Published in final edited form as:

Biochem J. 2013 May 15; 452(1): 87–95. doi:10.1042/BJ20121845.

Proteomic Analysis and Molecular Modeling Characterize the Iron-Regulatory Protein, Hemojuvelin/Repulsive Guidance Molecule c

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

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

Abstract

Hemojuvelin (HJV) plays a key role in iron metabolism in mammals by regulating expression of the liver-derived hormone, hepcidin, which controls systemic iron uptake and release. Mutations in HJV cause juvenile hemochromatosis, a rapidly progressing iron overload disorder in humans. HJV, also known as repulsive guidance molecule c (RGMc), is a member of the three-protein RGM family. RGMs are GPI-linked glycoproteins that share ~50% amino acid identity and several structural motifs, including the presence of 14 cysteines in analogous locations. Unlike RGMa and RGMb, HJV/RGMc is composed of both single-chain and two-chain isoforms. To date there is no structural information for any member of the RGM family. Here we have mapped the disulfide bonds in mouse HJV/RGMc using a proteomics strategy combining sequential mass spectrometry (MS) steps composed of electron transfer dissociation (ETD) and collision-induced dissociation (CID), in which ETD induces cleavage of disulfide linkages, and CID establishes disulfide bond assignments between liberated peptides. Our results identify an HJV/RGMc molecular species containing 4 disulfide linkages. We predict using *ab initio* modeling that this molecule is a single-chain HJV/RGMc isoform. Our observations outline a general approach using tandem MS and *ab initio* molecular modeling to define unknown structural features in proteins.

Keywords

hemojuvelin; repulsive guidance molecule; iron metabolism; mass spectrometry; molecular modeling; disulfide bonds; protein structure

Introduction

Iron plays an essential role in human physiology as a cofactor for oxygen transport and is critical for the function of enzymes that regulate cellular respiration and energy metabolism [1, 2]. Because detrimental consequences result from its deficiency or excess, iron homeostasis is tightly controlled, primarily by actions of the liver-derived hormone hepcidin [1–4]. Hereditary hemochromatosis is an autosomal recessive disorder of iron overloading that leads to chronic accumulation of iron throughout the body, causing dysfunction and damage to multiple tissues and organs [1, 2, 5]. Juvenile hemochromatosis represents an

Address correspondence to: Peter Rotwein, Department of Biochemistry and Molecular Biology, Oregon Health and Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97239-3098, Mail code L224, Tel. 503 494-0536, Fax 503-494-8393, rotweinp@ohsu.edu.

Author Contributions

M.N., L.D., U.S., and P.R. conceived of experiments; M.N. performed biochemical experiments; M.N. and L.D. performed mass spectrometry experiments. J.E. and U.S. performed molecular modeling studies. M.N. and P.R. analyzed data; M.N. and P.R. wrote the manuscript; M.N., U.S., and P.R. edited the manuscript. All authors approved the final version of the manuscript.

uncommon but rapidly progressing form of hereditary hemochromatosis, and is caused primarily by mutations in the HJV/HFE2 gene, which encodes hemojuvelin (HJV) [5, 6]. HJV helps regulate hepatic expression of hepcidin, in part by acting as a co-receptor for the growth factor, BMP6, in inducing hepcidin gene transcription [6–8]. More than 30 potentially inactivating mutations have been identified in the HJV gene in patients with juvenile hemochromatosis, including single nucleotide substitutions that alter individual amino acids and DNA frame-shifts that could give rise to truncated HJV proteins [5], and levels of hepcidin are reduced in these patients, leading to excessive and physiologically inappropriate iron absorption from the duodenum [5, 6].

HJV, also known as repulsive guidance molecule c (RGMc), is a member of the three-gene RGM family in mammals [9, 10]. RGMa and RGMb are each expressed in the central nervous system during development [11], and are involved in controlling axonal patterning and neuronal survival [12–16]. In contrast, HJV/RGMc is synthesized in the liver and in cardiac and skeletal muscle [11, 17, 18]. All three RGMs are GPI-linked plasma membrane-associated glycoproteins [9, 10], and share ~50% amino acid identity and several structural motifs [9, 10], including the presence of 14 cysteine residues in comparable locations [9, 10]. Unlike RGMa and RGMb, HJV/RGMc consists of two cell membrane species, a ~50 kDa single-chain protein and a two-chain form with ~30 and ~20 kDa subunits, with the former giving rise to soluble single-chain species of ~50 and ~40 kDa [9, 19, 20] (see Fig. 1A).

Disulfide bonds contribute to the integrity of the three-dimensional structure of proteins and facilitate proper folding [21, 22]. Mass spectrometric (MS) methods for identifying disulfide bonds have improved recently, and in several proteins disulfide linkages have been mapped using a tandem MS approach in which peptides are subjected to electron transfer dissociation (ETD) followed by collision-induced dissociation (CID) [23, 24]. ETD preferentially fragments disulfide bonds rather than the peptide backbone [25, 26]. By contrast, CID cleaves peptides at the amide bond, generating a series of y and b ions [27–29] and rarely fragments disulfide linkages. Since ETD cleaves disulfide bonds, a tandem MS approach may be applied to protein samples without prior chemical reduction [23], with subsequent CID-mediated fragmentation of peptides liberated by ETD leading to the identification of the cysteines involved in disulfide linkages [23].

