SLCO2A1	NM_005630	Solute carrier organic anion transporter family, member 2A1	3.1	2.9
PAQR8	NM_133367	Progesterin and adipoQ receptor family member VIII	2.9	2.9
DUSP1	NM_004417	Dual specificity phosphatase 1	2.9	2.2
SAMD11	NM_152486	Sterile alpha motif domain containing 11	2.8	2.5
SNAI1	NM_005985	Snail homolog 1 (Drosophila)	2.8	2.7
MCC	NM_001085377	Mutated in colorectal cancers	2.7	2.9
TBX3	NM_016569	T-box 3	2.6	2.8
SPRY4	NM_030964	Sprouty homolog 4 (Drosophila)	2.6	2.4
SCNN1A	NM_001038	Sodium channel, nonvoltage-gated 1 alpha	2.6	2.5
SSTR2	NM_001050	Somatostatin receptor 2	2.5	2.6
SNAI2	NM_003068	Snail homolog 1 (Drosophila)	2.5	2.5
RAB31	NM_006868	RAB31, member RAS oncogene family	2.5	2.3
PRICKLE1	NM_153026	Prickle homolog 1 (Drosophila)	2.4	2.6
PITPNC1	NM_181671	Phosphatidylinositol transfer protein, cytoplasmic 1	2.4	2.0
ANKRD1	NM_014391	Ankyrin repeat domain 1 (cardiac muscle)	2.3	2.0
FZD7	NM_003507	Frizzled homolog 7 (Drosophila)	2.2	2.4
MTERFD2	NM_182501	MTERF domain containing 2	2.1	2.0
FAM169A	NM_015566	Family with sequence similarity 169, member A	2.1	2.4
ETS2	NM_005239	V-ets erythroblastosis virus E26 oncogene homolog 2	2.1	2.1
GRK5	NM_005308	G protein-coupled receptor kinase 5	2.0	2.1
SPAG1	NM_003114	Sperm associated antigen 1	2.0	2.0
DNAJC22	NM_024902	DnaJ (Hsp40) homolog, subfamily C, member 22	2.0	2.2
DLL4	NM_019074	Delta-like 4 (Drosophila)	2.0	2.0
ZMIZ1	NM_020338	Zinc finger, MIZ-type containing 1	2.0	2.1

**Table 2.3 Effects of Noggin and RGMc40 on genes induced by BMP2**

Gene	Gene Name	Fold Change		
		BMP2 vs. un- treated	Noggin vs. BMP2	RGMc vs. BMP2
ID1	Inhibitor of DNA binding 1	13.2	- 29.4	- 2.0
ID3	Inhibitor of DNA binding 3	11.1	- 22.9	- 2.1
SLC6A19	Solute carrier family 6, member 19	7.2	- 6.4	- 2.9
SMAD6	SMAD family member 6	6.7	- 7.0	- 4.0
SMAD7	SMAD family member 7	4.3	- 4.6	- 2.1
GHR	Growth hormone receptor	4.1	- 3.8	- 1.4
BAMBI	BMP and activin membrane-bound inhibitor homolog	4.1	- 3.8	- 2.6
	urothelial cancer associated 1	4.1	- 4.2	- 2.2
HEY1	Hairy/enhancer-of-split related with YRPW motif 1	4.0	- 3.6	- 3.7
RASSF5	Ras association (RalGDS/AF-6) domain family member 5	3.9	- 4.3	- 3.0
SLC29A3	Solute carrier family 29, member 3	3.8	- 4.1	- 1.8
DSE	Dermatan sulfate epimerase	3.7	- 3.5	- 1.4
SMAD9	SMAD family member 9	3.7	- 3.9	- 2.1
PFKFB3	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	3.6	- 3.6	- 2.1
GPD1L	Glycerol-3-phosphate dehydrogenase 1-like	3.5	- 3.3	- 1.8
ATOH8	Atonal homolog 8	3.4	- 4.1	- 2.4
ID4	Inhibitor of DNA binding 4	3.4	- 4.0	- 2.2
GATA2	GATA binding protein 2	3.2	- 3.7	- 2.4
GATA3	GATA binding protein 3	3.1	- 3.5	- 2.3
SLCO2A1	Solute carrier organic anion transporter family, 2A1	3.1	- 2.7	- 1.8

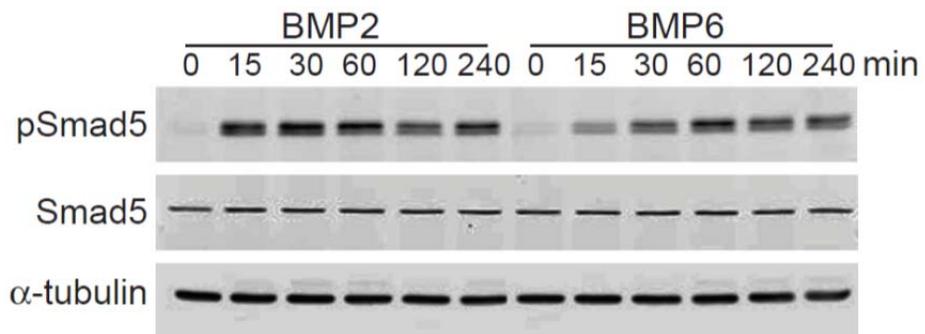
**Table 2.4 Effects of Noggin and RGMc40 on genes induced by BMP6**

Gene	Gene Name	Fold Change		
		BMP6 vs. un- treated	Noggin* vs. BMP6	RGMc vs. BMP6
ID1	Inhibitor of DNA binding 1	11.2	-	- 2.1
ID3	Inhibitor of DNA binding 3	9.7	-	- 2.5
SLC6A19	Solute carrier family 6, member 19	6.3	-	- 4.0
SMAD6	SMAD family member 6	6.2	-	- 4.4
SMAD7	SMAD family member 7	4.5	-	- 2.6
GHR	Growth hormone receptor	4.0	-	- 1.7
HEY1	Hairy/enhancer-of-split related with YRPW motif 1	3.9	-	- 4.0
ATOH8	Atonal homolog 8	3.9	-	- 3.5
BAMBI	BMP and activin membrane-bound inhibitor homolog	3.8	-	- 3.0
DSE	Dermatan sulfate epimerase	3.7	-	- 1.4
SLC29A3	Solute carrier family 29, member 3	3.6	-	- 2.2
SMAD9	SMAD family member 9	3.5	-	- 2.6
RASSF5	Ras association (RalGDS/AF-6) domain family member 5	3.5	-	- 3.2
GPD1L	Glycerol-3-phosphate dehydrogenase 1-like urothelial cancer associated 1	3.4	-	- 2.1
		3.3	-	- 2.4
PFKFB3	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	3.3	-	- 2.4
GATA2	GATA binding protein 2	3.1	-	- 2.4
GATA3	GATA binding protein 3	3.1	-	- 2.5

\*Noggin had no effect on any of these transcripts.

**Figure 2.1 BMP2 and BMP6 rapidly stimulate Smad phosphorylation in the Hep3B liver cell line.** Human Hep3B cells were incubated in serum-free medium with BMP2 or BMP6 at the indicated doses for up to 4 hr, and assayed for activation of Smad5. Immunoblots of whole cell protein lysates for phosphorylated Smad 5 (pSmad5), total Smad5, and  $\alpha$ -tubulin after incubation of cells with BMP2 or BMP6 (200 ng/ml) for 0 - 240 min.

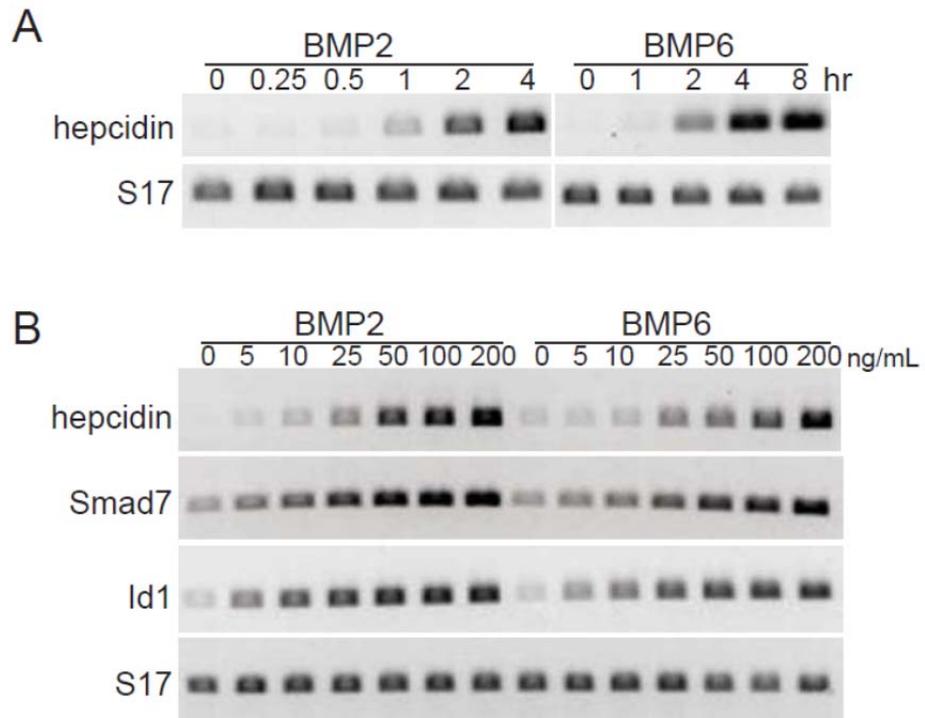
**Figure 2.1**



**Figure 2.2 BMP2 and BMP6 up-regulate BMP-responsive genes in Hep3B**

**cells.** Human Hep3B cells were incubated in serum-free medium with BMP2 or BMP6 at the indicated doses for up to 8 hr, and assayed for induction of gene expression. **A.** Measurement of hepcidin and S17 mRNAs by RT-PCR after incubation of cells with BMP2 or BMP6 (200 ng/ml) for 0 – 8 hr. **B.** Measurement of hepcidin, Smad7, Id1, or S17 mRNAs by RT-PCR after incubation of cells with graded doses of BMP2 or BMP6 (0 – 200 ng/ml) for 4 hr. Representative results are shown of 3 independent experiments.

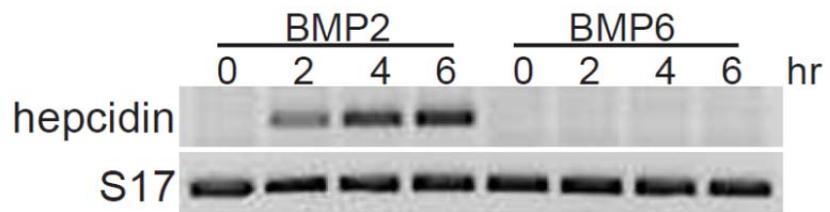
**Figure 2.2**



**Figure 2.3 BMP6 does not induce hepcidin gene expression in the mouse AML12**

**liver cell line.** Induction of hepcidin gene expression in AML12 cells. Time course of accumulation of hepcidin and S17 mRNAs by RT-PCR after incubation of AML12 cells with BMP2 or BMP6 (200 ng/mL) in serum-free medium for 0-6 hr.

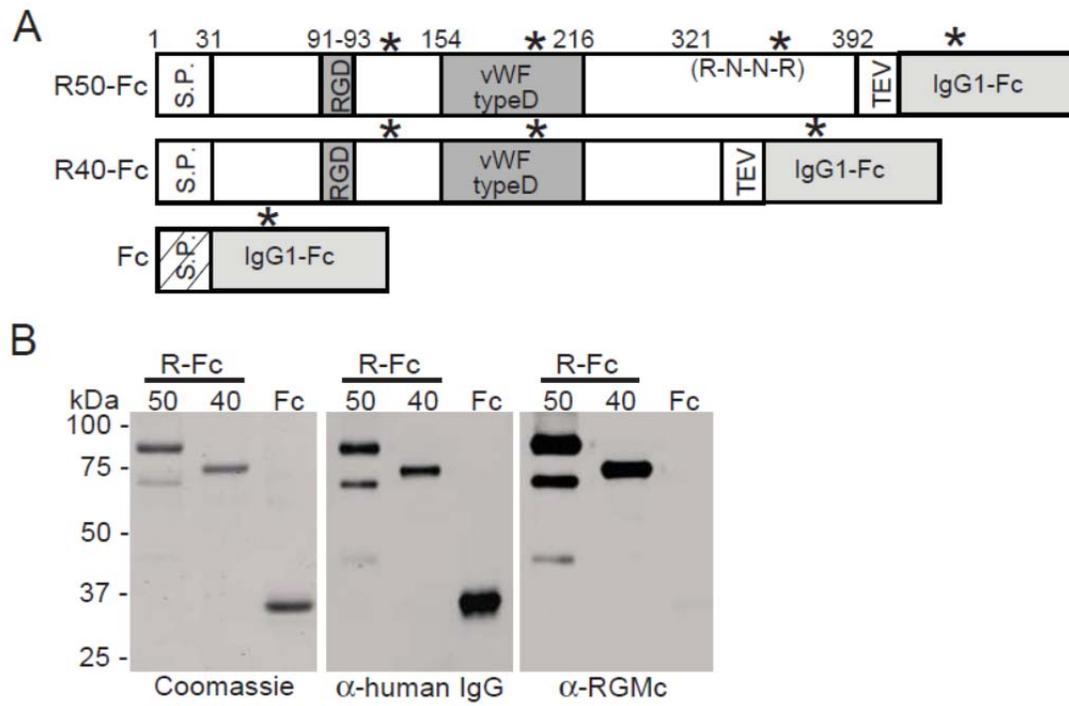
**Figure 2.3**



**Figure 2.4 Purification of soluble RGMc proteins from mammalian cells.**

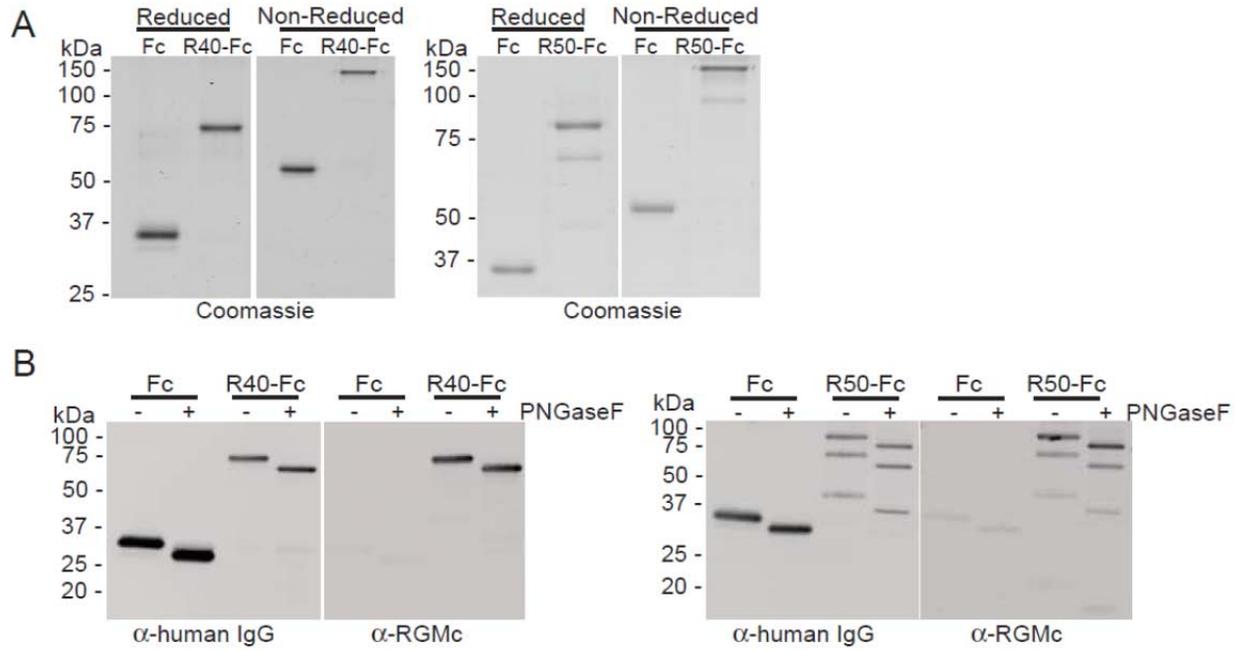
Recombinant mouse RGMc proteins were purified from HEK293 cell culture medium as described in Materials and Methods. **A.** Map of RGMc-IgG1-Fc fusion proteins showing locations of signal peptide (S.P.), RGD sequence, von Willebrand factor type D (vWF type D) domain, pro-protein convertase recognition site (R-N-R-R), TEV protease recognition site, IgG1-Fc domain, and N-linked glycosylation sites (asterisks). R50-Fc and R40-Fc are RGMc50-Fc and RGMc40-Fc, respectively. Also depicted is IgG1-Fc (Fc). **B.** Analysis after reducing SDS-PAGE of R50-Fc, R40-Fc, and Fc by staining with Coomassie blue (left panel), or after immunoblotting with anti-human IgG antibody ( $\alpha$ -human IgG, middle) or anti-RGMc antibody ( $\alpha$ -RGMc, right). Molecular weight markers are indicated

**Figure 2.4**



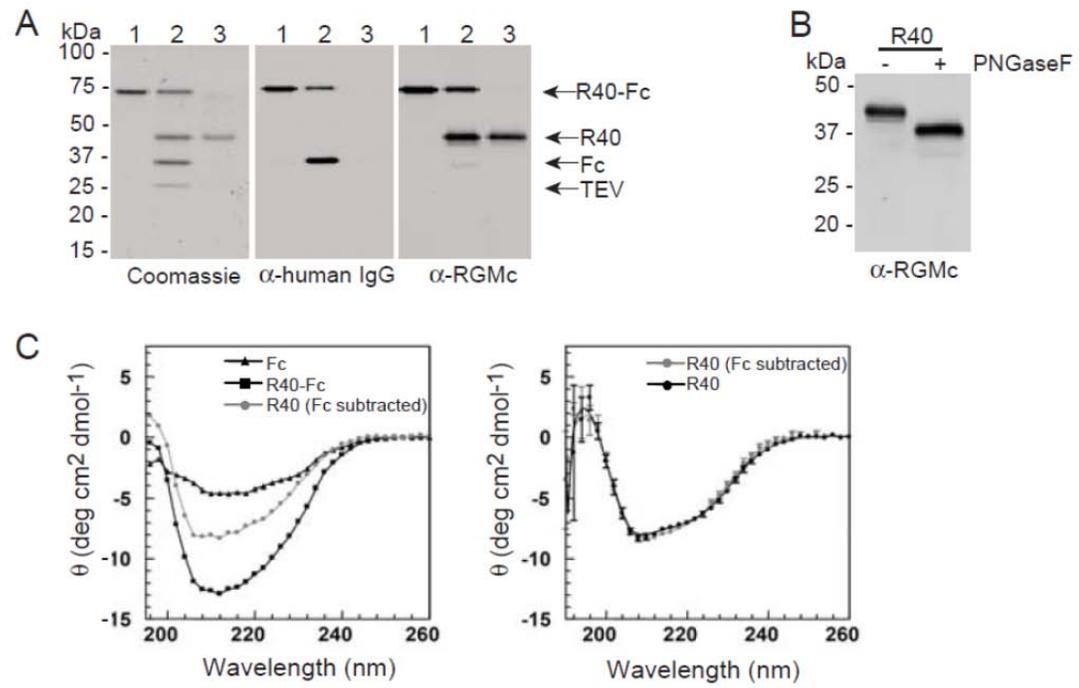
**Figure 2.5** **Characterization of soluble RGMc proteins.** **A.** RGMc-Fc fusion proteins are dimers. Coomassie blue stained gels showing purified Fc and R40-Fc (left panels) or R50-Fc (right) after separation by SDS-PAGE under reducing (left panel) or non-reducing (right) conditions. **B.** R40-Fc, R50-Fc and Fc are glycosylated. Immunoblots of PNGaseF treated purified Fc and R40-Fc (left panels) or R50-Fc (right) with antibodies to human IgG (left panel) or to RGMc (right). For **A-B** molecular weight markers are indicated

**Figure 2.5**



**Figure 2.6 Purification and characterization of RGMc40.** **A.** Purification of RGMc40 from the Fc fusion protein. Analysis after reducing SDS-PAGE by Coomassie blue staining (left panel), and immunoblotting with  $\alpha$ -human IgG (middle) or  $\alpha$ -RGMc (right). Lane numbers represent: (1) Purified R40-Fc. (2) R40-Fc after incubation with TEV for 3 hr at 20°C. (3) Purified R40 after sequential NiSepharose and Protein A Sepharose affinity chromatography (see Methods for details). **B.** R40 is glycosylated. Immunoblot of PNGaseF treated purified R40 with RGMc antibody. For **A-B** molecular weight markers are indicated. **C.** RGMc40 and RGMc40-Fc are structurally similar. Left panel: Results of 3 averaged CD spectra for purified RGMc40-Fc (black squares) and IgG1-Fc (black triangles) measured from 190 to 260 nm at 4°C (left). The gray circles and gray line depict a spectrum for RGMc40 that was calculated by subtracting data for IgG1-Fc from RGMc40-Fc. Right panel: Results of 5 averaged CD spectra for purified RGMc40 measured from 190 to 260 nm at 4°C (black circles) compared with the calculated spectrum from the left panel for RGMc40 (gray circles; corrected for IgG1-Fc). The standard deviation of the measurements is included at every fourth data point.

**Figure 2.6**

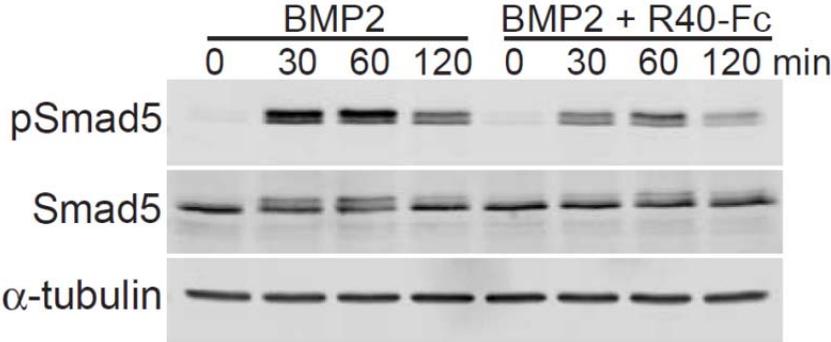


**Figure 2.7** Dose-dependent inhibition of BMP2- and BMP6-mediated signaling by RGMc50 and RGMc40. BMP2 or BMP6 (100 ng/ml) were pre-incubated for 3 hr at 20°C with various concentrations of Noggin-Fc (Nog-Fc), IgG1-Fc (Fc), or RGMc proteins, and after addition to cells for 2 hr, pSmad5, Smad5, and  $\alpha$ -tubulin were detected by immunoblotting. **A.** Nog-Fc: 2-20-fold molar excess, Fc: 100-fold molar excess, or RGMc50-Fc: 2-40-fold molar excess over BMP. **B.** Nog-Fc: 4-10-fold molar excess, Fc: 100-fold molar excess, or RGMc40-Fc: 2-40-fold molar excess over BMP. **C.** Nog-Fc: 10-fold molar excess, RGMc40-Fc (R40-Fc): 10-fold molar excess, RGMc40 (R40): 10-fold molar excess, or Fc: 100-fold molar excess over BMP.



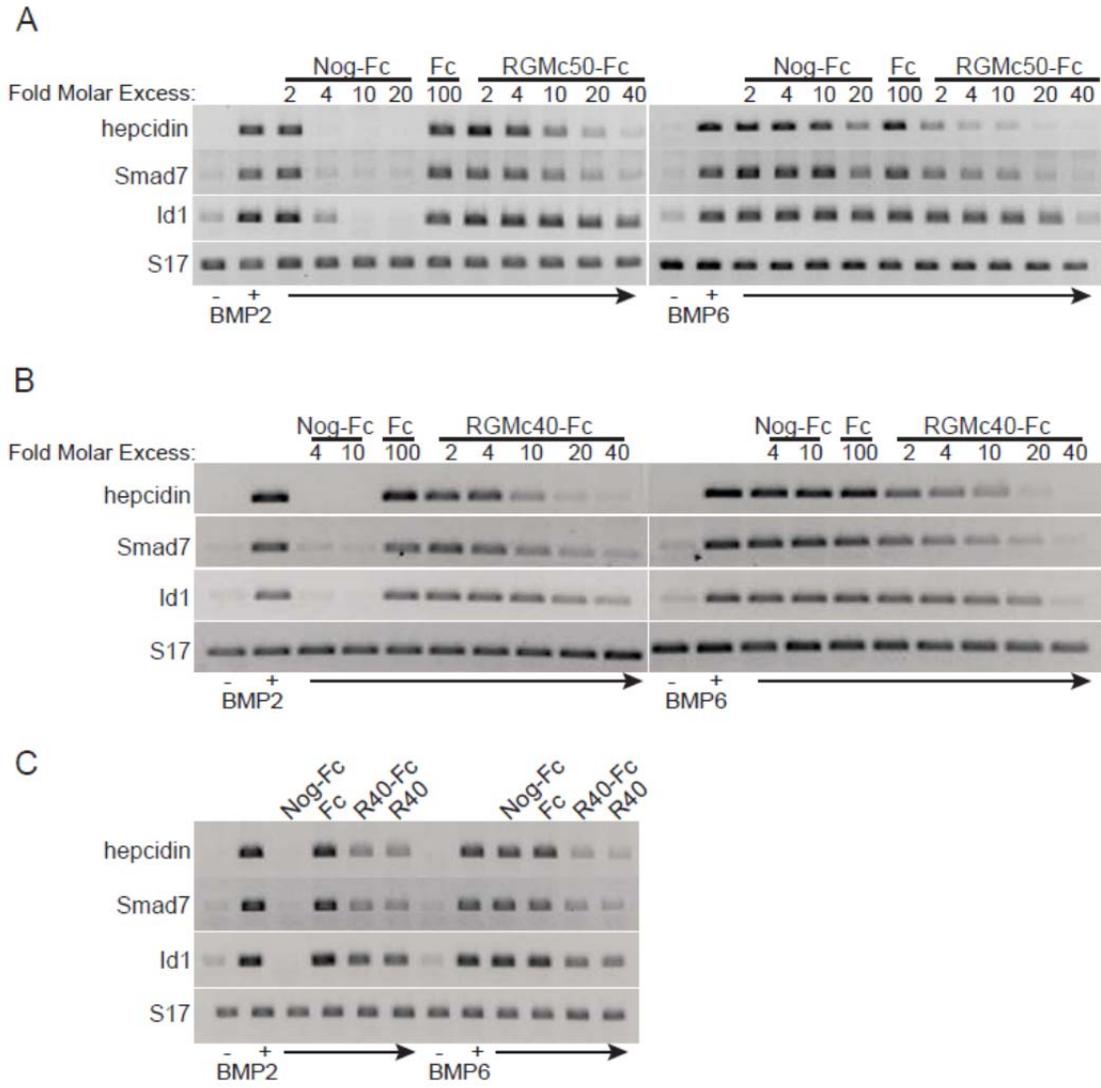
**Figure 2.8** Time course of RGMc40 inhibition of BMP2-mediated signaling. BMP2 (100 ng/mL) was pre-incubated for 3 hr at 20°C alone (left) or with 20-fold molar excess of RGMc40-Fc (R40-Fc, right). After addition to Hep3B cells for indicated times, pSmad5, Smad5, and  $\alpha$ -tubulin were detected by immunoblotting.

**Figure 2.8**



**Figure 2.9** Dose-dependent inhibition of BMP2 and BMP6-mediated gene expression by RGMc50 and RGMc40. BMP2 or BMP6 (100 ng/ml) were pre-incubated for 3 hr at 20°C with various concentrations of Noggin-Fc (Nog-Fc), IgG1-Fc (Fc), or RGMc proteins, and after addition to cells for 4 hr, hepcidin, Smad7, Id1 and S17 mRNA levels were analyzed by RT-PCR. **A.** Nog-Fc: 2-20-fold molar excess, Fc: 100-fold molar excess, or RGMc50-Fc: 2-40-fold molar excess over BMP. **B.** Nog-Fc: 4-10-fold molar excess, Fc: 100-fold molar excess, or RGMc40-Fc: 2-40-fold molar excess over BMP. **C.** Nog-Fc: 10-fold molar excess, RGMc40-Fc (R40-Fc): 10-fold molar excess, RGMc40 (R40): 10-fold molar excess, or Fc: 100-fold molar excess over BMP.

**Figure 2.9**



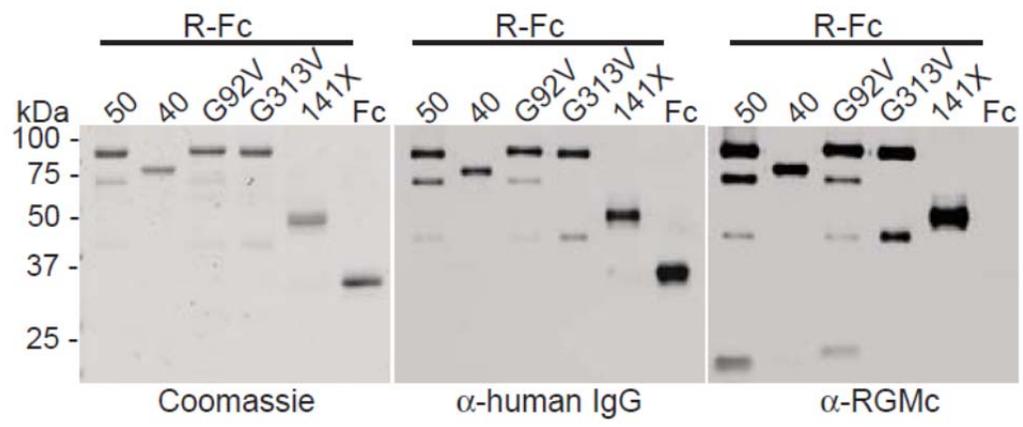
**Figure 2.10 RGMc40 inhibits BMP2- and BMP6-mediated gene expression. A.**

Results of microarray analyses using RNA isolated from Hep3B cells treated with BMP2 or BMP6 (100 ng/ml) for 4 hr. For the indicated samples growth factors were pre-incubated for 3 hr at 20°C with a 10-fold molar excess of Noggin-Fc or a 20-fold molar excess of RGMc40-Fc. Displayed is a heat map of genes identified as differentially expressed (> 2-fold change vs. un-treated; one-way ANOVA,  $p < 0.02$ ). The scale bar indicates the approximate fold-change vs. un-treated. See Supplemental Table 1 for additional gene information. **B.** Analysis of ATOH8, SLC6A19, DUSPI, GATA2, and S17 mRNA levels by RT-PCR after incubation of cells with BMP2 or BMP6 for 4 hr  $\pm$  a 10-fold molar excess of Noggin-Fc or a 20-fold molar excess of RGMc40-Fc administered as in **A**.



**Figure 2.11 Purification of mutant RGMc proteins.** Analysis of purified RGMc-Fc (R-Fc) fusion proteins or IgG1-Fc (Fc) after SDS-PAGE by Coomassie blue staining (left panel), or by immunoblotting for IgG1-Fc (middle), or RGMc (right). RGMc proteins are as follows: 50 – RGMc50, 40 – RGMc40, G92V – RGMcG92V substitution mutant, G313V – RGMcG313V substitution mutant, 141X – RGMc141X truncation mutant.

