3D Structure of the Biomarker Hepcidin-25 in its Native State

Dissertation
zur Erlangung des akademischen Grades
Doctor rerum naturalium (Dr.rer.nat.)
in der Wissenschaftsdisziplin “Analytische Chemie”

vorgelegt der
Mathematisch-Naturwissenschaftlichen Fakultät
Institut für Chemie
der Universität Potsdam

von

M.Sc. Marija Vranic

Potsdam, 17. March 2020

https://doi.org/10.25932/publishup-45929
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-459295

Diese Arbeit ist vom Oktober 2015 bis Oktober 2019 als kooperatives Projekt an der Universität Potsdam, der Bundesanstalt für Materialforschung und -prüfung und der Humboldt Universität zu Berlin im Rahmen der Graduiertenschule School of Analytical Sciences Adlershof entstanden.

Gutachter: Prof. Dr. Heiko Moeller
PD Dr. Michael G. Weller
Prof. Dr. Valentin Wittmann

Acknowledgments

Obtaining a PhD was an exciting journey, and there are many people to whom I owe a debt of gratitude.

First, I would like to express my sincere gratitude to my PhD supervisors Prof. Dr. Heiko Moeller and PD Dr. Michael G. Weller for the opportunity of working with them and their continuous support during my PhD research.

Most of all, I want to express my deepest appreciation to Prof. Dr. Heiko Moeller from whom I learned a lot during these four years, for all the scientific and non-scientific discussions, the positive attitude towards research, for helping me develop ideas and see things from a different perspective. I could not have imagined having a better and more understanding mentor for my PhD.

My sincere thanks to the many past and present members of the AK Moeller Group. My sincere thanks to Dr. Ines Starke and Sylvia Fürstenberg for helping me with MS measurements, Dr. Matthias Heydenreich and Angela Krtitschka for all support, suggestions and help with NMR measurements, Dr. Andreas Koch and Dr. Steffen Thomas for constructive discussion and all help in very different problems. I am especially grateful to Dr. Ruslan Nedielkov and Marcus for teaching me different techniques and the beauty of recombinant protein expression, to Marcel and Öznur, Monique, Christin, Stefan and Magdalena for stimulating discussions and encouragement, and for all the fun we had in the last four years.

I would also like to thank many people from the Weller group, especially to Dr. Ioana-Monica Abbas, for productive collaboration and discussion.

I would also like to extend my sincere gratitude to the SALSA team, Katharina Gliege, Dr. Virginia Merk, Katharina Schulthens, Kerstin Steinberg, Stefanie Sellon, Esther Santel, Kristin Kuschnerik and Pablo Lores Lareo for their help with all the administration work and the great time together. Also, a big thanks goes to our group secretary, Karis Linke, for making all administration work easy.

I am very grateful for spending the last four years surrounded by extraordinary people who made this journey, even in difficult situations, much easier. To all my friends, I will say thank you a lot for give me the chance to spend all this time with you, sharing good and bad times.

My greatest thanks go to my family for their emotional and financial support during my entire education and for creating many opportunities in my personal development!

Table of contents

Abstract

Hepcidin-25 (Hep-25) plays a crucial role in the control of iron homeostasis. Since the dysfunction of the hepcidin pathway leads to multiple diseases as a result of iron imbalance, hepcidin represents a potential target for the diagnosis and treatment of disorders of iron metabolism. Despite intense research in the last decade targeted at developing a selective immunoassay for iron disorder diagnosis and treatment and better understanding the ferroportin-hepcidin interaction, questions remain. The key to resolving these underlying questions is acquiring exact knowledge of the 3D structure of native Hep-25. Since it was determined that the N-terminus, which is responsible for the bioactivity of Hep-25, contains a small Cu(II)-binding site known as the ATCUN motif, it was assumed that the Hep-25-Cu(II) complex is the native, bioactive form of the hepcidin. This structure has thus far not been elucidated in detail. Owing to the lack of structural information on metal-bound Hep-25, little is known about its possible biological role in iron metabolism. Therefore, this work is focused on structurally characterizing the metal-bound Hep-25 by NMR spectroscopy and molecular dynamics simulations.

For the present work, a protocol was developed to prepare and purify properly folded Hep-25 in high quantities. In order to overcome the low solubility of Hep-25 at neutral pH, we introduced the C-terminal DEDEDE solubility tag. The metal binding was investigated through a series of NMR spectroscopic experiments to identify the most affected amino acids that mediate metal coordination. Based on the obtained NMR data, a structural calculation was performed in order to generate a model structure of the Hep-25-Ni(II) complex. The DEDEDE tag was excluded from the structural calculation due to a lack of NMR restraints. The dynamic nature and fast exchange of some of the amide protons with solvent reduced the overall number of NMR restraints needed for a high-quality structure. The NMR data revealed that the 20 C-terminal Hep-25 amino acids experienced no significant conformational changes, compared to published results, as a result of a pH change from pH 3 to pH 7 and metal binding. A 3D model of the Hep-25-Ni(II) complex was constructed from NMR data recorded for the hexapeptide-Ni(II) complex and Hep-25-DEDEDE-Ni(II) complex in combination with the fixed conformation of 19 C-terminal amino acids. The NMR data of the Hep-25-DEDEDE-Ni(II) complex indicates that the ATCUN motif moves independently from the rest of the structure. The 3D model structure of the metal-bound Hep-25 allows for future works to elucidate hepcidin's interaction with its receptor ferroportin and should serve as a starting point for the development of antibodies with improved selectivity.

Zusammenfassung

Hepcidin-25 (Hep-25) spielt eine entscheidende Rolle bei der Kontrolle der Eisenhomöostase. Da die Dysfunktion des Hepcidin-Signalweges aufgrund des Eisenungleichgewichts zu mehreren Krankheiten führt, stellt Hepcidin ein potenzielles Ziel für die Diagnose und Behandlung von Störungen des Eisenstoffwechsels dar. Trotz intensiver Forschung in den letzten zehn Jahren, die darauf abzielte, einen selektiven Immunoassay für die Diagnose und Behandlung von Eisenerkrankungen zu entwickeln und die Ferroportin-Hepcidin-Interaktion besser zu verstehen, bleiben Fragen offen. Der Schlüssel zur Lösung dieser grundlegenden Fragen liegt darin, genaue Kenntnisse über die 3D-Struktur des nativen Hep-25 zu erlangen. Da festgestellt wurde, dass der N-Terminus, der für die Bioaktivität von Hep-25 verantwortlich ist, eine kleine Cu(II)-Bindungsstelle enthält, die als ATCUN-Motiv bekannt ist, wurde angenommen, dass der Hep-25- Cu(II)-Komplex die native, bioaktive Form des Hepcidins ist. Diese Struktur ist bisher noch nicht im Detail untersucht worden. Aufgrund fehlender Strukturinformationen über metallgebundenes Hep-25 ist wenig über seine mögliche biologische Rolle im Eisenstoffwechsel bekannt. Daher konzentriert sich diese Arbeit auf die strukturelle Charakterisierung des metallgebundenen Hep-25 mittels NMR-Spektroskopie und Molekulardynamik Simulationen.

In der vorliegenden Arbeit wurde ein Protokoll zur Präparation und Reinigung von korrekt gefaltetem Hep-25 in hohen Mengen entwickelt. Um das Problem der geringen Löslichkeit von Hep-25 bei neutralem pH-Wert zu überwinden, haben wir einen C-terminalen DEDEDEDE Löslichkeits-Tag eingeführt. Die Metallbindung wurde durch eine Reihe von NMR-spektroskopischen Experimenten untersucht, um die Aminosäuren zu identifizieren, welche an der Metallkoordination beteiligt sind. Basierend auf den erhaltenen NMR-Daten wurde eine Strukturberechnung durchgeführt, um eine Modellstruktur des Hep-25-Ni(II)-Komplexes zu erzeugen. Der DEDEDE-Tag wurde aufgrund fehlender NMR- restraints von der Strukturberechnung ausgeschlossen. Die dynamische Natur und der schnelle Austausch eines Teils der Amid-Protonen mit dem Lösungsmittel reduzierten die Gesamtzahl der NMR-restraints, die für eine hochwertige Struktur erforderlich waren. Die NMR-Daten zeigten, dass die 20 C-terminalen Hep-25-Aminosäuren keine signifikanten Konformationsänderungen als Folge eines pH-Wechsels von pH 3 auf pH 7 und einer Metallbindung erfuhren. Ein 3D-Modell des Hep-25-Ni(II)-Komplexes wurde aus den NMR-Daten des Hexapeptid-Ni(II)-Komplexes und des Hep-25-DEDEDE-Ni(II)-Komplexes in Kombination mit der bekannten Konformation

der 19 C-terminalen Aminosäuren erstellt. Die NMR-Daten des Hep-25-DEDEDE-Ni(II)-Komplexes zeigen, dass sich das Ni-ATCUN-Motiv unabhängig vom C-Terminus bewegt. Die 3D-Modellstruktur des metallgebundenen Hep-25 ermöglicht es, in Zukunft die Interaktion von Hepcidin mit seinem Rezeptor Ferroportin zu untersuchen und soll als Ausgangspunkt für die Entwicklung von Antikörpern mit verbesserter Selektivität dienen.

Abbreviations

1D / 2D / 3D	One-, Two-, Three-Dimensional
ACN	Acetonitrile
APS	Ammonium persulfate
ATCUN	Amino terminal Cu(II)- and Ni(II)-binding motif
ATP	Adenosine triphosphate disodium salt
BMP	Bone morphogenetic protein
BSA	Bovine serum albumin
ccdB	Control of Cell Death B (gene)
CID	Collision-induced dissociation
Da	Dalton
DMT1	Divalent metal transport protein
DNA	Deoxyribonucleic acid
DNAse I	Deoxyribonuclease I
dNTP	Deoxynucleotide
DTT	Dithiothreitol
ESI	Electrospray ionization
Fe-Tf	Transferrin-bound iron
FPLC	Fast protein liquid chromatography
FPN	Ferroportin
GSSG	Oxidized Glutathione
GSH	Reduced Glutathione
GST	Glutathione S-transferase
HAMP	Hepcidin antimicrobial peptide
Hb	Hemoglobin
HCP1	Heme carrier protein 1
Hep-25	Hepcidin-25
HFE	Human hemochromatosis protein
HJV	Haemojuvelin
HPLC	High-performance liquid chromatography
HSQC	Heteronuclear single quantum coherence

His6 tag	Hexa-histidine tag
IL6	Interleukin-6
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRP	Iron regulatory proteins
IRE	Iron responsive elements
Leap1	Liver-expressed antimicrobial peptide
MALDI-TOF	Matrix-assisted laser desorption/ionization - Time-of-flight
MBP	Maltose-binding protein
MS	Mass spectrometry
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
OmpA	Outer membrane protein A
PCR	Polymerase chain reaction
PDB	Protein data bank
PelB	Periplasmic pectate lyase B
PSF	Protein structure file
RMSD	Root mean squared deviation
RNA	Ribonucleic acid
ROESY	Rotating frame Overhauser enhancement spectroscopy
Sap1	Secreted Aspartyl Proteinase 1
SDS PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
STAT 3	Signal transducer and activator of transcription 3
SUMO	Small ubiquitin-like modifier
TFA	Trifluoroacetic acid
TFR	Transferrin receptor
TOCSY	Total correlation spectroscopy
Trx	Thioredoxin
ULP1	Ubiquitin-like protein specific 1
UV–Vis	Ultraviolet-visible (spectroscopy)

Amino acids

Name	Three letter code	One letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic Acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucin	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

1 Introduction

1.1 Iron metabolism

Iron is a critical nutrient for various metabolic processes in the human body, including oxygen transport, deoxyribonucleic acid (DNA) synthesis, and electron transport [1, 2]. It is distributed in the body in three categories, functional, storage, and transport iron, and it is present as heme and non-heme iron. A human adult body contains up to 80% of iron present in heme- and non-heme-containing proteins. Out of that percentage, about 70% is found in the erythrocyte in the form of hemoglobin [3, 4]. The remaining iron, approximately 20%, is stored within macrophages and hepatocytes [3].

Iron metabolism is an essential pathway, and iron homeostasis is crucial, as iron deficiency or overload leads to the development of several pathologies. To maintain iron homeostasis, the organism controls iron metabolism at different systemic and cellular level levels and with diverse mechanisms [5]. Cellular and systemic iron imbalances are detrimental, and so these processes require tight regulation by a multitude of regulators, transporters, and enzymes, which are involved in iron absorption, distribution, storage, and recycling action [6].

The cellular iron level regulates the expression of proteins involved in the mechanisms of iron uptake, export, and intracellular use [6]. Cellular iron metabolism is mainly controlled by iron regulatory proteins (IRPs) that bind iron-responsive elements (IREs), which are located in the *mRNAs*, and block the translation of protein. Regulation systems for IRP and IRE were found in various of proteins involved in iron transport and storage [6, 7]. The regulation system for cellular IRP and IRE controls the expression of transferrin (Tf), ferroportin (FPN), and ferritin. For example, in a case of iron deficiency in cells, IRPs bind to the IRE of ferritin mRNA and inhibit the translation of ferritin. This results in the increased expression of transferrin receptor 1 (TfR1). The inhibition of ferritin further enables the cells to increase iron uptake by Tf and minimize the storage of iron in ferritin [7, 8].

Systemic iron homeostasis is regulated through signaling mechanisms between iron demand and iron supply [9]. Systemic iron levels control the release from macrophages, which recycle senescent red blood cells; the release of stored iron from hepatocytes; iron absorption from the environment; and iron export from absorptive cells to cells where it can be utilized or stored [10].

The interface between cellular and systemic iron homeostasis is observed in four main cell types: duodenal enterocytes, erythrocytes, macrophages, and hepatocytes [11]. Iron metabolism can be understood primarily based on how iron is handled by these four main cell types (Figure 1.1).

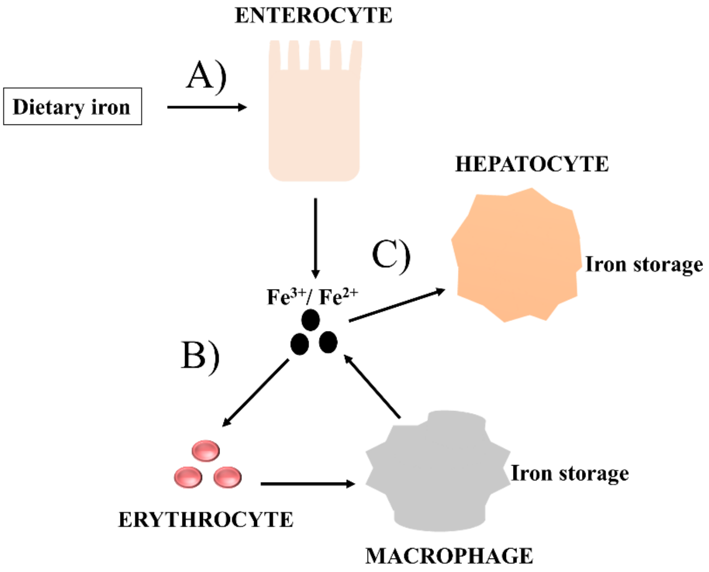


Figure 1.1: Iron distribution in the human body. A) The absorption of dietary iron occurs in the duodenal enterocytes. Following iron absorption, iron has two possible routes, and it can be utilized or stored. B) Most of the body's iron is utilized in the hemoglobin of developing erythroid precursors and mature red blood cells. Macrophages recycle iron from ingested senescent erythrocytes and rerelease it into plasma for storage or use. C) Excess of body iron is stored in hepatocytes, the leading iron storage site in the body. The sketch for the picture was taken from Kaplan at al. [12].

Iron absorption plays a crucial role in iron balance because of the deficiency of well-developed and regulated pathways for iron excretion [12]. Iron excretion occurs only during menstruation, the shedding of hair and skin cells, and the rapid turnover and excretion of enterocytes [13]. The absorption of dietary iron, in the form of heme and non-heme iron, occurs in the duodenum, as Figure 1.2 illustrates.

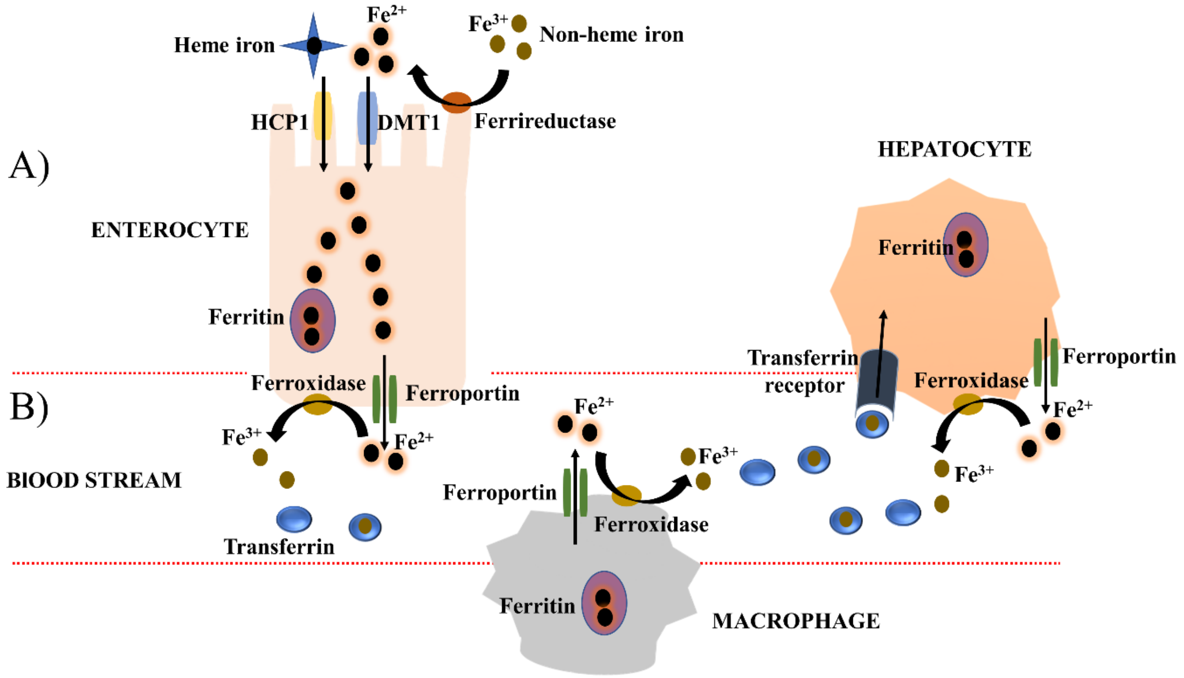


Figure 1.2: Cells and proteins involved in iron homeostasis. A) Most of the iron is absorbed in the duodenum. Non-heme iron mainly exists in Fe $^{3+}$ state. Before the uptake, the Fe^{3+} is first converted to Fe^{2+} by the inherent ferric reductase activity and transported into the cell by the divalent metal transporter 1 (DMT1). Heme enters the enterocyte presumably through the heme carrier protein (HCP). Iron is either further stored in cellular ferritin or released into the circulation by ferroportin. B) Ferroportin is the cellular iron exporter present on the surface of enterocytes, hepatocytes, and macrophage. Ferroxidase oxidizes Fe^{2+} to Fe^{3+} before loading into transferrin. Transferrin delivers iron through the circulation to the erythron of the bone marrow or hepatocytes. Transferrin receptor present on the surface of cells allows the entrance of transferrin-bound iron into cells. The sketch for the picture was taken from Steinbicker at al. [4].

Heme iron is mostly present in meat as part of hemoglobin or myoglobin, while non-heme iron consists of various forms of iron of both plant and animal origin [14]. Non-heme iron mainly exists in the Fe^{3+} state. Prior to cellular uptake, Fe^{3+} is converted to Fe^{2+} through inherent ferric reductase activity and transported inside enterocytes by a divalent metal transport protein (DMT1) [15, 16]. Heme absorption is fairly well understood. It has been suggested that heme is taken into the intestinal cell through the activity of heme carrier protein 1 (HCP1) [17]. In the cellular phase, heme is degraded by heme oxygenase, and iron is released. The Fe^{2+} inside the cell has two possible fates. It can either be stored as ferritin or transported into plasma through FPN [18]. However, the factors that decide whether iron is stored in the intestinal cell or transported into the plasma are not clear yet [19]. Iron is bound to Tf in the plasma and transported throughout the body via Tf. Transferrin-bound iron (Fe-Tf) is further taken up by transferrin receptors (TfR) present on the cell surface and further distributed for use and storage in a cell (Figure 1.2) [20]. The majority of body iron is passed to bone marrow, where most of it is further used to produce erythroid precursors and red blood cells [20, 21]. Excess iron, which is not immediately used in the body, is stored as ferritin in hepatocytes. These can accommodate both Fe-Tf and non-transferrin bound iron [22, 23]. Macrophages are responsible for the storage of recycled iron from senescent erythrocytes, also in the form of ferritin. In macrophages, old and damaged erythrocytes are lysed, and released hemoglobin is cleaved into heme and globin, and at the end, iron is liberated from heme by heme oxygenase [24]. Iron export from duodenal enterocytes, macrophages, and hepatocytes into plasma is achieved through the transmembrane iron exporter ferroportin (Figure 1.2) [25]. Recent studies indicate that ferroportin levels are post-translationally regulated by Hep-25, a circulating peptide hormone that binds to ferroportin, causing its internalization and degradation (see Section 1.3) [22].

1.2 The discovery of hepcidin

Hepcidin is a cationic peptide that contains 25 amino acids, and is today considered to be the central regulator of iron homeostasis [22]. The history of this small peptide begins almost 20 years ago. It was independently discovered in human plasma [26] and urine [27] by two groups. In both cases, the authors were focused on the discovery of novel cysteine-rich antimicrobial peptides. Isolated peptides presented antibacterial activity against *Escherichia coli, Staphylococcus aureus, Staphylococcus epidermidis*, and group *B Streptococcus,* and antifungal activity against *Candida albicans, Aspergillus fumigatus*, and *Aspergillus niger* [26, 27]. Krause et al. purified hepcidin from human blood and named it LEAP-1 (liver-expressed antimicrobial peptide) [26]. A 25-residue peptide containing four disulfide bonds was identified in human blood ultrafiltrate by a mass spectrometric assay for cysteine-rich peptides [26]. The primary structure of LEAP-1 was revealed by Edman degradation and had the following sequence: DTHFPICIFCCGCCHRSKCGMCCKT [26]. In addition, in order to investigate the expression pattern, the expression of the LEAP-1 gene was examined by real-time RT-PCR [26]. The expression was predominantly observed in the liver, and, to a much lower extent, in the heart [26]. Park et al. isolated the same cysteine-rich peptide hormone from human urine and named it hepcidin (hepatic bactericidal protein) [27]. The isolated peptide was characterized through amino acid sequencing and MALDI-TOF-MS, which revealed the mass, primary sequence, and presence of four internal disulfide bonds in hepcidin [27]. The results were identical to those reported by Krause et al. [26]. An N-terminally truncated variance of the 25-residue peptide, named hepcidin-20 (Hep-20), was identified in urine. Park et al. further obtained valuable insights into hepcidin expression [27]. The hepcidin gene encodes an 84-residue precursor protein known as a pre-prohepcidin [28]. The pre-prohepcidin sequence contained a putative 24-amino-acid peptide and a characteristic furin cleavage site, both present at the N-terminus. The first 24 amino acids of pre-prohepcidin contain an endoplasmic-reticulum-targeting signal that is cleaved by signal peptidase and produces a 60-amino-acid pro-hepcidin. Furthermore, pro-hepcidin is cleaved by furin to generate the biologically active 25-residue peptide (Hep-25) [27].

1.3 Bioactivity of hepcidin-25

Further research on the hepcidin gene led to the discovery of hepcidin's mRNA overexpression in the liver of iron-overloaded mice, but also in the liver of mice that died from the iron deficiency anemia [28]. However, Nicolas et al. also reported that hepcidin gene expression is reduced in the iron overload in the Usf2 mice model [29, 30]. These results suggest that hepcidin can act as a signaling molecule in iron regulation. The role of hepcidin in iron metabolism was later confirmed in humans. Inactivating mutations in the gene of hepcidin have been found in several patients with hemochromatosis [11, 31]. As a result, in the past few years, research has focused on better understanding hepcidin`s role in iron metabolism, the elucidation of the mechanism of Hep-25 activity as an iron regulator, and the factors implicated in Hep-25 regulation [32]. It was discovered that hepcidin acts by binding to ferroportin, inducing its internalization and degradation and thereby preventing the entry of iron into plasma [32, 11].

To investigate the functional importance of the N-terminal and C-terminal amino acids, as well as the disulfide-bonding pattern of Hep-25, researchers incorporated alterations into human Hep-25. Nemeth et al. measured the ability of Hep-25 to induce ferroportin endocytosis [33]. The significant effect on activity was only seen when five N-terminal amino acids were cleaved off, resulting in a complete loss of activity [33]. It was demonstrated that in addition to Hep-25, urine and plasma contain minor 20- and 22-amino-acid forms, truncated at the N-terminus [27, 34] (Figure 1.3). N-truncated peptides, representing metabolites of Hep-25, displayed a significantly reduced ability to cause ferroportin internalization [33–35]. Based on these previous experiments, it became evident that the N-terminal amino acids of Hep-25 play a crucial role in the interaction with ferroportin. However, molecular interactions between hepcidin and ferroportin have still not been fully explained.

Figure 1.3: Primary sequence of Hep-25 and N-terminally truncated metabolites of Hep-25: Hep-22 and Hep-20. N-terminally truncated peptides that are most likely metabolites of Hep-25 display significantly reduced ability to cause ferroportin internalization in comparison to Hep-25. Indicating that red-colored N-terminal amino acids are essential for the biological activity of Hep-25.

1.4 Regulation of hepcidin-25 expression

Normal iron homeostasis depends on the right balance between dietary iron absorption and body utilization. However, an imbalance leads to disorders of iron homeostasis, resulting in iron deficiency or overload, which are associated with various diseases, including the most common diseases in humans [36]. Diseases of iron deficiency are linked to anemia, which refers to a decrease in the levels of red blood cells or hemoglobin levels [36]. Iron overload diseases, such as hereditary hemochromatosis, are characterized by excessive dietary iron uptake, excessive transferrin-bound iron, and iron deposition in the liver, pancreas, plasma, and heart [36]. As an iron regulator, Hep-25 was found to be involved in the pathogenesis of many iron-related disorders in humans [11]. Dysregulation of hepcidin expression in the liver leads to abnormally low or high extracellular iron levels, as Figure 1.4 illustrates. On the one hand, a cytokine, produced due to inflammatory disorders and infections, stimulates excessive hepcidin production, which contributes to the development of anemia [32, 37]. Excessive hepcidin in the bloodstream induces ferroportin internalization and degradation. The loss of the iron exporter ferroportin from the cell surface then results in low iron levels in plasma and decreased erythropoiesis due to the decreased release of iron (Figure 1.4 A) [37]. On the other hand, low hepcidin production, due to mutations in the hepcidin gene or its putative regulators, contributes to the development of most types of hemochromatosis [37, 38]. Reduced hepcidin production causes increased iron import through ferroportin into plasma, high transferrin saturation, and iron overload in vital organs (Figure 1.4 B) [39].

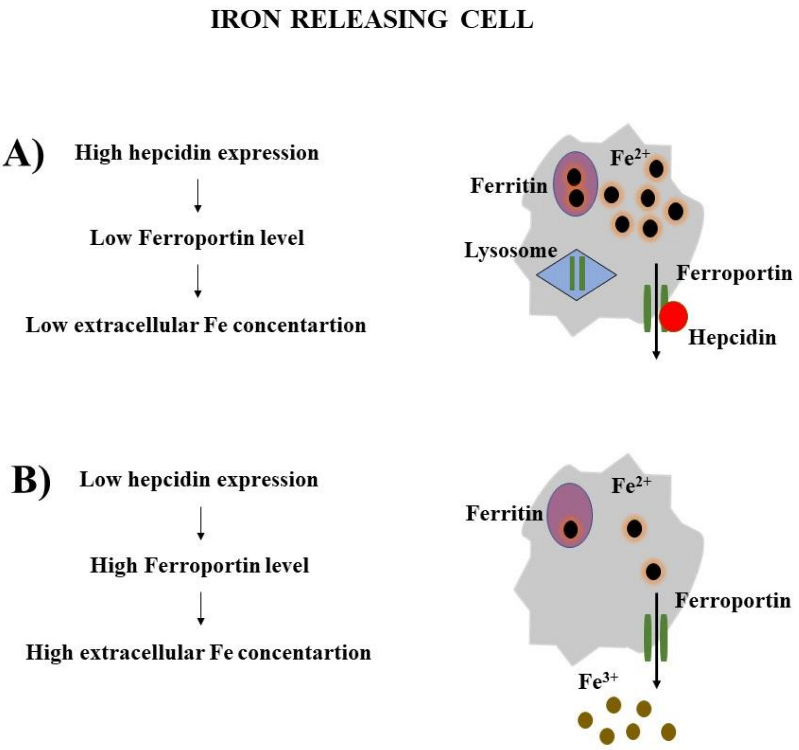


Figure 1.4: Hepcidin mediates regulation of iron homeostasis. Hepcidin controls the iron flow into plasma by inducing the endocytosis and proteolysis of the iron exporter ferroportin. Different stimuli affect hepcidin expression in the liver. A) High levels of hepcidin in the bloodstream result in the internalization and degradation of the iron exporter ferroportin. Loss of cell surface ferroportin results in macrophage iron loading, low plasma iron levels, and reduced level of red blood cells. B) Low hepcidin expression causes increased iron import into plasma, high transferrin saturation, and excess iron deposition in vital organs. The sketch for the picture was taken from De Domenico at al. [37].

Hepcidin and its regulators represent potential targets for the diagnosis and treatment of iron disorders [32]. It has been demonstrated that the mutation of human hemochromatosis protein (HFE), human hemochromatosis 2 protein (HEF2), TfR, or hepcidin antimicrobial peptide (HAMP) genes is responsible for dysregulated hepcidin expression [38, 40]. These genes encode HFE, hemochromatosis (HJV), TFR2, and Hep-25 expression, respectively. However, the signaling mechanism by which different stimuli affect hepcidin expression in the liver is not entirely clear yet. Various signaling pathways and factors regulate hepcidin production, including the bone morphogenetic protein (BMP) pathway, iron levels, and inflammation [11]. One mechanism includes the BMP pathway with its canonical signaling system, and it seems to be the critical signaling pathway for the regulation of Hep-25 expression. Bone morphogenetic proteins are cytokines that play essential roles in many processes, such as bone development or tissue repair and formation. The BMP pathway contains

HJV as a cofactor, which is the protein responsible for the iron overload condition known as juvenile hemochromatosis [38]. It was found that the BMP pathway leads to reduced Hep-25 expression and iron overload disease due to mutations in HFE, HEF2, TfR, and HAMP genes [37, 38]. A second regulatory mechanism senses iron levels through intracellular and extracellular iron sensors. For example, iron overload stimulates hepcidin production and excretion by the hepatocytes. Transferrin receptors and the membrane protein HFE seem to be involved in iron signaling through an incompletely understood pathway. The third type of regulation of Hep-25 production is transcriptionally regulated by inflammatory cytokines (IL-6) that increase hepcidin expression. Inflammatory mediators, predominately interleukin-6 IL-6, bound to IL-6 receptors (IL-6Rs) activate signal transducer transcription 3 (STAT-3). Activated STAT-3 binds to a regulatory element in the HAMP promoter and increases Hep-25 synthesis [37, 41]. It is unknown whether cross interference between these different pathways exists and whether it regulates hepcidin production.

1.5 Structure of hepcidin-25

A complete three-dimensional structural model could give further insight into Hep-25-ferroportin interaction and the iron-regulatory capacity. Hunter et al. previous reported on solution NMR structure of Hep-25 [42]. Investigation into synthetic human Hep-25 revealed the presence of a compact fold with β-sheet and β-hairpin loop elements, stabilized by four intramolecular disulfide bonds [42]. It has previously been confirmed that all eight Cys residues form intramolecular bonds; however, disulfide connectivity could not be determined with certainty (Section 1.2) [27]. Based on NMR data and structure calculations, two independent groups observed disulfide bond connectivity between Cys7-Cys23, Cys10-Cys22, and Cys11-Cys19, as well as a rare vicinal disulfide bond between Cys13-Cys14 [42, 43]. However, this disulfide bond connectivity was based on incomplete NMR data, as the resonances from two cysteines, Cys13 and Cys14, could not be detected. This was most likely due to exchange broadening [42]. The disulfide bond pattern between Cys7-Cys23, Cys10-Cys13, Cys11-Cys19, and Cys14-Cys22 that is currently accepted was suggested in 2009 by Jordan et al. (Figure 1.5) [44]. Partial reductive alkylation and NMR-based probability calculation were used for determination of disulfide mapping [44]. It is important to mention that the same disulfide connectivity was confirmed in Hep-25 by various sources, such as a human Hep-25 extracted and purified from urine, synthesized and folded Hep-25, and recombinant Hep-25 expressed in two different cell types, *Chinese hamster ovary* (CHO) and *Escherichia coli* [44]. It is evident that the complicated dynamic nature of the peptide prevents crystallization and degrades the quality of NMR constraints. Full sequential assignment, including amide protons of Cys13 and Cys14 residues that were broadened at room temperature, was achieved by temperature variation in order to overcome exchange broadening [44]. Differences in the chemical shifts between the low- and high-temperature forms of hepcidin also reveal that significant conformational changes occur in the loop region near Cys-13 and Cys-14 residues [44]. The dynamic loop region adopts two different, almost equally populated, conformations at ambient temperatures [44]. The structure of Hep-25 was calculated based on NMR data collected at two different temperature: 325K and 253K (Figure 1.5). At both temperatures, the structure of hepcidin consists of a stable β-sheet element and a β-hairpin loop with the currently accepted disulfide bond pattern (PDB code 2KEF) [44]. However, the N-terminal region of Hep-25, which was found to be necessary for binding with ferroportin, was found to be disordered under the conditions used for structural determination, particularly at acidic pH (pH=3).

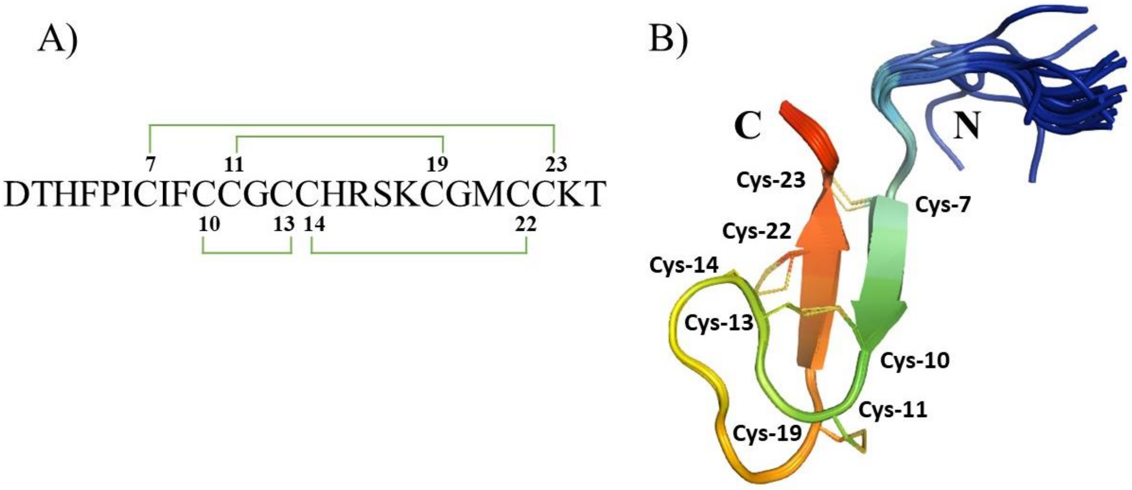


Figure 1.5: Solution NMR structure of Hep-25 at 325K (PDB code: 2KEF). A) The Hep-25 amino acid sequence with the currently accepted disulfide connectivity. Disulfide connectivity is indicated with green lines between following cysteines Cys7-Cys23, Cys10-Cys13, Cys11-Cys19, and Cys14-Cys22 [44]. B) The red part indicates the C-terminus, light yellow loop region, and blue N-terminus. Hep-25 contains β-sheet and β-hairpin loop elements, stabilized by four intramolecular disulfide bonds and the N-terminus which is not fully elucidated.

1.6 ATCUN motif

ATCUN motif (amino-terminal copper and nickel binding motif)was first identified 50 years ago in human serum albumin [45]. Later, the ATCUN motif was also identified in bovine serum albumin (BSA), neuromedins C and K, human sperm protamine P2a, and histatins [46]. The motif has been extensively characterized by UV-vis spectroscopy, NMR, electron paramagnetic resonance (EPR), and X-ray crystallography [45-49]. The ATCUN motif binding site is specific for the coordination of Cu^{2+} and Ni^{2+}. Four nitrogen atoms are involved in the coordination. They belong to the first three N-terminal residues that act as metal ligands. The complexation involves the free α-NH_2 nitrogen, two backbone amide nitrogen atoms of the second and third amino acid, and the imidazole δ-nitrogen atom of a histidine residue in the third position (Figure 1.6). The ATCUN motif can be defined as a structural feature with a preserved free NH_2-terminus and a histidine residue in the third position [45, 46].

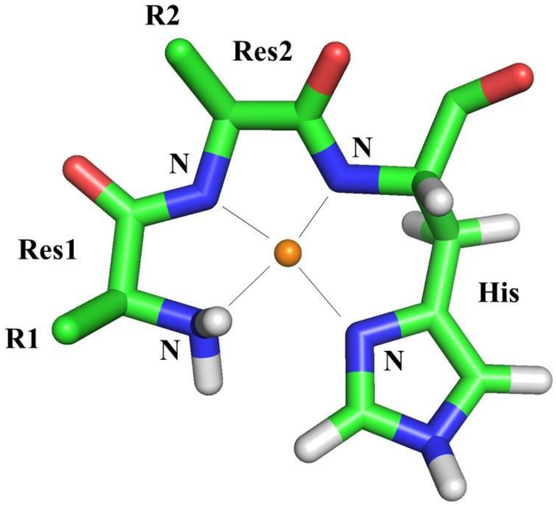


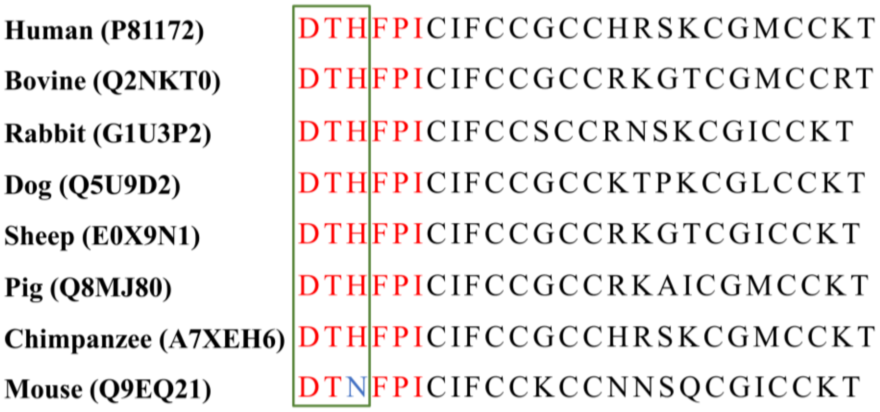
Figure 1.6: The amino-terminal copper and nickel binding (ATCUN) motif. The ATCUN motif is a structural motif which can be defined with three N-terminal residues, containing histidine always in the third position. Four nitrogens are involved in metal coordination, a free NH_2-terminus of the first residue, two amides from second and histidine residues, and nitrogen from the imidazole ring of histidine [45].

The crystal structure of the copper complex of the Gly-Gly-His tripeptide reveals a distorted square-planar geometry of the ATCUN-Cu(II) complex [46]. Laussac et al. studied the metal binding of human serum albumin by ^{1}H and ^{13}C NMR spectroscopy [47]. ^{13}C NMR data suggest that the carboxylate group of the aspartate residue in the Asp-Ala-His motif may interact with the metal ion, forming a penta-coordinated structure [47]. However, this theory was ruled out later by Hureau et al. by the X-ray structure of the [Cu(II)(DAHK)] complex, which revealed that the carboxyl oxygen from Asp1 is not involved in the metal coordination [48].

It is essential to mention that the ATCUN motif present in naturally occurring proteins was found to have a significant impact on their biological function [46]. For example, the albumin-Cu(II) complex is generally considered to be a major transport form of Cu^{2+} in blood. Apart from naturally occurring proteins, the ATCUN motif can be biotechnologically attached to proteins and equipped these with additional features, e.g.it can enable DNA, RNA, or protein cleavage, increase antimicrobial and antifungal activity, and it can be used for Cu-sensing [46, 49, 50].

1.7 Hepcidin-metal complex

In Section 1.2, it was mentioned that the breakdown products of Hep-25 (Hep-22 and Hep-20), which were found in urine and plasma, demonstrated the reduced ability to cause ferroportin internalization [33, 34]. Apart from N-terminal truncation, individual replacement of each of the six N-terminal residues with alanine reduces the bioactivity of Hep-25 [35]. All reported results reveal that six amino-terminal residues of Hep-25 play a fundamental role in its biological activity. Additionally, these N-terminal residues are highly *conserved* in Hep-25 from various species (Figure 1.7), which suggests a specific receptor interaction [51].



Human (P81172)	DTHFPICIFCCGCCHRSKCGMCCKT
Bovine (Q2NKT0)	DTHFPICIFCCGCCRKGTCGMCCRT
Rabbit (G1U3P2)	DTHFPICIFCCSCCRNSKCGICCKT
Dog (Q5U9D2)	DTHFPICIFCCGCCKTPKCGLCCKT
Sheep (E0X9N1)	DTHFPICIFCCGCCRKGTCGICCKT
Pig (Q8MJ80)	DTHFPICIFCCGCCRKAICGMCCKT
Chimpanzee (A7XEH6)	DTHFPICIFCCGCCHRSKCGMCCKT
Mouse (Q9EQ21)	DTNFPICIFCCKCCNNSQCGICCKT

Figure 1.7: A sequence alignment of hepcidin-25 from other species. Residues in red are highly conserved among the species, and residues in black are moderately conserved. The green square indicates the conserved ATCUN motif among all species apart from murine hepcidin-25. The difference is indicated with blue. The ATCUN motif from this species does not contain a His residue at the third position. The sketch for the picture was taken from Abbas at al. [51]. Hepcidin sequences were obtained from the UniProt (Universal Protein resource) databank.

Melino et al. first characterized the presence of an ATCUN motif, which has a high affinity for coordination of Cu^{2+} and Ni^{2+}, at the N-terminus of Hep-25 (see also Section 1.6) [52]. The characterization of the metal binding to Hep-25 and Hep-20 and the derivation of the peptides was performed by UV–Vis spectroscopy and circular dichroism (CD). Results clearly indicate that the first three N-terminal amino acids (DTH) are essential for the coordination of Cu^{2+} and Ni^{2+}. It was confirmed that the absence of histidine at the third position, as well as the absence of the ATCUN motif, resulted in a non-specific metal–peptide interaction [52]. Moreover, Konz et al. applied tandem mass spectrometry, which again confirmed the specific binding of Cu^{2+} to the N-terminal motif of Hep-25 [53]. The collision-induced dissociation (CID) experiments revealed the high stability of the first three N-terminal residues of Hep-25 coupled to one copper ion (DTH-Cu(II) complex) [53]. This binding motif was preserved during mass spectrometry (MS) fragmentation [53]. Since Hep-25 is a regulator of iron-homeostasis, it was not clear whether it binds iron or not. In 2010, Tselepis et al. tested the ability of human Hep-25 to form complexes with several transition metals (Cu^{2+}, Fe^{2+}, Fe^{3+}, Zn^{2+}, and Ni^{2+}) through Fourier-transform ion cyclotron resonance MS [54]. The results did not provide evidence for complex formation in the cases of Fe^{2+} and Fe^{3+}, while Cu^{2+}, Ni^{2+}, and Zn^{2+} all bound Hep-25. Cu^{2+} exhibited the highest affinity, and Zn^{2+} exhibited the lowest affinity [54]. Moreover, the authors determined the dissociation constant of the Hep-25-Cu(II) complex to be $<<10^{-6}$ M [54]. Recently, Plonka et al. employed potentiometric titration and (UV-Vis) spectroscopy to characterize copper complexation with the N-terminal hexapeptide of Hep-25 (DTHFPI) [55]. The hexapeptide DTHFPI demonstrated the strongest ATCUN-Cu(II) binding ever reported, with a dissociation constant of $10^{-14.66}$ M [55], even higher than the affinity of albumin for copper ($K_D = 10^{-12}$ M) [55, 56].

The ATCUN motif of human serum albumin (HSA) accommodates approximately 15% of the total plasmatic copper. The rest is bound to ceruloplasmin (65%), transcuprein (12%), and other serum components of low molecular weight [51]. McMillin et al. reported the ratio of bound and unbound fractions of copper. They determined this ration by separating the bound fraction through ultrafiltration, using a molecular weight cut-off of 30 kDa, which eliminates albumin, ceruloplasmin, and other proteins expected to bind copper [57]. The authors revealed that approximately 1–2 µM of unbound extracellular copper concentration is available for metal binding with small ligands such as Hep-25. Based on all results, the authors suggested that Hep-25 exists, at least sometimes, as a Cu(II) complex in blood [55, 57]. However, natural Hep-25 has not been found to contain metals due to the low pH needed for extraction and purification, which leads to a quantitative loss of any metals bound to the peptide. Although metal ions are lost under acidic pH, their presence should not be ignored at physiological or higher pH [51].

1.8 Specific Aims

Hepcidin plays a vital role in the control of iron homeostasis. Since the dysfunction of the hepcidin pathway leads to multiple diseases as a result of iron imbalance, hepcidin represents a potential target for the diagnosis and treatment of iron disorders. Despite knowledge of the 3D structure of Hep-25 [44], the ferroportin-hepcidin interaction has not yet been made clear, and selective immunoassay for iron disorder diagnosis and treatment has not yet been developed. Based on previous studies, it became evident that the N-terminal amino acids of Hep-25 play a crucial role in interaction with ferroportin [33–35]. Additionally, Melino et al. characterized the presence of an ATCUN motif, which has a high affinity for coordination of Cu^{2+} and Ni^{2+}, at the N-terminus of Hep-25 [52]. Therefore, it was assumed that the Hep-25-copper complex is perhaps the native, bioactive form of hepcidin. In order to investigate the possible role of metal-bound Hep-25 in iron metabolism, it is essential to elucidate in detail the 3D structure of metal-bound Hep-25.

Therefore, this research will be focused on the structural characterization of metal-bound Hep-25 through NMR spectroscopy and molecular dynamics simulations. NMR serves as an ideal tool for the determination of Hep-25's 3D structure, considering the Cu^{2+} and Ni^{2+} binding capacity of hepcidin's ATCUN motif. First, the metal binding of N-terminal ATCUN peptide fragment, comprising the six N-terminal amino acids (DTHFPI) from Hep-25, will be investigated through NMR spectroscopy. Then, the metal-hepcidin complex will investigated in detail through NMR spectroscopy. In order to track the changes induced by metal binding, the resonance assignment of Hep-25 will be the first step in metal-binding study. Based on the NMR data obtained, a structural calculation will be performed for the development of the 3D structure of Hep-25 in the metal-bound form.