As a ‘proof-of-principle’ of the utility of this tandem MS approach combining ETD and CID, we recently characterized the disulfide bonds found in mouse IGFBP-5, in which only 2 of 9 linkages had been identified previously [30], and were able to directly map 3 additional disulfide linkages. Moreover, with the assistance of unconstrained *ab initio* molecular modeling we assigned the other 4-disulfide bonds to a cysteine-rich region of the protein [31]. Here we have applied an analogous experimental plan to HJV/RGMc, a protein for which no structural information is available. Our results describe an initial molecular map for a single-chain HJV/RGMc species with 4 disulfide bonds, and show that a combined approach employing sequential MS and molecular modeling can be useful in obtaining new testable structural insights for a protein.

Materials and Methods

Materials

Dulbecco’s modified Eagle’s medium (DMEM), fetal bovine serum (FBS), ultralow IgG FBS, phosphate-buffered saline (PBS), Protein A Sepharose 4B, AcTEV Protease, and trypsin/EDTA solution were purchased from Invitrogen (Carlsbad, CA). TransIT-LT1 was from Mirus Bio (Madison, WI). PEG2000 maleimide, proteomics grade trypsin, and iodoacetamide were from Sigma-Aldrich (St. Louis, MO), and the furin convertase inhibitor,

decanoyl-RVKR-CMK, was from Alexis Biochemicals (San Diego, CA). GelCode Blue Stain Reagent was from Pierce Biotechnologies (Rockford, IL). NiSepharose 6 Fast Flow was from GE-Healthcare (Waukesha, WI). A polyclonal antibody to mouse RGMc has been described [19]. IR800-conjugated goat-anti-human IgG and IRD 800-conjugated anti-rabbit IgG were from Rockland Immunochemical (Gilbertsville, PA). Other chemicals and reagents were purchased from commercial suppliers.

Generation, expression, and purification of RGMc-IgG2-Fc fusion proteins

For generation of the full-length RGMc 50-Fc fusion protein, DNA containing codons 1–391 of mouse RGMc was sub-cloned using 5' EcoRI and 3' NotI sites into pcDNA3 (Invitrogen) in frame with the Fc region of human IgG2 (from pFuse-hIgG2 Fc1, Invitrogen), as described [32]. DNA comprising the codons of a TEV protease recognition site (glutamate-asparagine-leucine-tyrosine-phenylalanine-glutamine), plus several additional codons, was interposed in frame between RGMc and IgG2-Fc segments (see Fig. 2D), 5' to the NotI restriction enzyme site. RGMc truncation mutants were generated by PCR to replace codons after Q318 and C141 with the TEV recognition site and IgG2-Fc to generate RGMc 40-Fc and RGMc 141X-Fc, respectively. DNA sequencing was used to confirm all nucleotide changes. IgG2-Fc was prepared as described [32]. All fusion proteins and IgG2-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. Medium was changed to DMEM with 2% ultralow IgG FBS at 5 hr post-transfection, and conditioned medium was collected 48 hr later. The furin protease inhibitor decanoyl-RVKR-CMK was added to conditioned medium at a final concentration of 5 µM to minimize proteolytic cleavage of RGMc-Fc fusion proteins [20]. Protein purifications were performed by IgG affinity chromatography in batch mode. After incubation of conditioned medium with protein A Sepharose 4B for 16 hr at 4°C, bound proteins were eluted with 100 mM glycine pH 3.0 followed by immediate neutralization with 1 M Tris-Cl pH 8.0. Purification was assessed after SDS-PAGE by staining protein bands with GelCode Blue. Protein concentrations were estimated by comparison with bovine serum albumin standards electrophoresed in adjacent lanes.

Purification of RGMc 40

Purified RGMc 40-Fc was incubated with TEV protease (5 U/10 µg RGMc 40-Fc) for 3 hr at 20°C in buffer supplied by manufacturer. Samples were incubated sequentially with NiSepharose and Protein A Sepharose, each for 2 hr at 20°C to remove TEV and IgG2-Fc, respectively. Purification was assessed and protein concentrations estimated as described above.

Immunoblotting

Protein samples 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. The primary antibody to mouse RGMc was used at a 1:750 dilution. Secondary antibodies were used at 1:4,000. Results were visualized and images captured using the LiCoR Odyssey and version 3.0 analysis software.

Digestion of RGMc-Fc

RGMc fusion proteins (8 µg) were dried by vacuum centrifugation and re-suspended in buffer with 8 M urea, 1M Tris-Cl, pH 8.5, and 8 mM calcium chloride. Free cysteines were either un-modified or were alkylated by incubation with 20 mM iodoacetamide for 15 min protected from light at 20°C. Samples were diluted 4-fold, and trypsin (0.32 µg) was added,

and incubated for 16 hr at 37°C with shaking. Digestions were stopped by addition of 2 μ l of 88% formic acid, and samples were stored at -20°C until analysis.

