Cloning, expression and purification of the different human haptoglobin chains and initial characterization by mass spectrometry

Camille Lombard

The University of Toledo

Follow this and additional works at: http://utdr.utoledo.edu/theses-dissertations

Recommended Citation
Lombard, Camille, "Cloning, expression and purification of the different human haptoglobin chains and initial characterization by mass spectrometry" (2013). Theses and Dissertations. 138.
http://utdr.utoledo.edu/theses-dissertations/138

This Thesis is brought to you for free and open access by The University of Toledo Digital Repository. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of The University of Toledo Digital Repository. For more information, please see the repository's About page.
A Thesis
entitled
Cloning, Expression and Purification of the Different Human Haptoglobin Chains
and Initial Characterization by Mass Spectrometry
by
Camille Lombard
Submitted to the Graduate Faculty as partial fulfillment of the requirements for the
Master of Science Degree in Chemistry

_________________________________________
Dr. Wendell P. Griffith, Committee Chair

_________________________________________
Dr. Timothy C. Mueser, Committee Member

_________________________________________
Dr. Max O. Funk, Committee Member

_________________________________________
Dr. Patricia R. Komuniecki, Dean
College of Graduate Studies

The University of Toledo
May 2013
Haptoglobin (Hp) is an acute phase glycoprotein that exists in all vertebrates and can be found in the majority of body fluids. It is phenotypically variant in humans and presents in two allelic forms: Hp1 and Hp2. The two forms vary by the identity of their light chains: L₁ or L₂ for Hp1 and Hp2, respectively. Hp is most well known for its function in rapidly binding to free hemoglobin (Hb), which is subsequently removed from circulation, making this protein very important in the prevention of Hb-based oxidative damage and a natural bacteriostat. Despite much research efforts in the past few decades, there is little information available on the structure of human Hp and its complex with Hb. This is in large part attributed to its large size, high degree of glycosylation, structural heterogeneity, and difficulty to obtain (due to it being an acute phase protein). Recombinantly producing Hp and its individual chains in E. coli solves all of these problems.

This thesis presents the cloning, expression, purification, and initial characterization of the individual chains of human Hp: L₁, L₂, and H. The cloning, expression and purification of L₁ and L₂ were successful. Correct protein sequences were
confirmed using trypsin in-gel digestion, MALDI-TOF mass spectrometry, and tandem mass spectrometry. Cloning for the H-chain is ongoing as of the publication date of this thesis. Using chemical modification by iodoacetamide and charge state distributions in ESI mass spectra, results support L₁ contained the correct disulfide pattern, while L₂ did not. In order to obtain the proper folding and the correct disulfide pattern for L₂, it will be necessary to introduce a renaturation step.

In addition to the Hp project, this thesis included work done to elucidate the regions of disorder in the protein DnaT. DnaT is one of the primosomal proteins that is an essential component of the PriA replication restart pathway. Very little is known about this protein, its structure, and its actual mode of function. Several attempts to crystallize this protein have been met with inconclusive results thought to be due to regions of intrinsic disorder in the protein and degradation during the crystallization process. We were able to identify the cleavage sites in the degradation of the protein by sequence analysis and ESI MS of fresh and degraded protein samples. Using analysis of the charge state distributions in ESI MS, we provide definitive evidence of the intrinsic disorder of \textit{E. coli} DnaT. Through collision-induced dissociations studies and the top-down approach, we were able to identify that the N-terminal part of the protein was the flexible regions. This allowed the proposal of a sequence construct lacking this flexible region for future crystallization studies.
Acknowledgements

I would first like to thank my advisor, Dr. Wendell Griffith, for giving me the opportunity to work on this challenging yet incredibly interesting project. I would like to thank my committee members, Dr. Mueser and Dr. Funk, for their advice and expertise.

I am also thankful for the support from my family back home in France and my husband Nathan Banek. I must express my gratitude to my labmates both past and present, with a special thought for Kate who helped me a lot in operating the mass spectrometer instrument and kept me entertained. I also acknowledge my friends here in Toledo, especially the Amanda’s lab for the wonderful lunch times; as well as my friends back home.

Overall it was an excellent experience and it would not have been the same without all of these people.
Table of Contents

Abstract.................................................................................................................. iii
Acknowledgements ................................................................................................. v
Table of Contents ...................................................................................................... vi
List of Tables ........................................................................................................... xi
List of Figures.......................................................................................................... xii
List of Abbreviations ............................................................................................... xv
1 Introduction.......................................................................................................... 1
  1.1 Haptoglobin....................................................................................................... 1
  1.1 DnaT.................................................................................................................. 4
  1.2 Project Goals .................................................................................................... 5
    1.2.1 Haptoglobin.................................................................................................. 5
    1.2.2 DnaT .......................................................................................................... 6
  1.3 Organization of this Thesis ............................................................................... 6
2 Mass Spectrometry: Principles and Instruments Design .................................. 8
  2.1 Principles of Mass Spectrometry ..................................................................... 9
    2.1.1 Typical Instrument Construction ............................................................... 9
    2.1.2 Definition and Principle............................................................................. 11
    2.1.3 The Typical Mass Spectrum .................................................................. 13
## 2.2 Ionization Methods

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.1</td>
<td>Electrospray Ionization (ESI)</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Matrix-Assisted Laser Desorption/Ionization (MALDI)</td>
</tr>
</tbody>
</table>

## 2.3 Mass Analyzers

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3.1</td>
<td>Time of Flight Mass Analyzer</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Hybrid Quadrupole Time-of-Flight Mass Analyzer</td>
</tr>
<tr>
<td>2.3.3</td>
<td>TOF-TOF Mass Analyzer</td>
</tr>
</tbody>
</table>

## 2.4 Ion Activation Methods

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4.1</td>
<td>Overview of Ion Activation Methods</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Collision Induced Dissociation</td>
</tr>
<tr>
<td>2.4.3</td>
<td>Peptide and Protein Fragmentation and Fragment Nomenclature</td>
</tr>
</tbody>
</table>

## 2.5 Ion Mobility Mass Spectrometry

## 2.6 Instrumentation Used in this Research

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.6.1</td>
<td>Synapt HDMS (Waters Corp.)</td>
</tr>
<tr>
<td>2.6.2</td>
<td>UltrafleXtreme MALDI-TOF/TOF Mass Spectrometer</td>
</tr>
</tbody>
</table>

## 3 Applications of Mass Spectrometry to Protein Identification, Protein Structure and Protein Dynamics
3.2.3 Mapping of Disulfide Bridging Patterns and their Verification for Recombinant Proteins ................................................................. 35
3.2.4 Noncovalent Protein Complexes ...................................................... 37
3.2.5 MS as a Tool for Protein Crystallography ........................................ 39

4 Cloning, Expression and Purification of the Individual Subunits of Human Haptoglobin Chains ................................................................. 42

4.1 Material and Methods ................................................................................. 43
4.1.1 List of Materials and Reagents .............................................................. 43
4.1.2 Polymerase Chain Reaction (PCR) ......................................................... 44
4.1.3 Insertion of the GOI into the entry vector ............................................. 45
4.1.4 Insertion of the GOI into the expression vector ..................................... 46
4.1.5 Protein Expression ................................................................................ 47
4.1.6 Cell Lysis .............................................................................................. 48
4.1.7 Purification of Hp Chains ...................................................................... 49
4.1.8 Electrophoresis .................................................................................... 50
4.1.9 Trypsin Digestion ................................................................................ 52
4.1.10 In Silico Digest ................................................................................... 53
4.1.11 MALDI-TOF/TOF MS Experiments ................................................. 53

4.2 Results and Discussion ............................................................................. 54
4.2.1 Protein Cloning, Expression and Purification of the Light Chains (L1 and L2) of Human Haptoglobin ......................................................... 54
4.2.2 Protein Cloning of the H-chain of Human Haptoglobin ....................... 65

4.3 Conclusion ............................................................................................... 71
5 Initial Characterization of the Haptoglobin Chains by Mass Spectrometry ........... 72

5.1 Materials and Methods .................................................................................. 73

5.1.1 Sample Preparation for Mass Spectrometry Analysis ............................... 73

5.1.2 Reaction for Mapping of the Disulfide Bridges ........................................ 73

5.1.3 In-solution Trypsin Digestion ..................................................................... 74

5.1.4 Mass Spectrometry Analysis ....................................................................... 74

5.2 Results and Discussion .................................................................................. 75

5.2.1 Analysis of the Intact Light Chains ............................................................ 75

5.2.2 Establishing the Disulfide Pattern of the Light Chains .............................. 76

5.2.3 Molecular Modeling .................................................................................... 91

5.3 Conclusion ..................................................................................................... 95

6 Protein Disorder Determination of DnaT by Mass Spectrometry ................. 96

6.1 Material and Methods .................................................................................. 98

6.1.1 Protein Expression ....................................................................................... 98

6.1.2 Lysis and Ammonium Sulfate Precipitation ............................................. 98

6.1.3 Purification of DnaT ................................................................................... 99

6.1.4 Sample Preparation for Mass Spectrometry Experiments ....................... 100

6.1.5 Mass Spectrometry Experiments ............................................................... 101

6.2 Results and Discussion ................................................................................ 101

6.2.1 Expression and Purification of DnaT ......................................................... 101

6.2.2 DnaT: Degradation Products ................................................................. 103

6.2.3 Identification of the Flexible Domain of DnaT ........................................ 108

6.2.4 Study and Identification of DnaT Conformers in Solution ...................... 115
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Overview of some activation methods used in mass spectrometry</td>
<td>20</td>
</tr>
<tr>
<td>4.1</td>
<td>Reagents for the PCR reactions</td>
<td>45</td>
</tr>
<tr>
<td>4.2</td>
<td>List of the reagents used for the pENTR™/D-TOPO® reaction</td>
<td>45</td>
</tr>
<tr>
<td>4.3</td>
<td>List of reagents and conditions for the LR Cloning™ reaction</td>
<td>46</td>
</tr>
<tr>
<td>4.4</td>
<td>Composition of the buffer used for the cell lysis after expression of the haptoglobin subunits</td>
<td>48</td>
</tr>
<tr>
<td>4.5</td>
<td>Buffers composition for Ni affinity purification of Hp chains.</td>
<td>50</td>
</tr>
<tr>
<td>4.6</td>
<td>List of the peptides used to confirm expression products of L₁</td>
<td>59</td>
</tr>
<tr>
<td>5.1</td>
<td>List of the major peptides detected</td>
<td>81</td>
</tr>
<tr>
<td>5.2</td>
<td>List of the major peptides detected</td>
<td>84</td>
</tr>
<tr>
<td>5.3</td>
<td>List of the major peptides detected</td>
<td>88</td>
</tr>
<tr>
<td>5.4</td>
<td>List of the major peptides detected</td>
<td>90</td>
</tr>
<tr>
<td>6.1</td>
<td>Composition of Lysis buffer</td>
<td>99</td>
</tr>
<tr>
<td>6.2</td>
<td>Composition of TEDG buffer</td>
<td>99</td>
</tr>
<tr>
<td>6.3</td>
<td>Composition of HPLC purification buffers</td>
<td>100</td>
</tr>
<tr>
<td>6.4</td>
<td>Masses of the DnaT fragments correlated to their possible sequence</td>
<td>105</td>
</tr>
</tbody>
</table>
List of Figures

1-1 Schematic of the structure of the different human Hp phenotypes. ........................................ 3

2-1 Schematic of a typical mass spectrometer .................................................................................. 12

2-2 Schematic of the typical mass spectrometer ............................................................................ 13

2-3 Mechanism of the ion formation in ESI. Adapted from Banerjee et al. 41 .............................. 14

2-4 Biemann nomenclature of the most common peptide fragments for a generic tripeptide. ................................................................................................................................. 23

2-5 Schematic of the ESI Q/oa-TOF Synapt HDMS system from Waters Corp. (Reproduced with permission from the manufacturer.) ................................................................. 26

2-6 Schematic of the Bruker Daltonics UltraFleX MALDI TOF/TOF MS from Bruler Daltonics. (Adapted from Bruker Daltonics) ........................................................................ 27

3-1 Comparison of the simulated charge state distributions of an intrinsically disordered protein (left) and a folded protein (right) in solution under non-denaturing conditions. ... 33

4-1 SYBR GOLD™ stained 1% DNA agarose gels of L1 chains cloning steps......................... 54

4-2 DNA sequence of the newly synthesized expression clones...................................................... 55

4-3 Expression test for L1. The SDS PAGE gel was stained with Coomasie blue™ ......... 56

4-4 Sequences of Hp light chains including the His-tag and TEV protease site.............. 57

4-5 MALDI-TOF/TOF MS of in-gel tryptic digest of L1-chain ...................................................... 58

4-6 Tandem MALDI-TOF/TOF MS of the m/z 1708.88 from the trypsin digest of L1 chain.................................................................................................................................................. 59
4-7 Solubility test for L1 ........................................................................................................... 60
4-8 Ni-NTA purification chromatogram of L1 chain................................................................. 61
4-9 Sequences of Hp light chains including the His-tag and TEV protease site. .............. 62
4-10 MALDI-TOF/TOF MS of the tryptic digest of L1 after His-tag removal. ................. 64
4-11 SYBR GOLD™ stained 1% DNA agarose gels of H chains cloning steps.............. 65
4-12 Expression test for H-chain. The SDS gel was stained with Coomasie blue™. ...... 67
4-13 Sequences of Hp heavy chains including the His-tag and TEV protease site. ........... 67
4-14 MALDI-TOF/TOF MS of in gel tryptic digest of H-chain. ......................................... 68
4-15 DNA sequence alignment between the sequenced DNA sequence of H chain gene inserted into pDEST-C1 vector (top) and the theoretical one (bottom)....................... 69
4-16 SYBR GOLD™ stained 1% DNA agarose gels of H chains cloning steps.............. 70
5-1 ESI-MS of intact L1 chain. ..................................................................................................... 76
5-2 Schematic of the workflow for mapping of the disulfide bonds .................................... 77
5-3 Amino acid sequence of recombinant L1. ............................................................. 78
5-4 Amino acid sequence of recombinant L2........................................................................ 79
5-5 ESI-MS of the tryptic digest of L1 after reaction with TCEP followed by IAA. ........ 80
5-6 Tandem MS of the m/z 1136.5392 peak (z = +3) after MaxEnt3 transformation ....... 82
5-7 ESI-MS of the tryptic digest of L1 after reaction with IAA. ........................................ 83
5-8 Tandem MS of the m/z 1675.1041 peak (z = +4) after MaxEnt3 transformation...... 85
5-9 Tandem MS of the m/z 1170.6300 peak (z = +2) after MaxEnt3 transformation.... 86
5-10: ESI-MS of the tryptic digest of L2 after reaction with TCEP followed by IAA. .... 87
5-11 Tandem MS of the m/z 748.8302 peak (z = +2) after MaxEnt3 transformation.... 89
5-12 ESI-MS of the tryptic digest of L2 after reaction with IAA .................................. 89
5-13 Molecular modeling of haptoglobin chains. A: L₁, B: L₂ and C: H ......................... 92
5-14 Cartoon representation of the model of human Hp1 ............................................ 93
5-15 Surface representation of the model of the H-chain of human Hp ....................... 94
6-1 Chromatogram of DnaT purification by Q-sepharose ion exchange .................. 102
6-2 Chromatogram of DnaT purification by POROS HS ion exchange .................... 102
6-3 SDS PAGE of purified DnaT after POROS HS chromatography ....................... 103
6-4 Top: ESI-MS of the degraded DnaT sample ...................................................... 104
6-5 ESI-MS of freshly purified DnaT in 50% ACN containing 0.1% FA, 10 mM AmmAc pH 3 and 10 mM AmmAc pH7 ................................................................. 107
6-6 Schematic showing how CID MS experiments can be used to identify flexible regions in proteins ............................................................................................................. 108
6-7 ESI-MS of intact DnaT in ammonium acetate pH7 ............................................ 110
6-8 Tandem mass spectrum of the species at m/z 2149.044 (z = +9) ..................... 110
6-9 Representation of the fragmentation pattern obtained from the native species ..... 111
6-10 ESI-MS of the intact DnaT protein in denaturing solution ................................. 112
6-11 Tandem mass spectrum of m/z 1215.2977 (z = +16) specie (peak 1) .......... 112
6-12 Tandem mass spectrum of m/z 1208.6812 (z = +16) species (peak 2) .......... 113
6-13 Tandem mass spectrum of m/z 2416.2156 (z = +8) species (peak 4) ........ 113
6-14 Tandem mass spectrum of m/z 1631.4274 (z = +5) species (peak 5) ........ 114
6-15 Truncated DnaT sequence selected for new molecular cloning ....................... 114
6-16 Ion mobility plots of DnaT under different solution conditions .................... 116
List of Abbreviations

°C ......................... Degree celcius
%(w/v) .................... % weight/volume
%(v/v) ....................... % volume/volume

ACN ....................... acetonitrile
AIMS ..................... Aspiration ion mobility spectrometry
AmmAc ..................... Ammonium acetate
Amp ....................... Ampicillin
ATP ....................... Adenosine triphosphate

bp ...................... Base pair
BME ...................... β-mercaptoethanol

CAM ...................... Chloramphenicol
CID ........................ Collision-induced dissociation
CV ......................... Column volume

DC ......................... Direct current
DIMS ...................... Drift-time ion mobility spectrometry
DMS ....................... Differential mobility spectrometry
DNA ....................... Deoxyribonucleic acids
dNTP ...................... Deoxyribonucleoside triphosphates
dsDNA .................... Double-stranded DNA
DTT ....................... Dithiothreitol

ECD ....................... Electron capture dissociation
E. coli .................... Escherichia coli
EDTA ...................... Ethylenediaminetetraacetic acid
ESI ........................ Electrospray ionization
ETD ....................... Electron transfer dissociation
eV ........................ Electronvolt
\(\varepsilon_x\) .................... Extinction coefficient at x nm

FA ........................ Formic acid
FAIMS ..................... Field assymetric waveform ion mobility spectrometry
FTICR ..................... Fourier transform ion cyclotron resonance

GHz ....................... Gigahertz
gp-4..........................glycoprotein 4
g...............................gram

h...............................Hour
Hb.............................Hemoglobin
HBV..........................Hepatitis B virus
HDMS..........................High definition mass spectrometry
hHp...........................Human haptoglobin
Hp.............................Haptoglobin
Hz.............................Hertz

IGD..........................In gel digestion
IMMS..........................Ion mobility mass spectrometry
IMS............................Ion mobility spectrometry
IPTG..........................Isopropyl-β-D-thiogalactopyranoside
IRMPD........................Infrared multiphoton dissociation

Kan..........................Kanamycin
kbp..........................kilo base pair
Kd.............................Constant of dissociation
kDa..........................Kilodalton
kHz..........................Kilohertz
KOD..........................Kodakaraensis

LID..........................Laser induced dissociation
LB.............................Luria Broth
L..............................Liter

M.............................Molar
MDa..........................Mega Dalton
MALDI........................Matrix assisted laser desorption ionization
mg..........................Milligram
min..........................Minute
mL..........................Milliliter
μL..........................Microliter
mM..........................Millimolar
μM..........................Micromolar
MS............................Mass spectrometry
m/z..........................Mass/charge ratio

NMR..........................Nuclear magnetic resonance
nano-ESI.....................Nano-electrospray

oa-TOF........................Orthogonal time-of-flight
ODx..........................Optical density at x nm
Pa..........................Pascal
PCR..........................Polymerase chain reaction
PDB..........................Protein data bank
PEI..........................Polyethylenimine
pHp..........................Porcine haptoglobin
pI..........................Isoelectric point
PMSF..........................Phenylmethylsulfonyl fluoride
ppm..........................parts per million
PSD..........................Post-source decay
PTM..........................Post translational modification

Q..........................Quadrupole
QIT..........................Quadrupole ion trap
QqQ..........................Triple quadrupole
Q/TOF..........................Quadrupole time-of-flight
R..........................
RF..........................Radio frequency
rpm..........................Rotation per minute
RT..........................Room temperature
rTOF..........................Reflectron time-of-flight

SDS-PAGE.....................Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SID..........................Surface-induced dissociation
SSB..........................Single stranded binding protein
Strep..........................Streptomycin

TBP..........................TATA-box binding protein
TCEP..........................Tris-2-carboxyethylphosphine
TEDG..........................Tris, EDTA, DTT, glycerol
TFA..........................Trifluoroacetic acid
TFIIB..........................Transcription factor II B
Th..........................Thomson
TIS..........................Time ion selector
T_m.........................Melting temperature
TOF..........................Time of flight
TOF/TOF......................Hybrid time of flight/ time of flight
Tris..........................Tris(hydroxymethyl)aminomethane
TWIMS..........................Travelling wave ion mobility mass spectrometry

UV-Vis......................Ultraviolet-visible
Chapter 1

Introduction

1.1 Haptoglobin

The abundance of hemoglobin (Hb) in humans (ca. 1020 molecules per liter of blood) is a necessity caused by respiratory requirements, but it also poses a grave danger. Oxyhemoglobin, which is a highly oxidizable molecule, spontaneously undergoes an intramolecular oxidation-reduction reaction to generate the nonfunctional methemoglobin and biologically heinous superoxide.\(^1\) Indeed, any Hb molecule not confined within a red blood cell and its constituent antioxidant defense systems would have an oxidative potential that could wreak oxidative havoc to any of the surrounding tissue.\(^2\) Also, any free Hb in the blood quickly dissociates in to its constituent dimers and monomers, which the body tries to clear from circulation via the kidneys, which may result in severe renal damage. With the development of hemoglobin-based artificial blood as a viable and safer alternative to blood transfusion as well as the increasing world-wide incidence of diseases like malaria, free hemoglobin toxicity has become a very important issue. The free hemoglobin also serves as a source of iron for many pathogenic bacteria.\(^3\)
There is currently one known “oxidative damage control” mechanism by which vertebrates can remove this toxic free hemoglobin from circulation: haptoglobin. Haptoglobin (Hp) rapidly binds to free hemoglobin, which is subsequently transported to the liver and catabolized. Because of its importance in the prevention of hemoglobin-based oxidative damage and as a natural bacteriostat, Hp has been a target of extensive research efforts in the past. Despite much research efforts in the past few decades, there is only available a dearth of information on the structure of human Hp and its complex with Hb in large part due to its large size, high degree of glycosylation and structural heterogeneity.

Haptoglobin belongs to a group of genetically polymorphic serum α-2-glycoproteins that are present in all vertebrates and found in the majority of body fluids. In its simplest form, haptoglobin is a tetrachain structure that contains two light chains (designated as L₁ and L₂) and two heavy chains (designated as H), which are all connected by disulfide bridges with the two H-chains connected via the two L-chains between them. N-linked glycosylation accounts for approximately 20% of the mass of Hp and only occurs at 4 sites on the H-chains. The H-chains also contain the sites for binding to hemoglobin.³,⁵,⁶

Most vertebrates exhibit the simplest phenotypic forms.⁷,⁸,⁹,¹⁰,¹¹ Human Hp is present under two allelic forms: Hp1 and Hp2. The two forms vary by the identity of their light chains. The light chain in Hp2 (L₂) possesses one additional free cysteine residue than the light chain in Hp1 (L₁), which can form a disulfide bridge with another H/L dimer.¹² These two allelic forms give rise to three possible phenotypes: Hp1-1, Hp2-1 and Hp2-2. Due to the extra free cysteine residue on L₂, Hp2-1 and Hp2-2 form higher
order polymers (Figure 1-1).12, 13, 14 Humans are the only mammals having different forms of haptoglobin.15 The distribution of the different allelic forms, thus of the various phenotypes, varies upon geographic location.11, 16

![Diagram of Hp phenotypes](image)

**Figure 1-1:** Schematic of the structure of the different human Hp phenotypes. The legend is represented in the figure. Only the two simplest polymeric forms of Hp2-2 and Hp2-1 are represented.