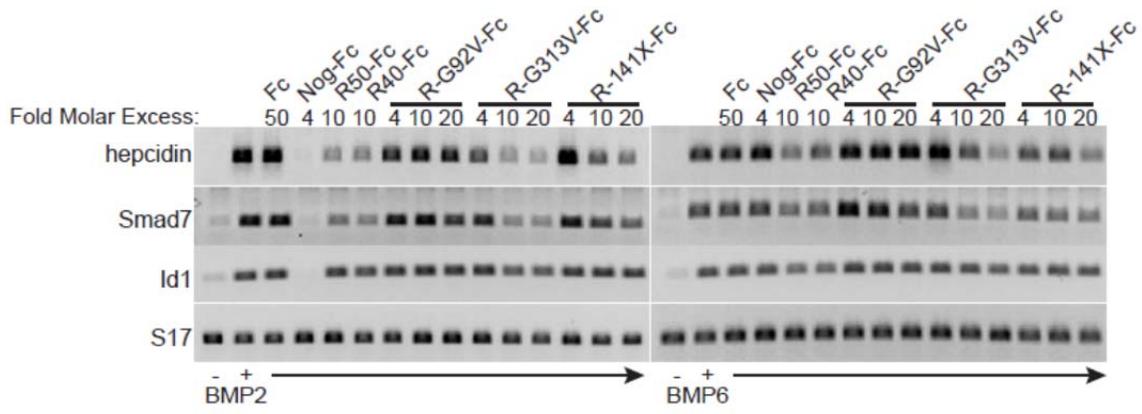
**Figure 2.11**



**Figure 2.12 Selected mutant RGMc proteins inhibit BMP-mediated gene**

**expression.** BMP2 or BMP6 (100 ng/ml) were pre-incubated for 3 hr at 20°C with 50-fold molar excess of IgG1-Fc, 4-fold molar excess of Noggin-Fc (Nog-Fc), 10-fold molar excess of R50-Fc or R40-Fc or 4-20-fold molar excess of R-G92V-Fc, R-G313V-Fc, or R-141X-Fc, and then added to Hep3B cells for 4 hr. Analysis of hepcidin, Smad7, Id1, and S17 mRNA levels by RT-PCR is shown.

**Figure 2.12**



# Chapter 3

## **Defining the Disulfide Bonds of Insulin-Like Growth Factor Binding Protein-5 by Tandem Mass Spectrometry with Electron Transfer Dissociation and Collision-Induced Dissociation**

Mahta Nili, Aditi Mukherjee, Ujwal Shinde, Larry David, and Peter Rotwein

Department of Biochemistry and Molecular Biology, Oregon Health & Science

University, Portland, OR 97239-3098, USA

MS experiment design, execution, and data analysis were performed by the author of this dissertation. A. Mukherjee expressed and purified IGFBP-5. U. Shinde built the IGFBP-5 homology model. L. David aided in design and execution of the MS experiments. P. Rotwein aided in design of experiments and planning of manuscript. (This work has been submitted to *J. Biological Chemistry* and is in revision).

### 3.1 SUMMARY

The six high-affinity insulin-like growth factor binding proteins (IGFBPs) comprise a conserved family of secreted molecules that modulate IGF actions by regulating their half-life and access to signaling receptors, and also exert biological effects that are independent of IGF binding. IGFBPs are composed of cysteine-rich amino- (N-) and carboxyl- (C-) terminal domains, along with a cysteine-poor central linker segment. IGFBP-5 is the most conserved IGFBP, and contains 18 cysteines, but only 2 of 9 putative disulfide bonds have been mapped to date. Using a mass spectrometry (MS) based strategy combining sequential electron transfer dissociation (ETD) and collision-induced dissociation (CID) steps, in which ETD fragmentation preferentially induces cleavage of disulfide bonds, and CID provides exact disulfide linkage assignments between liberated peptides, we now have definitively mapped 5 disulfide bonds in IGFBP-5. In addition, in conjunction with *ab initio* molecular modeling we are able to assign the other 4 disulfide linkages to within a GCGCCXXC motif that is conserved in five IGFBPs. Due to the nature of ETD fragmentation MS experiments were performed without chemical reduction of IGFBP-5. Our results not only establish a disulfide bond map of IGFBP-5 but also define a general approach that takes advantage of the specificity of ETD and the scalability of tandem MS, and the predictive power of *ab initio* molecular modeling to characterize unknown disulfide linkages in proteins.

### 3.2 INTRODUCTION

The two closely related peptide growth factors, insulin-like growth factor-I and -II (IGF-I and IGF-II), are necessary for normal growth and development in mammals and other vertebrates, and exert biological effects that promote proliferation, differentiation, and/or survival of a variety of cell and tissue types (146-148). In the circulation and in the extracellular space, IGFs are normally bound to one of six members of a conserved family of IGF binding proteins (IGFBPs), which modulate IGF actions by regulating IGF half-life and access to cell surface signaling receptors (149). Several studies also suggest that IGFBPs control other biological processes that are independent of their IGF binding properties (150, 151). Each IGFBP mediates both unique and overlapping actions based in part on tissue- and developmental-stage specific patterns of expression, and on different affinities for each IGF and for other bioactive molecules (151, 152).

The six IGFBPs are secreted proteins of 201-289 amino acids in length (151, 153) and share ~36% sequence identity (153). Each IGFBP contains highly conserved N- and C-terminal domains, along with a less conserved central linker segment (151, 154). Most IGFBPs have 12 and 6 cysteine residues in their N- and C-terminal domains, respectively, but lack cysteines in the linker region. Exceptions include IGFBP-4, with two cysteines in its linker segment (155), and IGFBP-6, with only 10 cysteines in its N-terminal domain (156). In addition, IGFBPs 1-5 share a cysteine-rich motif, GCGCCXXC (where X is any amino acid), within the N-terminal domain (151, 153, 157). Limited insights into the three-dimensional organization of IGFBPs have come from results of high-resolution x-ray crystallographic analyses of the isolated N-terminal

domain of IGFBP-4 and the C-terminal segments of IGFBP-1 and IGFBP-4 (158, 159). One consistent observation from these data is that IGFBPs lack inter-domain disulfide-bonds (157-160). However, as the structure of a full-length IGFBP has not been solved, possibly because of the disordered nature of the linker segment, this conclusion remains provisional.

IGFBP-5, a 252-amino acid mature protein with 18 cysteine residues (126), is the most conserved IGFBP in mammals (127); for example, human and mouse IGFBP-5 are 97% identical to one another (128). IGFBP-5 has been found to be a key component of the IGF signaling axis in tissue repair and regeneration, and is able to regulate osteogenesis (161-164), muscle differentiation (165-169), and kidney development (170), among other processes (126). Furthermore, its deficiency in mice has led to increased growth but diminished glucose tolerance (171). IGFBP-5 also has been shown to exert IGF independent actions (150, 172-175). As with other IGFBPs, the N-terminal domain (residues 1-84) of IGFBP-5 encodes the primary IGF-binding site, with the C-terminal region (residues 165-252) contributing in a secondary way to binding stability and affinity (176-178). NMR and protein crystallographic studies of a portion of the N-terminal segment of IGFBP-5 (amino acids 40 to 92) have demonstrated that this part of the protein appears to be organized into a tight globular structure that contains an anti-parallel  $\beta$ -sheet stabilized by two disulfide bonds linking Cys<sup>47</sup> to Cys<sup>60</sup> and Cys<sup>54</sup> to Cys<sup>80</sup> (176, 178). Five residues located near these disulfide-linked cysteines (K68, P69, L70, L73, and L74) have been shown via mutagenesis studies to be major contributors to high affinity binding of IGF-I and IGF-II (176-179).

Disulfide bonds contribute to the proper folding of proteins and to the integrity and stability of their three-dimensional structures (93). Mass spectrometric (MS) methods for identifying disulfide bonds have improved over the last few years, and several proteins recently have been mapped using a tandem MS approach in which peptides, including disulfide-linked species, are selected in the MS1 scan and then subjected to ETD (ETD-MS2) followed by CID (CID-MS3) (116, 119). ETD fragments peptides via electron transfer from a radical anion to a protonated peptide, causing cleavage between C $\alpha$ -N bonds which results in c and z ions (107). However, in a disulfide-linked peptide, ETD has been demonstrated to preferentially cleave the disulfide bond rather than the peptide backbone (111). In contrast, CID rarely dissociates disulfide bonds, and generally fragments peptide backbones at the amide bond generating a series of y and b ions (103). Traditional MS disulfide mapping methodologies have employed CID, but only to compare protease-digested peptides in proteins treated with or without reducing agents (113). However, since ETD preferentially cleaves disulfide bonds, this approach may be applied to protein samples without prior chemical reduction (116), as subsequent CID fragmentation of the peptides liberated by ETD will then identify the cysteines involved in disulfide bonds (116).

Here we have characterized the disulfide linkage map for mouse IGFBP-5 by using a tandem MS approach combining ETD and CID. Our results definitively identify 5 of 9 disulfide bonds in the protein, and determine that the other 4 linkages involve the four cysteines within the conserved GCGCCMTC motif (residues 32 to 39) in the N-terminal region of the protein. As CID spectra of peptides containing this cysteine-rich

motif could not precisely assign the disulfide linkages within the N-terminal segment of IGFBP-5, we employed unconstrained *ab initio* modeling to further refine the map. Overall, our results demonstrate the power of a combined approach employing both sequential MS and *ab initio* molecular modeling to identify and characterize disulfide bonds in a protein, and define a complete disulfide linkage map for IGFBP-5. In addition, we find that amino acid substitution mutations in N-terminal domain residues that are critical for maintaining ligand-binding affinity (K68, P69, L70, L73, and L74) have a minimal impact on the global tertiary structure of IGFBP-5.

### **3.3 MATERIALS and METHODS**

#### **3.3.1 Materials**

Fetal bovine serum (FBS), Dulbecco's modified Eagle's medium (DMEM), phosphate-buffered saline (PBS), and trypsin/EDTA were purchased from Invitrogen (Carlsbad, CA). Sequencing grade chymotrypsin was purchased from Roche (Indianapolis, IN). Proteomics grade trypsin and heparin agarose were from Sigma-Aldrich (St. Louis, MO); Criterion precast gels were purchased from BioRad (Hercules, CA). AquaBlock EIA/WIB solution was from East Coast Biologicals (North Berwick, ME). GelCode Blue Stain Reagent was purchased from Pierce Biotechnologies (Rockford, IL). NitroBind nitrocellulose was from GE Water & Process Technologies (Trevose, PA). Biotinylated human IGF-II was from GroPep (Adelaide, Australia). Polyclonal anti-IGFBP-5 antibody was purchased from Santa Cruz Biotechnology (Santa Cruz, CA). Secondary antibodies, AlexaFluor 680-conjugated goat-anti-mouse IgG and IR800-conjugated streptavidin were from Invitrogen and Rockland Immunochemical (Gilbertsville, PA), respectively. Other chemicals and reagents were purchased from commercial suppliers.

#### **3.3.2 Expression and Purification of IGFBP-5**

Recombinant adenoviruses encoding the tetracycline transactivator protein (Ad-tTA), mouse IGFBP-5, and a modified mouse IGFBP-5 with amino acid substitutions within the IGF binding region (N mutant) have been described (166). C3H10T1/2 mouse embryonic fibroblasts (CCL226; ATTC, Rockville, MD), incubated at 37°C in humidified air with 5% CO<sub>2</sub> in DMEM with 10% fetal calf serum, were infected at ~50% of

confluent density with Ad-tTA plus either wild-type or N-terminal mutant IGFBP-5 at a multiplicity of infection of 500. The following day medium was replaced with DMEM plus 2% fetal bovine serum. IGFBP-5 was purified from medium conditioned for 48 hr using heparin-affinity chromatography, as described (166), and was stored in aliquots at -80°C until use.

### **3.3.3 Immunoblotting and Ligand Blotting**

For immunoblotting, purified IGFBP-5 (500 ng/lane) was separated by reducing SDS-PAGE and transferred to nitrocellulose membranes. After blocking with 50% AquaBlock solution for 1 hr at 20°C, membranes were incubated sequentially with  $\alpha$ -IGFBP-5 (1:500 dilution) and AlexaFluor 680-conjugated goat-anti-mouse IgG (1:10,000). For ligand blotting, purified IGFBP-5 (100 ng/lane) was resolved by non-reducing SDS-PAGE and transferred to nitrocellulose membranes. After blocking in 50% AquaBlock, membranes were incubated with biotin-conjugated IGF-II (100 ng/ml) in 50% AquaBlock in PBS with 0.1% Tween-20 for 16 hr at 4°C. Membranes were then washed with PBS plus 0.1% Tween-20, followed by incubation with IR800-conjugated streptavidin (1:10,000). Results were visualized and images captured using the Odyssey® Infrared Imaging System and version 3.0 analysis software (LI-COR, Lincoln, NE).

### **3.3.4 Digestion of IGFBP-5**

Thawed IGFBP-5 (6  $\mu$ g) was incubated protected from light with iodoacetamide [5 mM] with shaking for 30 min at 20°C in buffer with 4 M urea, and then was separated

by non-reducing SDS-PAGE using Criterion precast gels. Alternatively, IGFBP-5 was incubated in the gel after electrophoresis was completed with or without iodoacetamide. Protein bands were stained with GelCode Blue, de-stained with double deionized water, excised, and incubated twice in 500  $\mu$ l of 50 mM ammonium bicarbonate, 50% (v/v) acetonitrile while shaking for 30 min at 20°C. Samples were dehydrated in 100% acetonitrile for 2 min, dried by vacuum centrifugation, and rehydrated with 10 mg/ml of trypsin or chymotrypsin in buffer containing 50 mM ammonium bicarbonate and 5 mM calcium chloride for 15 min on ice. Excess buffer was removed and replaced with 50  $\mu$ l of the same buffer without enzyme, followed by incubation for 16 hr at 37°C or 20°C, for trypsin and chymotrypsin, respectively, with shaking. Digestions were stopped by addition of 3  $\mu$ l of 88% formic acid, and after brief vortexing, the supernatant was removed and stored at -20°C until analysis.

### **3.3.5 Localization of Disulfide Bonds by Mass Spectrometry**

Peptides were injected onto a 1 mm x 8 mm trap column (Michrom BioResources, Inc., Auburn, CA) at 20 ml/min in a mobile phase containing 0.1% formic acid. The trap cartridge was then placed in-line with a 0.5 mm x 250 mm column containing 5 mm Zorbax SB-C18 stationary phase (Agilent Technologies Inc., Santa Clara, CA), and peptides separated by a 2-30% acetonitrile gradient over 90 min at 10 ml/min with a 1100 series capillary HPLC (Agilent Technologies). Peptides were analyzed using a LTQ Velos linear ion trap with an electron transfer dissociation (ETD) source (Thermo Scientific, San Jose, CA). Electrospray ionization was performed using a Captive Spray source (Michrom Bioresources, Inc.). Survey MS scans were followed by 7 data-

dependant scans consisting of collision-induced-dissociation (CID) and ETD MS2 scans on the most intense ion in the survey scan, followed by 5 MS3 CID scans on the 1<sup>st</sup> - 5<sup>th</sup> most intense ions in the ETD MS2 scan. CID scans used normalized collision energy of 35, and ETD scans used a 100 ms activation time with supplemental activation enabled. Minimum signals to initiate MS2 CID and ETD scans were 10,000, minimum signals for initiation of MS3 CID scans were 1000, and isolation widths for all MS2 and MS3 scans were 3.0 m/z. The dynamic exclusion feature of the software was enabled with a repeat count of 1, exclusion list size of 100, and exclusion duration of 30 sec. These experiments used inclusion lists to target specific cross-linked species for collection of ETD MS3 scans. Separate dta files for MS2 and MS3 scans were created by Bioworks 3.3 (Thermo Scientific) using ZSA charge state analysis. Matching of MS2 and MS3 scans to peptide sequences was performed by Sequest (V27, Rev 12, Thermo Scientific), using a database consisting of reversed yeast sequence entries, supplemented with the sequences of common contaminants, and the sequence of mouse IGFBP-5 (6182 entries total). The analysis was performed without enzyme specificity, a parent ion mass tolerance of 2.5, fragment mass tolerance of 1.0, and a variable mass of +16 for oxidized methionine residues. Searches of CID MS2 and MS3 data both specified matches to y and b ions. Results then were analyzed using the program Scaffold (V3\_00\_08, Proteome Software, Portland, OR) (180, 181) with minimum peptide and protein probabilities of 95% and 99% being used. IGFBP-5 peptides from MS3 results were sorted by scan number, and cysteine containing peptides were identified from groups of MS3 scans produced from the 5 most intense ions observed in ETD MS2 scans. The

identities of cysteine peptides participating in disulfide-linked species were further confirmed by manual examination of the parent ion masses observed in the survey scan and the ETD MS2 scan.

### **3.3.6 IGFBP-5 *ab initio* and Homology Modeling**

Structural models for an N-terminal segment, consisting of amino acids 5 to 41 and 1 to 84 of mouse IGFBP-5, were constructed using Rosetta *ab initio* modeling (123). Structures were generated using the standard Rosetta fragment server. The fragment selection procedure was performed *de novo* (without templates from existing structural homologues in the Protein Data Bank (PDB)) and also with access to PDB homologues. The 5,000 independent predicted structures from each search were subjected to clustering analysis. The centers of the five largest clusters were chosen as the best models, defined as having the lowest standard deviation of the mean among positions of  $\alpha$ -carbon atoms of all residues when compared with all other simulations in a cluster. A homology model of amino acids 5 to 84 of IGFBP-5 also was built using the alignment interface of SwissModel, which predicts structures reliably with a root mean square deviation  $< 2 \text{ \AA}$  for sequences with 50-60% identity (182), and using as a template the x-ray structure of the IGFBP-4 N-terminal domain (PDB Code 2DSR).

## **3.4 RESULTS**

### **3.4.1 Rational of Experimental Approach**

Traditional MS methodologies aimed at mapping the disulfide bonds of proteins require two separate experiments, with and without reducing agents. Recently strategies involving LC-MS combining ETD and CID have eliminated the need for chemical reduction as ETD preferentially cleaves disulfide bonds. A CID activation step following ETD allows for identification of peptides released after disulfide-linkage dissociation. To date, this method has been used to establish partial disulfide maps for immunoglobulin light chain and human growth hormone, and a complete map for tissue plasminogen activator, which contains 35 cysteines. However, these disulfide maps were established for proteins with known structures. Here we employed LC-MS combining ETD and CID to establish the disulfide linkages of IGFBP-5, a protein with 18 conserved cysteines.

### **3.4.2 An MS Approach to Mapping Disulfide Bonds in IGFBP-5**

There is little information on the disulfide bonding pattern of IGFBP-5, as only 4 of 18 cysteines in the protein have been mapped to date (176). Here we have applied a tandem MS approach combining ETD and CID to identify the linked cysteine residues in wild type mouse IGFBP-5 and in a N-terminal amino acid substitution mutant (Figure 3.1). Both proteins were purified after over-expression of recombinant adenoviruses in cultured mammalian cells (Figure 3.2). The 18 cysteines of IGFBP-5 potentially reside in 11 tryptic peptides, which could form up to 9 disulfide bonds (Figure 3.3). Purified IGFBP-5 was digested with trypsin under non-reducing conditions and subjected to a MS3 protocol with ETD (ETD-MS2) followed by CID (CID-MS3) (116). CID-MS2 also

was employed to help identify disulfide-linked peptides (Figure 3.1). The tryptic peptides analyzed by these methods covered 87% of IGFBP-5, and we could account for each of the 18 cysteines in the protein (data not shown). The m/z values for cysteine-containing tryptic peptides are found in Table 3.1.

IGFBP-5 used in our assays was prepared by over-expression in mammalian cells, and there was the possibility that these proteins underwent disulfide scrambling during or after biosynthesis. Although we could not prevent any re-arrangements that occurred during protein maturation, to limit any subsequent scrambling we alkylated IGFBP-5 with iodoacetamide. Therefore, any potentially free cysteine residues were blocked prior to SDS-PAGE and in-gel protease digestion of purified IGFBP-5. We obtained identical disulfide linkage assignments with and without alkylation (data not shown), and also observed highly similar chromatographic elution profiles for trypsin-digested IGFBP-5 and its N-terminal amino acid substitution mutant in multiple experiments (Figure 3.4). Therefore, based on the reproducibility of these results, we believe that our data accurately reflect the real disulfide map.

### **3.4.3 Mapping Disulfide Bonds in the C-terminal Domain of IGFBP-5**

ETD was used to fragment the linked tryptic peptides, P9, P10, and P11 within the C-terminal domain of IGFBP-5 (Figure 3.5A). Cysteine residues 221 and 223 in P10 are separated by a single amino acid, making it initially difficult to identify specific fragment ions in the CID-MS2 step to assign exact disulfide-linkages involving these two cysteines. However, in the CID-MS3 step following ETD-MS2 fragmentation in which P9 was released, we recovered an abundant fragment ion,  $y^{*4}$ , allowing us to assign the

linkage between Cys<sup>223</sup> in P10 and Cys<sup>243</sup> in P11 (Figure 3.5B). Based on these results we then can deduce that Cys<sup>210</sup> in P9 is bonded to Cys<sup>221</sup> in P10. The y and b ions from the CID-MS3 fragmentation step supporting the latter assignment are found in Table 3.2.

We next used CID-MS2 and ETD-MS2 to identify the remaining disulfide bond in the C-terminal domain, which we find linked Cys<sup>172</sup> and Cys<sup>199</sup> (Figure 3.6). CID-MS2 y and b ions are listed in Table 3.3. Taken together, results in Figures 3.5 and 3.6 show that there are three disulfide bridges within the C-terminal segment of IGFBP-5, and also indicate that there are no cysteine linkages that join the C- and N-terminal regions of the protein.

#### **3.4.4 Mapping Disulfide-Linkages in the N-terminal Domain of IGFBP-5.**

By CID-MS2 we detected 10 of the 12 cysteines found in the N-terminal domain of IGFBP-5 within 4 linked peptides connecting tryptic fragments P1, P2, P3, and P5 (Figure 3.7). To resolve these potentially highly intertwined disulfide bonds, we first analyzed the previously identified disulfide-linkage between Cys<sup>47</sup> in P3 and Cys<sup>60</sup> in P5 (176, 178) by ETD-MS2 followed by CID-MS3 of the precursor ion 865.9 (+7). Several y and b ions (e.g., b<sup>10</sup> and y<sup>8</sup>) generated from the P3 peptide precede Cys<sup>47</sup>, and support the linkage to Cys<sup>60</sup> (Figure 3.8A). Since neither P1 nor P2 was liberated after ETD fragmentation we reasoned that these peptides must be connected to other peptides in the grouping by two disulfide bonds. Analysis of CID-MS3 spectra following ETD-MS2 of linked peptides P1, P2, P3, and P5 provided evidence for two assignments involving P2 and P3, and P1 and P3. Recovery of peptides with a single cleaved disulfide bond

indicated that Cys<sup>25</sup> in P2 was linked to Cys<sup>39</sup> in P3 (Figure 3.8B, e.g., y<sup>17</sup> and b<sup>13</sup>), and that Cys<sup>7</sup> in P1 was bonded to Cys<sup>33</sup> in P3 (Figure 3.8C, e.g., y<sup>\*12</sup>, y<sup>\*8</sup>, y<sup>22</sup>).

To provide additional support for these findings IGFBP-5 was digested with chymotrypsin and subjected to the MS3 protocol. We identified two linked peptides (C4 and C5) that confirmed disulfide bonds between Cys<sup>47</sup> and Cys<sup>60</sup> (Figure 3.9A, Tables 3.1 and 3.4). Three other linked peptides containing six cysteine residues (Cys 7, 10, 18, 33, 35, 36) also were detected after CID-MS2 (C1, C2, C3, Figure 3.9B, Table 3.5). Digestion of IGFBP5 with chymotrypsin also should generate two cysteine-containing peptides of 4 amino acids each, GC<sup>25</sup>EL and TC<sup>39</sup>AL. The fact that neither of these peptides were found to be associated with the larger group of N-terminal chymotrypsin peptides (Figure 3.9B), suggests that Cys<sup>25</sup> and Cys<sup>39</sup> are disulfide bonded to one another, in agreement with analysis of the tryptic peptides (Figure 3.7). However, as we did not recover this small putative disulfide-linked peptide, we cannot definitively reach this conclusion.

As described above, disulfide bonds involving the N-terminal G<sup>32</sup>CGCCMTC<sup>39</sup> motif were not definitively established, because mapping the linkages between peptides with multiple cysteines by CID-MS3 requires ample spacing between individual cysteine residues to assign y and b ions. Unfortunately, neither protease digestion nor chemical cleavage strategies were able to separate these cysteines from one another. Because of these difficulties, we employed *ab initio* molecular modeling to collect information on all possible combinations of disulfide bonds within this region. For these experiments we limited our analyses to amino acids 5 - 41 of mouse IGFBP-5, which contained the 8

cysteines whose linkages could not be resolved completely by our tandem MS approach. We first employed *de novo ab initio* modeling using Rosetta (123), in which homologous structures in PDB, such as other IGFBPs, are not used as templates to guide predictions. Using this approach, we were able to generate multiple highly related predicted structures (Figure 3.10A) that were remarkably similar to the structures obtained using a homology-based search (Figure 3.10B). Both models also identified identical disulfide linkages involving the GCGCCMTC motif: Cys<sup>33</sup>-Cys<sup>7</sup>, Cys<sup>35</sup>-Cys<sup>10</sup>, Cys<sup>36</sup>-Cys<sup>18</sup>, and Cys<sup>39</sup>-Cys<sup>25</sup> (Figure 3.10C). Moreover, a *de novo ab initio* model of the complete N-terminal domain of IGFBP-5 (amino acids 1 - 84) aligned closely with predictions for IGFBP-5 based on the x-ray crystallographic structure of the N-terminal segment of IGFBP-4 (158, 159) (Figure 3.10D). Taken together, these results support the validity of using molecular modeling as part of a combined experimental approach with the MS3 protocol described here for delineating previously undefined disulfide linkages in proteins.

### **3.4.5 Amino Acid Substitution Mutations in the N-terminal Domain of IGFBP-5 that Reduce IGF Binding do not Alter Disulfide Bonds.**

Previous NMR studies had mapped the disulfide bond between Cys<sup>54</sup> in P4 and Cys<sup>80</sup> in P6, using a ‘mini-IGFBP-5’ protein as the starting material (178). We analyzed this linkage in full-length IGFBP-5 and compared it to results obtained with a N-terminal domain amino acid substitution mutant involving residues K68N, P69Q, L70Q, L73Q, and L74Q to determine whether disruption of this disulfide bond might account for the diminished IGF binding affinity of the latter protein (177, 179). Both wild type and N-mutant IGFBP-5 were digested with trypsin and subjected to the MS3 protocol (Figure

3.1). The Cys<sup>54</sup>-Cys<sup>80</sup> linkage was identified between peptides P4 and P6, precursor ion of m/z 352.8 (+4), by CID-MS2 and ETD-MS2 in both wild type and N-mutant IGFBP-5 (Figure 3.11, Table 3.6). In addition, since the complete elution profile of cysteine-containing peptides of the N-terminal IGFBP-5 mutant matches that of the wild type protein (Figure 3.4), these results imply that the overall tertiary structure of the N-terminal mutant is not perturbed.

## 3.5 DISCUSSION

### 3.5.1 IGFBP-5 is a Cysteine-Rich Protein with 9 Disulfide Bonds

In this study we have shown that the 18 cysteines in mouse IGFBP-5 form 9 disulfide bonds. Application of a tandem MS approach employing ETD and CID directly identified 5 disulfide linkages: Cys<sup>47</sup>-Cys<sup>60</sup>, Cys<sup>54</sup>-Cys<sup>80</sup>, Cys<sup>172</sup>-Cys<sup>199</sup>, Cys<sup>210</sup>-Cys<sup>221</sup>, and Cys<sup>223</sup>-Cys<sup>243</sup>, and the combination of MS with *ab initio* molecular modeling established the most likely arrangement of the other 4 disulfide pairs: Cys<sup>7</sup>-Cys<sup>33</sup>, Cys<sup>10</sup>-Cys<sup>35</sup>, Cys<sup>18</sup>-Cys<sup>36</sup>, and Cys<sup>25</sup>-Cys<sup>39</sup>. Taken together, our studies show that IGFBP-5 is composed of structurally independent N- and C-terminal domains, containing 6 and 3 disulfide bonds, respectively.

### 3.5.2 Structure of IGFBP Family Members

To date there has been no x-ray crystal structure reported for full-length IGFBP-5, nor for any other full-length IGFBP, although the complete C-terminal domains of IGFBP-1 and IGFBP-4 have been characterized (157-159, 176). Based on these structural data, on amino acid sequence similarity with IGFBP-1 and IGFBP-4, and on concordance in the location of cysteine residues between the two proteins, it is likely that the C-terminal segment of IGFBP-5 also is composed of a thyroglobulin type-1 fold consisting of an alpha helix and three-stranded antiparallel beta sheets held in a compact formation through the 3 disulfide bonds.

Amino acid sequencing and MS methods have been used previously to map some of the disulfide linkages in several IGFBPs. Protease cleavage followed by N-terminal

sequencing was employed to identify the 3 disulfide bonds in the C-terminal segment of IGFBP-2 (183), and electrospray ionization (ESI)-MS was used to identify all 8 disulfide linkages in IGFBP-6 (156). In both of these cases the characterized disulfide bonds in the C-terminal domain match the results that we have established here for IGFBP-5. Thus, in conjunction with x-ray crystallographic data for the C-terminal segments of IGFBP-1 and IGFBP-4, it is likely that the C-terminal domains of all five IGFBPs adopt a very similar overall conformation with only slight differences in secondary structural features.

### **3.5.3 C-terminal Disulfide Bonds of IGFBP-5**

Recently in a search for new antimicrobial peptides, Osaki et al. discovered in cell-conditioned tissue culture medium a disulfide-linked amidated peptide containing amino acids 193-214 derived from the C-terminal portion of IGFBP-5, in which Cys<sup>199</sup> was bonded to Cys<sup>210</sup> (184). Perhaps surprisingly, we also identified this Cys<sup>199</sup>-Cys<sup>210</sup> linkage in our analyses, but it was present as a very minor peptide species and was not detected in all protein samples evaluated (data not shown). In contrast, disulfide linkages Cys<sup>172</sup>-Cys<sup>199</sup>, Cys<sup>210</sup>-Cys<sup>221</sup>, and Cys<sup>223</sup>-Cys<sup>243</sup> were the dominant pairings found in every purified IGFBP-5 protein sample that we analyzed. Clearly, further studies will be needed to elucidate the biochemical mechanisms responsible for generation of this potentially alternative peptide from full-length IGFBP-5, and to define the structural features responsible for its novel biological properties.