Localization of disulfide bonds by mass spectrometry

Tryptic peptides were injected onto a 1 mm $\times$ 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 placed in-line with a 0.5 mm $\times$ 250 mm column containing 5 mm Zorbax SB-C18 stationary phase (Agilent Technologies Inc., Santa Clara, CA). Peptides were separated using a 2–30% acetonitrile gradient over 90 min at 10 ml/min with an 1100 series capillary HPLC (Agilent Technologies), and were analyzed using a LTQ Velos linear ion trap with an electron transfer dissociation (ETD) source (Thermo Scientific, San Jose, CA). Electro-spray ionization was performed with 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 first to fifth most intense ions from 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. The minimal signals to initiate MS2 CID and ETD scans were 10,000, the minimal signals for 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, an exclusion list size of 100, and exclusion duration of 30 sec. Bioworks 3.3 (Thermo Scientific) was used to create separate data files for MS2 and MS3 scans, using ZSA charge state analysis. Matching MS2 and MS3 scans to peptide sequences was performed by Sequest (V27, Rev 12, Thermo Scientific), with a database consisting of reversed yeast sequence entries, supplemented with the sequences of common contaminants, and the sequence of mouse RGMc (6182 entries total). The analysis was performed with a parent ion mass tolerance of 2.5, with fragment mass tolerance of 1.0, and with 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 with the program Scaffold (V3_00_08, Proteome Software, Portland, OR) [33, 34] with minimum peptide and protein probabilities of 95% and 99% being used. RGMc 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. Additional confirmation of the identities of cysteine-containing peptides participating in disulfide-linkages was obtained by manual examination of parent ion masses observed in the survey scan and in the ETD MS2 scan.

Molecular modeling of HJV/RGMc

Structural models of amino acids 36 to 320 of mouse RGMc were constructed using Rosetta *ab initio* modeling [35]. The fragment selection procedure was performed *de novo* (without access to structural homologues in the Protein Data Bank (PDB)) using the standard Rosetta fragment server. Decoy structures were generated with the “abinitiorelax” application. Experimentally determined disulfide bonds were enforced by the “fixdisulf” option and by constraining the distance between the cysteine C β atoms to a value between 3 and 5 Å. The 5,000 independent predicted structures 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 α -carbon atoms of all residues when compared with all other simulations in a cluster. The secondary structural depiction was performed using Chimera [36].

Results

An MS approach to mapping disulfide bonds in HJV/RGMc

There are no structural data on HJV/RGMc and no information on the number or organization of disulfide bonds in the protein, in part because there are several different HJV/RGMc protein species [9, 19, 20]. HJV/RGMc biosynthesis is complicated. The 420 amino acid mouse HJV/RGMc primary translation product undergoes co-translational processing of the NH₂-terminal signal peptide and post-translational processing of the COOH-terminal GPI addition sequence to yield a 361 amino acid glycoprotein [9, 19, 20] (Fig. 1). As shown by pulse-chase and cell-surface labeling studies, two distinct GPI-anchored and glycosylated HJV/RGMc molecules are targeted to the cell membrane and appear to undergo different fates [19]. Single-chain 50 kDa HJV/RGMc is released from the cell surface and accumulates in the extracellular fluid as 50 and 40 kDa soluble species with half-lives of over 24 hr [19]. The 40 kDa isoform represents a pro-protein convertase-mediated cleavage product of 50 kDa HJV/RGMc, and contains the first ~325 amino acids of the 50 kDa protein [9, 19, 20] (Fig. 1). In contrast, the other membrane-linked HJV/RGMc isoform, a disulfide-bonded heterodimer composed of NH₂- and COOH-terminal fragments, is not found in the extracellular fluid, and is short-lived, with a half-life of less than 3 hr [19] (Fig. 1).

To begin to evaluate the disulfide bonds in HJV/RGMc, we incubated purified recombinant RGMc 40, which contains 12 cysteines, with or without PEG2000 maleimide, which reacts with free sulfhydryl groups. After a 10-minute incubation, larger molecular mass species could be detected, ranging in size up to ~52 kDa (Supplemental Fig. 1). By contrast, under the same experimental conditions, a control protein, purified human IgG2 Fc fragment, which contains 6 cysteines, was not labeled by PEG2000 maleimide (Supplemental Fig. 1). We interpret these results to indicate that potentially up to 6 of 12 cysteines in RGMc 40 are not involved in disulfide linkages, and that as expected all 6 cysteines in IgG2 Fc are disulfide bonded [23].

We next digested purified HJV/RGMc fusion proteins with trypsin under non-reducing conditions and subjected it to a MS3 analytical protocol with ETD (ETD-MS2) followed by CID (CID-MS3). CID-MS2 also was used to help identify disulfide-linked peptides (Fig. 2A). For these studies we used two purified recombinant HJV/RGMc IgG2 Fc fusion proteins, RGMc 50-Fc and RGMc 141X-Fc, as described [32] (Fig. 2B and C). In initial analyses by CID-MS2 after digestion of RGMc 50-Fc by trypsin, we obtained up to 80% coverage of peptides spanning the entire fusion protein (Fig. 2D), thus validating the overall quality of the starting material.

In previous experiments we had found that both fusion proteins could inhibit the biological effects of BMP2 and BMP6 [32], implying that each was structurally intact, although it is possible that each could undergo disulfide scrambling, as they were prepared by over-expression in mammalian cells. Although we could not prevent any re-arrangements during biosynthesis, to limit any subsequent scrambling we alkylated RGMc with iodoacetamide to block free cysteine residues prior to in-solution trypsin digestion of the purified proteins. We obtained identical disulfide linkage assignments with and without alkylation (data not shown), and also observed highly reproducible chromatographic elution profiles in multiple MS1 experiments, including coverage of all predicted cysteine-containing tryptic peptides except for the 3 smallest ones (P1, P5, P12, Table 1). Therefore, we believe that our data accurately reflect the real disulfide map.