2 Methods

2.1 Recombinant protein expression

2.1.1 Fusion proteins and tagged proteins

High protein purity and solubility present an essential requirement for structural and biophysical studies. Therefore, fusion proteins are often used as a useful tool for protein engineering that enables the optimization of protein functions. The resulting fusion protein can offer a combination of attributes that enhance protein expression, reduce proteolytic degradation, improve folding, solubility, and/or simplify purification and detection. For that reason, a fusion protein can be a crucial tool for recombinant protein production. A variety of structures have been used as fusion motifs, including affinity, solubility, and secretory tags [58, 59].

Affinity tags are polypeptides that are usually fused either onto the N- or C-terminus of a target protein. The most commonly used tag is the polyhistidine tag, consisting of six (His6) or more consecutive histidine residues. The small ubiquitin-like modifier (SUMO) tag is a small ubiquitin-related modifier that enhances expression and solubility and is fused to the N-terminus of a target protein. Additionally, the SUMO fusion protein can be cleaved with a specific, highly active ubiquitin-like protein specific (ULP1) protease [60]. The secretory signal tag allows the export of a protein into the periplasmic space with instant cleavage from a protein or peptide. The cleavage is possible only if the secretory tag is present on the N-terminus of the target protein. The most commonly used secretory tags are the OmpA and PelB secretion tags [61].

2.1.2 FX-cloning

The addition of different fusion tags can optimize the expression and purification of a target protein. One of the most efficient, inexpensive cloning methods suitable for screening of a broad variety of constructs is fragment exchange (FX) cloning [62]. FX-cloning allows the quick exchange of a suicide protein (ccdB) sequence, present on the pExpression plasmid (pQE-80L), with the gene of interest, present on the pDelivery plasmid (pUC-like derivative). At the end of the FX-cloning, the gene of interest will be incorporated into an expression vector which contains the desired tags present at N- or C-terminus of the protein, as Figure 2.1 illustrates.

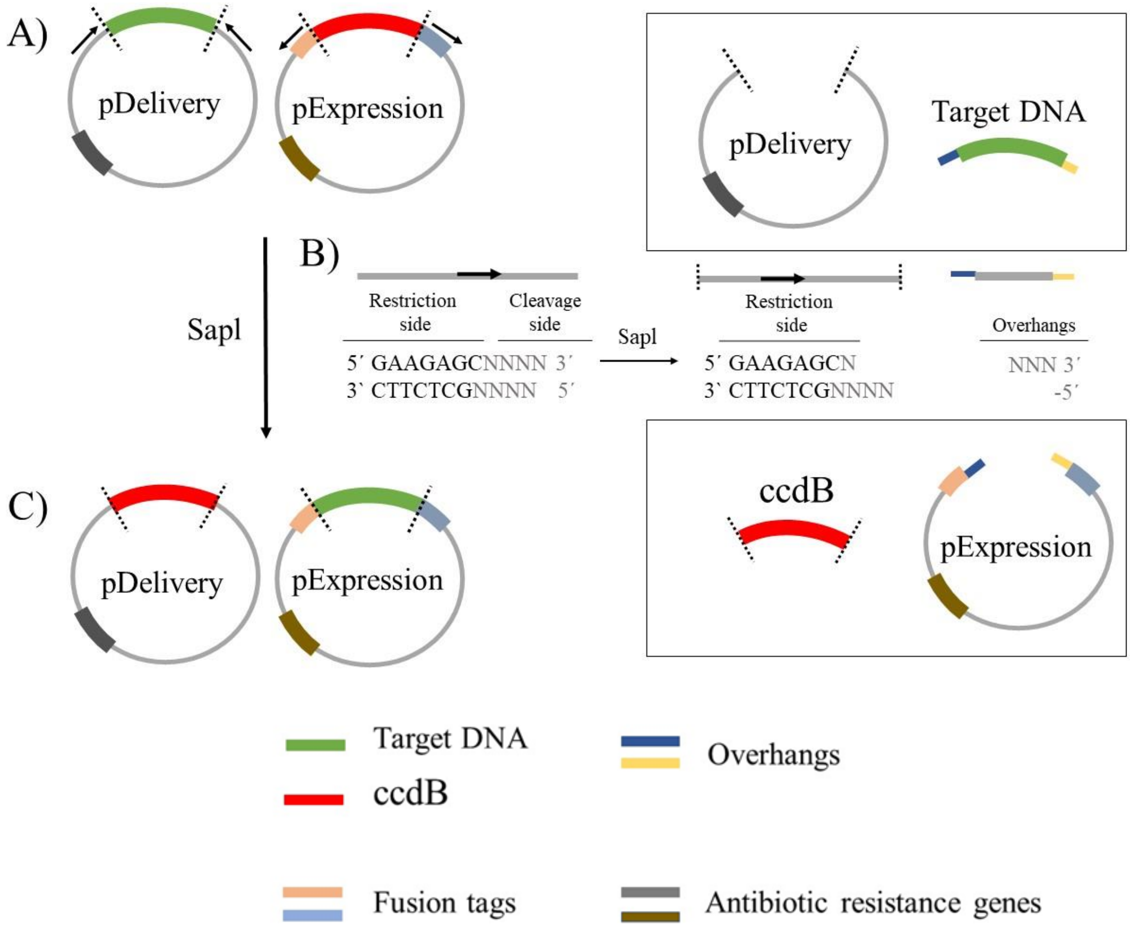


Figure 2.1: Schematic overview of the FX-cloning method [62]. A) Genes coding for the target DNA on pDelivery and a suicide protein (ccdB) on pExpression vector are colored in green and red, respectively. Black arrows indicate the direction of the restriction sites on plasmids. B) The SapI restriction enzyme recognizes restriction sites and cleaves double-stranded DNA, generating single-stranded overhangs on a target gene and pExpression vector. Overhangs generated upon cleavage are shown in blue and yellow and are present on pExpression vector and target protein (green). C) Complementary overhangs are ligated. After the exchange, a ccdB sequence (red) is now present on the pDelivery plasmid and the gene of interest (green) present on the pExpression plasmid.

The method is based on the use of the secreted aspartyl proteinase 1 (SapI) restriction endonuclease, which recognizes and binds specifically to the GAAGAGC sequence, known as a SapI recognition site. The SapI recognition site encompasses seven nucleotides is present on pDelivery and pExpression vector and located near to the ATG or GGG. As soon as the SapI restriction enzyme recognized its restriction site, it will cleave the double-strand of a circular DNA, producing three unpaired bases (ATG or GGG), known as "sticky" ends or overhangs. If

ATG and GGG, coding methionine and glycine respectively, are not the part of the target gene, it is mandatory to insert these into the target DNA sequence. These overhangs are incompatible with each other, which prevents self-ligation of the vector and insertion of the target in the undesired direction. Therefore, the direction of the SapI restriction site is significant in order to separate recognition sites from the gene of interest and pExpression vector. In the case of the of pDelivery vector, the SapI restriction site is oriented toward the gene of interest, so the respective recognition sites remain on the pDelivery vector and overhangs are present at the ends of the gene of interest. In the case of the pExpression vector, the direction of the cleavage sites is reversed, so that the respective recognition sites remain on the ccdB sequence and overhangs at the pExpression vector. After cleavage, overhangs present at the expression plasmid and the gene of interest are complementary and finally will be joined by ligation. The expression vector with an incorporated gene of interest is then transformed in *E. coli* and selected through growth on agar plate supplemented with specific antibiotics, to which the expression vector carries resistance. Since a ccdB protein causes cell death and pDelivery vector has an orthogonal antibiotic resistance to an expression vector, we have a high selectivity of colonies which contain the pExpression vector with an inserted gene of interest [62].

2.1.3 Polymerase Chain Reaction (PCR)

The generation of a plasmid suitable for FX-cloning and generation of mutated versions of the target gene can be achieved by the modification and amplification of the DNA sequence. Modifying DNA sequences has become feasible with polymerase chain reaction (PCR) [63, 64]. Site-directed mutagenesis is a PCR-based method which is widely used for substitutions, deletions, and insertions of nucleotides into plasmid DNA [63, 64]. The mutagenesis protocol comprises four crucial steps. PCR amplification of target plasmid in the presence of two mutagenic primers is the first step of mutagenesis. The second step is digestion of template DNA, which is parentally methylated DNA by DpnI. The last two steps include circularization of mutated PCR products by ligation with T4 DNA Ligase and transformation into *E. coli* [65, 66].

The key ingredients of PCR amplification are DNA polymerase, primers, template DNA, and nucleotides [67]. The first and essential step is the design of primers. The primers, with desired mutation(s), are designed so that they anneal back to back to the plasmid. Primers are generally 20–40 nucleotides in length and ideally have a GC content at the complementary part between 40–60% [67]. Additionally, primers should contain no more than three G or C base at the 3‘end of the primer, because GC-rich regions are much more prone to mispairing with other GC-rich regions than AT regions. Through specific forward and reversed primer design, we can substitute, delete, and insert the desired part of the target gene in the plasmid DNA. For example, primers for insertion of new DNA sequence into DNA plasmid contain 10-20 nucleotides at the 3' end that are complementary to a template DNA and the insertion piece of DNA at the 5' end (Figure 2.2). The insertion sequence will be split between the two primers and codon optimized for *E. coli* usage because the codon optimization increases a protein's expression yield. Half of the insert was added to the 5' end of the forward primer, and the other half was added to the 5' end of the reverse primer. Reversed and forward primers are designed so that 5' end of both primers is annealed back to back (Figure 2.2). Therefore, phosphorylation of primers, at the 5' end, is required later for the ligation of the PCR product into a vector [65, 66]. Before PCR amplification, it is also important to determine the melting temperature of the primers and heat stability of DNA polymerase. The melting temperature of the complement part of primers determines the annealing temperature of the PCR reaction. The heat-stability of a DNA polymerase plays a crucial role in PCR because high temperature is used repeatedly to denature the template DNA. For example, Q5 Hot Start High-Fidelity DNA polymerase is a high-fidelity, thermostable DNA polymerase able to withstand the denaturing condition required during the PCR. The PCR product generated using Q5 Hot Start High-Fidelity DNA polymerase have blunt ends, and ligation occurs between two blunt ends [67].

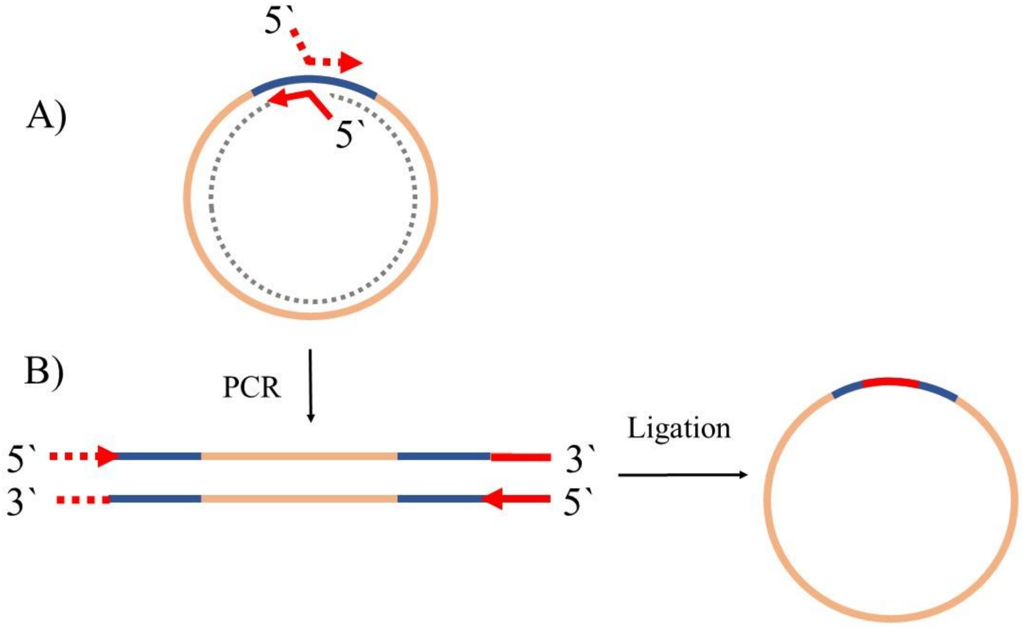


Figure 2.2: Forward and reversed primers required for the insertion of new DNA sequence into DNA plasmid. A) Primers, shown in red, contain 10-20 nucleotides at the 3' end that is complementary to a template DNA (blue) and the insertion piece of DNA at the 5' end. A DNA polymerase synthesizes a new DNA strand, during PCR, through the extension of primers, in the 5' to 3' direction by incorporating

nucleotides (dNTPs) that are complementary to the DNA template, grey point circle. B) After PCR, the elongated linear DNA strand is joined by ligation into a vector.

The DNA amplification is based on repeated cycles of heating and cooling that allow for DNA to be synthesized. Initialization heating at 98°C for 30 seconds is required for DNA polymerases heat activation and the initial denaturation. The denaturation step is the first cycling event and consists of heating the reaction chamber to 98°C for 5–10 seconds. This heating causes denaturation of the double-stranded DNA template or new DNA, made after the first PCR cycle, by breaking the hydrogen bonds between complementary bases. During the next, annealing step, the reaction temperature is lowered to 50–72°C for 10–30 seconds, so that reversed and forward primers can bind to their complementary sequences on the single-stranded template. The annealing temperature is determined by the melting temperature (T_m) of the selected primers. In the last step of the cycle, extension step, the reaction temperature is increased to 72°C and the DNA polymerase synthesizes a new DNA strand. The new DNA strand is synthesized through the extension of the primers, in the 5' to 3' direction by incorporating nucleotides (dNTPs) that are complementary to the DNA template. The precise time required for elongation depends on the length of the DNA target to amplify, and extension time is between 20 to 30 seconds per kb. The processes of denaturation, annealing, and elongation constitute a single cycle. Usually, 25–35 cycles are required to amplify the DNA target to millions of copies. Following the last PCR cycle, the final elongation is performed at a temperature of 72°C for 2 minutes to ensure that any remaining single-stranded DNA is fully elongated. The final step cools the reaction chamber to 4–10°C [67].

2.1.4. Protein expression

For the structural study, it is important to prepare a specific protein on a large scale, and this can be achieved using recombinant protein expression. Additionally, isotopic labeling is possible since for example *E. coli* cells can utilize $^{15}NH_4Cl$ as a sole source of nitrogen [68]. Recombinant production of proteins is based on the expression of the protein, of the gene that is encoded in an expression vector, in living organisms. In order to achieve a high protein yield, the selection of the host organism and expression plasmid is important. Therefore, the pET system was used as an expression vector for recombinant protein expression in *E. coli*. Target genes are cloned in pET plasmids, and their expression is under the control of the T7 promoter driven system from T7 bacteriophage, which induces a robust gene expression [69]. Generally, each plasmid contains an origin of replication (ori), the multiple cloning site, transcriptional and translational control sequences for protein expression and an antibiotic resistance gene that allows growth in the presence of a specific antibiotic. After plasmids are incubated with *E. coli*, those cells that take up the plasmid, through a process called transformation, can be selected by growing them in medium supplemented with specific antibiotics (e.g. ampicillin, tetracycline, kanamycin or chloramphenicol).

After plasmids are created, they are transformed into a host strain. Quite often, for heterologous expressions, the E. coli strain BL21 (DE3) is used as an expression organism in the pET expression system. The BL21 (DE3) strain has several advantages, such as deficiency in Lon and OmpT proteases. As a result, the heterologous proteins unknown to *E. coli* will not be digested during expression. In case the target protein contains a disulfide bridge, the modified BL21 (DE3) strain Origami B is used. Lack of thioredoxin reductase (trxB) and glutathione reductase (gor) genes, in Origami B host strains, leads to an oxidative environment within the cells which may lead to the formation of disulfide bridges in the cytoplasm. It is important to mention that DE3, of the BL21 (DE3) strain, indicates that the DE3 lysogen, of the λ phage, carries the gene for T7 RNA polymerase. The DE3 lysogen induces T7 polymerase expression under control of the *lac* operator. Since the synthesis of many heterologous proteins is toxic for the host bacterium, the *l*ac operator also regulates the recombinant protein expression. Therefore, a part of *the lac* operator integrated into the chromosomal genome of *E. coli* strain, one copy of the *lac* gene is inserted into the pET plasmid. That means that lac operon controls both the expression of T7 RNA polymerase and the gene of interest. T7 polymerase in conjunction with the T7 promoter and the lac operon present the currently most popular way of gene expression. Additionally, the T7 polymerase is very promoter-specific and has a very low error rate, which ensures high reproducibility in recombinant gene expression [69–71].

The *Lac* operon allows expression of the target protein only in the presence of the inducer isopropyl *β*-D-1-thiogalactopyranoside (IPTG). In the absence of the IPTG, the *lac* operator is inhibited by *lac* repressor protein and therefore prevents the transcription of genes in the *lac* operon. As soon as IPTG binds to the *lac* repressor and releases it from the *lac* operator in the *E. coli* chromosomal genome, T7 RNA polymerase will be transcribed and translated. At the same time, IPTG releases the repressor from the lac operator present at the pET plasmid and allows T7 RNA polymerase to bind the T7 promoter sequence present on the plasmid finally and triggers transcription of the target gene followed by protein biosynthesis [70]. The experimental procedure for protein production using the pET expression system is based on the growth of cells to a specific optical density at specific wavelength 600nm ($O.D._{600}$), between 0.6-0.8, and the gene expression of the target protein by IPTG induction.

2.2 Protein purification

In order to be able to elucidate the peptide/protein structure, it is not only enough to produce protein in high quantities, but it is also important to obtain the protein in a pure form. The protein purification strategy strongly depends on the production process used for protein or peptide production. Since the present work was focused on biotechnological protein expression in *E. coli*, the target protein after production was separated from the remaining cell components of the *E. coli* host organism, such as membranes, DNA, and other proteins. Affinity chromatography separates proteins from all other cell constituents based on an interaction between a protein and a specific ligand coupled to a chromatography matrix. Interaction between a ligand and its binding partner can be based on the interaction of lectin with polysaccharide or glycoprotein, glutathione with glutathione-S-transferase or GST fusion proteins and of metal ions with polyhistidine fusion proteins. For example, immobilized metal ion affinity chromatography is based on the chelating interaction between the histidine residues, polyhistidine tag present at a C- or N-terminus of the fusion protein, and immobilized nickel ions on a Sepharose matrix. The protein is released from the column by elution with imidazole, which competes with the polyhistidine tag for binding to the column. Increasing the imidazole concentration, the binding of imidazole to the Ni^{2+} chelate matrix will compete with the polyhistidine tag and release the polyhistidine tag of the target protein. is more competitive than the polyhistidine tag. Apart from affinity chromatography, a larger pool of proteins can be separated using various types of column chromatography, for example, size exclusion chromatography and ion exchange chromatography. Size exclusion chromatography separates molecules based on their hydrodynamic radius, a property proportional to their molecular weight. Ion-exchange chromatography separation is based on the electrostatic interaction of the ionic groups on the protein surface and the oppositely charged ionic group present on the surface of the stationary phase. Cation-exchange columns have a negatively charged surface, and conversely, anion-exchange columns have a positively charged stationary phase, these columns are used at pH values below or above the isoelectric point of the peptide, respectively. Sometimes, the purification scheme requires the use of both fast protein liquid chromatography (FPLC) and high-performance liquid chromatography (HPLC) to yield the necessary purity for downstream applications. High-performance liquid chromatography separates compounds based on their polarity.

Optimization of buffer conditions also plays an essential role in maintaining protein stability and function. During purification, many processes can occur that affect protein quality and yield: protein unfolding, aggregation, degradation, and loss of function. The stability and homogeneity of a protein sample are strongly influenced by the pH value, the ionic strength of the buffer, and the presence of additives. The pH value of the buffer should be chosen accordingly to the isoelectric point (pI) of the protein because it represents the pH where the solubility of protein is minimal. Therefore, the pH of the preparation buffer should be at least one pH unit away from the pI of the protein. Additionally, the amino acids composition dictates the presence of additives. The presence of cysteines could lead to aggregation, which could be prevented in the presence of reducing agents. The presence of hydrophobic regions in the protein sequence increase the interaction within a protein or with the neighboring protein rather than with the water molecules, thereby it could lead to insoluble protein aggregates. Chloride ions are slightly chaotropic and readily stabilize hydrophobic molecular surfaces in aqueous solutions by reducing hydrophobic interactions. Also, hydrophobic interactions can be reduced in the presence of denaturing agents or detergents. Since the production of ^{15}N and ^{13}C isotope-labeled protein can be very costly, optimizing of the purification conditions towards high protein yields, before isotope labeling expression and purification, is an important and often lengthy process.

2.3 NMR

Nuclear magnetic resonance spectroscopy provides information about physical and chemical properties of molecules like 3D structure, dynamics, and chemical environment.

One prerequisite for NMR is a nucleus having an angular momentum, expressed by the nuclear spin quantum number I (I= 1/2, 1, 3/2…). If spin quantum number I is not equal to zero ($I \neq 0$), the nuclear angular momentum gives rise to a magnetic moment (μ). The relationship between the nuclear angular momentum (P) and the magnetic moment, for each type of nucleus, is described through the gyromagnetic ratio (γ), a constant for any given nuclide [72, 73].

$$\mu = P * \gamma \tag{2.1}$$

For a spin of the I, there exist 2I+1 possible spin state (orientations of magnetic moment) with respect to an external magnetic field B_o. For example, for the proton (I= 1/2) there are two possible states denoted + ½ and –½, therefore, two possible orientations with respect to an external magnetic field B_o, either parallel or antiparallel. Transition between these states can be excited by photon having the proper energy ($h\nu$). A frequency equal to that energy of photon which would cause a transition between states is called Larmor frequency (ω_0).

$$\Delta E = h * \nu = \frac{h * \gamma * B_0}{2\pi} = \gamma * \hbar * B_0 = \hbar * \omega_0 \tag{2.2}$$

$$\omega_0 = \nu * 2\pi \tag{2.3}$$

After the radiofrequency (RF) pulse is irradiated, nuclear spins begin to precess with Larmor frequency and induce an oscillating voltage in the receiver coil, which decreases exponentially by spin-spin relaxation and is recorded as so-called free induction decay (FID). The Larmor frequency depends not only on the nuclear species but also on its electronic environment because the environment of the nuclei influences the effective magnetic field. Spins can also be influenced by the interactions with other spins, an effect called coupling that can be mediated either through chemical bonds (scalar coupling) or through space (dipolar coupling) [72, 73]. The frequency is usually not given as an absolute value but in comparison to a reference. The chemical shift δ is calculated as follows and is given in parts per million (ppm).

$$\delta = \frac{\nu_{sample} - \nu_{\mathrm{reference}}}{\nu_{\mathrm{reference}}} \tag{2.4}$$

By varying the pulse sequence and introducing evolution and mixing times in which magnetization is transferred from one nucleus to another, multidimensional spectra are generated whose cross-peaks provide information about the connectivity or confirmation of the molecule [72].

2.3.1 The resonance assignment of peptide

For a structural determination of the 3D structure of a peptide, a virtually complete assignment is mandatory. To this end, as many resonance assignments as possible of to ^{1}H, ^{13}C, and ^{15}N nuclei of the backbone as well as side chain will be compiled. Based on this, structural information will be gathered in form of intranuclear distances, dihedral angles, and further parameters like orientational constrains.

As the size of the molecule increases, one-dimensional spectra are often not sufficient for the resonance assignment, even smaller peptides cannot be assigned from 1D spectra. The large number of protons per residue and the resulting resonance overlap, require the use of two or sometimes also three-dimensional correlation spectra and isotopic labeling strategies. The principle of a two-dimensional experiment is transferred magnetization from one spin to another via the scalar coupling or dipolar coupling through space, between homonuclear (e.g.,

^{1}H–^{1}H) or heteronuclear (e.g., ^{1}H–^{13}C) spins. The low natural abundance of isotopes ^{13}C and ^{15}N in peptide makes the use of heteronuclear NMR spectroscopy difficult due to the reduced sensitivity. However, the sensitivity is enhanced with isotopic labeling. The choice of 2D pulse sequence strongly depends on which information is to be extracted from cross-peaks [72].

The sequential assignment consists of two stages, the first stage involves the identification of certain amino acid and the second stage involves the assignment of each amino acid to a particular residue in the amino acid sequence.

In the first stage of sequential assignment we can determine to which type of amino acids the spin system belongs. A spin system is a group of spins that are connected by a network of scalar spin-spin couplings, and it can be determined using total correlation spectroscopy (TOCSY) experiments [73]. In the ^{1}H-^{1}H TOCSY experiment, the magnetization is transferred via the scalar coupling through residues within the same spin system, but which may not share mutual couplings. Each amino acid has a characteristic pattern of chemical shift that is usually well resolved in the amide region of TOCSY spectra (Figure 2.3). Proline has no amide proton but displays signals in the aliphatic region of TOCSY spectra.

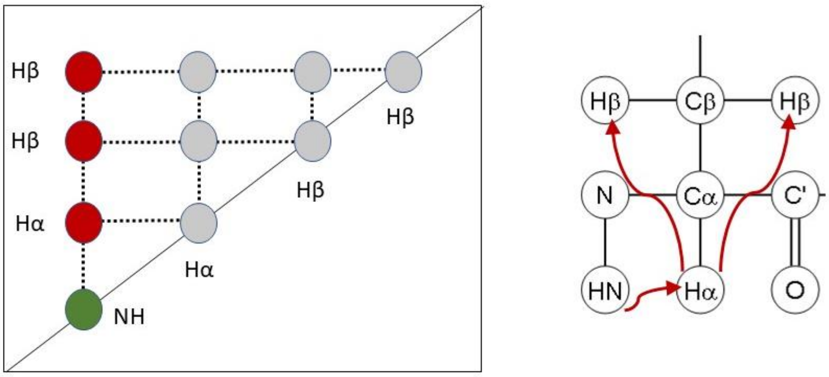
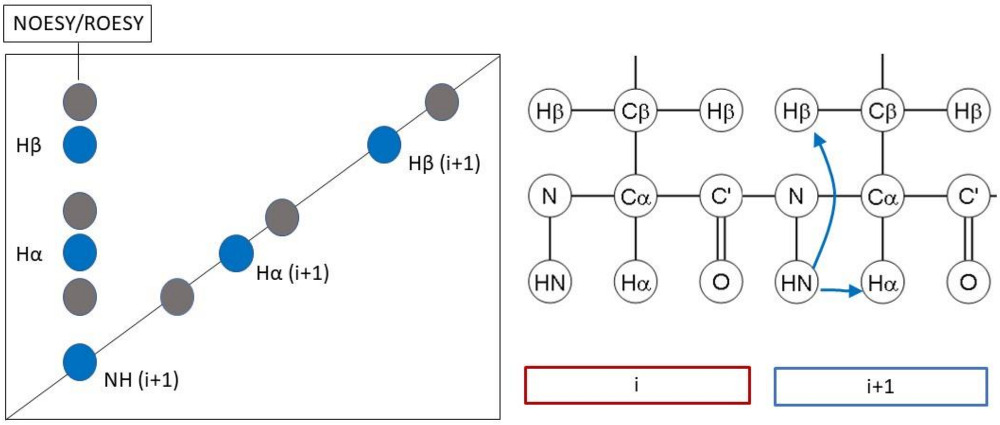


Figure 2.3: Amide region of the TOCSY spectrum. TOCSY experiments provide correlations between all nuclei within the same spin system, even though they do not all couple with each other. In a chain of protons, HN-Hα-Hβ, magnetization is transferred from HN to Hα and then further to Hβ. Therefore, we can determine to which type of amino acids the spin system belongs.

In the second stage, as we determined the spin systems, sequential resonance assignment is applied in order to connect spin systems in their sequential order. Since scalar couplings between protons of different amino acids are not observed, NOEs constitute a powerful tool to establish sequential connectivity, especially in non-labeled peptides. Cross peaks in the ^{1}H-^{1}H NOESY (Nuclear Overhauser Effect Spectroscopy) [74] or ROESY (Rotating-frame Overhauser Spectroscopy) [74] originate from the relaxation processes triggered by dipolar couplings, interactions of spins via space and hence only depend on the intranuclear distance. The NOESY or ROESY experiment correlates protons that are within a distance of 5-6 Å. The spectra will contain cross peaks from both intra-residue correlations (Figure 2.4 A) as well as inter-residue correlations (Figure 2.4 B), which allows us to find which residues are in the neighborhood with each other in the peptide chain. Therefore, NOEs may be found between all protons close in space and could lead to some ambiguity. For that reason, sequential correlations usually depend on the secondary structure in the peptide. In the case of α helical regions, we can observe NOEs between backbone amide protons H^N, because they are close in space. However, in the case of β-sheets, we observe NOE cross peaks between H^N and Hα from neighboring residues (NH_{i+1} - αH_i).

A) Intra-residue correlations



B) Inter-residue correlations

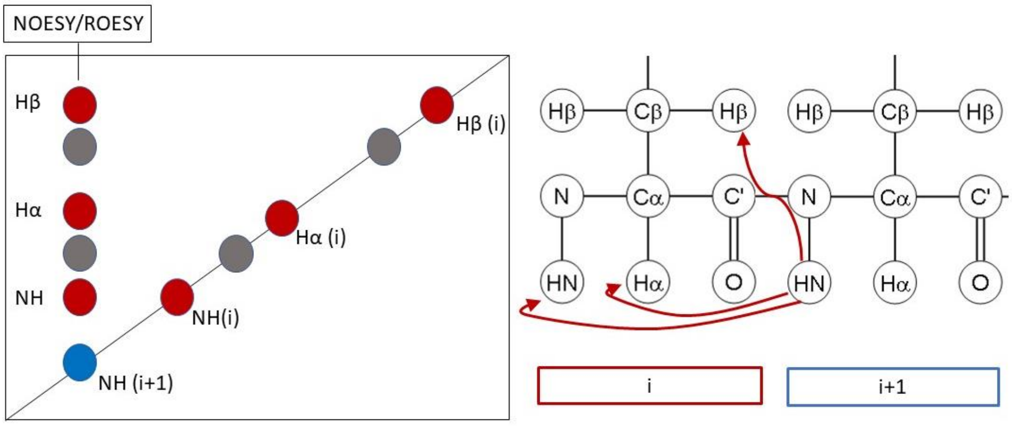


Figure 2.4: NOESY diagonal and cross peaks in the NH (F2)/ NH (F1) region. NOESY experiment allow us to find which residues are in the neighborhood with each other in the peptide chain through the interactions of spins via space. The spectra will contain cross peaks from both A) intra-residue correlations, blue signals, as well as B) inter-residue correlations, red signals. For example, the connectivity of a given amino acid in the sequence (i) to its following one (i+1) can be observed in the

NOESY because the distance of the amide proton of (i+1) to the HN, Hα or Hβ protons of (i) is smaller than 5 A in almost every case (B). Therefore, sequential cross signals to Hα(i), Hβ(i) etc. are observed at the frequency of HN(i+1) (B), red signals.

Advances in protein expression methods and the possibility for uniform labeling of recombinant proteins with ^{15}N offered us a chance for ^{15}N assignment with HSQC (Heteronuclear Single-Quantum Coherence) experiment. The HSQC experiment shows a correlation between protons and directly bound heteroatoms. Therefore, each amino acid, except proline and N-terminal amine, provides a peak in the HSQC spectra from backbone ^{1}H and ^{15}N correlation. So, we could assign a part of backbone amides and amides from the side chains of Arg, Trp, Asn, and Gln [74].

2.3.2 Chemical shift perturbation (CSP)

As it was mentioned in section 2.3., the larmor frequency of a nuclear spin depends on its chemical environment, for example, ligand binding induces chemical shift changes in the protein leading to a chemical shift perturbation (CSP). Therefore, chemical shift perturbation is a common technique for the detection and characterization of protein-ligand interactions or structural changes of proteins. This indicates that we could identify the ligand binding site of proteins and measure ligand affinity. In order to study the interaction of a protein with a ligand, ^{1}H-^{1}H TOCSY or ^{1}H-^{15}N-HSQC spectra are often used, since they show a large dispersion of the signals and, if resonance assignment is available, directly allow to locate the binding pocket. For this purpose, the unlabeled or ^{15}N-labeled protein is mixed with increasing concentrations of the unlabeled ligand.

As we add ligand to the protein, this induces chemical environment changes in the protein, whereby, depending on the exchange kinetics and the difference of the chemical shifts in the free and bound state, two different effects can be observed. Under the condition of a fast exchange between bound and free state, one observes a gradual shift of the signals with increasing ligand concentration [75]. In slow exchange, the original signal of the free form gradually decreases, and the signal of the bound complex appears to the same extent at another position in spectra [75].

2.3.3 Structure determination

Protein structure can be elucidated with NMR methods. To this end, structural information is required in addition to the resonance assignment. Once chemical shift assignments have been completed then it is necessary to obtain structural constraints such as secondary shifts [76], torsion angles [77], or distance information [78] between protons within a protein, which can be obtained from various NMR experiments, as described in the following sections.

In the first stages of NMR structure determination, we commonly assign the chemical shifts of the Cα, Cβ, CO, N, NH and Hα atoms. These chemical shifts are strongly correlated with local structure of protein and, therefore, it possible to derive secondary structure elements and dihedral φ and ψ angles from chemical shifts.

The secondary chemical shift of Cα, CO, Cβ Hα, HN and NH atoms are strongly linked to the secondary structure. The secondary chemical shift can be obtained by subtracting the observed chemical shift from the chemical shift expected in a random-coil peptide [76]. Once the secondary chemical shifts have been defined, we look for patterns of up- or downfield shifted atoms along the protein sequence in order to identify stretches of α-helix or β-strand [76]. For example, Cα atoms in α-helices will tend to have positive secondary chemical shifts and in β-strands will tend to have negative secondary chemical shifts. However, low resolution in information on secondary structure is accessible by analyzing the chemical shifts.

As the Ramachandran plot shows that the secondary structure of a protein is the result of following residues having similar φ and ψ torsion angles, with one area of the plot being linked to α-helices and another to β-strands. Therefore, it is additionally possible to derive not only the secondary structure from chemical shifts, but also the approximate dihedral angles [79]. This strategy has been implemented in a software package called TALOS+, which predict dihedral angle restraints for protein structure determination from chemical shifts [79]. TALOS+ is a database system for empirical prediction which is based on the comparison of Hα, Cα, Cβ, CO and N chemical shift assignments with database values collected for a suite of proteins of known structure.

Second step in structural determination involves determination of dihedral angles driven from scalar couplings because dihedral angles and $^3J_{HN\text{-}\alpha H}$ coupling constants of vicinal protons are related via the Karplus equation [77]. For structural calculation, in proteins, two angles phi (φ) and psi (ψ) in the backbone are evaluated, because they contain the most significant information content about the secondary structural elements, and thus the total folding. The phi angle (φ) describes the angle between CO, N, and Cα and psi angle (ψ) describe the angle between the surface spanned by N, Cα, and CO. The phi angle (φ) is determined from the angle HN-N-Cα-Hα by determining the coupling constant between HN and Hα. The coupling constant assumes different values depending on the dihedral angle according to the Karplus equation, as it was mentioned already. In order to avoid the ambiguous dihedral angle constrains, it is mandatory to acquire high-resolution *J*-resolved information.

However, by far the most important and widely used structural information is a distance relationship that can be measured due to the dipolar interaction of two spatially adjacent nuclei and the resulting nuclear Overhauser effect (NOE) [78]. The Nuclear Overhauser Effect (NOE) is the change in intensity of one resonance when the spin transitions of a dipolar coupled nucleus are somehow perturbed from their equilibrium populations. This perturbation can be achieved by selective inversion of the spin, by applying a selective 180° pulse to its resonance as in the transient NOE experiment. For example, if the spin S is perturbed from the equilibrium, equilibrium recovery will take place by the cross-relaxation, where the S magnetization is transferred to the I spin. Thus, the magnitude of the NOE observed for spin I is expressed as the percentage of relative intensity change between the equilibrium intensity (I_0) and that in the presence of the NOE (I). The correlation between two protons, based on NOE, depends on the distance between them and a signal is observed within a distance up to approximately 5-6 Å. The cross-peak intensities from 2D NOESY spectra can be quantified, since the intensity is inversely proportional to the sixth power of the distance of the nuclei, and thus indicates a distance value. Most commonly, unknown internuclear distances are estimated by comparison with NOESY cross peaks of nuclei having a known distance, according to the equation 2.5, where the intensity of cross peak (V_{IS}) is compared with the intensity of reference nuclei ($V_{reference}$) with known distance ($r_{reference}^{-6}$). Therefore, the intensity of cross peaks has to be calibrated [74].

$$r_{IS} = \sqrt[6]{\frac{V_{reference}}{V_{IS} * r_{reference}^{-6}}} \tag{2.5}$$

These peak intensities, however, cannot be used to determine exact distances since contributions can also arise from spin diffusion. Therefore, multiple NOESY spectra with different mixing times have to be acquired in order to identify and correct indirect NOEs, which arise from spin diffusion.

In the optimal case, distances in combination with dihedral angles are used in molecular dynamics-based structure calculation with programs such as Xplor-NIH [80].

2.4 NMR molecular structure determination using Xplor-NIH

Xplor-NIH is a software package for structure determination which combins various experimental data sources with known geometric data. Structure determination is achieved by seeking the minimum of a target energy function using a variety of optimization procedures such as molecular dynamics in cartesian and torsion angle space, Monte Carlo methods, and conventional gradient-based minimization. In the most common use of Xplor-NIH, an energy metric is minimized by a combination of simulated annealing using molecular dynamics in cartesian space, and gradient minimization [80]. The structure calculation based on Xplor-NIH program usually consists of five steps. The first step involves creating the protein structure file (PSF). The second and third step, calculating a first extended structure and calculating a family of structures using *ab initio* simulating annealing. The last two steps involve refinement of the family of structures and selection of low energy structures

In the initial step of Xplor-NIH, the PSF file can be loaded from a protein data bank (PDB) file or generated from the sequence. The resultant PSF file describes how atoms in the system are connected by bonds, angles, dihedrals and improper bond angles along with other properties such as partial charges and atom masses. Although the protein's covalent topology and associated parameters are established at this point, information on the location of each atom is still undefined. In order to minimize the energy function, a starting structure is necessary. Therefore, in the second step, an extended starting structure is generated. In the following step, *ab initio* simulated annealing creates a family of structures starting from the extended structure and experimental NMR restraints [81]. There are four main types of NMR restraint: distances, dihedral angles, hydrogen bonds, and orientational restraints, such as residual dipolar couplings (RDCs). Simulated annealing strategies then serve to solve the complicated optimization problem of combining all this information into a three-dimensional structure and find the configuration of atoms that minimizes the hybrid energy function. The hybrid energy function is total energy formulated as a linear combination of the geometric force field and experimental derived restraints [80]. The geometric force field is the sum of bonding (E_{bonds}, E_{angles} and $E_{improper}$) and non-bonding (E_{vdw}) interactions [80]. The experimental term is the sum of restraints from solution NMR experiments such as NOEs, hydrogen bonds and torsion angles [80]. Hence, the configuration of atoms that satisfies both experimental restraints and standard bond length and angles, present in the geometric force field, is achieved by the minimisation of the hybrid energy function. Upon calculation, the coordinate files contain useful information about energies, restraints, and violations. The next step includes refinement of the family of structures generated by simulated annealing. The refinement of the structures is repeated until there are no improvements in energy, violations, and structural features such as the Ramachandran plot, root-mean-square deviation (RMSD) and covalent geometry, which speak in favour of a higher-quality structure. The final step includes a selection of the final bundle of low energy structures and structures with defined NOE distances and dihedral angle restraint violations [80].

3 Results and Discussion

3.1 The hexapeptide DTHFPI

In order to investigate the 3D structure of the hepcidin-metal complex, the first step in this study was to confirm the metal binding of Hep-25 through NMR spectroscopy. Complexation involves free α-NH_2 nitrogen, two backbone nitrogen atoms of the second and third amino acid, and the imidazole nitrogen of a histidine residue in the third position. As the N-terminus of Hep-25 is composed of aspartic acid, threonine, and histidine in the third position, it became clear that it can be defined as an ATCUN motif. Although Melina et al. first identified the presence of an ATCUN motif at the N-terminus of Hep-25 [52], which binds $Cu^{2+}$and Ni^{2+} with high affinity, the structure of the hepcidin-metal complex was not detailed. All previous publications reported that the N-terminus of Hep-25 was found to be disordered in contrast to the rest of the molecule, which is highly rigid due to four intramolecular disulfide bridges [44]. In addition, the sequential alignment of Hep-25 sequences of various species revealed a highly conserved N-terminus [51], not only the ATCUN motif (Asp-Thr-His) but even the 6-residue peptide DTHFPI. In our initial study, we investigated the Cu^{2+} and Ni^{2+} binding of this peptide fragment, comprising the six N-terminal amino acids from Hep-25. For this purpose, a chemically synthesized hexapeptide was purchased from Peptides & Elephants.

3.1.1 Resonance assignment of hexapeptide

The first step of the metal biding study involved the complete resonance assignment, in order to easily track changes in the chemical shifts of a hexapeptide when a metal is added. Due to the size of hexapeptide, the overlap of signals was reduced. Therefore, sequence-specific assignments were achieved through an analysis of 1D-^{1}H, ^{1}H-^{13}C HSQC, 2D TOCSY, and 2D ROESY spectra. The NMR measurements were performed at room temperature, and a virtually complete assignment was obtained in an acidic solution (Figure 3.1). A peptide concentration of 1 mM in a 25 mM phosphate buffer, pH = 3.2, was used for the measurements. First, individual amino acid residues identifications were assigned from 2D TOCSY spectra, and side-chain resonances were assigned using a combination of 2D TOCSY and 2D ROESY spectra. Chains of neighbors could be defined and assigned to specific amino acids within the sequence through an analysis of 2D-ROESY spectra. An analysis of hexapeptide's NMR spectra revealed the presence of two sets of resonances. This effect was attributed to *cis/trans* proline isomerization, as the proline is present in the hexapeptide (DTHFPI) sequence at position five. The slow *cis-trans* proline isomerization, concerning the NMR chemical shift scale, induced the double set of signals. Two resonance frequencies were observed for almost every nuclear spin affected by the isomerization process. Proline *cis/trans* isomerization likely occurs in many folded proteins because both the *cis* and *trans* peptide bond conformations are thermodynamically accessible. Therefore, it was essential to distinguish the *cis* and *trans* isomers from each other in the hexapeptide. This was achieved by a comparison of the ^{13}C chemical shifts of proline, assigned in ^{1}H-^{13}C HSQC spectra, to those published previously by Lee et al. [82]. Additionally, NOESY cross peaks between Hα of phenylalanine (Phe-4) and Hα or Hδ resonances of Pro-5 confirmed the presence of *cis* or *trans* isomers, respectively. An analysis of NMR data yielded that the ratio of *trans/cis* isomers is about 70%/30%. Two sets of resonances were observed in NMR spectra at 298K as well as at 325K. Neither the temperature nor the pH value affects the *cis/trans* isomer ratio to a significant extent. A double set of resonances was assigned to specific amino acid in its the *cis* or *trans* form.

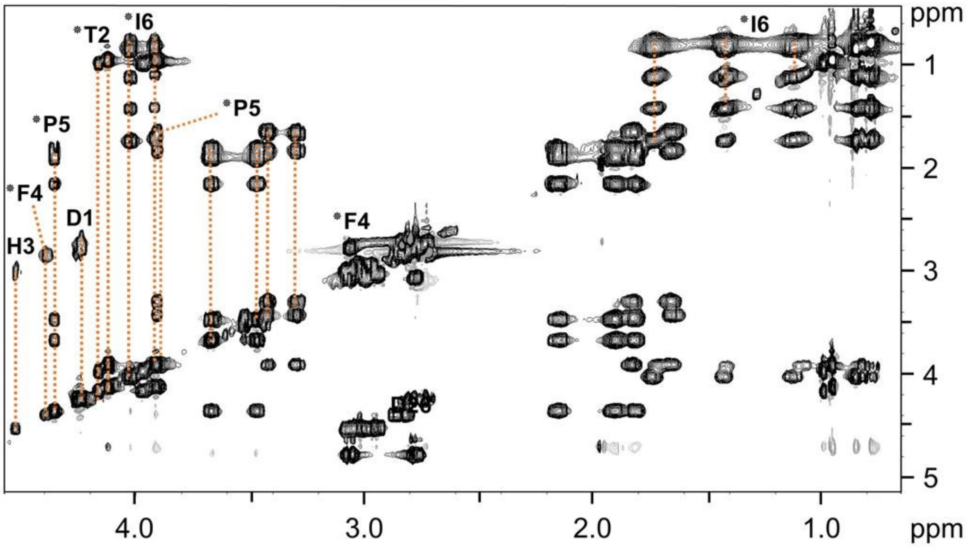


Figure 3.1: ^{1}H-^{1}H TOCSY spectra of hexapeptide (DTHFPI). In order to obtain a complete assignment, spectra of 1 mM sample of hexapeptide were recorded in 25 mM sodium phosphate, pH 3.2, 10% D_2O, T=298K. The spectrum shows the TOCSY aliphatic pattern assigned for each amino acid of the hexapeptide (DTHFPI). Resonance assignments are indicated with one-letter amino acid codes and residue numbers. The symbols in spectra encode followings: * The spectrum contains a double set of peaks for each amino acid because of *cis/trans* proline isomerization [44].

3.1.2 Metal coordination of the N-terminal hexapeptide

Following the resonance assignment of the hexapeptide, metal ions were titrated into solutions of the peptide until 1:1 molar ratio was reached. Coordination of a metal ion to the peptide affects its properties, often through a change in its 3D structure. In order to track changes, 1D-^{1}H- and 2D-TOCSY spectra were acquired after each titration step, thereby allowing us to trace complex formation-induced spectral changes. As was reported previously, the peptide reflects an affinity for Cu^{2+} as well as Ni^{2+}. In the case of Cu^{2+}, a paramagnetic complex is formed with the ATCUN motif, which caused drastic line broadening of residues near the coordination site. However, under sub-stoichiometric conditions, we observed that the exchange between complexed and uncomplexed peptide was slow on the NMR time scale. For that reason, we were able to record the reduced signal intensity of the uncomplexed peptide even in the presence of the paramagnetic Cu^{2+} complex. Two sets of signals occur in spectra at sub-stoichiometric hexapeptide to Cu^{2+}ratios. One set of signals corresponded to the metal-free peptide, displaying sharp NMR resonances, and another set of signals demonstrated broadened lines of amino acid residues remote from the coordination site (Figure 3.2). However, signals of the residues directly involved in metal binding or close to the coordination site completely vanished (Figure 3.2). TOCSY spectra taken at an equimolar concentration of Cu^{2+} and hexapeptide clearly demonstrated that resonance assignment could be only obtained for Ile-6, the residue located furthest away from the metal binding site. Additionally, sidechain protons Hβ and Hγ of proline (Pro-5) retained detectable intensity. Except for Ile-6 and Pro-5, all other residues were completely broadened out (Figure 3.3).