### 3.5.4 Disulfide-Linkages within the N-terminal domain of IGFBP-5

The N-terminal domain of IGFBP-4 consists of a series of disulfide bridges that leads to a globular base, structural features that may define the IGF binding motif (157). In the N-terminal segment of IGFBP-5, two disulfide bonds analogous to two of the six disulfides in IGFBP-4, Cys<sup>47</sup>-Cys<sup>60</sup> and Cys<sup>54</sup>-Cys<sup>80</sup>, had been identified previously using solution-based NMR and x-ray crystallography of a ‘mini’ IGFBP-5 N-terminal domain protein (amino acids 40 – 92) (176, 178). We now confirm these assignments in full-length IGFBP-5. Amino acids 32-39 within the N-terminal part of IGFBP-5 comprise a conserved motif of GCGCCMTC that is found (as GCGCCXXC) in IGFBPs 1 - 4 (151, 153). We establish here that the four cysteine residues in this motif form the disulfide bonds that connect with cysteines 7, 10, 18, and 25 at the extreme N-terminus of IGFBP-5, a conclusion reached in conjunction with the application of *de novo ab initio* modeling. Overall, as depicted in Fig. 5C, it is likely that the three-dimensional structure of the N-terminal domain of IGFBP-5 is very similar to IGFBP4, and we predict that IGFBPs 1 - 3 (151) will exhibit analogous structural features.

A series of engineered amino acid substitutions within the N-domain of IGFBP-5 (K68N, P69Q, L70Q, L73Q, L74Q) results in a nearly 100-fold decline in binding affinity for IGF-I and IGF-II (177, 179). Despite this major perturbation in IGF binding capability, our results show that the disulfide-binding pattern of the cysteines flanking these mutations is not compromised. Thus, lower affinity binding of IGFs to this mutant IGFBP-5 does not reflect lack of structural integrity, but rather represents a loss of key interactions between the two molecules.

### 3.5.6 MS Combining CID and ETD for Establishing Disulfide-Linkages

Traditional approaches for mapping disulfide bonds have relied on a strategy comparing data generated with and without reductive alkylation, in which peptides isolated from the protein of interest after single or multiple proteolytic digestions were subjected to MS or other analytical methods (113). As ETD causes preferential cleavage of disulfide bonds rather than the peptide backbone it can obviate the need for reducing agents (103, 116). A subsequent CID step then can facilitate identification of individual peptides (116, 119). Based on our current experience, we can envision the development of more optimized approaches for determining the location of disulfide bonds in proteins in which no structural data are available.

### 3.5.7 Summary

In summary, we have used an MS-based strategy combining ETD and CID steps coupled with *ab initio* molecular modeling to elucidate the disulfide-bond map for IGFBP-5. Our results represent an extension of recent observations employing tandem MS to identify disulfide linkages in an immunoglobulin light chain (116), human growth hormone (116), and in tissue plasminogen activator (119), three proteins in which the disulfide map had been known previously. Similar combinatorial approaches that also take advantage of the rapidly improving computational landscape of molecular modeling (185, 186) should be applicable to other proteins in which the number or pattern of disulfide bonds is unknown.

### **3.6 ACKNOWLEDGEMENTS**

We thank Tony Capps, Debra McMillen, and John Klimek for experimental and technical advice, and the assistance of the OHSU Shared Proteomics Resource. This work was supported by NIH research grants 5R01 DK042748-22 (to P. R.), 5R01 EY007755-18, 5P30 EY10572 (to L. D.), and NSF grant 0746589 (to U. S.). M. N. was supported by NIH pre-doctoral fellowship F31 HL095271-03.

**Table 3.1 m/z values of IGFBP-5 cysteine-containing peptides**

(Peptides P and C are from trypsin and chymotrypsin digests, respectively.)

Cysteine Residues	Peptide No.	Peptides
Cys7/10	P1	LGSFVHCEPCDEK
Cys18/25	P2	ALSMCPPSPLGCELVK
Cys33/35/36/39/47	P3	EPGCGCCMTCALAEGQSCGVYTER
Cys80	P5	CLPR
Cys7/10 + Cys18/25 + Cys33/35/36/39/47 + Cys80	P1-P2-P3-P5	LGSFVHCEPCDEK-ALSMCPPSPLGCELVK-EPGCGCCMTCALAEGQSCGVYTER-CLPR
Cys7/10 + Cys18/25 + Cys33/35/36/39/47	P1-P2-P3	LGSFVHCEPCDEK-ALSMCPPSPLGCELVK-EPGCGCCMTCALAEGQSCGVYTER
Cys7/10 + Cys33/35/36/39/47 + Cys80	P1-P3-P5	LGSFVHCEPCDEK-EPGCGCCMTCALAEGQSCGVYTER-CLPR
Cys18/25 + Cys33/35/36/39/47 + Cys80	P2-P3-P5	ALSMCPPSPLGCELVK-EPGCGCCMTCALAEGQSCGVYTER-CLPR
Cys54	P4	CAQGLR
Cys80	P6	GVCLNEK
Cys54 + Cys80	P4-P6	QESEGGPCR-AVYLPNCDR
Cys172	P7	QESEGGPCR
Cys199	P8	AVYLPNCDR
Cys172 + Cys199	P7-P8	QESEGGPCR-AVYLPNCDR
Cys210	P9	KQCKPSR
Cys243	P11	LPGMEYVDGDFQCHAFDSSNVE
Cys221/223	P10	GICWCVDK
Cys210 + Cys221/223	P9-P10	KQCKPSR-GICWCVDK
Cys221/223 + Cys243	P10-P11	GICWCVDK-LPGMEYVDGDFQCHAFDSSNVE
Cys210 + Cys221/223 + Cys243	P9-P10-P11	KQCKPSR-GICWCVDK-LPGMEYVDGDFQCHAFDSSNVE
Cys7/10	C1	GSFVHCEPCDEKAL
Cys18	C2	CPPSPL
Cys33/35/36	C3	VKEPGCGCCM
Cys7/10 + Cys18 + Cys33/35/36	C1-C2-C3	GSFVHCEPCDEKAL_CPPSPL-VKEPGCGCCM
Cys47	C4	AEGQSCGVY
Cys80	C5	RCLPRQDEEKPL
Cys47 + Cys80	C4-C5	AEGQSCGVY-RCLPRQDEEKPL

Peptide No.	m/z								
	+1	+2	+3	+4	+5	+6	+7	+8	+9
P1	1463.8	732.3	488.5	-	-	-	-	-	-
P2	1644.8	822.9	-	-	-	-	-	-	-
P3	2467.9	1234.5	-	-	-	-	-	-	-
P5	488.3	244.6	-	-	-	-	-	-	-
P1-P2-P3-P5	6051.6	3026.3	2017.9	1513.7	1211.1	1009.4	865.4	757.3	673.3
P1-P2-P3	5566.4	2783.7	1856.1	1392.3	1114.1	928.6	796.1	-	-
P1-P3-P5	4411.8	2206.4	1471.3	1103.7	883.2	736.1	631.1	-	-
P2-P3-P5	4593.0	2297.0	1531.7	1149.0	919.4	766.3	-	-	-
P4	647.3	324.2	-	-	-	-	-	-	-
P6	762.4	381.7	-	-	-	-	-	-	-
P4-P6	1406.7	703.8	469.6	352.4	-	-	-	-	-
P7	1033.4	517.2	-	-	-	-	-	-	-
P8	1050.5	525.8	-	-	-	-	-	-	-
P7-P8	2080.9	1041.0	694.3	521.0	-	-	-	-	-
P9	846.5	423.7	282.8	-	-	-	-	-	-
P11	2460.0	1230.5	820.7	-	-	-	-	-	-
P10	923.4	462.2	-	-	-	-	-	-	-
P9-P10	1766.9	883.9	589.6	442.5	354.2	-	-	-	-
P10-P11	3380.4	1690.7	1127.5	845.9	-	-	-	-	-
P9-P10-P11	4223.9	2112.4	1408.6	1056.7	845.6	704.8	604.3	-	-
C1	1534.7	767.8	512.2	-	-	-	-	-	-
C2	613.3	307.1	-	-	-	-	-	-	-
C3	1026.4	513.7	-	-	-	-	-	-	-
C1-C2-C3	3166.3	1583.7	1056.1	792.3	634.1	528.6	453.2	-	-

**Table 3.2 CID-MS3 y and b ions for IGFBP-5 C-terminal peptides**

MS3-CID for P9-P10 disulfide-linked peptides not containing P11 and P10-P11 disulfide-linked peptides not containing P9

<u>P9-P10 disulfide-linked peptides</u>									
		y	y+2	y+3	y+4	b	b+2	b+3	b+4
	G	-	-	-	-	-	-	-	-
	I	1709.8	855.9	570.9	428.5	171.1	-	-	-
C221	C (843.45)	1596.8	799.4	533.3	400.2	1117.6	559.8	373.5	280.4
	W	650.3	326.1	-	-	1303.7	652.8	435.6	326.9
C223	C	464.2	233.1	-	-	1406.7	704.3	469.9	352.7
	V	361.2	181.6	-	-	1505.7	753.9	502.9	377.4
	D	262.1	132.1	-	-	1620.8	811.4	541.3	406.2
	K	147.1	74.6	-	-	-	-	-	-
	K	-	-	-	-	-	-	-	-
	Q	1638.8	820.4	547.3	410.7	257.2	-	-	-
C210	C (920.4)	1510.7	756.4	504.6	378.7	1280.6	641.3	427.9	321.1
	K	487.3	244.7	-	-	1408.7	705.3	470.6	353.2
	P	359.2	180.6	-	-	1505.7	753.9	502.9	377.4
	S	262.2	132.1	-	-	1592.8	797.4	531.9	399.2
	R	175.1	88.6	-	-	-	-	-	-
<u>P10-P11 disulfide-linked peptides</u>									
		y	y+2	y+3	y+4	b	b+2	b+3	b+4
	G	-	-	-	-	-	-	-	-
	I	3323.4	1662.7	1108.8	831.8	171.1	-	-	-
C221	C	3210.3	1606.2	1071.1	803.6	274.1	138.1	92.4	69.5
	W	3107.3	1554.7	1036.8	777.8	460.2	231.1	154.4	116.1
C223	C (2457.00)	2921.2	1461.6	974.7	731.3	3020.2	1511.1	1007.7	756.1
	V	361.2	181.6	-	-	2119.3	1060.6	707.4	530.8
	D	262.1	132.1	-	-	3234.3	1618.2	1079.1	809.6
	K	147.1	74.6	-	-	-	-	-	-
	L	-	-	-	-	-	-	-	-
	P	3267.3	1634.7	1090.1	-	211.1	-	-	-
	G	3170.3	1586.1	1057.8	-	268.2	-	-	-
	M	3113.3	1557.6	1038.8	-	399.2	-	-	-
	E	2982.2	1492.1	995.1	-	528.3	-	-	-
	Y	2853.2	1427.6	952.1	-	691.3	-	-	-
	V	2690.1	1346.1	897.7	-	790.4	-	-	-
	D	2591.0	1296.5	864.7	-	905.4	-	-	-
	G	2476.0	1239.0	826.3	-	962.4	-	-	-
	D	2419.0	1210.5	807.3	-	1077.5	-	-	-
	F	2303.0	1152.5	768.7	-	1224.5	-	-	-
	Q	2156.9	1079.5	720.0	-	1352.6	-	-	-
C243	C (920.4)	2028.8	1015.4	677.3	-	2376.0	1189.0	793.0	595.0
	H	1005.4	-	-	-	2513.1	1257.5	838.7	629.3
	A	868.4	-	-	-	2584.1	1293.0	862.4	647.0
	F	797.3	-	-	-	2731.2	1366.6	911.4	683.8
	D	650.3	-	-	-	2846.2	1424.1	949.7	712.5
	S	535.2	-	-	-	2933.2	1467.6	978.7	734.3
	S	448.2	-	-	-	3020.3	1511.1	1007.8	756.1
	N	361.2	-	-	-	3134.3	1568.1	1045.8	784.6
	V	247.1	-	-	-	3233.4	1617.7	1078.8	809.3
	E	148.1	-	-	-	-	-	-	-

Values in ( ) represent peptide masses disulfide-linked to specific cysteine residues.  
 + represents remainder of peptides for which ion series cannot be generated.

**Table 3.3** CID-MS2 y and b ions for IGFBP-5 C-terminal peptides

MS2-CID for P7-P8 disulfide-linked peptides							
		y	y+2	y+3	b	b+2	b+3
	Q	-	-	-	-	-	-
	E	1952.9	977.4	652.0	258.1	-	-
	S	1823.8	912.9	608.9	345.1	-	-
	E	1736.8	869.4	579.9	474.2	-	-
	Q	1607.8	804.9	536.9	602.2	-	-
	G	1479.7	740.9	494.2	659.3	-	-
	P	1422.7	712.3	475.2	756.3	-	-
C172	C(1047.50)	1325.6	663.8	442.9	1906.8	954.4	636.6
	R	175.1	88.6	-	-	-	-
	A	-	-	-	-	-	-
	V	2009.9	1005.9	671.0	171.1	-	-
	Y	1910.8	956.4	637.9	334.2	-	-
	L	1747.8	874.9	583.6	447.3	-	-
	P	1634.7	818.3	545.9	544.3	-	-
	N	1537.6	769.8	513.5	658.4	-	-
C199	C(1030.42)	1423.6	712.8	475.5	1791.8	896.9	598.3
	D	290.1	146.1	-	1906.8	954.4	636.6
	R	175.12	88.6	-	-	-	-

Values in ( ) represent peptide masses disulfide-linked to specific cysteine residues.

**Table 3.4 y and b ions for establishing Cys<sup>47</sup>-Cys<sup>60</sup> disulfide-linkage**

MS2-CID for C4-C5 disulfide-linked peptides											
		y	y+2	y+3	y+4	y+5	b	b+2	b+3	b+4	b+5
	A	-	-	-	-	-	-	-	-	-	-
	E	2323.1	1162.6	775.4	581.8	465.6	201.1	-	-	-	-
	G	2194.1	1098.0	732.4	549.5	439.8	258.1	-	-	-	-
	Q	2137.1	1069.5	713.4	535.3	428.4	386.2	-	-	-	-
	S	2009.0	1005.5	670.7	503.3	402.8	473.2	-	-	-	-
C47	C (1480.8)	1922.0	962.0	641.7	481.5	385.4	2057.0	1029.5	686.7	515.3	412.4
	G	338.2	-	-	-	-	2114.0	1058.0	705.7	529.5	423.8
	V	281.2	-	-	-	-	2213.1	1107.5	738.7	554.3	443.6
	Y	182.1	-	-	-	-	-	-	-	-	-
	R	-	-	-	-	-	-	-	-	-	-
C60	C (910.4)	2238.1	1120.0	747.0	560.5	448.6	1170.5	586.3	391.2	293.6	235.1
	L	1224.7	613.3	409.2	-	-	1283.6	642.8	428.9	321.9	257.7
	P	1111.6	556.8	371.5	-	-	1380.7	691.3	461.2	346.2	277.1
	R	1014.5	508.3	339.2	-	-	1536.8	769.4	513.3	385.2	308.4
	Q	858.4	430.2	-	-	-	1664.8	833.4	555.9	417.2	334.0
	D	730.4	366.2	-	-	-	1779.8	890.9	594.3	446.0	357.0
	E	615.3	308.7	-	-	-	1908.9	955.4	637.3	478.2	382.8
	E	486.3	244.1	-	-	-	2037.9	1020.0	680.3	510.5	408.6
	K	357.3	179.6	-	-	-	2166.0	1084.0	723.0	542.5	434.2
	P	229.2	-	-	-	-	2263.1	1132.5	755.4	566.8	453.6
	L	132.1	-	-	-	-	-	-	-	-	-

**Table 3.5** y and b ions for establishing linkages between N-terminal chymotrypsin peptides.

MS2-CID for C1-C2-C3 disulfide-linked peptides														
		y	y+2	y+3	y+4	y+5	y+6	b	b+2	b+3	b+4	b+5	b+6	
C18	C (2553)	-	-	-	-	-	-	-	-	-	-	-	-	
	P	510.3	256.1	-	-	-	-	2754.1	1378.0	919.0	689.5	551.8	460.0	
	P	413.2	207.6	-	-	-	-	2851.1	1426.6	951.4	713.8	571.2	476.2	
	S	316.2	159.1	-	-	-	-	2938.2	1470.1	980.4	735.5	588.6	490.7	
	P	229.2	115.6	-	-	-	-	3035.2	1518.6	1012.7	759.8	608.0	506.9	
	L	132.1	67.1	-	-	-	-	-	-	-	-	-	-	
	G	-	-	-	-	-	-	-	-	-	-	-	-	
	S	3109.3	1555.7	1037.4	778.3	622.9	519.2	145.1	-	-	-	-	-	
	F	3022.3	1512.2	1008.4	756.6	605.5	504.7	292.1	-	-	-	-	-	
	V	2875.2	1438.6	959.4	719.8	576.0	480.2	391.2	-	-	-	-	-	
	H	2776.2	1389.1	926.4	695.0	556.2	463.7	528.3	-	-	-	-	-	
	C7	C (1631.7)	2639.1	1320.6	880.7	660.8	528.8	440.9	-	-	-	-	-	
	C10	C (1631.7)	2310.0	1156.0	771.0	578.5	463.0	-	2592.1	1297.0	865.0	649.0	519.4	433.0
		D	575.3	288.7	-	-	-	-	2707.1	1354.5	903.4	677.8	542.4	452.2
E		460.3	231.1	-	-	-	-	2836.1	1419.1	946.4	710.0	568.2	473.7	
K		331.2	166.6	-	-	-	-	2964.2	1483.1	989.1	742.1	593.8	495.0	
A		203.1	102.6	-	-	-	-	3035.3	1518.6	1012.8	759.8	608.1	506.9	
L		132.1	67.1	-	-	-	-	-	-	-	-	-	-	
V		-	-	-	-	-	-	-	-	-	-	-	-	
K		3067.3	1534.7	1023.4	767.8	614.5	512.2	228.2	-	-	-	-	-	
E		2939.2	1470.6	980.7	735.8	588.8	490.9	357.2	-	-	-	-	-	
P		2810.2	1406.1	937.7	703.5	563.0	469.4	454.3	-	-	-	-	-	
G		2713.1	1357.6	905.4	679.3	543.6	453.2	511.3	-	-	-	-	-	
C33		C (1529.7)	2656.1	1329.1	886.4	665.0	532.2	443.7	-	-	-	-	-	
C35		C	-	-	-	-	-	-	-	-	-	-	-	
C36		C (610.3)	863.4	432.7	-	-	-	-	-	-	-	-	-	
	M	150.1	-	-	-	-	-	-	-	-	-	-		

**Table 3.6** y and b ions for establishing Cys<sup>54</sup>-Cys<sup>80</sup> disulfide-linkage

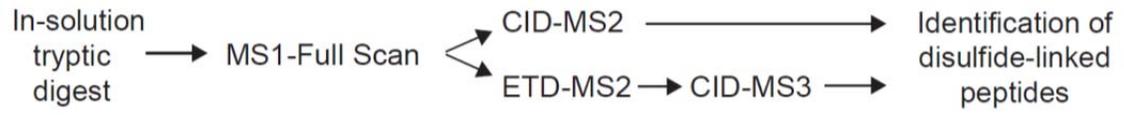
MS2-CID for P4-P6 disulfide-linked peptides							
		y	y+2	y+3	b	b+2	b+3
C54	C(759.4)	-	-	-	-	-	-
	A	544.3	273.2	182.4	934.5	468.2	312.5
	Q	473.3	237.6	158.8	1062.5	532.3	355.2
	G	345.2	173.6	116.1	1119.5	560.8	374.2
	L	288.2	145.1	97.1	1232.6	617.3	411.9
	R	175.1	88.6	59.4	-	-	-
C80	G	-	-	-	-	-	-
	V	1349.7	675.8	450.9	157.1	79.5	53.4
	C(644.3)	1250.6	626.3	417.9	904.4	453.2	302.5
	L	503.3	252.6	168.8	1017.5	509.7	340.2
	N	390.2	196.1	131.1	1131.5	566.8	378.2
	E	276.2	139.1	93.1	1260.6	-	-
	K	147.1	74.6	50.0	-	-	-

Values in ( ) represent peptide masses disulfide-linked to specific cysteine residues.

**Figure 3.1 Experimental plan for identifying the disulfide bonds of IGFBP-5.**

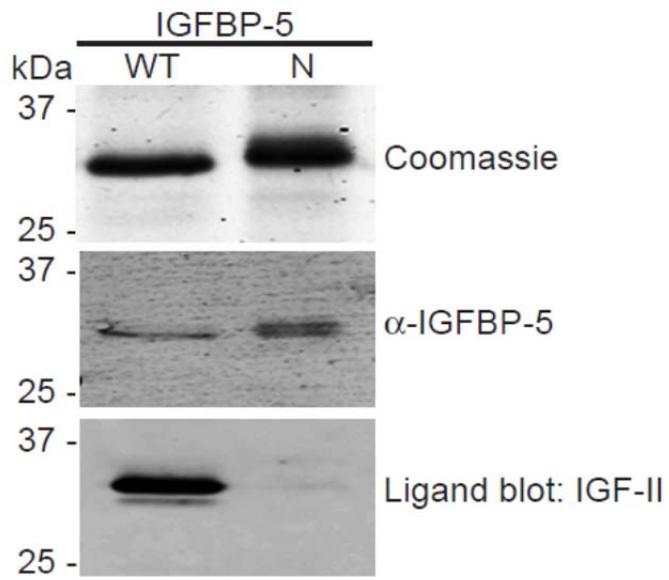
Analytical approach for identification of disulfide linkages in IGFBP-5 by MS. See Sections 3.3.4 and 3.3.5 for details.

**Figure 3.1**



**Figure 3.2 Purification and characterization of IGFBP-5.** Analysis of purified wild type and N-terminal mutant IGFBP-5 after non-reducing SDS-PAGE by staining with Coomassie blue (top), immunoblotting with anti-IGFBP-5 antibody (middle), and ligand blotting with biotinylated human IGF-II (bottom).

**Figure 3.2**



**Figure 3.3** **Expected cysteine-containing peptides of IGFBP-5.** List of the 11-tryptic peptides (P1-P11) in mouse IGFBP-5 that contain cysteine residues (cysteines are in **bold** script) along with their location in the N- or C-terminal domain of the protein. Numbers in parentheses denote the location of each cysteine in the amino acid sequence.

**Figure 3.3**

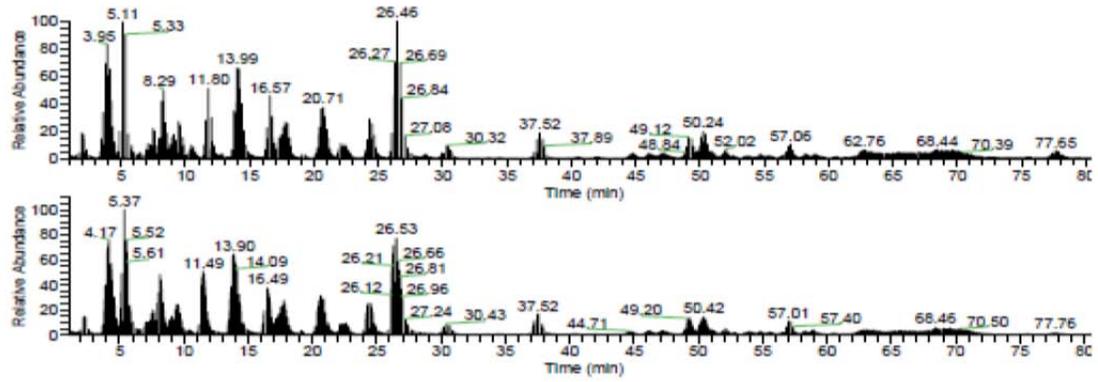
Expected Cysteine-Containing Tryptic Peptides

N	P1:	LGSFVHCEPCDEK	(7, 10)
	P2:	ALSMCPPSPLGCELVK	(18, 25)
	P3:	EPGCGCCMTCALAEQGSCGVYTER	(33, 35, 36, 39, 47)
	P4:	CAQGLR	(54)
	P5:	CLPR	(60)
	P6:	GVCLNEK	(80)
C	P7:	QESEQGPCR	(172)
	P8:	AVYLPNCDR	(199)
	P9:	R/KQCKPSR	(210)
	P10:	GICWCVDK	(221, 223)
	P11:	LPGMEYVDGDFQCHAFDSSNVE	(243)

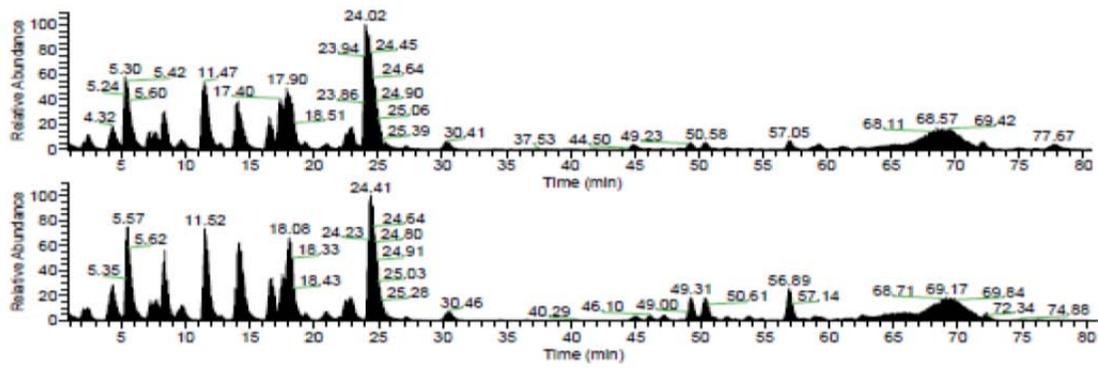
**Figure 3.4** Analysis of IGFBP-5 tryptic peptides by MS. Elution profiles after separation by liquid chromatography for 0 to 80 min of peptides/ions derived from trypsin digestion of wild type (top) or N-mutant IGFBP-5 (bottom). Results of two independent experiments for each IGFBP-5 species are shown and demonstrate the reproducibility of each protein's tryptic peptide profile.

Figure 3.4

IGFBP-5 WT



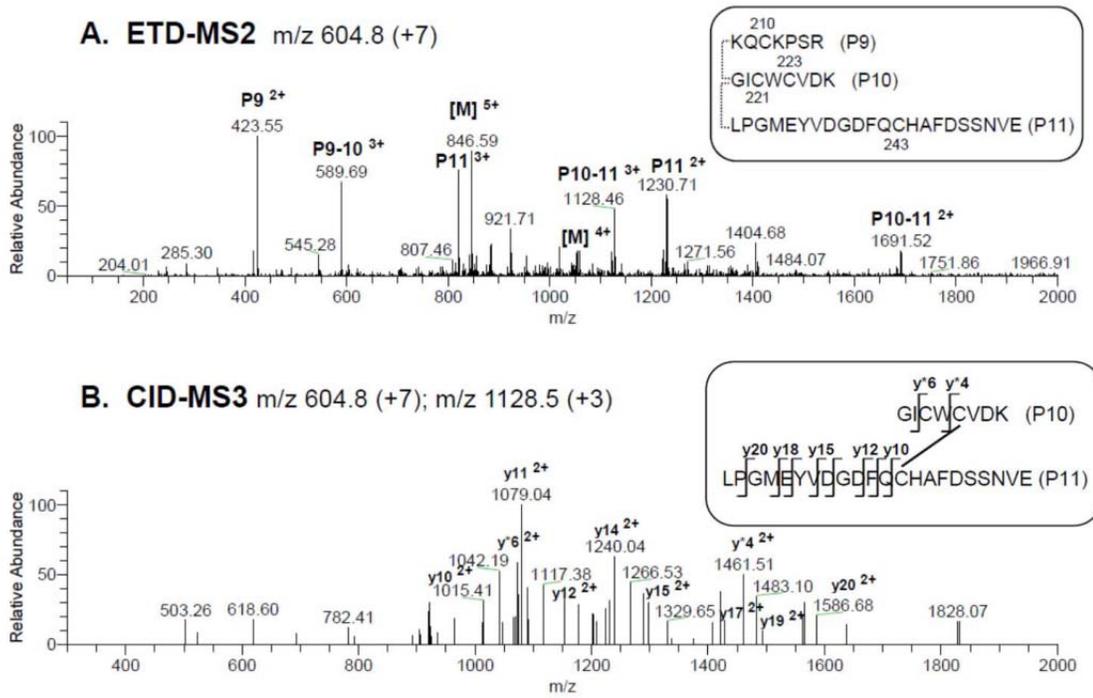
IGFBP-5 N



**Figure 3.5 Defining disulfide linkages in the C-terminal segment of IGFBP-5**

(I). Results demonstrate that peptides P9-P10-P11 are linked by disulfide bonds. See Table 3.1 for additional information. **A, B.** Analysis of P9-P10-P11. **A.** ETD of precursor ion,  $m/z$  604.8 (+7), corresponding to P9-P10-P11, and containing cysteine residues 210, 221, 223, and 243. **B.** CID of precursor ion,  $m/z$  1128.5 (+3), of disulfide-linked peptides P10 and P11. The y ions marked with an asterisk are from P10; other y ions are from P11. In **A**, [M] denotes the parent mass. Additional y ions were identified in **B**, but have been removed from the graphs for clarity.

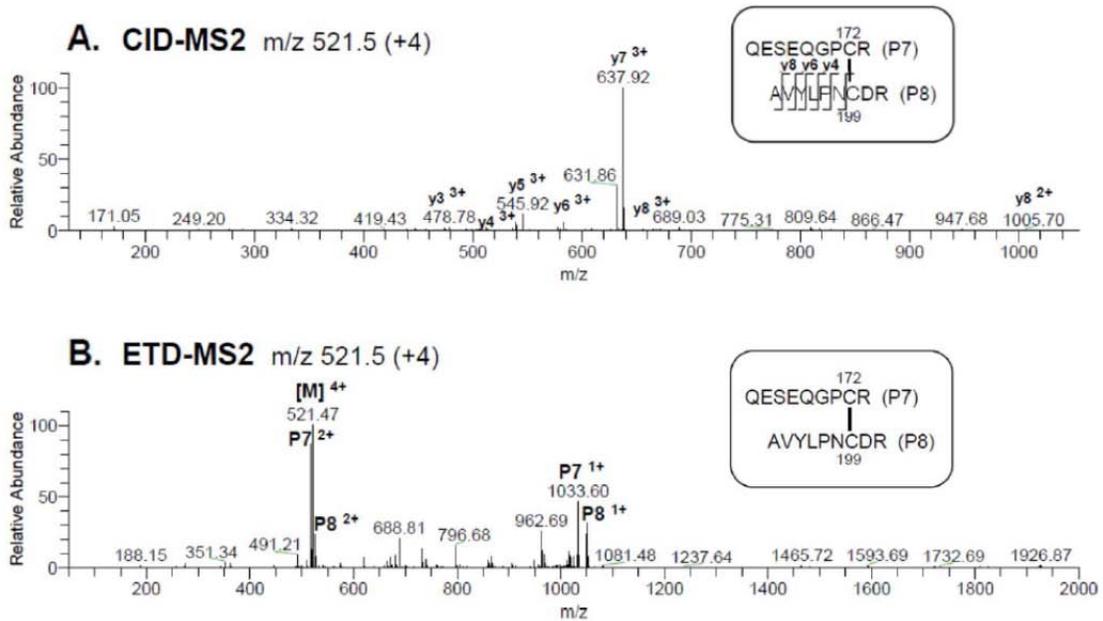
**Figure 3.5**



**Figure 3.6 Defining disulfide linkages in the C-terminal segment of IGFBP-5**

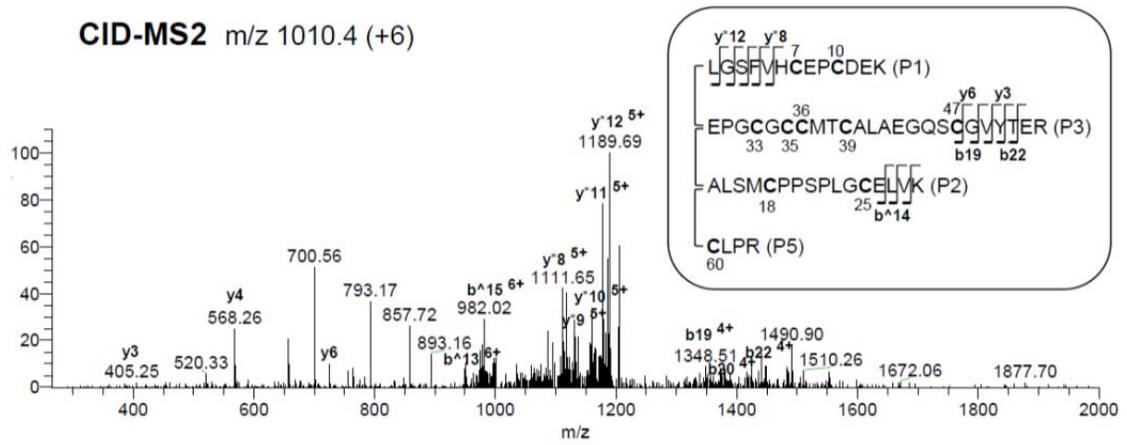
**(II).** P7-P8 are linked by disulfide bonds. See Table 3.1 for additional information. **A,**  
**B.** Analysis of P7-P8. **A.** CID of precursor ion  $m/z$  521.5 (+4) of disulfide-linked  
peptides containing cysteine residues 172 and 199. The y ions are from P8. **B.** ETD of  
precursor ion  $m/z$  521.5 (+4). Additional y ions were identified in **A**, but have been  
removed from the graphs for clarity. In **B**, [M] denotes the parent mass.