Mapping disulfide bonds in the NH₂-terminal part of HJV/RGMc

P2 and P4, linked tryptic peptides located within the NH₂-terminal segment of HJV/RGMc 50, were fragmented by ETD and CID. These peptides contain cysteine residues 39 and 82, respectively (Fig. 3A and B, parental ion *m/z* 684.7 (+3)). In the ETD-MS2 step, the ion, *m/z* 812.48 (+1) represents P4, and *m/z* 1242.56 (+1) corresponds to P2 (Fig. 3A). Fractionation by CID-MS2 confirmed the assignment of a disulfide linkage between Cys³⁹ in P2 and Cys⁸² in P4, as the disulfide bond was retained in several recovered *y* and *b* ions (e.g., *y**5, *y**4, *y*9, *y*7, *b*3, *b*5) (Fig. 3B). In addition, we observed an identical disulfide linkage to that found in the full-length protein when ETD-MS2 was applied to RGMc 141X, a truncation mutant based on the human disease-causing mutation with a frame-shift after Asp 149 [5] that contains the first 110 amino acids of mature HJV/RGMc 50 and 7 cysteines (Fig. 3C). For additional information, see Table 2 and Supplemental Table 1.

Identifying disulfide linkages within the central region of HJV/RGMc

A disulfide bond between Cys¹⁴¹ and Cys²²³ was identified within the central segment of HJV/RGMc (Fig. 4), flanking the vWF Type D domain in the protein. Tryptic peptide P7 encodes Cys¹⁴¹ and P10 contains Cys²²³ (Fig. 4A). ETD-MS2 of parental ion *m/z* 818.9 (+4) yielded ions *m/z* 1083.7 (+2) representing P7, and *m/z* 1108.4 (+1) corresponding to P10 (Fig. 4A). Fractionation by CID-MS2 confirmed this disulfide bond assignment, based on recovery of multiple *y* and *b* ions that contained the linkage (e.g., *b*19, *y*7, *y**8, *y**6, Fig. 4B, Table 2 and Supplemental Table 1).

Defining other disulfide bonds within the central part of HJV/RGMc

We found that tryptic peptides P8, P9, and P11 within the central segment of RGMc 50 were linked to each other (Fig. 5A). Application of CID-MS3 helped resolve these potentially intertwined disulfide bond-containing peptides. CID-MS3 dissociated the precursor ion *m/z* 1127.3 (+3), which contains P9 and P11 (Fig. 5B), and *m/z* 1272.5 (+3), which contains P8 and P11 (Fig. 5C). In each case, multiple *y* and *b* ions were identified that supported the assignments of linkage of Cys¹⁷⁸ in P9 to Cys³¹⁴ in P11 (e.g., *y*9 and *y*7, Fig. 6B) and Cys¹⁶⁰ in P8 to Cys³¹⁰ in P11 (e.g., *b*14 and *y**11, Fig. 6C). Also see Table 2 and Supplemental Table 2.

Molecular modeling of HJV/RGMc 50

We next used *ab initio* molecular modeling to expand on the information about HJV/RGMc obtained with our tandem MS experiments, focusing on amino acids 36 – 320 of mouse RGMc, which encode all of the disulfide bonds identified by our biochemical studies. We employed *de novo ab initio* modeling using Rosetta [35], as there are no homologous structures in PDB that could be used as templates to guide predictions. Using this approach, and constraining model building by maintaining the experimentally identified disulfide bonds, we were able to generate multiple closely related structures (Fig. 6A). The model predicts a tight globular protein composed of two domains, with the NH₂-terminal segment containing primarily α -helices, and the COOH-terminal domain consisting of a combination of α -helical segments and β -sheets (Fig. 6A, B). We found that disulfide linkages Cys¹⁶⁰ - Cys³¹⁰ and Cys¹⁷⁸ - Cys³¹⁴ are located adjacent to each other, and that there were no additional disulfide bonds beyond the four determined experimentally, observations consistent with our data showing several free cysteine residues in HJV/RGMc (Supplemental Fig. 1). In addition, the site of cleavage responsible for two-chain HJV/RGMc [19] maps to the beginning of a α -helical segment of the protein in close proximity to the Cys¹⁶⁰ - Cys³¹⁰ linkage, and does not appear to be accessible to proteolysis, thus providing supporting evidence that this represents the single-chain HJV/RGMc isoform (Fig. 6A). Taken together, our results support use of molecular modeling as a component of a

combined experimental approach with ETD-MS2 and CID-MS3 for delineating previously undefined disulfide linkages and inferring other structural features in HJV/RGMc.

Discussion

Recent discoveries have refined our understanding of the biochemical processes controlling systemic iron metabolism (reviewed in [1, 2, 5]). The liver-derived hormone, hepcidin, is the critical regulator of iron homeostasis through control of uptake from the duodenum and release of iron from macrophage stores [3, 4]. Hepcidin inhibits the actions of the iron exporter, ferroportin [3, 4]. Levels of systemic iron in turn regulate hepcidin biosynthesis by indirect mechanisms that influence its gene transcription [4, 5].