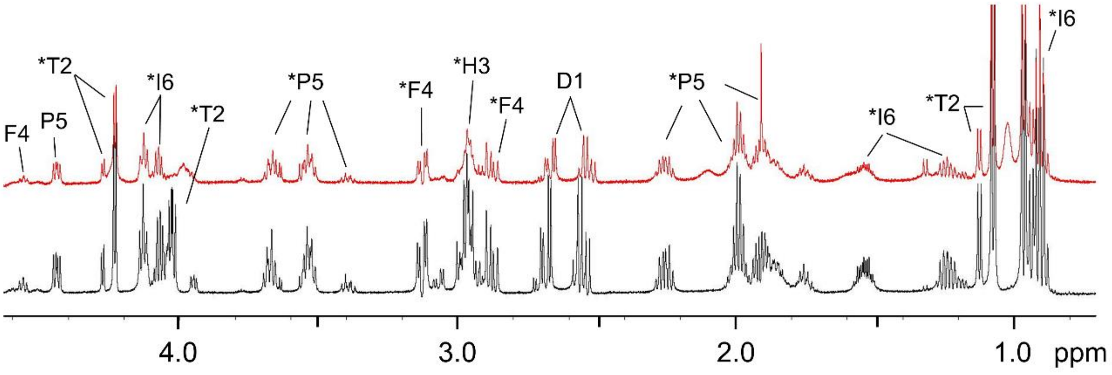


Figure 3.2: ^{1}H NMR spectra of the metal-hexapeptide complex in the presence of 0.5 equivalent of Cu^{2+} (red) and uncomplexed peptide (black). The red spectrum proves the slow Cu^{2+}exchange between complexed and uncomplexed peptide. Therefore, we were able to see the reduced signal intensity of an uncomplexed peptide and broadened signals from residue which are far away from the coordination site

even in the presence of paramagnetic Cu^{2+} complex. * spectra contain a double set of peaks for each amino acid because of cis/trans proline isomerization [44].

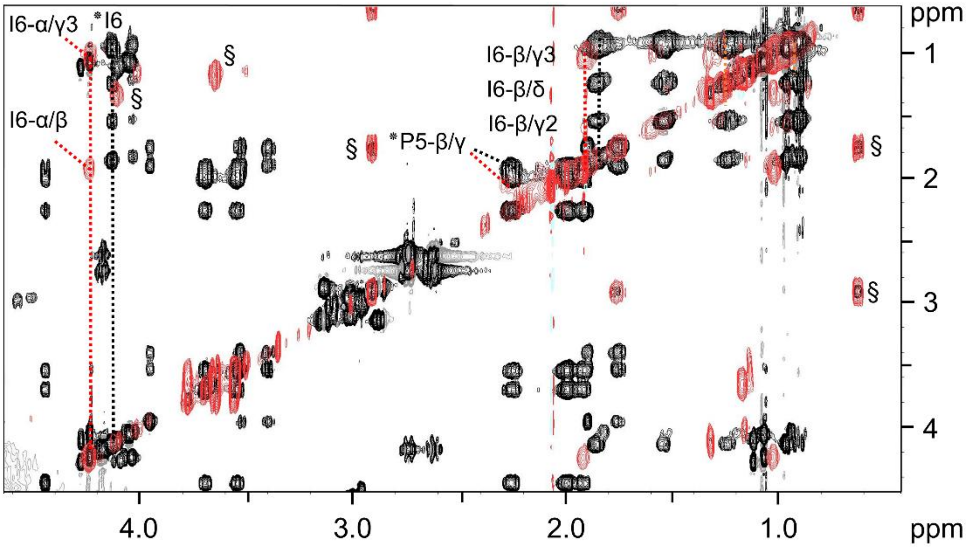


Figure 3.3: TOCSY spectra show the DTHFPI hexapeptide in the presence of Cu^{2+} (1:1) complex (red) and the absence of Cu^{2+} (black). Resonance assignment of the Cu^{2+}-hexapeptide complex could be only obtained for Pro-5 and Ile-6 amino acids, which were not so strongly affected by the paramagnetic effect of Cu^{2+}-hexapeptide complex. Hα and sidechain protons of Ile-6 and sidechain protons (Hβ and Hγ) of Pro-5 were visible in the TOCSY spectra. The symbols in spectra encode followings: § indicates the presence of the impurities, * spectra contains a double set of peaks for each amino acid because of *cis/trans* proline isomerization [44].

All information necessary for a high-resolution structure determination of Cu^{2+}-bound N-terminal hexapeptide was not observable because of paramagnetic line broadening. For further studies, we used Ni^{2+} as a diamagnetic probe instead of Cu^{2+} to better understand metal binding to the hexapeptide and obtain structural information. This process is further discussed in Section 3.8. It is known from previous literature that ATCUN motifs bind both metal ions and adopt square planar geometries [83, 84]. As Figure 3.4 illustrates, Ni^{2+} forms a stable complex with the DTHFPI hexapeptide. As with Cu^{2+}, the Ni^{2+} complex also demonstrates a slow exchange between complexed and uncomplexed peptide on the NMR time scale. Nickel coordination induces chemical shift perturbations because of changes in the electronic structure and stabilization of a specific conformation. The most strongly affected amino acids are those of the ATCUN motif, specifically aspartic acid (Asp-1), threonine (Thr-2), and histidine (His-3) (Figure 3.4) [51]. It is important to notice that upon complexation, two amino protons of Asp-1 became visible and demonstrated correlations to Asp-1's α– and β-protons in the TOCSY spectrum. This correlation was most likely visible because of the transfer of magnetization from nickel to two amino protons of Asp-1. Chemical shifts due to metal complexation decreased further away from the metal binding site, with the smallest effect observed at phenylalanine (Phe-4), proline (Pro-5), and isoleucine (Ile-6) [51].

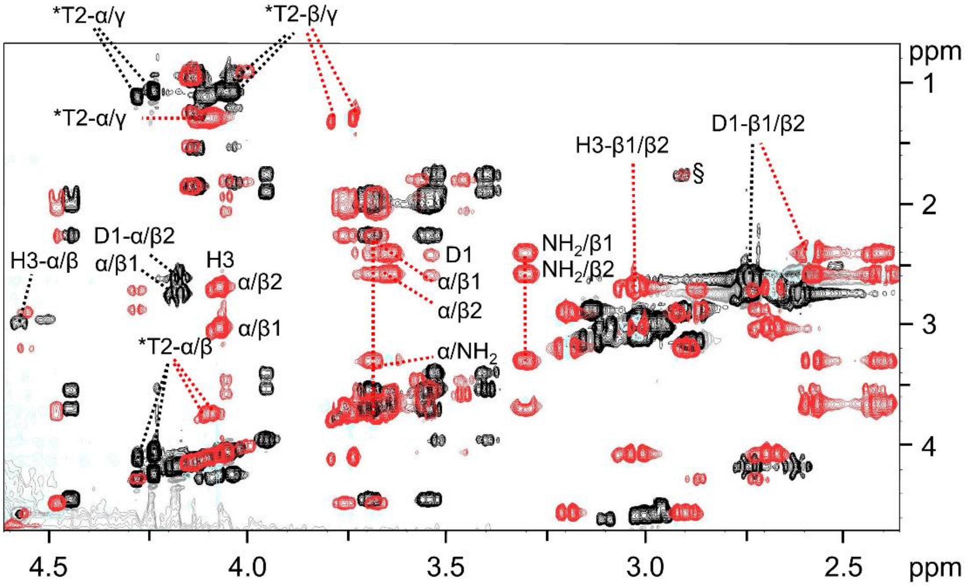


Figure 3.4: ^{1}H-^{1}H TOCSY spectra of the hexapeptide in the presence of Ni^{2+} (1:1) complex (red) and the absence of Ni^{2+} (black). The most strongly affected amino acids are those of the ATCUN motif, specifically Asp-1, Thr-2, and His-3. Additionally, upon Ni^{2+} binding, two amino protons of Asp-1 became visible and showed correlations to Asp-1's α– and β-protons in the TOCSY spectra. The symbols in spectra encode followings: * spectra contain a double set of peaks for each amino acid and suggest the presence of cis-trans proline isomerization [44], § indicates the presence of the impurities.

3.2 Oxidation and folding of liner hepcidin-25

As the major aim of this research was to determine the structure of Hep-25's metal complex by NMR, as a next step, we started with a chemically synthesized, linear Hep-25 (DTHFPICIFCCGCCHRSKCGMCCKT), which was purchased from Peptides & Elephants. As was mentioned in Section 1.4, Hep-25 contains eight cysteines in its sequence. Preparation of correctly folded, biologically active Hep, therefore, is challenging because four disulfide bonds have to be folded properly. Chemically synthesized material was folded according to the protocol adapted from Jordan et al. [44].The oxidation of linear human Hep-25 was carried out in the presence of folding solution at pH 7.5 for 16 hours (Figure 3.5). The folding solution consisted of 20% of acetonitrile and 0.25mM glutathione/glutathione disulfide (GSH/GSSG) redox system (GSSG/GSH=1/1). After 16 hours, the solution was acidified by the addition of trifluoroacetic acid (TFA) (0.1%) and lyophilized.

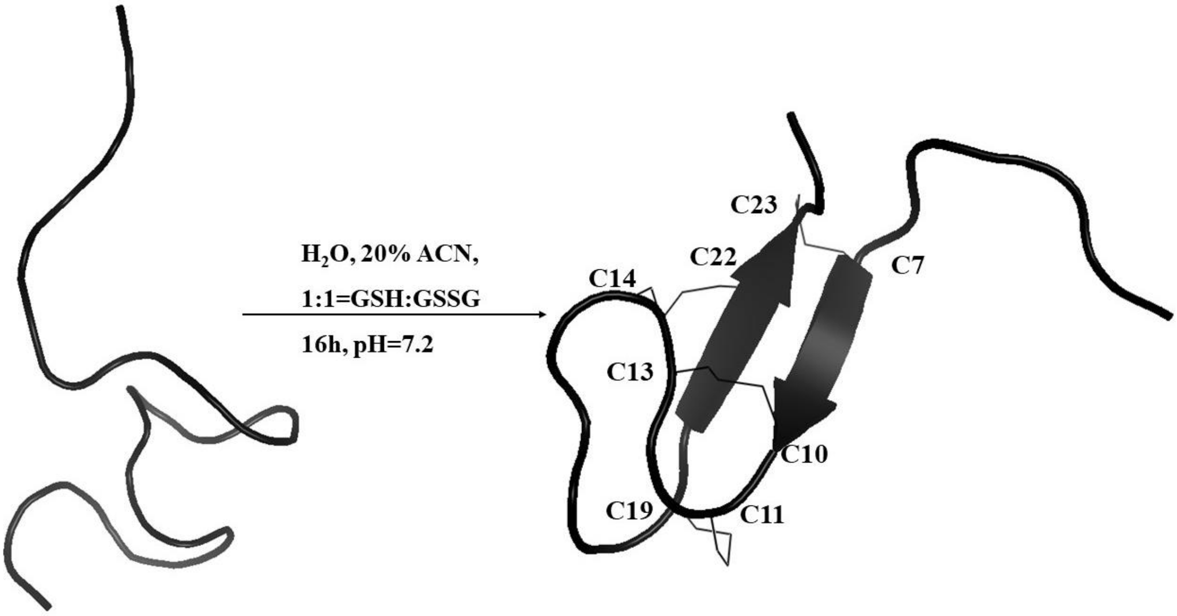


Figure 3.5: Oxidation of linear human Hep-25 was carried out in the presence of 20% of acetonitrile and 0.25mM glutathione/glutathione disulfide (GSH/GSSG) redox system (GSSG/GSH=1/1) at pH 7.5 for 16 h.

The lyophilized folding mixture was purified by HPLC. All preparative runs were carried out with a linear H_2O/ACN gradient, starting with 20% of ACN and going up to 60% of ACN. Both eluents were in the presence of 0.05% TFA. It could be observed that the folded product displayed an identical retention time of 24.5 minutes as the commercial Hep-25 standard, which reports the currently accepted disulfide connectivity between the following cysteines: Cys-7-Cys-23, Cys-10-Cys-13, Cys-11-Cys-19, and Cys-14-Cys-22. At this point, it could be concluded, as illustrated in Figure 3.6 A, that the peak with a retention time of approximately 24.5 minutes appeared to be folded Hep-25. The mass of this product, as determined by ESI-MS spectrometry, was 2787,0415Da, and the loss of 8 Da is consistent with the formation of four disulfide bonds (Figure 3.6 B). The accurate mass of folded Hep-25 was calculated to be 2787.0415 Da with the use of charge state deconvolution, which aligns with the theoretical value of 2787.0248.

RP-HPLC isolated folded Hep-25 in 22% yield. However, the preparation of larger amounts of folded Hep-25, as required for structural analyses, was considered costly. Even though labeling with ^{15}N would be possible via this approach, it would increase the costs beyond acceptable limits. In order to obtain larger amounts of peptide at reduced costs, in further study, we expressed Hep-25 in an *E. coli* strain as a recombinant fusion protein.

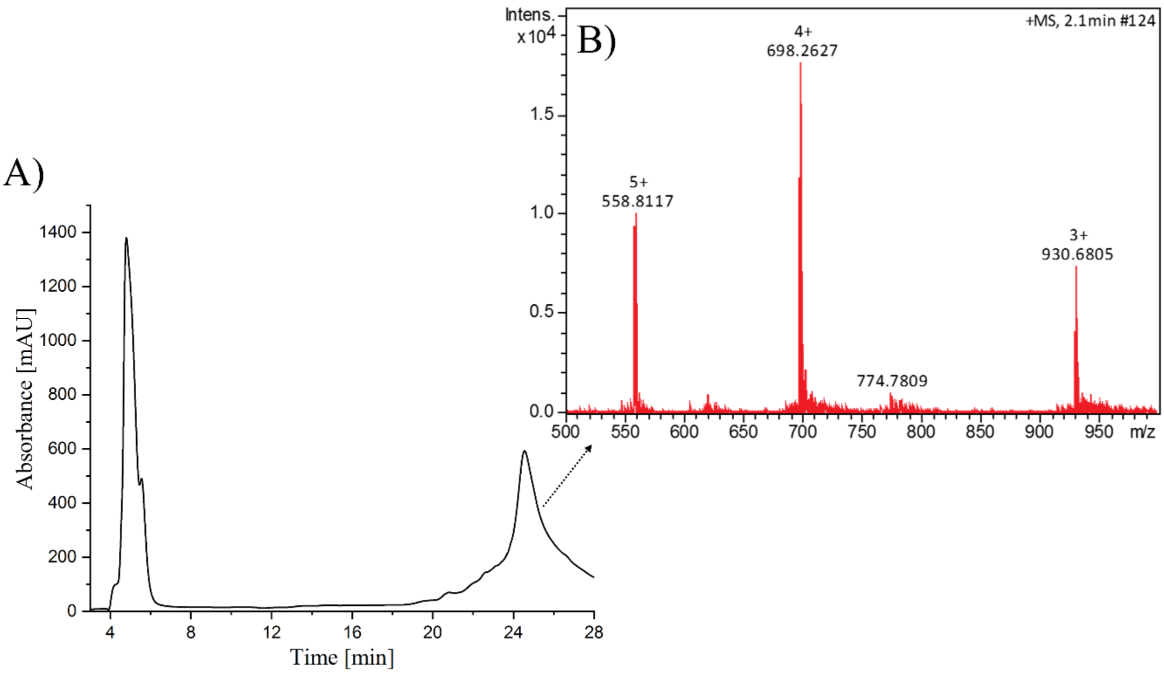


Figure 3.6: Purification of folded Hep-25 by oxidizing the chemically synthesized, linear peptide precursor. A) Preparative RP-HPLC profile of the crude peptide, purified after folding in the presence of glutathione/glutathione disulfide (GSH/GSSG) redox system (GSSG/GSH=1/1). The folded Hep-25 is eluted at 24.5 min in the 22% yield. B) The mass of the fraction with the retention time 24.5 min, as determined by ESI-MS spectrometry, was 2787.0415 Da and the loss of 8 Da is consistent with the formation of four disulfide bonds. Three major mass peaks of Hep-25 (M+5 (558.8148), M+4 (698.2620), M+3 (930.6837)) are present in ESI-MS.

3.3 Design of hepcidin-25 constructs

In order to solve the 3D structure of the Hep-25-metal complex, higher peptide concentration, and isotopic labeling of the peptide was considered to be necessary. High costs of purchased linear peptide and the low folding yield were the main reasons to design constructs encoding human Hep-25 for expression in *E. coli* (Section 3.2). The major problems during heterologous expression of eukaryotic proteins in *E. coli* are protein solubility, expression rate, and correct folding of the protein. These problems should be circumvented as the target was expressed as a fusion construct containing protein tags or short amino acid sequence [92]. Proteins or short amino acid sequences should have high solubility and an established folding mechanism in *E. coli*. The plasmid containing the genetically modified Hep-25 construct can be purchased or generated using various methods. In addition to the purchased plasmid, we used the FX cloning method and the insertion of the short amino acid sequences through site-directed mutagenesis, in order to obtain the desired fusion constructs.

3.3.1 PelB-Hep-25-His$_8$ and OmpT-Hep-25-His$_8$ constructs

One option for preparing recombinant peptides or proteins containing SS-bonds is to utilize the bacterial secretory pathway. To this end, the peptide of interest is fused to a secretory tag leading to the export of the fused peptide into the periplasmic space. There, the thiols are oxidized, and disulfide bonds are formed. Subsequently, the secretory tag should be cleaved off, optimally in a traceless fashion. Both steps or requirements can be achieved with the secretory signaling sequence of *E. coli.*

Hep-25 was fused to N-terminal secretory tag, also known as a signaling peptide, in order to express properly folded Hep-25 with four disulfide bonds and extract peptide from *E. coli* periplasmic space. The extraction from the periplasmic space is possible, as a secretory signal tag allows for the export of a protein into the periplasmic space with instant cleavage from the peptide. Another advantage of a secretory tag is that disulfide oxidoreductases and isomerases located in the periplasmic space catalyze the formation of the disulfide bonds [85, 86]. With this strategy, the properly folded Hep-25 can be purified in two steps, as the additional step for the cleavage of tag or folding of linear Hep-25 is not necessary. Several secretory sequences have been used for the efficient production of recombinant proteins in *E. coli*, including OmpT (outer membrane protein T) and PelB (pectate lyase B) [86]. In this work, secretory tags were fused to the target peptide with the FX-cloning method, which is described in detail in Section 2.1.2. This method is based on the exchange of the sequence of Hep-25 on a pDelivery vector, with a ccdB sequence, present on the pExpression vector. At this point, the idea was to confirm whether recombinant Hep-25 fused to secretory tag could be exported and then extracted from the periplasmic space. Therefore, we did not design a construct, but rather used a pExpression vector, which was already present in our FX-cloning library. The pExpression vector used in the present work contained the N-terminal secretory tag and a C-terminal histidine (His)$_8$ tag.

The constructs were then transformed into *E. coli* as PelB-Hep-25-His$_8$ and OmpT-Hep-25-His$_8$ (Figure 3.7). As was mentioned earlier, upon IPTG induction, Hep-25 should be expressed as Hep-25-His$_8$ in both cases. However, the expression and purification data revealed that the secretory production of Hep-25-His$_8$ in *E. coli* was not successful. (Section 3.4). Therefore, the secretory strategy was abandoned, and new fusion constructs were designed.

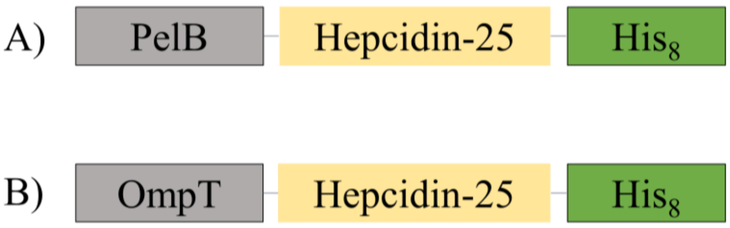


Figure 3.7: Schematic representation of the Hep-25 fusion construct in the presence of secretory tags (gray). Secretory tags export the protein into the periplasmic space with instant cleavage from Hep-25-His_8 and enable the formation of disulfide bonds by disulfide oxidoreductases and isomerases located in periplasmic space. A) PelB (pectate lyase B) and B) OmpT (outer membrane protein T) tags are fused N -terminally to Hep-25, and a His_8 tag (green) is fused C-terminally to Hep-25 (yellow). In both cases, the Hep-25 was fused to the secretory tag via the FX-cloning method.

3.3.2 The His_6-SUMO-Hep-25 construct

As an alternative to the secretory pathway, a whole range of other fusion tags are available. Fusion tags that enhance expression and increase protein solubility are, e.g., glutathione S-transferase (GST), maltose-binding protein (MBP), Thioredoxin (Trx), and small ubiquitin-like modifier (SUMO). Furthermore, fusion tags have been developed that facilitate purification of peptide/protein of interest, e.g., His-tags and Strep-tag. It should be noted that we have not attempted to express Hep-25 without tags, as it is common knowledge that peptides of small size have very low expression yields or are not expressed at all for several reasons.

For further studies, due to several advantages, we chose SUMO as the fusion partner [87] in combination with a histidine tag. It is known that a histidine tag facilitates generic purification. Additionally, SUMO has a few advantages, such as high solubility and a robust folding mechanism in *E. coli*, which improves recombinant protein expression [87]. Another advantage of the SUMO tag is the possibility of selective cleavage off the fusion partner. Several proteases, such as Thrombin, Factor Xa, and TEV protease, recognize short amino acid sequences. Thus, cleavage may occur within the protein of interest. In contrast, a SUMO tag could be selectively cleaved off the fusion partner with ULP1 (ubiquitin-like protein-specific protease) protease. The ULP1 protease has high selectivity because it recognizes the tertiary structure of SUMO rather than an amino acid sequence [87]. The protease cleaves precisely at the C-terminus of a SUMO tag so that the tag can be introduced at the N-terminus of the fusion protein only. For metal binding studies, this feature of the SUMO tag was of crucial importance. As the N-terminus of Hep-25 plays a key role in a metal binding study, cleavage of the fusion tag from the peptide was mandatory. Additionally, the histidine tag, attached to the N terminus of the SUMO tag, allows for easy purification of a recombinant protein from crude lysate. It is cleaved off together with SUMO tag. The His_6-SUMO-Hep-25 construct was codon-optimized and purchased as a GenScript DNA plasmid (Figure 3.8).

Figure 3.8: Schematic representation of the Hep-25 fusion construct. A His_6-tag and SUMO (green) is fused to the N-terminus of Hep-25 (yellow). As the N-terminus of Hep-25 plays a crucial role in the metal binding study, a His_6-SUMO tag was an ideal choice as it can be very selectively cleaved from the fusion partner with ULP1 (ubiquitin-like protein-specific protease) protease in a traceless fashion.

3.3.3 His_6-SUMO-Hep-25 constructs with additional solubility tags

The preparation of highly soluble and stable NMR samples remains a bottleneck for studies of many biological systems [88]. The application of solubility tags has been highly efficient in overcoming solubility problems and has enabled structural studies of challenging biological systems by NMR [89]. Due to the low solubility of Hep-25 at the pH=7.4 and pH=6.0, which is further discussed in Section 3.6, we decided to increase peptide solubility with short C-terminal peptide tags. These peptide tags are usually no longer than 15 residues and mostly comprise of one or two amino acids repeated a number of times. As a result, the solubility of the peptide could be increased without interference with the structure of the peptide of interest or a compromise of its activity. An extra step for the removal of the peptide tags after purification does not necessarily contrast with N-terminal protein fusion tags. In the case of Hep-25, N-terminal tags of any kind would interfere with metal binding and would, therefore, have to be removed prior to studies of structure and biological activity. Furthermore, protein fusion tags may interfere with studies because of their size. For example, they could degrade the quality of NMR spectra or block the binding site.

In the present study, we used solubility tags composed of five and six amino acids in order to reduce a possible interference with the structure of Hep-25. Solubility tags consisted of one or two alternating amino acids [89]. It is known that amino acid composition dictates the chemical and physical properties of peptides. The presence of extended sequences of hydrophobic amino acids in Hep-25 is responsible for low peptide solubility at neutral and basic pH. However, a slightly basic pH was mandatory for metal binding studies. For that reason, the idea was to add polar tags, and in this way, to increase hydrophilicity. It has been reported in previous literature that the presence of polar, especially polycationic and polyanionic amino acid tags enhances a protein's solubility [89]. The most investigated polycationic tags that have enhanced protein solubility consist of the polar basic amino acids arginine or lysine [89]. A comparison between arginine and lysine tags revealed that the solubility tag increased the protein solubility in both cases. However, arginine tags were more efficient than lysine tags of the same size in improving solubility, potentially due to the more hydrophilic character of arginine [89]. In the case of the polyanionic amino acid tag, aspartic acid and glutamic acid contribute equally to enhancing protein solubility and expression [89]. Polyionic tags, both positively and negatively charged, enhance solubility by introducing repulsive electrostatic interactions between similarly charged surfaces or by disrupting hydrophobic interactions between or within the same protein [89]. Therefore, we examined the effect of polar, polar polycationic, and polyanionic amino acid tags on Hep-25 solubility at neutral pH (see Section 3.6.3). We used the sequence of five serine residues as an uncharged polar tag. The polycationic tag was composed of five arginines in the row. Arginine is positively charged at pH 7 and the guanidinium group has only a weak donor capacity, so it does not form stable complexes with positive metal ions. Therefore, the arginine tag appeared as a perfect choice, as the focus of the study was to investigate the metal binding of Hep-25. In the case of the polyanionic tag, we used a stretch of six amino acids with aspartic acid and glutamic acid alternating (DEDEDE). However, the drawback of the aspartic acid and glutamic acid is that they could be potentially involved in the nickel and copper binding [90]. At pH values above 4.0, the oxygen donor atom of the carboxylic acid group is available for binding [91]. Therefore, it was necessary to prove that amino acids from these solubility tags interfere neither with the Hep-25 structure nor with the metal binding affinity of Hep-25 (Section 3.6.3 and 3.7.3).

Due to the advantages of the His$_6$-SUMO fusion tag mentioned above, the DNA sequence of additional solubility tags was inserted at the C-terminus of the His$_6$-SUMO-Hep-25 construct (Figure 3.9). The soluble tags were introduced at the C-terminus because the N-terminus of Hep-25 plays a vital role in metal binding, and, therefore, has to remain intact or unmodulated. The His$_6$-SUMO-Hep-25 vector was used as a starting point for the insertion of three different DNA sequences, which encode the solubility tags. The DEDEDE, SSSSS, and RRRRR tags were inserted during PCR amplification, according to the protocol in Section 4.2. A new DNA sequence was incorporated into the plasmid through site-directed mutagenesis using specifically designed forward and reverse primers. Data on DNA sequencing proved that all created plasmid constructs contained the desired mutations or insertions.

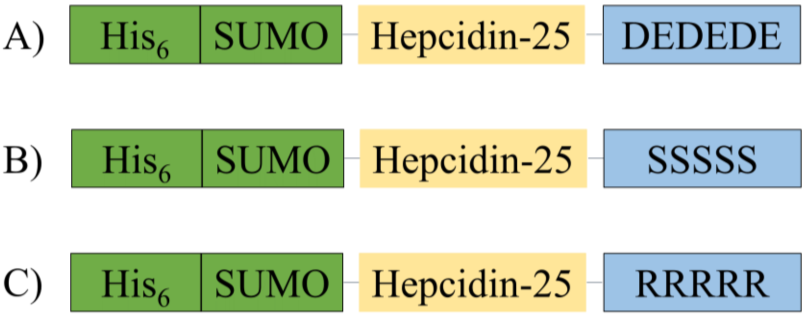


Figure 3.9: Schematic representation of His_6-SUMO-Hep-25-solubility tag constructs. A His_6-tag and SUMO (green) is fused to the N-terminus of Hep-25 (yellow). The additional solubility tag (blue) is fused directly to the C-terminus of Hep-25. Three different solubility tags were used in the study composed of A) negatively charged amino acids (DEDEDE), B) uncharged polar amino acids (SSSSS) and C) positively charged amino acids (RRRRR).

3.4 Expression and purification of hepcidin-25 constructs

The expression and purification of recombinant proteins are crucial steps in biochemical studies. Each step of the expression and purification thus have to be carefully designed and optimized in order to produce soluble and properly folded peptide. Once peptide fusion constructs have been created and cloned into a suitable expression vector, all desired constructs can be expressed in *E. coli*. However, the presence of several disulfide bonds, rare codon usage, and cytotoxicity are the major drawbacks of expression in *E. coli*. Here, due to a disulfide nature of Hep-25, two different approaches were designed in order to produce substantial yields of correctly oxidized and folded peptide. The first one includes a recombinant protein secretion into the periplasmic space and second intracellular expression. In both cases, the expression was induced under the control of strong T7 promoter in *E. coli* Origami B strain, which is a genetically engineered host for production of disulfide rich proteins.

Furthermore, secretory and intracellularly produced recombinant proteins were purified with the use of two main protocols. In the first case, secreted peptide should be already cleaved from PelB or OmpT tag and secreted into the periplasmic space. Additionally, secreted peptide should be oxidized because of disulfide oxidoreductases and isomerases located in the periplasmic space [85, 86]. The protocol was based on the extraction of Hep-25-His_8 by a cold osmotic shock from periplasmic space (Figure 3.10 A). The extracted proteins and peptides were further purified via affinity chromatography followed by preparative reversed-phase high-performance liquid chromatography (RP-HPLC). Unfortunately, the extraction and purification of Hep-25-His_8 were not successful. In the second protocol, intracellularly expressed constructs were isolated from *E. coli* cells by high-pressure homogenization (Emulsiflex) (Figure 3.10 B). His_6-SUMO-Hep-25 constructs should also be oxidized and at least partially folded because of oxidative conditions inside *E. coli* Origami B. However, purification via affinity chromatography and the cleavage of tags were performed under reducing conditions, followed with size exclusion chromatography. Therefore, Hep-25 was oxidized and folded in the last step of purification and purified via RP-HPLC, as Figure 3.10 B illustrates.

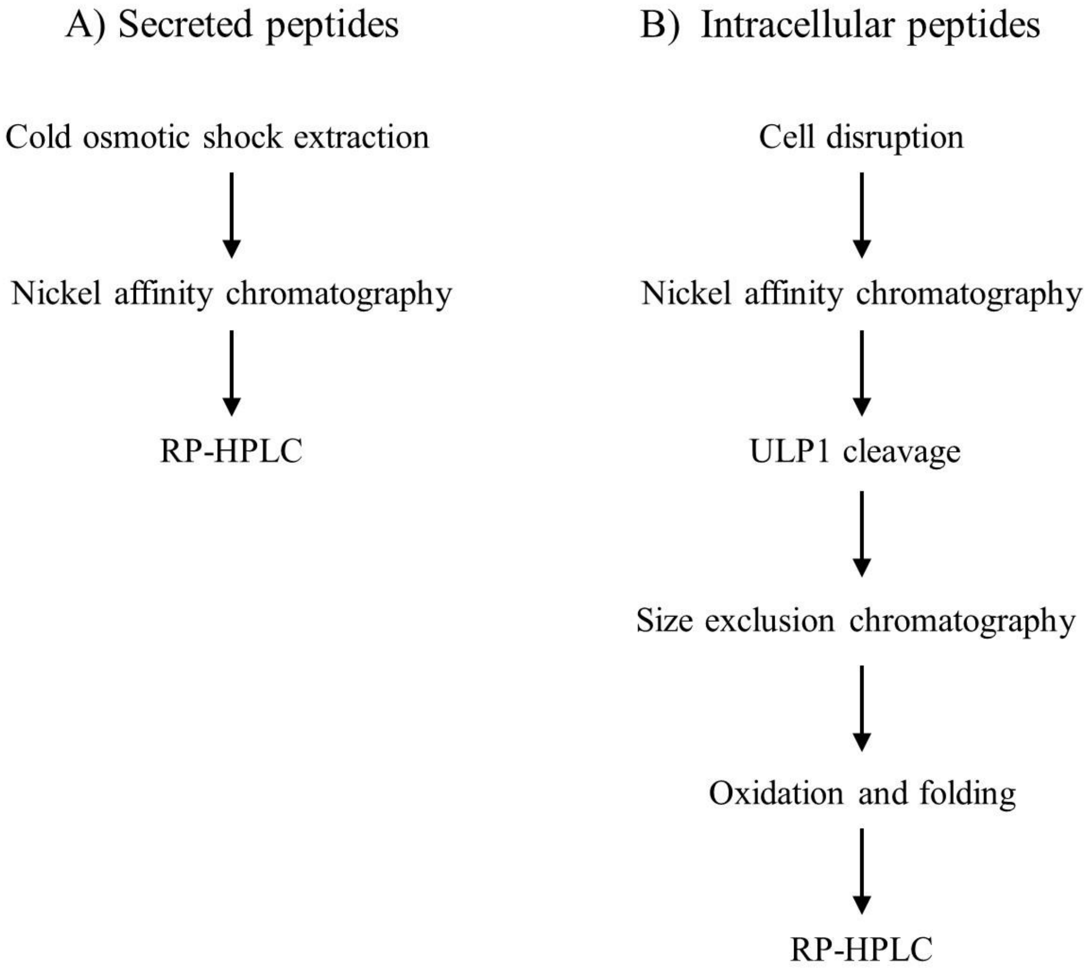


Figure 3.10: Purification scheme of Hep-25 fusion proteins. Purification of Hep-25 constructs consisted of two main protocols. A) The first one includes a recombinant protein secretion into the periplasmic space and involves the extraction of the secreted construct by a cold osmotic shock from periplasmic space. B) The second protocol includes release of intracellularly expressed constructs from E. coli cells by cell disruption. In both cases, the fusion protein was isolated via Ni^{2+} - affinity chromatography. Further purification steps were carried out, depending on the peptide construct. One protocol involved further direct purification of the fusion protein with HPLC (A). For the second protocol, the fusion

protein was first cleaved, cleaved products were separated by size exclusion chromatography, and then, after folding, the peptide was purified by HPLC (B).

All molecular constructs used in this work had an affinity tag with eight or six histidine residues at the C - or N-terminus, respectively. An affinity tag with histidine residues has a high affinity for nickel and, therefore, the tag is bound tightly to the nickel immobilized on the nickel affinity column. A histidine tag serves to separate the protein via the nickel affinity chromatography from all other soluble cell constituents. Constructs presented as SUMO fusion proteins were then digested with the ULP1 protease to cleave off SUMO of the target peptide. Conventionally, His-SUMO and ULP-His could then be separated from the target protein via affinity chromatography. However, a separation of His_6-SUMO and ULP1 protease from Hep-25 is not possible with nickel affinity chromatography due to the high affinity of hepcidin's ATCUN motif for nickel (see Section 3.1.2.). Therefore, size exclusion chromatography was performed to separate Hep-25 from all other proteins, followed by direct oxidation and folding of Hep-25. Following the peptide purification, in both protocols, preparative reversed-phase high-performance liquid chromatography (RP-HPLC) was performed to isolate pure folded Hep.

3.4.1 Expression of recombinant hepcidin-25

All Hep-25 constructs were expressed in *Escherichia coli* according to the same protocol in Section 4.6. DNA encoding a His6-SUMO-Hep-25 or His6-SUMO-Hep-25-solubility tag was a part of the pET15b expression vector. In the case of PelB-Hep-25-His_8 and OmpT-Hep-25-His_8, the target fusion construct was encoded on the pQE vector. Vectors were then introduced into an expression host cell. In practice, protein expression can be very challenging because so many factors can influence the process. The expression level and solubility of proteins could be affected by several factors, including protein structure and size, host strain, inducer concentration, and growing temperature. Each protein folds in its unique manner, a process that may be influenced by the choice of the expression host and other fractions. Therefore, in the first step, we chose an expression host, and then we optimized protein expression of Hep-25 constructs. For each construct, a plasmid was transformed first into BL21 (DE3) and then into Origami B strains. In all cases, upon isopropyl β-D-1-thiogalactopyranoside (IPTG) induction in the BL21 strain, the desired protein was not expressed. Hep-25 is an aggregation prone protein since it contains multiple disulfide bonds and, therefore, we could assume that the reductive environment of cytoplasm of BL21 (DE3) strain does not allow the desired protein to be expressed. Therefore, the desired protein was expressed into Origami B strain, which is a genetically engineered host for production of disulfide rich proteins. In the second step, the protein was expressed at different temperatures (at 20, 25, 30, and 37 °C) and in the presence of 0.25 or 1 mM IPTG. Lowering the induction temperature from 37 to 20 °C and IPTG concentration to 0.25mM resulted in higher peptide expression of all constructs. This result suggests that induction at a lower temperature can improve expression yield in *E. coli*. Therefore, protein expression was induced with 0.25mM IPTG at 0.6 OD_{600} and further incubated at 20 ° C for 16 hours. For the NMR studies, Hep-25 was expressed in unlabeled LB-medium and isotopically labeled forms in the minimal growth medium. As the only source of nitrogen and carbon, the minimal medium M9 contains ammonium chloride and glucose. All Hep-25 fused contracts were prepared using Origami B strains and expressed according to the same protocol for labeled and unlabeled media. All His6-SUMO-Hep-25 constructs were successfully expressed, as Figure 3.11 illustrates. However, in the case of PelB-Hep-25-His_8 and OmpT-Hep-25-His_8, expression and secretion of Hep-25-His_8 were not visible on SDS gel upon IPTG induction in both cases. As was mentioned earlier, upon IPTG induction, the secretory tag cleaved off the target peptide. For that reason, we were not sure whether the target peptide was not expressed or whether we were unable to see a band on the SDS gel because of the low molecular weight of Hep-25. This is despite the fact that we used Tricine-SDS gels which offer high resolution, especially for small proteins and peptides range. However, after the purification of Hep-25-His_8, described in Section 3.4.2, we could only conclude that expressions and purifications of the secreted peptides were not successful.

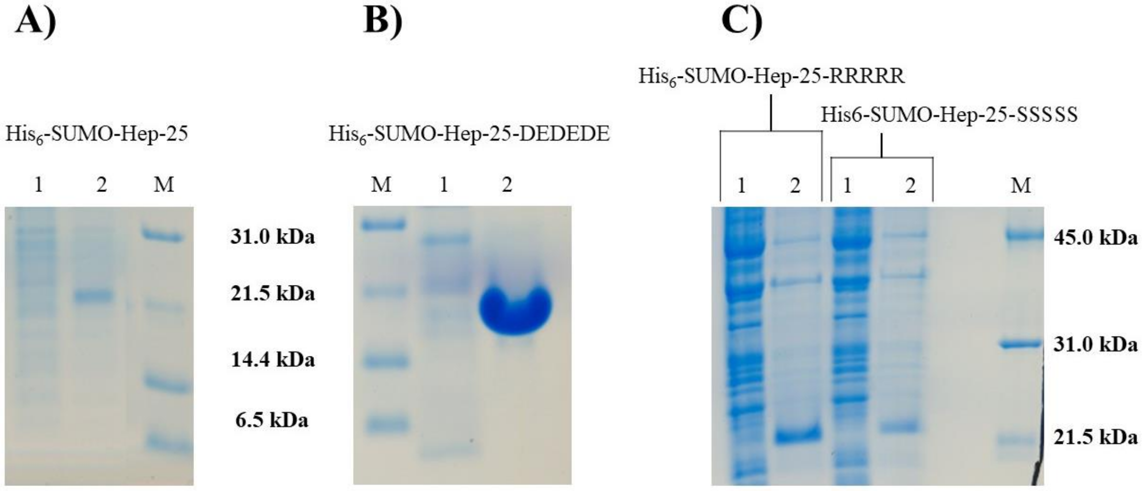


Figure 3.11: Expression of His_6-SUMO-Hep-25 constructs. Two samples from each expression of different constructs were loaded onto the gels. Band M: Broad-Range marker. Band 1: Cell growth at 37 °C Band 2: Target protein expression upon IPTG induction at 25 °C. SDS–PAGE of A) His6-SUMO-

Hep-25 expression. B) His6-SUMO-Hep-25-DEDEDE expression. C) His6-SUMO-Hep-25-RRRRR and His6-SUMO-Hep-25-SSSSS expression. Expression of all four proteins was visible on the SDS gel after IPTG induction.

The cell yield was comparable between expressed constructs (His6-SUMO-Hep-25, His5-SUMO-Hep-25-DEDEDE). Cell yield of the unlabeled sample was approximately 3g cells per liter of a nutritionally rich medium. In contrast, the yield of labeled protein was overall reduced by approximately 1.5g cells per liter due to the low growth of cells in the isotope-labeled medium.

3.4.2 Purification of PelB- and OmpT-Hep-25-Hep$_8$ constructs

Based on the SDS gel, we were not able to prove whether PelB-Hep-25-His$_8$ and OmpT-Hep-25-His$_8$ were expressed (Section 3.4.1). We assumed that due to low Hep-25-His$_8$ molecular weight and low quantity of expressed peptide, we might not have been able to detect a band on the SDS gel. Thus, we proceeded with peptide extraction and purification. As was mentioned in Section 3.4, the secretory tags, PelB, and OmpT, should lead to Hep-25-His$_8$ in the periplasmic space with the secretion peptide cleaved off upon export. The periplasmic space is located between the outer and inner membrane of *E. coli.* Membranes are semipermeable, and certain molecules pass through them in certain cases, such as upon osmotic shock. During osmotic shock, the *E. coli* cell is exposed to the hypotonic solution followed by the subsequent addition of hypertonic solution. Upon exposure of cells to osmotic shock, Hep-25-His$_8$, other periplasmic proteins, metabolites, and other cytoplasmic proteins should be released. Therefore, harvested cells were firstly suspended in a hypertonic solution, and then the hypotonic solution was added to release proteins and peptides from periplasmic space. The periplasmic solution was further purified through nickel affinity chromatography, since constructs contain the N-terminal ATCUN motif and C-terminal His$_8$ tag. The presence of C-terminal His$_8$ tag was not part of the construct design. Instead, we used PelB and OmpT constructs from our group FX-cloning library for test expression and purification. As a result, Hep-25 contained a C-terminal histidine (His)$_8$ tag. Finally, the collected fractions were purified via HPLC. However, we were not able to see the band of a Hep-25-His$_8$ on the SDS gel neither after extraction nor in fractions after affinity chromatography. The situation was the same in both cases with PelB and OmpT constructs. The optimization of the pH and salt concentration did not yield better results. Unfortunately, we were also not able to confirm the presence of Hep-25-His$_8$ through ESI-MS. As the focus of our study was the structural determination of the peptide-metal complex, we did not further optimize the production of periplasmic Hep-25-His$_8$. We could only propose two possible reasons for the unsuccessful extraction of the peptide from the periplasmic space. First, the heterologous peptide was secreted from *E. coli* into the culture medium rather than into periplasmic space. However, due to the high peptide dilution in the culture medium, we were not able to see peptide on the SDS gel and prove this. Second, the secretory production of heterologous proteins in *E. coli* was inefficient. Therefore, Hep-25-His$_8$ was not expressed or expressed in low amounts or degraded intracellularly or extracellularly. One of the possible reasons for ineffective expression is the presence of rare codons in PelB and OmpT DNA sequences, present in the DNA of constructs from our FX-constructs library. The frequency of codon usage plays a crucial role when a recombinant protein is expressed in different organisms. It is known that rare codon content reduces the efficiency of recombinant protein expression, and this content is used with low frequency in *E. coli* as compared to the usage in the original host. Therefore, rare codons should be optimized to commonly used codons in *E. coli* in order to enhance the expression efficiency of PelB-Hep-25-His$_8$ and OmpT-Hep-25-His$_8$.

3.4.3 Purification of His$_6$-SUMO-Hep-25

Since the purification protocol was new for the His$_6$-SUMO-Hep-25 construct, preliminary work revealed that cell disruption and SUMO digestion were critical steps in the purification. Therefore, the buffer conditions of those purification steps were optimized to increase the final protein yield.

Upon protein expression, the *E. coli* cells were first suspended in a 50mM phosphate buffer pH 7.4, in the presence of 300mM NaCl, and then disrupted by high-pressure homogenization (Emulsiflex). The cell lysate was then centrifuged at 100,000 g to pellet the membrane fragments and insoluble cell components. Since the His6-SUMO-Hep-25 construct should be oxidized and at least partially folded inside of Origami B strain, and is thus most likely not prone to aggregation, we assumed that oxidized and soluble protein may be released upon cell disruption. However, it turned out that after cell disruption and centrifugation protein was present in the form of insoluble protein. We had two possible hypotheses: a protein was expressed in the form of inclusion bodies or expressed as soluble proteins in *E. coli* and aggregated after cell lysis. Inclusion bodies are commonly referred to as highly aggregated protein formed during high-level expression. Aggregation of recombinant proteins during expression or after cell lysis is driven by the association of partially folded or misfolded peptide intermediates. In both cases, the recovery of protein from the aggregated state is typically accomplished by solubilization with denaturing agents and later refolding to obtain natively folded peptide. As various factors can lead to protein aggregation, the lysis buffer was optimized with various additives in order to extract the protein from cells in the soluble form (Table 3.1).

Table 3.1: Tested lyses buffer conditions

Additives	50mM phosphate buffer, 300mM NaCl pH 7.4					
DTT	5mM	20mM	20mM	20mM	20 mM	20mM
Triton X-100	-	2%			-	-
Tween 20	-	-	2%		-	-
Urea	-	-	-	-	6M	6M
Guanidine HCl	-	-	-	-	-	1M
Cell disruption						
	Insoluble protein	Partially soluble Protein				Soluble protein

In the presence of various salt concentrations, protein aggregates were solubilized to a small extent (10%) with this method. We assumed that the misfolding of cysteine residues caused aggregation. Native Hep-25 contains four well-defined disulfide bonds. However, the presence of the large His$_6$-SUMO tag could interfere with native peptide folding and caused aggregation. In order to solubilize and thus achieve a higher yield, a reducing agent was included in the lysis buffer. Dithiothreitol (DTT) was included in lysis buffer to reduce disulfide bonds. Subsequently, two different DTT concentrations of 5mM and 20mM were tested in order to isolate soluble protein from *E. coli* cells (Table 3.1). We noticed that a better result was achieved with 20mM DTT. However, still, one half of the protein was present in the cell pellet even after the addition of 20mM DTT. The result demonstrated that aggregation occurs not only because of free cysteine residues but presumably involves other intermolecular interactions between partially folded proteins. We could not use the higher DTT concentration, since a high concentration of reducing agents, above 5mM DTT, would result in nickel leakage from chelating chromatography resins. Even with 20mM DTT, it was necessary to dilute the sample prior to affinity chromatography to 5 mM DTT. Later on, the freshly lysed recombinant Hep-25 was kept in linear form and purified in the presence of 2.5 mM DTT, which was included in the purification buffer. The presence of DTT was necessary in order to prevent disulfide bond re-formation and subsequent protein aggregation and enable the protein to remain as a monomer.