Figure 3.6



**Figure 3.7 Cysteines in the N-terminal domain of IGFBP-5 are located in highly intertwined peptides.** CID-MS2 of precursor ion, m/z 1010.4 (+6), demonstrates that P1-P2-P3-P5, containing 10 cysteine residues, are linked by disulfide bonds. Ions marked with \* and ^ are from P1 and P2, respectively. Other ions are from P3.

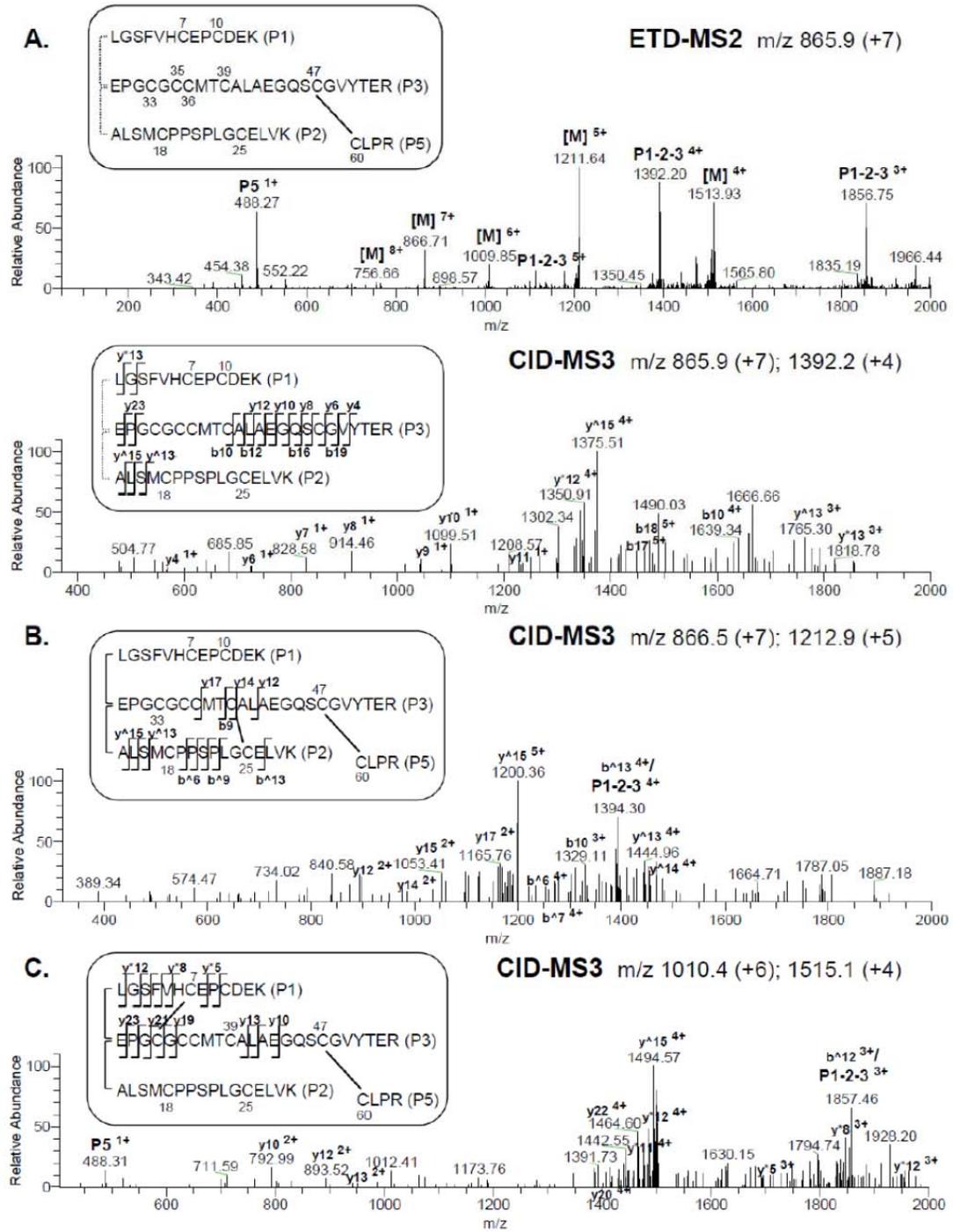
Figure 3.7



**Figure 3.8 Identifying disulfide bonds within the N-terminal domain of IGFBP-5.**

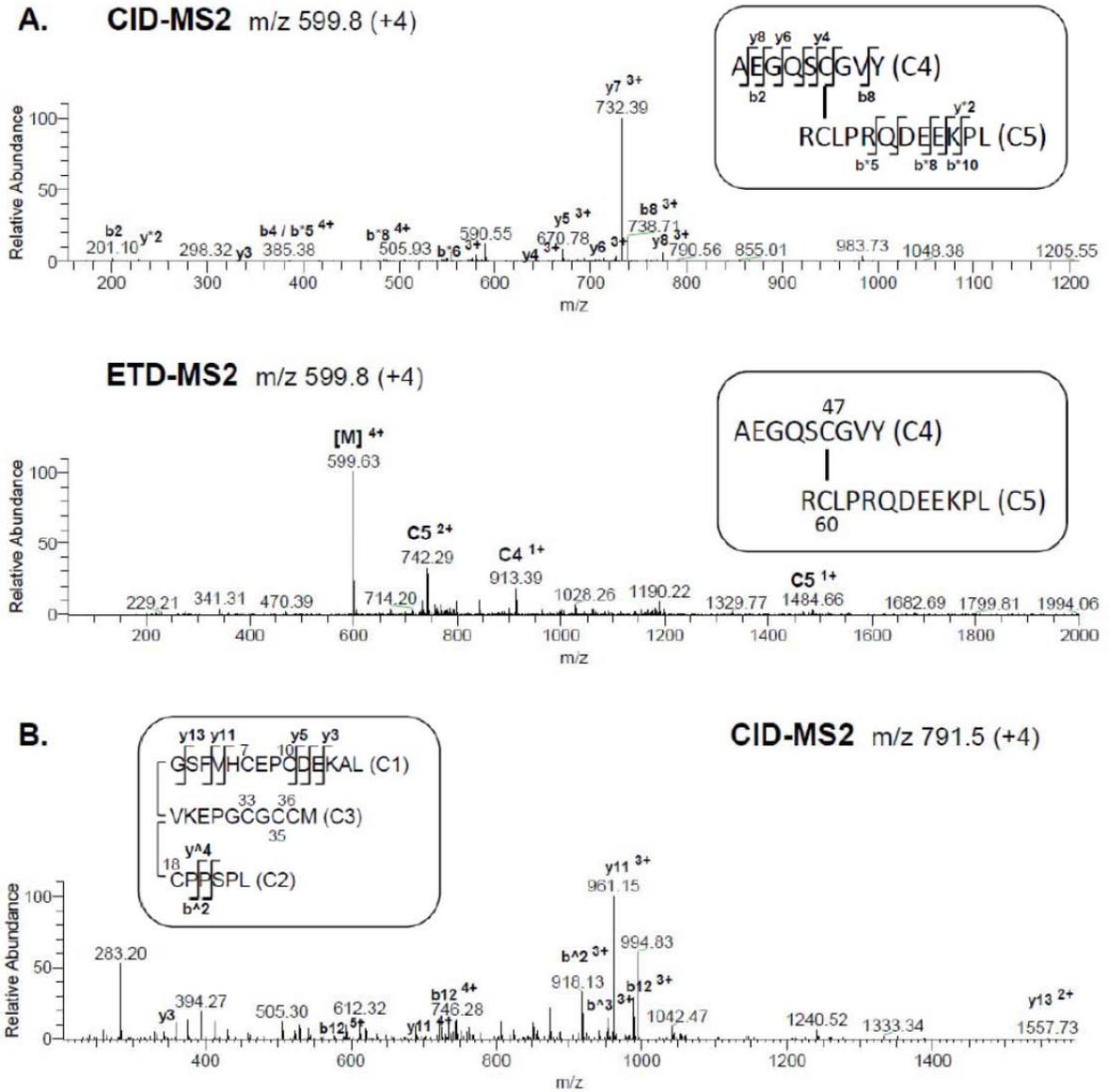
Cysteines 47 and 60 are linked. Top: ETD-MS2 of precursor ion,  $m/z$  865.9 (+7), corresponding to disulfide-linked peptides P1, P2, P3 and P5. [M] denotes the parent mass. Bottom: CID-MS3 of precursor ion,  $m/z$  1392.2 (+4), of disulfide-linked peptides P1, P2, and P3 (P5 has been dissociated from the other peptides). **B.** Cysteines 25 and 39 are linked. CID-MS3 of precursor ion 1212.9 (+5) containing linked peptides P2 and P3. **C.** Cysteines 7 and 33 are linked. CID-MS3 of precursor ion 1515.1 (+4) containing linked peptides P1 and P3. For **A-C**,  $y^*$  and  $b^*$  ions are from P1,  $y^\wedge$  and  $b^\wedge$  ions are from P2, and  $y$  and  $b$  ions are from P3 peptides. In CID spectra, additional  $y$  and  $b$  ions were identified, but were removed from the figures for clarity.

Figure 3.8



**Figure 3.9** Alternative steps to identify disulfide linkages in the N-terminal domain of IGFBP-5. Analysis of N-terminal peptides generated by chymotrypsin (C). **A.** CID-MS2 (top) and ETD-MS2 (bottom) of precursor ion, m/z 599.8 (+4), identifying the disulfide-linkage between Cys47 and Cys60 in peptides C4 and C5, respectively. [M] denotes the parent mass. Ions marked with \* are from C5; other ions are from C4. **B.** CID-MS2 of precursor ion, m/z 791.5 (+4) corresponding to peptides C1-C2-C3. Ions marked with ^ are from C2; other ions are from C1.

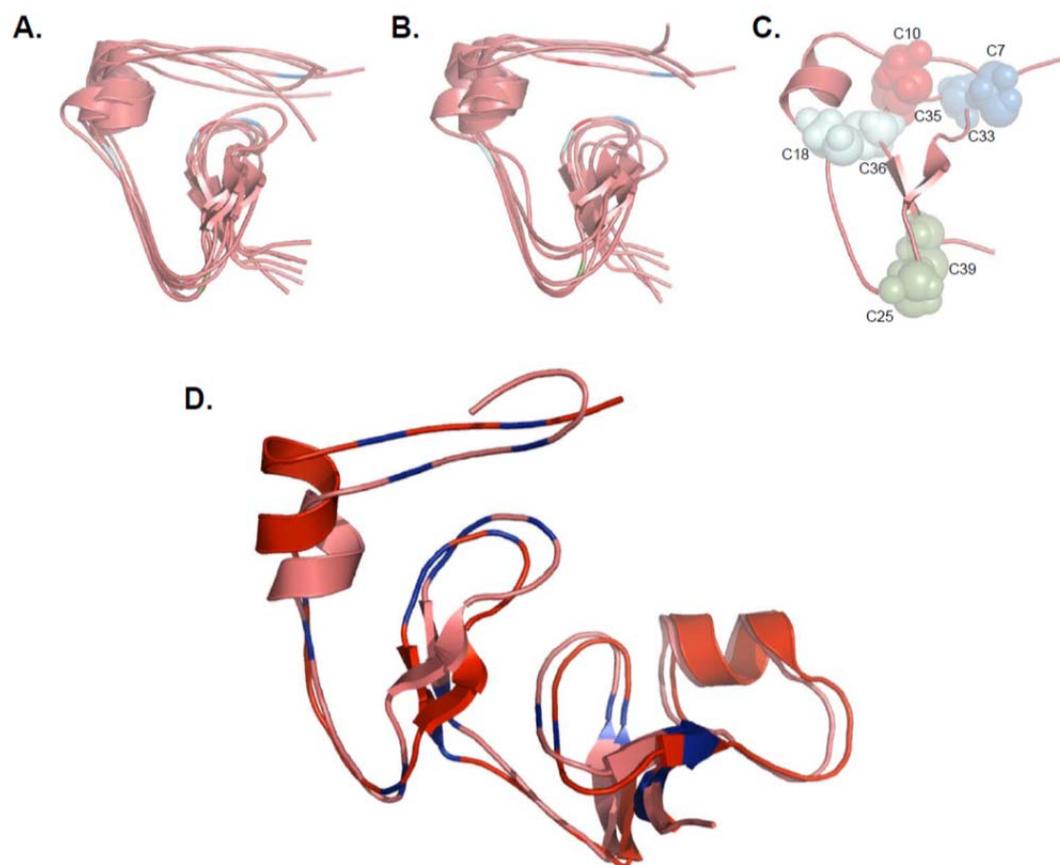
Figure 3.9



**Figure 3.10 Structural predictions and disulfide bond map of the N-terminal**

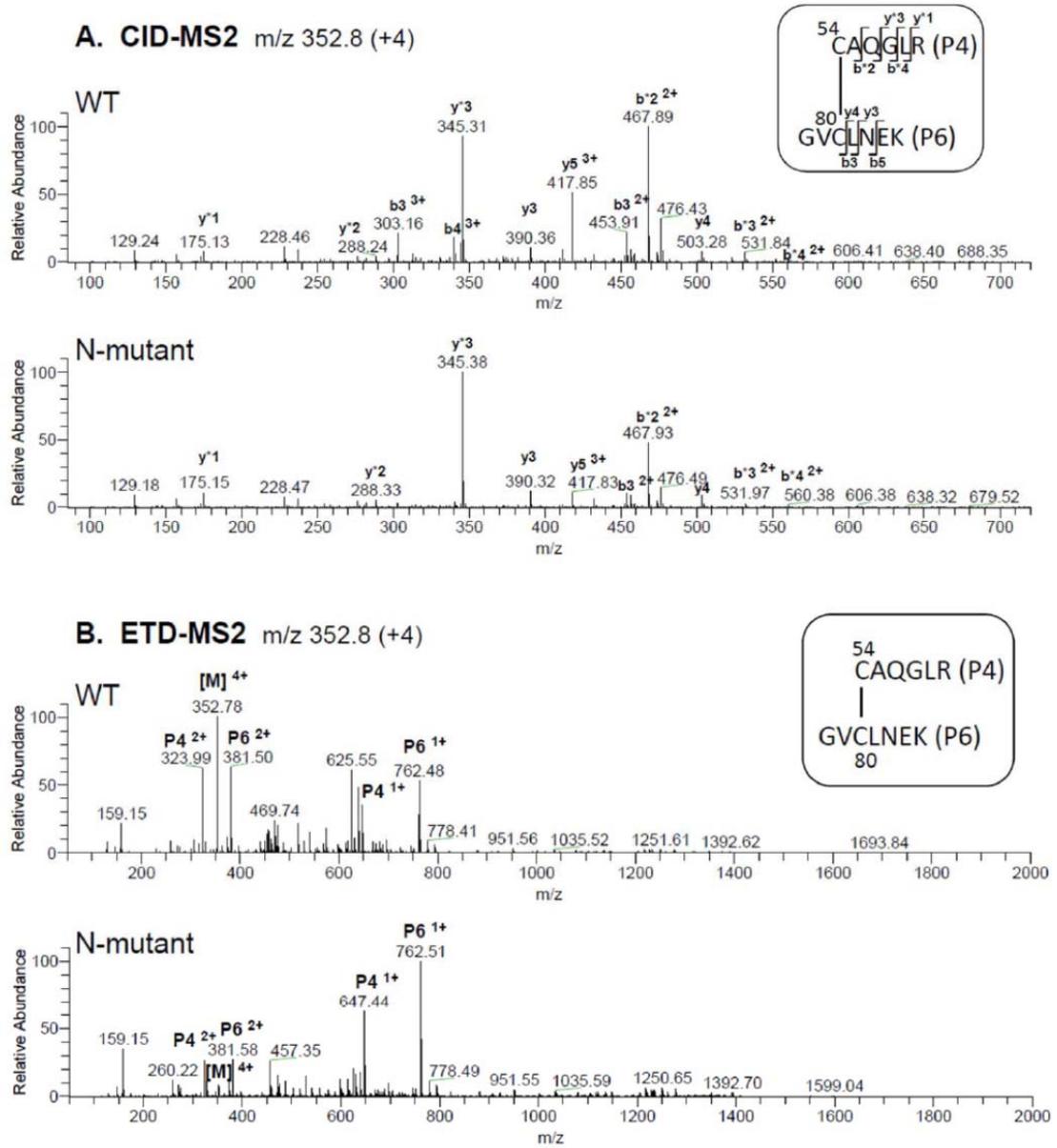
**domain of IGFBP-5.** **A.** *De novo ab initio* molecular modeling of amino acids 5 - 41 of the N-terminal domain of IGFBP-5. **B.** Homology-based *ab initio* molecular modeling of amino acids 5 - 41 of the N-terminal domain of IGFBP-5. In **A** and **B**, the centers of clusters of the largest groups of highly related structures are represented by individual lines and the location of cysteine residues is indicated with the same color scheme as in **C**. **C.** Location of disulfide bonds involving cysteines found within the conserved G<sup>32</sup>CGCCXXC<sup>39</sup> motif based on predictions in **A** and **B**. **D.** *De novo ab initio* model of the complete IGFBP-5 N-terminal domain (amino acids 1 - 84) aligned with a model of IGFBP-5 (residues 5 - 84) based on the x-ray crystallographic structure of the N-terminal segment of IGFBP-4. Cysteines are in blue. Note that both models are nearly identical.

**Figure 3.10**



**Figure 3.11 Amino acid substitution mutations in the N-terminal domain of IGFBP-5 do not alter disulfide-linkages.** **A.** CID-MS2 of precursor ions,  $m/z$  352.8 (+4), of disulfide-linked peptides P4-P6 from wild type (top) and N-terminal mutant (bottom) IGFBP-5 containing cysteine residues 54 and 80. The y and b ions marked with an asterisk are from P4; other y and b ions are from P6. Additional y and b ions were identified, but have been removed from the graph for clarity. **B.** ETD-MS2 spectra of precursor ions,  $m/z$  352.8 (+4) from wild type (top) and N-terminal mutant (bottom) IGFBP-5 containing cysteine residues 54 and 80. [M] represents parent mass.

**Figure 3.11**



## Chapter 4

### **Disulfide Mapping of Repulsive Guidance Molecule c / Hemojuvelin by LC-MS Reveals Two Oxidation States**

Mahta Nili, Ujwal Shinde, Larry David, and Peter Rotwein

Department of Biochemistry and Molecular Biology, Oregon Health & Science  
University, Portland, OR 97239-3098, USA

MS experiment design, execution, and data analysis reported in this chapter were performed by the author of this dissertation. U. Shinde performed the *ab initio* modeling. L. David aided in the design and execution of the MS experiments. P. Rotwein aided in the design of experiments and in the planning of the manuscript. (This work is in preparation).

## 4.1 SUMMARY

The aggressive early onset iron overload disorder, juvenile hemochromatosis (JH), is caused by inactivating mutations in RGMc/HJV. The complex biosynthesis and processing of RGMc/HJV gives rise to four distinct isoforms. Two soluble proteins are released from the cell, while two other isoforms, a full-length single-chain and a two-chain species, are attached to the cell membrane via a GPI-linkage. We have previously shown that single-chain RGMc preferentially binds members of the BMP family of proteins, while two-chain RGMc binds neogenin. No structural data is currently available for any RGMc isoforms. Here, we identify key disulfide bonds of RGMc using a recently established MS3 approach combining ETD and CID, and provide the first use of this methodology to assign unknown linkages. Our results indicate that RGMc exists as two distinct oxidation states. In one state, we find disulfide linkages connecting Cys<sup>160</sup>-Cys<sup>310</sup> and Cys<sup>178</sup>-Cys<sup>314</sup>, while in the second, Cys<sup>160</sup> is bonded to Cys<sup>178</sup>. In addition, disulfide mapping of the JH-associated RGMc truncation mutant, RGMc141X, in combination with the full-length protein, established a linkage connecting Cys<sup>39</sup>-Cys<sup>82</sup> in the N-terminus of the protein. Finally, via *ab initio* molecular modeling we provide putative RGMc structures for the N-terminal domain, and the two different oxidation states, that can represent single- and two-chain RGMc.

## 4.2 INTRODUCTION

The complex biosynthesis and processing of RGMc gives rise to four distinct isoforms. Two soluble species, a full-length 50 kDa and 40 kDa PC-cleavage product, are found in the circulation; two isoforms, single- and two-chain RGMc proteins, are linked to the cell membrane via GPI-linkage. Soluble RGMc isoforms directly bind to BMPs and negatively regulate their function (20, 68, 136). It has been reported that RGMc isoforms on the cell surface function as BMP co-receptors that enhance BMP-mediated actions (27, 34, 86). However, it has not been reported if both single- and two-chain RGMc carry out this role.

RGMc and the two other RGM family members, RGMa and RGMb, possess sites of internal cleavage (FGDP), and in RGMc this processing site resides between residues 163 and 166 (51). N-terminal sequencing established a proline residue is the first amino acid of the second chain, following cleavage of RGMa (51), but similar studies have not been performed for RGMc. It is not known what processing event gives rise to the two-chain RGM isoforms. It has been proposed that RGMs undergoes acid-based hydrolysis at this site (66), similar to a processing event that the protein mucin undergoes (187); however, this has not been demonstrated.

Two-chain RGMc binds to the transmembrane protein neogenin, but does not bind to members of the BMP family (68). Through its interaction with neogenin, RGMa has been shown to regulate the repulsive guidance of retinal axons and promote neuronal survival (62, 64). It has been proposed that binding to neogenin is necessary for RGMc

to be released from the cell surface (65), but no evidence supporting this hypothesis has been reported. Selected RGMc mutants, such as D165E and G313V, do not give rise to the two-chain isoform and do not bind to neogenin (68). However, a Gly to Val mutation in the RGD domain of RGMc, which does not compromise expression of the two-chain isoform, is tolerated for binding to neogenin (68).

No structural data has been published for RGMc, therefore it is not known if different RGMc isoforms possess different final protein conformations. Recently, the Karger laboratory published a novel MS3 methodology combining ETD and CID to map protein disulfide bonding patterns (116). ETD can induce breakage of disulfide bonds, thereby mapping studies can be performed with non-reduced proteins samples (103, 110, 116). CID following ETD activation allows for identification of peptides after disulfide linkages are dissociated (104). This method was validated by establishing the disulfide bonds of several proteins with resolved structures (116, 119).

In this study, we use an MS3 approach to map the disulfide bonds of RGMc. We also map the disulfide linkages in an RGMc truncation mutant that represents the first ~110 amino acids of the mature protein. Our results establish a key disulfide bond in the N-terminus of RGMc that is identified in both the full-length protein and the truncation mutant. We also identify two distinct RGMc oxidation states. *Ab initio* models support the existence of these two different oxidation forms of RGMc, which allow us to present putative distinct structures for single- and two-chain RGMc.

## **4.3 MATERIALS and METHODS**

### **4.3.1 Materials**

Dulbecco's modified Eagle's medium (DMEM), fetal bovine serum (FBS), ultralow IgG FBS, phosphate-buffered saline (PBS), Protein A Sepharose 4B, and trypsin/EDTA were purchased from Invitrogen (Carlsbad, CA). TransIT-LT1 was from Mirus Bio (Madison, WI). Proteomics grade trypsin and iodoacetamide was from Sigma-Aldrich (St. Louis, MO); Criterion precast gels were purchased from BioRad (Hercules, CA). The furin convertase inhibitor decanoyl-RVKR-CMK was purchased from Alexis Biochemicals (San Diego, CA); GelCode Blue Stain Reagent was purchased from Pierce Biotechnologies (Rockford, IL). Other chemicals and reagents were purchased from commercial suppliers.

### **4.3.2 Generation, Expression, and Purification of RGMc-IgG1-Fc Fusion Proteins**

The generation, expression, and purification of RGMc50- and RGMc141X-IgG1-Fc fusion proteins used in this study are described in detail in Sections 2.3.2 and 2.3.3.

### **4.3.3 Digestion of RGMc**

RGMc (8  $\mu$ g) was dried down by vacuum centrifugation and resuspended in buffer containing 8M urea, 1M Tris, pH 8.5, and 8 mM calcium chloride. Free cysteines were either un-modified or alkylated by the addition of 20 mM iodoacetamide for 15 min protected from light at 20°C. Samples were diluted 4-fold, and trypsin (0.32  $\mu$ g) was added and allowed to digest for 16 hr at 37°C with shaking. Digestions were stopped by addition of 2  $\mu$ l of 88% formic acid and were stored at -20°C until analysis.

#### **4.3.4 Localization of Disulfide Bonds by Mass Spectrometry**

Methods used to establish the disulfide linkages in this study are described in detail in Section 3.3.5.

#### **4.3.5 Modeling of RGMc**

*Ab initio* RGMc models were generated using Rosetta (120, 123). In the first step, 1000 independent structures were predicted prepared with the Robetta Fragment server (120, 123, 188). Structures were clustered for similarity based on their root mean square deviations. Representative models of the N-terminal domain (residues 32-141) and the central region of the protein (a.a. 142-319) that contained the disulfide linkages established in this study were selected.

## 4.4 RESULTS

### 4.4.1 MS Experimental Approach

RGMc contains 14 cysteines and the disulfide linkages between them have not been established. In this study, we applied a tandem MS approach combining ETD and CID to identify the disulfide bonds in full-length RGMc and in a truncation mutant, RGMc141X, which contains the first seven cysteines. For the purposes of this study, we consider RGMc141X to represent most of the N-terminal segment of the full-length protein. Purified RGMc proteins were digested with trypsin under non-reducing conditions and subjected to a MS3 protocol with ETD (ETD-MS2) in the second MS scan followed by CID (CID-MS3) in the third and final data-dependent scan (Figure 4.1).

The 14 cysteines of RGMc reside within 13 tryptic peptides. The disulfide-linked tryptic peptides that we observed in this study are listed in Figure 4.2, and their m/z values are displayed in Table 4.1. Of the peptides listed, only P1 and P2 are found in RGMc141X (Figure 4.2).

### 4.4.2 Disulfide mapping and structure prediction of the N-terminus of RGMc

ETD was used to fragment the linked tryptic peptides, P1 and P2, containing cysteines 39 and 82, respectively, in both full-length RGMc and RGMc141X (Figure 4.3). The fragmented ion, m/z 812.4 (+1), corresponds to P2; P1 is represented by 1242.5 (+1) (Figure 4.3). Further dissociation of these ions by CID allowed for identification of the individual peptides (data not shown).

*Ab initio* modeling was performed with the N-terminal residues (a.a. 33-141) of RGMc, which also represents the full sequence of the truncation mutant RGMc141X

(Figure 4.4). In the model, both Cys<sup>39</sup> and Cys<sup>82</sup> are located at the ends of  $\alpha$ -helices that are connected to one another through the disulfide linkage between these cysteines. The model indicates that a potential disulfide linkage connects cysteines 34 and 90; however this has not been confirmed with MS mapping studies. The three remaining cysteines do not participate in disulfide linkages.

#### 4.4.3 The two oxidation states of RGMc

Full-length RGMc exists as either a single- or two-chain species (58, 66). To date, no differential processing pathways or variations in structure have been reported to explain the biogenesis of these individual isoforms. Here we used our MS approach and identified two distinct oxidation states within RGMc (Figures 4.5 and 4.6). Our results were further validated with *ab initio* models (Figure 4.7) which demonstrate that both oxidation states are conformations that the protein can adopt.

ETD fragmented the disulfide-linked peptides P3-P4-P5,  $m/z$  728.2 (+7), to generate ions that represent P3 or P4 alone or linked to P5 (Figure 4.5A and Table 4.1). P5 contains two cysteines and therefore can form two disulfide bonds (Figure 4.2 and 4.5). Further dissociation of the fragment ion,  $m/z$  1127.3 (+3), by CID establishes not only that P4 or P5 are linked but also demonstrates that Cys<sup>178</sup> is linked to Cys<sup>314</sup>. Key ions that establish this linkage are  $y^7$ ,  $y^8$ ,  $y^9$ , and  $b^{12}$  (Figure 4.5B and Table 4.2). The fact that we can clearly assign this disulfide bond between Cys<sup>178</sup> and Cys<sup>314</sup> also indicates that Cys<sup>160</sup> and Cys<sup>310</sup> in P3 and P5, respectively, must be linked. We provide further support for this linkage with CID data dissociating the precursor ion,  $m/z$  1272.5 (+3) (Figure 4.5C and Tables 4.1 and 4.3).

In the second RGMc oxidation state identified, ETD was used to fragment the precursor ion,  $m/z$  598.0 (+5), that corresponds to the disulfide-linked peptides P3 and P4 (Figure 4.6A and Table 4.1). In this oxidation state, Cys<sup>160</sup> is linked to Cys<sup>178</sup>. The identity of the dissociated peptides are established by CID (Figure 4.6 A, B and Table 4.4).