HJV/RGMc, initially discovered as the mutated gene products responsible for juvenile hemochromatosis [17], is a key component in the mechanisms controlling hepcidin gene expression through its actions as a putative co-receptor for BMP6 in the liver [7, 8]. Single-chain but not two chain HJV/RGMc species can bind BMP2 or BMP6 directly [7, 37–39], and GPI-linked cell membrane-associated HJV/RGMc has been found to enhance the activity of BMP-dependent reporter genes [37] and boost BMP-mediated hepcidin gene transcription [37]. Although fusion proteins composed of RGMa or RGMb plus the IgG Fc domain also can bind selected BMPs [40, 41], and over-expressed RGMa or RGMb can increase BMP-stimulated gene expression [40, 41], the role of interactions with BMPs in the physiological effects of RGMa or RGMb have not been defined yet. In contrast, two-chain RGMa has been shown to control axonal guidance within the central nervous system through direct binding to the trans-membrane receptor neogenin [13, 14]. Of interest, cell-based and biochemical studies have found that two-chain HJV/RGMc binds to neogenin, but that single-chain HJV/RGMc does not [39, 42].

To date no x-ray crystallographic studies or any other structural analyses have been reported for any member of the RGM family. Here we find through application of a tandem MS strategy employing ETD and CID that HJV/RGMc 50 contains 4 disulfide linkages, involving Cys³⁹-Cys⁸², Cys¹⁴¹-Cys²²³, Cys¹⁶⁰-Cys³¹⁰, and Cys¹⁷⁸-Cys³¹⁴. Based on results of *ab initio de novo* molecular modeling studies, in which potential structures were constrained by our MS data, we predict that this isoform comprises the single-chain HJV/RGMc species. Modeling results additionally indicate that this molecule shares several structural features with the presumptive precursor of two-chain HJV/RGMc (data not shown), although the site that is cleaved to generate the two-chain protein appears not to be accessible in the single-chain isoform (Fig. 6A). We postulate that RGMb will be found to be structurally similar to single-chain HJV/RGMc, and that both molecules will be shown to form a three-dimensional complex with selected BMPs, and potentially with BMP receptors [43].

Traditionally, disulfide linkages have been identified in proteins through comparison of results obtained with and without reductive alkylation, with the peptides isolated from the protein after single or multiple proteolytic digestions being subjected to MS or other analyses [44]. Based on reports using tandem MS to identify disulfide bonds in proteins with known disulfide linkage maps, including an immunoglobulin light chain, human growth hormone, and tissue plasminogen activator [23, 24], we recently employed IGFBP-5 as a test case to show that a tandem MS approach coupling ETD and CID steps could be used to characterize disulfide bonds *de novo* [31], and in conjunction with molecular modeling results, could lead to predictions of three-dimensional protein structure. Our current observations now show that this experimental strategy is applicable to a protein in which there is no prior information on either the number or configuration of disulfide linkages. Our results further demonstrate that the iterative combination of targeted proteomics with

molecular modeling provides opportunities for gaining new insights on protein structure. We envision the development of more optimized MS-based approaches for determining the location of disulfide bonds and for defining other features in proteins in which no structural data are available.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Tony Capps, Debra McMillen, and John Klimek for technical advice, and appreciate the assistance of the OHSU Shared Proteomics Resource. HJV/RGMc models were generated using computing equipment in the laboratory of Michael Chapman funded by the OHSU Emerging Technology Fund. This work was supported by NIH research grants 5R01 DK042748-23 (to P. R.), 5R01 EY007755-18 and 5P30 EY10572 (to L. D.), and NSF grant 0746589 (to U. S.). M. N. was supported by NIH pre-doctoral fellowship F31 HL095271-04, and J. E. by a pre-doctoral fellowship from the American Heart Association (12PRE11470005).

Abbreviations

CID	collision-induced dissociation
ETD	electron transfer dissociation
HJV	hemojuvelin
IgG	immunoglobulin G
RGM	repulsive guidance molecule
TEV	tobacco etch virus

References

1. Andrews NC, Schmidt PJ. *Annu Rev Physiol.* 2007; 69:69–85.
2. Pantopoulos K, Porwal SK, Tartakoff A, Devireddy L. *Biochemistry.* 2012
3. Ganz T. *J Am Soc Nephrol.* 2007; 18:394–400.
4. Ganz T, Nemeth E. *Cold Spring Harb Perspect Med.* 2012; 2:a011668.
5. Santos PC, Krieger JE, Pereira AC. *Int J Mol Sci.* 2012; 13:1497–1511. [PubMed: 22408404]
6. Zhang AS. *Adv Nutr.* 2010; 1:38–45. [PubMed: 22043450]
7. Andriopoulos BJ, Corradini E, Xia Y, Faasse SA, Chen S, Grgurevic L, Knutson MD, Pietrangelo A, Vukicevic S, Lin HY, Babitt JL. *Nat Genet.* 2009; 41:482–487. [PubMed: 19252486]
8. Meynard D, Kautz L, Darnaud V, Canonne-Hergaux F, Coppin H, Roth MP. *Nat Genet.* 2009; 41:478–481. [PubMed: 19252488]
9. Severyn CJ, Shinde U, Rotwein P. *Biochem J.* 2009; 422:393–403. [PubMed: 19698085]
10. Corradini E, Babitt JL, Lin HY. *Cytokine Growth Factor Rev.* 2009; 20:389–398. [PubMed: 19897400]
11. Schmidtmer J, Engelkamp D. *Gene Expr Patterns.* 2004; 4:105–110. [PubMed: 14678836]
12. Monnier PP, Sierra A, Macchi P, Deitinghoff L, Andersen JS, Mann M, Flad M, Hornberger MR, Stahl B, Bonhoeffer F, Mueller BK. *Nature.* 2002; 419:392–395. [PubMed: 12353034]
13. Matsunaga E, Tauszig-Delamasure S, Monnier PP, Mueller BK, Strittmatter SM, Mehlen P, Chedotal A. *Nat Cell Biol.* 2004; 6:749–755. [PubMed: 15258591]
14. Rajagopalan S, Deitinghoff L, Davis D, Conrad S, Skutella T, Chedotal A, Mueller BK, Strittmatter SM. *Nat Cell Biol.* 2004; 6:756–762. [PubMed: 15258590]
15. Niederkofler V, Salie R, Sigrist M, Arber S. *J Neurosci.* 2004; 24:808–818. [PubMed: 14749425]