As was mentioned earlier, even with the addition of 20mM DTT, half of the protein was still present in an insoluble form. To make further progress in lysis buffer optimization, in a second step, we investigated the effect of several detergents and denaturing agents (Table 3.1). The addition of the detergents Triton X-100 and Tween 20 did not influence the solubility of the protein, so they were excluded from lysis buffer and further purification. In the case of the denaturing agents, urea and guanidium hydrochloride, the situation was different. The addition of urea into the lysis buffer supplemented with 20mM DTT significantly increased the amount of solubilized protein. However, complete protein solubilization was achieved with additional guanidine hydrochloride (Table 3.1).

The concentration of denaturing agents was optimized in order to achieve a soluble protein, without creating any adverse effect on the further purification steps. The critical point that was affected by a higher concentration of urea or guanidium hydrochloride was ULP1 protease cleavage. Guanidine hydrochloride and urea can denature a SUMO structure and ULP1 as well. Therefore, a ULP1 protease would not be able to recognize the tertiary structure of SUMO and cleave a tag. A previous study demonstrated that ULP1 cleaves His$_6$-SUMO constructs effectively in the presence of 2M urea [87]. Within this concentration of urea, the 3D structure of His$_6$-SUMO remains intact. Optimization of all purification steps revealed general aggregation in the absence of urea. Therefore, urea, together with DTT, was also included in the purification buffer.

Guanidine hydrochloride, in combination with urea and DTT, had a significant impact on protein solubility. However, it could not be present in the solution during ULP1 cleavage. In order to be sure that guanidine hydrochloride from lysis buffer reversibly denatured a SUMO tag and, thus, did not affect the 3D structure of His$_6$-SUMO and cleavage, we tested the effect of different concentrations of guanidine hydrochloride on the SUMO refolding. We investigated the effect of 0.4M, 0.5M, and 1M guanidine hydrochloride on His$_6$-SUMO-Hep-25 refolding. Prior to ULP cleavage, His$_6$-SUMO-Hep-25 was refolded by removing guanidine hydrochloride with size-exclusion chromatography. Comparison of protein bands on the SDS gel, before and after cleavage, suggested that in all three cases (0.4M, 0.5M, and 1M guanidine hydrochloride), the cleavage at 25ºC was completed (Figure 3.12). Therefore, 1M guanidine hydrochloride was found to be present in the lyses buffer, but excluded from further purification steps.

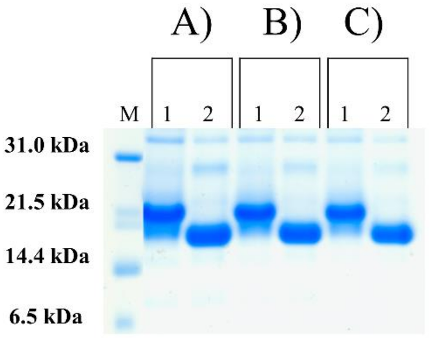


Figure 3.12: Effect of A) 0.4M, B) 0.5M and C) 1M guanidine hydrochloride on the yield of ULP1 cleavage. The letter and numbers encode followings: Band M: Broad-Range marker. Band 1: Before ULP1 cleavage, His6-SUMO-Hep-25 was firstly separated from guanidine hydrochloride by size-exclusion chromatography and then refolded from the guanidine-denatured state. Band 2: Refolded protein was cleaved in the presence of ULP1 protease. In all three cases, the cleavage at 25ºC was completed.

The final cell lysis protocol included protein solubilization under denaturing and reducing conditions in the presence of 300mM NaCl, 6M urea, 1M guanidine HCl, and 20mM DTT. The lysis buffer was also supplemented with deoxyribonuclease and protease inhibitors. A general protease inhibitor cocktail *(*SIGMA*FAST*™ Protease Inhibitor Cocktail Tablets, EDTA-Free) was used to inhibit degradation of His_6-SUMO-Hep-25 protein in the tissue or cell extracts, which is caused by various types of proteases. This optimized lysis buffer yielded complete His_6-SUMO-Hep-25 solubilization. The protein solution was diluted four times with phosphate buffer to a final concentration of 5mM DTT before Ni^{2+} affinity chromatography in order to avoid nickel leakage from chelating resins. The solubilized target protein was first separated by affinity chromatography from the rest of the soluble cell components, as described in Section 3.4. However, the imidazole elution of a protein from nickel affinity chromatography revealed that some amount of protein was cleaved by a protease. We noticed one main peak eluted from 11 to 14mL, followed by an additional shoulder peak elution from 15 to 19mL. According to the SDS-PAGE gel monitoring the elution, fractions from an elution volume of 11 to 19mL contained fully degraded His_6-SUMO-Hep-25 (see Figure 3.13). The SDS gel indicated that bands attributed to protein degradation only appear after cell disruption (see Figures 3.11 and 3.13). Therefore, it was found that degradation occurs during cell disruption and not during the expression itself. The estimated yield of degradation was 90% according to the SDS-PAGE of several purifications.

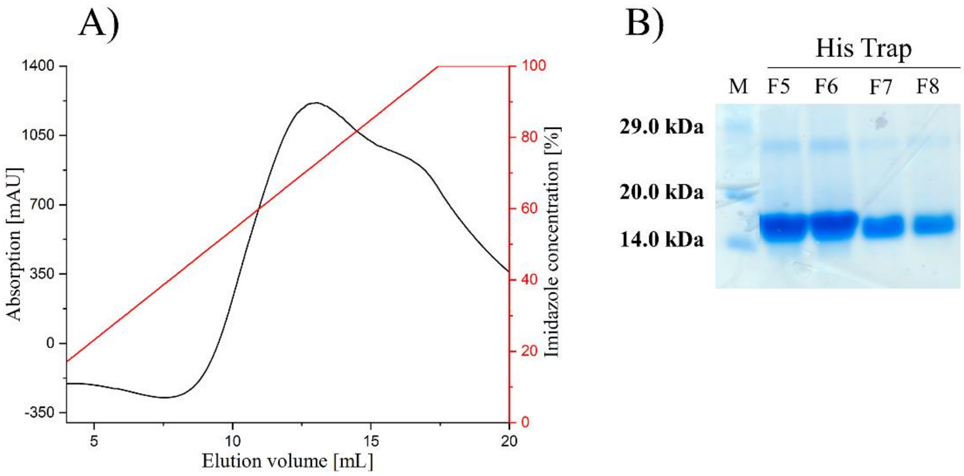


Figure 3.13: Proteolytic cleavage of His$_6$-SUMO-Hep-25. A) Ni^{2+} affinity chromatogram of His$_6$-SUMO-Hep-25 showing one main peak eluting from 11 to 14mL followed by an additional peak from 15 to 19mL. B) SDS-PAGE of fractions from the Ni^{2+} affinity chromatography. Band F5-F6: Fractions from the main peak from the chromatogram, containing proteolytically cleaved His$_6$-SUMO-Hep-25. Band F7-F8: Fractions from the shoulder peak which contained the proteolytically cleaved His$_6$-SUMO-Hep-25.

In order to prevent the proteolytic cleavage of His$_6$-SUMO-Hep-25, the cell lysis was additionally optimized. The lysis buffer was supplemented with the higher amount of the general protease inhibitor cocktail. Therefore, apart from the lysis buffer, protease inhibitor was additionally added to the cell lysate directly after cell disruption and, again, after centrifugation at 100,000 g. As is illustrated in Figure 3.14, the degradation effect was eliminated with an additional amount of the protease inhibitor cocktail, and the protein was successfully separated from all other cell constituents with Ni^{2+} affinity chromatography. His-Trap loading, and elution buffer were supplemented with DTT and urea.

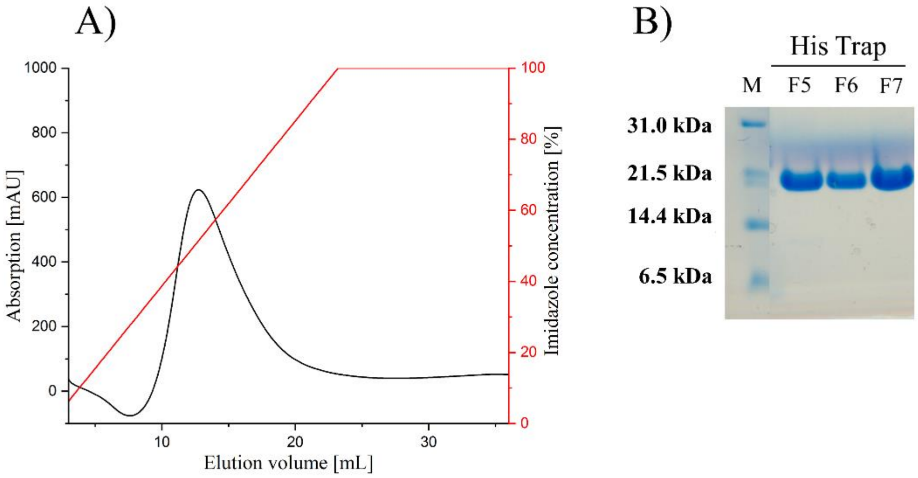


Figure 3.14: His_6-SUMO-Hep-25 purification by Ni^{2+} affinity chromatography. A) His_6-SUMO-Hep-25 was separated by affinity chromatography from the rest of soluble cell components that were released after cell lysis. The protein was eluted with approximately 200mM imidazole. B) SDS-PAGE of fractions from the Ni^{2+} affinity chromatography. The letter and numbers encode followings: Band M: Broad-Range marker. Band F5-F7: Fractions eluted between 11-14mL contained His_6-SUMO-Hep-25. From SDS gel is clear that with an additional amount of protease inhibitor, there were no decomposition products.

The next step of the purification protocol was SUMO cleavage. This turned out to be critical as well. Since we were aware from earlier purification steps that Hep-25 fusion construct tended to aggregate, the protein was digested with the ULP1 protease in the presence of two different DTT concentrations: 5mM and 20mM DTT. However, in both cases, the sample was incompletely cleaved because the aggregation presumably reduced the accessibility of the SUMO cleavage site to ULP1 protease. In order to eliminate inefficient cleavage, it was necessary to include the denaturing agents in order to solubilize aggregates. As the efficiency of cleavage is reduced with guanidine HCl, the fusion protein was digested with the ULP1 protease in the presence of 20mM DTT and 2M urea. Under these conditions, oligomers were dissociated into monomers, thereby allowing the ULP1 proteases to recognize the tertiary structure of the SUMO tag. Indeed, in the presence of these additives, the tag was completely cleaved off (Figure 3.15 A). However, Hep-25 precipitated after cleavage of His_6-SUMO tag, probably due to the aggregation of linear Hep-25. The protein precipitation was reversed with the additional heating of the sample at 42ºC with 3M guanidine HCl.

The usual way to separate His_6-SUMO and ULP1 from the target protein is purification via Ni^{2+} affinity chromatography. However, as was already mentioned, upon the cleavage, the N-terminus of Hep-25 was liberated, and the ATCUN motif would be able to bind the nickel of the nickel-Sepharose. Therefore, the separation of Hep-25 from His$_6$-SUMO and ULP1 protease via the nickel-Sepharose was not feasible. The liberated Hep-25 has finally separated from His$_6$-SUMO and ULP1 protease through size exclusion chromatography. The elution profile of Hep-25 from the Superdex 30 16/600 column is shown in Figure 3.15 B. Since the presence of additives would contribute to the misfolding of Hep-25 in the following folding step, the phosphate buffer supplemented with 150mM NaCl was used as mobile phase for size exclusion chromatography. The pH of the buffer was reduced to pH 7 because the isoelectric point (pI) of liberated Hep-25 is 8.10. Size exclusion chromatography demonstrated that under this condition, Hep-25 was partially aggregated. As the oxidation of Hep-25 aggregates had a meager folding yield, only fractions that contained Hep-25 monomer and dimer were combined and further oxidized. Calibration of the size exclusion column revealed that dimeric and monomeric peptide were eluted with a volume of 70 and 94mL, respectively. As the Hep-25 sample absorb weakly at 280 nm, size exclusion chromatography was monitored by UV absorption at 210nm. However, at 210nm, the peaks broadened, as the measurement of absorption at 210nm is far more prone to external influence. This is because peptide bonds, many solvents, and other chemicals absorb at this wavelength.

Linear Hep-25 was oxidized in the final step of purification. It was folded through the application of an adapted protocol based on Jordan et al. [44], as presented in Section 3.2. The complete formation of the intramolecular disulfide bonds was carried out overnight (over approximately 16 hours) at pH 7.0 in the presence of 0.25mM glutathione oxide-reducing system and 20% ACN. Reducing the pH from 7.5 to 7.0 yielded better, but still low, peptide folding. However, the low folding yield of the purchased and recombinant-purified linear Hep-25 proved the low solubility of Hep-25 even at pH 7.0. It later became apparent with the investigation of solubility (Section 3.6) that solubility of peptide played a crucial role during the folding. It is important to note that Hep-25 is a basic and hydrophobic peptide. Therefore, the use of an organic solvent (ACN) was also found to be necessary to enhance the solubility of the peptide during folding. Altogether, the pH and organic solvent did not significantly increase the folding yield. With the higher solubility of the peptide, the folding yield increased, which was observed with the folding of solubility tag constructs. This is discussed further in the text (Section 3.6). In order to separate the folded Hep-25 from glutathione and different possible folding products, the sample was further purified by preparative reversed phase HPLC (Figure 3.15 C). All preparative runs were carried out with a linear AB gradient, starting with 20% of eluent B and going up to 60% of eluent B at room temperature and at a flow rate of 3ml/min. Eluent A was 0.05 % trifluoroacetic acid (TFA) in H_2O, and B was 0.05% TFA in acetonitrile. The HPLC analysis demonstrated that only one single folding product was present in all cases. The retention time of the folded product was 24.5 minutes. However, the same retention time was not proof of correct disulfide bond connectivity. Therefore, NMR measurement was necessary to confirm the same disulfide connectivity as the currently accepted disulfide connectivity.

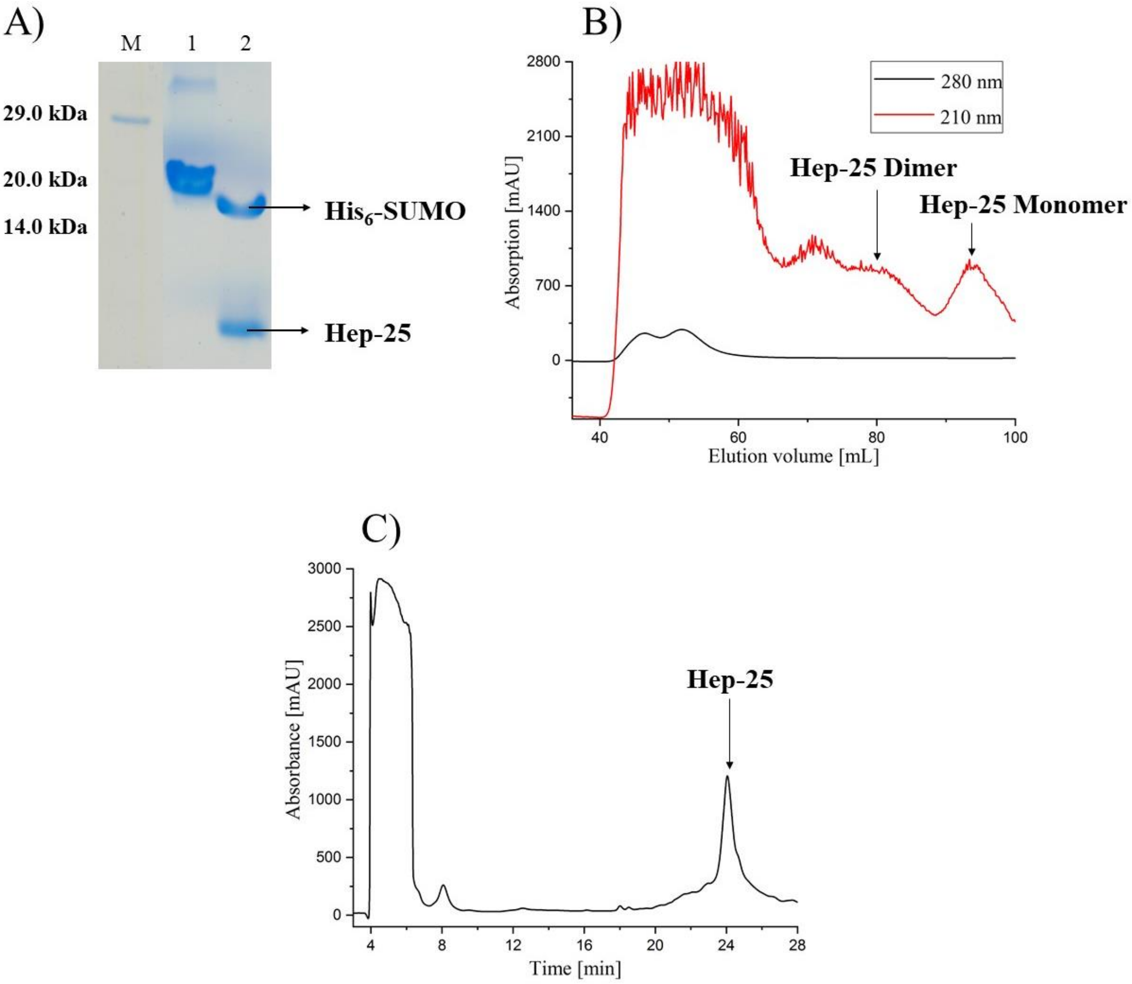


Figure 3.15: Purification of Hep-25 after ULP1 cleavage. A) SDS-PAGE of the individual purification steps. Band M: Broad-Range marker. Some marker bands diffused out of the gel during the long destaining with water. Band 1: After the first purification step, over the nickel-Sepharose, a large part of the unwanted cell components was removed, and the fusion protein is seen near 21kDa. Band 2: After digestion, two strong bands of the cleavage products His_6-SUMO and Hep-25 are visible. B) Elution profile of Hep-25 from the Superdex 30 16/600 column. ULP1, His_6-SUMO, and the aggregated Hep-25 were present in the exclusion volume. Dimeric and monomeric protein were eluted with an elution volume of approximately 70 and 94mL, respectively. In each step, the protein was monitored by UV absorption at 280nm, 210nm. C) Preparative RP-HPLC profile of the Hep-25, purified after oxidation and folding in the presence of glutathione/glutathione disulfide (GSH/GSSG) redox system

(GSSG/GSH=1/1). The folded Hep-25 is eluted at 24.5 min at an overall yield of approximately 0.7mg per liter of LB medium.

The same protocol was applied for the purification of unlabeled and ^{15}N labeled Hep-25. In each step of the purification, the protein was monitored through UV absorption at 280nm, 210nm. After RP-HPLC, approximately 0.7mg of Hep-25 and 0.6mg of ^{15}N-Hep-25 were isolated from 1L LB and M9 medium, respectively. The concentration of the sample was determined based on a measurement of the absorbance of the four disulfide bonds at 280nm because Hep-25 does not contain Tyr or Trp in the sequence. It should be noted that determining the concentration of Hep-25 is not precise. It is known that the absorbance of a protein/peptide at 280nm depends on the content of Trp, Tyr, and disulfide bonds, but cystine has a relatively low extinction coefficient (125 M^{-1}/cm^{-1}) in comparison to Tyr and Trp [93].

3.4.4 Purification of Hep-25 constructs containing an additional solubility tag (His$_6$-SUMO-Hep-25-solubility tag)

In order to obtain Hep-25 with its native, N-terminal ATCUN motif, but with better solubility for NMR investigations, additional solubility tags were fused to the Hep-25's C-terminus. Three constructs were investigated: His$_6$-SUMO-Hep-25-DEDEDE His$_6$-SUMO-Hep-25-SSSSS, and His$_6$-SUMO-Hep-25-RRRR. Upon expression, His$_6$-SUMO-Hep25-DEDEDE, -SSSSS, and -RRRRR were purified according to the same protocol for purification that was used in the case of His$_6$-SUMO-Hep25. The protocol consists of five steps: cell lysis, purification by affinity chromatography, ULP protease cleavage, purification by size exclusion chromatography, peptide oxidation, folding, and final purification by HPLC (Section 3.4.3). Unless otherwise noted, the same buffer conditions were applied as for the purification of His$_6$-SUMO-Hep-25. However, the pH of buffers was adjusted for all proteins processed here according to their isoelectric points (pI). All constructs were lysed in 50mM phosphate buffer supplemented with 300mM NaCl, 20mMDTT, 0.6M urea, and 1M guanidine HCl. Under these conditions and in the presence of an additional amount of protease inhibitor, all solubility tag constructs were extracted as soluble proteins from *E. coli* after cell disruption. The solubilized target protein, which contained the remaining soluble portion of the cells, was first purified by nickel affinity chromatography. Subsequently, the fusion protein was digested with the ULP1 protease in the presence of DTT and urea and further purified via size exclusion chromatography. Linear Hep-25-solubility tag was oxidized in the final step of purification in the presence of 0.25mM glutathione oxide/redox system and 20% acetonitrile. In order to separate the folded Hep-25 from glutathione, the sample was further purified by preparative reversed phase RP-HPLC (Figure 3.16).

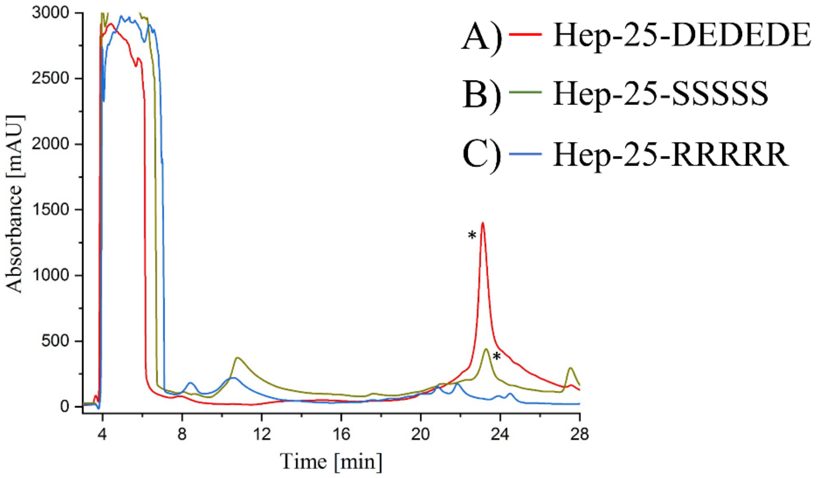


Figure 3.16: The oxidative folding of Hep-25-solubility tag constructs. Folding was carried out for 16h in the presence of reduced and oxidized glutathione, followed by an HPLC separation using a reverse-phase column. Asterisks indicate correctly folded species. The chromatogram showed a peak at ~23 min attributed to folded Hep-25-solubility tag and a broad peak at ~5 min that can be attributed to glutathione/glutathione disulfide (GSH/GSSG) redox system (GSSG/GSH=1/1). A) The yield of folded Hep-25-DEDEDE was 1.3mg per 1L medium (red) B) Folded Hep-25-SSSSS was eluted from the column at about 24 minutes, purified product yield was 0.3mg per 1L medium (green). C) In the case of Hep-25-RRRRR, chromatogram showed that the peptide was not folded (blue).

In the case of the negatively charged DEDEDE tag, the pH of buffers was the same as for the purification of His_6-SUMO-Hep-25. The negatively charged tag shifted the pI of the His_6-SUMO-Hep-25 construct and Hep-25 to the lower pI value of 5.61 and 4.63, respectively, in comparison to the His_6-SUMO-Hep-25 (pI = 6.40). However, purification of this construct revealed that digestion of His_6-SUMO-Hep25-DEDEDE with the ULP1 protease was a critical step that should be additionally optimized to increase a final protein yield. The cleavage yield was 60%, estimated from the SDS gel, in comparison to the total cleavage of His6-SUMO-Hep-25 construct. In both cases, the fusion protein was digested with the ULP1 protease in the presence of 20mM DTT and 2M urea. It was clear from the earlier study (Section 3.4.3) that a high concentration of urea or large aggregation could have been responsible for the low efficiency of ULP1 cleavage. It turned out that a high concentration of urea can affect cleavage. The screening of the cleavage yield in the presence of 1mM, and 2mM urea demonstrated that cleavage tolerated 1mM urea (Figure 3.17).

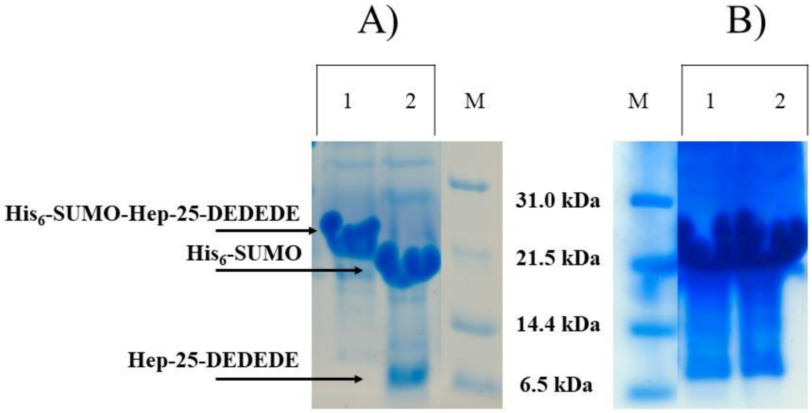


Figure 3.17: Effect of A) 1mM, B) 2mM urea on the yield of ULP1 cleavage of His$_6$-SUMO-Hep-25-DEDEDE. Band M: Broad-Range marker. Band 1: Before, Band 2: After ULP1 cleavage. His$_6$-SUMO-Hep-25-DEDEDE was only cleaved in the presence of 1mM urea.

Later on, complete cleavage of the Hep-25-DEDEDE construct was achieved in the presence of 1mM urea and 20mM DTT. Following the cleavage and size exclusion chromatography, liberated Hep-25-DEDEDE was folded and purified by HPLC under the same conditions as with Hep-25. Solubility plays an important role during protein folding. As the solubility of Hep-25 was increased with the DEDEDE tag, the folding yield was also higher. After all purification steps, the yield of the pure folded peptide was 1.3mg per liter of LB medium, approximately a twofold higher yield than for Hep-25 (Figure 3.16 A).

The presence of the uncharged serine tag did not influence the pI. Therefore, we used the same pH and buffered conditions as for the purification of His$_6$-SUMO-Hep-25. After all purification steps, a total yield of Hep-25-SSSS was 0.3mg per liter of LB medium (Figure 3.16 B). We obtained approximately twofold lower yield than for Hep-25.

In the case of RRRRR tag, the positively charged tag shifted the isoelectric point of His$_6$-SUMO-Hep-25 and Hep-25 to higher pI values, 8.14, and 9.37, respectively. The fusion tag was purified using the same buffer conditions, depending on the purification step, as in the case of His$_6$-SUMO-Hep-25, only adjusted to pH 7.0. The His$_6$-SUMO-Hep-25-RRRRR construct was extracted as a soluble protein and further purified via nickel affinity chromatography. Subsequently, the fusion protein was successfully digested with the ULP1 protease in the presence of 20mM DTT and 2M urea and further purified via size exclusion chromatography. However, the oxidation and purification by HPLC revealed that Hep-25-RRRR was not folded (Figure 3.16 C). The peptide was neither folded in the presence of the higher amount of salts (500mM NaCl) nor at higher pH, which was shifted from pH 7.0 to pH 8.0. We concluded that disulfide bond formation was strongly hindered because of the repulsive electrostatic interaction between positively charged parts of the peptide. It is important to note that Hep-25 is a small peptide, which contains positively charged His-15, Arg-16, and Lys-18 in the loop region (Figure 3.18 A) and His-3 at the N-terminal position. Folding of the peptide could position the RRRRR tag close to the positively charged residues of Hep-25. Therefore, the RRRRR tag may interact repulsively with the loop region (Figure 3.18 A) and/or His-3 (Figure 3.18 B) and disable proper formation of SS-bods and peptide folding.

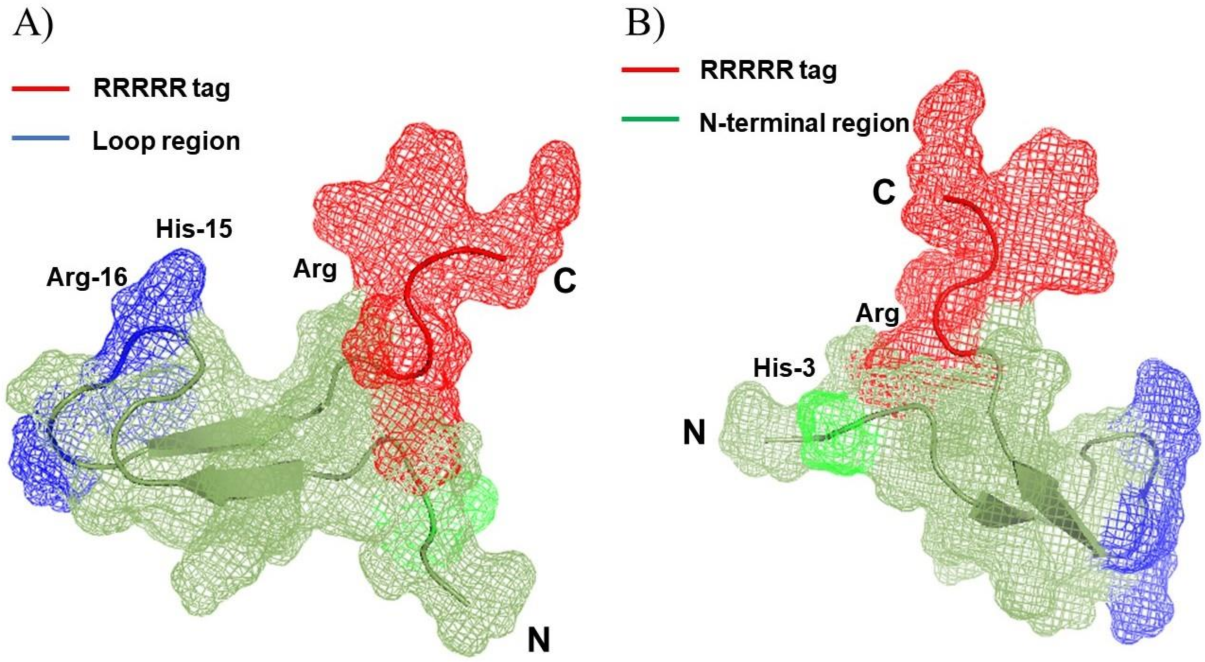


Figure 3.18: Repulsive electrostatic interaction between positively charged parts of the peptide and RRRRR tag. RRRRR tag, red color, may interact repulsively with A) loop region (His-15, Arg-16 and Lys-18), blue color, and/or B) His-3, green, and disable proper formation of SS-bods and peptide folding.

3.5 Disulfide bond characterization

3.5.1 ESI MS characterization of hepcidin constructs

Electrospray ionization (ESI) mass spectra of samples were taken after each HPLC purification. In the case of Hep-25, one single folding product was identified in several purifications, heaving the mass of folded Hep-25, 2787.0415 Da. Compared to the fully reduced, linear, precursor peptide, the mass difference of 8 Da after folding was represented by the loss of eight hydrogen atoms due to the formation of the four disulfide bridges. Three main ions of folded Hep-25 were observed by ESI-MS at mass-to-charge ratios (m/z) of 930.68 ($[M+3H]^{3+}$), 698.26 ($[M+4H]^{4+}$, and 558.81 ($[M+5H]^{5+}$) (Figure 3.19 A). Hep-25-DEDEDE was also further analyzed through ESI-MS. The appearance of species with an increased mass of 732 Da relative to the mass of the folded Hep-25 proved the presence of DEDEDE solubility tag at the Hep-25 (Figure 3.19 B). We were able to see the mass difference of 8 Da again after folding, corresponding to disulfide formation.

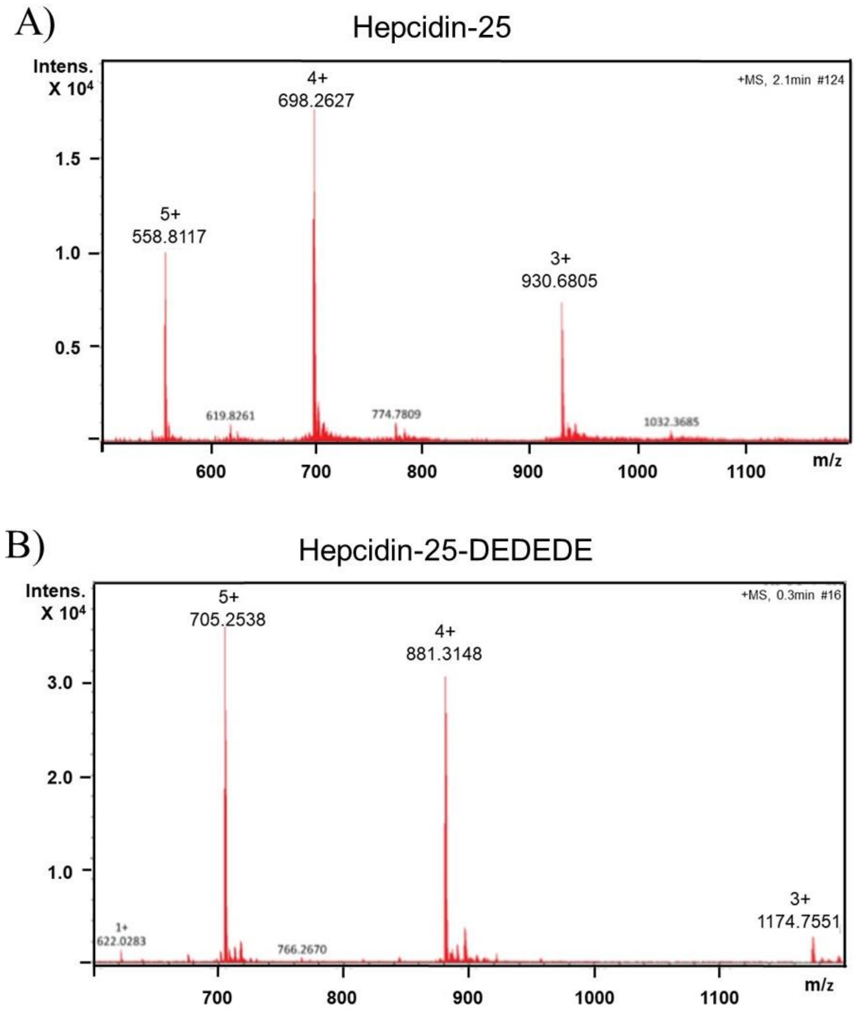


Figure 3.19: Mass of Hep-25 and Hep-25-DEDEDE, as determined by ESI-MS spectrometry, was 2787.0415 Da and 3522.2653 Da, respectively, and the loss of 8 Da is consistent with the formation of four disulfide bonds. Three major mass peaks of A) Hep-25 (M+5 (558.8148), M+4 (698.2620), M+3 (930.6837)) and B) Hep-25-DEDEDE (M+5 (705.2538), M+4 (881.3148), M+3 (1174.7551)) are present in ESI-MS.

These results demonstrated that four disulfide bridges in Hep-25 were also formed in the presence of the solubility tag and that only one single folding product was identified. It should be pointed out that by mass spectrometry, we can distinguish fully, correctly, oxidized from incorrectly oxidized Hepcidin, and in addition, Hep-monomer from partially present Hep-S-S-Hep-dimmers, oligomers. Overall, based on the mass spectra, we proofed that all Hep-25 constructs were folded and contained four disulfide bridges. However, based on mass spectra, we cannot conclude that the correct Cys - SH groups have been paired.

3.5.2 NMR characterization of disulfide bond connectivity

The correct intramolecular disulfide pattern is a critical issue in the analysis of the peptide. Based on mass spectra, we cannot conclude whether the disulphide bond connectivity is correct. The folding and oxidation could be completely non-native. As folding of Hep-25 can be challenging, and misfolding artefacts can occur because of the complex network of disulfide bonds, disulfide bond connectivity was further investigated by NMR to establish the location of bridged residues in the different Hep-25 constructs. As all peptide constructs reflectedhigh solubility in the phosphate buffer and water at pH=3 (Section 3.6.1), disulfide bond connectivity was further investigated at acidic pH. First, it was important to check whether the recombinant produced Hep-25 has the same disulfide pattern as the currently accepted disulfide connectivity assigned by Jordan et al. (Cys-7-Cys-23, Cys-10-Cys-13, Cys-11-Cys-19, and Cys-14-Cys-22) (Figure 3.20) [44].

Figure 3.20: Hep-25 amino acid sequence with the currently accepted disulfide connectivity. Disulfide connectivity is indicated with red lines between following cysteines Cys-7-Cys-23, Cys-10-Cys-13, Cys-11-Cys-19, and Cys-14-Cys-22.

The high similarity of proton chemical shift to those published by Jordan et al. [44] was a good starting point for the investigation into the intramolecular disulfide connectivity. The analyses of 2D TOCSY and ROESY spectra of folded Hep-25, acquired at 325K, proved the same chemical shift assignment and intramolecular disulfide connectivity as assigned by Jordan et al. [44]. We observed in 2D ROESY spectra only two out of four NOEs between the corresponding Hα protons. Disulfide Cys-7-Cys-23 connectivity is defined with NOE (1) and

Cys-11-Cys-19 with NOE (2) between the corresponding Hα protons (Figure 3.21). Further analyses of NOEs from β-proton resonances was complicated as this region is marked by a number of overlaps. Therefore, we were not able to see ROESY cross peaks between the following disulfide bonds: Cys-10-Cys-13 and Cys-14-Cys-22. The NOE interactions between Hα protons of Cys-7-Cys-23 and Cys-11-Cys-19 were also observed by Jordan et al. in 2D NOESY spectra [44]. However, Jordan et al. identified NOEs between Cys10-Cys13 and Cys14-Cys22 in the TOCSY-NOESY experiment [44]. Therefore, the presence of Cys-7-Cys-23 and Cys-11-Cys-19 NOEs proved the presence of currently accepted disulfide connectivity between following cysteine pairs: Cys-7-Cys-23, Cys-10-Cys-13, Cys-11-Cys-19, and Cys-14-Cys-22. ESI-MS and NMR data strongly validate the presence of one single folding product as a result of the oxidation of the linear 25-residue peptide. This result indicates that Hep-25 adapts only one favoured disulfide connectivity.

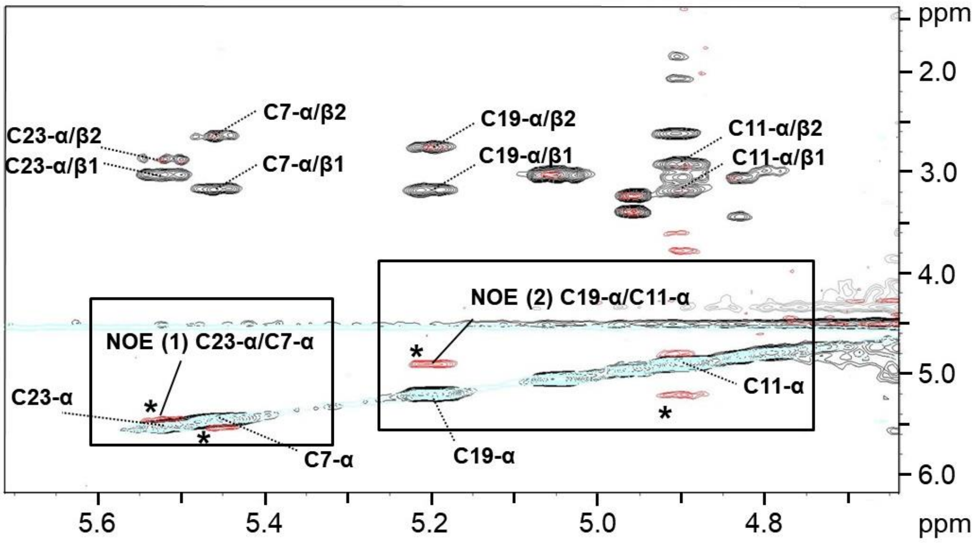


Figure 3.21: Superposition of TOESY spectra (black) with ROESY spectra (red) of Hep-25. The ^{1}H-^{1}H ROESY spectra of Hep-25 was recorded at 325K with a mixing time of 200ms. Resonance assignment of cysteine residues is labelled with dashed lines in the spectrum. We were able to observe ROESY cross peaks between two intra-cysteine out of four. Asterisks indicate cysteine NOEs between the Hα-Hα protons of Cys-7-Cys-23 (NOE 1) and Cys-11-Cys-19 (NOE 2) residues. The presence of these NOEs proved the presence of currently accepted disulfide connectivity between following cysteine pairs: Cys-7-Cys-23, Cys-10-Cys-13, Cys-11-Cys-19, and Cys-14-Cys-22.

We expressed and purified Hep-25 constructs containing C-terminal solubility tags, in order to overcome the low solubility of Hep-25 at neutral pH. All constructs were folded under the same conditions, according to the protocol explained in Section 3.4.4. As the C-terminal solubility tag could affect the folding of Hep-25 and as this would affect the whole 3D structure, we also set out to establish the correct disulfide patter for these constructs. Following resonance assignment and determining disulfide-bond connectivity's of Hep-25 at pH 3, we were able to investigate the effect of the DEDEDE and SSSSS tag on the Hep-25 structure. However, due to limited sensitivity/peptide concentration, we were not able to obtain high-quality NMR data on the Hep-25-SSSSS construct. Figure 3.22 illustrates the only effect of the DEDEDE tag on the Hep-25 structure.

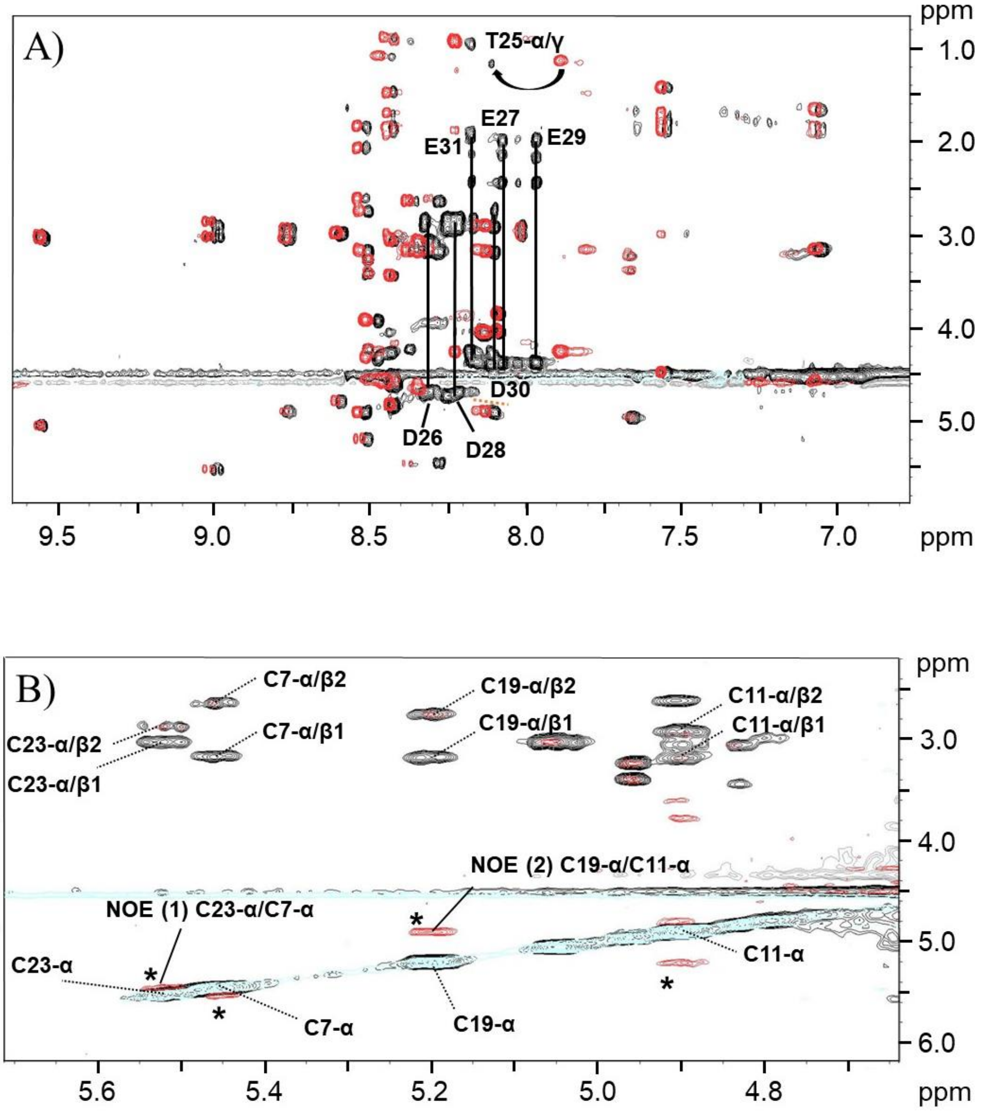


Figure 3.22: A) Superposition of ^{1}H-^{1}H TOCSY spectra of the Hep-25-DEDEDE (black) and Hep-25 (red). All signals of Hep-25-DEDEDE are comparable with signals of Hep-25 with small chemical shift perturbation in the amide region. Thr-25 was most affected residue due to its direct connection to the DEDEDE tag. Additional signals visible in the spectrum and label with dashed lines belong to Asp and Glu from DEDEDE tag. B) Superposition of ROESY spectra of Hep-25-DEDEDE tag (red) and TOCSY of Hep-25-DEDEDE (black) for verification of correct disulfide bond connectivity. The comparison

revealed the presence of same inter-cysteine NOEs observed for Hep-25. Asterisks indicate inter--cysteine NOEs between the Hα-Hα protons of Cys-7-Cys-23 (NOE 1) and Cys-11-Cys-19 (NOE 2) residues. Spectra were recorded at 325K with a mixing time of 200ms and pH 3.

However, in the cases of both constructs, a comparison of 2D TOCSY spectra revealed that the most strongly affected residue is the C-terminal Thr-25 that is directly connected to the solubility tag, which was expected. However, all other signals from Hep-25-solubility tags have similar chemical shifts to Hep-25 (Figure 3.22 A). The small temperature variation in two different NMR measurements could be responsible for these small chemical shift variations. In addition to signals, which belong to Hep-25 in TOCSY spectra, we observed the TOCSY pattern of Asp and Glu from the DEDEDE tag. The 2D ROESY of Hep-25-DEDEDE was compared with the one of Hep-25 to identify NOEs expected across each disulfide bond. In both cases, we were able again to observe two inter-cysteine NOEs, Cys-7-Cys-23 defined with NOE (1) and Cys-11-Cys-19 with NOE (2) between the corresponding Hα protons, at the same position as in case of Hep-25 (Figure 3.22 B). This result demonstrated that all constructs contained the same cross-disulfide NOEs, which reflects the same disulfide connectivity pattern in all cases. It became clear that even in the presence of the C-terminal solubility tag, the oxidation product always has the same disulfide pattern and documents how highly favored this disulfide pattern is. Based on the NMR data, we learned that the structure of peptide was preserved, even in the presence of solubility tags.