*Ab initio* modeling was performed to establish whether both oxidation states identified by MS resulted in feasible protein structures. The models were built with amino acids 142-319, which contains 5 cysteine residues (Figure 4.7). In a model representing the first oxidation state, the disulfide linkages Cys<sup>160</sup>-Cys<sup>310</sup> and Cys<sup>178</sup>-Cys<sup>314</sup> are located very close to one another, and may be forming a small cysteine-knot. The cleavage site necessary for generating two-chain RGMc, which is located at position D165E, appears to be inaccessible in this oxidation state. We therefore postulate that this model represents single-chain RGMc (Figure 4.7A). In a model of the second oxidation state, with Cys<sup>160</sup> linked to Cys<sup>178</sup>, the internal cleavage site at D165E resides in a surface loop that is potentially accessible to proteases. This model could represent two-chain RGMc (Figure 4.7B). Both oxidation states result in tight globular proteins with a combination of  $\beta$ -sheets and  $\alpha$ -helices. The fifth cysteine in the sequence, at position 223, maps to the end of an  $\alpha$ -helix in both models and does not form a disulfide bond (Figure 4.7).

## 4.5 DISCUSSION

In this chapter, the disulfide linkages that were established by MS3 combining ETD and CID include: Cys<sup>39</sup>-Cys<sup>82</sup> (N-terminus and RGMc141X); Cys<sup>160</sup>-Cys<sup>310</sup> and Cys<sup>178</sup>-Cys<sup>314</sup> (single-chain); and Cys<sup>160</sup>-Cys<sup>178</sup> (two-chain). *Ab initio* modeling supported the existence of two different RGMc oxidation states.

### 4.5.1 The N-terminus of RGMc represents a functional domain

RGMc binds to members of the BMP family (34, 36, 68, 70), and an RGD sequence (a.a. 91-93) in RGMc appears to play a role in this interaction, as a Glu to Val mutation at position 92 abolishes binding (68). RGMc141X (which possesses this motif) also can bind to BMPs and inhibit their function, however not as effectively as wild-type RGMc (136). Retaining the ability to modulate BMP function led us to believe that this protein forms a functional stand-alone domain. The RGMc N-terminal domain *ab initio* model shows that the RGD domain may exist in a surface loop that is accessible for binding to BMPs (Figure 4.4).

To define the disulfide linkages in the N-terminus of RGMc we applied MS3 methodology on the full-length protein and to a truncation mutant, RGMc141X. The 141X mutants represents the first 109 residues of RGMc. In both proteins we found that Cys<sup>39</sup> is linked to Cys<sup>82</sup> (Figure 4.3), suggesting that the mutant and the N-terminus of wild-type RGMc adopt a similar conformation. *Ab initio* models predict that the N-terminus forms a tight compact structure that is predominately composed of  $\alpha$ -helices. Cys<sup>39</sup> and Cys<sup>82</sup> are located in two different helices. A disulfide linkage between the

cysteine residues can provide a point of stabilization and hold these helices together in a rigid position.

*Ab initio* models also suggest that a disulfide linkage occurs between Cys<sup>34</sup> and Cys<sup>90</sup> (Figure 4.4). However, the first amino acid of mature RGMc has not yet been identified, and even though prediction algorithms claim that Gln<sup>33</sup> is the first mature amino acid, it is not clear if Cys<sup>34</sup> is found in the mature protein. It is possible that mature RGMc isoforms begin with Lys<sup>35</sup>. Attempts to perform N-terminal sequencing have been unsuccessful and MS data do not conclusively support a disulfide bond between Cys<sup>34</sup> and Cys<sup>90</sup>. Additional studies are necessary to determine if this disulfide linkage exists and provides another point of stabilization between helices in the N-terminus of RGMc.

#### **4.5.2 Single-chain RGMc possesses two disulfide bonds in close proximity**

Disulfide mapping studies demonstrated that an oxidation state of RGMc exists in which the linkages Cys<sup>160</sup>-Cys<sup>310</sup> and Cys<sup>178</sup>-Cys<sup>314</sup> form. *Ab initio* modeling predicts that in this oxidative form, the two disulfide linkages form a small cysteine knot and render the internal cleavage site that gives rise to the two-chain protein inaccessible (Figure 4.7A). In this model, two cysteines (Cys<sup>160</sup> and Cys<sup>178</sup>) within the central region of RGMc form disulfide bonds with cysteines (Cys<sup>310</sup> and Cys<sup>314</sup>) ~70 amino acids from the C-terminus of the protein; the majority of the sequence between these disulfide bonds (a.a. 179-309) adopts a very highly ordered conformation, harboring a combination of  $\alpha$ -helices and  $\beta$ -sheets.

RGMc has a PC cleavage site at R328 that must be accessible in the single-chain protein (74, 75), as PCs, such as furin, cleave GPI-linked single-chain RGMc at the cell surface and release a 40 kDa soluble protein that is very stable in the extracellular environment (74, 75). The two-chain protein is not a substrate for PCs (58, 74). Extended *ab initio* predictions may demonstrate that this site is only accessible in single-chain RGMc. Extending beyond predicted structures of the individual proteins, the interactions between RGMc and neogenin, a transmembrane protein, may preclude access to PCs.

#### **4.5.3 The internal cleavage site is accessible in two-chain RGMc**

Both single- and two-chain RGMc reside on the cell surface via GPI-linkage (58). The mechanism responsible for cleavage of RGMc to the two-chain is unknown. In order for two-chain RGMc to form, an internal cleavage site residing after D165 must be accessible. In MS disulfide mapping studies we identified an RGMc oxidation state in which Cys<sup>160</sup> is linked to Cys<sup>178</sup> (Figure 4.3). *Ab initio* modeling demonstrates that the residues between these cysteines may form a surface-accessible loop that contains the internal cleavage site (Figure 4.7B), and that the Cys<sup>160</sup>-Cys<sup>178</sup> disulfide bond is the critical linkage connecting the N- and C-terminal domains of two-chain RGMc. In the model, ~50% of the protein takes on ordered secondary structure, with a combination of  $\alpha$ -helices and  $\beta$ -sheets, while the remainder is disordered.

Studies with BMP and neogenin, have demonstrated binding preferences for different RGMc isoforms and selected RGMc mutants (68). Two-chain RGMc binds to

the transmembrane protein neogenin, but does not bind to BMPs (68). It is not known why this differential binding occurs. In addition, RGMc mutants that do not form the two-chain species, such as D165E and G313V, do not bind to neogenin (68). Both of these mutants bind to BMPs and inhibit their function; however, they are less effective than their wild-type counterparts (136). G313V is the most commonly occurring JH-linked mutation (59). This mutation is positioned in a location that could directly inhibit the formation of the disulfide bond between Cys<sup>310</sup> and Cys<sup>314</sup>, and thereby de-stabilize the two-chain species.

#### **4.5.4 Mapping the disulfide bonds of RGMc with MS3 combining ETD and CID**

To map the disulfide bonds of RGMc we used a MS3 methodology combining ETD and CID recently established by the Karger laboratory (116). No reported studies have used this approach to map the disulfide bonds of proteins *de novo*. The Karger group mapped the disulfide bonds of an immunoglobulin light chain, human growth hormone, and in tissue plasminogen activator, three proteins in which the three-dimensional crystal structure had been known previously (116, 119). Since ETD can induce breakage of disulfide bonds, the studies reported in this chapter were performed only with non-reduced protein samples. Using non-reduced samples is advantageous when disulfide bonds occur among three or more peptides. In the RGMc single-chain oxidation state, P3 and P4 were both linked to P5 (Figure 4.5), which contains the sequence CVGGC (Figure 4.2). There is not a good selection of proteases that will cleave the protein between the cysteine residues. With traditional MS mapping methods comparing reduced and non-reduced samples, the exact sites of linkages between P3 and

P5 and between P4 and P5 would be difficult to map. ETD allows for dissociation of one of the peptides, and the following MS3 CID step allows for identification of not only the peptides involved, but through analysis of specific y and b ions the exact site of linkages can be defined.

#### 4.5.6 Summary

In this study, by employing a MS3 methodology combining ETD and CID we established a disulfide bond in the N-terminus of RGMc linking Cys<sup>39</sup> to Cys<sup>82</sup>. This same linkage was found in the truncation mutant, RGMc141X, suggesting the mutant adopts an overall similar fold to the N-terminus of RGMc. We also observed that two different oxidation states of RGMc exist. In the first, or single-chain RGMc, disulfide linkages connecting Cys<sup>160</sup>-Cys<sup>310</sup> and Cys<sup>178</sup>-Cys<sup>314</sup> were established. The Cys<sup>160</sup>-Cys<sup>178</sup> linkage was in the second oxidation state, or two-chain RGMc. *Ab initio* modeling supported our MS studies by demonstrating that the different RGMc oxidation states can physically exist and that both form tight compact structures.

#### **4.6 ACKNOWLEDGEMENTS**

We thank David Kuninger, Aditi Mukherjee and the OHSU Shared Proteomics Resource for experimental and technical advice. This work was supported by National Institutes of Health research grants 5R01 DK042748-22 (to P. R.), 5R01 EY007755-18 and 5P30 EY10572 (to L. D.), and National Science Foundation grant 0746589 (to U. S.). M. N. was supported by a National Institutes of Health pre-doctoral fellowship (F31 HL095271-03).

**Table 4.1 m/z values of selected RGMc cysteine-containing tryptic peptides**

Cysteine Residues	Peptide No.	Peptides
Cys39	P1	CNAEYVSSTLR
Cys82	P2	SYALCTR
Cys39 + Cys82	P1-P2	CNAEYVSSTLR - SYALCTR
Cys 160	P3	APGFLHCASFGDPHVR
Cys 178	P4	SFHNQFHTCR
Cys310/314	P5	AFSAEQDLQLCVGGCPPSQR
Cys160 + Cys178	P3-P4	APGFLHCASFGDPHVR-SFHNQFHTCR
Cys160 + Cys178 + Cys310/314	P3-P4-P5	APGFLHCASFGDPHVR-SFHNQFHTCR- AFSAEQDLQLCVGGCPPSQR
Cys160 + Cys310/314	P3-P5	APGFLHCASFGDPHVR-AFSAEQDLQLCVGGCPPSQR
Cys178 + Cys310/314	P4-P5	SFHNQFHTCR-AFSAEQDLQLCVGGCPPSQR

Peptide No.	m/z									
	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
P1	1242.6	621.8	-	-	-	-	-	-	-	-
P2	813.4	407.2	-	-	-	-	-	-	-	-
P1-P2	2053.0	1027.0	685.0	514.0	-	-	-	-	-	-
P3	1710.8	855.9	570.9	428.5	-	-	-	-	-	-
P4	1276.6	638.8	426.2	319.9	-	-	-	-	-	-
P5	2106.0	1053.5	-	-	-	-	-	-	-	-
P3-P4	2984.4	1492.7	995.5	746.8	597.7	498.2	427.2	373.9	-	-
P3-P4-P5	5087.3	2544.2	1696.4	1272.6	1018.3	848.7	727.6	636.8	566.1	509.6
P3-P5	3813.8	1907.4	1271.9	954.2	763.6	636.5	-	-	-	-
P4-P5	3379.5	1690.3	1127.2	845.6	676.7	564.1	-	-	-	-

**Table 4.2 y and b CID-MS3 ions for establishing Cys<sup>178</sup>-Cys<sup>314</sup> linkage**

Values in ( ) represent peptide masses disulfide-linked to specific cysteine residues.

P4-P5 disulfide-linked peptides (Cys178-314)

<u>P4</u>	y	y+2	y+3	y+4	y+5	y+6	b	b+2	b+3	b+4	b+5	b+6
S	-	-	-	-	-	-	-	-	-	-	-	-
F	3292.5	1647.3	1098.5	824.1	659.5	549.8	235.1	-	-	-	-	-
H	3145.5	1573.7	1049.5	787.4	630.1		372.2	187.1				
N	3008.4	1505.2	1003.8	753.1	602.7		486.2	244.1				
Q	2894.4	1448.2	965.8	724.6	579.9		614.3	308.1				
F	2766.3	1384.2	923.1	692.6	554.3		761.3	381.7				
H	2619.2	1310.6	874.1	655.8			898.4	450.2	300.5			
T	2482.2	1242.1	828.4	621.5			999.4	500.7	334.1			
<b>C178</b>	<b>C(2103.0)</b>	2381.1	1191.6	794.7			3205.5	1603.7	1069.5	802.4	642.1	
R	175.1						-	-	-	-	-	-
<u>P5</u>	y	y+2	y+3	y+4	y+5	y+6	b	b+2	b+3	b+4	b+5	b+6
A	-	-	-	-	-	-	-	-	-	-	-	-
F	3308.5	1655.3	1103.8	828.1	662.7	552.4	219.1					
S	3161.5	1581.7	1054.8	791.4	633.3	527.9	306.1					
A	3074.4	1538.2	1025.8	769.6	615.9	513.4	377.2					
E	3003.4	1502.7	1002.1	751.9	601.7	501.6	506.2					
Q	2874.4	1438.2	959.1	719.6	575.9	480.1	634.3					
D	2746.3	1374.2	916.4	687.6	550.3	458.7	749.3					
L	2631.3	1316.6	878.1	658.8	527.3	439.5	862.4					
Q	2518.2	1260.1	840.4	630.5	504.6	420.7	990.5					
L	2390.1	1196.1	797.7	598.5	479.0	399.4	1103.5					
<b>C310</b>	<b>C</b>	2277.0	1139.5	760.0	570.3	456.4	380.5					
V	2174.0	1088.0	725.7	544.5	435.8	363.3	1305.6					
G	2075.0	1038.5	692.7	519.7	416.0	346.8	1362.6					
G	2018.0	1010.0	673.7	505.5	404.6	337.3	1419.7					
<b>C314</b>	<b>C(1273.6)</b>	1960.9	981.5	654.6	491.2	393.2	327.8	2796.3	1399.1	933.1	700.1	560.3
P	584.3	293.2					2893.3	1447.7	965.4	724.3	579.7	
P	487.3	244.6					2990.4	1496.2	997.8	748.6	599.1	
S	390.2	196.1					3077.4	1539.7	1026.8	770.4	616.5	
Q	303.2	152.6					3205.5	1603.7	1069.5	802.4	642.1	535.2
R	175.1	88.6					-	-	-	-	-	-

**Table 4.3 y and b CID-MS3 ions for establishing Cys<sup>160</sup>-Cys<sup>310</sup> linkage**

Values in ( ) represent peptide masses disulfide-linked to specific cysteine residues.

P3-P5 disulfide-linked peptides (Cys160-Cys310)

<u>P3</u>	y	y+2	y+3	y+4	y+5	y+6	b	b+2	b+3	b+4	y+5	y+6
A	-	-	-	-	-	-	-	-	-	-	-	-
P	3742.8	1872.4	1248.6	936.7	749.6	624.8	169.1	-	-	-	-	-
G	3645.7	1823.9	1216.2	912.4	730.1	608.6	226.1	-	-	-	-	-
F	3588.7	1795.4	1197.2	898.2	718.7	599.1	373.2	-	-	-	-	-
L	3441.6	1721.8	1148.2	861.4	689.3	574.6	486.3	-	-	-	-	-
H	3328.6	1665.3	1110.5	833.1	666.7	555.8	623.3	312.7	-	-	-	-
C(2103.0)	3191.5	1596.7	1064.8	798.9	639.3	-	2829.3	1415.7	944.1	708.3	-	-
A	985.5	493.7	329.5	-	-	-	2900.4	1451.2	967.8	726.1	-	-
S	914.5	458.2	305.8	-	-	-	2987.4	1494.7	996.8	747.9	-	-
F	827.4	414.7	276.8	-	-	-	3134.5	1568.2	1045.8	784.6	-	-
G	680.4	341.2	227.8	-	-	-	3191.5	1596.8	1064.8	798.9	-	-
D	623.3	312.7	208.8	-	-	-	3306.5	1654.3	1103.2	827.6	-	-
P	508.3	255.2	170.4	-	-	-	3403.6	1702.8	1135.5	851.9	-	-
H	411.3	206.6	-	-	-	-	3540.6	1771.3	1181.2	886.2	709.1	-
V	274.2	138.1	-	-	-	-	3639.7	1820.9	1214.2	910.9	728.9	607.6
R	175.1	88.6	-	-	-	-	-	-	-	-	-	-
<u>P5</u>	y	y+2	y+3	y+4	y+5	y+6	b	b+2	b+3	b+4	b+5	b+6
A	-	-	-	-	-	-	-	-	-	-	-	-
F	3742.7	1872.4	1248.6	936.7	749.5	624.8	219.1	-	-	-	-	-
S	3595.7	1798.8	1199.6	899.9	720.1	600.3	306.1	-	-	-	-	-
A	3508.6	1755.3	1170.5	878.2	702.7	585.8	377.2	-	-	-	-	-
E	3437.6	1719.8	1146.9	860.4	688.5	573.9	506.2	-	-	-	-	-
Q	3308.6	1655.3	1103.9	828.1	662.7	552.4	634.3	-	-	-	-	-
D	3180.5	1591.3	1061.2	796.1	637.1	531.1	749.3	-	-	-	-	-
L	3065.5	1533.7	1022.8	767.4	614.1	511.9	862.4	-	-	-	-	-
Q	2952.4	1477.2	985.1	739.1	591.5	493.1	990.5	-	-	-	-	-
L	2824.3	1413.2	942.4	707.1	565.9	471.7	1103.5	-	-	-	-	-
C(1707.8)	2711.2	1356.6	904.7	678.8	543.2	452.9	2914.4	1458.2	972.5	729.6	583.9	-
V	900.4	451.2	-	-	-	-	3013.4	1507.7	1005.5	754.4	603.7	-
G	801.4	401.7	-	-	-	-	3070.4	1536.2	1024.5	768.6	615.1	-
G	744.4	373.2	-	-	-	-	3127.5	1564.7	1043.5	782.9	626.5	-
C	687.3	344.7	-	-	-	-	3230.5	1616.2	1077.8	808.6	647.1	-
P	584.3	293.2	-	-	-	-	3327.5	1664.8	1110.2	832.9	666.5	-
P	487.3	244.6	-	-	-	-	3424.6	1713.3	1142.5	857.1	685.9	-
S	390.2	196.1	-	-	-	-	3511.6	1756.8	1171.5	878.9	703.3	-
Q	303.2	152.6	-	-	-	-	3639.7	1820.8	1214.2	910.9	728.9	-
R	175.1	88.6	-	-	-	-	-	-	-	-	-	-

**Table 4.4 CID-MS3 y and b ions for defining Cys<sup>160</sup>-Cys<sup>178</sup> disulfide-linkage**

<u>P3</u>	y	y+2	y+3	y+4	y+5	y+6	b	b+2	b+3	b+4	y+5	y+6
A	-	-	-	-	-	-	-	-	-	-	-	-
P	1639.8	820.9	547.6	410.9	329.0	274.3	169.1	-	-	-	-	-
G	1542.7	772.4	515.2	386.7	309.5	258.1	226.1	-	-	-	-	-
F	1485.7	743.9	496.2	372.4	298.1	248.6	373.2	-	-	-	-	-
L	1338.6	670.3	447.2	335.7	268.7	224.1	486.3	-	-	-	-	-
H	1225.6	613.8	409.5	307.4	246.1	205.3	623.3	312.7	-	-	-	-
<b>C160</b>												
C	1088.5	545.2	363.8	273.1	218.7	-	726.3	364.2	243.1	182.6	-	-
A	985.5	493.7	329.5	-	-	-	797.4	399.7	266.8	200.3	-	-
S	914.5	458.2	305.8	-	-	-	884.4	443.2	295.8	222.1	-	-
F	827.4	414.7	276.8	-	-	-	1031.5	516.7	344.8	258.9	-	-
G	680.4	341.2	227.8	-	-	-	1088.5	545.2	363.8	273.1	-	-
D	623.3	312.7	208.8	-	-	-	1203.5	602.8	402.2	301.9	-	-
P	508.3	255.2	170.4	-	-	-	1300.6	651.3	434.5	326.1	-	-
H	411.3	206.6	-	-	-	-	1437.6	719.8	480.2	360.4	288.5	-
V	274.2	138.1	-	-	-	-	1536.7	769.4	513.2	385.2	308.3	257.1
R	175.1	88.6	-	-	-	-	-	-	-	-	-	-
<u>P4</u>	y	y+2	y+3	y+4	y+5	y+6	b	b+2	b+3	b+4	b+5	b+6
S	-	-	-	-	-	-	-	-	-	-	-	-
F	1189.5	595.8	397.5	298.4	238.9	199.3	235.1	-	-	-	-	-
H	1042.5	522.2	348.5	261.6	209.5	-	372.2	187.1	-	-	-	-
N	905.4	453.7	302.8	227.4	182.1	-	486.2	244.1	-	-	-	-
Q	791.4	396.7	264.8	198.8	159.3	-	614.3	308.1	-	-	-	-
F	663.3	332.7	222.1	166.8	133.7	-	761.3	381.7	-	-	-	-
H	516.2	259.1	173.1	130.1	-	-	898.4	450.2	300.5	-	-	-
T	379.2	190.6	127.4	95.8	-	-	999.4	500.7	334.1	-	-	-
<b>C178</b>												
C	278.1	140.1	93.7	-	-	-	1102.5	552.2	368.5	276.6	221.5	-
R	175.1	-	-	-	-	-	-	-	-	-	-	-

**Figure 4.1 Experimental model for identifying the disulfide bonds of RGMc.**

Analytical approach for identification of disulfide linkages in RGMc by MS. See Sections 2.3.3 and 3.3.5 for details.

**Figure 4.1**



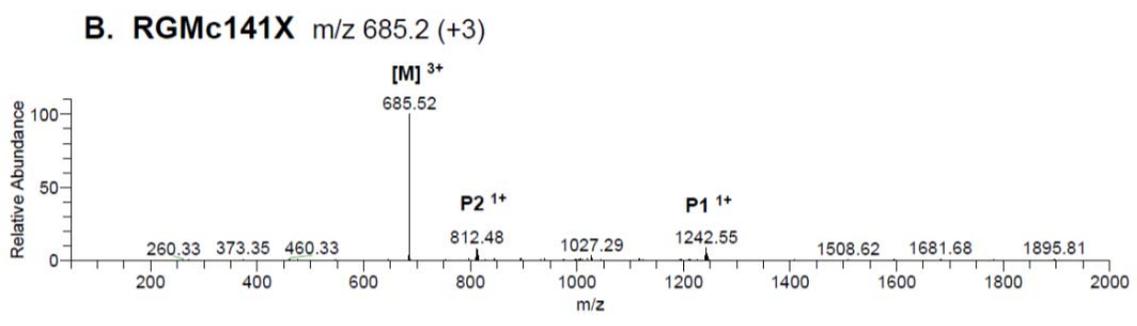
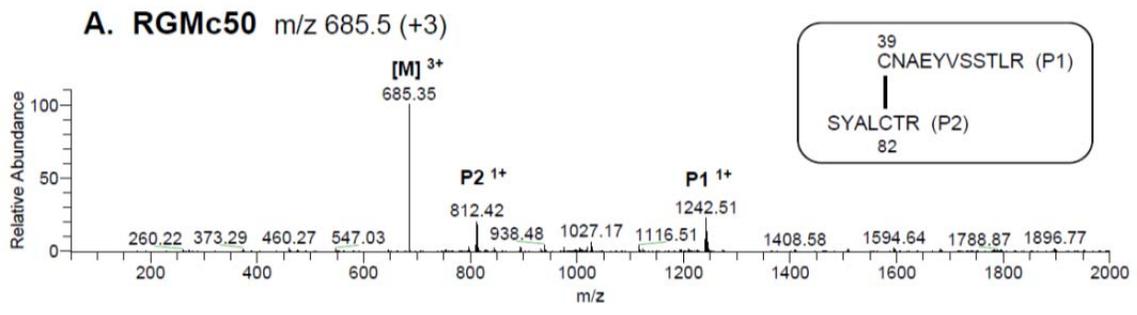
**Figure 4.2 Selected RGMc cysteine-containing tryptic peptides.** List of 5 selected tryptic peptides (P1-P5) in RGMc that contain cysteine residues (cysteines are in **bold** script). Numbers in parentheses denote the location of each cysteine in the amino acid sequence.

**Figure 4.2**

P1: **C**NAEYVSSTLR (39)  
P2: SYAL**C**TR (82)  
P3: APGFLH**C**ASFGDPHVR (160)  
P4: SFHNQFHT**C**R (178)  
P5: AFSAEQDLQL**C**VGG**C**PPSQR (310, 314)

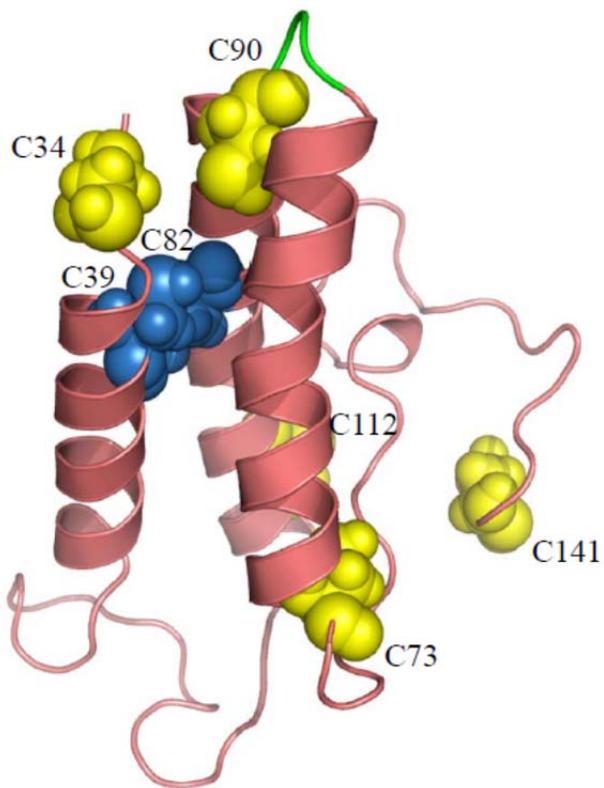
**Figure 4.3** The N-terminus of RGMc contains one disulfide bond linking Cys<sup>39</sup> to Cys<sup>82</sup>. Results demonstrate that peptides P1-P2, containing cysteine residues 39 and 82, are linked by disulfide bonds. ETD of precursor ion, m/z 685 (+3) of RGMc (**A**) and RGMc141X (**B**). [M] denotes the parent mass.

Figure 4.3



**Figure 4.4** *Ab initio* model of the N-terminal domain of RGMc. Cartoon representation of the *ab initio* model of the N-terminal domain of RGMc, containing amino acid residues 33-141. This model also represents the truncation mutant RGMc141X. Cylinders represent  $\alpha$ -helical regions; thin lines are unstructured regions. The RGD motif (residues 91-93) is depicted in green. Cysteines are depicted as space-filling spheres. C represents cysteines; all cysteines within the model are numbered. The free cysteines at positions 34, 73, 90, 112, and 141 are in yellow. The disulfide linked cysteines, Cys<sup>39</sup>-Cys<sup>82</sup>, are presented in blue.

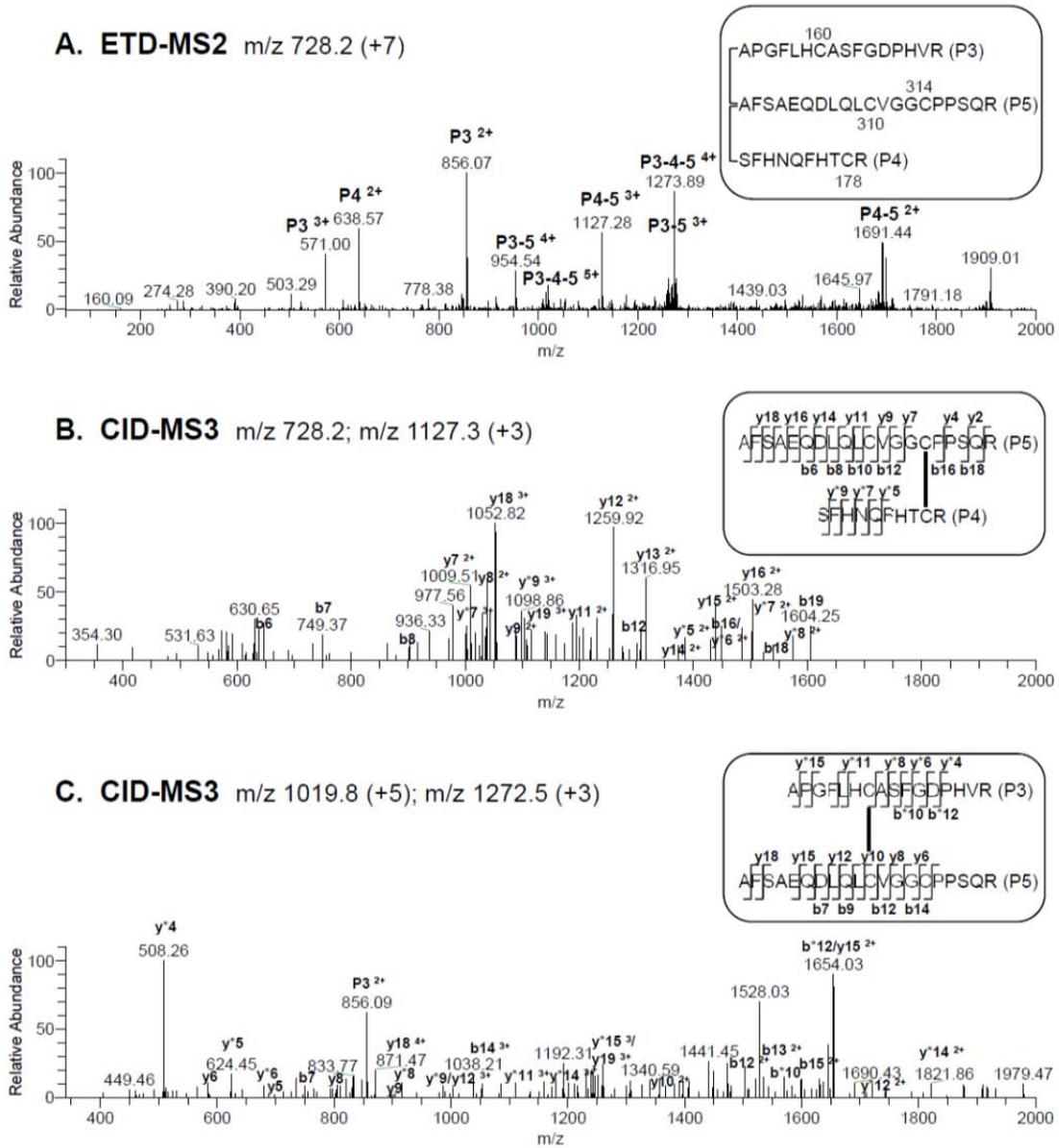
**Figure 4.4**



**Figure 4.5 The disulfide linkages, Cys<sup>160</sup>-Cys<sup>310</sup> and Cys<sup>178</sup>-Cys<sup>314</sup>, in RGMc.**

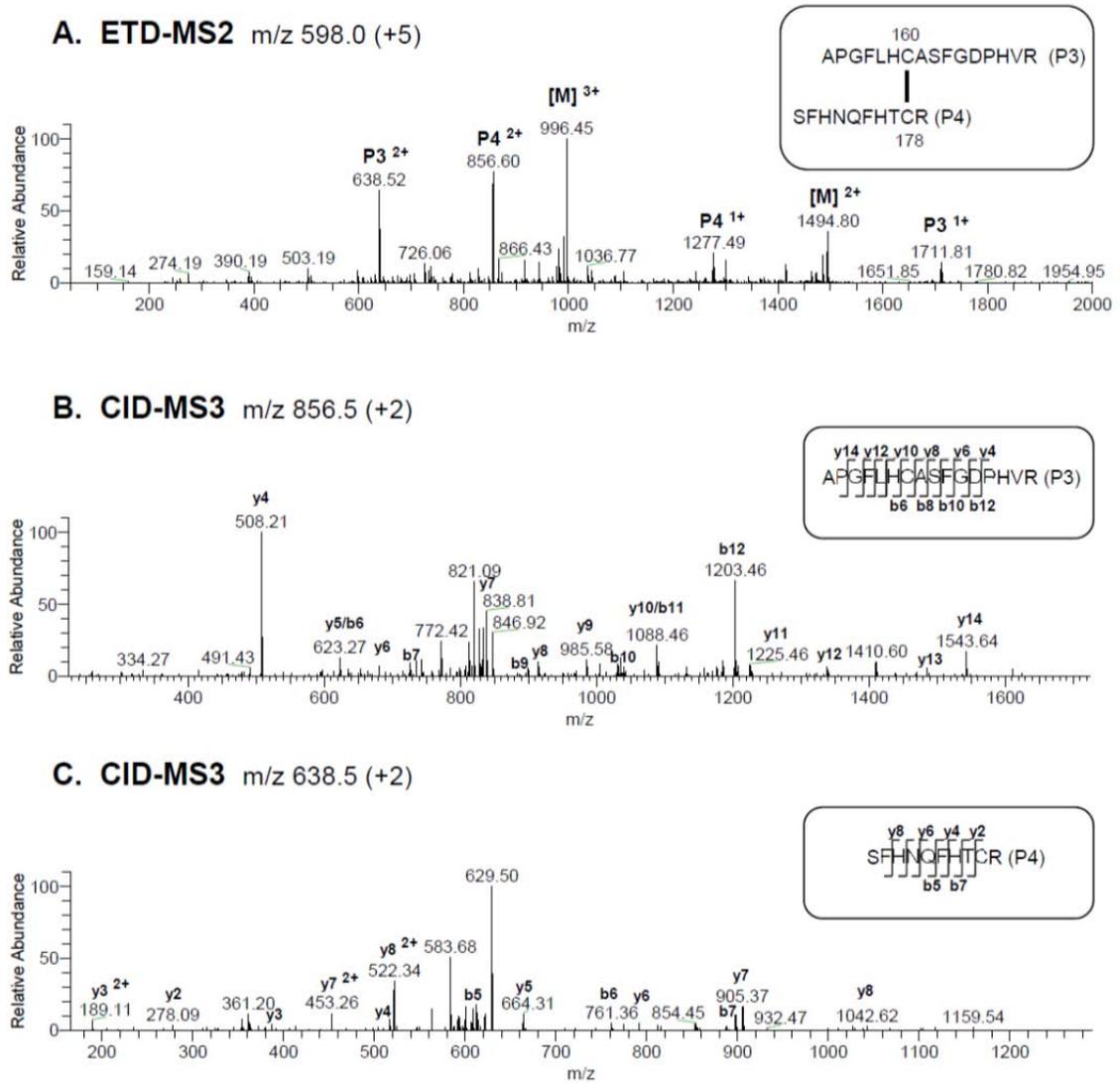
Results demonstrate that peptides P3-P4-P5 are linked by disulfide bonds. See Table 4.1 for additional information. **A - C.** Analysis of P3-P4-P5. **A.** ETD of precursor ion, m/z 728.2 (+7), corresponding to P3-P4-P5, and containing cysteine residues 160, 310, and 314. **B.** CID of precursor ion, m/z 1127.3 (+3), of disulfide-linked peptides P4 and P5. The y ions marked with an asterisk are from P4; other y ions are from P5. **C.** CID of precursor ion m/z 1272.5 (+3) of disulfide-linked peptides connected between cysteine residues 178 and 314. The y ions marked with an asterisk are from P3. In **A**, [M] denotes the parent mass. Additional y ions were identified in **B** and **C**, but have been removed from the graphs for clarity.