Until recently, very little was known about the structure of Hp and its complex with Hb. For the former, structural information were obtained through molecular modeling based on the sequence homology of Hp with serine proteases.17, 18, 19, 20 For the later, electron micrography data provided insight on the shape and the size of the complex.14 It was only last year that Andersen et al. solved the structure of porcine Hp in complex with Hb. Their findings were consistent with the early electron micrography
Hb binds as a heterodimer to Hp and the binding is mainly triggered by hydrophobic interactions on a large area on the surfaces of both counterparts, accounting for the low dissociation constant ($K_d \sim 10^{-15}$).\textsuperscript{3,21}

Hp has a number of other functions besides scavenging Hb. Hp has been shown to have intrinsic anti-oxidant potential.\textsuperscript{22,4} Its antioxidant activity has been demonstrated to be greater than that of vitamin E.\textsuperscript{22} More importantly, it has been demonstrated that Hp has ATP independent chaperone-like activity, resembling that of clusterin\textsuperscript{23} and prevents amyloids formation through interaction with prefibrillar protein species.\textsuperscript{24}

\section*{1.1 DnaT}

All the genetic information of any organisms is contained within its DNA, which encodes for the proteins responsible for the coordination of all cellular processes. Herein, survival and effective reproduction of an organism is highly dependent on its ability to replicate its genetic code. The bacterial genetic code is contained in a large, circular double-stranded piece of DNA. The replication is a semiconservative and bidirectional process; both strands of the DNA serve as templates for polymerase III. The elongation of the DNA chain is polarized. The deoxyribonucleotides (dNTPs) are added in the 5’ to 3’ direction. The initiation of the replication in \textit{E. coli} is sequence specific and occurs at a region called \textit{OriC}.\textsuperscript{25}

Notwithstanding, DNA damage can occur through a variety of mechanisms resulting in replication fork arrest.\textsuperscript{26} In order to ensure the survival of the organisms, several DNA repair mechanisms exist. Once the DNA has been repaired replication must restart in an origin independent manner. One of the machinery involved in the replication
restart pathway in *E. coli* is called Φ174-type primosome. It provides for the helicase and primase activity for replication.\footnote{27,28} Eight proteins are involved in the assembly. They are PriA, PriB, PriC, DnaB, DnaC, DnaG, DnaT and single stranded DNA binding proteins (SSB). The first protein to be assembled is PriA, which recognizes single-stranded DNA, such as a D-loop. It is structurally specific. The rest of the assembly occurs in the following order. PriB is loaded onto PriA to facilitate the complex formation with DnaT. Loading of DnaB is assisted by DnaC and requires the hydrolysis of one molecule of ATP. Subsequently DnaG is loaded onto the assembly and can start the synthesis of the primers in the lagging strand of the DNA.\footnote{29,30,31}

One of the necessary proteins composing this assembly is DnaT. It has been shown that DnaT is an essential component of the PriA replication restart pathway.\footnote{32} Cells containing a DnaT mutation are temperature-sensitive for growth and DNA synthesis. This indicates that DnaT is an important constituent for primosome formation and thus chromosomal replication.\footnote{33} Although, little is known about this protein, neither its actual function nor its structure have yet been determined. Very recently, Szymanski *et al.* showed that the oligomeric state of DnaT might be a trimer,\footnote{34,35} but previous experiments in our lab showed only the existence of dimer.

### 1.2 Project Goals

#### 1.2.1 Haptoglobin

As human Hp exists in three phenotypic forms, obtaining commercial samples of a single one can be non-cost effective. Additionally, it is an acute phase protein, making
it not readily available. Finally, Hp has quite a large size (100 kDa for the smallest form) and is heavily glycosylated, which represent a challenge for studies using common bioanalytical techniques. Hence, it was decided to obtain human Hp recombinantly. To eliminate the problem of size, the different subunits were cloned separately.

The goal of this project was to clone and express the recombinant subunits of human Hp and to characterize these expression products by mass spectrometry.

1.2.2 DnaT

DnaT has been studied by the Mueser laboratory for a number of years and several attempts to crystallize the protein have been met with inconclusive results. The difficulty in crystallizing the protein is thought to be due to regions of intrinsic disorder in the protein and degradation during crystallization. The goals of the study of DnaT are to elucidate the regions of disorder and their effect on the protein structure. Collision induced dissociation was used to address the first problem and ion mobility mass spectrometry for the second.

1.3 Organization of this Thesis

This thesis work has been divided into seven chapters, each addressing a specific aspect of the project. In chapters 2 and 3, the principles of mass spectrometry, the instrumentation and various applications relevant to this work are presented. Chapter 4 describes the cloning and expression of the various chains of human Hp, as well as the
challenges encountered. The fifth chapter presents the initial characterization of the two light chains as well as the conclusions that can be made about the structures of the recombinant chains. Finally, the last chapter presents the study of the intrinsically disordered domain of DnaT.
Chapter 2

Mass Spectrometry: Principles and Instruments Design

The study of the structure and structural dynamics of proteins and protein assemblies is of prime interest. In recent years mass spectrometry (MS) has become a very prominent bioanalytical technique in the analyses of these types of analytes. One unique feature about mass spectrometry is its ability to detect and characterize individual conformational states of proteins and protein complexes that may co-exist in solution at equilibrium. With recent developments and improvements in mass spectrometric instrumentation, these analyses can now achieve sensitivities of detection in the sub-femtomolar range, requiring only modest amounts of protein for analysis. This is much unlike other commonly used techniques like Nuclear Magnetic Resonance Spectroscopy (NMR) and X-ray crystallography. Consequently, mass spectrometry can allow for the analysis of proteins, protein complexes, and protein structural dynamics at concentrations at or even below endogenous levels. Ease of use as well as speed of analysis has also increased the attractiveness of this technique in biomolecular analysis. The application of mass spectrometry instrumentation is in large part dependent on components including the ionization source, mass analyzer. Ion activation and the capability for tandem mass spectrometry, and additional features like ion mobility serve to markedly increase the
applicability of this technique. This chapter will present the principles, instrument design and details on the operation of the mass spectrometers specific to this thesis research.

2.1 Principles of Mass Spectrometry

2.1.1 Typical Instrument Construction

As mentioned above, the typical mass spectrometer consists of an ionization source, followed by a mass analyzer and finally a detector (Figure 2-1). These components are generally kept under high vacuum. There are though some instrument designs where ionization occurs at atmospheric pressure, for example electrospray ionization (ESI) as will be discussed later.

Ionization, or conversion of gas phase analyte molecules into ions, is one of the most important processes in MS. This is accomplished in the source of the mass spectrometer. A number of ionization sources/methods are currently available for MS. The choice of source is dependent of the nature of the analyte and the type of information about the analyte that is sought. For example, electron ionization or chemical ionization is commonly used for small organic molecules, which tend to be thermally stable. More thermally labile molecules, such as biomolecules (proteins or peptides) require much “softer” ionization methods to ensure that they do not decompose during the ionization process. 36, 37

Two types of direct ionization sources exist: liquid phase and solid-state sources. In the former, the analytes are in solution and introduced as droplets into the source at atmospheric pressure. In the later, the analyte is typically mixed with a non-volatile
deposit or matrix, which can be a solid or a viscous liquid. The deposit is then irradiated by either energetic particles or photons to achieve desorption of the ions. Good examples of the liquid phase and solid-state direct ionization sources of interest to this thesis research are electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI), respectively. These will be discussed in detail later in this chapter.

Once the ions have been produced, they are separated in the mass analyzer according to their mass-to-charge ratio (m/z). The five main characteristics that define the performance of the mass analyzer are mass range limit, speed of analysis, ion transmission, mass accuracy and resolution. The mass range is the range of m/z values that can be measured by the mass analyzer. The speed of analysis represents the rate at which the mass analyzer can measure over a defined mass range. Ion transmission reflects the efficiency of the mass analyzer to transmit ions from the source to the detector. It is measured by the ratio of number of ions reaching the detector to the number of ions entering the mass analyzer. Mass accuracy, which represents how good the instrument is in determining a particular mass, is measured by the difference between the theoretical (or calculated) mass and the mass measured. Finally, resolution reflects the ability of the mass analyzer to separate two species with similar m/z values.

There are two main categories of mass analyzers: scanning analyzers, which allow for the transmission of selected ions; and simultaneous transmission mass analyzers which allow all the ions to go through at the same time. Several types of these mass analyzers are commercially available and present a range of performances. Nonetheless, it is worth noting that the mass analyzer having the overall best combination of
performance, resolution (into the millions) and mass accuracy (less than 5 ppm is possible) is Fourier transform ion cyclotron (FTICR). The unequalled performance of this mass analyzer though comes at great financial cost and it also has a limited mass range. Another interesting mass analyzer is time-of-flight (TOF), which offers a virtually unlimited mass range but with only a fraction of the high resolution and mass accuracy achievable with FTICR. However, by combining different mass analyzers, it is possible for one to exploit their advantageous qualities while reducing some liabilities. For instance, the transmission quadrupole mass analyzer provides poor resolution and mass accuracy while time-of-flight provides considerably higher. Neither of the two mass analyzers are capable of tandem mass spectrometry. Hybrid mass analyzers, connected in tandem, such as the triple quadrupole (QqQ), the TOF/TOF, and the Q-TOF all provide facile ion activation for fragmentation. The Q-TOF and TOF/TOF combinations also provide comparably higher resolutions and improved mass accuracy than the single transmission quadrupole or the simple time-of-flight alone. The Q-TOF and TOF/TOF mass analyzers will be described in greater details later in this chapter.  

2.1.2 Definition and Principle

A mass spectrometer is an instrument that is capable of converting gaseous molecules into ions, accelerating them into an electric field, separating them according to their mass-to-charge ratio (m/z, or Thomson Th), and detecting the relative amounts of each species. In mass spectrometry, the analyte must be ionized in order to be detected. The ions are produced in an ion source where the samples are vaporized, desolvated and ionized. These ions are then focused and transferred to a mass analyzer, which separates
them prior to detection. Finally, a computer or data processing system creates a suitable mass spectrum. A schematic of the typical design of a mass spectrometer is presented in Figure 2-1.36

![Schematic of a typical mass spectrometer](image)

**Figure 2-1:** Schematic of a typical mass spectrometer

The mean free path is the distance an ion can travel through the mass analyzer without collision with another gaseous molecule. The gas phase ions should have the highest possible mean free path, as the ions produced are often metastable. Thus instruments must function under high vacuum (10^-4 - 10^-6 Pa), as higher pressure leads to unwanted gaseous molecules, which could modify the path of the ions during transmission or mass analysis. This may result in fragmentation of the analyte through collisions thereby preventing the detection of the intact species.36
2.1.3 The Typical Mass Spectrum

In a typical mass spectrum, the detector signal in terms of ion abundance is plotted with respect to their mass-to-charge ratio \((m/z)\). On the y-axis, the ion abundance can be represented as either relative intensity (which is normalized relative to the most intense peak or base peak) or absolute ion count. As is shown in Figure 2-2, the height or the area under the peak in the mass spectrum provides quantitative information with the necessary controls. On the other hand, the position of the peak on the \(m/z\) scale provides qualitative information from which the mass of the analyte species can be determined (Figure 2-2).\(^{36}\) As will be presented in Chapter 3, the charge state distributions for proteins and other biomolecules (and hence where the series of peaks lie on the \(m/z\) scale) can give an indication as to the degree of unfolding or structural flexibility. This turns out to be an extremely useful tool in investigating the effect of denaturation, destabilization and intrinsic disorder in protein structures.

![Figure 2-2: Schematic of the typical mass spectrometer.](image)
2.2 Ionization Methods

2.2.1 Electrospray Ionization (ESI)

Electrospray ionization (ESI), which is the most widely used technique in the field of biomolecular analysis by mass spectrometry, was introduced in 1988 by John Fenn. ESI is a so-called “soft” ionization method, which produces multiply charged ions while preventing them from fragmentation during the ionization process. One key advantage of producing multiply charged ions is that it allows ESI to overcome the mass range limitation of some mass analyzers such as the transmission quadrupole. ESI is a very versatile ionization technique as it is suitable not only for the study of proteins, but also for a wide range of analytes including small molecules and nanoparticles.

In standard ESI, the sample is dissolved in a polar volatile solvent at low concentration (nM to μM range). The analytes are introduced into the instrument via a syringe, through a metal capillary with a weak flux (1 - 20 μL/min). A high voltage (2 - 6 kV) is applied at the tip of the capillary creating an electric field. The strong electric

![Figure 2-3: Mechanism of the ion formation in ESI. Adapted from Banerjee et al.](image-url)
field creates a charge accumulation on the surface of the liquid exiting the tip. It results in the formation of a Taylor cone followed by the production of a spray of highly charged droplets. The precursor droplets continue to shrink through solvent evaporation until the charge per unit volume is too high and reach a point that is known as the Rayleigh limit, when the surface tension (that holds the droplet together) is equal to the cumbic repulsion. This causes the droplets to explode creating new generations of offspring droplets. The process is repeated until completely desolvated gas phase ions result. The ESI source is usually kept at atmospheric pressure, thus a continuous gradient of pressure is used from the source to the detector in order to maintain a high vacuum into the analyzer.\cite{37,41} Despite the fact that ESI is widely used, the exact mechanism for its operation is not yet fully understood. In recent years, though a number of research groups have made considerable contributions to increased understanding of the chemical and physical processes that occur in ESI during the analyses of various types of analytes.\cite{42,43}

Recent developments to ESI achieved in the last decade have led to the creation of micro- and nano-electrospray sources, which have much lower flux (0.2 - 1 µL/min) than traditional ESI sources. The metallic capillary is replaced by a borosilicate glass capillary with a tip diameter of 1 - 4 µm. The initial droplets produced by nano-ESI are subsequently much smaller. Along with a much lower flow rate, nano-ESI allows for the reduction of sample consumption, improves the desolvation process, and increases sensitivity of detection. In large part to the reduction of the initial droplet size, nano-ESI allows for the use of aqueous solvents and/or lower temperature for the desolvation process, which facilitates the analysis of labile biological molecules and complexes.\cite{44}
2.2.2 Matrix-Assisted Laser Desorption/Ionization (MALDI)

MALDI, an improvement to the already existing Laser Desorption/Ionization (LDI), was introduced by Karas and Hillenkamp in 1987.\textsuperscript{45} Shortly after, in 1988, Tanaka demonstrated the use of this technique for the analysis of intact protein molecules exceeding 100 kDa.\textsuperscript{46} Compared with other soft ionization methods, MALDI benefits from easy sample preparation and a higher tolerance to nonvolatile salts, buffers and other contaminants.

As previously mentioned, MALDI was developed as an improvement to Laser Desorption/Ionization (LDI), which exhibited major fragmentation of analytes in excess of 500 Da. In both techniques the sample is deposited onto a metal support and desorbed by mean of laser irradiation. The successes of MALDI in desorption and ionization of large molecules resides in the use of a large excess of a matrix, which is co-crystallized with the analyte. MALDI matrices should meet the following requirements: they must absorb light in the UV range (having aromatic groups); they must be miscible and cocrystallize with the analyte; and they are typically acidic and capable of donating protons to the analyte.\textsuperscript{37}

MALDI presents several advantages for the analysis of biological samples. As MALDI produces mainly singly charged ions, it is a suitable technique for the analysis of complex mixtures. Also, as mentioned previously, MALDI can be more tolerant to contaminants such as glycerol, urea or nonvolatile salts that can be present in biological samples, and thus these can be analyzed without further desalting. One drawback to MALDI is that low molecular weight analytes (< 400 g/mol) produce peaks in the same region of the mass spectrum as the matrix peaks. As a result, these small molecule
analytes are likely to not be detected by this method. MALDI is a so-called pulsed ionization technique, which makes it well suited for combination with pulsed mass analyzers like time-of-flight (TOF).\textsuperscript{37}

2.3 Mass Analyzers

2.3.1 Time of Flight Mass Analyzer

Time-of-flight (TOF) mass analyzers were first described by Stephens in 1946.\textsuperscript{47} Conceptually, they are the simplest of all mass analyzers. Ions are separated, after initial acceleration, based on the time they take to go through a field free region called the “flight tube”. A packet containing ions with a range of $m/z$ values, but having the same kinetic energy, enters the flight tube. Each ion of a particular $m/z$ value travels through the flight tube with a specific velocity that depends on the mass. TOF mass analyzers are pulsed and are consequently most often paired with pulsed ionization sources like MALDI. However, ion optics features make TOF suitable for combination with a range of other ionization sources.\textsuperscript{38,48}

TOF mass analyzers are considered to have virtually no upper limit in mass range, which makes them well suited for the analysis of very large biomolecules. The sensitivity of TOF is typically in the range of pico to femtomoles. However, their resolution is low due to kinetic energy spread among ions with same $m/z$. Two major techniques have been employed to overcome this disadvantage. The first method is delayed pulse extraction. The ions are submitted to a short field free region before they are subjected to the initial extraction pulse. The second method is to introduce an electrostatic reflector, called
“reflectron”, after the flight tube. The reflectron, which consists in a series of equally spaced grid electrodes, is situated opposite the ion source and reflects ions orthogonally to an off-axis detector. Ions with higher velocity will penetrate deeper into the field of the reflectron, while slower ions will consequently spend much less time there. This corrects the dispersion in the initial kinetic energies of the ions leaving the ionization source with the same \( m/z \) resulting in their reaching the detector simultaneously. By essentially almost doubling the flight time of the ions, the reflectron greatly increases the resolution of the mass analysis, but at the expense of sensitivity of detection.\(^{38}\)

2.3.2 Hybrid Quadrupole Time-of-Flight Mass Analyzer

The most common quadrupole time-of-flight (Q-TOF) mass analyzers results from coupling electrostatic focusing devices in the source with a transmission quadrupole mass analyzer, a gas filled collision cell, and a time-of-flight mass analyzer. The transmission quadrupole mass analyzer functions in RF-only mode for MS experiments, which means that all of the ions whose \( m/z \) value falls within a particular range are transmitted.\(^{49}\)

In tandem MS experiments, on the other hand, this same mass analyzer is used to select and transmit the precursor ion of interest. Ions are then accelerated by a potential difference between 20 and 200 eV before they enter the collision cell, a second quadrupole filled with collision gas (typically He or Ar). Thereafter, the ions enter the TOF analyzer, typically an orthogonal TOF (oa-TOF). Initially, no voltage is applied to the repeller situated at the entrance of the TOF. The ions are pushed toward the analyzer in a direction orthogonal to their original trajectories following the application of a pulse
voltage to the repeller thereby creating a perpendicular electric field. Prior to entering the flight tube, the packet of ions is further accelerated by the orthogonal accelerator. Once the packet of ions is in the flight tube, no more voltage is applied to the repeller and another packet of ions starts filling the orthogonal accelerator for the subsequent scan. The flight cycle for each scan ends when the ion with the highest m/z reaches the detector after which a new cycle, with a subsequent packet of ions, starts by applying a pulse voltage to the repeller.

Q-TOF hybrid mass analyzers have been shown to be powerful tools in the sequencing of peptides and proteins. More importantly for this thesis research, they have well-documented success in the analysis of macromolecular assemblies such as non-covalent protein complexes. This thesis presents a number of applications of the Q-TOF mass spectrometry in protein analysis.

2.3.3 TOF-TOF Mass Analyzer

Hybrid TOF/TOF mass analyzers consist of two successive TOF mass analyzers with a collision chamber in between. The first TOF is a linear TOF while the second often has reflectron TOF (rTOF) capabilities. During tandem MS experiments, the ions are analyzed a first time by the first TOF analyzer. Selection of precursor ions and fragmentation are achieved by an ion gate and the collision cell, respectively. The fragments are thereafter analyzed by the second TOF mass analyzer. TOF/TOF mass spectrometers are also predominantly and conveniently used in proteomics analyses.38
2.4 Ion Activation Methods

2.4.1 Overview of Ion Activation Methods

The development of soft ionization techniques allows for the transfer into the gas phase and subsequent mass analysis of intact large molecules with little fragmentation allowing for the determination of the masses of these analytes, but thereby limiting the amount of structural information possible. Consequently, tandem MS rose as a necessary technique for the structural analysis of various types of analyte molecules. The crucial point of these experiments is the ion activation step. Ion activation methods can be divided in two categories: low energy and high energy activation. A list of some these is presented in Table 2.1. The table shows a few of the commonly used activation methods, their associated energy level, and the mass analyzers for which they are possible.