3.6 Solubility of hepcidin-25

3.6.1 Solubility of hepcidin-25 in aqueous and aqueous/organic solutions

The solubility of peptides at different pH values strongly depends on their amino acid content and sequence. The Hep-25 sequence comprises of three basic, one acidic, five polar, and 16 hydrophobic amino acids and reflects an amphipathic character (Figure 3.23).

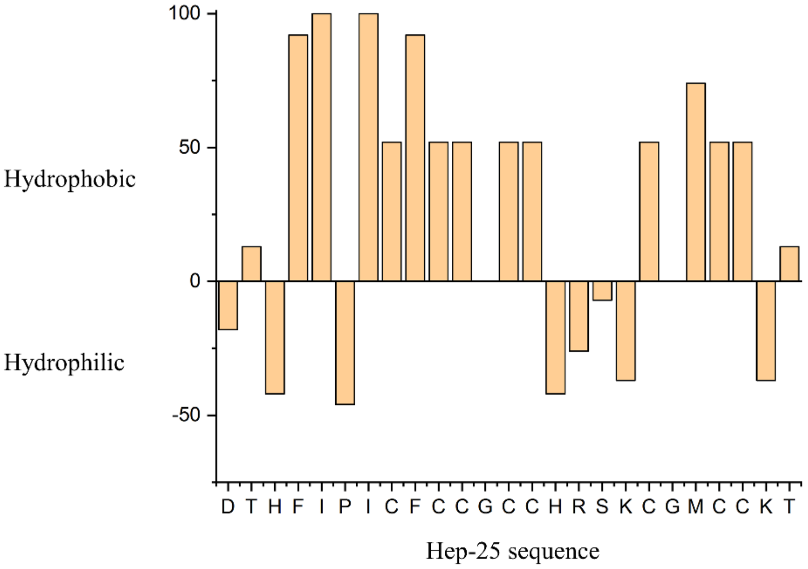


Figure 3.23: Hydrophobicity of Hep-25 at pH 7. The Hep-25 amino acid sequence is composed of approximately 70 % of hydrophobic residues. Values above are normalized so that the most hydrophobic residue has a value of 100 relative to glycine, which is considered neutral (0 value). The scales were extrapolated to residues which are more hydrophilic than glycine [94].

It was found that Hep-25 demonstrated better solubility at acidic pH, which is probably due, to a large extent, to the protonation of His. Thus, acidic pH, at the same time, strongly affects metal ion binding. However, in order to investigate metal binding, the peptide should also be soluble and stable under physiological conditions. We used 50mM phosphate buffer and 100mM NaCl at pH 7.40 as an initial NMR buffer. We were able to record the NMR data but with poor signal-to-noise ratio, as the peptide demonstrated reduced stability in the phosphate buffer and precipitated at pH 7.4. The low solubility of the Hep-25 was attributed to its isoelectric point pI of 8.2. At a pH of 7.5, the peptide had almost no net charge, apparently resulting in a lower solubility. For subsequent NMR experiments, we checked buffer conditions with different pH values that were more distant to the pI. Based on the previous study, it was known that a neutral or basic pH is required for complex stability [52]. Previous publications have also found that Hep-25-Cu(II) is not stable at a pH lower than 5.2 [52]. On the other hand, a pH higher than 8.2 was unsuitable for NMR because at higher pH, the exchange ratio of backbone amide protons is too fast. This leads to signal loss and results in the loss of information. First, the phosphate buffer was tested in the pH range from pH 6.0 to pH 7.0. In each case, however, peptide precipitation could be observed. In the next step, we investigated different buffer conditions at pH 6.0, as at this pH, the peptide should carry at least one positive net charge according to its isoelectric point (Table 3.2).

Figure 3.2: Tested NMR buffer conditions for solubilization of Hep-25

Buffer	Phosphate Buffer						HEPES
C =	50mM	50mM	50mM	50mM	25mM	25mM	50mM
pH	7.4	6.0	7.0	6.0	6.0	6.0	6.0
NaCl	100mM	100mM	100mM	300mM	50mM		
Solubility	–	–	–	–	+	+	–

– peptide precipitated
+ peptide precipitated after few hours

It turned out that Hep-25 precipitated even at pH 6.0, independent of the presence of higher and lower NaCl concentration in phosphate buffer. It also precipitated in the HEPES buffer. Results demonstrated that Hep-25 has poor solubility in neutral and slightly acidic aqueous solutions in contrast to its high solubility in the acidic solutions. However, as soon as the pH was adjusted from pH 6.0 to pH 3, the precipitated peptide was again soluble. The NMR data of the sample collected after the adjusting pH from neutral to acidic revealed that the structure of the peptide was not affected by precipitation and solubilization. Further optimization of the NMR buffer was limited as a great number of additives, such as glycerol and urea, would interfere with proton NMR experiments and/or peptide structure. Although additives would not be a problem for recording spectra, such as a ^{15}N-HSQC, on isotopically labeled samples, a great deal of important information for structural calculation from proton NMR experiments, such as NOESY, would be compromised by strong signals from the buffer. We assume that the low solubility of Hep-25 is not attributed only to the neutral pH, which was close to its isoelectric point, but also to the overall hydrophobic nature of the peptide. The amino acid sequence is composed of approximately 70% of hydrophobic amino acids (Figure 3.23). Therefore, the low solubility at neutral pH could probably be explained by the association/aggregation of Hep-25 that is promoted by intermolecular hydrophobic interactions between peptide molecules.

New buffer conditions have been developed for NMR measurements at neutral pH to compensate for all mentioned disadvantages. The new strategy was focused on the solubility of the peptide in organic solvents, due to better solubility of hydrophobic peptides in organic solvents. Moreover, 20% of ACN was already present and mandatory during Hep-25 folding (Section 3.4.3). Therefore, 20mM phosphate buffer (pH 6) was supplemented with dimethyl sulfoxide (DMSO) or ACN, the most widely used organic solvents in the research of biological systems. The effect of DMSO and ACN on peptide solubility were monitored by NMR (Table 3.3). However, the organic solvent should increase Hep-25's solubility without strongly affecting its 3D structure. Although the 10% of DMSO did not increase the peptide solubility, we did not investigate the higher DMSO concentration. We were focused more on the peptide solubility in 10%, 20%, 30%, 40%, and 50% of ACN in neutral solution. It turned out that, at pH=6, a significantly higher solubility of Hep-25 was achieved at the high content of ACN of 50%. Even under these solvent conditions, the peptide started to precipitate after only one day.

Table 3.3: Solubility of Hep-25 with organic solvents

25mM phosphate-buffer, pH 6.0

Organic solvents	DMSO	ACN				
	10%	10%	20%	30%	40%	50%
Solubility	–	–	–	–	+	++

–	peptide precipitated
+	peptide precipitated after few hours
++	soluble peptide (limited to approximately one day)

Hep-25 displayed better solubility at pH 6.0 in the presence of ACN in comparison to the same sample without ACN (Figure 3.24). However, we were able to see only a few residues, probably because of the speed of the exchange of amide protons. Although TOCSY and ^{15}N-HSQC spectra demonstrated higher sensitivity at pH=6.0 in the presence of 50% ACN, we did not observe cross-peaks in NOESY spectra. Since the cross-peak intensity of NOEs is directly related to the distances between protons, NOEs present an essential piece for the structure determination of metal bound Hep-25. Before we sought to check Hep-25's solubility at higher percentage of ACN, it was necessary to check the effect of ACN on the peptide structure. Since almost full ^{1}H backbone and side chain resonance assignment were reported in acidic solution (Section 3.7.1) [44], our study focused on investigating the effect of 50% acetonitrile on the structure at pH 3. Acetonitrile was added in small amounts to the peptide sample, and chemical shift changes were monitored. Figure 3.24 documents that Hep-25 experiences significant chemical shift perturbations of backbone as well as side chain resonances its content is increased on ACN to 50%. Based on inter-cysteine NOEs, we could conclude that ACN did not affect the disulfide connectivity. However, we did not further investigate the solubility of Hep-25 in higher ACN percentages because of the large chemical shift perturbations in the presence of 50% of ACN.

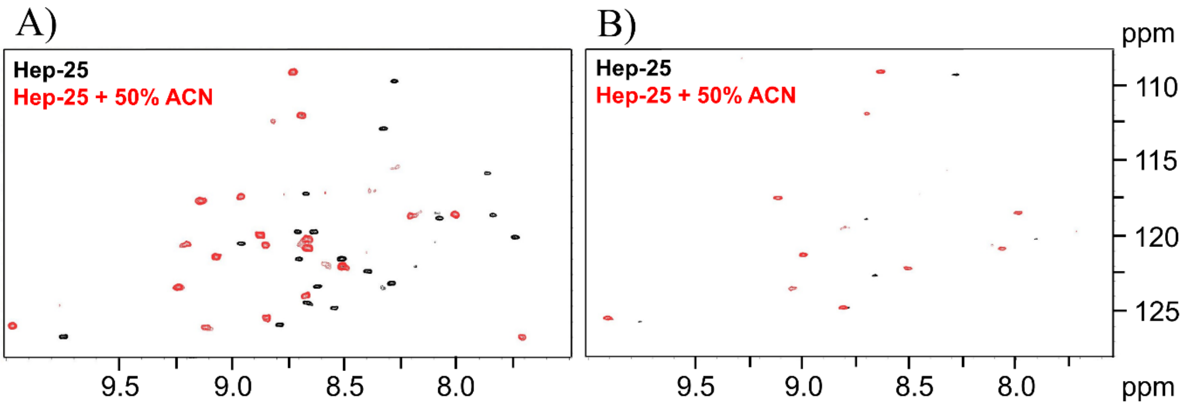


Figure 3.24: Solubility of Hep-25 in the presence of ACN. A) ^{15}N HSQC spectra of ^{15}N-labelled Hep-25 in phosphate buffer (black) and phosphate buffer with 50% ACN (red) at 325K, pH 3.0. Comparison of ^{15}N HSQC spectra showed that large chemical shift perturbation was observed in the presence of 50% of ACN. B) ^{15}N-HSQC spectra of ^{15}N-labelled Hep-25 in phosphate buffer (black) and phosphate buffer with 50% ACN (red) with 600MHz spectrometer, at 325K, pH 6.0. At pH 6.0, Hep-25 showed better solubility in the presence of ACN in comparison to the sample without ACN. Although the solubility was increased, we were able to see only a few residues presumably because of fast exchange of amide protons.

3.6.2 Solubility of hepcidin-25 in the presence of solubility tag

The lack of solubility, even in the presence of 50 % ACN, made metal binding studies with Hep-25 with available instrumentation (600MHz spectrometer, conventional-non-cryogenic-probes) impossible. The new approach was then oriented to increase the peptide solubility with the C-terminal peptide solubility tags. As was explained in Section 3.3.3, the peptide solubility tag should increase the hydrophilicity of the peptide and. at the same time, shift the pI away from neutral pH. Solubility tags were comprised of polar SSSSS, polyanionic DEDEDE, and polycationic RRRRR amino acids. We produced properly folded Hep-25 only in the presence of DEDEDE and SSSSS. Therefore, we investigate the effect of DEDEDE and SSSSS on the peptide solubility. The pH of buffers was adjusted between 6.0 and 7.4, depending on the isoelectric point of constructs. The negatively charged amino acids DEDEDE shifted the isoelectric point of Hep-25 to 4.63. This resulted in low solubility of Hep-25-DEDEDE at pH 6 but led to high solubility of Hep-25-DEDEDE at pH 7.5. Good solubility of Hep-25-DEDEDE in comparison to Hep-25 at pH 7.5 was further supported by analysis of 2D NOESY and TOCSY spectra (Section 3.7). In the case of the Hep-25, we observed low solubility leading to poor spectra intensity. In addition, the Hep-25 construct with the polar but uncharged solubility tag (SSSSS) and the same pI=8.21 as Hep-25 displayed similarly low solubility at pH=7.4, as well as at pH=6.0. Therefore, we focused on the most promising construct Hep-25-DEDEDE in our metal binding studies. The DEDEDE tag probably increased the solubility by introducing repulsive electrostatic interactions and/or by disrupting hydrophobic interactions between peptide molecules. Although Hep-25 had low solubility at neutral pH, we also investigated the metal binding of Hep-25. It was important to exclude the interference of DEDEDE tag in Hep-25-metal coordination (Section 3.7.3).

3.7 Resonance assignment and metal-binding studies

3.7.1 Resonance assignment at acidic pH

Chemical shift perturbation (CSP) experiments are often used to identify a binding site in peptides/proteins [76]. This type of experiment is based on ligand-induced changes in the chemical shifts of the peptide/protein. In order to easily observe the conformational changes upon metal binding, resonance assignments of the peptide were a prerequisite (Section 3.7.3). The resonance assignment of the peptide usually consists of two steps (Section 2.3.1). First, amino acid spin systems were identified through an analysis of bond connectivities in TOCSY. In a second step, in NOESY spectra, we identified neighboring residues in the amino acid sequence from through-space connectivities. In the case of ^{15}N labeled samples, backbone amide groups were identified in 2D ^{15}N-HSQC-TOCSY and ^{15}N-HSQC spectra. As was explained in Section 3.1.2, the complexation involves the free α-NH_2 nitrogen, two backbone nitrogen atoms of the second and third amino acid, and the imidazole nitrogen of a histidine residue in the third position. Therefore, assignment and then monitoring of ^{15}N resonances during metal titration should give detailed insights into the metal coordination. After unlabeled and ^{15}N-labeled peptide samples were prepared using protocols outlined in Sections 3.4.3 and 3.4.4, the resonance assignments of Hep-25 constructs were made firstly at pH 3 and then at neutral pH.

The acidic pH allowed an almost a full sequence-specific assignment, which is not possible at pH 7 due to the exchange of amid-broadening protons. Hep-25 constructs were dissolved in 25mM phosphate buffer (pH 3) and 100mM NaCl. The peptide was measured in H_2O/D_2O (95%/5%) to prevent the loss of signals of amide protons from an exchange with solvent deuterons. At ambient temperature, interpretation of NMR spectra of Hep-25 was complicated by the internal dynamics of the peptide, which was first noticed by Jordan et al. [44]. Therefore, in order to improve spectral quality, we investigated different temperatures in the range from 298K up to 325K. Comparison of ^{1}H NMR spectra demonstrates nearly complete chemical shift changes over the recorded temperature range. The Cys-14 HN resonance was the most affected amino acid. It appears as a broadened signal, at 9.161 ppm, at high temperatures and then completely disappears into the baseline at ambient temperatures; the same effect was noticed by Jordan et al. [44]. Therefore, spectra were acquired at 325 K in order to minimize line broadening caused by intermediate dynamics of the loop region [44]. Sequence-specific assignments were achieved through an analysis of 1D ^{1}H, 2D TOCSY, 2D NOESY, 2D NOESY, ^{1}H-^{15}N HSQC, and 2D ^{15}N HSQC-TOCSY spectra. We were able to transfer 88% of the published resonance assignment of protons at 352 K and for the first time, assigned 96% of backbone ^{15}N resonances (Figure 3.25 A and C). Even though we were able to see the signal of Cys14 at 9.161 ppm in ^{1}H NMR, the signal was not visible in TOCSY spectra. Pro-5 and Asp-1 at the N-terminus are not visible in the amide region. This was expected because Pro does not contain hydrogen on the amide group and the amino group of the N-terminal residue is usually not visible due to exchange with solvent. The side-chain assignment of Asp-1 and Cys-14 was also not possible due to signal overlaps. The sequential $H\alpha$-NH_{i+1} connectivities were revealed for almost the entire peptide chain except between the four amino acids Cys-13, Cys-14, His-15, and Arg-16 within the loop region. The HN/N correlation was archived through analysis of 2D HSQC-TOCSY spectra, where we observed a direct connection between HN/N correlation and, to addition, the TOCSY fingerprint pattern. Due to reasons related to sensitivity, we were not able to observe signals from Cys-13, Cys-14, His-15, and Ser-17 residues, located in the flexible loop region, and Cys-23 residue in HSQC-TOCSY spectra (Figure 3.25 B). However, apart from Cys-14, all other residues were visible in ^{1}H-^{15}N HSQC spectra (Figure 3.25 C).

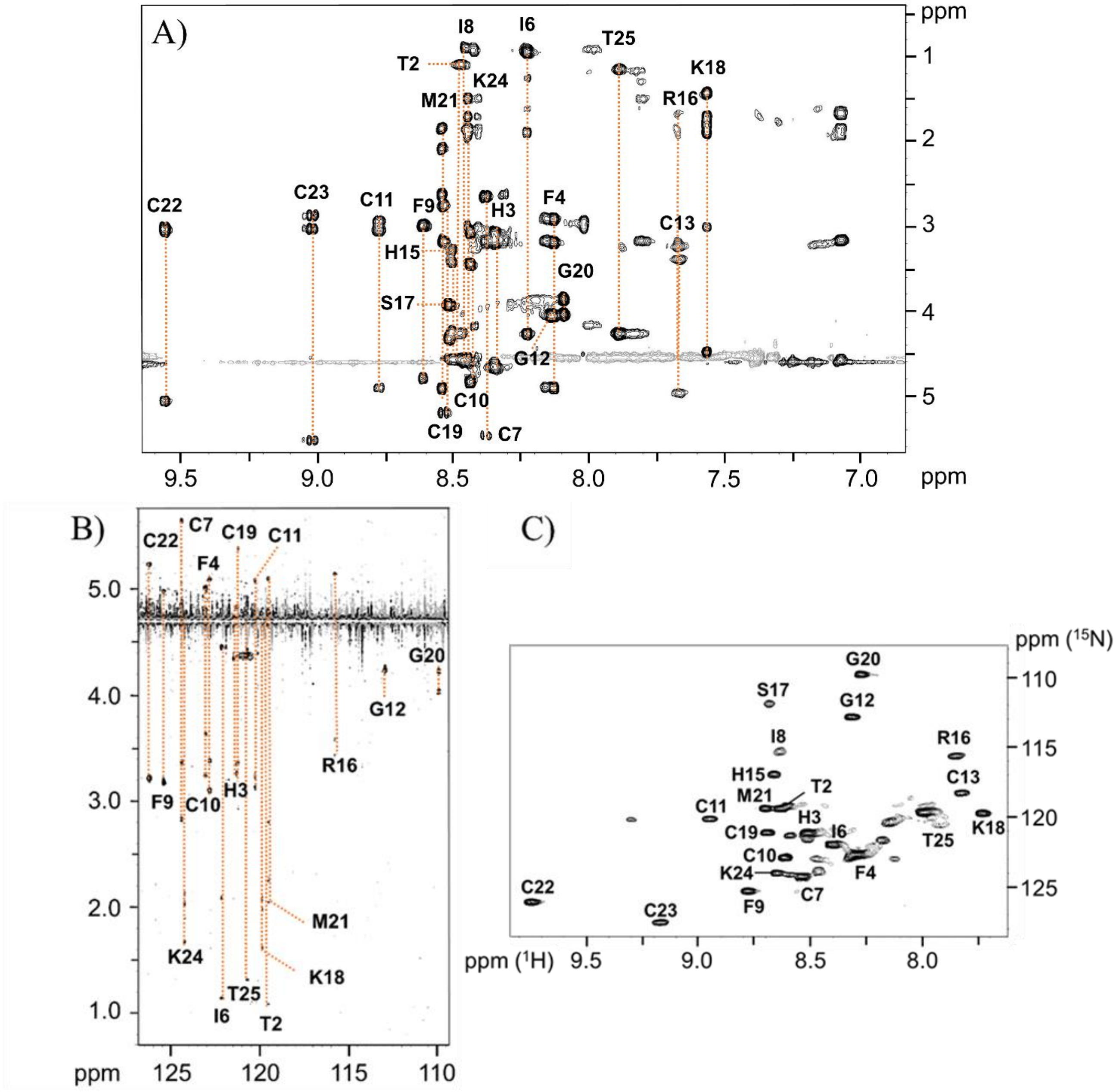


Figure 3.25: Resonance assignment of Hep-25 at 325 K and pH 3.0. (A) The fingerprint region of the TOCSY spectrum of Hep-25. Resonance assignments are indicated with one-letter amino acid codes and residue numbers. The amide proton of Cys-14 was not observed, presumably because of signal broadening. B) ^{15}N-aliphatic ^{1}H correlations observed in the 2D ^{15}N HSQC-TOCSY spectrum. For reasons of sensitivity and due to line broadening Cys-13, Cys-14, His-15, Ser-17, and Cys-23 residues were invisible in this experiment. C) the 2D ^{15}N HSQC spectrum. All expected correlations, apart from Cys-14, could be assigned.

The Hep-25-DEDEDE construct displayed high solubility at pH 7.5 and therefore was considered the most suitable candidate for further metal binding studies. As a result, the resonance assignment of this construct was the following step. It was already mentioned in Section 3.5.2 that superposition of Hep-25 and Hep-25-DEDEDE signals indicates that the structure of the Hep-25 in the presence of the solubility tag was preserved (Figure 3.22). Analyses of TOCSY and NOESY spectra, acquired at pH 3.2 and 325K, revealed the same chemical shift assignment as in the case of Hep-25, with additional signals originating from the DEDEDE tag. For that reason, Figure 3.26 includes labels of only these new signals, which belong to the DEDEDE solubility tag.

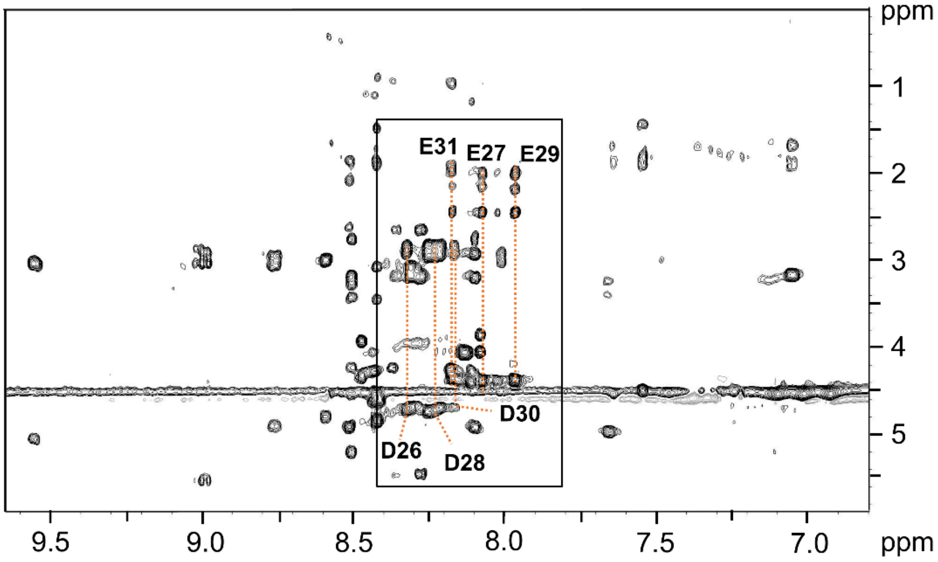


Figure 3.26: Resonance assignment of Hep-25-DEDEDE, at pH 3 and 325K, revealed the same chemical shift assignment as in the case of Hep-25 with additional signals originating from the DEDEDE tag. This is particularly clear from the comparison of TOCSY spectra of Hep-25-DEDEDE with spectra of Hep-25, as Figure 3.22 illustrates. In order to simplify the resonance assignment of Hep-25-DEDEDE, only signals originating to the DEDEDE tag are labeled.

As in the case of the hexapeptide, we were able to detect proline [Pro-5] *cis-trans* isomerization in Hep-25 and Hep-25-DEDEDE through NMR spectroscopy. In accordance with previous reports, due to the slow *cis-trans* proline isomerization with respect to the NMR chemical shift scale, we observed a double set of resonances for almost every nuclear spin affected by the isomerization process [42, 44]. Thus, a double set of signals was observed for the N-terminal residues. All residue between Thr-2 and Cys-7, which are close to Pro-5, were affected by isomerization. *Cis* and *trans* proline isomers in Hep-25 were distinguished by a comparison between the TOCSY spectra of Hep-25 and the hexapeptide (Figure 3.27.). As was determined from peak intensities, the ratio of the *cis/trans* isomers in hepcidin-25 is about 10%/90%. The lower *cis/trans* ratio was probably caused by the reduced flexibility of Hep-25 compared to the hexapeptide. The same *cis/trans* ratio was observed in the TOCSY spectra of Hep-25-DEDEDE. Neither the temperature nor the pH value affects the *cis/trans* isomer ratio as in the case of the hexapeptide. Proline *cis-trans* isomerization is a well-known phenomenon that is frequently observed, especially for proline residues at flexible parts of polypeptides [51].

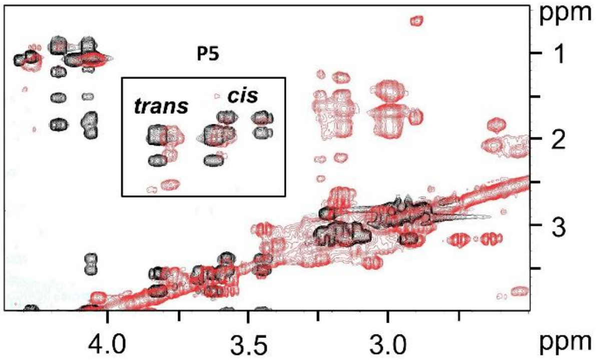


Figure 3.27: *Cis* and *trans* proline isomers in Hep-25. Isomerization ratio in Hep-25 was about 10% to 90% of *cis* to *trans* isomer. The isomerization ratio was determined from peak intensities by comparison of the Hep-25 TOCSY spectra with TOCSY spectra of hexapeptide, acquired at 298K and acidic pH.

3.7.2 Effect of pH on chemical shift and line broadening

Following the resonance assignment at acidic pH, we investigated the effect of pH on the Hep-25-DEDEDE with regard to chemical shift and line broadening predominantly at neutral pH. As a small peptide, Hep-25 easily allows solvent accessibility to the backbone amides under neutral conditions. Adjusting the pH from 3.0 to 7.5 increases the hydrogen exchange rates and, therefore, causes a decrease of the NMR peak intensity and even disappearance of many of the peaks from spectra. The ^{1}H-^{15}N HSQC and TOCSY spectra revealed that the backbone amide nitrogen of almost all regions of the peptide was exposed to the solvent, leading to the disappearance of amide signals. Amide protons of only eight residues, of Hep-25-DEDEDE, are visible in the amide region at pH=7.5. Four residues Phe-9, Cys-10, Gly-20, and Cys-22 from the β-sheet region, are protected from solvent accessibility because they are buried in well-characterized β-sheet region stabilized with disulfide bonds (Figure 3.28 and Figure 3.29) [44]. Extremely slow exchange rates of Phe-9, Cys-10, Gly-20, and Cys-22 amide protons at pH 7.5 were also present at pH 3, indicating that Hep-25 contains well-characterized β-sheet element (Figure 3.28) [44]. Beside these residues, backbone amides of four residues from DEDEDE tag, Asp-26, Glu-29, Asp-30, and Glu-31, are also visible (Figure

3.29). A reason for this could be that negatively charged C-terminal residues are stabilized by electrostatic interaction with positively charged residues in the loop region, i.e., His-15, Arg-16, and Lys-18. Due to line broadening of most amid signals, it was not possible to investigate the effect of the metal binding on backbone amide resonances. Additionally, this made the resonance assignment at neutral pH and structural determination of Hep-25-metal complex problematic. In the aliphatic region of spectra, we found, however, that Hep-25-DEDEDE displayed similar chemical shifts as at pH 3 (Section 3.8.2). This strongly indicates that the 3D structure of the C-terminal part starting at Cys-7 is largely unchanged under these two conditions. The resonance assignment of the aliphatic region of Hep-25-DEDEDE at pH 7.5 was then based on chemical shift comparisons with the one obtained at pH 3. We were able to assign 77% of ^{1}H-resonances of side chains, excluding Cys-14, His-15, and the C-terminal residues Cys-23, Lys-24, Asp-26, Glu-27, Asp-28, and Asp-30. The assignment of these residues was not possible because of the signal overlap in the aliphatic region. The same effect was observed at 298K and 325K. Figure 3.29 illustrates the resonance assignment of Hep-25-DEDEDE at pH 7.5, 325K. It is vital to notice that observed NOE patterns at pH 7.5 were quite similar to those at pH 3.0, further supporting the similarity of the structure at both acidic and neutral pH values. However, we see less NOESY cross signals at neutral pH due to line broadening. The inter-cysteine NOE between the Hα protons of Cys-11 and Cys-19 was the only visible inter-cysteine NOE at pH 7.5 (Figure 3.29 C).

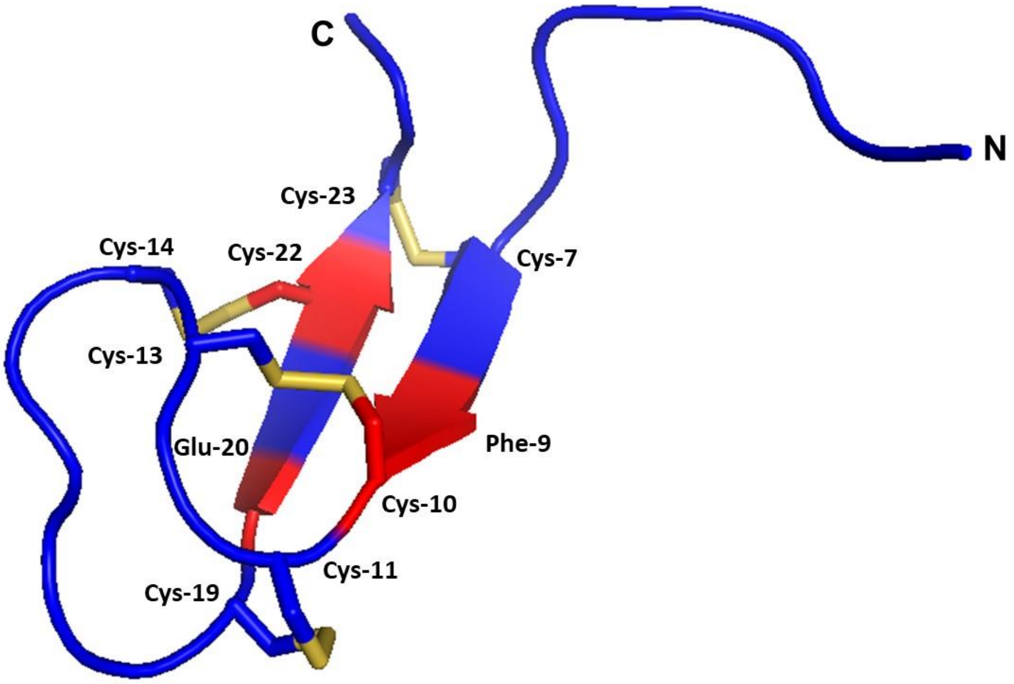


Figure 3.28: 3D structure of Hep-25. Four residues from the β-sheet region, Phe-9, Cys-10, Gly-20, and Cys-22, indicated in red, are protected from solvent accessibility because they are buried in well-

characterized β-sheet region stabilized with disulfide bonds Disulfide bonds between Cys-7-Cys-23, Cys-10-Cys-13, Cys-11-Cys-19 and Cys-14-Cys-22 residues that stabilize β-sheet region are indicated in yellow.

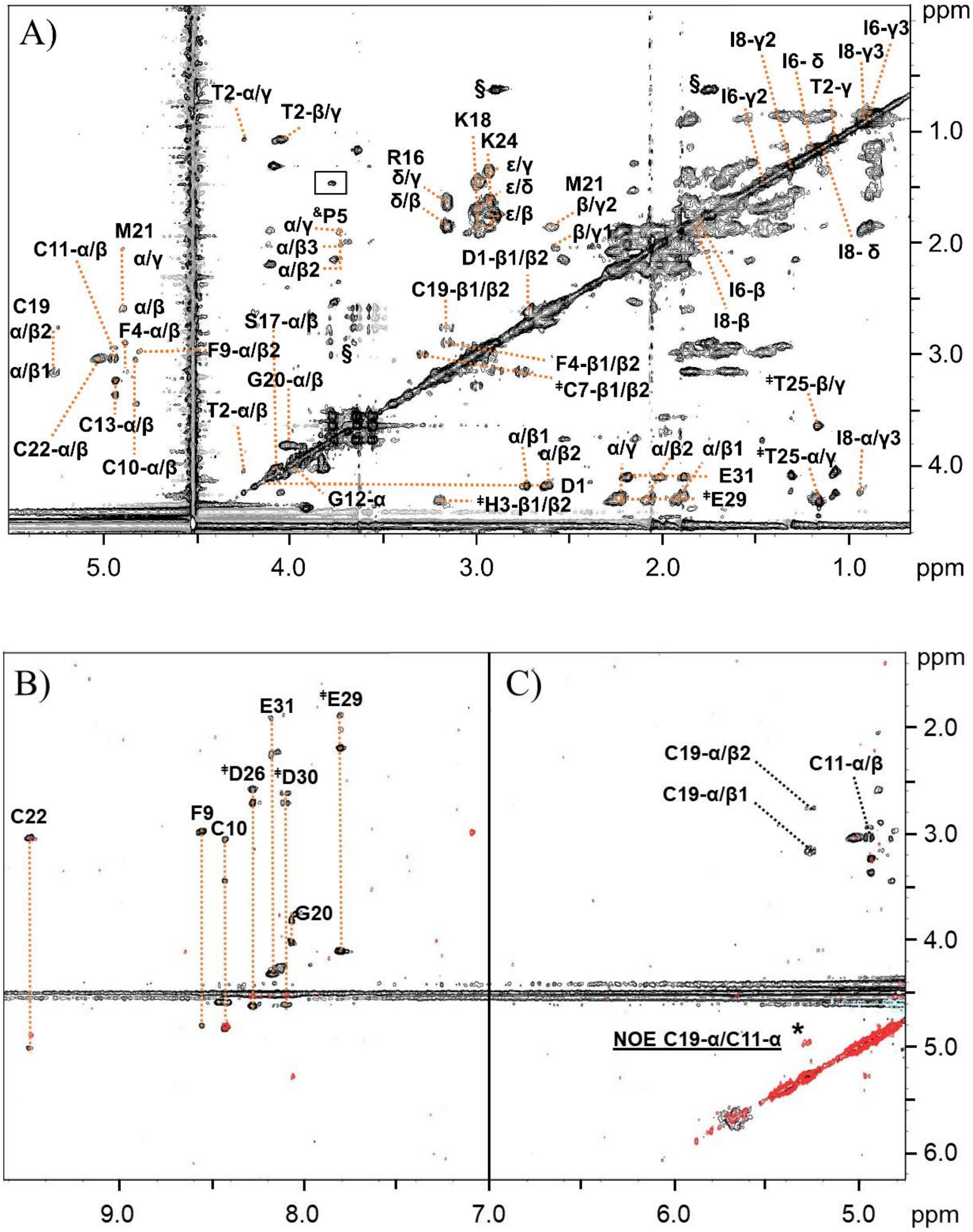
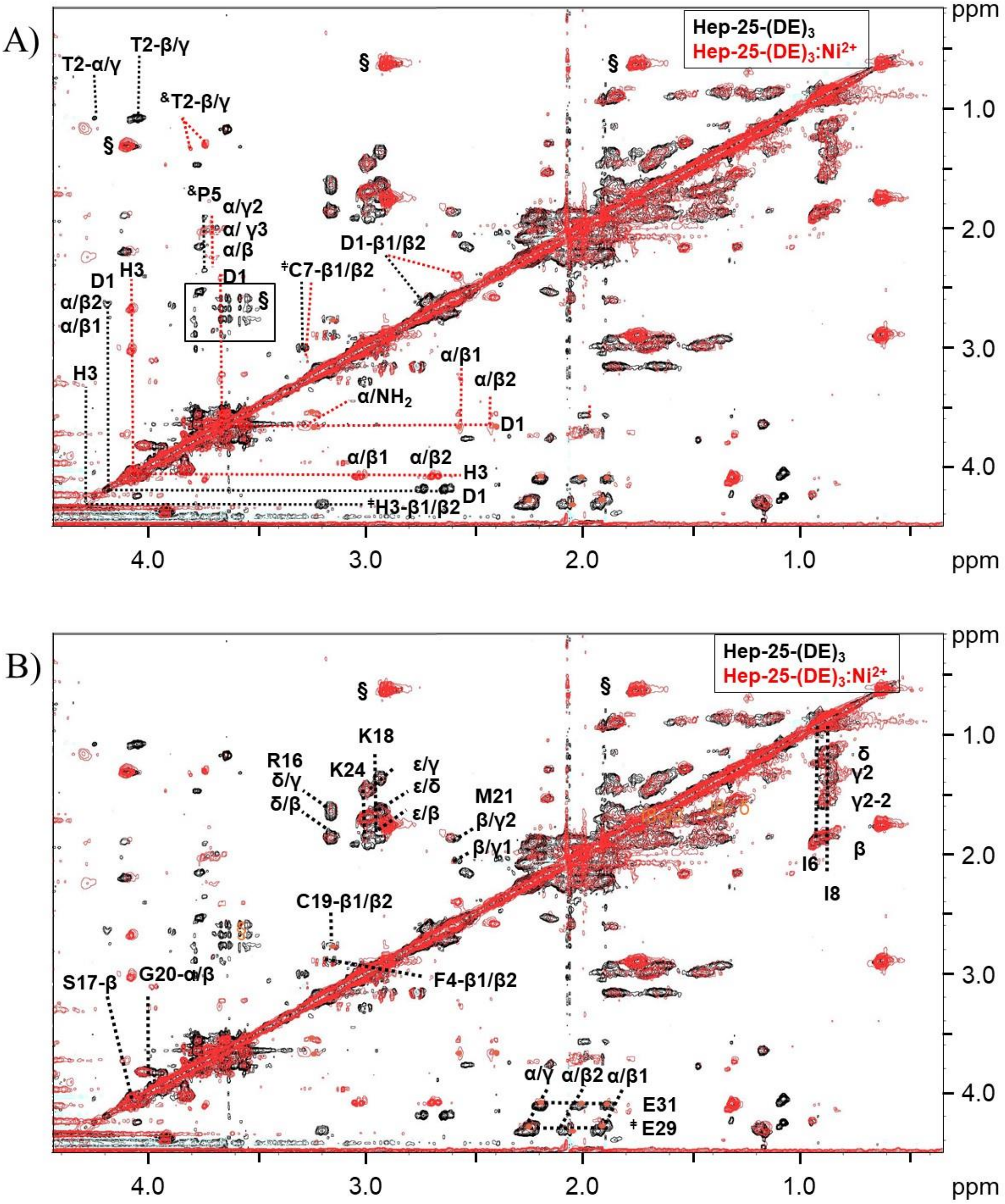


Figure 3.29: The assignment of Hep-25-DEDEDE at pH 7.5 and 325K. Resonance assignments are indicated with one-letter amino acid codes and residue numbers. A) 77% of residues were assigned in the aliphatic region of the TOCSY spectrum. His-15, Cys-14 and C-terminal residues Cys-23, Lys-24, Asp-26, Glu-27, Asp-28, and Asp-30 were not assigned due to signal overlap. ǂ residues that are assigned with uncertainty, & contains a double set of peaks due to cis-trans proline isomerization [44],

§ indicates the presence of impurities. Peaks indicated with the black square are not assigned. B) Superposition of TOCSY (black) and NOESY (red) spectra of Hep-25-DEDEDE. Due to fast amide exchange with the solvent, only eight residues were visible in the fingerprint region of the TOCSY spectrum, indicated by orange dotted lines: Phe-9, Cys-10, Gly-20 and Cys-22 from the well-characterized β-sheet element, and Asp-26, Glu-29, Asp-30 and Glu-31 from the DEDEDE tag. C) Superposition of Hα region of the TOCSY (black) and NOESY (red) spectra of Hep-25-DEDEDE We were able to observe only one inter-cysteine NOE between the Hα protons of Cys-11 andCys-19, as indicated by asterisks.

3.7.3 Metal coordination monitored by NMR spectroscopy

In order to extract valuable information and enable structure determination of the peptide-Ni(II) complex, we investigated the metal binding capacity of Hep-25-DEDEDE as this construct displayed better solubility at pH 7.5 in contrast to Hep-25. The metal was titrated into solutions of the peptide until a 1:1 molar ratio complex was reached. As was already clarified in Section 3.1.2, a high-resolution structure determination of Cu(II)-bound Hep-25 is impossible because of paramagnetic line broadening. Therefore, in the present study, we used Ni^{2+} as a diamagnetic alterative instead of Cu^{2+} to investigate the metal binding. It is known from previous literature that ATCUN motifs bind both metal ions and demonstrate square planar geometries [83, 84]. In order to follow chemical shift perturbations upon Ni^{2+} coordination, 1D ^1H- and 2D TOCSY spectra were acquired after each titration step at 325K. Upon coordination, Ni^{2+} causes chemical shift perturbations as a result of changes in the electronic structure and by stabilizing a specific conformation. The most strongly affected amino acids were those of the ATCUN motif, specifically Asp-1, Thr-2, and His-3 (Figure 3.30 A) [51]. The significant shift of these amino acids indicates that they are involved in metal binding. Besides this, as in case of the hexapeptide, upon complexation, two amino protons of Asp-1 became visible and reflected correlations to Asp-1's α– and β-protons in the TOCSY spectrum. The NMR data clearly demonstrates that chemical shift changes decrease further away from the metal binding site, with effects at Phe-4, Pro-5, Ile-6, and Cys-7. Apart from the mentioned residues, the rest of the signals in the aliphatic region of the TOCSY spectrum show slight chemical shift changes (Figure 3.30 B and C). This suggests that no major conformational changes are triggered by metal coordination.


A)
Hep-25-(DE)3
Hep-25-(DE)3:Ni2+
T2-α/γ
T2-β/γ
&T2-β/γ
§
&P5
α/γ2
α/ γ3
α/β
D1
H3
D1
α/β2
α/β1
‡C7-β1/β2
D1-β1/β2
H3
α/β1
α/β2
α/NH2
D1
α/β1
α/β2
H3
D1
‡H3-β1/β2
ppm
1.0
2.0
3.0
4.0
4.0
3.0
2.0
1.0
ppm
B)
Hep-25-(DE)3
Hep-25-(DE)3:Ni2+
§
K18
R16
δ/γ
δ/β
K24
ε/γ
ε/δ
ε/β
M21
β/γ2
β/γ1
δ
γ2
γ2-2
β
I6
I8
C19-β1/β2
F4-β1/β2
S17-β
G20-α/β
α/γ
α/β2
α/β1
E31
‡ E29
ppm
1.0
2.0
3.0
4.0
4.0
3.0
2.0
1.0
ppm

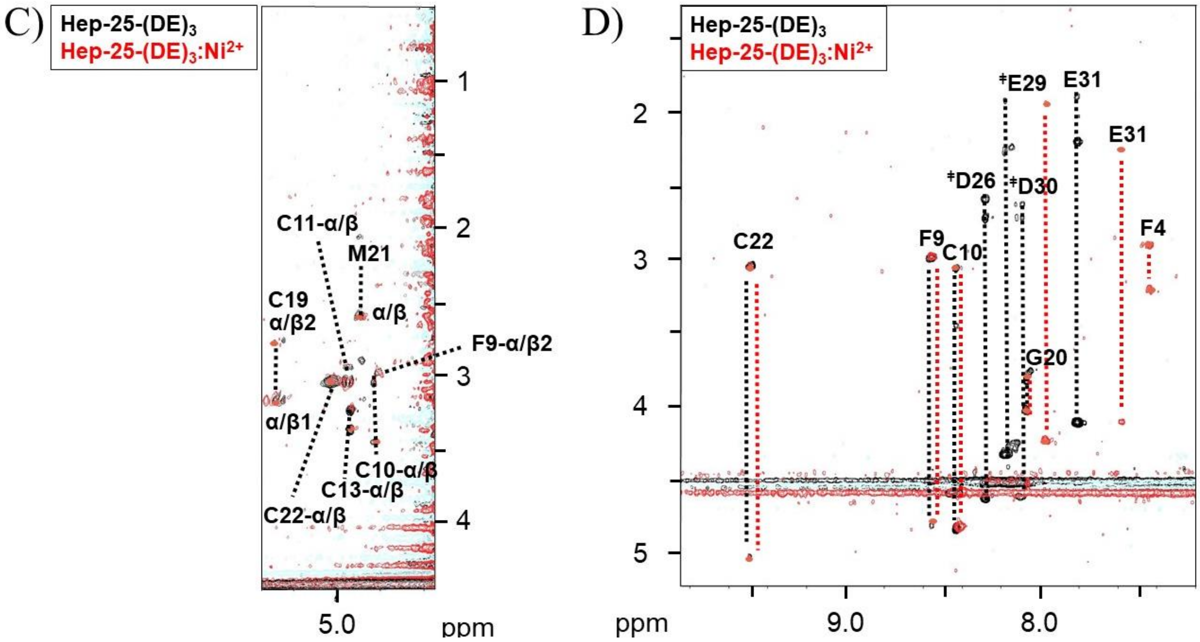


Figure 3.28: ^{1}H-^{1}H TOCSY spectra of the Hep-25-DEDEDE before (black) and upon Ni^{2+} coordination (red) at pH 7.5. Changes induced by Ni^{2+} binding are shown by red dash lines and black dash lines indicate chemical shifts that are not shifted after Ni^{2+} coordination. A) The most affected residues correspond to those of ATCUN motif, Asp-1, Thr-2, and His-3, with small effect at Phe-4, Pro-5, Ile-6, and Cys-7. B) and C) show all visible and unchanged/slightly shifted chemical shifts after Ni^{2+} coordination. D) The fingerprint region of the TOCSY spectrum shows that Phe-9, Cys-10, Gly-20, Cys-22, Glu-27, and Glu-31 are visible in the spectrum after metal coordination. Phe-9, Cys-10, Gly-20, and Cys-22 residues are not shifted after metal coordination in contrast to Glu-29 and Glu-31. § Indicates the presence of the impurities.$^{\&}$ A double set of peaks due to cis-trans proline isomerization [44].

Due to the fast exchange of amide protons at neutral pH and high flexibility of the loop region, the backbone amide protons of Phe-9, Cys-10, Gly-20, Cys-22, Asp-26, Glu-29, Asp-30, and Glu-31 were visible in the fingerprint region of the TOCSY spectrum (Figure 3.30 D). The data clearly demonstrated that Phe-9, Cys-10, Gly-20, and Cys-22 backbone amides are not shifted after Ni^{2+} coordination, supporting the fact that there are no major structural changes upon metal coordination, especially in the well-characterized β-sheet element (Figure 3.30). However, the TOCSY analysis underlines that the backbone amides of Asp-26, Glu-29, Asp-30, and Glu-31 are affected by Ni^{2+} binding. Amide protons of Glu-29 and Glu-31 are slightly shifted, and Asp-26 and Asp-30 are not visible in the spectra (Figure 3.30). The Asp and Glu from DEDEDE tag may bind Ni^{2+} with the carboxylate groups on their side chains. However, chemical shifts of Glu-29 and Glu-31 side chains were slightly shifted by the addition of Ni^{2+}. This indicates that these residues are not or are weakly involved in metal coordination. As we were not able to assign Asp-26, Glu-27, Asp-28, and Asp-30 side chains, our data cannot rule out the possibility of Ni^{2+} binding to these residues. It is important to note that possible Ni^{2+} coordination at C-terminus relates to the uncertain concentration of Hep-25-DEDEDE. It was already mentioned in Section 3.4.3 that the concentration of the sample was roughly determined to measure the absorbance of four disulfide bonds at 280. A disulfide bond has a relatively low extinction coefficient. Due to the uncertain concentration, a small excess of Ni^{2+} that was not bound to the ATCUN motif could be coordinated to C-terminal DEDEDE tag. A comparison between 2D TOCSY spectra of Ni^{2+} bound hexapeptide (DTHFPI) and Hep-25-DEDEDE revealed that spectra look similar, reflecting no significant chemical shift perturbation. In both cases, Asp-1, Thr-2, and His-3 from the ATCUN motif had similar chemical shifts. Therefore, the data indicates that the ATCUN motif has a high affinity for Ni^{2+} and that only an excess of unbound Ni^{2+} might interact with the C-terminal solubility tag. Neither Asp nor Glu from the DEDEDE tag interfere with Ni^{2+} coordination at N-terminus of Hep-25-DEDEDE.