**Figure 4.5**



**Figure 4.6** Cys<sup>160</sup> is linked to Cys<sup>178</sup> in a second RGMc oxidation state. **A.** ETD of precursor ion, m/z 598 (+5), corresponding to disulfide-linked peptides P3 and P4. [M] denotes the parent mass. **B.** CID of precursor ion, m/z 856.5 (+2), of P3 dissociated from P4. **C.** CID of precursor ion m/z 638.5 (+2) of P4 containing cysteine residues 178.

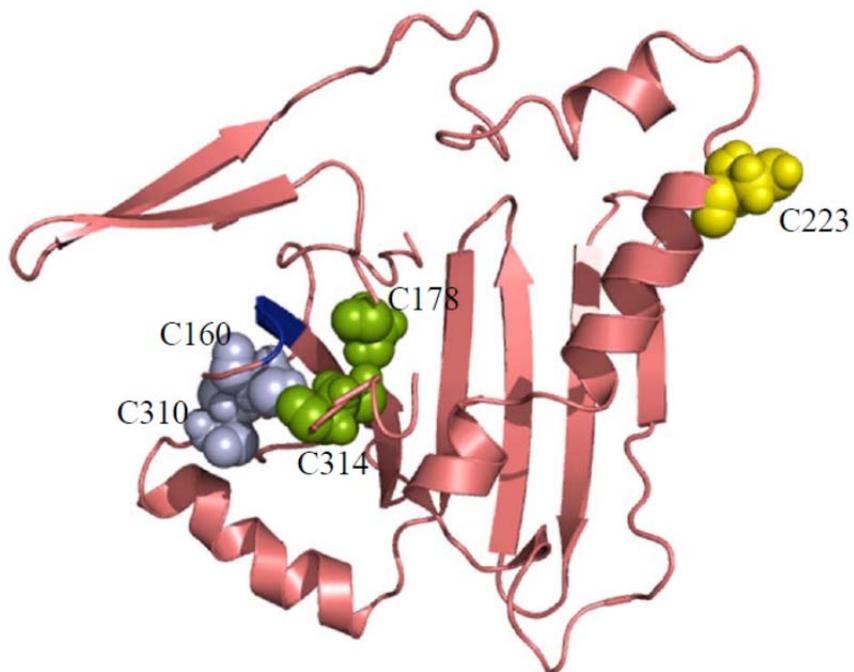
**Figure 4.6**



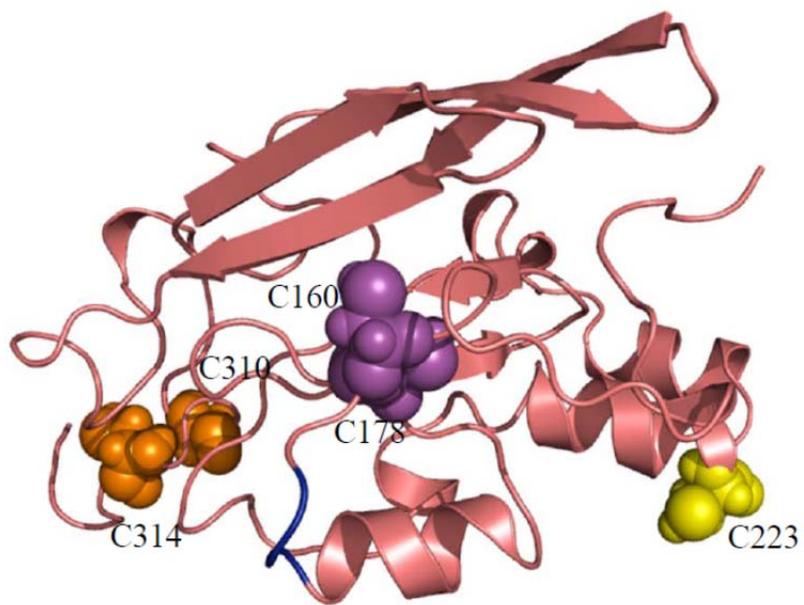
**Figure 4.7** *Ab initio* models of single and two-chain RGMc. Cartoon representations of the *ab initio* models of single- (**A**) and two-chain RGMc (**B**), containing amino acid residues 142-319. Cylinders represent  $\alpha$ -helical regions, thick lines with arrows,  $\beta$ -sheets, and thin lines, unstructured regions. Cysteines are depicted as space-filling spheres. The FGD residues (163-165) of the internal cleavage site (FGDP, with cleavage after D165E) are depicted in dark blue. The free cysteine at position 223 is in yellow. C represents cysteines; all cysteines within the model are numbered. **A.** The disulfide linked cysteines, Cys<sup>160</sup>-Cys<sup>310</sup> and Cys<sup>178</sup>-Cys<sup>314</sup>, are presented in light blue and green, respectively. **B.** Cys<sup>160</sup>-Cys<sup>178</sup> and Cys<sup>310</sup>-Cys<sup>314</sup> disulfide bonded cysteines are depicted in purple and orange, respectively.

**Figure 4.7**

**A.**



**B.**



# **Chapter 5**

## **Conclusions and Future Directions**

## **5.1 CONCLUSIONS**

The central goal of this dissertation was to elucidate the role of RGMc in iron regulation through a structure function approach. The main findings and conclusions of the studies reported in this dissertation, as well as their physiological implications, are briefly summarized below.

### **5.1.1 Soluble RGMc Inhibits BMP-mediated Signaling and Gene Expression**

The data presented in Chapter 2 of this dissertation characterizes the interaction between the soluble RGMc isoforms and two members of the BMP family, BMP2 and BMP6. Signaling and microarray results clearly demonstrate that RGMc is a potent inhibitor of BMPs. RGMc decreases BMP-mediated Smad phosphorylation in a dose-dependent manner. In addition, the expression of many BMP-response genes, such as BAMBI, GHR, and SMAD7 are reduced in the presence of RGMc. Although Noggin, a critical inhibitor of BMP2 and BMP4 during development, is a stronger antagonist of BMP2-mediated actions than RGMc, it is completely ineffective at modulating BMP6 signaling. Our studies demonstrate that RGMc is a broad-spectrum BMP antagonist, as members of the BMP2/4 and BMP5/6/7/8 sub-classes of growth factors are inhibited.

### **5.1.2 Structure-Function Insights from JH-linked RGMc Mutants**

Several studies in this dissertation based on selected mutants provide critical information about the biosynthesis, structure, and function of RGMc. The RGMc amino acid substitution mutants, G92V and G313V, and the truncation mutant RGMc141X aided in identifying functional domains that interact with BMPs (Chapter 2, Section 5.3).

Experiments with RGMc141X helped identify the disulfide bonds in the N-terminal domain of RGMc (Chapter 4). Finally, biosynthesis and processing analyses of the frameshift mutants, 378X, 354X, 321X, and 305X revealed the location of a key RGMc cleavage site for PCs (Appendix).

Previous studies from our lab demonstrated that single-chain RGMc preferentially binds to BMPs over the two-chain isoform (68). In addition, we showed that an amino acid substitution in the RGD domain, G92V, abolished binding to BMPs (68). Here we extended those studies by demonstrating that RGMcG92V was ineffective as a BMP inhibitor even at a 40-fold molar excess. However, two other RGMc mutants, G313V and 141X, inhibited BMP-mediated gene expression, but were slightly less potent than their wild-type counterparts.

The fact that RGMc141X was an effective BMP inhibitor led us to believe that this truncated protein formed an independent functional N-terminal domain. Attempts at expressing the truncated mutants 109X and 262X were unsuccessful (unpublished data), perhaps indicating that the last amino acid of the N-terminal domain resided closer to position 141. RGMc141X contains seven cysteines, and disulfide mapping studies revealed that Cys<sup>39</sup> and Cys<sup>82</sup> were linked to one another, a linkage that was also observed in full-length RGMc. No other disulfide bonds were detected in RGMc141X or in the N-terminus of the full-length protein, perhaps indicating that the overall secondary structure of these regions were the same in the two proteins.

The frameshift mutations, 378X, 354X, 321X, and 305X, located in the C-terminus of RGMc provided critical information about the site of cleavage of the full-length protein by PCs such as furin. Upon expression in Cos7 cells, RGMc305X and 321X gave rise to a single protein species, while RGMc354X and 378X gave rise to two protein species. These studies allowed us to narrow down the site of cleavage of RGMc to a series of PC-sites ending at R328. Through the use of these truncation mutants, which lack the GPI-anchor signal sequence, we were also able to demonstrate that GPI-linkage to the cell membrane was not necessary for processing of full-length RGMc50 to RGMc40.

### **5.1.3 Disulfide Mapping by MS Combining ETD and CID**

MS methods are routinely used to establish whole protein disulfide maps or individual disulfide linkages (113, 116). Traditional methods rely on comparison of the MS spectra of reduced and non-reduced protein samples (113). This approach has proven to be successful for peptides that contain single cysteines and can thereby be linked to one other peptide. Recently, specialized methods have allowed for establishment of disulfide bonds for peptides containing multiple cysteines that can give rise to highly intertwined peptides (116, 119).

Unlike the more conventional dissociation method, CID, which generates y and b ions, ETD gives rise to c and z ions (103, 105, 110). In addition, ETD induces the disulfide bond to break (107, 111). The Karger laboratory designed a novel MS3 methodology combining ETD and CID to map the disulfide bonds of an immunoglobulin

light chain, the growth hormone receptor, and the tissue plasminogen activator, three proteins for which the cysteine linkages were already established (116, 119).

In Chapter 3 of this dissertation, we employed a similar MS3 protocol combining ETD and CID to map the disulfide bonds of IGFBP-5. We showed that the 18 evolutionarily conserved cysteines were coordinated into nine disulfide bonds. Seven of the disulfide linkages, Cys<sup>7</sup>-Cys<sup>33</sup>, Cys<sup>25</sup>-Cys<sup>39</sup>, Cys<sup>47</sup>-Cys<sup>60</sup>, Cys<sup>54</sup>-Cys<sup>80</sup>, Cys<sup>172</sup>-Cys<sup>199</sup>, Cys<sup>210</sup>-Cys<sup>221</sup>, and Cys<sup>223</sup>-Cys<sup>243</sup>, were definitively mapped. The final two, which are predicted to connect Cys<sup>10</sup>-Cys<sup>35</sup> and Cys<sup>18</sup>-Cys<sup>36</sup> through homology models, were not established by this methodology. Although ETD dissociated those disulfide bonds, the fact that Cys<sup>35</sup> and Cys<sup>36</sup> were directly adjacent to one another did not allow for ion assignments between these cysteines, which is essential for establishing the exact linkage sites.

### 5.1.3 Disulfide Mapping of RGMc Reveals Multiple Oxidation States

Using a tandem MS approach combining ETD and CID, we mapped specific disulfide bonds of RGMc (Chapter 4). These disulfide mapping studies provide the first structural data for RGMc and for the RGM family. The RGMc141X truncation mutant, which contains seven cysteines, contains one disulfide bond linking Cys<sup>39</sup> and Cys<sup>82</sup>. The same disulfide linkage was established in full-length RGMc, while no other bonds were found in the N-terminus of the protein.

Full-length RGMc50 exists as either a single- or two-chain species. Using a two-pronged approach combining MS and *ab initio* modeling, we observed two principal

RGMc oxidation states. In the first, Cys<sup>160</sup> was linked to Cys<sup>178</sup>. The internal cleavage site that gives rise to the two-chain isoform is located at Asp165 which lies between these two cysteines. We postulate that the amino acids that reside between Cys<sup>160</sup> and Cys<sup>178</sup> lie in a surface loop that is accessible to the protease that cleaves RGMc at this site. In the second oxidation state, disulfide linkages connected Cys<sup>160</sup> to Cys<sup>310</sup> and Cys<sup>178</sup> to Cys<sup>314</sup>. In this single-chain species, we predict through *ab initio* modeling that the C-terminal region of the protein folds over the internal cleavage site to render this site inaccessible to proteases; however, the PC cleavage site at R328 must remain accessible, such that the full-length protein can be cleaved to produce RGMc40.

## 5.2 FUTURE DIRECTIONS

The results reported in this dissertation leave many unanswered questions open about the structure and function of RGMc. Specific topics for which future studies can be based on are addressed below.

### 5.2.1 Biosynthesis and Processing of RGMc Gives Rise to Distinct Isoforms

Biosynthesis and processing of RGMc gives rise to single- and two-chain cell-species (58, 66, 68, 73-75). Two RGMc isoforms are found on the cell surface via GPI-linkage (58), while two soluble isoforms, 50 and 40 kDa single-chain species, are found in the circulation (74, 75). Many questions remain regarding the biosynthesis and processing mechanisms that leads to the expression of four distinct RGMc isoforms remain unanswered.

The mechanisms responsible for two-chain RGMc have not been established. It has been suggested that cleavage at Asp165 is based on acid-hydrolysis but *in vitro* attempts to demonstrate this have been unsuccessful. It is possible that RGMc is cleaved by an un-identified protease, and that single- and two-chain RGMc traverse alternate intracellular trafficking pathways. In over-expression studies with RGMc proteins lacking GPI anchors, it appears that a region of the far C-terminus is necessary for efficient cleavage to the two-chain species. Expression of the C-terminal truncation mutant 378X produces primarily single-chain RGMc, while RGMc engineered to lack only the GPI anchor signal sequence (392X) is expressed as both single- and two-chain isoforms (74). Therefore, it appears that the GPI anchor is not necessary for the internal

cleavage event to take place. Mutational analysis, via signal amino acid substitutions and deletions, within RGMc between amino acids 378 and 392 could further resolve the critical region required for processing to the two-chain species.

Our lab has demonstrated that the single amino acid substitution mutants D165E and G313V do not give rise to two-chain RGMc (68). This is expected with RGMcD165E as this mutation abolishes the internal cleavage site. G313V is the most prevalent JH-linked mutation (59), and along with several C-terminal truncation mutants, 378X, 354X, 321X and 305X, only expresses as single-chain proteins (68, 74). It would be interesting to determine if other point mutations disrupt the internal cleavage event from taking place. RGMc is a glycoprotein that contains three N-linked glycosylation sites. It is not known if differential glycosylation can regulate folding, structure, or trafficking of RGMc to ultimately give rise to distinct protein species.

It is also unknown what mechanisms regulate the expression levels of the two cell surface isoforms and if there is a balance between them to maintain iron homeostasis. Cell-based iron loading or iron depletion studies have not revealed changes in RGMc mRNA or protein levels; however, release of RGMc from the cell surface may be regulated by iron (74). Many of the experiments to address these questions have been performed in hepatocarcinoma cell lines which may have lost key iron responses (73). Monitoring over-expression of RGMc in primary hepatocytes, in the presence and absence of iron, may provide clues as to how iron regulates expression of specific isoforms.

A study focusing on all three RGM family members could lend insight on the overall biosynthesis and processing of this protein family. A comprehensive study monitoring the expression of RGMa and RGMb has not been carried out. It is known that two-chain RGMa is expressed, but it is unclear if single-chain forms are present on the cell surface or in the circulation. RGMb does not have a PC cleavage site and therefore should not give rise to a soluble RGMc40 equivalent. RGMa and RGMb are expressed in the developing nervous system while RGMc is expressed in skeletal muscle, the heart and the liver, and it is not yet known if cell-type specific processing plays a role in the generation of specific RGM isoforms.

### **5.2.2 RGMc Structure**

Disulfide mapping studies presented in this dissertation that have revealed two primary RGMc oxidation states exist. In the first, or two-chain protein, Cys<sup>160</sup> is linked to Cys<sup>178</sup>; Cys<sup>160</sup>-Cys<sup>310</sup> and Cys<sup>178</sup>-Cys<sup>314</sup> disulfide linkages are present in the second single-chain oxidation state. Five JH-linked mutations either add in or subtract cysteines in the amino acid sequence. Two mutations of special interest in HJV are R176C and W191C which correspond to R169C and W184C, respectively, in RGMc. Both of these mutations insert cysteines near Cys<sup>160</sup> and Cys<sup>178</sup> which can disrupt the proper disulfide linkages that give rise to single- and two-chain RGMc. It would be interesting to study these mutants to determine if they are expressed, released from cells, and retain function as BMP inhibitors, and if so, establish the disulfide maps of these proteins.

Many attempts to express high yields of RGMc in bacterial systems were unsuccessful in our lab likely due to the numerous cysteines within the protein and the necessity for glycosylation. Perhaps turning to a baculovirus expression system could circumvent the problems met with glycosylation requirements. Higher RGMc yields could provide the driving force to attempt to define its xray crystal structure; however measures to individually purify only the single- or two-chain species would likely be necessary.

### **5.2.3 Establishment of disulfide linkages by MS combining ETD and CID**

The MS data presented in this dissertation establish a disulfide map for IGFBP-5 and several key disulfide linkages of RGMc. Several steps could be followed to enhance these findings. We attempted to perform a MS4 protocol in which two sequential ETD steps were followed by CID. ETD is ~ 20% as efficient as CID, therefore additional ETD dissociation steps resulted in very low ion abundance for analysis. To improve this method, highly intertwined peptides need to be enriched after digestion such that they are the major ion being analyzed during individual runs. This could lead to identification of the peptides in the final CID step following the sequential ETD steps. Unlike CID, ETD will induce breakage of highly charged peptides. As the charge of an ion species increases, the mass resolution decreases. To combat this problem, performing MS3 or MS4 experiments with ETD on an orbitrap would provide ever higher confidence of correct peptide assignment.

The MS methods employed in this dissertation could be used to answer significant biological questions. Disulfide scrambling, or altered disulfide linkages, which both can occur in proteins as a result of aging, can be monitored by MS. In addition, protein oxidation states of different cellular compartments, such as the ER or mitochondrial intermembrane space, can be monitored after variable perturbations.

#### **5.2.4 RGM Family Members are BMP Co-receptors that Bind to Neogenin**

All three members of the RGM family have been reported to function as BMP co-receptors that up-regulate BMP mediated actions (27, 34, 36, 71). To date, it has been shown that upon BMP stimulation and over-expression of RGMs, a slight increase in activity of a BMP response element results (34, 36, 71). However, it has never been shown that RGMs enhance Smad phosphorylation or increase BMP-mediated gene expression. In-depth studies that monitor basic signaling and gene expression need to be performed to validate this co-receptor hypothesis.

RGMA mediates its actions through the transmembrane protein neogenin (62, 64, 83), but the role of neogenin in the biological effects of RGMb and RGMc is unclear. The signaling pathways that result from this interaction are still undefined. Our lab has shown that two-chain RGMc preferentially binds neogenin, while the single-chain isoform binds to BMPs (68). We do not know if this holds true for RGMA or RGMb. It is still not clear if RGMs associate with larger protein complexes on the cell surface, or if varied RGM isoforms have different binding properties *in vivo*. It is also not known if

different RGM isoforms localize to the various regions of the cell surface such that differential interactions can take place.

### **5.2.5 RGMc in Iron Regulation**

Humans with mutations in hemojuvelin and mice lacking RGMc have very low levels of hepcidin and experience iron overload. To gain a better understanding of RGMc physiology, one could take RGMc knockout mice and establish a model system where the protein can be expressed under a controlled manner in the liver at specific points during development to see if replenishment with RGMc can reverse the pathological phenotype. In addition, it would be interesting to determine whether RGMa or RGMb could be substituted in for RGMc. These studies may tease out the biological actions of RGMc, that are necessary for proper development of iron regulatory pathways.

### **5.3 DISSERTATION SUMMARY**

The structure-function studies reported in this dissertation provide a framework for understanding the role of RGMc in iron regulation. The major findings of this dissertation are (i) soluble RGMc is a potent and broad spectrum inhibitor of BMPs, (ii) novel tandem MS methodologies combining ETD and CID can be used to map the disulfide bonds of proteins with unresolved structures, such as IGFBP-5, and (iii) mature RGMc exists in varied oxidation states. Future work is necessary to further elucidate the biological roles of RGMc and the resultant pathophysiological processes that occur upon mutation of the protein.

## References

1. Andrews, N.C., and Schmidt, P.J. 2007. Iron homeostasis. *Annu Rev Physiol* 69:69-85.
2. Hentze, M.W., Muckenthaler, M.U., Galy, B., and Camaschella, C. 2010. Two to tango: regulation of Mammalian iron metabolism. *Cell* 142:24-38.
3. Gunshin, H., Mackenzie, B., Berger, U.V., Gunshin, Y., Romero, M.F., Boron, W.F., Nussberger, S., Gollan, J.L., and Hediger, M.A. 1997. Cloning and characterization of a mammalian proton-coupled metal-ion transporter. *Nature* 388:482-488.
4. McKie, A.T., Barrow, D., Latunde-Dada, G.O., Rolfs, A., Sager, G., Mudaly, E., Mudaly, M., Richardson, C., Barlow, D., Bomford, A., et al. 2001. An iron-regulated ferric reductase associated with the absorption of dietary iron. *Science* 291:1755-1759.
5. Wrighting, D.M., and Andrews, N.C. 2008. Iron homeostasis and erythropoiesis. *Curr Top Dev Biol* 82:141-167.
6. Donovan, A., Lima, C.A., Pinkus, J.L., Pinkus, G.S., Zon, L.I., Robine, S., and Andrews, N.C. 2005. The iron exporter ferroportin/Slc40a1 is essential for iron homeostasis. *Cell Metab* 1:191-200.
7. Cherukuri, S., Potla, R., Sarkar, J., Nurko, S., Harris, Z.L., and Fox, P.L. 2005. Unexpected role of ceruloplasmin in intestinal iron absorption. *Cell Metab* 2:309-319.

8. Bothwell, T.H., and Charlton, R.W. 1979. Current problems of iron overload. *Recent Results Cancer Res* 69:87-95.
9. Ganz, T. 2011. Heparin and iron regulation, 10 years later. *Blood* 117:4425-4433.
10. Pigeon, C., Ilyin, G., Courselaud, B., Leroyer, P., Turlin, B., Brissot, P., and Loreal, O. 2001. A new mouse liver-specific gene, encoding a protein homologous to human antimicrobial peptide hepcidin, is overexpressed during iron overload. *J Biol Chem* 276:7811-7819.
11. Nicolas, G., Bennoun, M., Devaux, I., Beaumont, C., Grandchamp, B., Kahn, A., and Vaulont, S. 2001. Lack of hepcidin gene expression and severe tissue iron overload in upstream stimulatory factor 2 (USF2) knockout mice. *Proc Natl Acad Sci U S A* 98:8780-8785.
12. Rivera, S., Nemeth, E., Gabayan, V., Lopez, M.A., Farshidi, D., and Ganz, T. 2005. Synthetic hepcidin causes rapid dose-dependent hypoferremia and is concentrated in ferroportin-containing organs. *Blood* 106:2196-2199.
13. McKie, A.T., Marciani, P., Rolfs, A., Brennan, K., Wehr, K., Barrow, D., Miret, S., Bomford, A., Peters, T.J., Farzaneh, F., et al. 2000. A novel duodenal iron-regulated transporter, IREG1, implicated in the basolateral transfer of iron to the circulation. *Mol Cell* 5:299-309.
14. Donovan, A., Brownlie, A., Zhou, Y., Shepard, J., Pratt, S.J., Moynihan, J., Paw, B.H., Drejer, A., Barut, B., Zapata, A., et al. 2000. Positional cloning of zebrafish ferroportin1 identifies a conserved vertebrate iron exporter. *Nature* 403:776-781.

15. Abboud, S., and Haile, D.J. 2000. A novel mammalian iron-regulated protein involved in intracellular iron metabolism. *J Biol Chem* 275:19906-19912.
16. Nemeth, E., Tuttle, M.S., Powelson, J., Vaughn, M.B., Donovan, A., Ward, D.M., Ganz, T., and Kaplan, J. 2004. Hepcidin regulates cellular iron efflux by binding to ferroportin and inducing its internalization. *Science* 306:2090-2093.
17. Nicolas, G., Chauvet, C., Viatte, L., Danan, J.L., Bigard, X., Devaux, I., Beaumont, C., Kahn, A., and Vaulont, S. 2002. The gene encoding the iron regulatory peptide hepcidin is regulated by anemia, hypoxia, and inflammation. *J Clin Invest* 110:1037-1044.
18. Nemeth, E., Rivera, S., Gabayan, V., Keller, C., Taudorf, S., Pedersen, B.K., and Ganz, T. 2004. IL-6 mediates hypoferremia of inflammation by inducing the synthesis of the iron regulatory hormone hepcidin. *J Clin Invest* 113:1271-1276.
19. Ganz, T., Olbina, G., Girelli, D., Nemeth, E., and Westerman, M. 2008. Immunoassay for human serum hepcidin. *Blood* 112:4292-4297.
20. Lin, L., Valore, E.V., Nemeth, E., Goodnough, J.B., Gabayan, V., and Ganz, T. 2007. Iron transferrin regulates hepcidin synthesis in primary hepatocyte culture through hemojuvelin and BMP2/4. *Blood* 110:2182-2189.
21. Ramey, G., Deschemin, J.C., and Vaulont, S. 2009. Cross-talk between the mitogen activated protein kinase and bone morphogenetic protein/hemojuvelin pathways is required for the induction of hepcidin by holotransferrin in primary mouse hepatocytes. *Haematologica* 94:765-772.

22. Bartnikas, T.B., Andrews, N.C., and Fleming, M.D. 2011. Transferrin is a major determinant of hepcidin expression in hypotransferrinemic mice. *Blood* 117:630-637.
23. Feder, J.N., Penny, D.M., Irrinki, A., Lee, V.K., Lebron, J.A., Watson, N., Tsuchihashi, Z., Sigal, E., Bjorkman, P.J., and Schatzman, R.C. 1998. The hemochromatosis gene product complexes with the transferrin receptor and lowers its affinity for ligand binding. *Proc Natl Acad Sci U S A* 95:1472-1477.
24. Kawabata, H., Fleming, R.E., Gui, D., Moon, S.Y., Saitoh, T., O'Kelly, J., Umehara, Y., Wano, Y., Said, J.W., and Koeffler, H.P. 2005. Expression of hepcidin is down-regulated in Tfr2 mutant mice manifesting a phenotype of hereditary hemochromatosis. *Blood* 105:376-381.
25. Gao, J., Chen, J., Kramer, M., Tsukamoto, H., Zhang, A.S., and Enns, C.A. 2009. Interaction of the hereditary hemochromatosis protein HFE with transferrin receptor 2 is required for transferrin-induced hepcidin expression. *Cell Metab* 9:217-227.
26. Kautz, L., Meynard, D., Monnier, A., Darnaud, V., Bouvet, R., Wang, R.H., Deng, C., Vaulont, S., Mosser, J., Coppin, H., et al. 2008. Iron regulates phosphorylation of Smad1/5/8 and gene expression of Bmp6, Smad7, Id1, and Atoh8 in the mouse liver. *Blood* 112:1503-1509.
27. Andriopoulos, B., Jr., Corradini, E., Xia, Y., Faasse, S.A., Chen, S., Grgurevic, L., Knutson, M.D., Pietrangelo, A., Vukicevic, S., Lin, H.Y., et al. 2009. BMP6 is

- a key endogenous regulator of hepcidin expression and iron metabolism. *Nat Genet* 41:482-487.
28. Xia, Y., Babitt, J.L., Sidis, Y., Chung, R.T., and Lin, H.Y. 2008. Hemojuvelin regulates hepcidin expression via a selective subset of BMP ligands and receptors independently of neogenin. *Blood* 111:5195-5204.
  29. Lee, D.H., Zhou, L.J., Zhou, Z., Xie, J.X., Jung, J.U., Liu, Y., Xi, C.X., Mei, L., and Xiong, W.C. Neogenin inhibits HJV secretion and regulates BMP induced hepcidin expression and iron homeostasis. *Blood*.
  30. Silvestri, L., Pagani, A., Nai, A., De Domenico, I., Kaplan, J., and Camaschella, C. 2008. The serine protease matriptase-2 (TMPRSS6) inhibits hepcidin activation by cleaving membrane hemojuvelin. *Cell Metab* 8:502-511.
  31. Shi, Y., and Massague, J. 2003. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* 113:685-700.
  32. Waite, K.A., and Eng, C. 2003. From developmental disorder to heritable cancer: it's all in the BMP/TGF-beta family. *Nat Rev Genet* 4:763-773.
  33. Massague, J., Seoane, J., and Wotton, D. 2005. Smad transcription factors. *Genes Dev* 19:2783-2810.
  34. Babitt, J.L., Huang, F.W., Wrighting, D.M., Xia, Y., Sidis, Y., Samad, T.A., Campagna, J.A., Chung, R.T., Schneyer, A.L., Woolf, C.J., et al. 2006. Bone morphogenetic protein signaling by hemojuvelin regulates hepcidin expression. *Nat Genet* 38:531-539.