Table 2.1: Overview of some activation methods used in mass spectrometry

<table>
<thead>
<tr>
<th>Method</th>
<th>Method Acronym</th>
<th>Energy Range</th>
<th>Type of Mass Analyzer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-source decay</td>
<td>PSD</td>
<td>Low</td>
<td>RE-TOF</td>
</tr>
<tr>
<td>Collision-induced dissociation</td>
<td>CID</td>
<td>Low</td>
<td>QqQ, IT, RE-TOF, QqLIT, FTICR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High</td>
<td>TOF-TOF, sectors</td>
</tr>
<tr>
<td>Surface-induced dissociation</td>
<td>SID</td>
<td>Low</td>
<td>QqQ, IT, FTICR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High</td>
<td>TOF-TOF, RE-TOF</td>
</tr>
<tr>
<td>Electron capture dissociation</td>
<td>ECD</td>
<td>Low</td>
<td>FTICR, IT</td>
</tr>
<tr>
<td>Electron transfer dissociation</td>
<td>ETD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infrared multiphoton dissociation</td>
<td>IRMPD</td>
<td>Low</td>
<td>IT, FTICR</td>
</tr>
</tbody>
</table>

Early experiments into fragmentation of analytes involved the study of metastable ion self-decomposition. Indeed, in MALDI, the ions produced are often metastable and
decompose through a mechanism called post-source decay (PSD). This phenomenon can be amplified by increasing the energy of the laser, a method which is known as laser-induced dissociation (LID). Several other ion activation techniques have been developed over the years with collision-induced dissociation (CID) being the most widespread. CID involves the collision of the molecules with a target gas and will be discussed in greater detail later in this chapter. Other techniques that do not involve the use of higher pressure (collision gas), which can compromise the low vacuum pressure of the instrument have been developed.\textsuperscript{51}

One such method, which is closely related to CID as far as the type of fragments produced is surface-induced dissociation (SID) where the ions collide with a solid surface, usually composed of an inert material. Other techniques such as infra-red multiphoton dissociation (IRMPD) and electron capture dissociation (ECD) or electron transfer dissociation (ETD) involve the interaction of the analyte with either multiple photons (from a laser) or electrons, respectively. IRMPD involves the excitation of the ions by multiple photons. It has been performed by using lasers of different wavelengths. ECD and ETD are both only currently available on trapping analyzers such as FTICR or the quadrupole ion trap (QIT).\textsuperscript{50, 51}

### 2.4.2 Collision Induced Dissociation

Collision-induced dissociation or CID is the most common activation method and can be achieved on most of the current mass spectrometer formats. After precursor ion selection, the internal energy of the precursor ions is increased through collision with inert gas contained within a chamber known as the CID cell. Fragmentation of the
precursor ions occurs if the energy of the collision is sufficiently high that the ions are excited beyond their threshold for dissociation. The CID process can be divided in two categories: low energy collisions or high energy collisions.\textsuperscript{51}

Low-energy collisions are found mainly in quadrupole or ion trap type instruments. The energy of the collisions is around 1 - 100 eV. The gas pressure is an important factor as the degree of dissociation of higher mass ions increases as the gas pressure increases. An increase in mass of the neutral gas target will have the same effect (for example using Ar as opposed to He). Ion activation times of approximately tens of milliseconds can be used, thus increasing the number of collisions. This activation technique falls into the category of “slow-heating” processes, as its time scale is relatively long. Despite the lower energy input per collision, the product ion yields are high due to the occurrence of multiple collisions.\textsuperscript{51}

High-energy collisions are common to TOF and sector instruments where the precursor ions have a high translational energy. The energy of the collisions is in the keV range. The product ions obtained from the two possible energy regimes for CID are often different. For instance, high-energy collisions of peptides in a four-sector instrument yield principally \(v\) and \(w\) ions as there is an increase in side chain fragmentation. Low-energy CID yields mainly \(b\)- and \(y\)- ions and the spectra obtained are usually not as complex. These two techniques can thereby be seen as complementary considering the different information that each provides.\textsuperscript{51}
2.4.3 Peptide and Protein Fragmentation and Fragment Nomenclature

Fragmentation of peptides and proteins allows for the determination of their sequence and gaining insight into their structure. Upon activation, peptide ions fragment producing two possible fragments: those derived from the cleavage of one or two bonds on the chain itself, and those that also undergo cleavage of the amino acid lateral chain. The nomenclature suggested by Biemann in the early 90s allows for the characterization of the various fragment-types that can be obtained (Figure 2-4).

![Biemann nomenclature of the most common peptide fragments for a generic tripeptide.](image)

In the main chain of a typical peptide there are three types of bonds can be cleaved: Cα-C; C-N or N-Cα. N-terminal fragments are labeled a_n, b_n or c_n, respectively for these cleavage sites if the positive charge remains at the N-terminal side of the peptide. Similarly, C-terminal fragments are classified as x_n, y_n or z_n, respectively for these cleavage sites if the positive charge remains at the C-terminal side (Fig 2-4). The mass difference between each consecutive ion in a ladder series allows for the amino acid
sequence to be determined. Indeed, most of the 20 amino acid residues have a singular elemental composition and thus distinctive masses. Exceptions are the isomeric residues Ile and Leu and the isobaric Gln and Lys. It is impossible to differentiate between the isomeric residues just based on mass. Additionally, sufficiently high resolution is required to discriminate between Gln and Lys. The type of fragment ions observed in the tandem mass spectrum depends on several factors such as the peptide sequence or the ion activation method used. As mentioned previously, low-energy CID, mostly used in the course of this work, produces predominantly b- and y- type ions.\textsuperscript{50}

\section{2.5 Ion Mobility Mass Spectrometry}

Ion mobility mass spectrometry (IMMS) was developed in the late 1950s by McDaniel, but it was not until the early 1970s that the first commercial IMMS instrument was made available. The combination of ion mobility with MS allows for the separation of the ions based on their molecular shape, size and charge. This adds yet another dimension on the range of applications of mass spectrometry as an analytical tool. Although the method is not new, its popularity in the field of biomolecular analysis has only increased in the last 20 years mainly due to the demonstration of its utility in differentiation of coexistent protein conformers.\textsuperscript{52}

There are currently four different methods of achieving ion mobility with mass spectrometry: drift-time ion mobility (DIMS); aspiration ion mobility spectrometry (AIMS); differential-mobility spectrometry (DMS), also called field asymmetric waveform ion mobility (FAIMS); and the most recent traveling-wave ion mobility mass
spectrometry (TWIMS). All of these methods are based on similar principles where more elongated structures interact with the buffer gas longer and thus have an increased drift time compared to more compact structures. Only TWIMS was used in this thesis research and as such, additional details on the operation of IMS will be provided only for this method.

TWIMS is performed using a stacked ring ion guide. A direct current (DC) potential is superimposed onto a RF electrode. After a defined period of time, it is then switched to an adjacent electrode and the process is repeated along the ion guide. This creates a moving electric field or “travelling wave” on which the ions “surf”. The increase of pressure in the guide, by introduction of an inert gas (commonly N₂), can produce some form of ion mobility separation under the right conditions of wave height, wave velocity and IMS gas pressure. The ions with lower mobility roll over the wave more often than higher mobility ions, and are thereby retained longer in the TriWave. This IMS technique has been successfully coupled with Q/oa-TOF instrumentation. Applications of TWIMS to protein conformation analysis will be presented in Section 3.2.2. of this thesis.

2.6 Instrumentation Used in this Research

2.6.1 Synapt HDMS (Waters Corp.)

The majority of the mass spectrometry experiments presented in this thesis was conducted on a first generation Synapt High Definition mass spectrometry system (HDMS), manufactured by Waters Corp. The instrument is a hybrid quadrupole
orthogonal acceleration TOF mass spectrometer that is equipped with a nano-ESI source and capable of TWIMS (Figure 2-5).

Figure 2-5: Schematic of the ESI Q/oa-TOF Synapt HDMS system from Waters Corp. (Reproduced with permission from the manufacturer.)

The nano-ESI source (1) has a so-called “Z shape”, which allows for increased sensitivity as neutral and solvent molecules are sucked away by the vacuum system as the ionized molecules are being transferred into the mass spectrometer. The first feature of the mass analyzer is the T-Wave ion guide (2), which transports the ions from the source at atmospheric pressure to the quadrupole at very low pressure. The construction and mode of operation of T-Wave RF-only ion guide was previously presented in Section 2.5. The quadrupole (3) plays the role of either an ion guide or an ion mass selector for tandem MS experiments. It is followed by a series of three T-Wave ion guides, also called the TriWave (4), which contains both the ion mobility cell and the collision cell.
Ions are finally repelled into the orthogonal TOF mass analyzer (5), which can be used in “V-mode” or in “W-mode” for increased resolution and mass accuracy.

2.6.2 UltrafleXtreme MALDI-TOF/TOF Mass Spectrometer (Bruker Daltonics)

The UltrafleXtreme MALDI-TOF/TOF mass spectrometer contains a gridless MALDI source (1) with delayed extraction electronics; a first TOF mass analyzer (2); a high resolution time ion selector (TIS) (3); a second source (4); a “lift” device (4), which increases the potential energy of the ions; and finally a second TOF mass analyzer with rTOF capabilities (5) (Figure 2-6).

![Figure 2-6](image)

**Figure 2-6**: Schematic of the Bruker Daltonics UltraFleX MALDI TOF/TOF MS from Bruker Daltonics. (Adapted from Bruker Daltonics)

The Smartbeam-II laser technology, which can operate at up to 1 kHz; the FlashDetector™; and a 4 GHz digitizer coupled with fast electronics allows for ultra-high
data acquisition speed and enables mass resolution of 40,000 and a mass accuracy of 1 ppm. The “ion family” formed in the source, which is comprised of precursor ions and all resulting fragments from post-source-decay, reaches the TIS at the same time. The TIS deflects these “ion families”, besides the one selected, by switching off the voltage at the gate when the desired family passes through. All the ions from a same family have the same velocity after acceleration in the source. The TIS also enables high resolution when selecting the families.\textsuperscript{55}

After the TIS device, a second source with lift technology allows for the fragmentation of the selected precursor ions. The lift cell is composed of four grids forming three separate cells. The first cell is the true potential lift where the potential between the two first grids is increased from zero to 19 kV, resulting in an increase of the potential energy of the ions. The ions enter the second, or the focusing cell, where the potential is maintained at 19 kV until the ions are all in the cell upon which the potential of the third grid is reduced to 2 - 3 kV. Once in the last cell, the ions are then accelerated to full speed and transmitted to the second TOF mass analyzer.\textsuperscript{55}
Chapter 3

Applications of Mass Spectrometry to Protein Identification, Protein Structure and Protein Dynamics

Due to its versatility, mass spectrometry possesses a wide range of applications in various “omics” fields. This section focuses on its application to proteomics and structural proteomics. Indeed, mass spectrometry is a high throughput method, making it highly convenient for protein identification as results can be obtained in little time and with very low sample consumption. Moreover, mass spectrometry is also a method of choice in the study of intact proteins and protein assemblies since in principle it is not limited in terms of mass (like NMR) and does not require protein crystallization, (unlike X-ray crystallographic analysis). However, mass spectrometry does not provide the level of detail that X-ray crystallography can, but has proven useful in helping to determine the proper conditions to obtain crystal growth.

3.1 Proteomics and protein identification

The primary structure of a protein is the amino acid sequence. Mass spectrometry has risen in the past few decades as a powerful technique for proteins sequencing, as it is
a fast and reliable method. Two approaches can be taken to determine the sequence of proteins: top-down or bottom-up. The top-down method implies that the intact protein is being introduced into the mass spectrometer while in the bottom-up method the protein is first digested with a suitable protease. Both techniques present different advantages and disadvantages, which will be discussed.

3.1.1 The top-down approach

The top-down approach is quite recent and so far its range of applications is mainly related to the identification and characterization of post-translational modifications (PTM) in proteins.\textsuperscript{56} The combined ability to analyze partial protein sequences as well as the masses of intact proteins can provide more information than could be obtained separately. The main advantage of the “top-down” approach is that it does not require any extensive separation or digestion prior to analysis. Plus, analyzing intact protein virtually enables sequencing of the entire entity, allowing for a more complete characterization of the protein and any of its associated PTMs. The primary sequence of the protein of interest can be determined by database search or by “de novo” sequencing. In addition, the masses of the intact proteins are spread out over a wider mass range, potentially making the analysis of complex mixtures easier than the analysis of digested proteins.\textsuperscript{57, 58}

The instrumentation primarily used for the top-down approach is trapping instruments such as Fourier transform ion cyclotron resonance (FTICR) mass spectrometers. These are compatible with electron capture dissociation (ECD) and electron transfer dissociation (ETD), the preferred fragmentation methods for top-down
analysis. Indeed, both dissociation techniques are low energy activation methods, methods, which allow for the conservation of the PTMs. Also fragmentation occurs more randomly than with collision-induced dissociation improving the sequence accessibility.$^{58}$

However, the top-down approach has many drawbacks. First, it requires costly instrumentation. Secondly, the identification of PTMs is difficult and only possible for extensive PTMs as the sensitivity is not optimal. Lastly, the source required for the analysis is ESI, which produces multiply charged ions, thus the fragments are also multiply charged, which can make data analysis tedious. Nonetheless, a recent technique, which couples collision-induced dissociation with ion mobility mass spectrometry, seems to partially answer some of these limitations, by using ion mobility to separate the different fragments.$^{56}$

### 3.1.2 Bottom-up approach

The bottom-up approach is the most widely used in proteomics analysis. This method involves the digestion of the protein of interest prior to its analysis. Commonly trypsin is used as it cleaves C-terminally to lysine and arginine residues, unless followed by a proline. As these amino acids intrinsically possess a positive charge, it increases the ionization efficiency of the resulting peptides and thus the sensitivity of the technique. Unlike the top-down approach, virtually any tandem instrument can be used, as well as fragmentation technique.$^{57, 58}$ The masses of the peptides produced are unique to each protein and is called peptide mass fingerprint. The protein of interest is identified using a database searching utility (such as Mascot)$^{59}$ or by comparison with a mass list from an
“in-silico” digest. To increase the confidence of protein identification, each peptide can also be fragmented and sequenced by tandem MS.\textsuperscript{60}

One of the main advantages of the bottom-up approach is that it allows for the characterization of discrete modifications on a single amino acid, such as PTMs. Plus, when coupled with separation techniques such as HPLC, it allows for the analysis of complex mixtures. However, it is sometimes not possible to assign, with complete certainty, the origin of the peptides.\textsuperscript{60, 57}

There are still some limitations associated with this technique. One is that not all the peptides from the digest are usually analyzed; consequently, information on some regions of the protein is lost. Moreover, the technique can be time consuming, and despite being automated, its throughput is not yet optimal.\textsuperscript{58}

3.2 Protein structural analyses

3.2.1 Charge state distributions

Not long after the introduction of electrospray ionization (ESI), Loo and co-workers realized that changes in the solvent composition induced dramatic changes in the charge state distributions in the mass spectra of proteins. Indeed, more tightly folded proteins have less solvent accessible surface area; hence they display a narrower charge state distribution than unfolded protein with relatively much lower average charge. Under mild denaturing conditions, non-native and native protein states oftentimes coexist, in which case the charge state distribution becomes bimodal. This property can be utilized
to monitor large conformational changes in the protein or to investigate the dynamics of protein folding.61

Several factors can influence the charge state distribution. The first is the solvent composition. Organic solvents are known to lead to more unfolded/denatured protein structures relative to aqueous solvent. Additionally, most proteins are stable at neutral pH in which they adopt tightly folded native-like conformations. They are destabilized at either low or high pH at which they are unfolded or disordered. The effect is highly protein specific and also solvent specific. The second important factor is the presence of structure stabilizing elements such as disulfide bridges. The presence of disulfide bonds is responsible for the stability of many proteins, thus reducing their sensitivity to either solvent conditions or pH. The charge state distribution will not be affected as readily upon change of pH or addition of denaturant, with little or no change in their conformation.62,63

Figure 3-1: Comparison of the simulated charge state distributions of an intrinsically disordered protein (left) and a folded protein (right) in solution under non-denaturing conditions.
Another interesting case is the one of intrinsically disordered proteins. They display very unique mass spectra under native conditions. Indeed, even at neutral pH they often present multimodal distributions, indicating the existence of more than one conformation in solution. Figure 3-1 shows the comparison between the simulated charge state distributions of an intrinsically disordered protein and a naturally folded protein both at near physiological pH.

3.2.2 IMMS

It was only since 1995 that ion mobility started to be applied to the study of proteins. At this time, the main interest was to understand the desolvation process of large molecules. Studies showed compelling results between the observed collisional cross section of proteins and the reported X-ray crystallographic or NMR data, confirming the potential of IMMS for the studies of proteins and assemblies. As IMMS allows for the separation in the drift tube of the folded and unfolded species, one can foresee its utility in the study of protein conformations, dynamics and how it can help understand the folding mechanism of such entities.

Several studies illustrate the utility of IMMS in the study of protein conformational dynamics. Recently, the utility of IMMS in the study of intrinsically disordered proteins was demonstrated. One of the first experiments on α-synuclein shows great promise in the studies of such proteins, and has been thereafter used with success for the study of other systems like the tumor antigen p53. Additionally, IMMS has proven successful in monitoring conformational changes in the High Mobility Group A (HGMA) protein upon phosphorylation. Lastly, IMMS has been used to monitor protein
folding process. By placing an ion-trap mass analyzer in front of the ion mobility cell Badman et al. were able to accumulate ions in the trap and release them in a time-wise manner into the mobility cell, thereby allowing them to study the time dependent unfolding processes of various proteins.  

Since the development of TWIMS and its coupling to the Q/TOF, IMMS allowed for the analysis of intact protein assemblies. It was successfully used to measure the degree of compactness and the mechanism of dissociation of numerous protein complexes. Lorenzen et al. studied the conformational changes of the portal complex of the bacteriophage P22 upon binding of gp-4 using IMMS. Finally, the biggest assembly analyzed by IMMS so far is the Hepatitis B virus capsid (HBV). The two conformations of the protein complexes of 3 MDa and 4 MDa were successfully separated by ion mobility.

IMMS shows great promise for studying proteins and protein assemblies as depicted by numerous examples of its success in monitoring conformation dynamic changes which play key roles in the interactions within and between biomolecules.

3.2.3 Mapping of Disulfide Bridging Patterns and their Verification for Recombinant Proteins

Cysteine residues are unique among the 20 naturally occurring amino acids as they are able to form covalent disulfide bond linkages with other cysteine residues. Hence, they have an important biological significance and are often involved in the stabilization and in defining proteins structures. Additionally, recombinant proteins are often misfolded as they are synthesized in noncognate organisms, and may display
incorrect disulfide bond arrangements. Therefore, being able to identify cross-linked as well as free cysteine residues is crucial to the study of protein structure, but can be very challenging. Several methods have been developed over the years in order to accurately define the disulfide bond pattern of proteins.\textsuperscript{76, 77}

One of the first and most commonly employed techniques is chemical labeling of cysteine residues. Although there are various reagents available, the most frequently used is iodoacetamide or its derivatives. By a simple experiment, it is possible to map the free cysteine residues through reaction with these reagents, assuming that only the free residues are available to react. However, these alkylating agents are not specific to only thiols.\textsuperscript{78, 79} Other reagents, such as selenium derivative compounds have also been used and showed better reactivity and selectivity towards cysteine thiol moieties.\textsuperscript{80} These labeling experiments usually involve digestion by a specific protease prior to MS analysis. In order to establish the disulfide bridge patterns of proteins with the labeling technique, either tandem MS or dual labeling is necessary. The former presents liabilities where if several disulfides bonds are present they can become scrambled in solution resulting in inaccurate conclusions being drawn. The later does not allow for the exact characterization of the amino acid involved in the disulfide cross-linking. The liabilities associated with both techniques can be overcome using computational analysis.\textsuperscript{81}

More recently top-down MS was applied to the determination of disulfide patterns. The fragmentation methods used for this approach are ETD or CAD, requiring a trapping instrument. This technique has proven successful for the determination of complex disulfide patterns. The main liability of this technique is that the instrumentation required, such as FTICR is not readily available.\textsuperscript{82, 83}
Fragmentation using CID in negative ionization mode gives similar fragment patterns to positive mode. However, there are some differences allowing for the identification of particular amino acids or crosslinks. Indeed, CID of intramolecular and intermolecular disulfide bridged peptides results in very unique patterns in the mass spectrum. In the case of intramolecular disulfide bonds, an intense peak corresponding to the anion \([\text{M-H}-\text{H}_2\text{S}_2]^-\) can be observed. Intermolecular linked peptides, can lead to up to four different fragments in the case of symmetrical crosslinking or up to eight different fragments for non-symmetrical crosslinking. This technique can provide useful information of the sequences if the peptides involved in the linkage are known since little sequence information can be deduced from this experiment.\(^8^4\)

### 3.2.4 Noncovalent Protein Complexes

The study of protein complex assemblies is important as these structures play a key role in cellular molecular pathways. However, the study of protein complexes is often a challenge for structural biology due to their large size, heterogeneous composition, flexibility and asymmetric structure. The development of ESI MS allowed for the observation that noncovalent interactions can be maintained in the gas phase. Mass spectrometry has proven to be a valuable tool for the analysis of complexes assemblies. It allowed for the determination of the mechanism of assembly of the GroEL-GroES complex.\(^8^5\) Additionally, ESI enabled the analysis of very large and heterogeneous complexes such as intact ribosomes. These studies allow the understanding of the mechanism of assembly and dissociation of the ribosome and to have insight on the stabilization factors of this “machine” for different bacteria.\(^8^6,8^7\)
One important characteristic of these complexes is their stoichiometry. As mass spectrometry allows for the determination of the mass of the complex, it is an ideal tool for the determination of this property. It requires first the determination of the masses of the constituent species, if unknown. One method is to analyze the assembly under denaturing conditions (aqueous solution containing organic solvent and acid) in order to disrupt the complex to its single components. This technique is very well suited to the analysis of homogenous complexes but can be a bottleneck for the analysis of heterogeneous assemblies as the data interpretation is much more complicated. However, the development of IMMS allows overcoming this limitation as it enables separation of the species in a second dimension. An alternative is to apply CID to the complex assembly. Upon collision with the inert gas, the assembly releases one or several subunits that appear in the lower m/z range while the “stripped complex” appears in the higher m/z range. This technique has proven successful, but one cannot ensure the dissociation of each monomeric species which is again a limitation for the analysis of hetero-oligomers. Lastly, LC separation coupled with MS analysis allows for the separation of each subunit prior to analysis. The challenge with this technique is to obtain good separation of the various subunits. This technique has been successfully used in our lab.

One interesting feature of the study of non-covalent complexes using ESI MS is the charge state distribution, through which it is possible to determine in some part the types of interactions involved. Indeed, high affinity protein assembly formation generally leads to the exclusion of a larger surface area between the subunits; thereby the complex distribution appears at high m/z, away from the one of the monomeric species. Inversely, nonspecific or low affinity protein complexes display a lower exclusion of the solvent.
exposed area, hence the charge state distribution of the oligomeric and monomeric species overlap.\textsuperscript{90} One way to further distinguish between specific and nonspecific oligomeric species is CID. Fragmentation data of specific oligomers display the “stripped complex” due to asymmetric charge partitioning (described earlier), which is not observed for nonspecific complexes.\textsuperscript{64} Lastly, oligomers formed through crosslinking (disulfide bridges for instance) may also display a nonspecific-like charge state distribution. However, it can be differentiated as the oligomeric state is preserved upon denaturation.

### 3.2.5 MS as a Tool for Protein Crystallography

Mass spectrometry has played an important role as a complementary and supporting technique for protein crystallography. Protein X-ray crystallography is the gold standard for structure determination. However, it requires large quantities of high purity protein and the optimal conditions for crystal growth. Mass spectrometry has proven to be very useful in confirmation of the identity and the purity of the protein sample.\textsuperscript{91} Digestion of the sample and peptide mass fingerprinting can be performed to confirm the identity of recombinant proteins. This technique allows for the detection of possible errors that occurred during the PCR step of the cloning process or mistranslation errors.\textsuperscript{92} In addition, mass spectrometry allows for the identification of unwanted protein modifications such as the oxidation of methionine residues. The nature of the error can be identified by peptide mapping and the precise site of modification can be pinpointed using tandem MS.\textsuperscript{93} Mass spectrometry can also be used to detect and in some cases quantify the extent of desired modifications,\textsuperscript{94,95} thereby ensuring the purity of the protein
with higher confidence that other common techniques such as gel electrophoresis, dynamic light scattering or analytical ultracentrifugation.\textsuperscript{96}

Furthermore, success in producing crystals that are packed into a well-ordered lattice is dependent on the degree of conformational motion of the protein, which may limit the success of obtaining diffraction quality crystals. Although a number of computer-aided approaches are available, there are inherent ambiguities remaining concerning the precise boundaries of the folding domains. Limited proteolysis in combination with mass spectrometry can provide a higher level of detail and often resolve the ambiguities raised by other methods.\textsuperscript{97, 98}

Mass spectrometry can also be utilized for the analysis and identification of the crystal composition. MALDI-MS has been successfully used on intact crystals and allowed for the screening of samples containing the trimeric species formed between the TATA-box binding protein (TBP) from human or \textit{Arabidopsis thaliana} and the human transcription factor II B (TFIIB) in complex with DNA.\textsuperscript{99} Additionally, proteins sometimes crystallize over extended periods of time (weeks or months), during which the protein is susceptible to degradation or modification. These alterations can be detrimental to the crystallization success.\textsuperscript{100}

Mass spectrometry can also play an important role at other stages of the crystallographic process, for instance in the determination of the incorporation of selenomethionine into protein sequences which is necessary to guarantee successful phasing for structure determination.\textsuperscript{101, 102} It has also been demonstrated that mass spectrometry could help in structure refinement. It can fill in information for areas where there is poor density in the electron density maps.\textsuperscript{103} In combination with molecular
crosslinking, it also provides restraint information about the structure of the protein, which helps to resolve conformational ambiguity.\textsuperscript{104}
Chapter 4

Cloning, Expression and Purification of the Individual Subunits of Human Haptoglobin Chains

The choice of recombinant protein as a source of material, resides in the intrinsic properties of haptoglobin that can make it challenging to study. Indeed, human haptoglobin exists in three different phenotypes: Hp 1-1, Hp 2-2 and Hp 2-1. The phenotypes are randomly distributed among populations; hence it is challenging to obtain a sample containing a unique phenotypic form. Moreover, haptoglobin is an acute phase protein, meaning that it is only produced in large quantities upon hemolysis. As a consequence, haptoglobin extracted from human blood is quite pricy ($1000/10 mg for mixed phenotypes). In addition, haptoglobin is glycosylated in mammals, which causes difficulties for analysis by mass spectrometry.