16. Samad TA, Srinivasan A, Karchewski LA, Jeong SJ, Campagna JA, Ji RR, Fabrizio DA, Zhang Y, Lin HY, Bell E, Woolf CJ. *J Neurosci.* 2004; 24:2027–2036. [PubMed: 14985445]
17. Papanikolaou G, Samuels ME, Ludwig EH, MacDonald ML, Franchini PL, Dube MP, Andres L, MacFarlane J, Sakellaropoulos N, Politou M, Nemeth E, Thompson J, Risler JK, Zaborowska C, Babakaiff R, Radomski CC, Pape TD, Davidas O, Christakis J, Brissot P, Lockitch G, Ganz T, Hayden MR, Goldberg YP. *Nat Genet.* 2004; 36:77–82. [PubMed: 14647275]
18. Kuninger D, Kuzmickas R, Peng B, Pintar JE, Rotwein P. *Genomics.* 2004; 84:876–889. [PubMed: 15475267]
19. Kuninger D, Kuns-Hashimoto R, Kuzmickas R, Rotwein P. *J Cell Sci.* 2006; 119:3273–3283. [PubMed: 16868025]
20. Kuninger D, Kuns-Hashimoto R, Nili M, Rotwein P. *BMC Biochem.* 2008; 9:1–9. [PubMed: 18173839]
21. Mamathambika BS, Bardwell JC. *Annu Rev Cell Dev Biol.* 2008; 24:211–235. [PubMed: 18588487]
22. Kosuri P, Alegre-Cebollada J, Feng J, Kaplan A, Ingles-Prieto A, Badilla CL, Stockwell BR, Sanchez-Ruiz JM, Holmgren A, Fernandez JM. *Cell.* 2012; 151:794–806. [PubMed: 23141538]
23. Wu SL, Jiang H, Lu Q, Dai S, Hancock WS, Karger BL. *Anal Chem.* 2009; 81:112–122. [PubMed: 19117448]
24. Wu SL, Jiang H, Hancock WS, Karger BL. *Anal Chem.* 2010; 82:5296–5303. [PubMed: 20481521]
25. Zubarev RA, Horn DM, Fridriksson EK, Kelleher NL, Kruger NA, Lewis MA, Carpenter BK, McLafferty FW. *Anal Chem.* 2000; 72:563–573. [PubMed: 10695143]
26. Chrisman PA, Pitteri SJ, Hogan JM, McLuckey SA. *J Am Soc Mass Spectrom.* 2005; 16:1020–1030. [PubMed: 15914021]
27. Hunt DF, Yates JR 3rd, Shabanowitz J, Winston S, Hauer CR. *Proc Natl Acad Sci USA.* 1986; 83:6233–6237. [PubMed: 3462691]
28. Biemann K. *Methods Enzymol.* 1990; 193:455–479. [PubMed: 2074832]
29. Mikesch LM, Ueberheide B, Chi A, Coon JJ, Syka JE, Shabanowitz J, Hunt DF. *Biochim Biophys Acta.* 2006; 1764:1811–1822. [PubMed: 17118725]
30. Kalus W, Zweckstetter M, Renner C, Sanchez Y, Georgescu J, Grol M, Demuth D, Schumacher R, Dony C, Lang K, Holak TA. *EMBO J.* 1998; 17:6558–6572. [PubMed: 9822601]
31. Nili M, Mukherjee A, Shinde U, David L, Rotwein P. *J Biol Chem.* 2012; 287:1510–1519. [PubMed: 22117064]
32. Nili M, Shinde U, Rotwein P. *J Biol Chem.* 2010; 285:24783–24792. [PubMed: 20530805]
33. Humphries JD, Byron A, Bass MD, Craig SE, Pinney JW, Knight D, Humphries MJ. *Sci Signal.* 2009; 2:ra51. [PubMed: 19738201]
34. Keshamouni VG, Jagtap P, Michailidis G, Strahler JR, Kuick R, Reka AK, Papoulias P, Krishnapuram R, Srirangam A, Standiford TJ, Andrews PC, Omenn GS. *J Proteome Res.* 2009; 8:35–47. [PubMed: 19118450]
35. Bonneau R, Strauss CE, Rohl CA, Chivian D, Bradley P, Malmstrom L, Robertson T, Baker D. *J Mol Biol.* 2002; 322:65–78. [PubMed: 12215415]
36. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. *J Comput Chem.* 2004; 25:1605–1612. [PubMed: 15264254]
37. Babitt JL, Huang FW, Wrighting DM, Xia Y, Sidis Y, Samad TA, Campagna JA, Chung RT, Schneyer AL, Woolf CJ, Andrews NC, Lin HY. *Nat Genet.* 2006; 38:531–539. [PubMed: 16604073]
38. Halbrooks PJ, Ding R, Wozney JM, Bain G. *J Mol Signal.* 2007; 2:4. [PubMed: 17615080]
39. Kuns-Hashimoto R, Kuninger D, Nili M, Rotwein P. *Am J Physiol Cell Physiol.* 2008; 294:C994–C1003. [PubMed: 18287331]
40. Samad TA, Rebbapragada A, Bell E, Zhang Y, Sidis Y, Jeong SJ, Campagna JA, Perusini S, Fabrizio DA, Schneyer AL, Lin HY, Brivanlou AH, Attisano L, Woolf CJ. *J Biol Chem.* 2005; 280:14122–14129. [PubMed: 15671031]