3.8 Structure Calculations

3.8.1 Structure of hexapeptide-nickel complex

For the structure calculation of the DTHFPI-Ni(II) complex, the amino acid sequence, internuclear distances, and the dihedral angles were included as restraints in the Xplor-NIH structure calculation (Table 3.4). Distance information between the protons within the hexapeptide-Ni(II) complex was calculated based on ROESY cross peak volumes, which were corrected for multiplicity and calibrated in reference to the cross peak volumes of Hβ1/Hβ2 of Phe-4 corresponding to a distance of 1.8 Å. For all internuclear distances, we defined upper and lower distance bounds with a margin of (d±1Å) to consider offset-dependence and other distance-uncertainty specific to the ROESY experiment [95, 96]. A total of 36 ROE distance restraints were derived from the ROESY spectrum. In addition to the ROE-derived distances, 6 dihedral angle constraints (ϕ and ψ) were included in the calculation. Backbone dihedral angle (φ, ψ, and ω) data were obtained from ^{1}HN, ^{1}Hα,^{13}Cα, and ^{13}Cβ chemical shifts based on the use of the software TALOS [79]. The correct square planar geometry of Ni^{2+} coordination, known to be present in metal-bound ATCUN motifs, was maintained through an imposition of 14 restraints, including bond distances, bond angles, and dihedral angles [48, 49, 83, 84]. A family of 100 structures was calculated using *ab initio* simulated annealing starting from the hexapeptide (DTHFPI) sequence (Section 2.4). Preliminary structures calculated by simulated annealing were then refined using the simulated annealing refinement. The structure refinement was applied to generate a family of 10 structures, and after five cycles of refinements, there were no further improvements in energy. The final family of 10 structures was analysed to find the low energy structures. Figure 3.31 illustrates the superposition of 10 refined structures of the Ni^{2+}-hexapeptide complex.

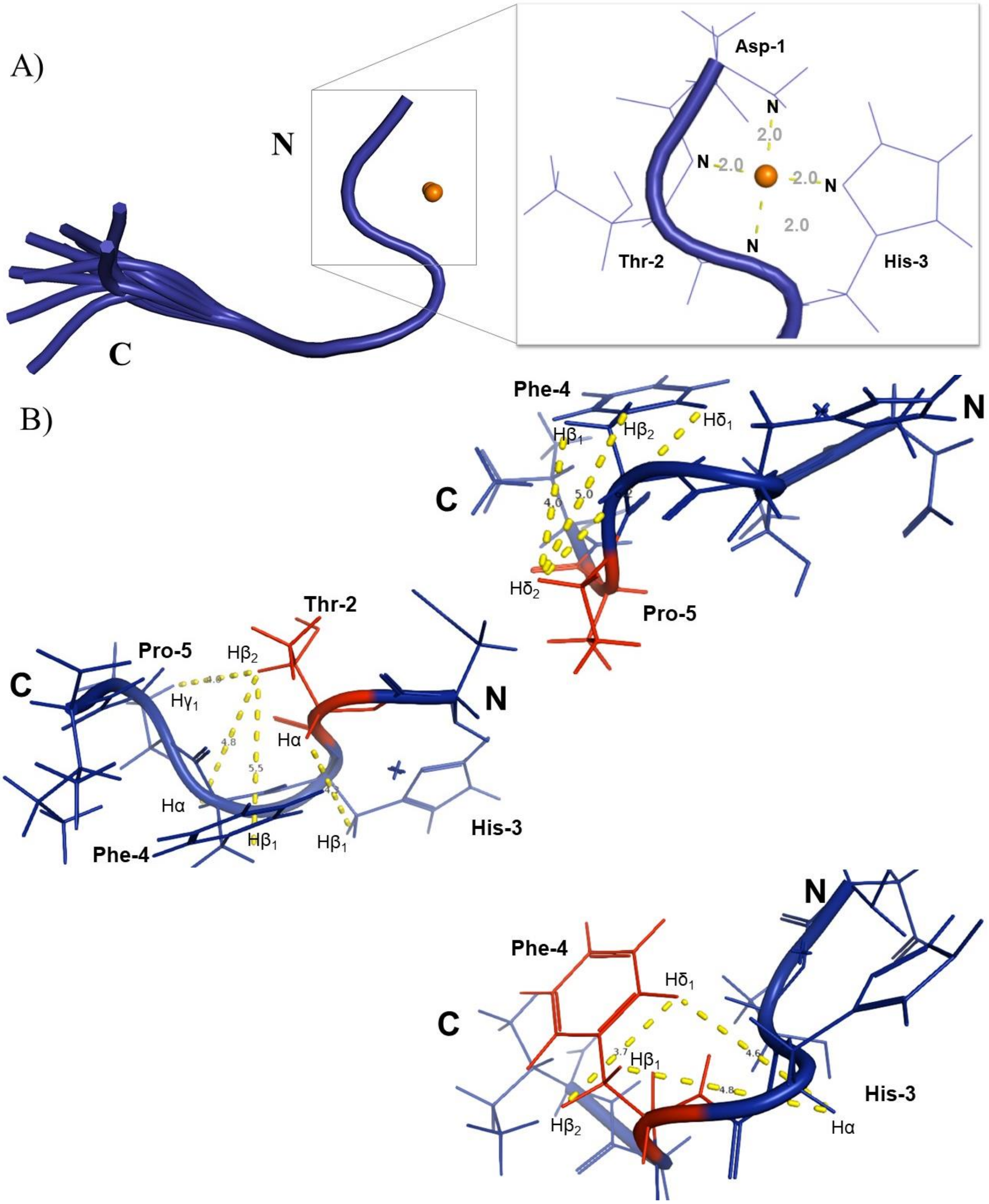


Figure 3.29: A) Superposition of ten refined structures of the Ni^{2+}-hexapeptide complex calculated with

Xplor-NIH. For better clarity, the hydrogen atoms were hidden. The structure of the peptide backbone is very similar among the individual calculated structures, which is evident in the low RMSD (Table 3.4). Metal coordinated amino acids are very well defined by restraints imposing the square-planar coordination and by additional NOEs and dihedral angles. Two C-terminal amino acids show some structural heterogeneity and most likely possess conformational flexibility [51]. B) A total of 36 ROE distance restraints were derived from the ROESY spectrum and included in the structure calculation. For better clarity, only intermolecular distances are shown.

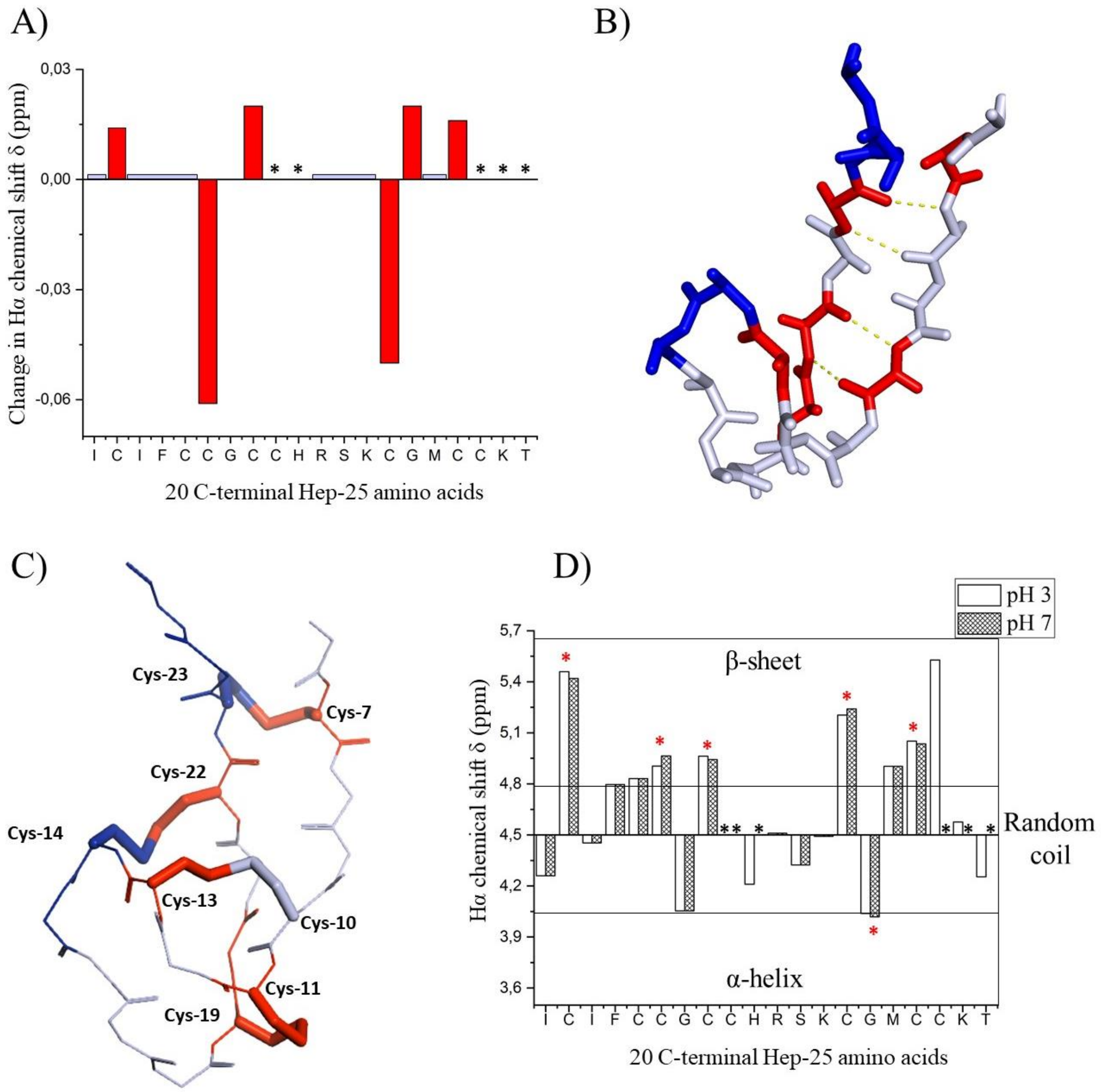
All structures contain well-defined metal bound N-terminal resides, Asp-1, Thr-2, and His-3. The two C-terminal amino acids reflect some structural heterogeneity. The experimental constraints of these residues indicate that this part of the structure is likely to possess conformational flexibility [51]. Overall, all the structure quality features of DTHFPI-Ni(II) complex were evaluated with the use of Protein Structure Validation Server (PSVS) and are summarized in Table 3.4.

Table 3.4. Structure determination statistics of DTHFPI-Ni(II) complex [51]

NMR distances and dihedral angel constrains	
Total number of distance constraints	
Total ROE	36
Total dihedral angle restraints	6
Structure statistics	
RMS deviation	
Distance constraints (Å) violation	0.026±0.001
Dihedral angle constraints (◦) violation	4.328±0.064
Deviation from idealized geometry	
Bond lengths (Å)	0.028
Bond angles (◦)	3.3
RMSD - Backbone heavy atoms(Å)	0.63
Ramachandran Plot Summary from Procheck	
Most favored regions	75.0%
Generously allowed regions	25.0%

3.8.2 Similarity between the C-terminal region at acidic and neutral/slightly basic pH

The structure of Hep-25 contains at high temperature and pH 3 the well-characterized C-terminus and a disordered N-terminus. The C-terminal 20 amino acids build up a stable β-sheet element and a β-hairpin loop, which are stabilized with four disulfide bridges [44, 51]. We investigated the structural similarity of the C-terminal 20 amino acids of Hep-25 at pH 3 and pH 7 (Figure 3.32). A comparison of TOCSY spectra of Hep-25 at pH 3 and pH 7.5 indicated that the aliphatic region of nine amino acids out of 20 reflect the same chemical shifts at both pH values. The effect of pH on Cys-14, His-15, Cys-23, Lys-24, and Thr-25 residues could be assessed as they were not assigned at pH 7.5. However, the six residues demonstrated shift deviations with a total range between 0.02 and 0.06 ppm.



E)

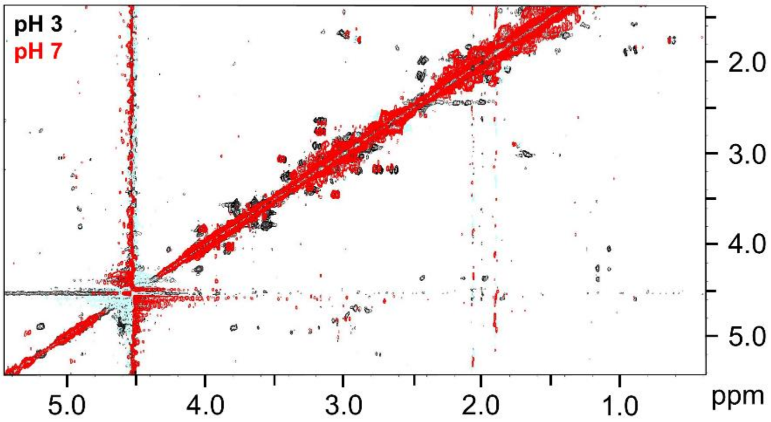


Figure 3.30: Structural comparison of 20 C-terminal residues of Hep-25 at pH 3 and pH 7.5. A) The plot of the Hα chemical shift changes as a function of the residue number of the C-terminal residues. Differences in chemical shifts between Hep-25 at pH 3 and pH 7.5 show that six residues are shifted approximately between 0.02 and 0.06 ppm, indicated with red rows. Residues that are not assigned are indicated with asterisks and not affected residues are indicated with gray bars. B) Structure of the C-terminus of Hep-25 as determined by Jordan et al. at pH 3 [44]. Five cysteine residues (Cys-7, Cys-11, Cys-13, Cys-19, and Cys-22) and Gly-20 that shift upon changing the pH from 3 to 7.5 are indicated with red color. Cys-7 is connected to the flexible N-terminus, Cys-13 is the part of the flexible loop region, and Cys-22 and Glu-20 belong to β-sheet. The shift changes of Cys-22 and Glu-20 residues are most likely connected with the break of the hydrogen bonds between β-sheets, Ile-8/Cys-22 and Cys-10/Glu-20. Light yellow dashed lines indicate the hydrogen bonds between antiparallel β-sheets. Cys-11 and Cys-19 are located in the region between β-sheets and β-hairpin loop. Residues colored gray are not affected, and residues colored blue are not assigned due to signal overlap. C) Disulfide bonds between Cys-7-Cys-23, Cys-10-Cys-13, Cys-11-Cys-19 and Cys-14-Cys-22 residues that stabilize the whole C-terminus, as established by Jordan et al. [44], are indicated with sticks. D) Peptide secondary structure from Hα chemical shifts. The plot shows the Hα chemical shift at pH 3 and pH 7.5 as a function of the residue number of Hep-25. Chemical shifts at pH 3 are indicated with white bars and at pH 7.5 with dashed bars. Chemical shift perturbation of six C-terminal residues, induced with the change of pH, shifted the structure to a more ordered structure. Shifted residues are indicated with red asterisks and not assigned residues with black asterisks. E) Superposition of NOESY spectra of the Hep-25-DEDEDE at pH 3 (black) and pH 7.5 (red). Very similar NOE patterns are observed at both pH values indicating that Hep-25-DEDEDE has the same 3D structure at both acidic and neutral pH values.

As Figure 3.32 illustrates, five cysteine residues (Cys-7, Cys-11, Cys-13, Cys-19, and Cys-22) and Gly-20 shifted after the change of pH from 3 to 7.5. Furthermore, we speculate the possible reasons for the shift of these residues. Cys-7 shifted approximately 0.02 ppm, most likely due to its proximity to flexible N-terminus. We assumed that Glu-20 and Cys-22 from the β-sheet shifted 0.02 ppm due to ionization of this residue, which causes the hydrogen bond between Ile-8/Cys-22 and Cys-10/Glu-20 residues to break. Antiparallel β-sheets are most likely built up by Ile-8, Phe-9, Cys-10, Gly-20, Met-21, and Cys-22 and stabilized via disulfide and hydrogen bonds at acidic pH. However, even with broken hydrogen bonds at neutral pH, β-sheets are stabilized with four disulfide bonds (Figure 3.32 B and C). We assumed that the conformational changes at the β-hairpin loop are observed with the shift of Cys-13 of 0.02 ppm from the loop region. The most affected residues are disulfide connected Cys-11 and Cys-19, which shifted approximately 0.06 ppm. These shifts most likely indicate that Cys-11 and Cys-19 reflect a direct connection between the β-sheet element and the β-hairpin loop (Figure 3.32 C). Therefore, the small conformational changes at the β-sheet and β-hairpin loop probably have a significant impact on these residues.

In order to conclude whether the chemical shift perturbation of six residues (Cys-7, Cys-11, Cys-13, Cys-19, Cys-22, and Gly-20) has a significant impact on the secondary structure, we investigated sequence-specific random coil chemical shifts [97, 98]. We compared Hα chemical shifts assigned at pH 3 with those at pH 7.5 (Figure 3.32 D), as the secondary elements, β-strand, random coil, and α-helix can be estimated from Hα chemical shifts [97, 98]. If these consecutive residues demonstrate secondary chemical shifts of the same secondary structure, that segment is thought to adapt this structure with high probability [97, 98]. It was noticed that the most affected residues Cys-11 and Cys-19 shifted toward the β-sheet secondary structure. Gly-20 occurred at pH 7.5 in the α-helical region. Small shifts of Cys-7, Cys-13, and Cys-22 residues toward random coil region aligned with the position of these three residues at the structure. Cys-7 and Cys-22 are connected to the flexible N- and C-terminal region, respectively. Moreover, Cys-22 made a disulfide bond with Cys-14, which, together with Cys-13, belongs to the β-hairpin loop. Overall, chemical shift perturbation of six C-terminal residues, induced by the change of pH, had a small effect on the secondary structure and apparently shifted the structure to a more stable secondary structure. Moreover, similar NOE patterns, at pH 3 and 7.5, supported the fact that the C-terminal 20 amino acids of Hep-25 does not reflect major conformational changes caused by pH changes (Figure 3.32 E). In addition, as at pH 3, the β-sheet element is characterized at neutral pH by extremely slow exchange rates of Phe-9, Cys-10, Gly-20, and Cys-22 amide protons (Figure 3.30) [44]. All of the data supported the structural similarity of β-sheet element and a β-hairpin loop at pH 3 and pH 7.5.

3.8.3 Model structure of Hep-25-Ni(II) complex

Apart from structural similarity at acidic and neutral pH, the NMR data suggested that the structure of 19 C-terminal residues of Hep-25 reflected no major conformational change triggered by metal coordination (Figure 3.30). Extremely slow exchange rates of Phe-9, Cys-10, Gly-20, and Cys-22 amide protons are evident also after Ni^{2+} coordination and reveal that the β-sheet element conformation is rigid (Figure 3.30). Additionally, the data clearly revealed that Phe-9, Cys-10, Gly-20, and Cys-22 backbone amides did not shift after Ni^{2+} coordination, supporting the fact that there are no major structural changes upon metal coordination, especially in the well-characterized β-sheet element.

The structure of metal-bound Hep-25 was calculated by simulated annealing with r^{-6} averaging of distances with Xplor-NIH. The low number of restraints was determined for 19 C-terminal residues of Hep-25-Ni(II) complex, which disabled a high-quality structure determination. However, 19 C-terminal residues of Hep-25-Ni(II) complex was proposed to be similar to the Hep-25 conformation, as reported by Jordan et al.(see Section 3.8.2) [44]. Therefore, we generated the structure by joining the 3D structure of the Ni^{2+} bound hexapeptide with the 3D structure of the 19 C-terminal amino acids, as determined by Jordan et al. [44]. The 19 C-terminal amino acids were kept fixed by appropriate coordinate restraints during the simulated annealing calculations [51]. As no NOEs were obtained for the DEDEDE tag, the C-terminal appendage was omitted from the Xplor-NIH structural calculation. The NMR data indicates that the tag interferes neither with Hep-25's structure (Section 3.5.2) nor with the metal binding of the ATCUN motif (Section 3.7.3). We applied 16 distance restraints of Hep-25-DEDEDE-Ni(II) complex obtained in NOESY spectra of Hep-25-DEDEDE-Ni(II) complex (Figure 3.33), and 36 distance and 6 dihedral angel restraints were transferred from the hexapeptide-Ni(II) complex. However, six NOE restrains out of 16 determined for Hep-25-DEDEDE-Ni(II) complex corresponded to the same restraints as for hexapeptide-Ni(II) complex with approximately 0,1 – 1 Å distance variations. The N-terminus represents a flexible part of the Hep-25 structure. Therefore, the presence of long-range distances between the N-terminus and the rest of the structure should define the spatial position of the N-terminus. Unfortunately, only one distance restraint defined the spatial position of the N-terminus, as we observed a long distance NOE between Hβ/Hβ protons of His-3 and Cys-11 (Figure 3.33 B).

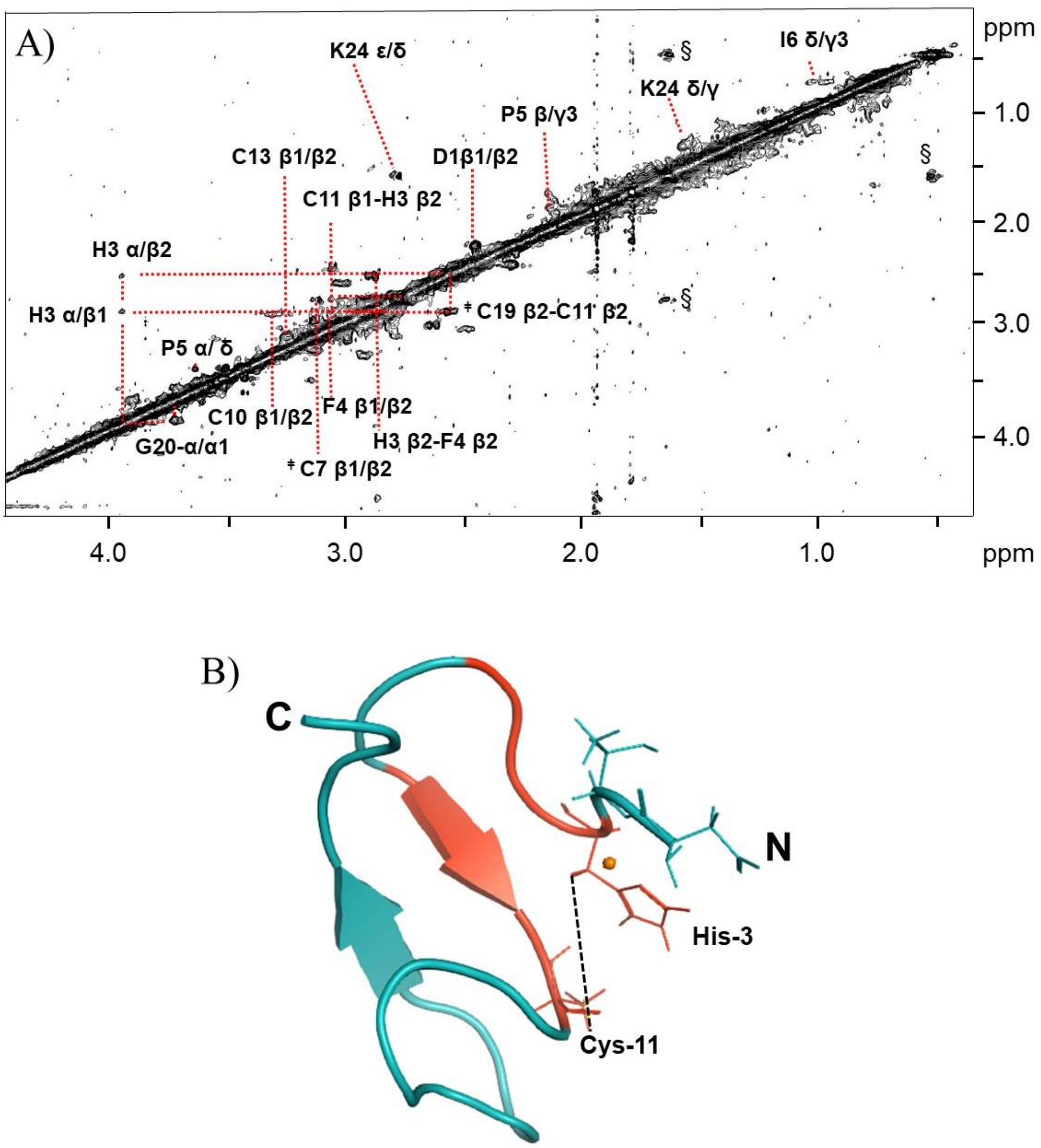


Figure 3.33: A) The NOESY spectrum of the Hep-25-DEDEDE- Ni(II) complex at pH 7 and 325K, mixing time 200 ms. We obtained 16 NOEs of Hep-25-DEDEDE-Ni(II) complex. B) However, only one long distance NOE between Hβ/Hβ protons of His-3 and Cys-11 was visible in the spectrum and defined the spatial position of the N-terminus. Residues indicated with orange color represent the important but missing NOEs.

The square planar coordination was imposed by a restraint of the distances and dihedral angles, as was already described for the hexapeptide-Ni(II) complex (Section 3.8.1) [51]. The structure was generated with Pro-5 in the *trans* configuration, as it was determined based on the NMR data that the *trans* configuration is thermodynamically strongly preferred (~90%) (Section 3.7.1). Ten structures were chosen to represent the ensemble of NMR structures based on the lowest target function and minimal distance and torsional angle restraint violations in the final stage (Figure 3.34). The structure of the Hep-25-Ni(II) complex with the lowest final restraint violation energy was defined as the most suitable structure. The C-terminus was kept fixed by appropriate coordinate restraints during the simulated annealing calculations. The N-terminus is located close to the well-defined β-sheet and β-hairpin loop.

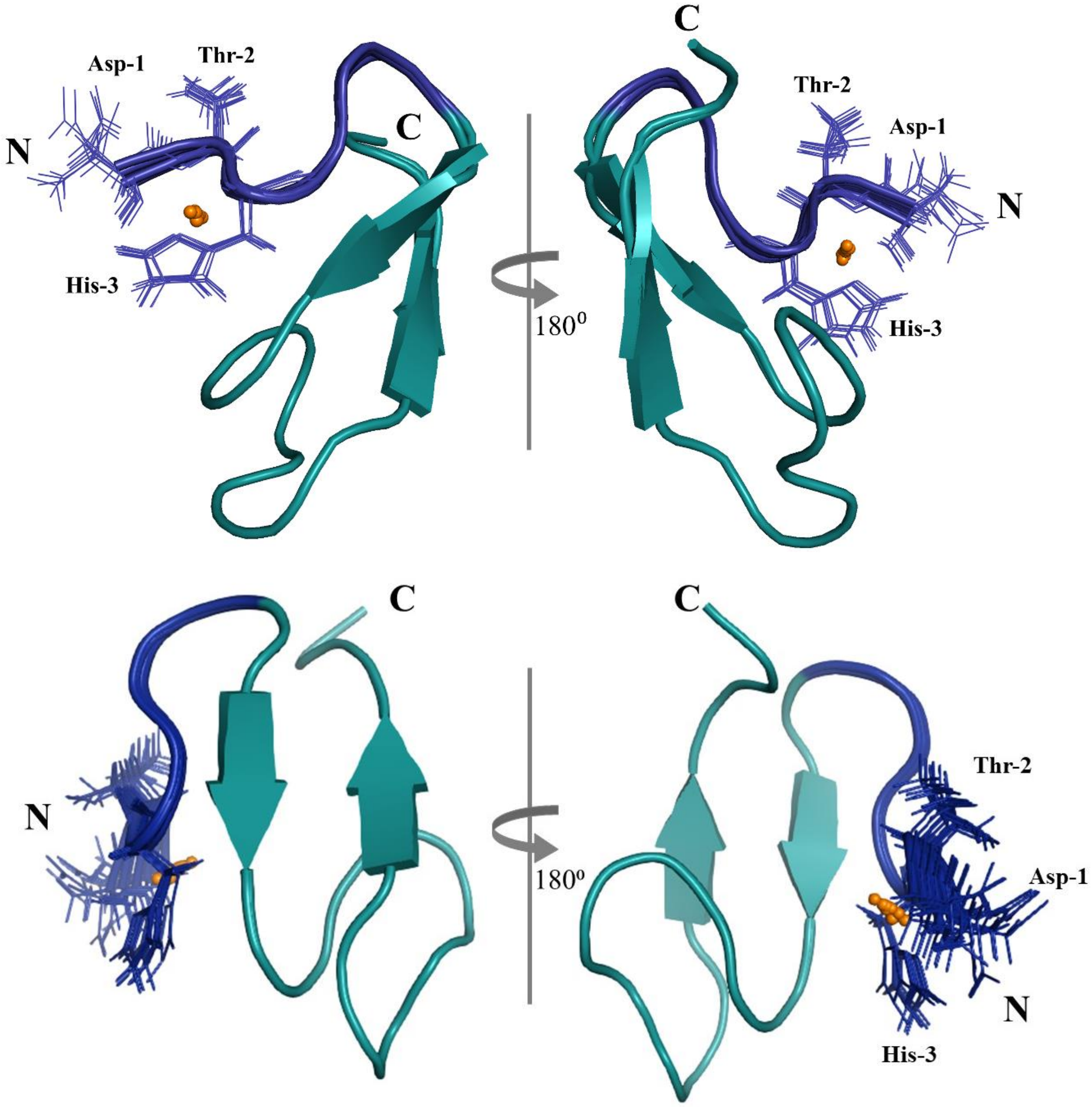


Figure 3.31: Superposition of ten refined model structures of the Hep-25-Ni(II) complex calculated with Xplor-NIH. For better clarity, the hydrogen atoms were hidden. The structure of the peptide backbone is very similar among the individual structures, predominantly of the C-terminus, as the C-terminus was kept fixed by appropriate coordinate restraints during the simulated annealing calculations. The fixed region of Hep-25 is colored turquoise. Metal coordinated N-terminal amino acids are very well defined by restraints imposing the square-planar coordination and by additional NOEs and dihedral angles. Residues represented with lines are involved in Ni^{2+} coordination. The N-terminus is close to the well-defined C-terminal β-sheets and β-hairpin loop.

3.9 Conclusion

The Hepcidin pathway plays an essential role in the control of iron homeostasis [32]. Since the dysfunction of the hepcidin pathway leads to multiple diseases as a result of iron imbalance, hepcidin represents a potential target for the diagnosis and treatment of iron disorders. This thesis focuses on the structural characterization of the biologically active Hep-25, evaluating the hypothesis that the native Hep-25 may contain a Cu2+ ion. This study used NMR spectroscopy as an approach to investigate the metal binding of the ATCUN motif present on the N-terminus of Hep-25.

First, the solution structure of the N-terminal hexapeptide, DTHFPI, complexing Ni2+ was determined. The internuclear distance restraints and dihedral angle restraints were incorporated into the structure calculation with Xplor-NIH. Additionally, the square-planar coordination, known to be present in metal-bound ATCUN motifs, was imposed by additional distance and dihedral angle restraints.

Secondly, the metal-hepcidin complex was investigated in detail through NMR spectroscopy. Since the purchased linear Hep-25 was folded in poor yield, molecular constructs were designed. The protocol for expression and purification of recombinant fusion Hep-25 in high quantity was optimized. Recent reviews reported that heterologous production of recombinant Hep-25 in E. coli largely results in the accumulation of the protein in inclusion bodies [44, 110, 111, 112]. Here, we established an efficient purification protocol for the production of Hep-25 in soluble form in E. coli. After the oxidation and folding of Hep-25, only one folding product was observed and contained the same intramolecular disulfide connectivity as those assigned by Jordan et al. (Cys-7-Cys-23, Cys-10-Cys-13, Cys-11-Cys-19, and Cys-14-Cys-22) [44].

The results of the present study also emphasize that the peptide reflects low solubility in different solvents and buffers at neutral pH due to its hydrophobic property. This could also be the reason that the 3D structure of Hep-25 was not earlier investigated at neutral pH. This result is important for future work, as the high solubility of the sample at neutral pH was mandatory for metal binding studies. The solubility of Hep-25 was increased with the introduction of the polar DEDEDE C-terminal solubility tag.

The metal binding of Hep-25-DEDEDE was investigated. As expected, the most affected residues belonged to the ATCUN motif, which consists of Asp1, Thr-2, and His-3. Chemical shift changes decreased further away from the metal binding site, with a small effect observed at Phe-4, Pro-5, Ile-6, and Cys-7. The rest of the structure of Hep-25 displayed no major conformational change triggered by metal coordination. However, backbone amides of C- terminal residues Asp-26, Glu-29, Asp-30, and Glu-31 from the DEDEDE tag also shifted upon Ni^{2+} addition. Therefore, it can be speculated that the nickel ion binds to side chains of DEDEDE tag due to the presence of oxygen donor atom. Previous studies have reported that the Asp and/or Glu could be involved in the nickel and copper binding, as at pH values above 4.0, the oxygen donor atom of the carboxylic acid group is available for binding [91, 92]. In addition, Laussac et al. employed NMR spectroscopy to study the ATCUN motif of BSA and suggested that the carboxylate group of the aspartate residue in the Asp-Ala-His motif may interact with the metal ion, forming a penta-coordinated structure [47]. It is important to note that the possible Ni^{2+} coordination at C-terminus relates to the uncertain concentration of Hep-25-DEDEDE, as it was proven that DEDEDE tag does not interfere with the ATCUN motif. A small excess of Ni^{2+} that was not bound to the ATCUN motif can be coordinated to the C-terminal DEDEDE tag. Abbas et al. quantified a species corresponding to Hep-25 complexed with two Cu^{2+} ions at pH 11 by MS [51], indicating that at high pH and in the presence of an excess of metal ion Hep-25 has two binding positions. Even though we investigated metal binding at neutral pH, the presence of the extra oxygen donor atoms of Asp and Glu residues and excess of Ni^{2+} ions clearly support our hypothesis that excess of Ni^{2+} can be coordinated to the C-terminal DEDEDE tag of the Hep-25-DEDEDE.

In the final stage, the NMR data recorded for the N-terminus (DTHFPI) complexed with Ni^{2+} and Hep-25-DEDEDE-Ni(II) complex allowed for the construction of a 3D model of Hep-25-Ni(II) complex. Due to high backbone amide exchange with solvent and a low number of NMR data restraints obtained at pH 7.5, the structure determination of Hep-25-Ni(II) complex was difficult. The NMR study of Hep-25 at pH 3 and pH 7.5 indicated that few facts speak in favor of similar structures at both pH. Similar chemical shifts of residues in the aliphatic region and similar observed NOE patterns at pH 7.5 to those at pH 3.0 support the similarity between the structures. In particular, 19 C-terminal amino acids are well-characterized and consist of antiparallel β-sheets and β-hairpin loop stabilized with four disulfide bonds. The similarity of these regions at both pH support the model structure of Hep-25-Ni(II) complex.

In the future, the 3D model structure of metal-bound Hep-25 allows for further research to elucidate hepcidin's interaction with its receptor ferroportin so that researchers can learn more about the signaling mechanism of hepcidin in the iron homeostasis. The interaction between hepcidin and ferroportin could be investigated through STD NMR [113] or based on a measurement of transferred NOEs [113] if the hepcidin - ferroportin interaction is excessively strong or weak for STD NMR experiment. However, the flexible nature of Hep-25 and low solubility at neutral pH may be obstacles for interaction studies. Even though Hep-DEDEDE is soluble at neutral pH, the further optimization of construct and solubility should be considered, as DEDEDE tag may interfere with the interaction of ferroportin and hepcidin. Moreover, the structure should serve as a starting point for the development of antibodies with improved selectivity. In this case, the DEDEDE tag may interfere with the development of antibodies, which can lead to misinterpreted data.

4 Experimental part

4.1 Chemicals and enzymes

Unless otherwise noted, all chemicals used during this study are summarized in Table 4.1. Compounds used throughout this work are presented with their corresponding grade of purity and supplier number.

Table 4.1: List of chemicals

Chemical	Purity	Supplier number
	Acros Organics	
Magnesium chloride hexahydrate	99%	197530010
Bromophenol blue	N/A	151350050
	AppliChem	
Agar	Biological gradient	A0949
	Bio-Rad	
Broad Range standards	N/A	1610317
Natural Polypeptide standards	N/A	1610326
	Cambridge Isotope	
Celtone base powder	98%	CGM-1030P
	Carl Roth	
Carbenicillin	≥88%	6344
Kanamycin	≥88%	T832
Dimethyl sulfoxide	99.5%	A994
Tryptone	Biological grade	8952.2
Yeast extract	Biological grade	2904
Sodium chloride	≥99.5%	3957.1
Sodium dihydrogen phosphate	≥99%	T879
Potassium chloride	≥99.5%	6781
Magnesium sulfate heptahydrate	≥99%	T888
Glucose monohydrate	≥99.5%	6780
Sodium hydroxide	≥99%	6771
Glycerol 86%	≥98%	4043
Urea	≥99.5%	X999
TEMED	99%	2367.3
Ammonium peroxodisulfate	≥98%	9592

Acrylamide	≥98%	7871.2
N, N-Methylene bisacrylamide	≥98%	7867
Roti®-Mark standard	N/A	X879
EDTA	≥99%	8043
SDS	≥99%	CN30
TRIS	≥99%	AE15
Tricin	≥99%	6977
HEPES	≥99.5%	HN78
Methanol	≥99%	8388
Phosphoric acid	Extra pure	9079
Ammonium sulfate	≥99.5%	3746
Coomassie brilliant blue	N/A	9598
Guanidinium chloride	≥99.5%	0037
Glutathione (reduced)	≥98%	6378
Sucrose	≥99.5%	4621
Nickel sulfate hexahydrate	≥99%	T111
	Euriso-top	
Deuterium oxide	99,90%	D216L
	New England Biolabs	
CutSmart Buffer	≥98%	B7204S
DpnI	≥98%	R0176S
T4 Ligase	≥98%	M0202S
SapI	≥98%	R0569S
10X Ligase Buffer	≥98%	B0202S
	Peptides&Elephants	
DTHFPI	99.9%	lot 1211N01
	Riedel-de Haen	
Calcium chloride dihydrate	≥99%	31307
Potassium dihydrogen phosphate	≥99%	30407
	Sigma-Aldrich	
Ammonium-^{15}N chloride	≥99%	299251
100x BME Vitamins	N/A	B6891
SIGMA*FAST*™ protease inhibitors	N/A	S8830
DNase I	N/A	DN25
Trifluoroacetic acid	N/A	1.08178
Glutathione (oxidized)	≥98%	G4251
Sodium 3-(Trimethylsilyl)-1-propansulfonate	97%	178837
Copper sulfate pentahydrate	LC-MS grade	31293
	VWR	
Acetonitrile	LC-MS grade	20048.320
1,4-Dithiothreitol	≥99%	A1101.0025

4.2 Site-directed mutagenesis

In order to generate the cloning plasmid for FX-cloning and mutagenize the target protein, we needed to modify and amplify the DNA sequence. Site-directed mutagenesis can be used to create specific, targeted, changes in double stranded plasmid DNA. Site-directed mutagenesis allows insertion of new DNA sequence into plasmid DNA using specifically designed forward and reverse primers. Therefore, the first step involved a design of primers. All primers were designed according to the method explained in the section 2.2.3. The PCR was run using either the pET15b His$_6$-SUMO-Hep-25 or pMC plasmid (Thermo Fisher Scientific) as a template DNA, depending on the purpose. In the case of His$_6$-SUMO-Hep-25, the DNA sequence was purchased from GenScript. The DNA sequence was codon-optimized, synthesized and subcloned into pET15b plasmids. The His$_6$-SUMO Hepcidin-25 vector was used as a template for insertion of three different gene sequences, which encode the solubility tags, DEDEDE, SSSSS and RRRRR tags. Therefore, we designed a pair of primers for each insertion (Table 4.2.). The solubility tag was inserted into the C-terminus of His$_6$-SUMO-Hep-25. In the case of pMC plasmid, we designed a pair of primers for insertion of the gene encoding the Hep-25 into the pDelivery vector (Table 4.2.), which was further used for Fx-cloning (Section 4.3.).

Table 4.2: Forward and Reversed primers

	Hep-25 insertion - pDeliveri plasmid - T_{ann}= 66°C
Fw	ctgccatcgtagcaagtgcggcatgtgctgcaagacctaaGGGTGAAGAGCGGTACCT
Rev	cagccgcagcagaaaatgcaaatcgggaagtgggtatcCATAGAAGAGCGAGCTCTGG
	DEDEDE insertion - His$_6$-SUMO-Hepcidin-25 plasmid - T_{ann}= 67°C
Fw	gaggatgaaTAACTCGAGGATCCGGCTG
Rev	atcttcatcGGTCTTGCAGCACATGCC
	RRRRR insertion - His$_6$-SUMO-Hepcidin-25 plasmid - T_{ann}= 67°C
Fw	cgccgtcgCTAACTCGAGGATCCGGCTG
Rev	gcgacgGGTCTTGCAGCACATGCC

SSSSS insertion - His$_6$-SUMO-Hepcidin-25 plasmid - T_{ann}= 68°C	
Fw	agctcgtcgTAACTCGAGGATCCGGCTG
Rev	cgagctGGTCTTGCAGCACATGCC
New DNA sequence 🟥	Template DNA 🟩

Each primer was created according the protocol explained in the method section 2.2.3. For each reversed and forward primer, we checked the melting temperature (T_m) of complementars part by NEB Tm Calculator [99], in order to determine the annealing temperature (T_{ann}) of the PCR reaction (Table 4.2. and 4.3.). It is important to calculate annealing temperature for each primer, because too high annealing temperature preclude primers to bind to the template. Additionally, in each case we checked both primers for self-dimers, hetero-dimers (IDT Oligo DNA Analyzer) [100] and for secondary binding sites at the template plasmid (DNA Align Tool) [101]. The complementary part of primers had GC content between 40–60%, but whole primers contented no more than 3 G/C base at the 3‘end of the primer.

Table 4.3: PCR

Step	T [°C]	[s]
Initial denaturation	98	30
25–35 Cycles	98	5–10
	*50–72	10–30
	72	$^{\&}$20–30 s/kb
Final Extension	72	120
Hold	4–10	

* Depending on melting temperature of primers

$^{\&}$ Approximately 20/30 seconds per kb DNA

All primers were phosphorylated according to the first step in protocol 1. The second step was amplification using Q5 Hot Start High-Fidelity DNA polymerase, primers, template DNA and nucleotides (Protocol 1, step 2), according to PCR cycling conditions in table 4.2. Each PCR had a specific annealing temperature and extension time. The extension time is determined by the size of DNA which will be amplified. In the case of insertion of solubility tag to the His$_6$-SUMO Hepcidin-25 pET15b vector, the extension time was 3 minutes and 2 seconds. In the case of insertion of Hep-25 to the pMC vector, the extension time was 1 minute and 30 seconds. Following the PCR amplification, we used the Monarch® PCR Cleanup kit to quickly and easily purify DNA from PCR. The incubation of purified DNA with DpnI and ligase allowed the removal of the template DNA and rapid circularization of the PCR product, respectively (Protocol 1, step 3 and 4).

Protocol 1:

First step: Phosphorylation 25µl Reaction		Second step: PCR reaction 50µl Reaction	
Water	18.0µl	Water	25.0µl
10x T4 PNK Buffer	2.5µl	5x Q5 Buffer	10.0µl
(100µM) reversed primer	0.5µl	(10mM) dNTPs	1.0µl
(100µM) forward primer	0.5µl	(2µM) Primer mix	12.5µl
(10mM) ATP	2.5µl	DNA template (~ 1ng)	1.0µl
T4 PNK	0.5µl	Q5 Polymerase	0.5µl
Conditions:	37°C – 10 min 65°C – 20 min	Conditions:	Table 4.1

Third step: DpnI 10µl Reaction		Fourth step: T4 ligation 15µl Reaction	
10 x CutSmart Buffer	1µl	PEG, 25%	3.08µl
Water	*	(10 mM) ATP	1.54µl
PCR-DNA (50ng)	*	T4 Ligase	0.77µl
DpnI	0.1µl		
Conditions:	37°C – 15 min 80°C – 20 min	Conditions:	25°C – 10 min

* Depending on the concentration of DNA after PCR reaction

DNA concentration was estimated by measuring the absorbance at 260nm, adjusting the A260 measurement for turbidity (measured by absorbance at 320nm), multiplying by the dilution factor, and using the relationship that an A260 of 1.0 the concentration is equal to 50µg/mL pure DNA [102].

$$C(\mu\mathrm{g/mL}) = (\mathrm{A260} - \mathrm{A320}) * \mathrm{dilutionfactor} * 50\mu\mathrm{g/mL} \tag{4.1}$$

In order to evaluate the DNA purity, absorbance was measured at 230nm, 260nm and 280nm.

The purity was calculated as a ratio of the absorbance at 260nm divided by the ration at 230nm and 280nm [102].

$$\text{DNA purity } (A260/A230) = (A260 - A320) \div (A230 - A320) \quad (4.2)$$

$$\text{DNA purity } (A260/A280) = (A260 - A320) \div (A280 - A320) \quad (4.3)$$

4.3 FX-cloning

FX-cloning allows the quick exchange of a ccdB sequence, present on the expression plasmid (pQE), with the gene of interest, present on the pDelivery plasmid (pMC) (Section 2.1.2.). Therefore, in the first step, the gene encoding Hepcidin-25 was inserted into the pMC vector as explained in the Section 4.2. The PCR product contained the gene encoding the Hep-25 and SapI restriction sites, as well as, ATG and GGG overhangs, inserted at ends of Hepcidin-25 DNA sequence. The pExpression vector containing target fusion tags and ccdB was produced by Roche. The next step was subcloning of a Hep-25 into a pExpression vector according to the Protocol 2. In the initial step, 100ng of the pDelivery and 100ng of pExpression vector were mixed in 1µl of 10x Cutsmart buffer. The volume of reaction was adjusted with MilliQ H_2O to 10µL and the digestion was started by the addition of 1µL of 0.20µL of SapI(2U). The mixture was digested for 15 min at 37°C. The enzyme was subsequently heat-inactivated by a 20 min incubation at 65°C and cooled to room temperature. Fragments were ligated by the addition of 1.166µL of 10x Ligase buffer and 0.5µl T4 Ligase and incubate at 25°C for 10 min.