Various host systems can be used in molecular cloning for the proliferation of the gene of interest. In this research, a bacterial host, *E. coli*, was used for gene amplification and expression. Bacterial hosts are often used for production of protein in order to carry out structural analysis or molecular assays. The main advantages of using bacterial host are the simplicity of the cultures, the fact that expression is scalable, and low cost. However, some drawbacks are inherent to this method. The first is that protein solubility can be an issue. Second, though some minor PTMs are possible, glycosylated proteins
cannot be produced by this system. And finally, it is often challenging to express functional mammalian proteins. Nevertheless, protein cloning is a method of choice to obtain large quantities of a pure protein at a reasonable cost.

This Chapter presents the cloning of the individual subunits of haptoglobin. The cloning of light chains (L₁ and L₂) was successful and led to soluble proteins. The cloning of the heavy chain (H) was rather challenging. The difficulties and the troubleshooting for the H-chain cloning will also be discussed.

4.1 Material and Methods

4.1.1 List of Materials and Reagents

Sequence optimized genes for the individual human haptoglobin subunits were purchased from Geneart (Germany) and the corresponding DNA primers were obtained in lyophilized form from Integrated DNA technology (IDT) (Coralville, IA). The gateway® cloning kits were purchased from Invitrogen (Invitrogen, Carlsbad, CA). The DNA purification kits, minilute or miniprep were obtained from QIGEN (Valencia, CA). All reagents were purchased from Sigma-Aldrich® (St. Louis, MO) or Fisher Scientific. Solvents used for mass spectrometry analysis were all of HPLC-grade or higher.
4.1.2 Polymerase Chain Reaction (PCR)

PCR is commonly used in molecular biology to amplify a gene sequence of interest by *in vitro* enzymatic replication assisted by a DNA polymerase. Plasmids containing genes of human haptoglobin chains L₁, L₂ and H were purchased from Geneart with the codon usage adapted to the codon bias of *E. coli* genes. The sequences are provided in Appendix A. OmniMAX™ *E. coli* strain cells were transformed with the plasmid to be amplified. The amplified plasmid was then extracted using a miniprep purification kit. As pENTR uses TOPO-assisted directional cloning for the insertion of the gene of interest (GOI), the forward primers were designed with a CACC overhang at the 5’ end (Table A.3).

The primer solutions were prepared by dissolving the lyophilized DNA in Tris EDTA pH 8.0 (TE solution) to a concentration of 250 µM. The PCR reaction was set up in a 0.2 mL low retention PCR tube, by mixing the DNA template that contains the GOI to be amplified, a thermo resistant DNA polymerase (Taq DNA polymerase), deoxyribonucleoside triphosphates (dNTP’s) and a buffer solution providing an optimum chemical environment for the enzyme activity and stability. The volumes of each component are listed in Table 4.1. The mixture was placed on a PCR Mastercycler personal from Eppendorff Scientific Inc. The programs described in Table A.2 were used.

The PCR reaction products were purified using the minilute gel extraction kit. Purification of the PCR product was necessary to eliminate traces of the remaining reagents that can interfere with subsequent downstream applications. Samples were stored at −20 °C.
Table 4.1: Reagents for the PCR reactions. A: PCR reaction of L₁ and L₂. B: PCR reaction of H

<table>
<thead>
<tr>
<th>Reagents</th>
<th>Volumes (µL)</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR-grade H₂O</td>
<td></td>
<td>22</td>
<td>5.8</td>
</tr>
<tr>
<td>Primer Mix (10 µM)</td>
<td></td>
<td>7</td>
<td>0.6</td>
</tr>
<tr>
<td>DNA Template</td>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>KOD Buffer</td>
<td></td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>MgSO₄ (20 mM)</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>dNTP</td>
<td></td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>KOD DNA Polymerase</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total Volume</td>
<td></td>
<td>50</td>
<td>20</td>
</tr>
</tbody>
</table>

4.1.3 Insertion of the GOI into the entry vector

In order to produce an expression clone using the Gateway® cloning method, an entry clone is first required. The TOPO® cloning reaction allows for the insertion of the PCR product into the entry vector pENTR™/D-TOPO® using TOPO assisted directional cloning (Invitrogen). The conditions of the reaction are summarized in Table 4.2.

Table 4.2: List of the reagents used for the pENTR™/D-TOPO® reaction

<table>
<thead>
<tr>
<th>Reagents</th>
<th>Volume (µL)</th>
<th>Incubation conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salt</td>
<td>1</td>
<td>5 min at RT</td>
</tr>
<tr>
<td>H₂O</td>
<td>3 - 3.5</td>
<td></td>
</tr>
<tr>
<td>pENTR/D-TOPO vector</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>PCR product</td>
<td>0.5 - 1</td>
<td></td>
</tr>
<tr>
<td>Final Volume (µL)</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>
After the reaction, TOP10® E. coli cell strains were transformed with the reaction products, and the next day the plasmids were extracted from different colonies using the miniprep extraction kit.

### 4.1.4 Insertion of the GOI into the expression vector

Finally, the expression clone was created by swapping the death gene present in the expression vector pDEST-C1 (ccdB gene) and the GOI previously inserted into the entry clone. The transfer between the two genes was achieved using a transposition reaction catalyzed by the LR clonase™ enzyme (Invitrogen). The setup of the reaction is presented in Table 4.3. After the reaction, competent OMNImax® E. coli cell strains were transformed with the reaction products, and the next day the plasmids were extracted from different colonies using the miniprep extraction kit.

**Table 4.3**: List of reagents and conditions for the LR Cloning™ reaction.

<table>
<thead>
<tr>
<th>Reagents</th>
<th>Volumes (µL)</th>
<th>Incubation Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>pDEST-C&lt;sub&gt;1&lt;/sub&gt; Vector</td>
<td>1</td>
<td>RT for 1h</td>
</tr>
<tr>
<td>pENTR/D-TOPO-GOI plasmid</td>
<td>1 - 7</td>
<td></td>
</tr>
<tr>
<td>1x TE buffer pH8</td>
<td>8 - 2</td>
<td></td>
</tr>
<tr>
<td>LR clonase mix</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Proteinase K</td>
<td>1</td>
<td>37 °C for 10min</td>
</tr>
</tbody>
</table>
4.1.5 **Protein Expression**

4.1.5.1 **Small Scale Protein Expression**

The expression plasmid previously extracted by miniprep was transformed into the cloning hosts; either BL21 DE3 or Rosetta DE3 pLysS *E. coli* cell strains. The following day, colonies were picked up on the plate selecting for the expression plasmid, and transferred into 5 mL of 25 mg/mL LB medium and 1 mM of streptomycin and the antibiotic selecting for the cell line. Two different cell lines were used, one with no resistance and the other with chloramphenicol resistance (Table A.2). The cultures were incubated at 37 °C in a New Brunswick Scientific Innova 4000 Incubator shaker. When their OD$_{600}$ reached 0.4 - 0.6, the protein expression was induced by adding 1 mM IPTG. The cultures were incubated for 2 to 4 h. Both culture samples were centrifuged at 10,000 rpm for 1 min. The supernatant was discarded and cell pellets were collected. 50 µL of 1x BugBuster™ Protein Extraction Reagent (Novagen, Darmstadt, Germany) and 1x SDS PAGE sample buffer (BioRad, Hercules, CA) were added to the samples for SDS PAGE analysis.

4.1.5.2 **Large Scale Protein Expression**

A small amount of cells taken from a previously prepared glycerol stock were transferred to a flask containing 50 mL of 25 g/L LB medium and 1 mM of Streptomycin and the antibiotic selecting for the cell line. The culture was incubated overnight at 37 °C in a shaker. The following morning, the overnight growth was added to 1 L of 25 g/L LB
media, 1 mM streptomycin and the antibiotic selecting for the cell line. The cells were
grown at 37 °C in the shaker. When their OD$_{600}$ reached 0.4 - 0.6, the expression was
induced by adding 1 mM IPTG. The culture was left to grow at 37 °C for three more
hours. Then, the cells were harvested by centrifugation (4000 rpm at 4 °C for 20 min)
using a Beckman Coulter™ TJ-25 centrifuge. The supernatant was discarded. The cell
pellet was then recovered and weighed.

4.1.6 Cell Lysis

To each gram of cells, 10 mL of lysis buffer (composition provided in Table 4.4)
were added. To the lysis buffer, a pinch of lysozyme, 1x phenylmethylsulfonyl fluoride
(PMSF) (protease inhibitor) and 10 mM β-mercaptoethanol (BME) were added. The
samples were incubated on ice for 30 min to 1 hr. Then, they were sonicated for 2 min on
ice and spun down (20 min at 10,000 rpm at 4 °C). The supernatant was thereafter
collected and divided into 5 to 15 mL aliquots to each of which 15 - 20% v/v glycerol
was added. The lysate aliquots and the cell debris were stored at −80 °C.

Table 4.4: Composition of the buffer used for the cell lysis after expression of the haptoglobin
subunits.

<table>
<thead>
<tr>
<th>Lysis buffer</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris-HCl pH 7.5</td>
<td>25 mM</td>
</tr>
<tr>
<td>NaCl</td>
<td>50 mM</td>
</tr>
<tr>
<td>BME</td>
<td>10 mM</td>
</tr>
</tbody>
</table>
### 4.1.7 Purification of Hp Chains

Since all three Hp chains included a N-term His-tag, the purification was achieved by Nickel affinity chromatography.

After lysis, 0.3% Polyethyleneimine (PEI) solution was added in order to precipitate the DNA from the sample. The mixture was homogenized for 2 min before centrifugation (10,000 rpm for 40 min at 4 °C). The supernatant was collected for purification. The column was first washed with buffer B (see Table 4.5 for composition) containing high concentration of imidazole to prevent the contamination by any other unwanted protein that may have remained on the column. After which the column was equilibrated with buffer A (Table 4.5) containing low imidazole concentration. The protein was loaded onto the column, after clarification of the lysate by centrifugation (10,000 rpm for 15 min at 4 °C) and filtration through a Millipore™ syringe filter. The column was washed with 5 column volumes (CV) of buffer A to elute the proteins that would not specifically bind to the resin. The protein of interest was eluted using an isocratic flow of buffer B at 2 mL/min. The fractions were collected in glass tubes and those containing the protein of interest were pooled and concentrated using the Millipore™ a 10 kDa molecular weight cut off filters.

The His-tag was consequently removed by adding the appropriate volume of TEV protease at 5 mg/mL to the sample in 1:20 protease:protein ratio. The sample was dialyzed using a SnakeSkin® dialysis tubing (Thermo Scientific) with a 3 kDa (L₁) or 10 kDa (L₂) cut off, against buffer A for 4 h to overnight. The reaction was loaded onto the Ni-Sepharose affinity column after equilibration with buffer A. The protein of interest was eluted with an isocratic flow of buffer A while the TEV and the His-tag were
thereafter eluted with an isocratic flow of buffer B. The pooled fractions containing the protein of interest was concentrated using a 15 mL Millipore™ centrifugal unit (MILLIPORE) with a 10 kDa (for H and L₂) or 3 kDa (for L₁) cut off, until a concentration of 5 mg/mL was reached. The samples were then divided into 200 - 500 µL aliquots after 80% glycerol was added up to a final concentration of 15% and the samples were flash-frozen in dry ice, and stored at –80 °C.

Table 4.5: Buffers composition for Ni affinity purification of Hp chains.

<table>
<thead>
<tr>
<th></th>
<th>Buffer A</th>
<th>Buffer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaH₂PO₄</td>
<td>20 mM</td>
<td>20 mM</td>
</tr>
<tr>
<td>NaCl</td>
<td>0.5 M</td>
<td>0.5 M</td>
</tr>
<tr>
<td>Imidazole</td>
<td>20 mM</td>
<td>500 mM</td>
</tr>
<tr>
<td>BME</td>
<td>10 mM</td>
<td>10 mM</td>
</tr>
<tr>
<td>Conductivity</td>
<td>35 mS/cm</td>
<td>50 mS/cm</td>
</tr>
</tbody>
</table>

4.1.8 Electrophoresis

4.1.8.1 DNA Agarose Gel

DNA-agarose gel was used to confirm each step of the cloning process. They were also used as mean to separate the PCR products in order to purify the desired one. The matrix used for DNA separation was 1% w/v agarose. The solvent used to dissolve the agarose powder was composed of Tris, EDTA and acetic acid (TEA). The samples were prepared as follows: 3 µL of 2x loading buffer is added to 3 µL of sample. The products of the reactions were confirmed by size comparison with a standard DNA ladder.
4.1.8.2  SDS-PAGE

SDS PAGE was used to confirm the success of protein expression or to estimate the purity of proteins preparations. In the case of protein expression, the samples were prepared as follows: 500 µL of the cell sample was collected before induction (t = 0 h) and 250 µL was collected after expression (t = 3 h), the samples were centrifuged (5 min at 10,000 rpm) and 50 µL of BugBuster™/loading buffer were added to the pellet. To confirm solubility or efficiency of the lysis, 10 µL of the lysis mixture was collected and centrifuged for 5 min at 10,000 rpm. The pellet and supernatant were separated and 10 µL of loading solution were added to each. For purified protein, 10 µL of loading buffer were added to 10 µL of concentrated protein sample. Loading of 10 µL was typically used.

The samples were loaded onto a 15% SDS polyacrylamide separating gel with a 5% stacking gel and run at 200 V for 70 - 90 min. The gel was stained with a solution of Coomassie blue™ (Bio-Rad) for one hour and destained overnight with a solution composed of 30% methanol and 10% glacial acetic acid. The proteins were identified by mass comparison with a standard protein ladder (Precision Plus Protein™ Dual Xtra Standard, Biorad), ranging from 5 to 250 kDa.
4.1.9 Trypsin Digestion

4.1.9.1 In-Gel Digestion

In-gel protein digestion was carried out according to the manufacturer’s suggested protocol (Sigma Aldrich). The gel band of interest was excised using a surgical razor blade and transferred to a 0.6 mL low retention eppendorf microcentrifuge tube. The staining was removed by two consecutive washes with acetonitrile and ammonium bicarbonate solution for 20 min at 37 °C. The sample was then dried to completeness for 20 min in a speed vac. Then, 20 µL of trypsin reaction, buffer (5% acetonitrile in 50 mM ammonium bicarbonate containing 1 mM HCl) and 0.4 µL of 1 µg/µL proteomics grade trypsin were added. The digest was carried out overnight at 37 °C.

4.1.9.2 In Solution Digest

After purification the protein samples were desalted using Amicon™ filters from MILIPORE™ with a solution of 10 mM ammonium acetate. 10 µL of the sample were diluted to 50 µL using 50 mM ammonium bicarbonate pH 8 after which 0.5 µL of 1 µg/µL proteomics-grade trypsin was added. The overnight digestion was carried out at 37 °C.
4.1.10 *In Silico* Digest

The mass lists of the expected tryptic peptides from the various Hp chains were generated *in silico* using ProteinProspector, a free online software made available by the University of California, San Francisco (prospector.ucsf.edu/prospector/mhhome.htm). The function “MS digest” was used to create the peptide mass lists, while “MS product” was used to help processing tandem mass spectra.

4.1.11 MALDI-TOF/TOF MS Experiments

The identities of the expressed proteins were confirmed after IGD using the UltraflleXtreme MALDI-TOF/TOF mass spectrometer equipped with a smartbeam II laser (Bruker Daltonics, Billerica, MA). A saturated solution of α-cyano-4-hydroxycinnamic acid (CHCA) dissolved in 50% ACN containing 0.1% TFA was used as the matrix. The sample and matrix were spotted onto the MALDI target using the sandwich method where 1 µL of sample is spotted between two layers containing 1 µL of matrix. The instrument was operated in reflectron mode and externally calibrated in the range 400 - 4000 m/z using the peptide calibration standard #206195 from Bruker Daltonics (Angiotensin I, Angiotensin II, Substance P, Bombesin, ACTH clip 1-17, ACTH clip 18-39 and Somatotensin 28). The data were collected in positive ion mode. Approximately 3000 - 8000 shots were summed per data acquisition using 1000 Hz acquisition speed. Data were processed using FlexAnalysis 3.3 software (Bruker Daltonics).
4.2 Results and Discussion

4.2.1 Protein Cloning, Expression and Purification of the Light Chains (L₁ and L₂) of Human Haptoglobin

4.2.1.1 Molecular Cloning of the Light Chains L₁ and L₂

The first step of the cloning process was amplifying the gene of interest by PCR. The results are shown in Figure 4-1 and Figure A-4 for L₁ and L₂, respectively. The sizes of the PCR products were compared against the 10 kbp DNA ladder (Lane L of Figure 4-1 and Figure A-4). Both products appear to have the correct size, 282 bp for L₁ and 463 for L₂ (Lane A on the figures). The PCR product was then inserted into the entry vector pENTR™/D/TOPO® through the TOPO® assisted reaction (Lane B on the figures). By comparison with the supercoiled DNA ladder (Lane L’ on the figures) one can be confident that the proper entry clones were created. The L₁ chain was expected to be around 2.9 kbp while L₂ should be around 3 kbp.

Figure 4-1: SYBR GOLD™ stained 1% DNA agarose gels of L₁ chains cloning steps. From left to right are the product from the PCR, the pENTR™/D/TOPO® reaction product and finally the LR Clonase® reaction product. The products of interest are indicated with a red box.
The final step was the creation of the expression clones. They were created by reacting the entry clones and the donor vector through the LR Clonase® reaction through ligation and recombination. The attachment regions from the entry clone (attL) and the one of the donor vector (attR) are recognized and swapped by the enzyme allowing for the creation of the proper expression clone. Based on the results from the DNA gel (Lane C), the reaction was successful. The expected numbers of base pairs for L1 and L2 were 4 kpb and 4.2 kbp, respectively. The newly created gene sequences are presented in Figure 4-2.

**Figure 4-2:** DNA sequence of the newly synthesized expression clones. A is the DNA sequence of L1 and B for L2. The start codons are indicated in grey and the inserted GOI in blue.

### 4.2.1.2 Expression and solubility test of the Light chains

The expression clones were transformed into chemically competent Rosetta DE3 pLysS and BL21 DE3 for L1 and L2, respectively. The colonies grown against the proper antibiotic (Streptomycin) were incubated in LB media in order to perform small scale
expression tests. Once the OD_{600} of the cultures reached 0.6, 500 µL aliquots were taken out and the expression was induced by adding 1 mM IPTG. The cells were left to grow for an additional three hours. 250 µL were then removed for gel analysis. Both aliquots were loaded onto a SDS-PAGE gel for comparison. The results are presented in Figure 4-3 and Figure A-5. The overexpressed proteins are indicated in the figures by a red box. A His-tag followed by a linker was inserted at the N-terminal domain of the proteins, coming from pDEST-C1. The sequences and the parameters of the expressed protein are represented in Figure 4-4 and Table A.4, respectively. Within the error expected for SDS-PAGE, both L_1 and L_2 seem to have a reasonable level of expression. The masses of the proteins including the His-tag are 14 kDa and 20 kDa for L_1 and L_2, respectively (Table A.4).

**Figure 4-3:** Expression test for L_1. The SDS PAGE gel was stained with Coomasie blue™. The overexpressed protein is circled in red.
Figure 4-4: Sequences of Hp light chains including the His-tag and TEV protease site. A is the sequence of L₁ and B the one of L₂. The His-tag and the TEV protease site with a linker in between are in italics. The cysteine residues are labeled in red; the one involved in intra disulfides bond are red underlined.

The identities of the expressed proteins were confirmed by trypsin In-Gel Digestion (IGD), followed by MALDI-MS analysis (Figure 4-5 and Figure A-9). The peaks corresponding to the tryptic peptides generated by in silico digest using ProteinProspector online software are indicated on the spectra with their position in the amino acid sequence of the protein. Table 4.6 and Table A-5 summarize the peptide list confirming the identity of the proteins, the sequences and the mass difference between the calculated and experimental m/z values of each identified peptide. Since the light chains account for 25% for L₁ and 37% for L₂ of the total mass of Hp monomeric species (one light chain and one heavy chain) Mascot platform could not positively identified the proteins. As a consequence, MS/MS experiments were necessary to increase the confidence in identification of the recombinant proteins. A representative tandem mass spectrum is provided in Figure 4-6. Both b- and y- type ions were identified confirming the expected peptide sequence as indicated. Other tandem MS spectra are presented in Appendix A. The percent mass coverage for L₁ and L₂ were 49% and 67%, respectively.
Some of the peptides were found to contain intra-disulfide bonds. The unassigned peaks are likely to be from other *E. coli* proteins of similar size to the protein of interest that were unresolved by SDS PAGE or from trypsin autolytic activity.

**Figure 4-5:** MALDI-TOF/TOF MS of in-gel tryptic digest of L₁-chain. The sequence positions of the positively identified peptides are indicated.
Table 4.6: List of the peptides used to confirm expression products of L\textsubscript{1}. \(m/z\): experimentally measured monoisotopic \(m/z\) value of peptide ion; \(m/z_{\text{calc}}\): calculated \(m/z\) value for the monoisotopic peak; \(\Delta M\): monoisotopic mass error of measurement. “*” indicates that these peptides involved intra disulfides bonds. Non Ambiguous sequence were determined by tandem MS experiments.