41. Babitt JL, Zhang Y, Samad TA, Xia Y, Tang J, Campagna JA, Schneyer AL, Woolf CJ, Lin HY. J Biol Chem. 2005; 280:29820–29827. [PubMed: 15975920]
42. Yang F, West APJ, Allendorph GP, Choe S, Bjorkman PJ. Biochemistry. 2008; 47:4237–4245. [PubMed: 18335997]
43. Xia Y, Babitt JL, Sidis Y, Chung RT, Lin HY. Blood. 2008; 111:5195–5204. [PubMed: 18326817]
44. Gorman JJ, Wallis TP, Pitt JJ. Mass Spectrom Rev. 2002; 21:183–216. [PubMed: 12476442]

Figure 1. Biosynthesis and processing of HJV/RGMc

Overview of the biosynthesis and processing of HJV/RGMc proteins. The primary translation product undergoes co- and post-translational processing at its NH₂ and COOH ends respectively, to yield two GPI-linked cell membrane-associated glycoproteins. Single-chain RGMc is further cleaved from the cell membrane, resulting in two soluble isoforms of 50 and 40 kDa, RGMc 50 and RGMc 40, respectively. The fate of two-chain cell membrane associated RGMc is unknown. Locations of amino acid domains and cysteines found in mature RGMc isoforms are indicated for the primary translation product (SP - signal peptide, RGD - arginine-glycine-aspartate motif, vWF Type D – von Willebrand factor Type D sequence, PC - cleavage site for pro-protein convertases, GPI - recognition sequence for addition of glycosphosphatidyl inositol anchor). Mature RGMc isoforms are labeled as follows: black squiggles - asparagine-linked sugars, jagged line - GPI-anchor, S-S - disulfide bonds. PPC represents pro-protein convertase.

Figure 2. Experimental approach for mapping disulfide linkages in HJV/RGMc

A. Analytical approach for identification of disulfide linkages in RGMc by MS. See ‘Materials and Methods’ for details. **B.** RGMc fusion proteins. Top panel: primary mouse RGMc translation product with locations of amino acid domains and cysteines found in mature protein isoforms indicated; second panel: single chain RGMc 50; third and fourth panels: RGMc-IgG2 Fc fusion proteins: RGMc 50-Fc and RGMc 141X-Fc (SP - signal peptide, RGD - arginine-glycine-aspartate motif, vWF Type D – von Willebrand factor Type D sequence, PC - cleavage site for pro-protein convertases, GPI - recognition sequence for addition of glycosphosphatidyl inositol anchor, TEV site - recognition and cleavage sequence for tobacco etch virus protease). **C.** Analysis of purified IgG2 Fc, RGMc 50-Fc, and RGMc 141X-Fc after reducing SDS-PAGE by staining with GelCode blue. Black arrows point to single-chain RGMc-Fc proteins and the gray arrows to two-chain RGMc 50-Fc. **D.** Identification of tryptic peptides from RGMc 50-Fc by CID-MS2. All identified peptides are underlined. Cysteines are highlighted (dark gray for RGMc, light gray for IgG2 Fc) and the linker region between RGMc 50 and IgG2 Fc is italicized (the TEV cleavage site is in bold script). Results are a composite of 3 independent experiments with 70–75% coverage per experiment.

Figure 3. Defining disulfide linkages in the NH₂-terminal part of HJV/RGMc
Peptides P2 and P4 are linked by a disulfide bond. See Supplemental Table 1 for additional information. **A.** ETD-MS2 of precursor ion, m/z 684.7 (+3), corresponding to P2–P4, and containing cysteine residues 39 and 82 in RGMc 50. **B.** CID-MS2 of precursor ion, m/z 684.7 (+3), of disulfide-linked peptides P2 and P4 in RGMc 50. The y ions marked with an asterisk are from P4; other y and b ions are from P2. **C.** ETD-MS2 of precursor ion m/z 685.2 (+3) of disulfide-linked peptides containing cysteine residues 39 and 82 from RGMc 141X. In **A** and **C** [M] denotes the parent mass. All labeled ions are from RGMc 50 or RGMc 141X.