35. Babitt, J.L., Zhang, Y., Samad, T.A., Xia, Y., Tang, J., Campagna, J.A., Schneyer, A.L., Woolf, C.J., and Lin, H.Y. 2005. Repulsive guidance molecule (RGMa), a DRAGON homologue, is a bone morphogenetic protein co-receptor. *J Biol Chem* 280:29820-29827.
36. Samad, T.A., Rebbapragada, A., Bell, E., Zhang, Y., Sidis, Y., Jeong, S.J., Campagna, J.A., Perusini, S., Fabrizio, D.A., Schneyer, A.L., et al. 2005. DRAGON, a bone morphogenetic protein co-receptor. *J Biol Chem* 280:14122-14129.
37. Canalis, E., Economides, A.N., and Gazzerro, E. 2003. Bone morphogenetic proteins, their antagonists, and the skeleton. *Endocr Rev* 24:218-235.
38. Balemans, W., and Van Hul, W. 2002. Extracellular regulation of BMP signaling in vertebrates: a cocktail of modulators. *Dev Biol* 250:231-250.
39. Piccolo, S., Agius, E., Leyns, L., Bhattacharyya, S., Grunz, H., Bouwmeester, T., and De Robertis, E.M. 1999. The head inducer Cerberus is a multifunctional antagonist of Nodal, BMP and Wnt signals. *Nature* 397:707-710.
40. Iemura, S., Yamamoto, T.S., Takagi, C., Uchiyama, H., Natsume, T., Shimasaki, S., Sugino, H., and Ueno, N. 1998. Direct binding of follistatin to a complex of bone-morphogenetic protein and its receptor inhibits ventral and epidermal cell fates in early *Xenopus* embryo. *Proc Natl Acad Sci U S A* 95:9337-9342.
41. Dionne, M.S., Skarnes, W.C., and Harland, R.M. 2001. Mutation and analysis of Dan, the founding member of the Dan family of transforming growth factor beta antagonists. *Mol Cell Biol* 21:636-643.

42. Hsu, D.R., Economides, A.N., Wang, X., Eimon, P.M., and Harland, R.M. 1998. The *Xenopus* dorsalizing factor Gremlin identifies a novel family of secreted proteins that antagonize BMP activities. *Mol Cell* 1:673-683.
43. Tortoriello, D.V., Sidis, Y., Holtzman, D.A., Holmes, W.E., and Schneyer, A.L. 2001. Human follistatin-related protein: a structural homologue of follistatin with nuclear localization. *Endocrinology* 142:3426-3434.
44. Kusu, N., Laurikkala, J., Imanishi, M., Usui, H., Konishi, M., Miyake, A., Thesleff, I., and Itoh, N. 2003. Sclerostin is a novel secreted osteoclast-derived bone morphogenetic protein antagonist with unique ligand specificity. *J Biol Chem* 278:24113-24117.
45. Nakayama, N., Han, C.E., Scully, S., Nishinakamura, R., He, C., Zeni, L., Yamane, H., Chang, D., Yu, D., Yokota, T., et al. 2001. A novel chordin-like protein inhibitor for bone morphogenetic proteins expressed preferentially in mesenchymal cell lineages. *Dev Biol* 232:372-387.
46. Abreu, J.G., Ketpura, N.I., Reversade, B., and De Robertis, E.M. 2002. Connective-tissue growth factor (CTGF) modulates cell signalling by BMP and TGF-beta. *Nat Cell Biol* 4:599-604.
47. Canalis, E., Smerdel-Ramoya, A., Durant, D., Economides, A.N., Beamer, W.G., and Zanotti, S. Nephroblastoma overexpressed (Nov) inactivation sensitizes osteoblasts to bone morphogenetic protein-2, but nov is dispensable for skeletal homeostasis. *Endocrinology* 151:221-233.

48. Truksa, J., Peng, H., Lee, P., and Beutler, E. 2006. Bone morphogenetic proteins 2, 4, and 9 stimulate murine hepcidin 1 expression independently of Hfe, transferrin receptor 2 (Tfr2), and IL-6. *Proc Natl Acad Sci U S A* 103:10289-10293.
49. Wang, R.H., Li, C., Xu, X., Zheng, Y., Xiao, C., Zerfas, P., Cooperman, S., Eckhaus, M., Rouault, T., Mishra, L., et al. 2005. A role of SMAD4 in iron metabolism through the positive regulation of hepcidin expression. *Cell Metab* 2:399-409.
50. Meynard, D., Kautz, L., Darnaud, V., Canonne-Hergaux, F., Coppin, H., and Roth, M.P. 2009. Lack of the bone morphogenetic protein BMP6 induces massive iron overload. *Nat Genet* 41:478-481.
51. Severyn, C.J., Shinde, U., and Rotwein, P. 2009. Molecular biology, genetics and biochemistry of the repulsive guidance molecule family. *Biochem J* 422:393-403.
52. Monnier, P.P., Sierra, A., Macchi, P., Deitinghoff, L., Andersen, J.S., Mann, M., Flad, M., Hornberger, M.R., Stahl, B., Bonhoeffer, F., et al. 2002. RGM is a repulsive guidance molecule for retinal axons. *Nature* 419:392-395.
53. Oldekamp, J., Kramer, N., Alvarez-Bolado, G., and Skutella, T. 2004. Expression pattern of the repulsive guidance molecules RGM A, B and C during mouse development. *Gene Expr Patterns* 4:283-288.
54. Schmidtmer, J., and Engelkamp, D. 2004. Isolation and expression pattern of three mouse homologues of chick Rgm. *Gene Expr Patterns* 4:105-110.

55. Niederkofler, V., Salie, R., Sigrist, M., and Arber, S. 2004. Repulsive guidance molecule (RGM) gene function is required for neural tube closure but not retinal topography in the mouse visual system. *J Neurosci* 24:808-818.
56. Samad, T.A., Srinivasan, A., Karchewski, L.A., Jeong, S.J., Campagna, J.A., Ji, R.R., Fabrizio, D.A., Zhang, Y., Lin, H.Y., Bell, E., et al. 2004. DRAGON: a member of the repulsive guidance molecule-related family of neuronal- and muscle-expressed membrane proteins is regulated by DRG11 and has neuronal adhesive properties. *J Neurosci* 24:2027-2036.
57. Kuninger, D., Kuzmickas, R., Peng, B., Pintar, J.E., and Rotwein, P. 2004. Gene discovery by microarray: identification of novel genes induced during growth factor-mediated muscle cell survival and differentiation. *Genomics* 84:876-889.
58. Kuninger, D., Kuns-Hashimoto, R., Kuzmickas, R., and Rotwein, P. 2006. Complex biosynthesis of the muscle-enriched iron regulator RGMc. *J Cell Sci* 119:3273-3283.
59. Papanikolaou, G., Samuels, M.E., Ludwig, E.H., MacDonald, M.L., Franchini, P.L., Dube, M.P., Andres, L., MacFarlane, J., Sakellaropoulos, N., Politou, M., et al. 2004. Mutations in HFE2 cause iron overload in chromosome 1q-linked juvenile hemochromatosis. *Nat Genet* 36:77-82.
60. Ruoslahti, E. 1996. RGD and other recognition sequences for integrins. *Annu Rev Cell Dev Biol* 12:697-715.
61. Sadler, J.E. 1998. Biochemistry and genetics of von Willebrand factor. *Annu Rev Biochem* 67:395-424.

62. Matsunaga, E., and Chedotal, A. 2004. Repulsive guidance molecule/neogenin: a novel ligand-receptor system playing multiple roles in neural development. *Dev Growth Differ* 46:481-486.
63. Matsunaga, E., Tauszig-Delamasure, S., Monnier, P.P., Mueller, B.K., Strittmatter, S.M., Mehlen, P., and Chedotal, A. 2004. RGM and its receptor neogenin regulate neuronal survival. *Nat Cell Biol* 6:749-755.
64. Rajagopalan, S., Deitinghoff, L., Davis, D., Conrad, S., Skutella, T., Chedotal, A., Mueller, B.K., and Strittmatter, S.M. 2004. Neogenin mediates the action of repulsive guidance molecule. *Nat Cell Biol* 6:756-762.
65. Zhang, A.S., Anderson, S.A., Meyers, K.R., Hernandez, C., Eisenstein, R.S., and Enns, C.A. 2007. Evidence that inhibition of hemojuvelin shedding in response to iron is mediated through neogenin. *J Biol Chem* 282:12547-12556.
66. Zhang, A.S., West, A.P., Jr., Wyman, A.E., Bjorkman, P.J., and Enns, C.A. 2005. Interaction of hemojuvelin with neogenin results in iron accumulation in human embryonic kidney 293 cells. *J Biol Chem* 280:33885-33894.
67. Zhang, A.S., Yang, F., Wang, J., Tsukamoto, H., and Enns, C.A. 2009. Hemojuvelin-neogenin interaction is required for bone morphogenic protein-4-induced hepcidin expression. *J Biol Chem* 284:22580-22589.
68. Kuns-Hashimoto, R., Kuninger, D., Nili, M., and Rotwein, P. 2008. Selective binding of RGMc/hemojuvelin, a key protein in systemic iron metabolism, to BMP-2 and neogenin. *Am J Physiol Cell Physiol* 294:C994-C1003.

69. Zhang, A.S., Yang, F., Meyer, K., Hernandez, C., Chapman-Arvedson, T., Bjorkman, P.J., and Enns, C.A. 2008. Neogenin-mediated hemojuvelin shedding occurs after hemojuvelin traffics to the plasma membrane. *J Biol Chem* 283:17494-17502.
70. Halbrooks, P.J., Ding, R., Wozney, J.M., and Bain, G. 2007. Role of RGM coreceptors in bone morphogenetic protein signaling. *J Mol Signal* 2:4.
71. Xia, Y., Yu, P.B., Sidis, Y., Beppu, H., Bloch, K.D., Schneyer, A.L., and Lin, H.Y. 2007. Repulsive guidance molecule RGMa alters utilization of bone morphogenetic protein (BMP) type II receptors by BMP2 and BMP4. *J Biol Chem* 282:18129-18140.
72. Celec, P. 2005. Hemojuvelin: a supposed role in iron metabolism one year after its discovery. *J Mol Med (Berl)* 83:521-525.
73. Lin, L., Goldberg, Y.P., and Ganz, T. 2005. Competitive regulation of hepcidin mRNA by soluble and cell-associated hemojuvelin. *Blood* 106:2884-2889.
74. Kuninger, D., Kuns-Hashimoto, R., Nili, M., and Rotwein, P. 2008. Pro-protein convertases control the maturation and processing of the iron-regulatory protein, RGMc/hemojuvelin. *BMC Biochem* 9:9.
75. Lin, L., Nemeth, E., Goodnough, J.B., Thapa, D.R., Gabayan, V., and Ganz, T. 2008. Soluble hemojuvelin is released by proprotein convertase-mediated cleavage at a conserved polybasic RNRR site. *Blood Cells Mol Dis* 40:122-131.

76. Silvestri, L., Pagani, A., and Camaschella, C. 2008. Furin-mediated release of soluble hemojuvelin: a new link between hypoxia and iron homeostasis. *Blood* 111:924-931.
77. Maxson, J.E., Chen, J., Enns, C.A., and Zhang, A.S. 2010. Matriptase-2- and proprotein convertase-cleaved forms of hemojuvelin have different roles in the down-regulation of hepcidin expression. *J Biol Chem* 285:39021-39028.
78. Maxson, J.E., Enns, C.A., and Zhang, A.S. 2009. Processing of hemojuvelin requires retrograde trafficking to the Golgi in HepG2 cells. *Blood* 113:1786-1793.
79. Hooper, N.M. 2001. Determination of glycosyl-phosphatidylinositol membrane protein anchorage. *Proteomics* 1:748-755.
80. Silvestri, L., Pagani, A., Fazi, C., Gerardi, G., Levi, S., Arosio, P., and Camaschella, C. 2007. Defective targeting of hemojuvelin to plasma membrane is a common pathogenetic mechanism in juvenile hemochromatosis. *Blood* 109:4503-4510.
81. Pagani, A., Silvestri, L., Nai, A., and Camaschella, C. 2008. Hemojuvelin N-terminal mutants reach the plasma membrane but do not activate the hepcidin response. *Haematologica* 93:1466-1472.
82. Wilson, N.H., and Key, B. 2007. Neogenin: one receptor, many functions. *Int J Biochem Cell Biol* 39:874-878.
83. Wilson, N.H., and Key, B. 2006. Neogenin interacts with RGMa and netrin-1 to guide axons within the embryonic vertebrate forebrain. *Dev Biol* 296:485-498.

84. Yang, F., West, A.P., Jr., Allendorph, G.P., Choe, S., and Bjorkman, P.J. 2008. Neogenin interacts with hemojuvelin through its two membrane-proximal fibronectin type III domains. *Biochemistry* 47:4237-4245.
85. Kautz, L., and Nemeth, E. 2010. Novel tools for the evaluation of iron metabolism. *Haematologica* 95:1989-1991.
86. Babitt, J.L., Huang, F.W., Xia, Y., Sidis, Y., Andrews, N.C., and Lin, H.Y. 2007. Modulation of bone morphogenetic protein signaling in vivo regulates systemic iron balance. *J Clin Invest* 117:1933-1939.
87. Huang, F.W., Pinkus, J.L., Pinkus, G.S., Fleming, M.D., and Andrews, N.C. 2005. A mouse model of juvenile hemochromatosis. *J Clin Invest* 115:2187-2191.
88. Niederkofler, V., Salie, R., and Arber, S. 2005. Hemojuvelin is essential for dietary iron sensing, and its mutation leads to severe iron overload. *J Clin Invest* 115:2180-2186.
89. Ludwiczek, S., Theurl, I., Bahram, S., Schumann, K., and Weiss, G. 2005. Regulatory networks for the control of body iron homeostasis and their dysregulation in HFE mediated hemochromatosis. *J Cell Physiol* 204:489-499.
90. Gleeson, F., Ryan, E., Barrett, S., Russell, J., and Crowe, J. 2007. Hepatic iron metabolism gene expression profiles in HFE associated hereditary hemochromatosis. *Blood Cells Mol Dis* 38:37-44.
91. Theurl, I., Ludwiczek, S., Eller, P., Seifert, M., Artner, E., Brunner, P., and Weiss, G. 2005. Pathways for the regulation of body iron homeostasis in response to experimental iron overload. *J Hepatol* 43:711-719.

92. Krijt, J., Vokurka, M., Chang, K.T., and Necas, E. 2004. Expression of Rgmc, the murine ortholog of hemojuvelin gene, is modulated by development and inflammation, but not by iron status or erythropoietin. *Blood* 104:4308-4310.
93. Mamathambika, B.S., and Bardwell, J.C. 2008. Disulfide-linked protein folding pathways. *Annu Rev Cell Dev Biol* 24:211-235.
94. Sevier, C.S., and Kaiser, C.A. 2002. Formation and transfer of disulphide bonds in living cells. *Nat Rev Mol Cell Biol* 3:836-847.
95. Thornton, J.M. 1981. Disulphide bridges in globular proteins. *J Mol Biol* 151:261-287.
96. Pace, C.N., Grimsley, G.R., Thomson, J.A., and Barnett, B.J. 1988. Conformational stability and activity of ribonuclease T1 with zero, one, and two intact disulfide bonds. *J Biol Chem* 263:11820-11825.
97. Riemer, J., Bulleid, N., and Herrmann, J.M. 2009. Disulfide formation in the ER and mitochondria: two solutions to a common process. *Science* 324:1284-1287.
98. Ostergaard, H., Tachibana, C., and Winther, J.R. 2004. Monitoring disulfide bond formation in the eukaryotic cytosol. *J Cell Biol* 166:337-345.
99. Sevier, C.S., and Kaiser, C.A. 2006. Conservation and diversity of the cellular disulfide bond formation pathways. *Antioxid Redox Signal* 8:797-811.
100. Tu, B.P., Ho-Schleyer, S.C., Travers, K.J., and Weissman, J.S. 2000. Biochemical basis of oxidative protein folding in the endoplasmic reticulum. *Science* 290:1571-1574.

101. Carr, S.A., Annan, R.S. 1997. *Overview of Peptide and Protein Analysis by Mass Spectrometry. Current Protocols in Molecular Biology; 10.21.1-10.21.27.*
102. Jonsson, A.P. 2001. Mass spectrometry for protein and peptide characterisation. *Cell Mol Life Sci* 58:868-884.
103. Mikesch, L.M., Ueberheide, B., Chi, A., Coon, J.J., Syka, J.E., Shabanowitz, J., and Hunt, D.F. 2006. The utility of ETD mass spectrometry in proteomic analysis. *Biochim Biophys Acta* 1764:1811-1822.
104. Wells, J.M., McLuckey, S.A. 2005. *Collision-Induced Dissociation (CID) of Peptides and Proteins. Methods in Enzymology; Vol. 42.*
105. Coon, J.J., Syka, J.E., Shabanowitz, J., and Hunt, D.F. 2005. Tandem mass spectrometry for peptide and protein sequence analysis. *Biotechniques* 38:519, 521, 523.
106. Kolker, E., Higdon, R., and Hogan, J.M. 2006. Protein identification and expression analysis using mass spectrometry. *Trends Microbiol* 14:229-235.
107. Wiesner, J., Premisler, T., and Sickmann, A. 2008. Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications. *Proteomics* 8:4466-4483.
108. Domon, B., and Aebersold, R. 2006. Mass spectrometry and protein analysis. *Science* 312:212-217.
109. Syka, J.E., Coon, J.J., Schroeder, M.J., Shabanowitz, J., and Hunt, D.F. 2004. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* 101:9528-9533.

110. Good, D.M., Wirtala, M., McAlister, G.C., and Coon, J.J. 2007. Performance characteristics of electron transfer dissociation mass spectrometry. *Mol Cell Proteomics* 6:1942-1951.
111. Chrisman, P.A., Pitteri, S.J., Hogan, J.M., and McLuckey, S.A. 2005. SO<sub>2</sub>-\* electron transfer ion/ion reactions with disulfide linked polypeptide ions. *J Am Soc Mass Spectrom* 16:1020-1030.
112. Loo, J.A., Edmonds, C.G., Udseth, H.R., and Smith, R.D. 1990. Effect of reducing disulfide-containing proteins on electrospray ionization mass spectra. *Anal Chem* 62:693-698.
113. Gorman, J.J., Wallis, T.P., and Pitt, J.J. 2002. Protein disulfide bond determination by mass spectrometry. *Mass Spectrom Rev* 21:183-216.
114. Zhang, Z., Pan, H., and Chen, X. 2009. Mass spectrometry for structural characterization of therapeutic antibodies. *Mass Spectrom Rev* 28:147-176.
115. Yen, T.-Y., Macher, B.A. 2006. *Determination of Glycosylation Sites and Disulfide Bond Structures Using LC/ESI-MS/MS Analysis. Methods in Enzymology; Vol. 415.*
116. Wu, S.L., Jiang, H., Lu, Q., Dai, S., Hancock, W.S., and Karger, B.L. 2009. Mass spectrometric determination of disulfide linkages in recombinant therapeutic proteins using online LC-MS with electron-transfer dissociation. *Anal Chem* 81:112-122.

117. Srebalus Barnes, C.A., and Lim, A. 2007. Applications of mass spectrometry for the structural characterization of recombinant protein pharmaceuticals. *Mass Spectrom Rev* 26:370-388.
118. Wang, Y., Lu, Q., Wu, S.L., Karger, B.L., and Hancock, W.S. 2011. Characterization and comparison of disulfide linkages and scrambling patterns in therapeutic monoclonal antibodies: using LC-MS with electron transfer dissociation. *Anal Chem* 83:3133-3140.
119. Wu, S.L., Jiang, H., Hancock, W.S., and Karger, B.L. 2010. Identification of the unpaired cysteine status and complete mapping of the 17 disulfides of recombinant tissue plasminogen activator using LC-MS with electron transfer dissociation/collision induced dissociation. *Anal Chem* 82:5296-5303.
120. Das, R., and Baker, D. 2008. Macromolecular modeling with rosetta. *Annu Rev Biochem* 77:363-382.
121. Eswar, N., Eramian, D., Webb, B., Shen, M.Y., and Sali, A. 2008. Protein structure modeling with MODELLER. *Methods Mol Biol* 426:145-159.
122. Jones, D.T., Taylor, W.R., and Thornton, J.M. 1992. A new approach to protein fold recognition. *Nature* 358:86-89.
123. Bonneau, R., Strauss, C.E., Rohl, C.A., Chivian, D., Bradley, P., Malmstrom, L., Robertson, T., and Baker, D. 2002. De novo prediction of three-dimensional structures for major protein families. *J Mol Biol* 322:65-78.
124. Schwede, T., Sali, A., Eswar, N., Peitsch, M.C. 2008. *Computational Structural Biology - Methods and Applications*.

125. Petsko, G.A., Ringe, D. 2004. *Protein Structure and Function: From Sequence to Consequence*; 4-7.
126. Beattie, J., Allan, G.J., Lochrie, J.D., and Flint, D.J. 2006. Insulin-like growth factor-binding protein-5 (IGFBP-5): a critical member of the IGF axis. *Biochem J* 395:1-19.
127. Schneider, M.R., Wolf, E., Hoeflich, A., and Lahm, H. 2002. IGF-binding protein-5: flexible player in the IGF system and effector on its own. *J Endocrinol* 172:423-440.
128. James, P.L., Jones, S.B., Busby, W.H., Jr., Clemmons, D.R., and Rotwein, P. 1993. A highly conserved insulin-like growth factor-binding protein (IGFBP-5) is expressed during myoblast differentiation. *J Biol Chem* 268:22305-22312.
129. Huang, F.W., Rubio-Aliaga, I., Kushner, J.P., Andrews, N.C., and Fleming, M.D. 2004. Identification of a novel mutation (C321X) in HJV. *Blood* 104:2176-2177.
130. Lee, P.L., Beutler, E., Rao, S.V., and Barton, J.C. 2004. Genetic abnormalities and juvenile hemochromatosis: mutations of the HJV gene encoding hemojuvelin. *Blood* 103:4669-4671.
131. Gehrke, S.G., Pietrangelo, A., Kascak, M., Braner, A., Eisold, M., Kulaksiz, H., Herrmann, T., Hebling, U., Bents, K., Gugler, R., et al. 2005. HJV gene mutations in European patients with juvenile hemochromatosis. *Clin Genet* 67:425-428.
132. Aguilar-Martinez, P., Lok, C.Y., Cunat, S., Cadet, E., Robson, K., and Rochette, J. 2007. Juvenile hemochromatosis caused by a novel combination of hemojuvelin G320V/R176C mutations in a 5-year old girl. *Haematologica* 92:421-422.

133. Koyama, C., Hayashi, H., Wakusawa, S., Ueno, T., Yano, M., Katano, Y., Goto, H., and Kidokoro, R. 2005. Three patients with middle-age-onset hemochromatosis caused by novel mutations in the hemojuvelin gene. *J Hepatol* 43:740-742.
134. Daraio, F., Ryan, E., Gleeson, F., Roetto, A., Crowe, J., and Camaschella, C. 2005. Juvenile hemochromatosis due to G320V/Q116X compound heterozygosity of hemojuvelin in an Irish patient. *Blood Cells Mol Dis* 35:174-176.
135. Hentze, M.W., Muckenthaler, M.U., and Andrews, N.C. 2004. Balancing acts: molecular control of mammalian iron metabolism. *Cell* 117:285-297.
136. Nili, M., Shinde, U., and Rotwein, P. 2010. Soluble repulsive guidance molecule c/hemojuvelin is a broad spectrum bone morphogenetic protein (BMP) antagonist and inhibits both BMP2- and BMP6-mediated signaling and gene expression. *J Biol Chem* 285:24783-24792.
137. Beutler, E. 2006. Hemochromatosis: genetics and pathophysiology. *Annu Rev Med* 57:331-347.
138. Rosenzweig, B.L., Imamura, T., Okadome, T., Cox, G.N., Yamashita, H., ten Dijke, P., Heldin, C.H., and Miyazono, K. 1995. Cloning and characterization of a human type II receptor for bone morphogenetic proteins. *Proc Natl Acad Sci U S A* 92:7632-7636.
139. ten Dijke, P., Yamashita, H., Sampath, T.K., Reddi, A.H., Estevez, M., Riddle, D.L., Ichijo, H., Heldin, C.H., and Miyazono, K. 1994. Identification of type I

- receptors for osteogenic protein-1 and bone morphogenetic protein-4. *J Biol Chem* 269:16985-16988.
140. Kingsley, D.M. 1994. The TGF-beta superfamily: new members, new receptors, and new genetic tests of function in different organisms. *Genes Dev* 8:133-146.
141. Lavery, K., Swain, P., Falb, D., and Alaoui-Ismaili, M.H. 2008. BMP-2/4 and BMP-6/7 differentially utilize cell surface receptors to induce osteoblastic differentiation of human bone marrow-derived mesenchymal stem cells. *J Biol Chem* 283:20948-20958.
142. Keller, S., Nickel, J., Zhang, J.L., Sebald, W., and Mueller, T.D. 2004. Molecular recognition of BMP-2 and BMP receptor IA. *Nat Struct Mol Biol* 11:481-488.
143. Zimmerman, L.B., De Jesus-Escobar, J.M., and Harland, R.M. 1996. The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4. *Cell* 86:599-606.
144. Winkler, D.G., Yu, C., Geoghegan, J.C., Ojala, E.W., Skonier, J.E., Shpektor, D., Sutherland, M.K., and Latham, J.A. 2004. Noggin and sclerostin bone morphogenetic protein antagonists form a mutually inhibitory complex. *J Biol Chem* 279:36293-36298.
145. Aschaffenburg, R., Lewis, M., Phillips, D.C., Press, E.M., Smith, S.G., Sutton, B.J., and Mountford, C.W. 1979. Crystallographic studies of immunoglobulins: crystallization of the Fc fragment of rabbit IgG with and without cleavage of the inter-chain disulphide bridge. *J Mol Biol* 135:1033-1036.

146. Nakae, J., Kido, Y., and Accili, D. 2001. Distinct and overlapping functions of insulin and IGF-I receptors. *Endocr Rev* 22:818-835.
147. Samani, A.A., Yakar, S., LeRoith, D., and Brodt, P. 2007. The role of the IGF system in cancer growth and metastasis: overview and recent insights. *Endocr Rev* 28:20-47.
148. Maki, R.G. 2010. Small is beautiful: insulin-like growth factors and their role in growth, development, and cancer. *J Clin Oncol* 28:4985-4995.
149. Clemmons, D.R. 1998. Role of insulin-like growth factor binding proteins in controlling IGF actions. *Mol Cell Endocrinol* 140:19-24.
150. Mohan, S., and Baylink, D.J. 2002. IGF-binding proteins are multifunctional and act via IGF-dependent and -independent mechanisms. *J Endocrinol* 175:19-31.
151. Daza, D.O., Sundstrom, G., Bergqvist, C.A., Duan, C., and Larhammar, D. 2011. Evolution of the insulin-like growth factor binding protein (IGFBP) family. *Endocrinology* 152:2278-2289.
152. Bach, L.A., Headey, S.J., and Norton, R.S. 2005. IGF-binding proteins--the pieces are falling into place. *Trends Endocrinol Metab* 16:228-234.
153. Hwa, V., Oh, Y., and Rosenfeld, R.G. 1999. The insulin-like growth factor-binding protein (IGFBP) superfamily. *Endocr Rev* 20:761-787.
154. Firth, S.M., and Baxter, R.C. 2002. Cellular actions of the insulin-like growth factor binding proteins. *Endocr Rev* 23:824-854.

155. Zhou, R., Diehl, D., Hoeflich, A., Lahm, H., and Wolf, E. 2003. IGF-binding protein-4: biochemical characteristics and functional consequences. *J Endocrinol* 178:177-193.
156. Neumann, G.M., and Bach, L.A. 1999. The N-terminal disulfide linkages of human insulin-like growth factor-binding protein-6 (hIGFBP-6) and hIGFBP-1 are different as determined by mass spectrometry. *J Biol Chem* 274:14587-14594.
157. Sitar, T., Popowicz, G.M., Siwanowicz, I., Huber, R., and Holak, T.A. 2006. Structural basis for the inhibition of insulin-like growth factors by insulin-like growth factor-binding proteins. *Proc Natl Acad Sci U S A* 103:13028-13033.
158. Sala, A., Capaldi, S., Campagnoli, M., Faggion, B., Labo, S., Perduca, M., Romano, A., Carrizo, M.E., Valli, M., Visai, L., et al. 2005. Structure and properties of the C-terminal domain of insulin-like growth factor-binding protein-1 isolated from human amniotic fluid. *J Biol Chem* 280:29812-29819.
159. Siwanowicz, I., Popowicz, G.M., Wisniewska, M., Huber, R., Kuenkele, K.P., Lang, K., Engh, R.A., and Holak, T.A. 2005. Structural basis for the regulation of insulin-like growth factors by IGF binding proteins. *Structure* 13:155-167.
160. Fernandez-Tornero, C., Lozano, R.M., Rivas, G., Jimenez, M.A., Standker, L., Diaz-Gonzalez, D., Forssmann, W.G., Cuevas, P., Romero, A., and Gimenez-Gallego, G. 2005. Synthesis of the blood circulating C-terminal fragment of insulin-like growth factor (IGF)-binding protein-4 in its native conformation. Crystallization, heparin and IGF binding, and osteogenic activity. *J Biol Chem* 280:18899-18907.