<table>
<thead>
<tr>
<th>Sequence position</th>
<th>(m/z)</th>
<th>(m/z_{\text{calc}})</th>
<th>(\Delta M)</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-18</td>
<td>2015.88</td>
<td>2015.85</td>
<td>0.03</td>
<td>AHHHHHHVGT GSNNDDDDK</td>
</tr>
<tr>
<td>1-24</td>
<td>2695.23</td>
<td>2695.20</td>
<td>0.03</td>
<td>AHHHHHHVGT GSNNDDDKSTSLYK</td>
</tr>
<tr>
<td>81-95</td>
<td>1708.88</td>
<td>1708.85</td>
<td>0.03</td>
<td>LRTEGDGVYT LNNEK</td>
</tr>
<tr>
<td>83-95</td>
<td>1439.69</td>
<td>1439.66</td>
<td>0.03</td>
<td>TEGDVYTLN NEK</td>
</tr>
<tr>
<td>96-100</td>
<td>688.43</td>
<td>688.38</td>
<td>0.05</td>
<td>QWINK</td>
</tr>
<tr>
<td>101-117*</td>
<td>1741.88</td>
<td>1741.88</td>
<td>0.00</td>
<td>AVGDLPECE AVCGPK</td>
</tr>
<tr>
<td>106-117*</td>
<td>1271.64</td>
<td>1271.63</td>
<td>0.01</td>
<td>LPECEAVCGK PK</td>
</tr>
</tbody>
</table>

Figure 4-6: Tandem MALDI-TOF/TOF MS of the \(m/z\) 1708.88 from the trypsin digest of L\textsubscript{1} chain. The ladder series of b- ions and the identified y- ions are indicated.
After confirmation of the identities of the expressed proteins, a medium scale bacterial growth for expression was carried out (100 mL) in order to study the solubility of the proteins and determine the appropriate lysis buffer. The first lysis buffer tested (Table 4.4) was recommended by Dr. Mueser. Both light chains were soluble as shown in Figure 4-7 and Figure A-15 for the L₁ and L₂, chain respectively. The quantity of protein in the lysate (soluble portion) was sufficient for subsequent experiments to be carried out. This result was different from what was previously reported. Indeed Lai et al. reported obtaining insoluble proteins when expressing the different haptoglobin chains in *E. coli.*¹⁰⁵ This was likely due to them not using optimized DNA sequence for *E. coli* codon usage. Using optimized sequences could be the key to success in attaining soluble proteins.

![Figure 4-7: Solubility test for L₁. The SDS PAGE gel was stained with Coomasie blue™. The overexpressed protein found in the lysate is circled in red.](image)

**Figure 4-7:** Solubility test for L₁. The SDS PAGE gel was stained with Coomasie blue™. The overexpressed protein found in the lysate is circled in red.
4.2.1.3 Large Scale Expression and Purification

Once the identity of the expressed material and its solubility were ascertained, large scale growth for protein expression was carried out. After a three-hour induction the cells were collected by centrifugation and lysis was performed on one third of the culture. The rest of the cell culture was stored as induced stock at –80 °C. The proteins were produced with a N-terminal His-tag for facile purification using Ni-affinity chromatography. The buffers used for the purification were as recommended by the resin supplier (GE healthcare) but with the addition of BME, which was necessary in order to prevent the chains from oligomerizing, since they theoretically possess two and three free cysteine residues for L1 and L2, respectively. The free cysteine residues are presented in red in Figure 4-4. Representative chromatograms are represented in Figure 4-8 and Figure A-16. After purification the fractions highlighted in red were pooled and concentrated using a Millipore™ filter centrifugal unit down to 5 to 7 mL (about 2 mg/mL).

![Ni-NTA purification chromatogram of L1 chain. Green and purple traces correspond to absorbance at 280 and 260, respectively. The fractions that were pooled together are highlighted in red on the figure](image-url)

Figure 4-8: Ni-NTA purification chromatogram of L1 chain. Green and purple traces correspond to absorbance at 280 and 260, respectively. The fractions that were pooled together are highlighted in red on the figure
The His-tags were removed in order to eliminate any possible risk that their presence influences the structure of the protein. As presented in Figure 4-4, the proteins contain a TEV protease cleavage site for facile His-tag removal. TEV protease specifically recognizes the “ENLYQG” sequence and cleaves C-terminally to glutamine residue, leaving a glycine residue at the N-terminus of the protein sequence. To the freshly purified sample, 200 µL of 5 mg/mL of TEV protease was added. The mixture was dialyzed against buffer A for 4 hours to overnight at 4 °C. The reaction was then loaded onto the Ni affinity column previously equilibrated with buffer A. The desired protein was eluted using an isocratic flow of 100% buffer A and collected while the TEV protease as well as the His-tag are eluted using an isocratic flow of 100% buffer B. The fraction containing the desired protein was concentrated to 5 mg/mL. Typically, 10 mg of protein per liter of culture was obtained. The new protein sequences after tag removal are shown in Figure 4-9. The glycine residue coming from the TEV protease cleavage site is indicated in *italics*. The properties of the final proteins are summarized in Table A.6.

| A          | 1 GVDGNDVTD IADDGCCKPP EIAHGYVEHS VRYQCKNYYK |
|           | 41 LRTEDGTVY2 LNNEKQWINK AVGDKLPECE AVCGKPKNPAN |
|           | 81 NPVQ |

| B          | 1 GVDGNDVTD IADDGCCKPP EIAHGYVEHS VRYQCKNYYK |
|           | 41 LRTEDGTVY2 LNNEKQWINK AVGDKLPECE ADDGCPKPPE |
|           | 81 IAHGYVEHSV RYQCKNYYKL RTEGDGVTY2L NNEKQWINKA |
|           | 121 VGDHLECEEA VCGKPKNPAN PVQ |

**Figure 4-9:** Sequences of Hp light chains including the His-tag and TEV protease site. A: sequence of L1 and B: sequence of L2. The glycine residues left after His-tag cleavage are in *italics*. The cysteine residues are labeled in red; the one involved in intra disulfides bond are underlined.
The samples containing the proteins of interest were concentrated down to 2 mL at 5 mg/mL, 500 µL were aliquot for further experiments, and to the left over, glycerol was added up to 15 to 20 % v/v and the sample were flash frozen in dry ice and kept at –80 °C. The protein aliquots were further concentrated and desalted using an Amicon™ centrifugal unit (Millipore) and washed with a solution of 10 mM ammonium acetate containing 10 nM L-cysteine. A 10 µL aliquot of the sample was removed to perform trypsin in-solution digest in order to confirm the identity of the purified protein and to confirm the removal of the His-tag. MALDI TOF/TOF MS was used to analyze the tryptic digest; the results are presented in Figure 4-10. Table 4.7 summarizes the peptide list confirming the identities of the proteins, the sequences and the mass difference between the calculated and experimental m/z values of each identified peptide. The peaks corresponding to the His-tag (2015.85 or 2695.20 m/z) are not present in the mass spectrum confirming the success of the TEV protease reaction. The mass coverage obtained is 81% of the L₁ chain; hence one can be confident in the identity of the protein purified.
Figure 4-10: MALDI-TOF/TOF MS of the tryptic digest of L₁ after His-tag removal.

Table 4.7: List of the peptides used to confirm purification products of L₁. m/z: experimentally measured monoisotopic m/z value of peptide ion; m/z_{calc}: calculated m/z value for the monoisotopic peak; ΔM: monoisotopic mass error of measurement.

<table>
<thead>
<tr>
<th>Sequence position</th>
<th>m/z</th>
<th>m/z_{calc}</th>
<th>ΔM</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-32</td>
<td>3349.85</td>
<td>3349.54</td>
<td>0.31</td>
<td>GVDSDGNDVDTD IADDGCPKPP EIAHGYVEHS VR</td>
</tr>
<tr>
<td>41-55</td>
<td>1708.96</td>
<td>1708.85</td>
<td>0.11</td>
<td>LRTEGDGVYT LNNEK</td>
</tr>
<tr>
<td>43-55</td>
<td>1439.77</td>
<td>1439.66</td>
<td>0.11</td>
<td>TEGDGVTILN NEK</td>
</tr>
<tr>
<td>56-60</td>
<td>688.43</td>
<td>688.38</td>
<td>0.05</td>
<td>QWINK</td>
</tr>
<tr>
<td>61-77*</td>
<td>1741.92</td>
<td>1741.88</td>
<td>0.04</td>
<td>AVGDKLPECE AVCGKPK</td>
</tr>
<tr>
<td>66-77*</td>
<td>1271.70</td>
<td>1271.63</td>
<td>0.07</td>
<td>LPECEAVCGK PK</td>
</tr>
</tbody>
</table>
4.2.2 Protein Cloning of the H-chain of Human Haptoglobin

4.2.2.1 Molecular Cloning of the H-chain

The H-chain gene was amplified by PCR. The size of the PCR product was compared against the 100 bp DNA ladder (Figure 4-11, Lane L). The product appeared to have the correct size, 804 bp (Lane A). The PCR product was then inserted into the entry vector pENTR™/D/TOPO® through the TOPO® assisted reaction (Figure 4-11, Lane B). By comparison with the supercoiled DNA ladder (Lane L’) one can be confident that the proper entry clones were created. The H-chain was expected to be around 3.4 kbp. The expression clone was created through the LR Clonase® reaction. As done previously for the L₁ and L₂ chains, the gene was inserted into the pDEST-C1 vector. According to

Figure 4-11: SYBR GOLD™ stained 1% DNA agarose gels of H chains cloning steps. From left to right are the product from the PCR, the pENTR™/D/TOPO® reaction and finally the LR Clonase® reaction. The products of interest are circled with a red box.
the results (Figure 4-11 Lane C), the reaction resulted the proper product size. The expected number of base pair is 4.5 kbp.

4.2.2.2 Small Scale Expression Test of H chain

The expression clone was transformed into chemically competent Rosetta DE3 pLysS for small scale expression test. No expression was observed. Thus, another expression host was chosen. The plasmid was then transformed into Tuner DE3 cells. The same procedure followed for the L₁ and L₂ chains was carried out. Expression was obtained as shown in Figure 4-12. The overexpressed proteins are indicated in the figures by a red box. A His-tag followed by a linker sequence coming from pDEST-C1 was inserted at the N-terminal domain of the proteins. The sequences and the parameters of the expressed protein are represented in Figure 4-13 and Table A.4. Within the error associated with SDS PAGE, the H-chain seemed to have the correct mass of approximately 32 kDa (Table A.4).
**Figure 4-12:** Expression test for H-chain. The SDS gel was stained with Coomasie blue™. The overexpressed protein is circled in red.

---

**Figure 4-13:** Sequences of Hp heavy chains including the His-tag and TEV protease site. The His-tag and the TEV protease site with a linker in between are in italic. The cysteine residues are labeled in red; the one involved in intra disulfides bond are underlined and the asparagine residues potentially N-glycosylated are in blue and underlined.
The band was excised, and trypsin in-gel digestion was performed and the resultant peptides analyzed by MALDI TOF/TOF MS. The spectrum is presented in Figure 4-14. From the digest, none of the peaks observed could be assigned to the H-chain sequence apart from the His-tag labeled with the “▲”. The identity of the peak was confirmed by tandem MS. At this point it was assumed that the expression clone was incorrect and troubleshooting ensued.
Troubleshooting of H Chain Cloning

The first step for trouble shooting was to send the gene out for sequencing to the Micro-genomics plant at Ohio State University (Columbus, OH). A section of the sequenced gene (top) and the theoretical sequence (bottom) have been aligned using nucleotide Blast (Figure 4-15). The expression clone was shown to possess thirteen additional base pairs when compared to the theoretical one (highlighted in red in Figure 4-15). Consequently, the reading of the gene is off-frame, explaining both the absence of expression in Rosetta DE3 and the expression of the wrong protein in Tuner DE3 cells, which still contained the expected His-tag. The codon start ATG is highlighted in grey, and the swap region between the attL1 from pDEST-C1 and the attR1 from pENTR™/D/TOPO® is indicated by the red connector. Considering where the mistake is located, two possible sources of error can be found. The additional base pairs could have been introduced during the PCR process, or it could have come from a preexisting error.
in the pENTR™/D/TOPO® vector. Thereby, new primers for H-chain with a higher melting temperature (T_m) to increase their specificity (Table A.2) and a new TOPO® reaction kit were ordered.

4.2.2.4 New Molecular Cloning of the H Chain

The H chain gene was amplified by PCR. The size of the PCR product was compared against the 100 bp DNA ladder (Lane L). The product appears to have the correct size, 804 bp (Lane A). The PCR product was then inserted into the entry vector pENTR™/D/TOPO® through the TOPO® assisted reaction (Lane B). By comparison with the supercoiled DNA ladder (Lane L’) one can be confident that the proper entry clones were created. H-chain is expected around 3.4 kbp.

![Image of gel](image.png)

**Figure 4-16:** SYBR GOLD™ stained 1% DNA agarose gels of H chains cloning steps. From left to right are the product from the PCR and the pENTR™/D/TOPO® reaction product. The products of interest are circled with a red box.
4.3 Conclusion

The \( L_1 \) and \( L_2 \) chains of human Hp were successfully cloned and expressed and soluble material was obtained. The cloning of the H chain was more difficult and after initial troubleshooting the process is yet incomplete.
Chapter 5

Initial Characterization of the Haptoglobin Chains by Mass Spectrometry

The amino acid sequences of recombinant eukaryotic proteins expressed in \textit{E. coli} may be identical to the wild type, yet they are often inactive or less active. \textit{E. coli} is a prokaryote hence it cannot perform many of the posttranslational modifications that these proteins require. One example of posttranslational modification relevant to this thesis is disulfide bond formation. The bacterial cytoplasm is a reductive environment, which inhibits disulfide bond formation. It is only after extraction of the protein of interest from this media that cysteine residues can become crosslinked and lead to the appropriate folded structure. However, when several cysteine residues are present, the formation of the proper linkages cannot be assured. As a result, it is necessary to study the folding of the recombinant proteins, to ensure the formation of the proper disulfide bridges and function.\textsuperscript{77}

The recombinant L\textsubscript{1} and L\textsubscript{2} chain of human haptoglobin contain numerous cysteine residues. Some are involved in intrachain disulfide bridging while some other are involved in the crosslinking to other chains. As discussed in the Section 3.2.3 several methods can be applied to determine the disulfide pattern of the protein of interest using mass spectrometry. In this section, the initial characterization of the L\textsubscript{1} and L\textsubscript{2} chains is
presented. Intact L₁ was analyzed and an attempt was made to characterize the disulfide patterns of both L₁ and L₂.

5.1 Materials and Methods

5.1.1 Sample Preparation for Mass Spectrometry Analysis

Proteins produced, as described in Chapter 3 of this thesis were used for the following experiments. Nonvolatile salts were removed by buffer exchange using Amicon™ Ultra 0.5 centrifugal devices from Millipore™ with a solution of 10 mM ammonium acetate containing 10 nM L-cysteine in order to prevent oligomerization through disulfide bond formation. Proteins in which sulfhydryl groups have been reacted (Section 5.1.2) were washed with a solution of 10 mM ammonium acetate.

5.1.2 Reaction for Mapping of the Disulfide Bridges

10 μL of the desalted protein solution were aliquoted into two 0.6 mL eppendorf tubes. To one aliquot, 0.5 μL of 1 M tris-2-carboxyethylphosphine (TCEP) was added and left to react for 30 min at room temperature. Afterward, 5 μL of 200 mM iodoacetamide (IAA) were added and the mixture was incubated in the dark overnight at room temperature. The reaction with only IAA was carried out in the same way, but without the step involving TCEP.
5.1.3 In-solution Trypsin Digestion

Typically, 10 µL aliquots of protein sample were added to an equivalent volume of 50 mM ammonium bicarbonate solution. 0.5 µL of a 1 mg/mL trypsin was added and digestion was carried out at 37 °C for 6 hours to overnight.

5.1.4 Mass Spectrometry Analysis

Data collection was carried out on a Synapt HDMS quadrupole time-of-flight mass spectrometer equipped with a nanospray source (Waters Corp.) in positive ion mode. For all data sets 300 scans were averaged. The following parameters were applied for MS experiments: capillary voltage, 3 - 3.5 kV; sampling cone, 40 kV; extraction cone, 4.0 V; cone gas flow, 0 mL/h; trap collision energy (CE), 6.0 V; transfer CE, 4.0 V; trap gas flow, 1.5 mL/min. For tandem MS data collection the same parameters were used with the exception of the trap CE which was increased to 15 - 60 V. Mass spectra were calibrated externally in positive ion mode in the range of 100 ≤ m/z ≤ 5000 with a solution of sodium cesium iodide. Data were processed using Masslynx 4.1 software (Waters Corp., Manchester, UK).
5.2 Results and Discussion

5.2.1 Analysis of the Intact Light Chains

All of the recombinant chains contain free cysteine residues, which are normally involved in interchain disulfide bonds. In order to keep the chains from forming polymers, the samples were washed with a solution containing 10 nM of L-cysteine in 10 mM ammonium acetate. The inclusion of L-cysteine maintains the protein in a low oligomerization state.

First the intact L<sub>1</sub> sample was analyzed in both 50% acetonitrile (ACN) in water containing 0.1% formic acid (FA), and ammonium acetate pH 7 (Figure 5-1). The predominant species is the monomeric L<sub>1</sub> chain. Some dimer can also be observed. The distribution of the dimer resembles the one of a nonspecific dimer, as its peaks are observed in between the peaks for the monomer. However, since the dimerization of the light chain occurs through disulfide bridging, and that only a small surface area of the protein is excluded from the solvent upon dimer formation, it is likely that the observed dimer is actually specific.

Additionally, no shift in the charge state distribution is observed between the two different conditions. This is due to the presence of the intrachain disulfide bridge which stabilizes the structure of the protein. These supports the notion that the L<sub>1</sub> chain is properly folded with properly formed disulfide bonds, however additional experiments were required to confirm this conclusion.
5.2.2 Establishing the Disulfide Pattern of the Light Chains

5.2.2.1 Principle of the Experiment

As the different haptoglobin chains possess several cysteine residues, some involved in intra-chain disulfide bridging and others in inter-chain disulfide bonds (Figure 5-3 and Figure 5-4), it is important to identify the disulfide pattern of the recombinant chains. Determining the disulfide pattern will certainly help to confirm the proper folding of the protein. As mentioned in Chapter 4, many proteins foreign to the
organism used to produce them, do not display proper folding after expression. Several approaches to identify the disulfide patterns of the recombinant and native proteins exist as previously presented in Chapter 3.

The method chosen to verify the proper formation of the disulfide bonds was the labeling technique. Despite the fact that labeling with iodoacetamide can lead to incomplete reaction and/or reaction with other amino acids, it was still chosen as labeling reagent. The different chains were labeled either directly after washing the sample or after reaction with TCEP. The conditions for the reduction of the chain with TCEP were chosen based on a previous study which showed that these conditions allow for partial reduction of proteins, allowing for disruption of interchain disulfide bond only. The workflow for the experiment is presented in Figure 5-2.

**Figure 5-2:** Schematic of the workflow for mapping of the disulfide bonds
In both cases, we assumed that the cysteine residue involved in the crosslinking to the H-chain was free \((C_{73} \text{ for } L_1 \text{ and } C_{132} \text{ for } L_2)\) and that the disulfide bond responsible for the eventual oligomerization are formed through the correct cysteine residues \((C_{16} \text{ for both chains and } C_{75} \text{ in the case of } L_2)\) (Figure 5-3 and Figure 5-4). Concerning the first case, the hypothesis is that under limited reaction with TCEP only the disulfides involved in the oligomerization of the light chains will be reduced and thereby modified with iodoacetamide \((C_{16} \text{ and } C_{75} \text{ for } L_2)\).\(^{82}\) The cysteine residues involved in oligomerization and in interchain bridging with the H-chain \((C_{73} \text{ for } L_1 \text{ and } C_{132} \text{ for } L_2)\) should also be modified.

In the second case, involving the reaction with iodoacetamide only, we expected modification of the cysteine residues that should be free \((C_{73} \text{ for } L_1 \text{ and } C_{132} \text{ for } L_2)\). Only the cysteine involved in the interchain bridging with H-chain should be modified. The predicted outcome of the experiment is presented in Table B.1.

![Figure 5-3: Amino acid sequence of recombinant L₁. The disulfide pattern is represented on the figure. The yellow connector indicates the cysteine residues theoretically involved in intra disulfide bonds, also labeled in yellow. The cysteine residues labeled in red are those involved in interchain disulfide bonds, either with another light chain \((C_{16})\) or with the heavy chain \((C_{73})\). The sequence position these cysteine residues is indicated above them in the sequence.](image-url)
Figure 5-4: Amino acid sequence of recombinant L2. The disulfide pattern is represented on the figure. The yellow connector indicates the cysteine residues theoretically involved in intra disulfide bonds, also labeled in yellow. The cysteine residues labeled in red are the one involved in interchain disulfide bonds, either with another light chain (C_{16} or C_{75}) or with the heavy chain (C_{132}). The sequence position these cysteine residues is indicated above them in the sequence.

5.2.2.2  Disulfide Pattern of Recombinant L1

After the reaction, the samples were washed again with 100 mM AmmAc solution to remove any traces of reagents that could interfere with the results and then digested overnight with trypsin. The samples were analyzed on the Synapt HDMS in positive ion mode.

The results for L1 for the first reaction (TCEP followed by IAA) are presented in Figure 5-5. The peptides identified are indicated by numbers, which are summarized in Table 5.1. The first notable result is that the reaction was incomplete and some of the cysteine residues were not labeled as expected and some crosslinked peptides could be observed. Additionally, apart from the incomplete reaction, all the cysteine residues seemed to be labeled, whether it was a cysteine residue involved in intra- or interchain crosslinking. This result indicates that the intrachain disulfide bond is either solvent accessible, as the protein was kept in near physiological conditions (100 mM ammonium...
acetate pH 7) or the protein is misfolded. Furthermore, most of the identified peptides were fragmented by collision induced dissociation (CID) in order to confirm their identities and the location of the eventual modifications. One example of the tandem MS spectrum of the modified peptides is presented in Figure 5-6. The other spectra are presented in Appendix B.

**Figure 5-5:** ESI-MS of the tryptic digest of L1 after reaction with TCEP followed by IAA. A complete list of the peptides identified and the sequence coverage are provided in Table 5.1. The numbers of the peptides refer to the number assigned to each peptide in Table 5.1.
Table 5.1: List of the major peptides detected. The peptides are indicated on the mass spectrum in Figure 5-5. \( m/z \): \( m/z \) value of the peptide ion; \( z \): charge of the peptide ion; \((M+H)_{\text{meas}}\): mass of the singly protonated peptides calculated from the monoisotopic experimentally measured \( m/z \) value; \((M+H)_{\text{calc}}\): mass of the singly protonated peptides calculated from the sequence. The peptides containing a cysteine labeled with iodoacetamide are indicated with \( C_x^* \). “♦” indicates the peptides involved in intradisulfide bonds.