Figure 4. Defining disulfide bonds in the central segment of HJV/RGMc
Peptides P7 and P10 are joined by a disulfide linkage. See Supplemental Table 1 for additional information. **A.** ETD-MS2 of precursor ion, m/z 818.9 (+4), corresponding to P7-P10, and containing cysteines 141 and 223 in RGMc 50. The parent mass is depicted by [M]. **B.** CID-MS2 of precursor ion, m/z 818.9 (+4), of disulfide-linked peptides P7 and P10 in RGMc 50. The y ions marked with an asterisk are from P10; other y and b ions are from P7. All labeled ions are from RGMc 50.

Figure 5. Defining other disulfide linkages in HJV/RGMc

Peptides P8-P11-P9 are linked together by disulfide bonds. See Supplemental Table 2 for additional information. **A.** ETD-MS2 of precursor ion, m/z 728.2 (+7), corresponding to P8-P11-P9, and containing cysteine residues 160, 178, 310, and 314. **B.** CID-MS3 of precursor ion, m/z 1272.5 (+3), of disulfide-linked peptides P8 and P11. The source ion from ETD-MS2, m/z 1019.8 (+5), is indicated. The y and b ions labeled with an asterisk are from P8; the others are from P11. **C.** CID-MS3 of precursor ion, m/z 1127.3 (+3), corresponding to disulfide-linked peptides P11 and P9. The source ion from ETD-MS2, m/z 728.2 (+7), is

indicated. The y ions marked with an asterisk are from P9; other y and b ions are from P11. All labeled ions are from RGMc 50.

A **B**

Figure 6. Structural predictions and disulfide bond map of HJV/RGMc

Ab initio modeling of amino acids 36 - 320 of mouse RGMc. **A.** Model of putative single-chain RGMc, with four disulfide bonds indicated. Cysteines comprising disulfide linkages are labeled and color-coded, α -helices are depicted as cylinders, β -sheets as flattened segments with arrows, and unstructured regions as thin lines. Other features indicated include an RGD sequence, the location of the internal cleavage site that generates two-chain RGMc, and the NH_2 - and COOH -termini of the protein (N-term and C-term, respectively) (see Experimental Procedures for details). **B.** Secondary structure diagram of the *ab initio* model depicted in panel A, with α -helices and β -sheets indicated. The RGD motif is labeled in green and the internal cleavage site is in blue. Experimentally derived disulfide bonds are indicated as dashed lines, with the colors of cysteine residues and lines corresponding to those in panel A.

Table 1

m/z Values of RGMc Cysteine-Containing Tryptic Peptides

Cysteines	Peptide #	Peptides	+1	+2	+3	+4	+5
Cys 34	P1	QCK					
Cys 39	P2	CNAEYVSSTLR	1242.6	621.8	-	-	-
Cys 73	P3	GGLASGGLCR	890.4	445.7	-	-	-
Cys 82	P4	SYALCTR	813.4	407.2	-	-	-
Cys 90	P5	TCR	379.2	190.1	-	-	-
Cys 112	P6	GDLAFHSAVHGIEDLMIQHNC ⁺ SR	2550.2	1275.6	850.7	638.3	510.8
Cys 141	P7	GPALPGAGAPLTPDP ⁺ CDYEAR	2165.0	1083.0	-	-	-
Cys 160	P8	APGFLHCASF ⁺ GD ⁺ PHVR	1710.8	855.9	570.9	428.5	-
Cys 178	P9	SFHNQFHTCR	1276.6	638.8	426.2	319.9	-
Cys 223	P10	NMQECIDQK	1108.5	554.7	-	-	-
Cys 310/314	P11	AFSAEQDLQLCVGG ⁺ CPPSQR	2106.0	1053.5	-	-	-
Cys 340	P12	LCK	363.2	182.1	-	-	-
Cys 354	P13	EGLPVEDAYFQSCVFDVSVSGDPNFTVAAQTALDDAR	3933.8	1967.4	-	-	-

Table 2

m/z values of RGMc disulfide-linked tryptic peptides

Peptide #	Peptides	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
P2-P4	CNAEYVSS ¹ TLR - SYAL ² CTR	2053.0	1027.0	685.0	514.0	-	-	-	-	-	-
P7-P10	GPALPGAGPAPLTPDP ³ C ⁴ YEAR-NMQECIDQK	3270.5	1635.8	1090.8	818.4	654.9	-	-	-	-	-
P8-P9-P11	APGFLHCASF ⁵ GDPHVR-SFHNQFHT ⁶ CR-AFSAEQDLQL ⁷ CVGGC ⁸ PPSQR	5087.3	2544.2	1696.4	1272.6	1018.3	848.7	727.6	636.8	566.1	509.6
P8-P11	APGFLHCASF ⁵ GDPHVR-AFSAEQDLQL ⁷ CVGGC ⁸ PPSQR	3813.8	1907.4	1271.9	954.2	763.6	636.5	-	-	-	-
P9-P11	SFHNQFHT ⁶ CR-AFSAEQDLQL ⁷ CVGGC ⁸ PPSQR	3379.5	1690.3	1127.2	845.6	676.7	564.1	-	-	-	-