161. Conover, C.A. 2008. Insulin-like growth factor-binding proteins and bone metabolism. *Am J Physiol Endocrinol Metab* 294:E10-14.
162. Mukherjee, A., and Rotwein, P. 2007. Insulin-like growth factor binding protein-5 in osteogenesis: facilitator or inhibitor? *Growth Horm IGF Res* 17:179-185.
163. Mukherjee, A., and Rotwein, P. 2008. Insulin-like growth factor-binding protein-5 inhibits osteoblast differentiation and skeletal growth by blocking insulin-like growth factor actions. *Mol Endocrinol* 22:1238-1250.
164. Salih, D.A., Mohan, S., Kasukawa, Y., Tripathi, G., Lovett, F.A., Anderson, N.F., Carter, E.J., Wergedal, J.E., Baylink, D.J., and Pell, J.M. 2005. Insulin-like growth factor-binding protein-5 induces a gender-related decrease in bone mineral density in transgenic mice. *Endocrinology* 146:931-940.
165. Cobb, L.J., Salih, D.A., Gonzalez, I., Tripathi, G., Carter, E.J., Lovett, F., Holding, C., and Pell, J.M. 2004. Partitioning of IGFBP-5 actions in myogenesis: IGF-independent anti-apoptotic function. *J Cell Sci* 117:1737-1746.
166. Mukherjee, A., Wilson, E.M., and Rotwein, P. 2008. Insulin-like growth factor (IGF) binding protein-5 blocks skeletal muscle differentiation by inhibiting IGF actions. *Mol Endocrinol* 22:206-215.
167. Ning, Y., Schuller, A.G., Bradshaw, S., Rotwein, P., Ludwig, T., Frystyk, J., and Pintar, J.E. 2006. Diminished growth and enhanced glucose metabolism in triple knockout mice containing mutations of insulin-like growth factor binding protein-3, -4, and -5. *Mol Endocrinol* 20:2173-2186.

168. Salih, D.A., Tripathi, G., Holding, C., Szeszak, T.A., Gonzalez, M.I., Carter, E.J., Cobb, L.J., Eisemann, J.E., and Pell, J.M. 2004. Insulin-like growth factor-binding protein 5 (Igfbp5) compromises survival, growth, muscle development, and fertility in mice. *Proc Natl Acad Sci U S A* 101:4314-4319.
169. Ren, H., Yin, P., and Duan, C. 2008. IGFBP-5 regulates muscle cell differentiation by binding to IGF-II and switching on the IGF-II auto-regulation loop. *J Cell Biol* 182:979-991.
170. Hise, M.K., Mantzouris, N.M., Lahn, J.S., Sheikh, M.S., Shao, Z.M., and Fontana, J.A. 1994. IGF binding protein 5 and IGF-I receptor regulation in hypophysectomized rat kidneys. *Am J Physiol* 266:F147-154.
171. Gleason, C.E., Ning, Y., Cominski, T.P., Gupta, R., Kaestner, K.H., Pintar, J.E., and Birnbaum, M.J. 2010. Role of insulin-like growth factor-binding protein 5 (IGFBP5) in organismal and pancreatic beta-cell growth. *Mol Endocrinol* 24:178-192.
172. Kim, K.S., Seu, Y.B., Baek, S.H., Kim, M.J., Kim, K.J., Kim, J.H., and Kim, J.R. 2007. Induction of cellular senescence by insulin-like growth factor binding protein-5 through a p53-dependent mechanism. *Mol Biol Cell* 18:4543-4552.
173. Sorrell, A.M., Shand, J.H., Tonner, E., Gamberoni, M., Accorsi, P.A., Beattie, J., Allan, G.J., and Flint, D.J. 2006. Insulin-like growth factor-binding protein-5 activates plasminogen by interaction with tissue plasminogen activator, independently of its ability to bind to plasminogen activator inhibitor-1, insulin-like growth factor-I, or heparin. *J Biol Chem* 281:10883-10889.

174. Tripathi, G., Salih, D.A., Drozd, A.C., Cosgrove, R.A., Cobb, L.J., and Pell, J.M. 2009. IGF-independent effects of insulin-like growth factor binding protein-5 (Igfbp5) in vivo. *FASEB J* 23:2616-2626.
175. Yasuoka, H., Yamaguchi, Y., and Feghali-Bostwick, C.A. 2009. The pro-fibrotic factor IGFBP-5 induces lung fibroblast and mononuclear cell migration. *Am J Respir Cell Mol Biol* 41:179-188.
176. Zeslawski, W., Beisel, H.G., Kamionka, M., Kalus, W., Engh, R.A., Huber, R., Lang, K., and Holak, T.A. 2001. The interaction of insulin-like growth factor-I with the N-terminal domain of IGFBP-5. *EMBO J* 20:3638-3644.
177. Imai, Y., Moralez, A., Andag, U., Clarke, J.B., Busby, W.H., Jr., and Clemmons, D.R. 2000. Substitutions for hydrophobic amino acids in the N-terminal domains of IGFBP-3 and -5 markedly reduce IGF-I binding and alter their biologic actions. *J Biol Chem* 275:18188-18194.
178. Kalus, W., Zweckstetter, M., Renner, C., Sanchez, Y., Georgescu, J., Grol, M., Demuth, D., Schumacher, R., Dony, C., Lang, K., et al. 1998. Structure of the IGF-binding domain of the insulin-like growth factor-binding protein-5 (IGFBP-5): implications for IGF and IGF-I receptor interactions. *EMBO J* 17:6558-6572.
179. Shand, J.H., Beattie, J., Song, H., Phillips, K., Kelly, S.M., Flint, D.J., and Allan, G.J. 2003. Specific amino acid substitutions determine the differential contribution of the N- and C-terminal domains of insulin-like growth factor (IGF)-binding protein-5 in binding IGF-I. *J Biol Chem* 278:17859-17866.

180. Humphries, J.D., Byron, A., Bass, M.D., Craig, S.E., Pinney, J.W., Knight, D., and Humphries, M.J. 2009. Proteomic analysis of integrin-associated complexes identifies RCC2 as a dual regulator of Rac1 and Arf6. *Sci Signal* 2:ra51.
181. Keshamouni, V.G., Jagtap, P., Michailidis, G., Strahler, J.R., Kuick, R., Reka, A.K., Papoulias, P., Krishnapuram, R., Srirangam, A., Standiford, T.J., et al. 2009. Temporal quantitative proteomics by iTRAQ 2D-LC-MS/MS and corresponding mRNA expression analysis identify post-transcriptional modulation of actin-cytoskeleton regulators during TGF-beta-Induced epithelial-mesenchymal transition. *J Proteome Res* 8:35-47.
182. Schwede, T., Kopp, J., Guex, N., and Peitsch, M.C. 2003. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* 31:3381-3385.
183. Forbes, B.E., Turner, D., Hodge, S.J., McNeil, K.A., Forsberg, G., and Wallace, J.C. 1998. Localization of an insulin-like growth factor (IGF) binding site of bovine IGF binding protein-2 using disulfide mapping and deletion mutation analysis of the C-terminal domain. *J Biol Chem* 273:4647-4652.
184. Osaki, T., Sasaki, K., and Minamino, N. 2011. Peptidomics-based discovery of an antimicrobial peptide derived from insulin-like growth factor-binding protein 5. *J Proteome Res* 10:1870-1880.
185. Fleishman, S.J., Corn, J.E., Strauch, E.M., Whitehead, T.A., Karanicolas, J., and Baker, D. 2011. Hotspot-Centric De Novo Design of Protein Binders. *J Mol Biol.*
186. Khatib, F., Dimaio, F., Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., Zabranska, H., Pichova, I., Thompson, J., Popovic, Z., et al. 2011. Crystal

- structure of a monomeric retroviral protease solved by protein folding game players. *Nat Struct Mol Biol* 18:1175-1177.
187. Lidell, M.E., Johansson, M.E., and Hansson, G.C. 2003. An autocatalytic cleavage in the C terminus of the human MUC2 mucin occurs at the low pH of the late secretory pathway. *J Biol Chem* 278:13944-13951.
188. Kim, D.E., Chivian, D., and Baker, D. 2004. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* 32:W526-531.
189. Lanzara, C., Roetto, A., Daraio, F., Rivard, S., Ficarella, R., Simard, H., Cox, T.M., Cazzola, M., Piperno, A., Gimenez-Roqueplo, A.P., et al. 2004. Spectrum of hemojuvelin gene mutations in 1q-linked juvenile hemochromatosis. *Blood* 103:4317-4321.
190. Cui, Y., Hackenmiller, R., Berg, L., Jean, F., Nakayama, T., Thomas, G., and Christian, J.L. 2001. The activity and signaling range of mature BMP-4 is regulated by sequential cleavage at two sites within the prodomain of the precursor. *Genes Dev* 15:2797-2802.
191. Kaup, M., Dassler, K., Weise, C., and Fuchs, H. 2002. Shedding of the transferrin receptor is mediated constitutively by an integral membrane metalloprotease sensitive to tumor necrosis factor alpha protease inhibitor-2. *J Biol Chem* 277:38494-38502.
192. Park, C.H., Valore, E.V., Waring, A.J., and Ganz, T. 2001. Hepcidin, a urinary antimicrobial peptide synthesized in the liver. *J Biol Chem* 276:7806-7810.



# Appendix

## **Juvenile Hemochromatosis-linked RGMc/Hemojuvelin C-Terminal Truncation Mutants are Cleaved into Smaller Isoforms**

Mahta Nili, David, Kuninger, and Peter Rotwein

Department of Biochemistry and Molecular Biology, Oregon Health & Science  
University, Portland, OR 97239-3098, USA

The author of this dissertation and David Kuninger equally contributed to the experiments and data analysis in this appendix. This work has been previously published (Kuninger, D., Kuns-Hashimoto, R., Nili, M., Rotwein, P. (2007) *BMC Biochemistry* 9:9).

## **A.1 SUMMARY**

In this study, we address questions regarding the biosynthesis and processing of RGMc. Through cell-based approaches we show that RGMc is expressed as a full-length 50 kDa isoform and a 40 kDa soluble species truncated at its C-terminus. Both 50 and 40 kDa RGMc are found in the blood of humans and mice, and in extra-cellular fluid of cultured cells. Several JH frameshift mutations that result in C-terminally truncated RGMc proteins exist. We show that these RGMc mutants are expressed and released by transfected *Cos7* cells, similar to the wild-type protein. Based on the location of the frameshift mutation in relation to a PC recognition site, we show that mutants that contain this site exist as two protein species while mutants that are truncated before the recognition site exist as a single isoform. Thus we provide evidence for a role for PCs in the biogenesis of RGMc.

## A.2 INTRODUCTION

RGMc plays a central role in systemic iron balance, however its mechanisms of action are unknown. Inactivating mutations in the RGMc gene cause JH, a rapidly progressing iron storage disorder with severe systemic manifestations. The liver-derived hormone, hepcidin, is an essential regulator of iron homeostasis that acts by controlling intestinal iron absorption and recovery from macrophages (135). Hepcidin binds to the membrane iron transporter, ferroportin, leading to its degradation (16). In hemochromatosis, hepcidin levels are low, and dietary iron uptake is excessive (137). Soluble RGMc has been proposed to inhibit hepcidin production through the BMP signaling pathway (20, 73).

RGMc undergoes complex biosynthetic steps leading to GPI-linked membrane-bound and soluble forms of the protein, including both 50 and 40 kDa single-chain species. Several JH frameshift mutations give rise to proteins truncated at their C-termini. Due to the lack of GPI-linkages, these proteins only exist as soluble single-chain species.

In this study, we demonstrate that soluble RGMc proteins, both wild-type and C-terminal truncation mutants that retain the PC recognition site, are cleaved by PCs to 40 kDa RGMc. Therefore, these RGMc proteins can exist as two soluble isoforms, a full-length species as the coding region dictates and a 40 kDa isoform. We also show that shorter truncation mutants that do not contain the PC recognition site are released as single protein species. It is possible that these RGMc truncation mutants that are released from

cells may be able to carry out the same role in iron regulation as their wild-type counterparts.

### **A.3 MATERIALS and METHODS**

#### **A.3.1 Materials**

Fetal bovine serum (FBS), Dulbecco's modified Eagle's medium (DMEM), trypsin/EDTA, were purchased from Invitrogen (Carlsbad, CA). The furin convertase inhibitor decanoyl-RVKR-CMK and okadaic acid were purchased from Alexis Biochemicals (San Diego, CA); protease inhibitor tablets were from Roche Applied Sciences (Indianapolis, IN). Restriction enzymes, buffers, ligases, and polymerases were purchased from Roche Applied Sciences, BD Biosciences-Clontech (Palo Alto, CA), and Fermentas (Hanover, MD). Protein A agarose, sodium orthovanadate and dexamethasone were from Sigma-Aldrich (St. Louis, MO); CNBr-activated Sephadex 4B were purchased from Amersham-Pharmacia (Piscataway, NJ). AquaBlock EIA/WIB solution was from East Coast Biologicals (North Berwick, ME). GelCode Blue Stain Reagent was purchased from Pierce Biotechnologies (Rockford, IL). TransIT-LT1 was from Mirus Bio (Madison, WI). NitroBind nitrocellulose was from GE Water & Process Technologies (Trevose, PA). Secondary antibodies, AlexaFluor 680-conjugated goat-anti-rabbit IgG and AlexaFluor 680-conjugated goat-anti-mouse IgG, were purchased from Invitrogen. Other chemicals and reagents were purchased from commercial suppliers.

#### **A.3.2 Cell Culture and Transient Transfections**

All cells were incubated at 37°C in humidified air and 5% CO<sub>2</sub>. Cos7 (CRL-1651) and Hep3B (HB-8064) cells purchased from American Type Culture Collection were maintained in DMEM plus 10% FCS and transfected at ~70% confluent density

with TransIT-LT1. Established murine C3H10T1/2 cells (CCL-226) and C2 myoblasts (58) were grown on gelatin coated dishes in DMEM plus 10% FCS and then incubated at confluent cell density in muscle differentiation-promoting medium, as described (58).

### **A.3.3 Purification of RGMc**

An antibody affinity column was prepared by coupling 1.5 mg of antigen-purified rabbit anti-RGMc IgG to CNBr-activated Sephadex 4B. Serum obtained from two healthy humans (ages 35 and 25) and two male mice (ages 10 and 12 weeks) was diluted into 10 mM TrisHCl, pH 7.4, 0.05% Tween-20, and protease inhibitors. Samples were precleared with protein-A agarose for 4 h at 4°C, then incubated with affinity resin for 16 h at 4°C. Following washes (10 column volumes of 10 mM TrisHCl, pH 7.4, 0.05% Tween-20), proteins were eluted with 0.5 ml of 100 mM glycine, pH 2.5, and neutralized with 1M TrisHCl, pH 8.0. A total of 50 µl was used for detection by SDS-PAGE and immunoblotting.

### **A.3.4 Generation and Expression of RGMc C-Terminal Truncation Mutants**

We previously cloned a mouse RGMc cDNA from skeletal muscle cells (57). Mouse RGMc truncation mutants were made by PCR by replacing codons after R378, C354, S321 and Q305 with a 6X His epitope tag and stop codon. These alterations correspond respectively to human JH-associated frame-shift mutations R385X, C361fsX366, S328fsX337 and Q312X (59, 131, 133, 189). DNA sequencing was used to confirm all nucleotide changes, and the regions with mutations were subcloned into HA-RGMc in pcDNA3 (57). Transient transfections were performed using 2 µg of DNA/35

mm dish, and decanoyl-RVKR-CMK or DMSO were added 18 h later. Cells and medium were harvested after an additional 24 h.

### **A.3.5 Protein Extraction and Immunoblotting**

Whole cell protein lysates were prepared by washing the cells with PBS, scraping them into 200  $\mu$ l RIPA buffer (150 mM NaCl, 50 mM Tris-Cl, 1% NP-40, 0.5% NaDOC, 0.1% SDS, pH 7.5) and drawing them through a 21 gauge needle 5 times such that foaming was minimized. Samples were centrifuged at 14K RPM for 10 minutes at 4°C to remove insoluble material. Protein concentrations of lysates were determined using the Pierce BCA protein assay kit, and aliquots were stored at -80°C until use. Protein samples (25  $\mu$ g/lane) were resolved by SDS-PAGE and transferred to nitrocellulose membranes. After blocking with 50% AquaBlock solution for 1 hr at 20°C, membranes were incubated sequentially with primary and secondary antibodies. Primary antibodies were used at the following dilutions: anti-RGMc, 1:1,000; anti-HA, 1:500; anti-His, 1:500. Secondary antibodies were used at 1:10,000. Results were visualized and images captured using the LiCOR Odyssey and version 3.0 analysis software.

## **A.4 RESULTS**

### **A.4.1 Rational of Experimental Approach**

RGMc undergoes complex biosynthetic steps leading to membrane-bound and soluble forms of the protein, including both 50 and 40 kDa single-chain species. Here, through biochemical and cell-based approaches, we examine the processes that regulate the release of these soluble isoforms. We show that both RGMc50 and 40 are found in the blood of humans and mice, and in extra-cellular fluid of cultured cells. We further show through truncation mapping studies that full-length RGMc50 is cleaved near an evolutionarily conserved PC-recognition site into a 40 kDa soluble species truncated at its C-terminus.

### **A.4.2 Soluble RGMc is Cleaved by PCs**

RGMc is produced by hepatocytes and striated muscle (57, 73, 88). RGMc also accumulates in medium conditioned by muscle cells as 50 and 40 kDa single-chain proteins (58), suggesting either direct secretion or release from the plasma membrane. Here we investigate mechanisms responsible for the appearance of 40 kDa RGMc in extra-cellular fluid. To determine if full-length RGMc is a substrate for PCs, we first purified the soluble protein by antibody affinity chromatography from culture medium from differentiating muscle cells, and from mouse and human blood. Both 50 and 40 kDa RGMc species were detected from all three sources (Figure A.1). Surprisingly, serum concentrations differed substantially between two healthy humans, and between two healthy male mice of the same age and on the same diet, where in one mostly the 50 kDa form of RGMc was detected (Figure A.1, bottom, lane c vs. d). Further study will be required to assess the mechanisms responsible for this variation. When muscle cells, which produce RGMc (58), were incubated with the peptide PC inhibitor, RVKR, the 50

kDa species became the only RGMc protein in the medium (Figure A.2, top), implicating PC activity in processing of endogenous RGMc. Similar results were seen in undifferentiated mouse C2 myoblasts and in human Hep3B cells expressing HA-RGMc (Figure A.2, bottom), although it should be noted that the same dose of RVKR was less effective in Hep3B cells in preventing accumulation of 40 kDa RGMc, implying that there is more PC activity in this cell line.

#### **A.4.4 The Location of PC Cleavage is Mapped by Analysis of JH-linked RGMc Truncation Mutants**

RGMc contains an evolutionarily conserved PC-recognition site between amino acid residues 318 and 327. Several frame-shift mutations in human RGMc associated with JH are located near this PC-recognition motif (59, 131, 133, 189). We mimicked four of these mutants, and generated a series of recombinant NH<sub>2</sub>-terminal HA-tagged mouse RGMc proteins with a 6× His tag at each COOH-terminus (Figure A.3). The mutant proteins expressed in Cos-7 cells, and accumulated in culture medium, but not on the cell membrane, as they lacked a GPI attachment sequence (Figure A.4, and data not shown). A doublet was seen in the medium of cells producing the R378X and C354X mutants, but only a single immunoreactive band was observed with S321X and Q305X. As all protein species were detected with an antibody to the N-terminal HA tag, this result suggests that the smaller mutant proteins (and 40 kDa RGMc derived from RGMcΔGPI) may lack C-terminal residues (Figure A.4). After incubation of cells with RVKR, the larger member of the protein doublet for R378X and C354X increased in abundance in the medium, and was recognized by an antibody to the C-terminal His tag, while the smaller of the doublet bands was not (Figure A.5). Similar results were seen for RGMcΔGPI (Figure A.5). In contrast, detection of S321X and Q305X with the His antibody was constant and was unaffected by RVKR. Taken together with observations

in Figures A.1 and A.2, the results with truncation mutants demonstrate that PC activity potentially removes a C-terminal segment of RGMc, and indicate that the cleavage site is located between amino acids 321 and 354 of the mouse protein, in agreement with identification of a conserved PC motif in this region.

## **A.5 DISCUSSION**

Soluble RGMc50 and RGMc40 are circulating in the blood of humans and mice and are released by cultured muscle cells. In our present work, we have employed biochemical and cell-based approaches and protein mapping studies to assess PC cleavage of full-length RGMc into RGMc40. Our results are discussed below.

### **A.5.1 RGMc is Cleaved by PCs**

The key observation in this appendix section is that a furin-like PC cleaves 50 kDa full-length single-chain RGMc at a conserved site to produce a 40 kDa soluble species. We find that PC activity leads to the accumulation of 40 kDa RGMc in medium of skeletal muscle cells, which endogenously synthesize RGMc (58), and in medium of several cell types expressing recombinant RGMc. Our detection of both 50 and 40 kDa RGMc in human and mouse serum supports a physiological significance of its PC-mediated proteolysis. A recent report showed that soluble 40 kDa RGMc blunted stimulation of hepcidin gene expression by BMP2 in cultured cells (73). This observation indicates that PC activity may influence the biological actions of RGMc.

### **A.5.2 Mapping Studies Localize RGMc Cleavage Site**

Mutations in HJV/RGMc result in the rapidly progressive iron overload disorder, JH (59, 131, 133, 189), implicating this protein in the regulation of whole-body iron metabolism. Several frameshift mutations in the human HJV/RGMc gene that cause JH are predicted to encode truncated RGMc proteins (59, 189). We studied four of these truncated RGMc proteins that terminate at amino acid residues Q305, S321, C354, and

R378. Two of these mutations, Q305 and S321, are located in the RGMc gene 5' to of an evolutionarily conserved predicted PC-recognition site while the latter two mutations, C354 and R378, are 3' to this site. We now find that when the shorter mutants, Q305 and S321, are expressed in Cos-7 cells they give rise to a single protein species; the longer mutants, C354 and R378, give rise to two soluble protein species. These findings map the RGMc cleavage site to its C-terminus at the PC-recognition site.

### **A.5.3 Possible Implications of these Findings on the Role of RGMc in Iron Regulation**

Recent reports from other groups have implicated PCs in the cleavage of RGMc (75, 76). Using Hek293 stably expressing human hemojuvelin/RGMc, Lin et al have found that a PC inhibitor blocks that appearance of a soluble form of the protein (sHJV), with a larger species (ecto-HJV), probably equivalent to 50 kDa mouse RGMc, accumulating instead (75). Silvestri et al have made similar observations in transfected HeLa cells (76). Our detection of both 50 and 40 kDa RGMc in human and mouse serum further supports the physiological significance of its PC-mediated proteolysis. As other studies indicate that furin may be involved in processing and maturation of several iron regulating proteins, including hepcidin, BMPs, and the soluble transferrin receptor (190-192), these observations when taken together imply a diverse and potentially important role for PCs in iron homeostasis.

### **A.5.6 Summary**

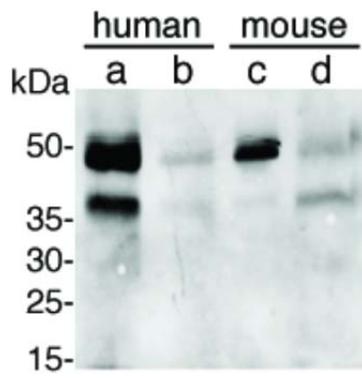
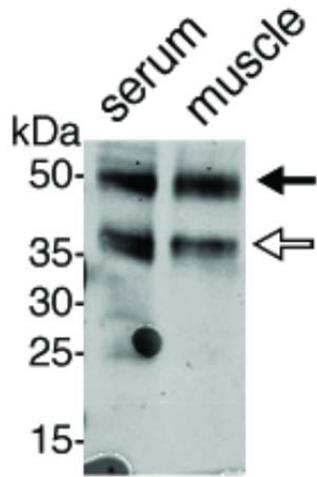
In summary, we show that RGMc accumulates in extra-cellular fluid of cultured cells and in mouse and human serum as 50 and 40 kDa protein species, and demonstrate that a PC inhibitor blocks accumulation of the 40 kDa protein. Finally, through mapping with JH-linked RGMc truncation mutants we are able to delineate the site of C-terminal cleavage of RGMc. Our results define a key role for PCs in the regulation of RGMc that has implications for the physiological effects of this critical iron regulatory protein.

## **A.6 ACKNOWLEDGEMENTS**

We thank Dr. Ujwal Shinde for advice. These studies were supported by NIH grant RO1 DK42748 (P. R.).

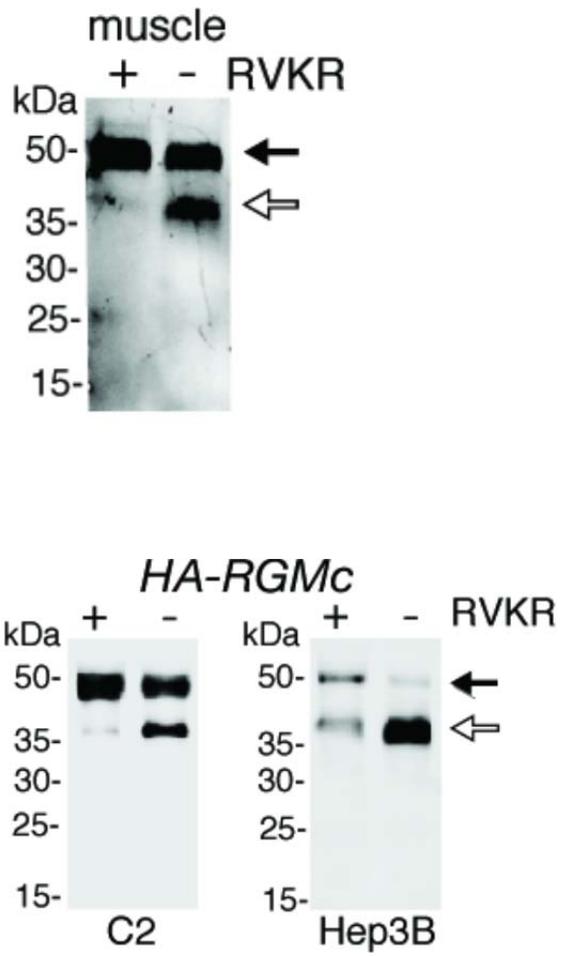
**Figure A.1 Endogenous RGMc exists as soluble 50 and 40 kDa isoforms. Top:**  
Immunoblot of affinity-purified RGMc from mouse serum and muscle cell conditioned medium, with 50 (black arrow) and 40 kDa (white) isoforms indicated. **Bottom:**  
Immunoblot of affinity-purified RGMc from blood from two humans (a, b) and two mice (c, d).

**Figure A.1**



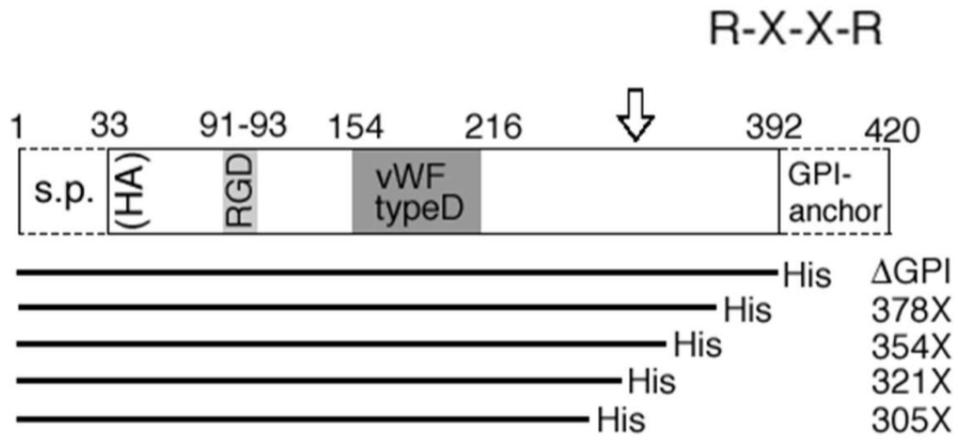
**Figure A.2** A PC inhibitor prevents accumulation of soluble 40 kDa RGMc. **Top:** Immunoblot of RGMc from muscle cell conditioned medium (50  $\mu$ l) incubated without or with the small molecule PC inhibitor, RVKR. **Bottom:** Immunoblot of RGMc from conditioned medium (25  $\mu$ l) of C2 myoblasts or Hep3B cells transfected with HA-RGMc and incubated  $\pm$  decanoyl-RVKR-CMK. Black arrows indicate 50 kDa RGMc and white arrows 40 kDa.

Figure A.2



**Figure A.3 Map of mouse RGMc C-terminal truncation mutants.** Locations of signal peptide (s.p.), RGD sequence, von Willebrandt type D domain, GPI-anchor motif, and putative PC site (R-X-X-R). RGMc $\Delta$ GPI-His and RGMc frame-shift truncation mutants are pictured below. Each contains a C-terminal His tag.

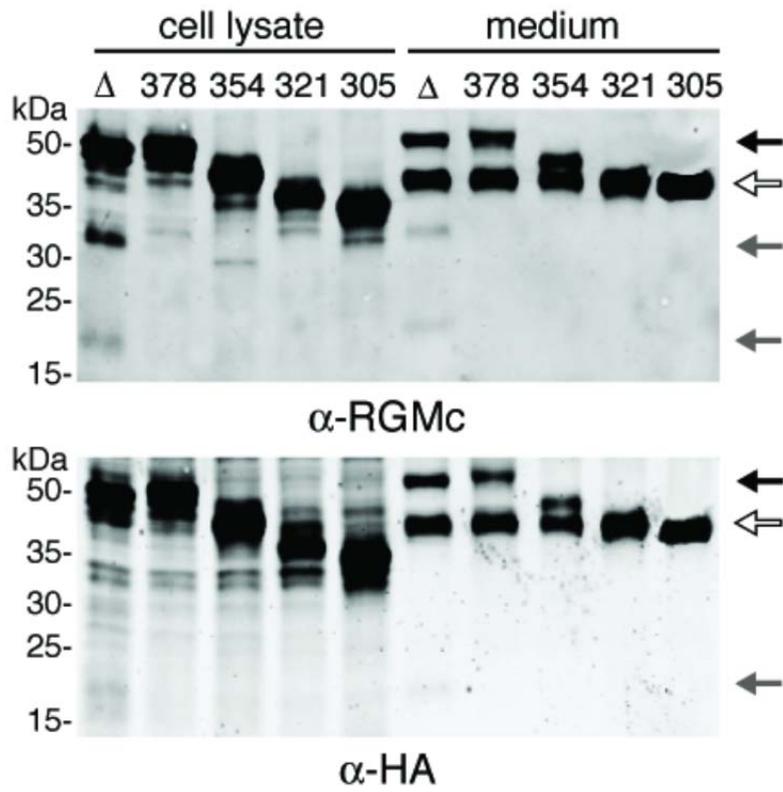
Figure A.3



**Figure A.4 Juvenile Hemochromatosis-linked RGMc C-terminal truncation**

**mutants are cleaved into smaller isoforms.** Detection of RGMc truncation mutants by immunoblotting of protein lysates (left) and conditioned medium (right) from transiently transfected Cos-7 cells using RGMc (upper panel) or HA (lower) antibodies. Black arrows indicate 50 kDa single-chain RGMc and white arrows 40 kDa, and gray arrows mark NH<sub>2</sub>- and COOH-terminal fragments resulting from intra-molecular cleavage.

**Figure A.4**



**Figure A.5 A PC inhibitor affects processing of selected RGMc C-terminal truncation mutants.** Detection of RGMc truncation mutants by immunoblotting of conditioned medium from transiently transfected Cos-7 cells  $\pm$ RVKR using RGMc (upper panel) or His (lower) antibodies. Black arrows indicate 50 kDa single-chain RGMc and white arrows 40 kDa, and gray arrows mark NH<sub>2</sub>- and COOH-terminal fragments resulting from intra-molecular cleavage.

**Figure A.5**