<table>
<thead>
<tr>
<th>Peak #</th>
<th>m/z</th>
<th>z</th>
<th>AA position &amp; cross linked</th>
<th>((M+H)_{\text{meas}})</th>
<th>((M+H)_{\text{calc}})</th>
<th>(\Delta(M))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>288.2134</td>
<td>+1</td>
<td>41 – 42</td>
<td>288.2134</td>
<td>288.2030</td>
<td>-0.01</td>
</tr>
<tr>
<td>2</td>
<td>688.3865</td>
<td>+1</td>
<td>56 – 60</td>
<td>688.3865</td>
<td>688.3777</td>
<td>-0.01</td>
</tr>
<tr>
<td>3</td>
<td>720.3500</td>
<td>+2</td>
<td>43 – 55</td>
<td>1439.7000</td>
<td>1439.6649</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>1439.7025</td>
<td>+1</td>
<td></td>
<td>1439.7025</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>739.4150</td>
<td>+1</td>
<td>78 – 84</td>
<td>739.4150</td>
<td>739.3733</td>
<td>-0.04</td>
</tr>
<tr>
<td>5</td>
<td>854.9477</td>
<td>+2</td>
<td>41 – 55</td>
<td>1708.8954</td>
<td>1708.8501</td>
<td>-0.05</td>
</tr>
<tr>
<td>6*</td>
<td>871.4694</td>
<td>+2</td>
<td>61 – 75</td>
<td>1741.9388</td>
<td>1743.8768</td>
<td>-1.94</td>
</tr>
<tr>
<td>7</td>
<td>929.4891</td>
<td>+2</td>
<td>61 – 75 (( C_{69}^* ) &amp; ( C_{73}^* ))</td>
<td>1856.9782</td>
<td>1858.0168</td>
<td>-1.04</td>
</tr>
<tr>
<td>8</td>
<td>1136.5392</td>
<td>+3</td>
<td>1 – 32 (( C_{16}^* ))</td>
<td>3407.6176</td>
<td>3406.5604</td>
<td>-1.06</td>
</tr>
<tr>
<td></td>
<td>1704.8151</td>
<td>+2</td>
<td></td>
<td>3407.6302</td>
<td></td>
<td>-1.07</td>
</tr>
<tr>
<td>9</td>
<td>1297.2833</td>
<td>+3</td>
<td>1 – 32 + 33 – 36</td>
<td>3889.8499</td>
<td>3888.7829</td>
<td>-1.07</td>
</tr>
<tr>
<td></td>
<td>1945.4257</td>
<td>+2</td>
<td></td>
<td>3889.8514</td>
<td></td>
<td>-1.07</td>
</tr>
<tr>
<td>10</td>
<td>1675.2853</td>
<td>+4</td>
<td>1 – 32 + 1 – 32 (( C_{16}^* ) – ( C_{16}^* ))</td>
<td>6698.1412</td>
<td>6697.0780</td>
<td>-1.06</td>
</tr>
<tr>
<td></td>
<td>2234.0581</td>
<td>+3</td>
<td></td>
<td>6699.1743</td>
<td></td>
<td>-2.10</td>
</tr>
</tbody>
</table>
Figure 5-6: Tandem MS of the m/z 1136.5392 peak \( \Delta z = +3 \) after MaxEnt3 transformation. The ladder of \( y \)-ions is indicated on the spectrum. The cysteine residue modified by carbamidomethylation is indicated by \( C^* \).

The result obtained for the second reaction (IAA only) is markedly different than previously. The data are presented in Figure 5-7 and a summary of the peptides identified is presented in Table 5.2. Several crosslinked peptides were identified, in particular the peptides at m/z 1675.1041 and at m/z 1710.6300. The former corresponds to the peptides containing the cysteine residue involved in the dimerization and the later the one containing the cysteine residues involved in the intrachain disulfide bond. The sequence of both peptides was determined by tandem MS (Figure 5-8 and Figure 5-9). Also, the peptides containing the cysteine residues involved in the intrachain crosslinking was found to be modified by carbamidomethylation. This result is expected as the peptide 61
– 77 contains two cysteine residues, one involved in the intrachain crosslinking and one involved in the linkage to the H-chain. The tandem MS experiment confirmed the position of both the crosslinking and the modification (Figure 5-9). Cysteine C₆₀ is involved in the crosslinking while the C₇₅ is modified. This result is consistent with the expected outcome for the reaction (Figure 5-3 and Table B.1).

However, some miscrosslinked peptides can be observed such as peptide #9 (Table 5.2), due to either the presence of some L₁ species misfolded in solution or to the possible scrambling of disulfides bonds in solution.

**Figure 5-7:** ESI-MS of the tryptic digest of L₁ after reaction with IAA. A complete list of the peptides identified and the sequence coverage are provided in Table 5.2. The numbers of the peptides refer to the number assigned to each peptide in Table 5.2.
**Table 5.2**: List of the major peptides detected. The peptides are indicated on the mass spectrum in Figure 5-7. \( m/z \): \( m/z \) value of the peptide ion; \( z \): charge of the peptide ion; \((M+H)_{\text{meas.}}\): mass of the singly protonated peptides calculated from the monoisotopic experimentally measured \( m/z \) value; \((M+H)_{\text{calc}}\): mass of the singly protonated peptides calculated from the sequence. The peptides containing a cysteine labeled with iodoacetamide are indicated with \( C_x^* \). “♦” indicates the peptides involved in intradisulfide bond.

<table>
<thead>
<tr>
<th>Peak #</th>
<th>( m/z )</th>
<th>( z )</th>
<th>AA position &amp; cross linked</th>
<th>((M+H)_{\text{meas.}})</th>
<th>((M+H)_{\text{calc}})</th>
<th>( \Delta(M) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>288.2197</td>
<td>+1</td>
<td>41 – 42</td>
<td>288.2197</td>
<td>288.2030</td>
<td>-0.02</td>
</tr>
<tr>
<td>2</td>
<td>688.4156</td>
<td>+1</td>
<td>56-60</td>
<td>688.4156</td>
<td>688.3777</td>
<td>-0.04</td>
</tr>
<tr>
<td>3</td>
<td>720.3500/1439.7600</td>
<td>+2 +1</td>
<td>43 – 55</td>
<td>1439.7000/1439.7600</td>
<td>1439.6649</td>
<td>-0.04/-0.1</td>
</tr>
<tr>
<td>4</td>
<td>739.4150</td>
<td>+1</td>
<td>78-84</td>
<td>739.4150</td>
<td>739.3733</td>
<td>-0.04</td>
</tr>
<tr>
<td>5</td>
<td>854.9693</td>
<td>+2</td>
<td>41 – 55</td>
<td>1708.9386</td>
<td>1708.8501</td>
<td>-0.09</td>
</tr>
<tr>
<td>6*</td>
<td>871.4929</td>
<td>+2</td>
<td>61 – 75</td>
<td>1741.9858</td>
<td>1743.8768</td>
<td>1.89</td>
</tr>
<tr>
<td>7</td>
<td>1136.5765/1704.3583</td>
<td>+3 +2</td>
<td>1 – 32 ( (C_{16}^*) )</td>
<td>3407.7295/3407.7166</td>
<td>3406.5604</td>
<td>-1.17/-1.16</td>
</tr>
<tr>
<td>8</td>
<td>1170.6300</td>
<td>+2</td>
<td>33 – 36 + 61 – 77 ( (C_{35}^* C_{69} &amp; C_{73}^*) )</td>
<td>2340.2600</td>
<td>2340.1907</td>
<td>-0.07</td>
</tr>
<tr>
<td>9</td>
<td>1297.3231</td>
<td>+3</td>
<td>1 – 32 + 33 – 36</td>
<td>3889.9693</td>
<td>3888.7829</td>
<td>-1.19</td>
</tr>
<tr>
<td>10</td>
<td>1675.1041/2233.1172</td>
<td>+4 +3</td>
<td>1 – 32 + 1 – 32 ( (C_{16}^* C_{16}) )</td>
<td>6697.4164/6697.3516</td>
<td>6697.0780</td>
<td>-0.34/-0.27</td>
</tr>
</tbody>
</table>
Figure 5-8: Tandem MS of the \(m/z\) 1675.1041 peak \((z = +4)\) after MaxEnt3 transformation. The \(y\)-ions identified are indicated on the spectrum. The sequence of the cross-linked peptide including its fragmentation pattern is presented in the figure.
Figure 5-9: Tandem MS of the m/z 1170.6300 peak (z = +2) after MaxEnt3 transformation. The y- and b- ions identified are indicated on the spectrum. The sequence of the cross linked peptide including its fragmentation pattern is presented in the figure. The carbamidomethylated cystein residue is indicated by a C* in the sequence.

5.2.2.3 Disulfide Pattern of Recombinant L2

The same process was followed for establishing L2 disulfide pattern. However, none of the reactions gave relevant results. The reaction with TCEP followed by IAA modified all the cysteine residues (Figure 5-10). The results are summarized in Table 5.3. Tandem MS experiments were done to confirm the sequence of some modified peptides. Figure 5-11 shows an example of the fragmentation of a peptide containing a glutamate residue modified with iodoacetamide. It has been reported that excess of IAA can lead to the modification of amino acids other than cysteine.79, 106 The second reaction did not
provide conclusive results either. Most of the cysteine residues were modified by iodoacetamide (Table 5.4). Tandem MS experiments were carried out on the majority of the peptides and the data are presented in Appendix B. The outcome of the reactions could be due to the fact that dithiotreitol was added during the concentration and washing step of the sample as L$_2$ was forming insoluble polymers, thus preventing the disulfide bonds to remain correctly formed. This highlights the importance of maintaining lower concentration for L$_2$ as well as carrying out additional procedures for renaturation. One possible method is to dialyze dilute solutions of protein against decreasing concentration of dithiothreitol (DTT) or any disulfide reducing agent over an extended period of time (perhaps days) in order to allow for proper folding of the protein.$^{107}$

![Figure 5-10: ESI-MS of the tryptic digest of L$_2$ after reaction with TCEP followed by IAA. A complete list of the peptides identified and the sequence coverage are provided in Table 5.3. The numbers of the peptides refer to the number assigned to each peptide in Table 5.3.](image-url)
Table 5.3: List of the major peptides detected. The peptides are indicated on the mass spectrum in Figure 5-10. \( m/z \): \( m/z \) value of the peptide ion; \( z \): charge of the peptide ion; \( (M+H)_{\text{meas.}} \): mass of the singly protonated peptides calculated from the monoisotopic experimentally measured \( m/z \) value; \( (M+H)_{\text{calc.}} \): mass of the singly protonated peptides calculated from the sequence. The peptides containing a cysteine labeled with iodoacetamide are indicated with C\(_x\)* and the one containing a different amino acid being modified X\(_x\)*. “♦” indicates the peptides involved in intradisulfide bond.

<table>
<thead>
<tr>
<th>Peak #</th>
<th>( m/z )</th>
<th>( z )</th>
<th>AA position</th>
<th>( [M+H]_{\text{meas.}} )</th>
<th>( [M+H]_{\text{calc.}} )</th>
<th>( \Delta M )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>288.2009</td>
<td>+1</td>
<td>41 – 42; 100 – 101</td>
<td>288.2009</td>
<td>288.2030</td>
<td>-0.002</td>
</tr>
<tr>
<td>2</td>
<td>587.2899</td>
<td>+1</td>
<td>37 – 40; 96 – 99</td>
<td>587.2899</td>
<td>587.2824</td>
<td>-0.01</td>
</tr>
<tr>
<td>3</td>
<td>688.3672</td>
<td>+1</td>
<td>56 – 60; 115 – 119</td>
<td>688.3865</td>
<td>688.3777</td>
<td>-0.01</td>
</tr>
<tr>
<td>4</td>
<td>720.3278;</td>
<td>+2</td>
<td>102 – 114</td>
<td>1439.6556; 1439.6747</td>
<td>1439.6649; 1439.6649</td>
<td>-0.01; -0.004</td>
</tr>
<tr>
<td></td>
<td>1439.6605</td>
<td>+1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>739.3549</td>
<td>+1</td>
<td>137 – 143</td>
<td>739.3549</td>
<td>739.3733</td>
<td>-0.02</td>
</tr>
<tr>
<td>6</td>
<td>748.8302</td>
<td>+2</td>
<td>102 – 114 (D(_{106})*)</td>
<td>1496.6604</td>
<td>1496.7349</td>
<td>-0.07</td>
</tr>
<tr>
<td>7</td>
<td>790.8879</td>
<td>+2</td>
<td>41 – 54</td>
<td>1580.7758</td>
<td>1580.7915</td>
<td>-0.02</td>
</tr>
<tr>
<td>8</td>
<td>796.3832</td>
<td>+1</td>
<td>137 – 143 (C-term*)</td>
<td>796.3832</td>
<td>796.4433</td>
<td>-0.06</td>
</tr>
<tr>
<td>9</td>
<td>854.9153</td>
<td>+2</td>
<td>100 – 114</td>
<td>1708.8306</td>
<td>1708.8501</td>
<td>-0.02</td>
</tr>
<tr>
<td>10</td>
<td>856.4147</td>
<td>+1</td>
<td>37 – 42; 96 – 101</td>
<td>856.4147</td>
<td>856.4676</td>
<td>-0.05</td>
</tr>
<tr>
<td>11*</td>
<td>871.4476</td>
<td>+2</td>
<td>120 – 136</td>
<td>1741.8952</td>
<td>1743.8768</td>
<td>-1.98</td>
</tr>
<tr>
<td>12</td>
<td>950.0739;</td>
<td>+3</td>
<td>66 – 91</td>
<td>2848.2217; 2847.2608</td>
<td>2848.3029</td>
<td>-0.33; -1.04</td>
</tr>
<tr>
<td></td>
<td>1424.1304</td>
<td>+2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1106.4824;</td>
<td>+3</td>
<td>61 – 91</td>
<td>3317.4472; 3317.4698</td>
<td>3318.5518</td>
<td>-1.10; -1.08</td>
</tr>
<tr>
<td></td>
<td>1659.2349</td>
<td>+2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1143.1498</td>
<td>+3</td>
<td>61 – 91 (C(<em>{69}) * or C(</em>{75}) *)</td>
<td>3427.4494</td>
<td>3427.5989</td>
<td>-0.15</td>
</tr>
<tr>
<td>15</td>
<td>1161.8271</td>
<td>+3</td>
<td>61 – 91 (C(<em>{69}) * &amp; C(</em>{75}) *)</td>
<td>3482.4813</td>
<td>3484.6689</td>
<td>-2.19</td>
</tr>
<tr>
<td>16</td>
<td>1297.2303</td>
<td>+3</td>
<td>1 – 32 + 92 – 95</td>
<td>3889.6909</td>
<td>3888.7829</td>
<td>-0.91</td>
</tr>
<tr>
<td>17</td>
<td>1311.5869</td>
<td>+1</td>
<td>43 – 54</td>
<td>1311.5869</td>
<td>1311.6063</td>
<td>-0.02</td>
</tr>
<tr>
<td>18</td>
<td>1675.4662</td>
<td>+4</td>
<td>1 – 32 + 1 – 32</td>
<td>6698.8648</td>
<td>6697.0780</td>
<td>1.79</td>
</tr>
<tr>
<td>19</td>
<td>1689.7146</td>
<td>+4</td>
<td>1 – 32 + 1 – 32 (D*)</td>
<td>6754.8584</td>
<td>6754.1480</td>
<td>-0.71</td>
</tr>
</tbody>
</table>
**Figure 5-11:** Tandem MS of the \( m/z 748.8302 \) peak (\( z = +2 \)) after MaxEnt3 transformation. The ladder of \( y \)-ions is indicated on the spectrum. The aspartic acid residue modified by carboxamidomethylation is indicated by \( D^* \).

**Figure 5-12:** ESI-MS of the tryptic digest of \( L_2 \) after reaction with IAA. A complete list of the peptides identified and the sequence coverage are provided in Table 5.4. The numbers of the peptides refer to the number assigned to each peptide in Table 5.4.
Table 5.4: List of the major peptides detected. The peptides are indicated on the mass spectrum in Figure 5-12. m/z: m/z value of the peptide ion; z: charge of the peptide ion; (M+H)\textsubscript{meas}: mass of the singly protonated peptides calculated from the monoisotopic experimentally measured m/z value; (M+H)\textsubscript{calc}: mass of the singly protonated peptides calculated from the sequence. The peptides containing a cysteine labeled with iodoacetamide are indicated with C\textsubscript{x} * and the one containing a different amino acid being modified X\textsubscript{x} *. “♦” indicates the peptides involved in intradisulfide bond.

<table>
<thead>
<tr>
<th>Peak #</th>
<th>m/z</th>
<th>z</th>
<th>AA position</th>
<th>[M+H]\textsubscript{meas}</th>
<th>[M+H]\textsubscript{calc}</th>
<th>∆M</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>288.2072</td>
<td>+1</td>
<td>41 – 42, 100 – 101</td>
<td>288.2072</td>
<td>288.2030</td>
<td>0.004</td>
</tr>
<tr>
<td>2</td>
<td>489.2704</td>
<td>+1</td>
<td>61 – 65, 120 – 124</td>
<td>489.2704</td>
<td>489.2667</td>
<td>0.004</td>
</tr>
<tr>
<td>3</td>
<td>587.2890</td>
<td>+1</td>
<td>37 – 40, 96 – 99</td>
<td>587.2890</td>
<td>587.2824</td>
<td>0.007</td>
</tr>
<tr>
<td>4</td>
<td>688.3672</td>
<td>+1</td>
<td>56 – 60, 115 – 119</td>
<td>688.3672</td>
<td>688.3777</td>
<td>-0.01</td>
</tr>
<tr>
<td>5</td>
<td>720.3278</td>
<td>+2</td>
<td>102 – 114</td>
<td>1439.6556</td>
<td>1439.6649</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>1439.6747</td>
<td>+1</td>
<td></td>
<td>1439.6747</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>739.3549</td>
<td>+1</td>
<td>137 – 143</td>
<td>739.3549</td>
<td>739.3733</td>
<td>-0.02</td>
</tr>
<tr>
<td>7</td>
<td>854.9153</td>
<td>+2</td>
<td>100 – 114</td>
<td>1708.8306</td>
<td>1708.8501</td>
<td>-0.02</td>
</tr>
<tr>
<td>8</td>
<td>871.4476</td>
<td>+2</td>
<td>120 – 136</td>
<td>1741.8952</td>
<td>1743.8768</td>
<td>-1.98</td>
</tr>
<tr>
<td>9</td>
<td>929.9725</td>
<td>+2</td>
<td>120 – 136 (C\textsubscript{128} – C\textsubscript{132})</td>
<td>1858.945</td>
<td>1858.0168</td>
<td>0.93</td>
</tr>
<tr>
<td>10</td>
<td>949.7783</td>
<td>+3</td>
<td>66 – 91</td>
<td>2847.3349</td>
<td>2848.3029</td>
<td>-0.97</td>
</tr>
<tr>
<td></td>
<td>1424.1304</td>
<td>+2</td>
<td></td>
<td>2847.2608</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>973.2065</td>
<td>+4</td>
<td>1 – 32, 92 – 95</td>
<td>3889.8260</td>
<td>3889.8103</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>1297.2701</td>
<td>+3</td>
<td></td>
<td>3889.7536</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1945.3768</td>
<td>+2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1106.1760</td>
<td>+3</td>
<td>61 – 91</td>
<td>3316.5280</td>
<td>3318.5518</td>
<td>-2.02</td>
</tr>
<tr>
<td></td>
<td>1658.7544</td>
<td>+2</td>
<td></td>
<td>3316.5088</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1136.5143</td>
<td>+3</td>
<td>1 – 32 (C\textsubscript{16} *)</td>
<td>3407.5429</td>
<td>3407.5642</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>1704.2821</td>
<td>+2</td>
<td></td>
<td>3407.5642</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1311.6288</td>
<td>+1</td>
<td>43 – 54</td>
<td>1311.6288</td>
<td>1311.6063</td>
<td>0.02</td>
</tr>
<tr>
<td>15</td>
<td>1675.7682</td>
<td>+4</td>
<td>1 – 32, 1 – 32 (C\textsubscript{16} – C\textsubscript{16})</td>
<td>6698.8648</td>
<td>6697.0780</td>
<td>1.99</td>
</tr>
</tbody>
</table>
5.2.3 Molecular Modeling

In order to facilitate the interpretation of the future data, it was decided to develop a structural model for human haptoglobin. The PDB file corresponding to the homology model was generated using Swiss model\textsuperscript{108, 109, 110} (http://swissmodel.expasy.org/) and PyMol (http://www.pymol.org/) was used for visualization. The automated mode was chosen, meaning that the software will automatically find the protein template that the model will be based on by searching for the highest sequence homology, among proteins of known structure.

The model for Hp1 is based on the recently published structure of porcine Hp\textsuperscript{111} (pdb: 4F4O) while Hp2 is based on the C1R protease (pdb: 1GPZ). Hp2 could not be generated using the same model as Hp1, as porcine Hp exists only in one phenotype, Hp1. The sequence identity between human Hp1 and porcine Hp is about 80%; and 30% between Hp2 and C1R. The sequence alignments are presented in the Appendix B.

In the models, all the disulfide bridges are formed for the various chains. L\textsubscript{1} possesses two free cysteine residues: one at the N-terminal involved in the dimerization and one at the C-terminal involved in the bridging with H-chain. The L\textsubscript{2} chain possesses three free cysteine residues, one involved in bridging with the H-chain and the other two in the oligomerization. The H-chain also possesses the appropriate intrachain disulfide bonds as well as the free cysteine residue involved in the crosslinking with the light chain. Concerning the secondary structure, the light chains are solely β strands and loops while H-chain shows a mix between α helices and β strands. Based on the model, the H-chain is more structured that the light chains.
In Figure 5-13, the regions of $L_1$ and $L_2$ involved in the dimerization are represented in red. They have been reported in both structures by extension from the interaction regions identified in porcine Hp. On the H-chain the loop domain involved in the binding to CD163 receptor is indicated in orange. The potentially N-glycosylated asparagine residues are represented in green. They are all solvent exposed in the model.

Figure 5-13: Molecular modeling of haptoglobin chains. A: $L_1$, B: $L_2$ and C: H. The cysteine residues are indicated as yellow sticks. The regions in red on $L_1$ (A) and $L_2$ (B) represent the region involved in the dimerization of haptoglobin according to the crystal structure of porcine Hp (PDB: 4F4O). The potential N-glycosylation sites are indicated as green sticks on H (C) and the orange loop is the binding domain for the CD163 receptor.
The model obtained is very similar to the template (Figure 5-14). The H-chains differ by the central helix which is longer in the porcine Hp. The main difference between the two structures is the light chain. The N-terminal β strand involved in dimerization is much shorter in the model of the human Hp than in the porcine structure.

**Figure 5-14:** Cartoon representation of the model of human Hp1 (A) and the structure of porcine Hp1 (pdb: 4F4O) (B). The light chain is represented in cyan and the heavy chain in grey. The cysteine residues are represented by yellow sticks and the potential N-glycosylation sites by green sticks.
Figure 5-15 is a model representation of the human H-chain only. The Hb binding site identified from the porcine H-chain was transposed to the model of the human which is represented in pink. The binding site for Hb on H-chain is quite extended and consists of mostly hydrophobic interactions, which justifies the high binding affinity ($K_d \approx 10^{-15}$ M). Additionally, the representation shows that the potentially glycosylated asparagine residues are solvent exposed, which further supports the validity of the model.

**Figure 5-15:** Surface representation of the model of the H-chain of human Hp. The glycosylation sites in blue and the binding site of CD163 receptor in green. The amino acids represented in pink are the one theoretically involved in the binding with Hb.
5.3 Conclusion

The analysis of the intact recombinant L₁ chain resulted in data that are consistent with proteins stabilized by disulfide bonds, which indicates that this chain is likely to have a native folding.