Protocol 2:

Subcloning 10 µl Reaction		T4 ligation	
pDelivery DNA (100ng)	*	10X Ligase Buffer	1.166µl
pExpression DNA (100ng)	*	T4 Ligase	0.5µl
Water	*	Conditions:	25°C – 10 min
SapI	0.2µl		
10x CutSmart Buffer	1µl		
Conditions:	37°C – 15 min 65°C – 20 min		

* Depending on the concentration of pDelivery and pExpression DNA

A reaction mixture was transformed into BL21 (DE3) Tuner competent cells (Experimental section 4.5.). As a pExpression vector conferred Carbenicillin resistance, competent cells were plated on LB-agar supplemented with 50µg/mL of Carbenicillin and colonies were obtained after overnight incubation at 37°C. Since, a ccdB sequence causes cell death and pDelivery vector had Kanamycin resistance, we had a high selectivity of colonies which contained the pExpression vector with inserted Hepcidin-25. The gene encoding the Hep-25 was incorporated into a pExpression vector which contained the desired N-terminal PelB and Ompt tag and a C-terminal His_6 tag. In the last step, the plasmid was purified using Monarch® Plasmid kit and transformed into BL21 (DE3) Origami B competent cells (Experimental section 4.5.).

4.4 Plasmid sequencing

DNA sequencing is the process of determining the nucleotide order of a given DNA fragment, e.g. by the Sanger method. DNA sequencing of plasmids was performed by the Eurofins genomics company via capillary electrophoresis using ABI's 3730xl DNA analyser. Applied Biosystem utilizes the process of labeling the terminators, commonly called 'dye terminator sequencing'. The major advantage of this approach is the complete sequencing set can be performed in a single reaction, rather than the four needed with the labeled-primer approach. Dye terminator sequencing is accomplished by labeling each of the dideoxynucleotide chain-terminators with a separate fluorescent dye, which fluoresces at a different wavelength.

4.5 Transformation

Escherichia coli BL21 (DE3) Origami B or BL21 (DE3) Tuner served as host for the amplification of plasmids. All plasmids containing the genetic information of a target protein were transformed into competent cells. The genetic information of His_6-SUMO-constructs were present on the pET15b-inducible vector. PelB-Hep-25-His_6 and OmpT-Hep-25-His_6 genetic information were incorporated in the pQE expression vector through FX-cloning. The respective sequences for the proteins can be taken from the appendix. In each case between 1-10µl of the plasmid solution was added in 100µl of competent cells *E. coli* BL21 (DE3) Origami B or BL21 (DE3) Tuner and incubated on ice for 30 minutes. The cell suspension was then incubated at 42°C for 30 seconds. In next step, we added 800µl of warm SOC medium into cell suspension and the mixture was shaken for 1h a at 37°C. Transformed cells were selected on

LB agar plates containing 50µg/ml Carbenicillin, overnight at 37°C. A selected culture was inoculated in 5ml of LB medium and shaken overnight at 37°C. In 920µl of the overnight culture was added 80µl of dimethyl sulfoxide (DMSO). A culture was frozen with liquid nitrogen and stored at 80°C. All media used for plasmid transformation into competent cells are summarized in Table 4.4.

Table 4.4: Transformation medium:

SOC medium:
20g/L tryptone
5g/L yeast extract
10mM NaCl
2.5mM KCl
autoclaved and cooled
10mM $MgCl_2$
10mM $MgSO_4$.
20mM glucose
pH titrated to 7.0 with NaOH

LB medium:
10g/L tryptone
5g/L yeast extract
10g/L NaCl
autoclaved and cooled
20mL/L 20 % glucose (autoclaved)
50mg/L Carbenicillin

LB agar plates:
10g/L tryptone
5g/L yeast extract
10g/L NaCl
15g/L agar
autoclaved and cooled
20mL/L 20 % glucose (autoclaved)
50mg/L Carbenicillin

4.6 Protein expression

All proteins were produced using the pET expression system (NOVAGEN). Depending on whether isotopic labelling with ^{15}N was necessary or not, cell culture was carried out in a minimal medium (M9 medium), containing ^{15}N enriched nutrients, or in a complete LB medium (Table 4.5). In both media, cell transformation and induction of gene expression proceeded in the same way:

1. After transformation and overnight incubation on an agar plate, single colonies were inoculated in 20mL of LB medium.
2. The precultures were shaken overnight at 37°C.
3. The overnight cultures were centrifuged at 3,000g and 25°C for 10 minutes.
4. The supernatant was discarded, and the cell pellet was suspended in 1mL of LB medium and inoculated into 1L growth medium (LB or M9)
5. The cells were shaken at 37°C to an OD_{600} of 0.6.
6. Gene expression was induced with 0.25mM IPTG and the temperature was lowered to 25°C.
7. The cell suspension was shaken for overnight at 25°C.
8. The cell suspension was centrifuged for 15 minutes at 10,000g.
9. The supernatant was discarded, the cell pellet was either processed directly for purification, or flash frozen with liquid nitrogen and stored at -80°C as explained in section 4.5.

Table 4.5: Expression medium

LB medium see section 4.5.

M9 medium:
1g/L NaCl
3g/L KH_2PO_4
6g/L Na_2HPO_4
1g/L $^{15}NH_4Cl$
pH titrated with NaOH to 7.4 and autoclaved
2mM $MgSO_4$ (sterile filtered, 0.2µm cellulose acetate membrane)
100µM CaCl2 (sterile filtered, 0.2µm cellulose acetate membrane)
10mL/L 100x BME Vitamins
20mL/L 20 % glucose (autoclaved)
100mg/L Celtone (^{15}N labeled)
50mg/L Carbenicillin

4.7 Purification of protein constructs

Chromatographic steps were performed on an Äkta Purifier (GE HEALTHCARE) and HPLC (Agilent). In all cases, the cell lysate was first applied to a nickel Sepharose column (GE HEALTHCARE). Unspecifically bound proteins were washed from the column with 4% of elution buffer until a stable UV baseline had established. The elution of the protein from the nickel matrix was performed in each experiment with imidazole, applying the linear gradient of elution buffer. Depending on the protein, further purification steps are explained in the protein-specific sections. The Amicon Ultra-15 3kDa membrane (MILLIPORE) was used to concentrate the solutions to a certain protein concentration. The membrane had a cut-off of 3kDa, and it was centrifugated at 3000g, at 4°C. In all cases the protein solution was evaporated prior to purification on HPLC. The protein concentrations were determined photometrically by measuring the absorbance at 280nm. The concentration of the protein was calculated using the Lambert-Beer law:

$$A = \varepsilon * c * d \qquad (4.4)$$

A: measured absorption	c: concentration of the protein
ε: molar extinction coefficient unit	d: thickness of the cuvette (1cm)

The molar extinction coefficient ε of the proteins was determined using the web-based software ProtParam [103]. The values for the proteins are compiled in the appendix.

The expression, purification steps and the purity of the protein solutions were monitored by tricine-sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) in the presence of 10 % polyacrylamide (acrylamide solution A) for proteins bigger than 5kDa and 16% polyacrylamide (acrylamide solution B) in presence of 6M urea for peptides. For SDS-PAGE, first the 10% gels and 15% were prepared according to the composition listed in Table 4.6 [104].

Table 4.6: Composition of the SDS-gel page.

	Stacking gel	Resolving gel	
		10% gel	16% gel
Acrylamide solution A	0.4mL	2mL	—
Acrylamide solution B	0.25mL	—	2.mL
gel buffer 3x	0.75mL	1mL	2.5mL
60% glycerol in H_2O	—	1mL	—
Urea	—	—	1.65g
deionized water	2mL	1mL	1mL
TEMED	2μL	3μL	3μL
10% (w/v) APS	20μL	30μL	30μL

Acrylamide solution A: 97% Acrylamide and 3% N, N- Methylene bisacrylamide
Acrylamide solution B: 94% Acrylamide and 6% N, N- Methylene bisacrylamide

A sample of 10 μL was mixed with the same amount of loading buffer (Table 4.7.) and incubated for 10 minutes at 95°C to denature the protein. As the loading buffer contains sodium dodecyl sulfate (SDS), the denatured state of the protein will be solubilized and stabilized by SDS. The SDS protein complexes, regardless of the intrinsic charge of the protein, are strongly negatively charged. Therefore, the proteins are separated in the electric field only according to their molecular size. The gel was run at 15W approximately 40 minutes in the presence of cathode and anode buffer (Table 4.7). The gel chamber was filled with cathode buffer and the electrophoresis chamber with anode buffer. All buffers used for the SDS-PAGE are listed in Table 4.7.

Table 4.7: SDS-PAGE buffers

	Loading buffer	3xGel buffer	Anode buffer	Cathode buffer
Bromophenol blue	0.04% (w/v)	—	—	—
Glycerol	20% (v/v)	—	—	—
DTT	2% (w/v)	—	—	—
EDTA	25mM	—	—	—
SDS	4% (w/v)	0.3% (w/v)	—	1% (w/v)
TRIS	100mM	300mM	200mM	100mM
Tricin	—	—	—	100mM
pH	6.8	8.45	8.90	8.25

Low and broad range molecular weight standards were used to estimate the molecular weights. We used a protein-molecular weight marker from Roti®-Mark (Broad-Range SDS-PAGE standard) and Bio Rad (Broad-Range and Natural Polypeptide SDS-PAGE standards), as it is show in the Table 4.8.

Table 4.8: Protein-molecular weight marker

Roti®-Mark		Bio Rad			
Broad-Range		Broad-Range		Natural Polypeptide	
Protein	kDa	Protein	kDa	Protein	kDa
Myosin	212	Myosin	200,0	Triosephosphate isomerase	26,6
B-galactosidase	118	B-galactosidase	116,3		
Serum albumin	66	Phosphorylase b	97,4	Myoglobin	17,9
Ovalbumin	43	Serum albumin	66,2	α-Lactalbumin	14,4
Carbonic anhydrase	29	Ovalbumin	45,0	Aprotinin	6,5
Trypsin Inhibitor	20	Carbonic anhydrase	31,0	Insulin b chain	3,5
Lysozyme	14	Trypsin Inhibitor	21,5	Bacitracin	1,4
		Lysozyme	14,4		
		Aprotinin	6,5		

Protein bands on the gel were stained in the presence of staining solution containing 20% (vol.) methanol, 10% (vol.) phosphoric acid, 10% (wt) ammonium sulphate and 0.12% (wt) Coomassie brilliant blue G-250. After the staining process, the band intensity was further enhanced by destaining a gel in the deionized water.

4.7.1 Cell lysis and purification of His$_6$-SUMO constructs

All His$_6$-SUMO constructs were purified according to the same protocol. Table 4.9 contains only buffer conditions for His$_6$-SUMO-Hep-25 purification. Unless otherwise noted, other His$_6$-SUMO constructs were purified under the same buffer conditions, as for His$_6$-SUMO-Hep-25, but different pH values depending on the isoelectric point of constructs.

Table 4.9: Purification buffers of His$_6$-SUMO-Hep-25

Lyses buffer

300 mM NaCl
50 mM phosphate buffer
pH adjusted to 7.4 with HCl

HisTrap loading buffer

300 mM NaCl
50 mM phosphate buffer
5 mM DTT
1 M Urea
pH adjusted to 7.4 with HCl

HisTrap elution buffer

300 mM NaCl
50 mM phosphate buffer
5 mM DTT
1 M Urea
500 mM imidazole
pH adjusted to 7.4 with HCl

Gelfiltration buffer

150 mM NaCl
50 mM phosphate buffer
pH adjusted to 7.0 with HCl

Folding buffer

0.2 mM reduced glutathione (GSH)
0.2 mM oxidized glutathione (GSSG) [XX]
pH 7 adjusted with NH_4OH 30%

Following expression, about 3g of cells were thawed on ice and resuspended in 10ml of lyses buffer and cooled on ice before and after digestion. Lyses buffer was supplemented with 20 mM DTT, 0.4 M guanidium HCl, 200µL 0.5M $MgCl_2$/0.1M $CaCl_2$, 2mL of protease inhibitors (one tablet of SIGMA*FAST*™ protease inhibitors dissolved in 10mL of H_2O) and a little DNase I of the tip of the spatula. Cells were disrupted via a high-pressure homogenizer Emulsiflex C3 (AVESTIN). After lysis, cells were ultra-centrifuged at 100,000g for 30 min at 4°C to separate cell membrane fragments from the soluble portion of the cells. The supernatant was diluted to 40mL with the lyses buffer and passed through a 0.45µm glass fiber (GF)/polyester (PET) CHROMAFIL filter (Carl Roth). The 40 mL of supernatant containing His$_6$-Sumo-Hep-25 constructs was loaded onto a 5mL nickel Sepharose column (GE HEALTHCARE). The column was firstly equilibrated with five volumes of loading buffer in the presence of 4% elution buffer. His$_6$-Sumo-Hep-25 fusion protein was eluted with the linear gradient of elution buffer. All fractions, which contained Hep-25 fusion protein, were combined and concentrated using an Amicon 3-kDa molecular weight cutoff membrane (MILLIPORE).

At the same time, in order to dilute imidazole concentration, the buffer was changed by concentrating the protein solution with lyses buffer (Table 4.9). The His_6-SUMO tag was enzymatically cleaved by incubation with 0.4mg of ULP1 protease. His_6-SUMO tag was cleaved in the presence of 1M urea and 5mM DTT. Only in the case of His_6-SUMO-Hep-25-DEDEDE, the mixture was incubated for 20h at room temperature in presence of 0.5M urea and 5mM DTT. The next day, the solution was boiled 4 h at 40°C in the presence of 3M guanidium HCl. After cleavage, the solution was charged onto a preequilibrated Superdex 30 16/600 gel-filtration column (GE HEALTHCARE), elution was achieved with gel-filtration buffer. Following the gel-filtration, the linear Hep-25 was oxidized overnight (16-24 h) in the presence of 20% ACN and folding buffer under stirring at 100 rpm using a magnetic stirrer. Then, the solution was acidified to pH 2.5 by addition of TFA (0.1%), evaporated and dissolved in mixture of H_2O/ACN = 80%/20%. The solution was further loaded onto a 250/10 Macherey-Nagel™ Nucleosil™ C18 column (120Å, 5µm) for preparative HPLC purification. All preparative runs were carried out with a linear AB gradient starting with 20% of eluent B and going up to 60% of eluent B (99.95% ACN and 0.05% TFA) and 40% of eluent A (99.95% H_2O and 0.05% TFA). Fractions containing folded hepcidin were collected, evaporated and the protein solution was available for further experiments. The purified protein solution was collected in fractions of 2mL, the flow rate of nickel-affinity chromatography was 2mL/min and of gel-filtration was 1mL/min. During HPLC purification, protein solution was collected in fractions of 3mL, the flow rate was 3mL/min. In each step protein was monitored by UV absorption at 210nm and 280nm.

4.7.2 Extraction of protein from periplasmic space

Designed constructs contained a N-terminal OmpT or PelB periplasmic tag, which is cleaved from the target protein upon expression. Periplasmic proteins or signal proteins are secreted into the periplasmic space located between the outer and inner membrane of *E. coli*. In order to purify proteins secreted into the periplasmic space using nickel affinity chromatography, the His_6 tag was engineered to the C-terminus of the target protein. All signal peptide constructs were purified according to the same protocol. Table 4.10 contains all buffer used for extraction of protein from periplasmic space.

Table 4.10: Buffers for extraction of protein from periplasmic space

Hypertonic solution
30mM Tris-HCl
20% sucrose
1mM EDTA
pH 7.0

Hypotonic solution
5mM $MgSO_4$
pH 7.0

HisTrap loading buffer
300mM NaCl
50mM phosphate buffer
pH adjusted to 7.0 with HCl

HisTrap elution buffer
300mM NaCl
50mM phosphate buffer
500mM imidazole
pH adjusted to 7.0 with HCl

Following expression, after thawing, the cell pellet was resuspended in the cold hypertonic solution and incubated on ice for 30 min with gentle agitation. Cells were centrifuged at 100,000g for 30 min and the supernatant collected. Following centrifugation, the cell pellet was again washed in the cold hypotonic solution and incubated on ice for 30 min with gentle agitation and centrifuged at 100,000g for 30 min and the supernatant collected. Both supernatants were combined and passed through a 0.45µm glass fiber (GF)/polyester (PET) CHROMAFIL filter (Carl Roth), before it was loaded onto a nickel affinity column. The column was firstly equilibrated with five volumes of loading buffer in the presence of 4% elution buffer. His$_6$-SUMO-Hepcidin fusion protein was eluted with the linear gradient of elution buffer. Following the affinity chromatography, peak fractions were evaporated and dissolved in a mixture of H_2O/ACN = 80%/20%. The solution was further loaded onto a 250/10 Macherey-Nagel™ Nucleosil™ C18 column (120Å, 5µm) for preparative HPLC purification. All preparative runs were carried out with a linear AB gradient starting with 20% of eluent B and going up to 60% of eluent B (99.95% ACN and 0.05% TFA) and 40% of eluent A (99.95% H2O and 0.05% TFA). During HPLC purification, protein solution was collected in fractions of 3mL, the flow rate was 3mL/min. In each step protein was monitored by UV absorption at 210nm and 280nm. All buffers used for the extraction of protein from periplasmic space are listed in Table 4.10.

4.8 Nuclear magnetic resonance (NMR) spectroscopy

All NMR measurements were taken on a BRUKER AVANCE III 600MHz (Bruker Biospin GmbH, Rheinstetten, Germany) equipped with a 5 mm double resonance broadband probe (BBI probe). Since all samples were measured in aqueous buffers, D_2O was added to each sample to a final concentration of 5% (vol.) for the field lock. The samples were measured either at 325K or 298K. The temperature was calibrated with Thermocouple PE 100. Unless otherwise stated, all measurements were carried out with a sample volume of 500µL in a 5 mm NMR tube (Boroeco-5-7). Chemical shifts were referenced to internal 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) at 0.0 ppm. Acquired data were processed and analyzed using Bruker Topspin (v3.5) and the CARA software (v1.9.1) [105].

4.8.1 NMR resonance assignment and the structural information

For further NMR studies, an unlabelled and ^{15}N-labeled Hep-25 was prepared at a concentration of approximately 0.6mM. In the case of Hep-25-DEDEDE and -SSSSS, the concentration was 1mM and 0.3mM, respectively. In the case of hexapeptide 2.5mM or 2.7mM sample was used for NMR measurements. Peptide samples were dissolved in a 25mM phosphate buffer.^{15}N resonances were indirectly referenced by multiplying the absolute referenced proton resonance frequency in MHz by the factor 0,101329118 to obtain absolute ^{15}N resonance frequency [106]. Chemical shift assignment of Hep-25 was achieved by analyzing 1D-^{1}H, ^{1}H-^{13}C HSQC, 2D ^{1}H-^{15}N HSQC, 2D ^{1}H TOCSY, 2D ^{1}H-^{15}N HSQC-TOCSY, 2D ^{1}H NOESY and 2D ^{1}H-ROESY spectra. Assignment of hexapeptide was achieved by analyzing 1D-^{1}H, ^{1}H-^{13}C HSQC, 2D ^{1}H TOCSY and 2D ^{1}H ROESY. The ROESY and NOESY mixing time of Hep-25 was 200ms and ROESY mixing time of hexapeptide was 400ms. In case of Hep-25, NMR spectra and Hep-25-metal complex were recorded at 298K and 325K. The spectra of hexapeptide and hexapeptide-metal complex were, however, acquired only at 298K.

4.8.2 Metal coordination monitored by NMR spectroscopy

Metal titrations were performed by adding small amounts of metal salts solutions to the peptide sample. Ni^{2+} or Cu^{2+} titrations were performed by adding 10mM $NiSO_4 \cdot 6H_2O$ or $CuSO_4$ stock solution, respectively, to the 25 mM phosphate-buffered solution of the peptide, at pH of 7.5, until a 1:1 complex was formed. The titration was monitored by 1D-^{1}H NMR and 2D TOCSY experiments.

4.9 Molecular modeling

Peptide-metal complex structures were calculated with Xplor-NIH v2.43 [80, 107]. Internuclear distances, a peptide sequence and dihedral angles were used as input files for the structure calculations. Modifications of the Xplor-NIH parameter and topology files were made to incorporate the nickel or copper ion in a square planar complex, using distance and torsion angle restraints, involving the four nitrogen atoms. Distance restraints were derived from ROESY or NOESY cross-peak volumes, which were multiplicity corrected and calibrated with reference to cross peak volumes corresponding to known distances in the peptide [74]. Chemical shifts were translated into backbone dihedral angle restraints by using the software TALOS+ [79]. Calculations were performed according to the standard protocol of Xplor-NIH (Section 2.4). The PDB structure of Hep-25 (PDB code 2KEF) [44] or amino acid sequence of hexapeptide was used as a starting structure for simulated annealing. After refinement, the 10 lowest energy structures, from a total of 100, were selected and manually analysed. The quality of the final structure was evaluated using Protein Structure Validation Software (PSVS) [108]. All figures were created using PYMOL or VMD-Xplor [109].

5 References

1. T. Ganz, Cellular iron: ferroportin is the only way out. *Cell Metab* **2005**; *1*:155–157.

2. N. Abbaspour, R. Hurrell, R. Kelishadi, Review on iron and its importance for human health. *J Res Med Sci* **2014**; *19*:164–174.

3. A. Lawen, D. J. R. Lane, Mammalian iron homeostasis in health and disease: uptake, storage, transport, and molecular mechanisms of action. *Antioxid Redox Signal* **2013**; *18:*2473–2507.

4. A. U. Steinbicker, M. U. Muckenthaler, Out of balance-systemic iron homeostasis in iron related disorders. *Nutrients* **2013**; *5*:3034–3061.

5. J. Chifman, R. Laubenbacher, S. V. Torti, A systems biology approach to iron metabolism. *Adv. Exp. Med. Biol.* **2014**; *844*:201–225.

6. T. A. Rouault, The role of iron regulatory proteins in mammalian iron homeostasis and disease. *Nat Chem Biol* **2006**; *2*:406–414.

7. M. U. Muckenthaler, B. Galy, M. W. Hentze, Systemic iron homeostasis and the iron-responsive element/iron-regulatory protein (IRE/IRP) regulatory network. *Annu Rev Nutr* **2008**; *28*:197–213.

8. M. Hentze, S. Caughman, T. Rouault, J. Barriocanal, A. Dancis, J. Harford, R. Klausner, Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. *Science* **1987**; *238*:1570–1573.

9. T. Ganz, Hepcidin and iron regulation, 10 years later. *Blood* **2011**; *117*:4425–4433.

10. N. C. Andrews, Disorders of iron metabolism. *N Engl J Med* **1999**; *341*:1986–1995.

11. T. Ganz, E. Nemeth, Hepcidin and iron homeostasis. *Biochim. Biophys. Acta* **2012**; *1823*:1434–1443.

12. J. Kaplan, D. M. Ward, The essential nature of iron usage and regulation. *Curr Biol* **2013**; *23*: 642-646.

13. J. C. Hunt, C. A. Zito, L. K. Johnson, Body iron excretion by healthy men and women. *Am J Clin Nutr* **2009**; *89(6)*:1792–1798.

14. M. Zielińska-Dawidziak, Plant ferritin-a source of iron to prevent its deficiency. *Nutrients* **2015**; *7*:1184–1201.

15. E. Nemeth, T. Ganz, Regulation of iron metabolism by hepcidin. *Annu Rev Nutr* **2006**; *26*: 323–342.

16. A. T. McKie, D. Barrow, G. O. Latunde-Dada, A. Rolfs, G. Sager, E. Mudaly, M. Mudaly, C. Richardson, D. Barlow, A. Bomford, T. J. Peters, K. B: Raja, S. Shirali, M. A. Hediger, F. Farzaneh, R. J. Simpson, An iron-regulated ferric reductase associated with the absorption of dietary iron. *Science* **2001**; *291*: 1755–1759.

17. S. Gulec, G. J. Anderson, J. F. Collins, Mechanistic and regulatory aspects of intestinal iron absorption. *Am J Physiol Gastrointest Liver Physiol* **2014**; *307*: 397-409.

18. S. Abboud, D. J. Haile, A movel mammalian iron-regulated protein involved in intracellular iron metabolism. *J. Biol. Chem.* **2000**; 26:19906–19912.

19. E. L. MacKenzie, K. Iwasaki, Y. Tsuji, Intracellular iron transport and storage: from molecular mechanisms to health implications. *Antioxid Redox Signal* **2008**; *10:*997–1030.

20. K. Dassler, M. Zydek, K. Wandzik, M. Kaup, H. Fuchs, Release of the soluble transferrin receptor is directly regulated by binding of its ligand ferritransferrin. *J. Biol. Chem.* **2006**; *281*: 3297–3304.

21. R. S. Hillman, P. A. Henderson, Control of marrow production by the level of iron supply. *J. Clin. Invest.* **1969***; 48*:454–460.

22. N. C. Andrews, P. J. Schmidt, Iron homeostasis. *Annu Rev Physiol* **2007**; *69*:69–85.

23. M. W. Hentze, M. U. Muckenthaler, N. C. Andrews, Balancing Acts: Molecular control of review mammalian iron metabolism. *Cell Press* **2004**; *117*: 285–297.

24. D. Bratosin, J. Mazurier, J. P. Tissier, J. Estaquier, J. J. Huart, J. C. Ameisen, D. Aminoff, J. Montreuil, Cellular and molecular mechanisms of senescent erythrocyte phagocytosis by macrophages. *Biochemie* **1998**; *80*:173–195.

25. M. D. Knutson, M. R. Vafa, D. J. Haile, M. Wessling-Resnick, Iron loading and erythrophagocytosis increase ferroportin 1 (FPN1) expression in J774 macrophages. *Blood* **2003**; *102*: 4191–4197.

26. A. S. N. Krause, M. Hans-Jurgen, A. Schulz, F. Wolf-Georg, P. Schulz-Knappe, K. Adermann, LEAP1, a novel highly disulfidebonded human peptide, exhibits antimicrobial activity. *FEBS Letters* **2000**; *480*: 147-150.

27. C. H. Park, E. V. Valore, A. J. Waring, T. Ganz, Hepcidin, a urinary antimicrobial peptide synthesized in the liver. *J. Biol. Chem.* **2001**; *276*: 7806–7810.

28. C. Pigeon, G. Ilyin, B. Courselaud, P. Leroyer, B. Turlin, P. Brissot, O. Loréal, A new mouse liver-specific gene, encoding a protein homologous to human antimicrobial peptide hepcidin, is overexpressed during iron overload. *J. Biol. Chem.* **2001**; *276*: 7811–7819.

29. G. Nicolas, M. Bennoun, I. Devaux, C. Beaumont, B. Grandchamp, A. Kahn, S. Vaulont, 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.* **2001**; *98*: 8780–8785

30. G. Nicolas, M. Bennoun, A. Porteu, S. Mativet, C. Beaumont, B. Grandchamp, M. Sirito, M. Sawadogo, A. Kahn, S. Vaulont, Severe iron deficiency anemia in transgenic mice expressing liver hepcidin. *Proc. Natl. Acad. Sci. U.S.A.* **2002**; *99:* 4596–4601

31. K. Pantopoulos, Inherited disorders of iron overload. *Frontiers in nutrition* **2018**; *5*: 1-11.

32. E. Nemeth, M. S. Tuttle, J. Powelson, M. B. Vaughn, A. Donovan, D. M. Ward, T. Ganz, J. Kaplan, Hepcidin regulates cellular iron efflux by binding to ferroportin and inducing its internalization. *Science* **2004**; *306*: 2090–2093.

33. E. Nemeth, G. C. Preza, C. L. Jung, J. Kaplan, A. J. Waring, T. Ganz, The N-terminus of hepcidin is essential for its interaction with ferroportin: structure-function study. *Blood* **2006**; *107*: 328–333.

34. E. H. J. M. Kemna, H. Tjalsma, V. N. Podust, D. W. Swinkel, Mass spectrometry-based hepcidin measurements in serum and urine: analytical aspects and clinical implications. *Clin Chem* **2007**; *53*: 620–628.

35. R. J. Clark, C. C. Tan, G. C. Preza, E. Nemeth, T. Ganz, D. J. Craik, Understanding the structure/activity relationships of the iron regulatory peptide hepcidin. *Chem Biol* **2011**; *18*: 336–343.

36. S. Dev, J. L. Babitt, Overview of iron metabolism in health and disease. Hemodialysis international. *International Symposium on Home Hemodialysis* **2017**; *21*: S6-S20.

37. I. De Domenico, D. M. Ward, J. Kaplan, Hepcidin regulation: ironing out the details. *The J. Clin. Invest.* **2007**; *117*: 1755–1757.

38. J. L. Babitt, H. Y. Lin, Molecular mechanisms of hepcidin regulation: implications for the anemia of CKD. *Am J Kidney Dis* **2010**; *55*: 726–741.

39. J. Chifman, R. Laubenbacher, S. V. Torti, A systems biology approach to iron metabolism. *Adv. Exp. Med. Biol.* **2014**; *844*: 201–225.

40. J. L. Babitt, F. W. Huang, Y. Xia, Y. Sidis, N. C. Andrews, H. Y. Lin, Modulation of bone morphogenetic protein signaling in vivo regulates systemic iron balance. *J. Clin. Invest.* **2007**; *117*: 1933–1939.

41. E. Nemeth, S. Rivera, V. Gabayan, C. Keller, S. Taudorf, B. K. Pedersen, T. Ganz, IL-6 mediates hypoferremia of inflammation by inducing the synthesis of the iron regulatory hormone hepcidin. *J. Clin. Invest* **2004**; *113*: 1271–1276.

42. H. N. Hunter, D. B. Fulton, T. Ganz, H. J. Vogel, The solution structure of human hepcidin, a peptide hormone with antimicrobial activity that is involved in iron uptake and hereditary hemochromatosis. *J. Biol. Chem.* **2002**; *277*: 37597–37603.

43. X. Lauth, J. J. Babon, J. A. Stannard, S. Singh, V. Nizet, J. M. Carlberg, V. E. Ostland, M. W. Pennington, R. S. Norton, M. E. Westerman, Bass hepcidin synthesis, solution structure, antimicrobial activities and synergism, and in vivo hepatic response to bacterial infections. *J. Biol. Chem.* **2005**; *280*: 9272–9282.

44. J. B. Jordan, L. Poppe, M. Haniu, T. Arvedson, R. Syed, V. Li, H. Kohno, H. Kim, P. D. Schnier, T. S. Harvey, L. P. Miranda, J. Cheetham, B. J. Sasu, Hepcidin revisited, disulfide connectivity, dynamics, and structure. *J. Biol. Chem.* **2009**; *284*: 24155–24167.

45. N. Camerman, A. Camerman, B. Sarkar, Molecular design to mimic the copper(II) transport site of human albumin. The crystal and molecular structure of copper(II) – glycylglycyl-L-histidine-N-methyl amide monoaquo complex. *Can J Chem* **1976**; *54:* 1309–1316.

46. C. S. B. Hartford, Amino Terminal Cu(II)- and Ni(II)-Binding (ATCUN) Motif of proteins and peptides metal binding, DNA cleavage, and other properties. *Acc Chem Res* **1997**; *30*: 123–130.

47. J. Laussac, B. Sarkar, 13Carbon-nuclear magnetic resonance investigation of the Cu(I1)-binding to the native sequence peptide representing the Cu(II) - transport site of human albumin. *J. Biol. Chem* **1980**; *255*: 7563–7568.

48. C. Hureau, H. Eury, R. Guillot, C. Bijani, S. Sayen, P.-L. Solari, E. Guillon, P. Faller, P. Dorlet, X-ray and solution structures of Cu(II) GHK and Cu(II) DAHK complexes: influence on their redox properties. *Chemistry* **2011**; *17*: 10151–10160.

49. R. Sankararamakrishnan, S. Verma, S. Kumar, ATCUN-like metal-binding motifs in proteins: identification and characterization by crystal structure and sequence analysis. *Proteins* **2005**; *58*: 211–221.

50. P. Gonzalez, K. Bossak, E. Stefaniak, C. Hureau, L. Raibaut, W. Bal, P. Faller, N-terminal Cu-binding motifs (Xxx-Zzz-His, Xxx-His) and their derivatives: chemistry, biology and medicinal applications. *Chemistry* **2018**; *24*: 8029–8041.

51. I. M. Abbas, M. Vranic, H. Hoffmann, A. H. El-Khatib, M. Montes-Bayón, H. M. Möller, M. G: Weller, Investigations of the Copper Peptide Hepcidin-25 by LC-MS/MS and NMR. I*nt J Mol Sci* **2018**; *19*: 1–16.

52. S. Melino, L. Garlando, M. Patamia, M. Paci, R. Petruzzelli, A metal binding site is present in the amino terminal region of the bioactive iron regulator hepcidin25. *J Peptide res* **2005**; *66*: 65–71

53. T. Konz, M. Montes-Bayón, A. Sanz-Medel, Elemental labeling and isotope dilution analysis for the quantification of the peptide hepcidin-25 in serum samples by HPLC-ICP-MS. *Anal Chem* **2012**; *84*: 8133–8139.

54. C. Tselepis, S. J. Ford, A. T. McKie, W. Vogel, H. Zoller, R. J. Simpson, J. Diaz Castro, T. H. Iqbal, D. G. Ward, Characterization of the transition-metal-binding properties of hepcidin. *Biochem J* **2010**; *427*: 289–296.

55. D. Płonka, W. Bal, The N-terminus of hepcidin is a strong and potentially biologically relevant Cu(II) chelator. *Inorganica Chimica Acta* **2017**; *472*: 76–81

56. M. Muñoz, I. Villar, J. A. García-Erce, An update on iron physiology. *World J Gastroenterol* **2009**; *15*: 4617-4626

57. G. A. McMillin, J. J. Travis, J. W. Hunt, Direct measurement of free copper in serum or plasma ultrafiltrate. *Am. J. Clin. Pathol.* **2009**; *131*: 160-165.

58. G. L. Rosano, E. A. Ceccarelli, Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol* **2014**; *5*: 1-17

59. K Terpe, Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* **2003**; *60*: 523–533.

60. M. P. Malakhov, M. R. Mattern, O. A. Malakhova, M. Drinker, S. D. Weeks, T. R. Butt, SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J Struct Funct Genomics* **2004**; *5*: 75–86.

61. J. H. Choi, S. Y. Lee, Secretory and extracellular production of recombinant proteins using Escherichia coli. *Appl Microbiol Biotechnol* **2004**; *64*: 625–635.

62. E. R. Geertsma, R. Dutzler, A versatile and efficient high-throughput cloning tool for structural biology. *Biochemistry* **2011**; *50*: 3272–3278.

63. A. Samadikuchaksaraei, Polymerase chain reaction for biomedical applications. *In Tech*: Iran, **2016**

64. J. Reikofski, B. Y. Tao, Polymerase chain reaction (PCR) techniques for site-directed mutagenesis. *Biotechnology Advances* **1992**; *10*: 535–547.

65. https://international.neb.com/applications/cloning-and-synthetic-biology/site-directed-mutagenesis. Accessed 27.03.2019.

66. M. Laible, K. Boonrod, Homemade site directed mutagenesis of whole plasmids. *J Vis Exp* **2009**; *27*: 1-3.

67. https://international.neb.com/protocols/pcr-using-q5-high-fidelity-dna-polymerase. Accessed 27.03.2019.

68. J. Marley, M. Lu, C. Bracken, A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* **2001**; *20*: 71–75

69. F. W. Studier, B. A. Moffatt, Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **1986**; *189*: 113–130

70. F. W. Studier, A. H. Rosenberg, J. J. Dunn, J. W. Dubendorff, Use of T7 RNA polymerase to direct expression of cloned genes. *Meth Enzymol* **1990**; *185*: 60–89

71. A. H. Rosenberg, B. N. Lade, D. S. Chui, S. W. Lin, J. J. Dunn, F. W. Studier, Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* **1987**; *56*: 125–135

72. T. D. W. Claridge, High-resolution NMR techniques in organic chemistry, *Elsevier Science*, UK, **2009**.

73. J. Keeler, Understanding NMR spectroscopy, *John Wiley and Sons*, UK, **2010**.

74. K. Würthrich, NMR of proteins and nucleic acids, *Wiley*, USA, **1986**.

75. M. P. Williamson, Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc.* **2013**; *73*: 1–16.

76. D.S. Wishart, B.D. Sykes and F.M. Richards, The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **1992**; 31: 164-1651.

77. M. Karplus, Vicinal proton coupling in nuclear magnetic resonance. *J Am Chem Soc* **1963**; *85*: 2870–2871.

78. A. W. Overhauser, Polarization of nuclei in metals. *Phys Rev* **1953**; *92*: 411–415.

79. G. Cornilescu, F. Delaglio, A. Bax, Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* **1999**; *13*: 289–302

80. C. D. Schwieters, J. Kuszewski, N. Tjandra, G. M. Clore, The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **2003**; *160*: 65–73

81. J. Kuszewski, M. Nilges, A. T. Brünger, Sampling and efficiency of metric matrix distance geometry: A novel partial metrization algorithm. *J Biomol NMR* **1992**; *2*: 33–56.

82. Y.C. Lee, P. L. Jackson, M. J. Jablonsky, D. D. Muccio, R. R. Pfister, J. L. Haddox, C. I. Sommers, G. M. Anantharamaiah, M. Chaddha, NMR conformational analysis of *cis* and *trans* proline isomers in the neutrophil chemoattractant, N-acetyl-proline-glycine-proline. *Biopolymers* **2001**; *58*: 548–561.

83. W. Bal, M. Jeżowska-Bojczuk, K. S. Kasprzak, Binding of nickel(II) and copper(II) to the N-terminal sequence of human protamine HP2. *Chem Res Toxicology* **1997**; *10*: 906-914

84. G. Gasmi, A. Singer, J. Forman-Kay, B. Sarkar, NMR structure of neuromedin C, a neurotransmitter with an amino terminal CuII-, NiII-binding (ATCUN) motif. *J. Pept. Res.* **1997**; *49*:500–509

85. M. Berkmen, Production of disulfide-bonded proteins in *Escherichia coli*. *Protein Expr Purif* **2012**; *82*: 240–251.

86. J. T. Sockolosky, F. C. Szoka, Periplasmic production via the pET expression system of soluble, bioactive human growth hormone. *Protein Expr Purif* **2013**; *87*: 129–135.

87. M. P Malakhov, M. R. Mattern, O. A. Malakhova, M. Drinker, S. D. Weeks, T. R. Butt, SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genomics* **2004**; *5*: 75–86.

88. P. Zhou, G. Wagner, Overcoming the solubility limit with solubility-enhancement tags: successful applications in biomolecular NMR studies. *J. Biomol. NMR* **2010**; 46: 23–31.

89. V. Paraskevopoulou, F. H. Falcone, Polyionic Tags as Enhancers of Protein Solubility in Recombinant Protein Expression. *Microorganisms* **2018**; *6*: 2-17.

90. J. Ikhsan, J. D. Wells, B. B. Johnson, M. J. Angove, The effect of aspartic acid on the binding of transition metals to kaolinite. *J Colloid Interface Sci* **2004**; *273*: 6–13.

91. T. Yasui, T. Ama, Metal Complexes of Amino Acids. VIII. Carbon-13 nuclear magnetic resonances of cobalt(III) complexes containing L-aspartic and L-glutamic acids. *Bull Chem Soc Jpn* **1975**; *48*: 3171–3174.

92. D. M. Francis, R. Page, Strategies to optimize protein expression in *E. coli*. *Curr Protoc Protein Sci* **2010**; *5*: 1-29.

93. A. Aitken, M. Learmonth, Protein determination by UV absorption. *In: J. M. Walker, The Protein Protocols Handbook. Humana Press*, Totowa, **1996**.

94. O. D. Monera, T. J. Sereda, N. E. Zhou, C. M. Kay, R. S. Hodges, Relationship of sidechain hydrophobicity and alpha-helical propensity on the stability of the single-stranded amphipathic alpha-helix. *J Pept Sci* **1995**; *1*: 319–329.

95. Ad Bax, Correction of cross-peak intensities in 2D spin-locked NOE spectroscopy for offset and Hartmann-Hahn effects. *J. Magn. Reson.* **1988**; *77*: 134–147

96. A. Bax, D. G. Davies, Practical aspects of two-dimensional transverse NOE spectroscopy. *J. Magn. Reson.* **1985**; *63*: 207–213

97. K. Tamiola, B. Acar, F. A. A. Mulder, Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. *J. Am. Chem. Soc.*: **2010**; 132: 18000–18003

98. P. M. S. Mielke1, V.V. Krishnan, Characterization of protein secondary structure from NMR chemical shifts. *Prog Nucl Magn Reson Spectrosc.* **2009**; 54: 141–165

99. http://tmcalculator.neb.com. Accessed 21.04.2019.

100. https://eu.idtdna.com/pages/tools/oligoanalyzer. Accessed 21.04.2019.

101. https://embnet.vital-it.ch/software/LALIGN_form.html. Accessed 21.04.2019.

102. https://www.promega.de/en/resources/pubhub/enotes/how-do-i-determine-the concentration-yield-and-purity-of-a-dna-sample. Accessed 21.04.2019.

103. E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, A. Bairoch, Protein Identification and Analysis Tools on the ExPASy Server. *In: John M. Walker, The Proteomics Protocols Handbook, Humana Press*, Totowa, **2005**: 571-607

104. H. Schägger, Tricine-SDS-PAGE. *Nat Protoc* **2006**; *1*: 16–22.

105. R. Keller, The computer aided resonance assignment tutorial. *CANTINA Verlag*: Germany, **2004**.

106. D. H. Live, D. G. Davis, W. C. Agosta, D. Cowburn, Long range hydrogen bond mediated effects in peptides: nitrogen-15 NMR study of gramicidin S in water and organic solvents. *J Am Chem Soc* **1984**; *106*: 1939–1941

107. C. D. Schwieters, J. J. Kuszewski, G. M. Clore, Using Xplor–NIH for NMR molecular structure determination. *Prog Nucl Magn Reson Spectrosc* **2006;** *48*: 47–62.

108. A. Bhattacharya, R. Tejero, G. T. Montelione, Evaluating protein structures determined by structural genomics consortia. *Proteins* **2007**; *66*: 778–795.

109. C. D. Schwieters, G. M. Clore, The VMD-XPLOR visualization package for NMR structure refinement. *J Magn Reson* **2001**; *149*: 239–244.

110. B. Gagliardo, A. Faye, M. Jaouen, J. Deschemin, F. Canonne-Hergaux, S. Vaulont, M. Sari, Production of biologically active forms of recombinant hepcidin, the iron-regulatory hormone. *FEBS Journal* **2008**; *275*: 3793–3803.

111. H. Zhang, Q. Yuan, Y. Zhu, R. Ma, Expression and preparation of recombinant hepcidin in Escherichia coli. *Protein Expr Purif*. **2005**; *41*:409-416.

112. V. Sadr, B. Saffar, R. Emamzadeh, Functional expression and purification of recombinant Hepcidin25 production in Escherichia coli using SUMO fusion technology. *Gene* **2017**; *610*: 112-117.

113. H. Möller, N. Serttas, H. Paulsen, J. M. Burchell, J. Taylor-Papadimitriou, B. Meyer, NMR-based determination of the binding epitope and conformational analysis of MUC-1 glycopeptides and peptides bound to the breast cancer-selective monoclonal antibody SM3. *FEBS* **2002**; 269: 1444-1455.

6 Appendix

6.1 NMR data of hexapeptide and Hep-25

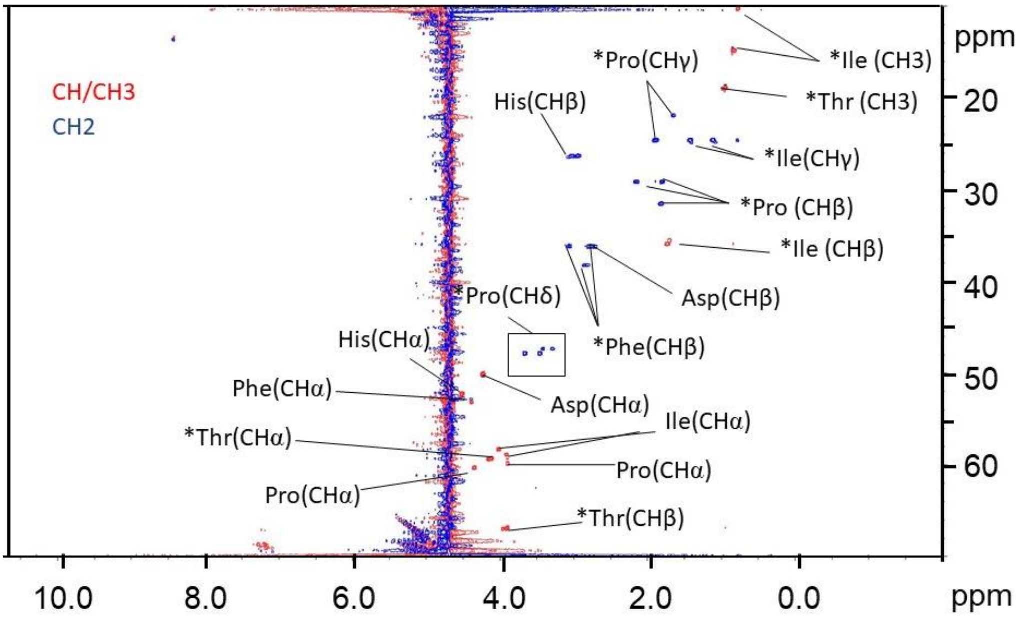


Figure 6.1: ^{1}H-^{13}C HSQC spectrum of hexapeptide DTHFPI. The spectrum shows a 1 mM sample of hexapeptide recorded in 25 mM sodium phosphate, pH 3.2, 10% D2O, T=298K. Positive peaks (CH2) are indicated blue and negative peaks (CH/CH3) are indicated red.

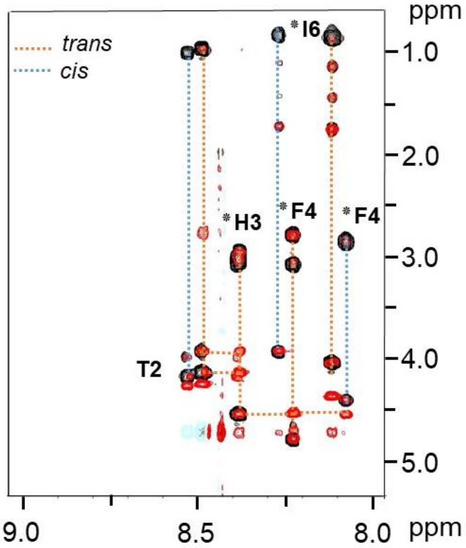


Figure 6.2: Superposition of amid region of ROESY (red) and TOCSY (black) spectra of Hexapeptide at pH 3 and 298K, mixing time 400 ms. * spectra contain a double set of peaks for each amino acid and suggest the presence of cis-trans proline isomerization [44].