The disulfide patterns of both light chains were analyzed using a method involving chemical modification of the cysteine residues by carbamidomethylation. The conclusion from these experiments is very different for the two cases. For L₁ data support that there is likely a mixture of properly folded and misfolded protein, while for L₂ it is difficult to draw any conclusion. It is likely that the protein is not properly folded and lacks proper disulfide pattern. Additionally, we were able to build a good model for human Hp and its individual chains, useful for further studies on Hp.
Chapter 6

Protein Disorder Determination of DnaT by Mass Spectrometry

During the lifetime of a cell many external parameters may result in the damage of DNA and thereby challenge the integrity of processes like DNA replication. In E. coli, the replication mechanism starts at a defined origin known as OriC. Replication arrest can occur at a region away from this origin. As the regular process of replication is highly regulated and sequence specific, it is necessary for the cell to rely on a different machinery to restart the replication. This assembly involved in the replication restart pathway is called the replication restart/replication rescue machinery and is non-OriC dependent. Seven different proteins comprise this assembly: PriA, PriB, DnaT, single stranded DNA binding proteins (SSBs), DnaB and DnaC (termed the preprimosome assembly) and finally upon the addition of DnaG, the primosomal assembly is formed. This assembly is also known as the ΦX174 primosomal assembly and provides the helicase and primase activity necessary to restart replication at the fork. PriA binds to DNA at a D-loop structure, and unwinds the double-stranded DNA (dsDNA) to allow SSB proteins to assist in the remodeling of the fork. Upon binding of PriA to DNA, PriB can bind to PriA to stabilize the protein-DNA complex. Thereafter, DnaT binds to PriA. DnaT is also able to bind to PriA alone, but its concentration has to be ten times higher.
for this process to occur. DnaB is then loaded and interacts with the PriA/PriB/DnaT ternary complex. This step requires the hydrolysis of one molecule of ATP by DnaC. DnaB unwinds the lagging strand DNA template in an opposite direction to PriA, allowing for the replisome to be stabilized. Finally, DnaG is loaded onto the lagging strand of the DNA through its interaction with DnaB. It is responsible for the synthesis of the primers in both strands of the fork, which allows for their extension by DNA polymerase III.

Very little is known about DnaT structure and its mode of binding. It is thought that it binds to the PriA-PriB complex as a monomer or trimer. Its real function also remains obscure albeit being an essential constituent of the replication restart mechanism.

Dr. Mueser’s former students have been studying this pathway and this particular protein by X-ray crystallography for many years, but so far attempts at structure determination of DnaT have been unsuccessful. Indeed, X-ray diffraction data collected on DnaT revealed that two crystalline species coexisted, even though the crystal trays were set up with pure protein. Preliminary experiments led to the discovery that DnaT was cleaved into two species during the crystallization process. Hence, it appeared necessary to conduct experiments in order to determine the part of cleavage. Based on the hypothesis that DnaT possesses a more flexible domain which could account for the fragmentation of the protein, top-down MS seemed suitable to map this region. In fact, less structured regions of a protein do fragment more readily that compact domains, allowing for the determination of these more flexible regions. Consequently, tandem MS experiments on intact DnaT proteins were conducted.
6.1 Material and Methods

6.1.1 Protein Expression

A small amount of cells from Rosetta blue DE3 strain were taken from a previously made glycerol stock and were transferred to a flask containing 150 mL of 25 g/L LB medium, 1 mM Ampicillin, and Chloramphenicol. The culture was incubated overnight at 37 °C in a New Brunswick Scientific Innova 4000 Incubator shaker. The following morning, 50 mL of the culture were added to each of three 1 L volumes of 25 g/L LB media containing 1 mM Ampicillin and Chloramphenicol. The cells were grown at 37 °C in the shaker. When their OD\textsubscript{600} reached 0.4 - 0.6, the expression was induced by adding 1 mM IPTG. The culture was left to grow at 37 °C for three additional hours. The cells were then harvested by centrifugation at 4000 rpm at 4 °C for 20 min using a Beckman Coulter™ TJ-25 centrifuge. The supernatant was discarded and the cell pellet was recovered and weighed.

6.1.2 Lysis and Ammonium Sulfate Precipitation

For each gram of cells pellet, 10 mL of lysis buffer (Table 6.1) was added. The mixture was then stirred for 1 hour and sonicated for 2 min, both on ice, followed by centrifugation for 20 min at 10,000 rpm and 4 °C. The supernatant or lysate was recovered. At this point 0.3% PEI solution was added and the suspension was mixed for a few minutes. The precipitate was separated by centrifugation at 10,000 rpm at 4 °C for 30 min, and the supernatant decanted and kept on ice. Proteins were precipitated by adding 3
M (NH₄)₂SO₄ to the supernatant, and the mixture was stirred on ice for 30 min. The precipitate was recovered by centrifugation at 10,000 rpm for 25 min. TEDG buffer (Table 6.2) was added until the entire pellet redissolved.

**Table 6.1:** Composition of Lysis buffer

<table>
<thead>
<tr>
<th>Lysis</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris-HCl pH 7.5</td>
<td>25 mM</td>
</tr>
<tr>
<td>KCl</td>
<td>300 mM</td>
</tr>
<tr>
<td>EDTA</td>
<td>5 mM</td>
</tr>
<tr>
<td>BME</td>
<td>2 mM</td>
</tr>
<tr>
<td>Sucrose</td>
<td>10% w/v</td>
</tr>
</tbody>
</table>

**Table 6.2:** Composition of TEDG buffer

<table>
<thead>
<tr>
<th>TEDG</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EDTA</td>
<td>1 mM</td>
</tr>
<tr>
<td>Tris-HCl pH 7.5</td>
<td>50 mM</td>
</tr>
<tr>
<td>BME</td>
<td>5 mM</td>
</tr>
<tr>
<td>Glycerol</td>
<td>15% w/v</td>
</tr>
</tbody>
</table>

### 6.1.3 Purification of DnaT

Purification involved two ion exchange chromatography steps. The first was a low resolution anion exchange column Q-Sepharose (Amersham Biosciences). The column was equilibrated with buffer A. Before loading the sample, the conductivity of the solution was adjusted to about 8 mS/cm. Recombinant DnaT was eluted using an increasing percentage of buffer B containing a high salt concentration. Buffer contents are summarized in Table 6.3. The fractions containing the protein of interest were pooled and dialyzed overnight against buffer A at 4 °C with medium stirring using SnakeSkin® dialysis tubing (Thermo Scientific) with a 10 kDa cut off.
The next step was a high resolution cation exchange column POROS HS (Applied Biosystems). The same buffers as step one were used. The sample was also eluted with an increasing gradient of buffer B. The fractions collected were then pooled together and concentrated using a Millipore™ centrifugal unit with a 10 kDa cut off to a final volume of 1 mL. Finally, 15% glycerol was added and the sample was divided in two 500 mL aliquots and flash frozen in dry ice before storage at –80 °C.

**Table 6.3: Composition of HPLC purification buffers**

<table>
<thead>
<tr>
<th></th>
<th>Buffer A</th>
<th>Buffer B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris-HCl</td>
<td>25 mM</td>
<td>25 mM</td>
</tr>
<tr>
<td>NH₄Cl</td>
<td>50 mM</td>
<td>1.5 M</td>
</tr>
<tr>
<td>Conductivity</td>
<td>8 mS/cm</td>
<td>120 mS/cm</td>
</tr>
</tbody>
</table>

6.1.4 Sample Preparation for Mass Spectrometry Experiments

Prior to mass spectrometry analysis, the sample was thawed and salts were removed by buffer exchange using an Amicon™ Ultra 0.5 centrifugal filter from Millipore™ with a 10 kDa cut out against 100 mM ammonium acetate. The DnaT sample was then diluted to a final concentration of either 5 µM in 50% acetonitrile (ACN) containing 0.1% formic acid (FA) (denaturing solution) or 10 µM in 10 mM ammonium acetate (AmmAc) pH 7.
6.1.5 Mass Spectrometry Experiments

Data collection was carried out on a Synapt HDMS quadrupole time-of-flight mass spectrometer equipped with a nanospray source (Waters) in positive ion mode. For all data sets 300 scans were averaged. The following parameters were applied for MS experiments: capillary voltage, 3 - 3.5 kV; sampling cone, 40 kV; extraction cone, 4.0 V; cone gas flow rate, 0 - 40 mL/h; trap collision energy (CE), 6.0 V; transfer CE, 4.0 V; trap gas flow rate, 1.5 mL/min. For tandem MS data collection the same parameters were used with the exception of the trap CE which was increased to 15 - 60 V. For the IMS experiments the IMS gas flow was 20 - 22 mL/min. Mass spectra were calibrated externally in positive ion mode in the range of $100 \leq m/z \leq 5000$ with a solution of sodium cesium iodide. Data were processed using Masslynx 4.1 software (Waters Corp.).

6.2 Results and Discussion

6.2.1 Expression and Purification of DnaT

DnaT protein was expressed and purified as described previously in the experimental section. First, Q-sepharose cation exchange chromatography was run (Figure 6-1) followed by a POROS HS anion exchange (Figure 6-2). The purity of the final protein was confirmed by SDS PAGE (Figure 6-3). The protein appears to be over 95% pure, thus no additional purification step was required for MS analysis.
Figure 6-1: Chromatogram of DnaT purification by Q-sepharose ion exchange

Figure 6-2: Chromatogram of DnaT purification by POROS HS ion exchange
6.2.2 DnaT: Degradation Products

Over the summer, a student was in charge of the expression and purification of DnaT in order for our lab to conduct the experiments required. However, the purification was delayed, and the protein was kept refrigerated at 4 °C for about a week. Upon ESI-MS analysis in 50% acetonitrile (ACN) containing 0.1% formic acid (FA), five main species were identified (Figure 6-4); two larger species, two smaller species and the intact protein. Using ProtParam tool online software (http://web.expasy.org/protparam/) the masses of the fragments were correlated to potential portions of the DnaT sequence. Based on this analysis, it seemed that two of the fragment species corresponded to the N-terminal domain while the two others corresponded to the C-terminal domain of DnaT (Table 6.4).
Figure 6-4: Top: ESI-MS of the degraded DnaT sample in 50% ACN containing 0.1% FA. Bottom: deconvolution of the ESI-MS of degraded DnaT data. The colors for each charge state distribution on the mass spectrum is correlated to the same color as the masses presented in the deconvoluted mass spectrum (bottom).
**Table 6.4:** Masses of the DnaT fragments correlated to their possible sequence. $M_{\text{calc}}$: calculated monoisotopic mass of the fragment from the protein sequence. $\Delta(M)$: mass error between the experimental mass and the $M_{\text{calc}}$.

<table>
<thead>
<tr>
<th>Mass</th>
<th>AA number</th>
<th>$M_{\text{calc}}$</th>
<th>$\Delta(M)$</th>
<th>Possible Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>7037.0005</td>
<td>1 - 68</td>
<td>7036.0</td>
<td>1</td>
<td>SSRVLTPDVV GIDALVHDHQ TVLAKAEGGV VAVFANNAPA FYAVTPARLA ELALEEKLR PGSDVAL</td>
</tr>
<tr>
<td>8154.0005</td>
<td>1 - 77</td>
<td>8153.2</td>
<td>0.8</td>
<td>SSRVLTPDVV GIDALVHDHQ TVLAKAEGGV VAVFANNAPA FYAVTPARLA ELALEEKLA RSGSDVALDD QLQEPQ</td>
</tr>
<tr>
<td>11190.0010</td>
<td>78 - 178</td>
<td>11188.7</td>
<td>1.3</td>
<td>AAPVAVPMGK FAMYPDWQPD ADFIRLAAALW GVALREPVTTE EELASFIAYW QAEGKVFHHV QQWQKLARSL QIGRASNGGL PKRDVNTVSEP DSQIPPGFR G</td>
</tr>
<tr>
<td>12307.5010</td>
<td>69 - 178</td>
<td>12305.8</td>
<td>1.7</td>
<td>DDQLQEPQQA APVAVPMGKF AMYPDWQPDA DFIRLAAALWG VALREPVTTE ELASFIAYWQ AEGKVFHHVQ WQWQKLARSLQ IGRASNGGLP KRDVNTVSEP DSQIPPGFRG</td>
</tr>
</tbody>
</table>
Additionally, freshly purified DnaT was analyzed under different conditions: 50% acetonitrile (ACN) containing 0.1% FA, 10 mM ammonium acetate pH 3 and 10 mM ammonium acetate pH 7. The mass spectra are shown in Figure 6-5, where only minor traces of degradation were detected. Moreover, two different charge state distributions are observed even at near-physiological pH conditions accounting for two different protein conformations in solution. One is narrow and centered in the low m/z range, characteristic of more denatured protein. The other charge state distribution is broader and centered in the higher m/z range, accounting for a more native-like protein structure. The amount of unfolded species decreases as the pH of the solution decreases. At pH 7 the more folded species is predominant. Also, a homodimeric species can be observed, at low abundance at pH 3, which increases at pH 7. This result contradicts the recent findings by Szymanski et al. who observed trimeric species in analytical centrifugation experiments. They also found that the trimer is formed through weak interactions of a monomer with a dimer formed through strong interactions. Additionally, after removal of a C-terminal region, they only observed dimerization. From these experiments they concluded that the dimerization occurs at the N-terminal domain of DnaT, after which the last monomer associate through interaction with its C-terminus.35
From these experiments, one can conclude that the degradation of DnaT is time dependent and the phenomenon illustrated in the spectrum in Figure 6-4 could be similar to what occurred during the crystallization process. Moreover, DnaT appears to have some intrinsically disordered regions that could prevent the protein from uniformly packing into the crystal lattice. In order to establish what these regions are, collision induced dissociation (CID) experiments on the intact protein were carried out.
6.2.3 Identification of the Flexible Domain of DnaT

Tandem MS experiments on intact protein, akin to the top-down approach for protein sequencing, are based on the hypothesis that more flexible regions of a protein, fragment more readily than more structured regions. This hypothesis is similar to the one made for limited proteolysis experiments, however without the limitation of protease specificity (Figure 6-6). In the case of DnaT, the naturally occurring fragmentation over time is thought to happen due to the presence of a more unstructured region.

Figure 6-6: Schematic showing how CID MS experiments can be used to identify flexible regions in proteins.
The first experiment was carried out under near-native conditions, ammonium acetate pH 7. The mass spectrum of the intact species obtained for this experiment is shown in Figure 6-7. The peak corresponding to \( m/z \ 2149.044 \) (\( z = +9 \)), accounting for native state protein, was selected for fragmentation. The fragmentation results are shown in Figure 6-8. Mainly b- ions were identified, although some y-ions could also be observed. Fragmentation occurred readily at the N-terminal domain of DnaT and also a little at the C-terminal domain. One possible conclusion that can be drawn is that the more disordered region is localized at the N-terminal domain of the protein up to around the 33\(^{\text{rd}}\) amino acid residue in DnaT sequence. The fragmentation that occurred at the C-terminal region of DnaT might be due to the innate flexibility existing in many protein sequences (Figure 6-9).

These results are somewhat different from what was observed recently by Szymanski et al. Upon limited proteolysis experiments with trypsin they observed a cleavage at the C-terminus of DnaT after K\(_{159}\). The observed results could be different due to the method used, and the lysine residue at the N-terminal domain may not be solvent accessible reducing the probability of cleavage. In addition, the fragmentation data could reveal the existence of the N-terminal region involved in the dimerization.\(^{34,35}\)
Figure 6-7: ESI-MS of intact DnaT in ammonium acetate pH7. The species fragmented is indicated by a red arrow.

Figure 6-8: Tandem mass spectrum of the species at $m/z$ 2149.044 ($z = +9$). The peak corresponding to the parent ion is indicated by a red star and b- and y-ions are indicated with red or green arrows, respectively.
Figure 6-9: Representation of the fragmentation pattern obtained from the native species (Figure 6-8) on the DnaT sequence. The b-ions identified are indicated by the red connectors while the y-ions are represented by the green connectors.

As folded species was also observed in denaturing conditions (organic solvent containing acid), it was necessary to perform the experiment on the species observed in this condition to confirm that the fragmented species was natively folded. The results are shown in Figure 6-10 and Figure 6-11 through Figure 6-14. In the mass spectrum of the intact species, two species were observed due to adduct formation. Both species were confirmed as DnaT from the tandem MS data. A third smaller species was also identified, its mass matched one of the fragments observed in the degraded sample. The tandem MS data presented in Figure 6-14 confirmed the sequence of the species. The fragmentation patterns for each of the species are presented in Appendix C.
Figure 6-10: ESI-MS of the intact DnaT protein in denaturing solution. The fragmented peaks are indicated by red arrows and their charges above the corresponding peaks.

Figure 6-11: Tandem mass spectrum of \textit{m/z} 1215.2977 (\textit{z} = +16) specie (peak 1). The peak corresponding to the parent ion is indicated by a red star and \textit{b}- and \textit{y}-ions are indicated by a red or a green arrow, respectively.
**Figure 6-12:** Tandem mass spectrum of $m/z$ 1208.6812 ($z = +16$) species (peak 2). The peak corresponding to the parent ion is indicated by a red star and b- and y-ions are indicated with red or green arrows, respectively.

**Figure 6-13:** Tandem mass spectrum of $m/z$ 2416.2156 ($z = +8$) species (peak 4). The peak corresponding to the parent ion is indicated by a red star and b- and y-ions are indicated with red or green arrows, respectively.
Figure 6-14: Tandem mass spectrum of m/z 1631.4274 (z = +5) species (peak 5). The peak corresponding to the parent ion is indicated by a red star and b- and y-ions are indicated with red or green arrows, respectively.

From the fragmentation data, we conclude that the N-terminal domain of DnaT is the flexible region. The first thirty or so amino acids might be responsible for the intrinsic disorder observed in this protein. In order to revisit the crystallization of DnaT, a new DNA construct was designed coding for the sequence presented in Figure 6-15. Primers were purchased from IDT, and the new gene was obtained by PCR, using the original construct as template. The primers were designed with a CACC and an additional TEV protease cleavage site 5’ overhang (Table C.2).

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GVVAVFANNA PAFYAVTPAR LAELLALEEK LARGGANCEDVAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DDQLYQEPQA APVAVPMKGF AMYPDWQPDQ DFIRLAALWQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VALREPVTTE ELASFIAYWQ AEGKVFHHVQ WQQLARSLQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGRASNGGLP KRDVNTVSEP DSQIPPFRG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6-15: Truncated DnaT sequence selected for new molecular cloning
6.2.4 Study and Identification of DnaT Conformers in Solution

In order to study the different conformers observed from the charge state distribution of the ESI mass spectra, ion mobility mass spectrometry (IMMS) studies were performed. IMMS, allow for the separation of the analytes not only according to their $m/z$ but also according to their molecular shapes and size.

The data presented in Figure 6-16 shows the ion mobiligram of DnaT in different conditions: A, denaturing conditions; B, ammonium acetate pH 3; and C, ammonium acetate pH 7. Previous studies on different protein system gave similar data. In all three data sets one can clearly see the two different conformers. The species observed at the lower drift time and low $m/z$ range are consistent with a disordered protein while the species at higher drift time and larger $m/z$ values are closer to native like conformation.

In denaturing solvent, at $z = +11$ and $+12$ two coexisting species can be seen, one having a higher drift time, corresponding to a more denatured state, and one with a shorter drift time, accounting for a more folded state. For the data collected in ammonium acetate pH 3 or pH 7 both conformers are very distinct from each other and no co-population of the conformational state can be observed. Another interesting feature is that the species corresponding to the unfolded protein has a narrow drift time compared to the one corresponding to the more compact structure. Additionally the dimeric species observed at pH 7 has a similar mobility to the native state, consistent with the fact that the dimerization occurs through the interaction of two compact monomers.
Figure 6-16: Ion mobility plots of DnaT under different solution conditions. A: 50% ACN containing 0.1% FA, B: 10 mM AmmAc pH 3 and C: AmmAc pH 7.
6.3 Conclusion

The flexible region of DnaT was determined using CID MS. From this experiment we were able to come up with a new truncated sequence for DnaT for which the 28th first amino acids have been removed. The cloning of this new sequence is currently in process. From the same experiments, other conclusions can be drawn. Firstly, under native condition, DnaT seems to exist as two different conformers. These conformers were distinguishable by IMMS supporting the hypothesis that the two structures are drastically different. Also, the data shows the presence of dimers, but no trimers or other larger oligomeric species. The later contradicts recent results stating that DnaT was a trimer.\textsuperscript{34, 35}
Chapter 7

Summary and Future Directions

The primary goals of the first part of this thesis were the cloning and expression of the individual chains of human haptoglobin in *E. coli*, as well as their initial characterization by mass spectrometry. Results show that cloning of the L₁ and L₂ chains was successful and led to expression of soluble protein. The identity of the recombinant L₁ and L₂ chains were confirmed using a common mass spectrometry-based proteomics platform including trypsin in-gel digestion, peptide mass fingerprinting and comparison with theoretical mass lists created from their individual amino acid sequences. Chemical modification by carbamidomethylation (using iodoacetamide) was used to map the free and disulfide bonded cysteine residues in each light chain. Results were consistent that a fraction of each chain was misfolded and lacked proper disulfide pattern. This was most evident with the L₂ chain and highlights the importance of introducing an additional refolding/renaturation step to ensure proper folding of the protein chains. Due to problems that arose in the cloning step for the H-chain, expression products were inconsistent with the expected sequence. This step has been troubleshooted and improved cloning is currently underway. Molecular models for each of the human Hp chains were
created. These models will be used in our future structural studies on the individual human haptoglobin chains.