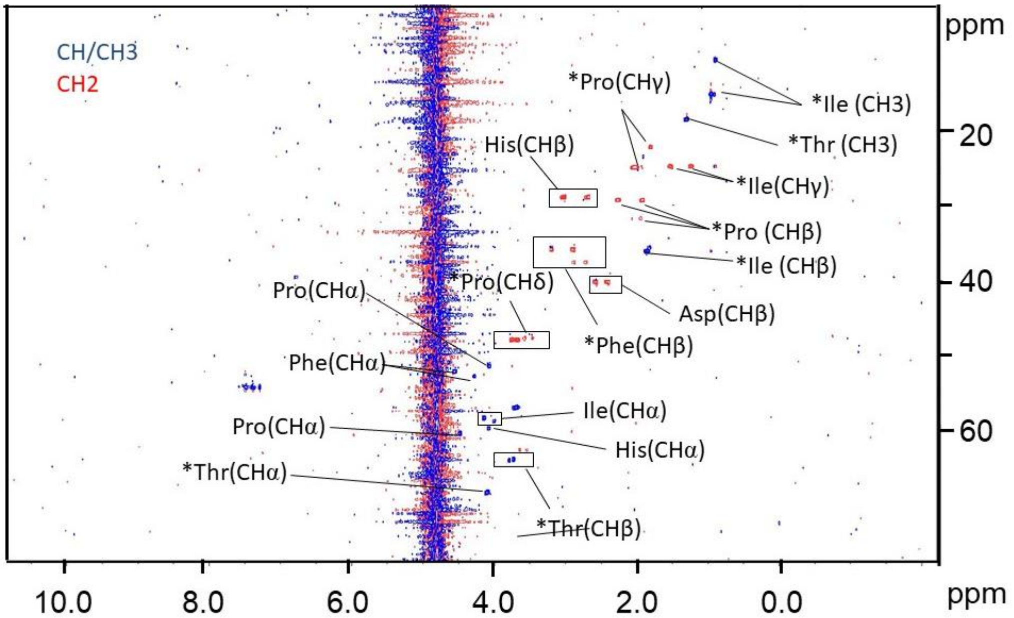


Figure 6.3: ^{1}H-^{13}C HSQC spectrum of hexapeptide DTHFPI. The spectrum shows a 1 mM sample of hexapeptide recorded in 25 mM sodium phosphate, pH 7.4, 10% D2O, T=298K. Positive peaks (CH2) are indicated blue and negative peaks (CH/CH3) are indicated red.

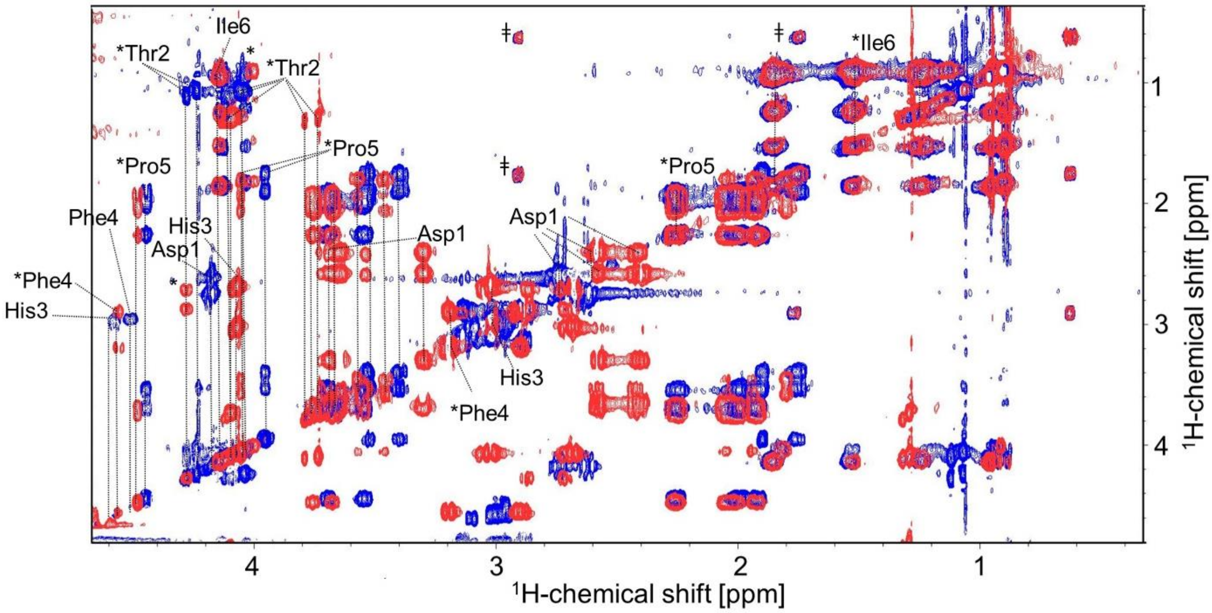


Figure 6.4: ^{1}H-^{1}H TOCSY spectra of the hexapeptide-Ni^{2+} complex [51]. The resonance assignment of the hexapeptide in the presence of Ni^{2+} (1:1 complex, red) and in the absence of Ni^{2+} (blue). * spectra contain a double set of peaks for each amino acid and suggest the presence of cis-trans proline isomerization [44]. ‡ Indicates the presence of the impurities.

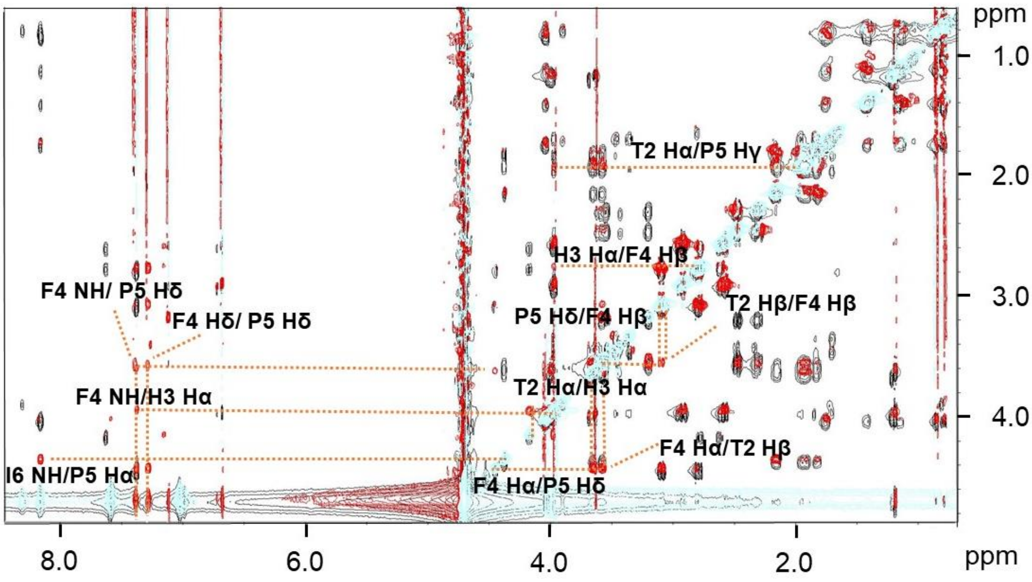


Figure 6.5: Superposition of ROESY (red) and TOCSY (black) spectra of Hexapeptide-Ni(II) complex, mixing time 400 ms.

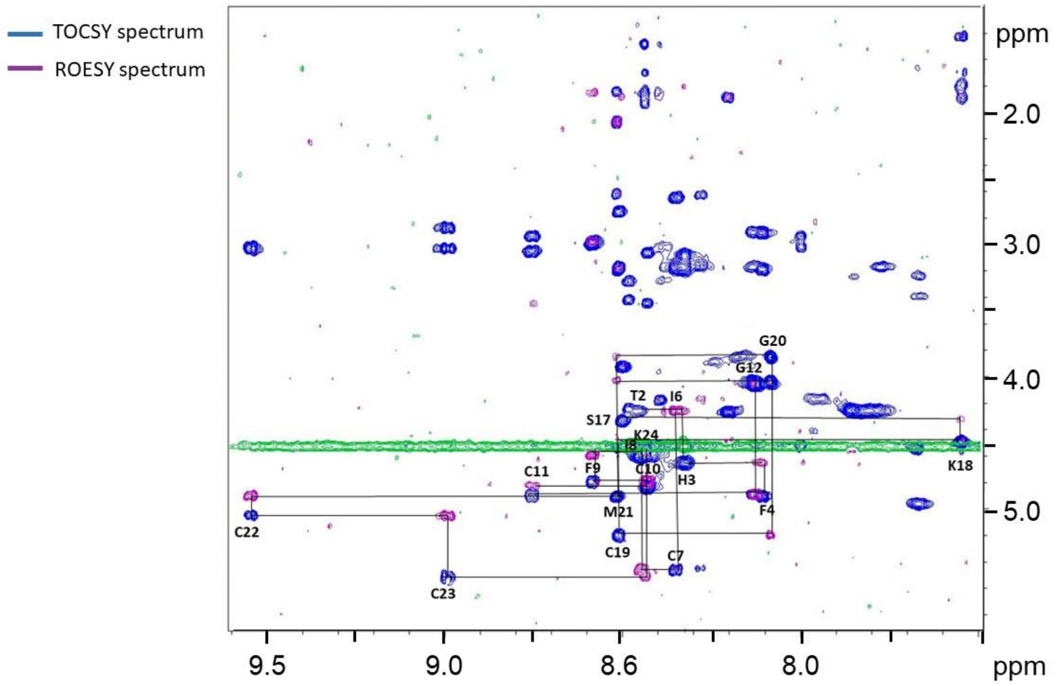


Figure 6.6: Superposition of amid region of ROESY (purple) and TOCSY (blue) spectra of Hep-25 at pH 3 and 325K, mixing time 200 ms.

6.2 Sequences and data of protein constructs

OmpT-Hepcidin-25-His$_8$

MRAKLLGIVLTTPIAISSFAMDTHFPICIFCCGCCHRSKCGMCCKTGHHHHHHHH

Molecular weight: 6.036 kDa

Theoretical pI: 8.75

Extinction coefficient: 500

Expression system: BL21(DE3) Origami B

Plasmid: pQE

Pelb-Hepcidin-His$_8$

MKYLLPTAAAGLLLLAAQPAMDTHFPICIFCCGCCHRSKCGMCCKTGHHHHHHHH

Molecular weight: 6.162 kDa

Theoretical pI: 8.50

Extinction coefficient: 500

Expression system: BL21(DE3) Origami B

Plasmid: pQE

His$_6$-SUMO-Hepcidin-25:

MGSSHHHHHHSSGLVPRGSHMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFF
KIKKTTPLRRLMEAFAKRQGKMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQ
IGDTHFPICIFCCGCCHRSKCGMCCKT

Molecular weight: 16.204 kDa

Theoretical pI: 6.40

Extinction coefficient: 1490

Expression system: BL21(DE3) Origami B

Plasmid: pET15b

His_6-SUMO-Hepcidin-25-DEDEDE:

MGSSHHHHHHSSGLVPRGSHMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFF
KIKKTTPLRRLMEAFAKRQGKMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQ
IGDTHFPICIFCCGCCHRSKCGMCCKTDEDEDE

Molecular weight: 16.937 kDa
Theoretical pI: 5.61
Extinction coefficient: 1490
Expression system: BL21(DE3) Origami B
Plasmid: pET15b

His_6-SUMO-Hepcidin-25-RRRRR:

MGSSHHHHHHSSGLVPRGSHMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFF
KIKKTTPLRRLMEAFAKRQGKMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQ
IGDTHFPICIFCCGCCHRSKCGMCCKTRRRRR

Molecular weight: 16.985 kDa
Theoretical pI: 8.14
Extinction coefficient: 1490
Expression system: BL21(DE3) Origami B
Plasmid: pET15b

His_6-SUMO-Hepcidin-25-SSSSS:

MGSSHHHHHHSSGLVPRGSHMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFF
KIKKTTPLRRLMEAFAKRQGKMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQ
IGDTHFPICIFCCGCCHRSKCGMCCKTSSSSS

Molecular weight: 16:639 kDa
Theoretical pI: 6.40
Extinction coefficient: 1490

Expression system: BL21(DE3) Origami B

Plasmid: pET15b

Hepcidin-25:

DTHFPICIFCCGCCHRSKCGMCCKT

Molecular weight: 2.797 kDa

Theoretical pI: 8.22

Extinction coefficient: 500

Hepcidin-25-DEDEDE:

DTHFPICIFCCGCCHRSKCGMCCKTDEDEDE

Molecular weight: 3.530 kDa

Theoretical pI: 4.63

Extinction coefficient: 500

Hepcidin-25-RRRRR:

DTHFPICIFCCGCCHRSKCGMCCKTRRRRR

Molecular weight: 3.578 kDa

Theoretical pI: 8.14

Extinction coefficient: 500

Hepcidin-25-SSSSS:

DTHFPICIFCCGCCHRSKCGMCCKTSSSSS

Molecular weight: 3.233 kDa

Theoretical pI: 8.22

Extinction coefficient: 500

6.3 Resonance assignment

Table 6.1: Proton chemical shifts of hexapeptide at pH=3, 298K

	H	Hα1	Hβ1	Hβ2	Hδ1	Hδ2	Hε	Hγ	Hγ2
D1									
cis		4.359	2.889						
trans		4.359	2.889						
T2									
cis	8.644	4.260	4.080		1.090				
trans	8.602	4.230	4.010		1.050				
H3									
cis	8.495	4.640	3.124			7.115	8.467		
trans	8.495	4.640	3.124			7.070	8.429		
F4									
cis	8.187	4.308	2.837						
trans	8.341	4.752	3.177	2.888					
P5									
cis		4.009	1.927		3.529	3.404			1.770
trans		4.463	2.720	1.930	3.775	3.571			2.006
I6									
cis	8.380	3.916	1.170	0.812				1.082	1.404
trans	8.228	4.124	1.843	0.940				1.535	1.240

Table 6.2: Carbon chemical shifts of hexapeptide at pH=3, 298K

	Cα	Cβ	Cδ1	Cγ	Cγ2
D1					
cis	50.193	36.140			
trans	50.061	36.206			
T2					
cis	59.384	67.018			
trans	59.430	66.952	18.854		
H3					
cis	52.305	26.350			
trans	52.371	26.310			
F4					
cis		38.186		38.186	
trans		36.272		36.272	
P5					
cis		59.871	31.522	47.422	21.889
trans		60.354	29.147	48.016	24.594
I6					
cis	58.871	35.908	14.764	24.528	24.660
trans	58.375	36.074	14.764	24.594	24.660

Table 6.3: Chemical shifts of Hepcidin-25 at pH=3, 325 K

	NH	H	Hα1	Hα2	Hβ1	Hβ2	Hδ1	Hδ2	Hε	Hγ	Hγ2	Hγ3
D1												
T2	119.28	8.401	4.253		4.000					1.078		
H3	120.10	8.312	4.649		3.159	3.104						
F4	122.64	8.105	4.898		2.905	3.186						
P5						2.190	3.598	3.772				2.009
I6	121.95	8.220	4.264		1.886					1.238	1.605	0.928
C7	124.20	8.339	5.474		3.167	2.639						
I8	115.32	8.449	4.453		1.850					1.108	1.541	0.888
F9	125.22	8.589	4.793		2.987							
C10	122.82	8.428	4.827		3.446	3.055						
C11	120.01	8.779	4.908		3.048	2.932						
G12	112.77	8.124	4.053									
C13	119.65	7.670	4.951		3.228	3.393						
C14												
H15	117.50	8.482	4.211		3.242	3.381						
R16	115.63	7.661	4.560		1.911		1.661			3.269		
S17	111.91	8.514	4.323		3.921							
K18	119.65	7.566	4.492		1.879	1.793	1.696		3.008			1.415
C19	120.07	8.501	5.196		3.174	2.741						
G20	109.75	8.088	4.033	3.851								
M21	119.32	8.503	4.910		1.835	2.072						2.620
C22	126.00	9.553	5.055		3.032							
C23	127.45	9.007	5.528		2.871	3.023						
K24	124.02	8.446	4.576		1.934	1.838	1.700					1.480
T25	119.75	7.886	4.254							1.133		

6.4 Distance and angle restraints

Table 6.4: ROE restraints of Hexapeptide-Ni(II) complex

Amino acid	distances
T2	
assign (T2 - Hα) (T2 - Hβ)	2.5
assign (T2 - Hα) (T2 – Hγ2)	2.9
assign (T2 - Hα) (H3 – Hα)	5.0
assign (T2 - Hβ) (F4 – Hβ1)	3.6
assign (T2 - Hα) (P5 - Hγ)	2.7
assign (T2 - Hβ) (F4 - Hα)	3.3
H3	
assign (H3 - Hα) (H3 - Hβ2)	2.1
assign (H3 - Hα) (H3 - Hβ1)	1.9
assign (H3 - Hα) (F4 - Hβ1)	3.4
assign (H3 - Hβ1) (H3 - Hβ2)	1.8
assign (H3 - Hβ1) (H3 - Hδ1)	2.3
F4	
assign (F4 - Hβ1) (F4 - Hβ2)	1.8
F4 aromatic	
assign (F4 - HN) (F4 - Hα)	2.5
assign (F4 – Hδ2) (F4 - Hα)	2.4
assign (F4 - Hδ2) (F4 - Hβ1)	2.4
assign (F4 - Hδ2) (F4 - Hβ2)	2.7
assign (F4 - HN) (F4 - Hβ1)	3.1
assign (F4 - HN) (H3 - Hα)	3.1
assign (F4 - HN) (P5 - Hδ2)	3.0
assign (F4 – Hδ2) (P5 - Hδ2)	3.3
P5	
assign (P5 - Hα) (P5 - Hβ1)	3.0
assign (P5 - Hβ2) (P5 - Hβ1)	2.2
assign (P5 - Hβ2) (P5 - Hγ2)	5.0

assign (P5 - Hδ2) (F4 - Hβ1)	5.0
assign (P5 - Hδ2) (F4 - Hα)	3.8
assign (P5 - Hδ2) (P5 - Hγ2)	2.8
I6	
assign (I6 - HN) (P5 - Hα)	2.5
assign (I6 - Hα) (I6 - Hβ)	2.5
assign (I6 - Hα) (I6 - Hγ)	3.2
assign (I6 - Hα) (I6 – Hδ1)	2.6
assign (I6 - Hα) (I6 - Hγ2)	3.1
assign (I6 - Hβ) (I6 - Hγ)	2.8
assign (I6 - Hβ) (I6 – Hδ1)	2.4
assign (I6 - Hγ) (I6 - Hγ2)	2.5
assign (I6 - Hγ2) (I6 – Hδ1)	3.1

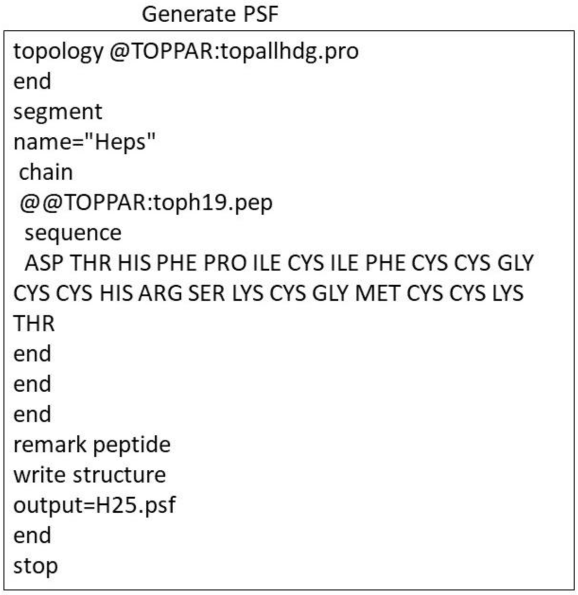
Table 6.5: Dihedral angle restraints

	Dihedral angle
Phi (φ) restraints:	
H3	
assign (T2 - C) (H3 - N) (H3 - Cα) (H3 - C)	70.0
F4	
assign (H3 - C) (F4 - N) (F4 - Cα) (F4 - C)	90
P5	
assign (F4 - C) (P5 - N) (P5 - Cα) (P5 - C)	65.0
Psi (ψ) restraints:	
H3	
assign (H3 - N) (H3 - Cα) (H3 - C) (T2 - N)	-30.6
F4	
assign (F4 - N) (F4 - Cα) (F4 - C) (P5 - N)	125.6
P5	
assign (P5 - N) (P5 - Cα) (P5 - C) (I6 - N)	131.6

Table 6.6: NOE restraints of Hepcidin-25-Ni(II) complex	
Amino acid	distances
assign (C13 – Hβ1) (C13 – Hβ2)	1.81
assign (C11 – Hβ1) (H3 – Hβ2)	3,40
assign (D1 – Hβ1) (D1 – Hβ2)	1.80
assign (C19 – Hβ1) (C11 – Hβ2)	2.04
assign (C7 – Hβ1) (C7 – Hβ2)	1.82
assign (K24 - Hδ) (K24 - Hγ)	4.00
assign (C10 – Hβ1) (C10 – Hβ2)	1.90
assign (K24 - Hε) (K24 - Hδ)	2.50
assign (G20 – Hα1) (G20 – Hα2)	2.16
assign (H3 - Hα) (H3 - Hβ2)	2.20
assign (H3 - Hα) (H3 - Hβ1)	2.01
assign (H3 - Hβ1) (F4 - Hβ1)	1.95
assign (F4 - Hβ1) (F4 - Hβ2)	1.80
assign (P5 - Hβ2) (P5 - Hγ2)	5.00
assign (P5 - Hδ2) (F4 - Hα)	3.80
assign (I6 - Hγ2) (I6 – Hδ1)	4.00

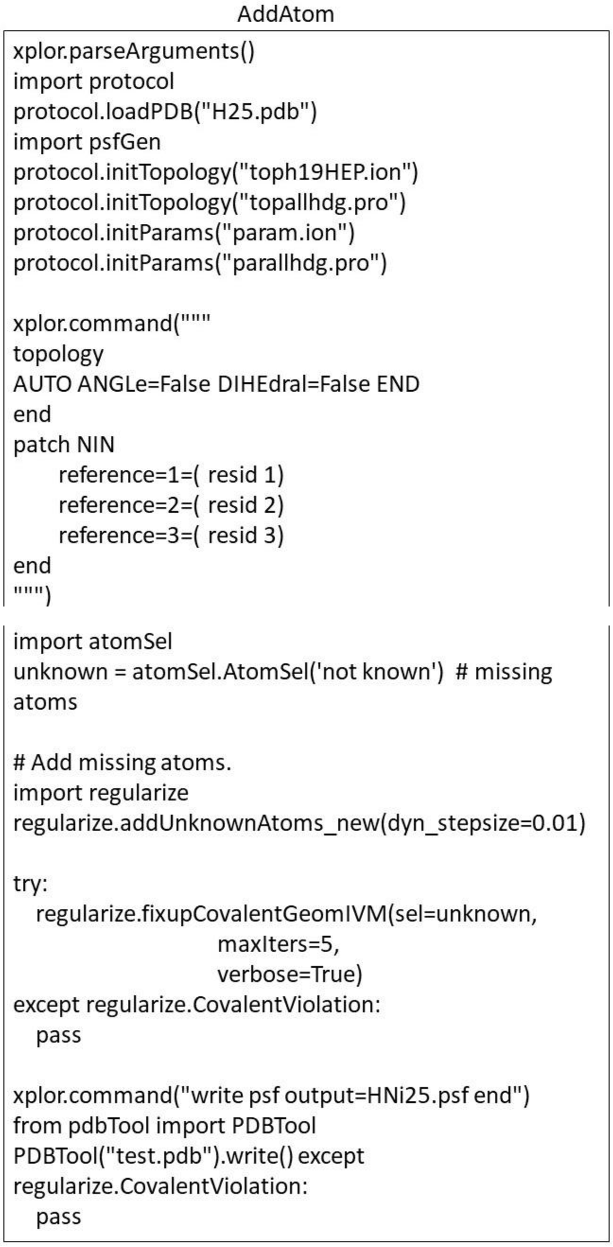
6.5 XPLOR scripts

Generate PSF	AddAtom



Generate PSF

```
topology @TOPPAR:topallhdg.pro
end
segment
name="Heps"
 chain
 @@TOPPAR:toph19.pep
  sequence
  ASP THR HIS PHE PRO ILE CYS ILE PHE CYS CYS GLY
CYS CYS HIS ARG SER LYS CYS GLY MET CYS CYS LYS
THR
end
end
end
remark peptide
write structure
output=H25.psf
end
stop
```



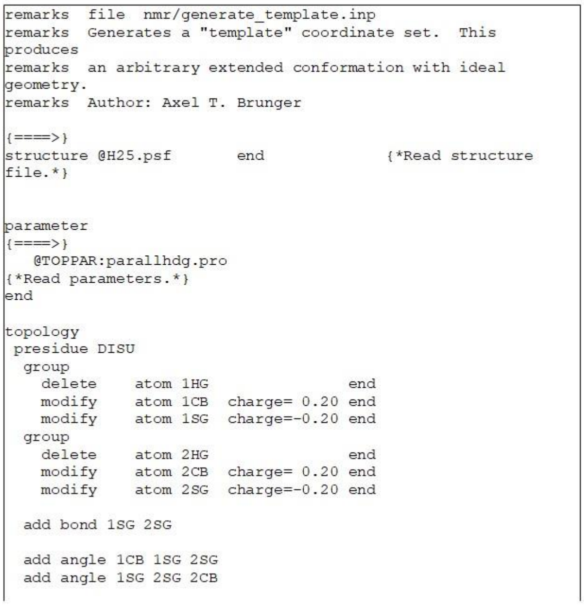
AddAtom

```
xplor.parseArguments()
import protocol
protocol.loadPDB("H25.pdb")
import psfGen
protocol.initTopology("toph19HEP.ion")
protocol.initTopology("topallhdg.pro")
protocol.initParams("param.ion")
protocol.initParams("parallhdg.pro")

xplor.command("""
topology
AUTO ANGLe=False DIHEdral=False END
end
patch NIN
      reference=1=( resid 1)
      reference=2=( resid 2)
      reference=3=( resid 3)
end
""")
import atomSel
unknown = atomSel.AtomSel('not known')  # missing
atoms

# Add missing atoms.
import regularize
regularize.addUnknownAtoms_new(dyn_stepsize=0.01)

try:
   regularize.fixupCovalentGeomIVM(sel=unknown,
                        maxIters=5,
                        verbose=True)
except regularize.CovalentViolation:
   pass

xplor.command("write psf output=HNi25.psf end")
from pdbTool import PDBTool
PDBTool("test.pdb").write() except
regularize.CovalentViolation:
   pass
```

Generate template

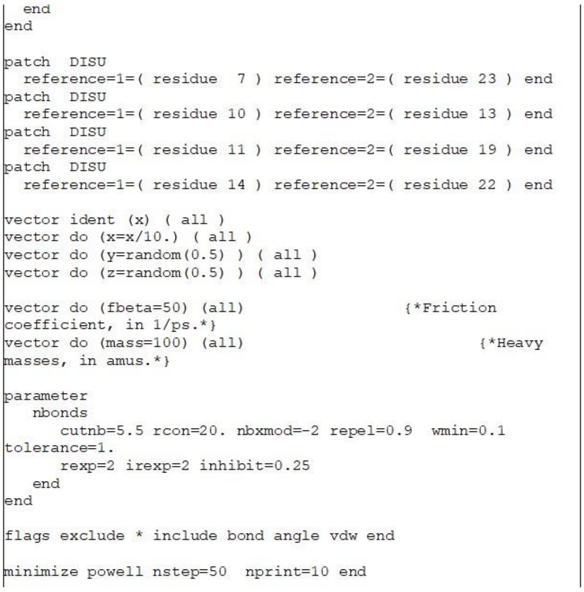
1.

Generate template



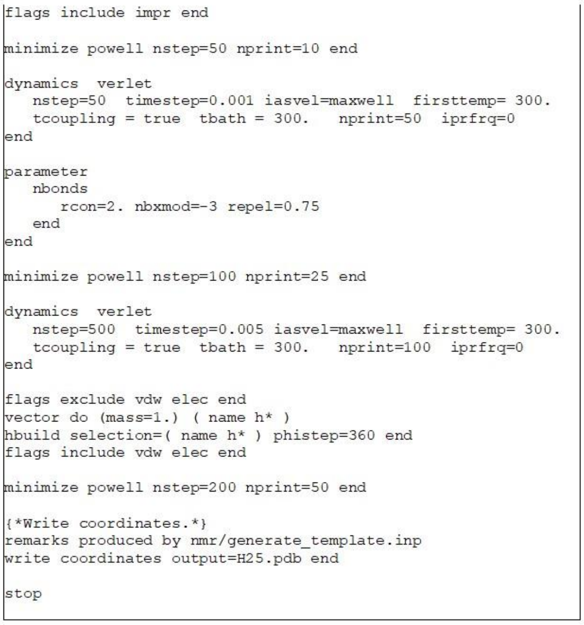
```
remarks  file  nmr/generate_template.inp
remarks  Generates a "template" coordinate set.  This
produces
remarks  an arbitrary extended conformation with ideal
geometry.
remarks  Author: Axel T. Brunger

{===>}
structure @H25.psf       end              {*Read structure
file.*}


parameter
{===>}
   @TOPPAR:parallhdg.pro
{*Read parameters.*}
end

topology
 presidue DISU
  group
    delete    atom 1HG               end
    modify    atom 1CB  charge= 0.20 end
    modify    atom 1SG  charge=-0.20 end
  group
    delete    atom 2HG               end
    modify    atom 2CB  charge= 0.20 end
    modify    atom 2SG  charge=-0.20 end

  add bond 1SG 2SG

  add angle 1CB 1SG 2SG
  add angle 1SG 2SG 2CB
```

2.



```
  end
end

patch  DISU
  reference=1=( residue  7 ) reference=2=( residue 23 ) end
patch  DISU
  reference=1=( residue 10 ) reference=2=( residue 13 ) end
patch  DISU
  reference=1=( residue 11 ) reference=2=( residue 19 ) end
patch  DISU
  reference=1=( residue 14 ) reference=2=( residue 22 ) end

vector ident (x)  ( all )
vector do (x=x/10.)  ( all )
vector do (y=random(0.5) )  ( all )
vector do (z=random(0.5) )  ( all )

vector do (fbeta=50) (all)                 {*Friction
coefficient, in 1/ps.*}
vector do (mass=100) (all)                         {*Heavy
masses, in amus.*}

parameter
   nbonds
      cutnb=5.5 rcon=20. nbxmod=-2 repel=0.9  wmin=0.1
tolerance=1.
      rexp=2 irexp=2 inhibit=0.25
   end
end

flags exclude * include bond angle vdw end

minimize powell nstep=50  nprint=10 end
```

3.

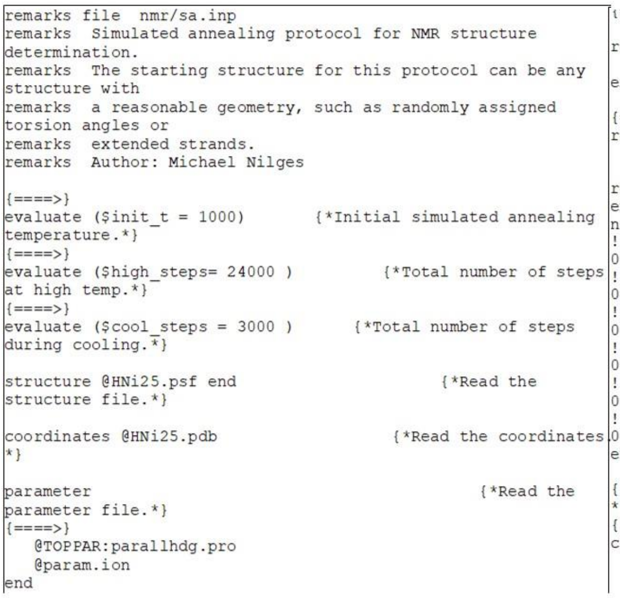


```
flags include impr end

minimize powell nstep=50 nprint=10 end

dynamics  verlet
   nstep=50  timestep=0.001 iasvel=maxwell  firsttemp= 300.
   tcoupling = true  tbath = 300.   nprint=50  iprfrq=0
end

parameter
   nbonds
      rcon=2. nbxmod=-3 repel=0.75
   end
end

minimize powell nstep=100 nprint=25 end

dynamics  verlet
   nstep=500  timestep=0.005 iasvel=maxwell  firsttemp= 300.
   tcoupling = true  tbath = 300.   nprint=100  iprfrq=0
end

flags exclude vdw elec end
vector do (mass=1.) ( name h* )
hbuild selection=( name h* ) phistep=360 end
flags include vdw elec end

minimize powell nstep=200 nprint=50 end

{*Write coordinates.*}
remarks produced by nmr/generate_template.inp
write coordinates output=H25.pdb end

stop
```

Simulated annealing

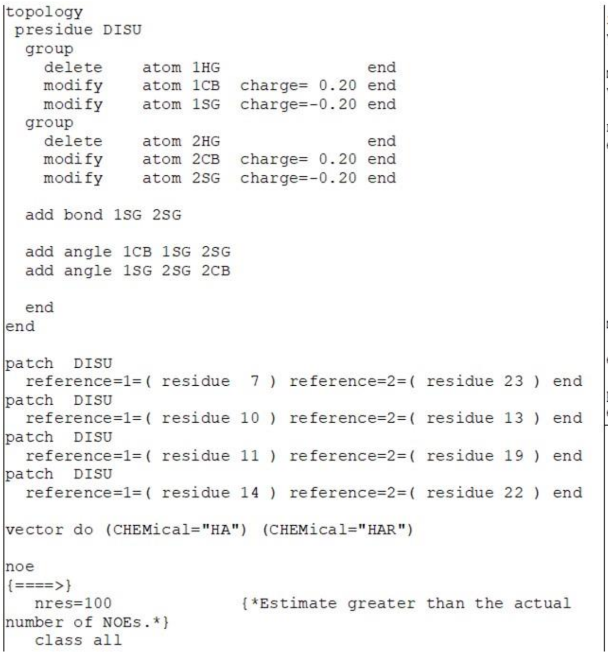
Simulated annealing

1.



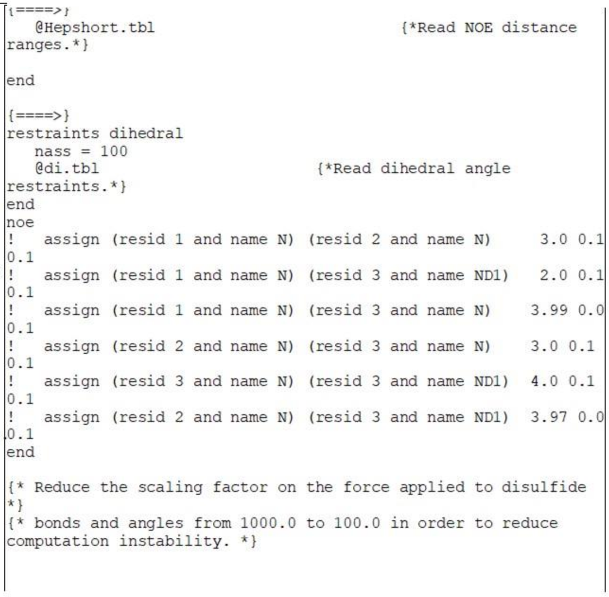
```
remarks file  nmr/sa.inp
remarks  Simulated annealing protocol for NMR structure
determination.
remarks  The starting structure for this protocol can be any
structure with
remarks  a reasonable geometry, such as randomly assigned
torsion angles or
remarks  extended strands.
remarks  Author: Michael Nilges

{====>}
evaluate ($init_t = 1000)       {*Initial simulated annealing
temperature.*}
{====>}
evaluate ($high_steps= 24000 )         {*Total number of steps
at high temp.*}
{====>}
evaluate ($cool_steps = 3000 )      {*Total number of steps
during cooling.*}

structure @HNi25.psf end                     {*Read the
structure file.*}

coordinates @HNi25.pdb                 {*Read the coordinates
*}

parameter                                        {*Read the
parameter file.*}
{====>}
   @TOPPAR:parallhdg.pro
   @param.ion
end
```

2.



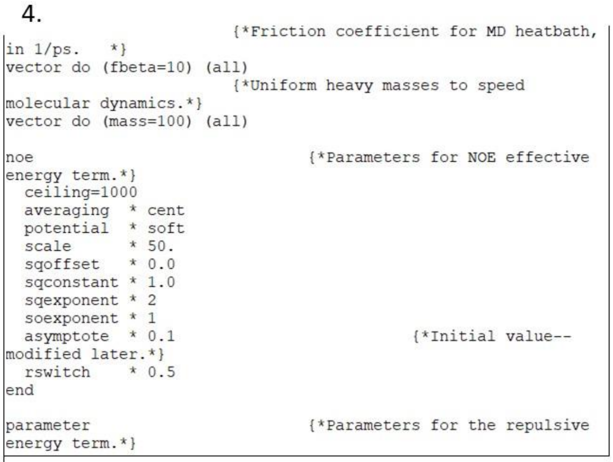
```
topology
 presidue DISU
  group
    delete    atom 1HG                 end
    modify    atom 1CB  charge= 0.20 end
    modify    atom 1SG  charge=-0.20 end
  group
    delete    atom 2HG                 end
    modify    atom 2CB  charge= 0.20 end
    modify    atom 2SG  charge=-0.20 end

  add bond 1SG 2SG

  add angle 1CB 1SG 2SG
  add angle 1SG 2SG 2CB

  end
end

patch  DISU
  reference=1=( residue  7 ) reference=2=( residue 23 ) end
patch  DISU
  reference=1=( residue 10 ) reference=2=( residue 13 ) end
patch  DISU
  reference=1=( residue 11 ) reference=2=( residue 19 ) end
patch  DISU
  reference=1=( residue 14 ) reference=2=( residue 22 ) end

vector do (CHEMical="HA") (CHEMical="HAR")

noe
{====>}
   nres=100              {*Estimate greater than the actual
number of NOEs.*}
   class all
```

3.



```
{====>}
   @Hepshort.tbl                        {*Read NOE distance
ranges.*}

end

{====>}
restraints dihedral
   nass = 100
   @di.tbl                       {*Read dihedral angle
restraints.*}
end
noe
!   assign (resid 1 and name N) (resid 2 and name N)     3.0 0.1
0.1
!   assign (resid 1 and name N) (resid 3 and name ND1)   2.0 0.1
0.1
!   assign (resid 1 and name N) (resid 3 and name N)    3.99 0.0
0.1
!   assign (resid 2 and name N) (resid 3 and name N)    3.0 0.1
0.1
!   assign (resid 3 and name N) (resid 3 and name ND1)  4.0 0.1
0.1
!   assign (resid 2 and name N) (resid 3 and name ND1)  3.97 0.0
0.1
end

{* Reduce the scaling factor on the force applied to disulfide
*}
{* bonds and angles from 1000.0 to 100.0 in order to reduce
computation instability. *}
```

4.

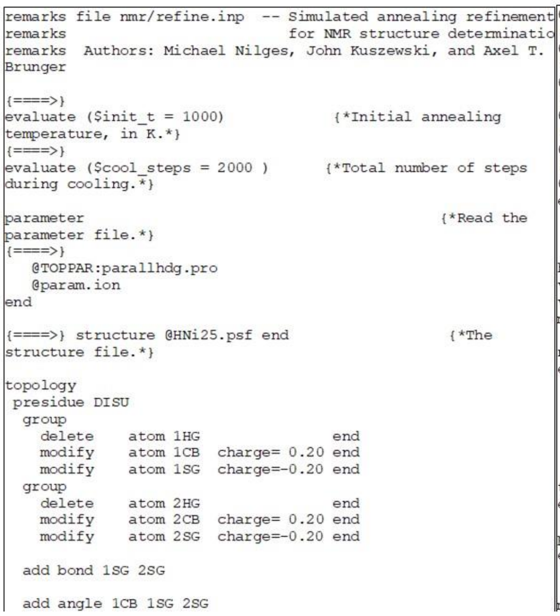


```
                                {*Friction coefficient for MD heatbath,
in 1/ps.   *}
vector do (fbeta=10) (all)
                                {*Uniform heavy masses to speed
molecular dynamics.*}
vector do (mass=100) (all)

noe                           {*Parameters for NOE effective
energy term.*}
  ceiling=1000
  averaging  * cent
  potential  * soft
  scale      * 50.
  sqoffset   * 0.0
  sqconstant * 1.0
  sqexponent * 2
  soexponent * 1
  asymptote  * 0.1                       {*Initial value--
modified later.*}
  rswitch    * 0.5
end

parameter                     {*Parameters for the repulsive
energy term.*}
```

Simulated annealing refinement

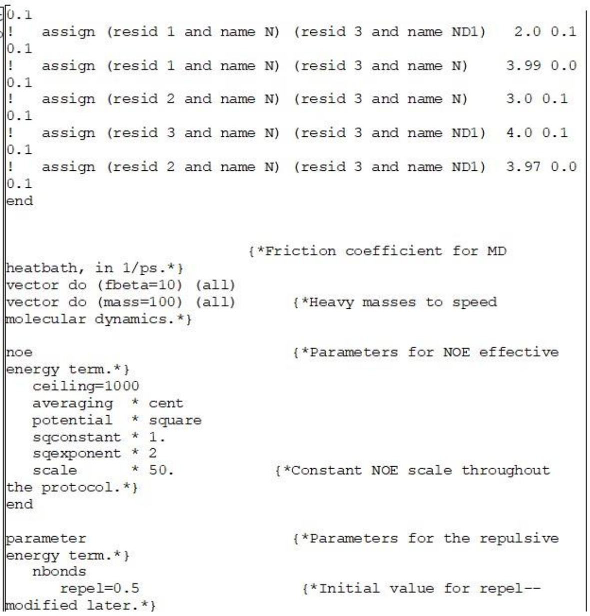
Simulated annealing refinement

1.



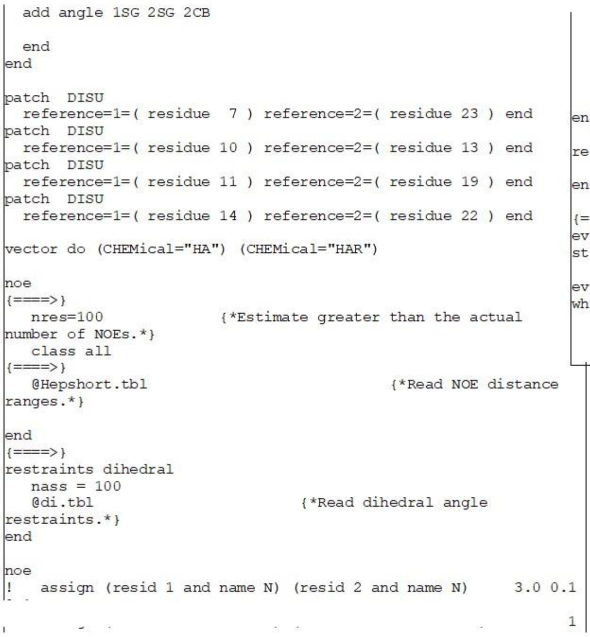
```
remarks file nmr/refine.inp  -- Simulated annealing refinement
remarks                         for NMR structure determinatio
remarks  Authors: Michael Nilges, John Kuszewski, and Axel T.
Brunger

{===>}
evaluate ($init_t = 1000)            {*Initial annealing
temperature, in K.*}
{===>}
evaluate ($cool_steps = 2000 )      {*Total number of steps
during cooling.*}

parameter                                        {*Read the
parameter file.*}
{===>}
   @TOPPAR:parallhdg.pro
   @param.ion
end

{===>} structure @HNi25.psf end                   {*The
structure file.*}

topology
 presidue DISU
  group
    delete    atom 1HG              end
    modify    atom 1CB  charge= 0.20 end
    modify    atom 1SG  charge=-0.20 end
  group
    delete    atom 2HG              end
    modify    atom 2CB  charge= 0.20 end
    modify    atom 2SG  charge=-0.20 end

  add bond 1SG 2SG

  add angle 1CB 1SG 2SG
```

3.



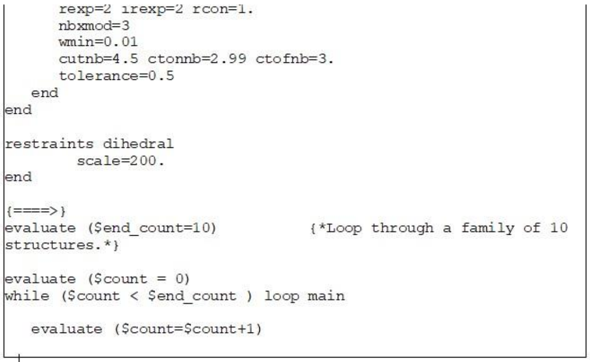
```
0.1
!   assign (resid 1 and name N) (resid 3 and name ND1)   2.0 0.1
0.1
!   assign (resid 1 and name N) (resid 3 and name N)    3.99 0.0
0.1
!   assign (resid 2 and name N) (resid 3 and name N)    3.0 0.1
0.1
!   assign (resid 3 and name N) (resid 3 and name ND1)  4.0 0.1
0.1
!   assign (resid 2 and name N) (resid 3 and name ND1)  3.97 0.0
0.1
end


                           {*Friction coefficient for MD
heatbath, in 1/ps.*}
vector do (fbeta=10) (all)
vector do (mass=100) (all)      {*Heavy masses to speed
molecular dynamics.*}

noe                             {*Parameters for NOE effective
energy term.*}
   ceiling=1000
   averaging  * cent
   potential  * square
   sqconstant * 1.
   sqexponent * 2
   scale      * 50.          {*Constant NOE scale throughout
the protocol.*}
end

parameter                       {*Parameters for the repulsive
energy term.*}
   nbonds
      repel=0.5                  {*Initial value for repel--
modified later.*}
```

2.



```
  add angle 1SG 2SG 2CB

  end
end

patch  DISU
  reference=1=( residue  7 ) reference=2=( residue 23 ) end
patch  DISU
  reference=1=( residue 10 ) reference=2=( residue 13 ) end
patch  DISU
  reference=1=( residue 11 ) reference=2=( residue 19 ) end
patch  DISU
  reference=1=( residue 14 ) reference=2=( residue 22 ) end

vector do (CHEMical="HA") (CHEMical="HAR")

noe
{===>}
   nres=100            {*Estimate greater than the actual
number of NOEs.*}
   class all
{===>}
   @Hepshort.tbl                        {*Read NOE distance
ranges.*}

end
{===>}
restraints dihedral
   nass = 100
   @di.tbl                     {*Read dihedral angle
restraints.*}
end

noe
!   assign (resid 1 and name N) (resid 2 and name N)     3.0 0.1

                                                              1
```

4.



```
      rexp=2 irexp=2 rcon=1.
      nbxmod=3
      wmin=0.01
      cutnb=4.5 ctonnb=2.99 ctofnb=3.
      tolerance=0.5
   end
end

restraints dihedral
        scale=200.
end

{===>}
evaluate ($end_count=10)          {*Loop through a family of 10
structures.*}

evaluate ($count = 0)
while ($count < $end_count ) loop main

   evaluate ($count=$count+1)
```