In the DnaT project, we were able to show that degradation of the protein occurs over the extended period required for crystallization. The sites of cleavage were identified using mass spectrometry. ESI MS was successfully used to definitively show that recombinant DnaT has a region with significant flexibility and intrinsic protein disorder. This region was found to be within the N-terminal 33 amino acid residues. IMMS showed that DnaT, even under near-physiological pH conditions, contained a significant fraction/subpopulation of intrinsically disordered protein. Two conformers were identified: one with the expected native folding (lower charge state) and one with a large degree of flexibility (higher charge states). Also, the highest order state of oligomerization that was detected for this protein was dimer. This is inconsistent with recent findings by Szymanski, where the functional state of the protein was shown to be a trimer. No evidence of trimer was found in our data. Current and future directions for this project include the cloning of a ΔN construct, lacking the flexible N-terminal region, which should hugely improve the odds of producing diffraction quality crystals and solution of the structure of DnaT. Our hope is to provide some insight into the mode of function of DnaT in the replication restart process including oligomerization state and the structural dynamics involved, especially within the context of binding to other proteins like PriA.
References


Appendix A

Cloning of Haptoglobin Chains: Supplementary Information

**Figure A-1:** DNA sequence map of L₁ chain. The amino acid sequence is represented in blue under the DNA sequence, the protein sequence was designed with a TEV protease cleavage site at the N-terminal. The gene was optimized by GeneArt.
Figure A-2: DNA sequence map of L2 chain. The amino acid sequence is represented in blue under the DNA sequence, the protein sequence was designed with a TEV protease cleavage site at the N-terminal. The gene was optimized by GeneArt.
Figure A-3: DNA sequence map of H chain. The amino acid sequence is represented in blue under the DNA sequence, the protein sequence was designed with a TEV protease cleavage site at the N-terminal. The gene was optimized by GeneArt.
Table A.1: PCR programs for L chains (A) and H chain (B)

(A)

<table>
<thead>
<tr>
<th>Program</th>
<th>Temperature (°C)</th>
<th>Time (s)</th>
<th>Number of cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activation</td>
<td>95</td>
<td>2 min</td>
<td></td>
</tr>
<tr>
<td>Denaturation</td>
<td>95</td>
<td>20 s</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>52</td>
<td>10 s</td>
<td></td>
</tr>
<tr>
<td>Extension</td>
<td>70</td>
<td>10 s</td>
<td></td>
</tr>
<tr>
<td>Additional extension</td>
<td>70</td>
<td>2 min</td>
<td>22</td>
</tr>
<tr>
<td>Total time</td>
<td></td>
<td></td>
<td>52 min</td>
</tr>
</tbody>
</table>

(B)

<table>
<thead>
<tr>
<th>Program</th>
<th>Temperature</th>
<th>Time (s)</th>
<th>Number of cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activation</td>
<td>95</td>
<td>2 min</td>
<td></td>
</tr>
<tr>
<td>Denaturation</td>
<td>95</td>
<td>20 s</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td>60</td>
<td>30 s</td>
<td>30</td>
</tr>
<tr>
<td>Extension</td>
<td>70</td>
<td>50 s</td>
<td></td>
</tr>
<tr>
<td>Additional extension</td>
<td>70</td>
<td>10 min</td>
<td></td>
</tr>
<tr>
<td>Total time</td>
<td></td>
<td></td>
<td>1 h 39 min</td>
</tr>
</tbody>
</table>
Table A.2: Plasmids and E. coli cells resistance

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Cloning host</th>
<th>Expression hosts</th>
</tr>
</thead>
<tbody>
<tr>
<td>HpL1</td>
<td>TOP10</td>
<td>BL21 DE3®</td>
</tr>
<tr>
<td>HpL2</td>
<td>Omnimax®</td>
<td>Rosetta DE3 Lys S®</td>
</tr>
<tr>
<td>H</td>
<td>Lys S® Tuner DE3</td>
<td></td>
</tr>
</tbody>
</table>

Antibiotic resistances

<table>
<thead>
<tr>
<th>Antibiotic resistances</th>
<th>Amp</th>
<th>Kan</th>
<th>Amp</th>
<th>Kan</th>
<th>Strep</th>
<th>Strep</th>
<th>Tet</th>
<th>None</th>
<th>Cam</th>
<th>None</th>
</tr>
</thead>
</table>

Table A.3: Primer sequences for the different proteins.

<table>
<thead>
<tr>
<th>Primers sequences</th>
<th>T_m (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward Primer L_1 &amp; L_2</td>
<td>CACCGAAATCTGTATTTTCAGGGTGTG</td>
</tr>
<tr>
<td>Reverse Primer L_1</td>
<td>TTACTGAACCGGATTTGCCCG</td>
</tr>
<tr>
<td>Reverse Primer L_2</td>
<td>TTATTGAACCGGATTTGCCCGATT</td>
</tr>
<tr>
<td>Forward Primer H</td>
<td>CACCGAGAACCTGTATTTTCAGGGTAT T</td>
</tr>
<tr>
<td>Reverse Primer H</td>
<td>TTAATTTTCTGCAATGGTTTTCGTAC</td>
</tr>
<tr>
<td>Forward Primer H*</td>
<td>CACCGAGAACCTGTATTTTCAGGGTATTTCTGGG</td>
</tr>
<tr>
<td>Reverse Primer H*</td>
<td>TTAATTTTCTGCAATGGTTTTCGTACCAATC</td>
</tr>
</tbody>
</table>

"*" indicates the primers designed for the new cloning of the H chain.
Figure A-4: SYBR GOLD™ stained 1% DNA agarose gels of L₁ chains cloning steps. From left to right are the product from the PCR, the pENTR™/D/TOPO® reaction and finally the LR Clonase® reaction. The products of interest are circled with a red box.

Figure A-5: Expression test for L₂. The SDS PAGE gel was stained with Coomasie brilliant blue™. The overexpressed protein is circled in red.
Table A.4: Protein parameters of the haptoglobin chain containing a His-tag, a linker and a TEV protease cleavage site, as reported in Figure 4-4.

<table>
<thead>
<tr>
<th>Parameters calculated</th>
<th>L₁</th>
<th>L₂</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular Weight</td>
<td>13623.8535</td>
<td>20377.3926</td>
<td>31696.6621</td>
</tr>
<tr>
<td>pI</td>
<td>6.32</td>
<td>6.37</td>
<td>6.79</td>
</tr>
<tr>
<td>Charge at pH 7</td>
<td>–3.3</td>
<td>–3.9</td>
<td>–1.6</td>
</tr>
<tr>
<td>$\varepsilon_{280}$ (M$^{-1}$ cm$^{-1}$)</td>
<td>14650 (w/o S-S)</td>
<td>26740 (w/o S-S)</td>
<td>45090 (w/o S-S)</td>
</tr>
<tr>
<td></td>
<td>14890 (w/ S-S)</td>
<td>27100 (w/ S-S)</td>
<td>45330 (w/ S-S)</td>
</tr>
<tr>
<td>$\varepsilon_{280}$ (mg/mL)$^{-1}$ cm$^{-1}$</td>
<td>1.0753 (w/o S-S)</td>
<td>1.3122 (w/o S-S)</td>
<td>1.4225 (w/o S-S)</td>
</tr>
<tr>
<td></td>
<td>1.0929 (w/ S-S)</td>
<td>1.3299 (w/ S-S)</td>
<td>1.4301 (w/ S-S)</td>
</tr>
</tbody>
</table>

Figure A-6: Tandem MALDI-TOF/TOF MS of the m/z 1271.64 from the trypsin digest of L₁ chain. The ladder series of y-ions and the identified b-ions are indicated.
Figure A-7: Tandem MALDI-TOF/TOF MS of the $m/z$ 1741.88 from the trypsin digest of L1 chain. The ladder series of b- and y-ions are indicated.

Figure A-8: Tandem MALDI-TOF/TOF MS of the $m/z$ 2695.23 from the trypsin digest of L1 chain. The ladder series of b- ions and the identified y- ions are indicated.
Figure A-9: MALDI-TOF/TOF MS of in gel tryptic digest of L₂. The sequence positions of the positively identified peptides are indicated.
**Table A.5:** List of the peptides used to confirm expression products of L2. \( m/z \): experimentally measured monoisotopic \( m/z \) value of peptide ion; \( m/z \)calculated: calculated \( m/z \) value for the monoisotopic peak; \( \Delta M \): monoisotopic mass error of measurement.

<table>
<thead>
<tr>
<th>Sequence position</th>
<th>m/z</th>
<th>( m/z )calc</th>
<th>( \Delta M )</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-18</td>
<td>2015.94</td>
<td>2015.85</td>
<td>0.09</td>
<td>AHHHHHVGT GSNDDDDK</td>
</tr>
<tr>
<td>1-24</td>
<td>2695.36</td>
<td>2695.20</td>
<td>0.16</td>
<td>AHHHHHVGT GSNDDDDKST SLYK</td>
</tr>
<tr>
<td>81-94</td>
<td>1580.87</td>
<td>1580.79</td>
<td>0.08</td>
<td>LRTEGDRVYT LNDK</td>
</tr>
<tr>
<td>81-95 or 140-154</td>
<td>1708.94</td>
<td>1708.88</td>
<td>0.06</td>
<td>LRTEGDRVYT LNDKK</td>
</tr>
<tr>
<td></td>
<td>1708.85</td>
<td></td>
<td>0.09</td>
<td>LRTEGDRVYT LNNEK</td>
</tr>
<tr>
<td>83-95 or 142-154</td>
<td>1439.75</td>
<td>1439.70</td>
<td>0.05</td>
<td>TEGDVYTLN DKK</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1439.66</td>
<td>0.09</td>
<td>TEGDVYTLN NEK</td>
</tr>
<tr>
<td>101-131</td>
<td>3316.74</td>
<td>3316.55</td>
<td>0.19</td>
<td>AVGDKLPECE ADDGCPKPPE IAHGYVEHSV R</td>
</tr>
<tr>
<td>106-131 or 140-164</td>
<td>2846.47</td>
<td>2846.30</td>
<td>0.17</td>
<td>LPECEADDGC PKPEIAHGY VEHSVVR LRTEGDRVYT LNNEKQWINKVINK AVGDK</td>
</tr>
<tr>
<td>160-176</td>
<td>1741.95</td>
<td>1741.88</td>
<td>0.07</td>
<td>AVGDKLPECE AVCGKPK</td>
</tr>
<tr>
<td>165-176</td>
<td>1271.68</td>
<td>1271.63</td>
<td>0.05</td>
<td>LPECEAVCGK PK</td>
</tr>
</tbody>
</table>

"*" indicates that these peptides involved intra disulfides bonds. Non Ambiguous sequence were determined by tandem MS experiments.
Figure A-10: Tandem MALDI-TOF/TOF MS of the $m/z$ 1439.70 from the trypsin digest of $L_2$ chain. The ladder series of $y$- ions and the identified $b$- ions are indicated.

Figure A-11: Tandem MALDI-TOF/TOF MS of the $m/z$ 1708.94 from the trypsin digest of $L_2$ chain. The ladder series of $y$- and $b$- ions are indicated.
Figure A-12: Tandem MALDI-TOF/TOF MS of the $m/z$ 2695.36 from the trypsin digest of L$_2$ chain. The ladder series of b- ions and the identified y- ions are indicated.

Figure A-13: Tandem MALDI-TOF/TOF MS of the $m/z$ 2846.47 from the trypsin digest of L$_2$ chain. The ladder series of y- ions and the identified b- ions are indicated.
Figure A-15: Solubility test for L₂. The SDS PAGE gel was stained with Coomassie blue™. The overexpressed protein found in the lysate is circled in red.

Figure A-16: Ni-NTA purification chromatogram of L₂ chain. The green trace corresponds to A₂₈₀ and the purple trace to A₂₆₀. The fractions that were pulled together are highlighted in red on the figure.
**Table A.6:** Protein parameters of the haptoglobin chains after HIS tag cleavage by TEV protease as shown in.

<table>
<thead>
<tr>
<th>Parameters Calculated</th>
<th>L₁</th>
<th>L₂</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular Weight</td>
<td>9249.2354</td>
<td>16002.7734</td>
<td>27322.0449</td>
</tr>
<tr>
<td>pH</td>
<td>5.41</td>
<td>5.88</td>
<td>6.79</td>
</tr>
<tr>
<td>Charge at pH 7</td>
<td>−2.7</td>
<td>−3.3</td>
<td>−1.0</td>
</tr>
<tr>
<td>( \varepsilon_{280} ) (M⁻¹ cm⁻¹)</td>
<td>12090 (w/o S-S)</td>
<td>24180 (w/o S-S)</td>
<td>42530 (w/o S-S)</td>
</tr>
<tr>
<td></td>
<td>12330 (w/ S-S)</td>
<td>24540 (w/ S-S)</td>
<td>42770 (w/ S-S)</td>
</tr>
<tr>
<td>( \varepsilon_{280} ) ((mg/mL)⁻¹ cm⁻¹)</td>
<td>1.3071 (w/o S-S)</td>
<td>1.5110 (w/o S-S)</td>
<td>1.5079 (w/o S-S)</td>
</tr>
<tr>
<td></td>
<td>1.3331 (w/ S-S)</td>
<td>1.5335 (w/ S-S)</td>
<td>1.5153 (w/ S-S)</td>
</tr>
</tbody>
</table>

**Figure A-17:** Tandem MALDI-TOF/TOF MS of the \( m/z \) 2695.19 from the trypsin digest of H chain. The ladder series of the b-ions identified are indicated.
Appendix B

Initial Characterization of Recombinant Haptoglobin Chains Supplementary Data

Table B.1: Predicted peptides crosslinking or containing carbamidomethylated cysteine residue for L1 chain. The modified cysteine residues are indicated by \( C_x^* \). \( M_{calc} \): calculated mass of the modified or crosslinked peptides.

<table>
<thead>
<tr>
<th>Chain</th>
<th>Cross linked peptides</th>
<th>( M_{calc} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>33-36 (C_{35}) &amp; 66-77 (C_{69})</td>
<td>1812.87</td>
</tr>
<tr>
<td></td>
<td>33-36 &amp; 61-77</td>
<td>2283.12</td>
</tr>
<tr>
<td></td>
<td>1-32 (C_{16}) &amp; 1-32</td>
<td>6697.08</td>
</tr>
<tr>
<td></td>
<td>66-77 (C_{73}^*)</td>
<td>1330.70</td>
</tr>
<tr>
<td></td>
<td>60-77 (C_{73}^*)</td>
<td>1800.95</td>
</tr>
<tr>
<td></td>
<td>60-77 (C_{69}^* &amp; C_{73}^*)</td>
<td>1858.02</td>
</tr>
<tr>
<td></td>
<td>65-77 (C_{69}^* &amp; C_{73}^*)</td>
<td>1387.77</td>
</tr>
<tr>
<td></td>
<td>1-32 (C_{16}^*)</td>
<td>3406.61</td>
</tr>
<tr>
<td></td>
<td>33-36 (C_{35}) &amp; 66-77 (C_{69} &amp; C_{73}^*)</td>
<td>1869.94</td>
</tr>
<tr>
<td></td>
<td>33-36 (C_{35}) &amp; 61-77 (C_{69} &amp; C_{73}^*)</td>
<td>2340.19</td>
</tr>
</tbody>
</table>
Table B.2: Predicted peptides crosslinking or containing carbamidomethylated cysteine residue for L₂ chain. The modified cysteine residues are indicated by Cₓ *. M_<sub>calc</sub>: calculated mass of the modified or crosslinked peptides.

<table>
<thead>
<tr>
<th>Chain</th>
<th>Cross linked peptides position/carbamidomethylated peptides</th>
<th>M_&lt;sub&gt;calc&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>L₂</td>
<td>33-36 (Cₓ₃₅*) &amp; 66-91 (Cₓ₆₉*)</td>
<td>3387.54</td>
</tr>
<tr>
<td></td>
<td>33-36 (Cₓ₃₅*) &amp; 61-91 (Cₓ₆₉*)</td>
<td>3857.79</td>
</tr>
<tr>
<td></td>
<td>92-95 (Cₓ₉₄*) &amp; 125-136 (Cₓ₁₂₇*)</td>
<td>1812.87</td>
</tr>
<tr>
<td></td>
<td>92-95 (Cₓ₉₄*) &amp; 120-136 (Cₓ₁₂₇*)</td>
<td>2283.12</td>
</tr>
<tr>
<td></td>
<td>1-32 (Cₓ₁₃*) &amp; 1-32 (Cₓ₁₃*)</td>
<td>6697.08</td>
</tr>
<tr>
<td></td>
<td>66-91 (Cₓ₇₅*)</td>
<td>2905.37</td>
</tr>
<tr>
<td></td>
<td>61-91 (Cₓ₇₅*)</td>
<td>3375.62</td>
</tr>
<tr>
<td></td>
<td>120-136 (Cₓ₁₂₇*)</td>
<td>1800.95</td>
</tr>
<tr>
<td></td>
<td>125-136 (Cₓ₁₃₂*)</td>
<td>1330.70</td>
</tr>
<tr>
<td></td>
<td>61-91 (Cₓ₆₉* &amp; Cₓ₇₅*)</td>
<td>3432.69</td>
</tr>
<tr>
<td></td>
<td>66-91 (Cₓ₆₉* &amp; Cₓ₇₅*)</td>
<td>2962.44</td>
</tr>
<tr>
<td></td>
<td>120-136 (Cₓ₁₂₇* &amp; Cₓ₁₃₂*)</td>
<td>1858.02</td>
</tr>
<tr>
<td></td>
<td>125-136 (Cₓ₁₂₇* &amp; Cₓ₁₃₂*)</td>
<td>1387.77</td>
</tr>
<tr>
<td></td>
<td>1-32 (Cₓ₁₆*)</td>
<td>3406.61</td>
</tr>
<tr>
<td></td>
<td>33-36 (Cₓ₃₅*) &amp; 66-91(Cₓ₆₉ &amp; Cₓ₇₅*)</td>
<td>3444.61</td>
</tr>
<tr>
<td></td>
<td>33-36 (Cₓ₃₅*) &amp; 61-91 (Cₓ₆₉ &amp; Cₓ₇₅*)</td>
<td>3914.86</td>
</tr>
<tr>
<td></td>
<td>92-95 (Cₓ₉₄*) &amp; 125-136 (Cₓ₁₂₇ &amp; Cₓ₁₃₂*)</td>
<td>1869.94</td>
</tr>
<tr>
<td></td>
<td>92-95 (Cₓ₹₄*) &amp; 120-136 (Cₓ₁₂₇ &amp; Cₓ₁₃₂*)</td>
<td>2340.19</td>
</tr>
</tbody>
</table>
**Figure B–1:** Tandem MS of the m/z 288.2134 peak \((z = +1)\) from L₁ tryptic digest. The ladder of the y-ions is indicated on the spectrum.

**Figure B–2:** Tandem MS of the m/z 688.3865 peak \((z = +1)\) L₁ tryptic digest. The ladder of the y-ions is indicated on the spectrum as well as the identified b-ions.
**Figure B–3:** Tandem MS of the $m/z$ 739.4150 peak ($z = +1$) L$_1$ tryptic digest. The ladder of the y-ions is indicated on the spectrum as well as the identified b-ions.

**Figure B–4:** Tandem MS of the $m/z$ 720.3500 peak ($z = +2$) L$_1$ tryptic digest. The ladder of the y-ions is indicated on the spectrum as well as the identified b-ions.
**Figure B–5:** Tandem MS of the $m/z$ 854.9477 peak ($z = +2$) L1 tryptic digest. The ladder of the y- and b- ions is indicated on the spectrum.

**Figure B–6:** Tandem MS of the $m/z$ 871.4694 peak ($z = +2$) L1 tryptic digest. The ladder of the y- and b- ions is indicated on the spectrum.
**Figure B–7:** Tandem MS of the m/z 688.3672 peak (z = +1) L2 tryptic digest. The ladder of the y- ions is indicated on the spectrum as well as the identified b- ions.

**Figure B–8:** Tandem MS of the m/z 720.3278 peak (z = +2) L2 tryptic digest. The ladder of the y- ions is indicated on the spectrum as well as the identified b- ions.
**Figure B–9**: Tandem MS of the $m/z$ 739.3549 peak ($z = +1$) L$_2$ tryptic digest. The ladder of the $y$-ions is indicated on the spectrum as well as the identified $b$-ions.

**Figure B–10**: Tandem MS of the $m/z$ 790.8879 peak ($z = +2$) L$_2$ tryptic digest. The ladder of the $y$-ions is indicated on the spectrum as well as the identified $b$-ions.
Figure B–11: Tandem MS of the $m/z$ 1311.6288 peak ($z = +1$) L$_2$ tryptic digest. The ladder of the y- ions is indicated on the spectrum as well as the identified b- ions.

Figure B–12: Tandem MS of the $m/z$ 854.9153 peak ($z = +2$) L$_2$ tryptic digest. The ladder of the b- ions is indicated on the spectrum as well as the identified y- ions.
**Figure B–13:** Tandem MS of the \( m/z \) 871.4476 peak (\( z = +2 \)) L\(_2\) tryptic digest. The ladder of the b- ions is indicated on the spectrum as well as the identified y- ions.

**Figure B–14:** Tandem MS of the \( m/z \) 1106.1760 peak (\( z = +3 \)) L\(_2\) tryptic digest. The ladder of the y- ions is indicated on the spectrum as well as the identified b- ions.
Figure B–15: Tandem MS of the m/z 1424.1304 peak (z = +2) L2 tryptic digest. The ladder of the y- ions is indicated on the spectrum as well as the identified b- ions.

Figure B–16: Tandem MS of the m/z 1136.5143 peak (z = +3) L2 tryptic digest. The ladder of the y- ions is indicated on the spectrum as well as the identified b- ions.
**Figure B–17:** Tandem MS of the m/z 1297.2701 peak (z = +3) L2 tryptic digest. The identified y- and b- ions are indicated. The sequence of the cross-linked peptides including its fragmentation pattern is indicated in the figure.

**Figure B–18:** Tandem MS of the m/z 1675.7682 peak (z = +4) L2 tryptic digest. The identified y-ions are indicated. The sequence of the crosslinked peptides including its fragmentation pattern is indicated in the figure.
Figure B–19: Sequence alignment between human Hp1 (hHp1) and porcine Hp (pHp). The residues similar to both sequences are indicated by “*” and the residues with similar properties by “:”. 
Figure B–20: Sequence alignment between human Hp2 (hHp2) and porcine Hp (pHp). The residues similar to both sequences are indicated by “*” and the residues with similar properties by “.”.
Figure B–21: Sequence alignment between human Hp2 (hHp2) and C1R protease. The residues similar to both sequences are indicated by “*”.
Appendix C

DnaT Supplementary Data

Table C.1: Protein parameters of DnaT

<table>
<thead>
<tr>
<th>Parameters Calculated</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular Weight</td>
<td>19323.8750</td>
</tr>
<tr>
<td>pI</td>
<td>5.31</td>
</tr>
<tr>
<td>Charge at pH 7</td>
<td>-5.1</td>
</tr>
<tr>
<td>$\varepsilon_{280}$ (M$^{-1}$ cm$^{-1}$)</td>
<td>27880</td>
</tr>
<tr>
<td>$\varepsilon_{280}$ ((mg/mL)$^{-1}$ cm$^{-1}$)</td>
<td>1.4428</td>
</tr>
</tbody>
</table>

Figure C–1: Mass deconvolution of the mass spectrum of the intact DnaT in AmmAc pH 7 presented in Figure 6-7:
Figure C–2: Mass deconvolution of the mass spectrum of the intact DnaT in 50% ACN containing 0.1% FA presented in Figure 6-10.

Figure C–2: Tandem mass spectrum of the species at m/z 2429.7083 (z = +8). The peak corresponding to the parent ion is indicated by a red star and b and y ions are indicated by respectively a red or a green arrow.
**Figure C–3:** Representation of the fragmentation pattern obtained from the fragmentation in denaturing solution. A: fragmentation pattern of peak 1 \( m/z \) 1215.2977; B: fragmentation pattern of peak 2 \( m/z \) 1208.6812; C: fragmentation pattern of peak 3 \( m/z \) 2429.7083; D: fragmentation pattern of peak 4 \( m/z \) 2416.2156. The b-ion fragments are indicated by red arrow and the y-ions fragments with green arrows.

**Table B.2:** Sequence of the designed primers for the cloning of the ΔN-DnaT protein.

<table>
<thead>
<tr>
<th>Primsers sequences</th>
<th>( T_m ) (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward Primer DnaT</td>
<td>CACCGAGAACCTGTATTTTCAGGG CACCCTCTGAATCCTGGTGGAGG 68.8</td>
</tr>
<tr>
<td>Reverse Primer DnaT</td>
<td>TTACCTCTGAATCCTGGGAAATTGGCT 62.7</td>
</tr>
</tbody>
</table>