



# LUND UNIVERSITY

## Exploring small heat shock protein chaperones by crosslinking mass spectrometry

Lambert, Wietske

2012

[Link to publication](#)

*Citation for published version (APA):*

Lambert, W. (2012). *Exploring small heat shock protein chaperones by crosslinking mass spectrometry*. [Doctoral Thesis (compilation), Biochemistry and Structural Biology]. Department of Chemistry, Lund University.

*Total number of authors:*

1

### General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117  
221 00 Lund  
+46 46-222 00 00

# Exploring small heat shock protein chaperones by crosslinking mass spectrometry

Wietske Lambert



**LUND**  
UNIVERSITY

ISBN 978-91-7422-299-9

Copyright © Wietske Lambert

Department of Biochemistry and Structural Biology

Lund University, P.O. Box 124, SE-221 00 Lund, Sweden

Printed in Sweden by Media-Tryck, Lund 2012

# Abstract

Together with other molecular chaperones, small heat shock proteins are key components of the protein quality control system, which is comprised of several hundred proteins and acts to maintain proteome homeostasis in the cell. Small heat shock proteins bind unfolding proteins at an early stage, to prevent these from further unfolding and aggregating. Partially unfolded proteins are being held in a refolding competent state, to be refolded by other chaperones or degraded by the degradation machinery. In the stress response, small heat shock proteins are among the most highly upregulated, preparing the cell to absorb large quantities of partially unfolded proteins. In this way, they form the first line of defence against the threat of protein aggregation under stress conditions. The polydispersity and dynamics of the large small heat shock protein oligomers have complicated their structural and functional characterization. In particular, the molecular mechanism of substrate protein protection remains poorly understood.

The work described in this thesis aims to characterize the molecular interactions between the plant small heat shock protein Hsp21 and model substrate proteins by crosslinking mass spectrometry. The model substrate proteins citrate synthase and malate dehydrogenase, both especially vulnerable to temperature-induced aggregation, were protected from aggregation by Hsp21 and therefore used to investigate the Hsp21-substrate interactions that confer protection. To be able to study the transient Hsp21-substrate interaction by crosslinking mass spectrometry, a workflow was developed based on isotope-labelled lysine-specific crosslinking, nano-LC MALDI-TOF/TOF mass spectrometry, and data analysis with the specialized software FINDX. During the development of this workflow, interactions within Hsp21 itself were characterized as a way to evaluate the method and to learn more about the conformation of Hsp21 in absence of substrate. The interpretation of the identified Hsp21-Hsp21 crosslinks required structural information on the Hsp21 oligomer, which was obtained by single particle negative stain electron microscopy. The combination of these data with native mass spectrometry and homology modelling, led to a structure model of the Hsp21 dodecamer. The in-depth analysis of Hsp21-Hsp21 crosslinks provided a framework for further application of the crosslinking mass spectrometry workflow to the Hsp21-substrate interactions. Finally, Hsp21-substrate crosslinks were identified that support the view that unfolding substrate proteins interact with the intrinsically disordered N-terminal region of the small heat shock protein Hsp21.

# List of papers

A doctoral thesis at a university in Sweden is produced as a monograph or as a collection of papers. The latter is the commonly used format of a doctoral thesis within the fields of life science. In this case, an introductory part together with a summary of the collection of papers precedes the actual papers.

This thesis is based on the following papers, which are referred to by their roman numerals in the text:

- I. Ahrman, E., Lambert, W., Aquilina, J.A., Robinson, C.V., and Emanuelsson, C. (2007). Chemical cross-linking of the chloroplast localized small heat-shock protein, Hsp21, and the model substrate citrate synthase. *Protein Sci* 16, 1464-1478.
- II. Lambert, W., Koeck, P.J.B., Ahrman, E., Purhonen, P., Cheng, K., Elmlund, D., Hebert, H., and Emanuelsson, C. (2011). Subunit arrangement in the dodecameric chloroplast small heat shock protein Hsp21. *Protein Sci* 20, 291-301.
- III. Lambert, W., Soderberg, C.A.G., Rutsdottir, G., Boelens, W.C., and Emanuelsson, C. (2011). Thiol-exchange in DTSSP crosslinked peptides is proportional to cysteine content and precisely controlled in crosslink detection by two-step LC-MALDI MSMS. *Protein Sci* 20, 1682-1691.
- IV. Söderberg, C.A.G., Lambert, W., Kjellström, S., Wiegandt, A., Peterson-Wulff, R., Månsson, C., Rutsdottir, G., and Emanuelsson, C. (2012). Identification of crosslinks within and between proteins by MALDI-MS and the software FINDX to reduce the amount of MSMS-data to acquire for validation. *Submitted*.
- V. Lambert, W., Rutsdottir, G., Bernfur, K., Kjellström, S., and Emanuelsson, C. (2012). Probing the transient interaction between the small heat shock protein Hsp21 and a model substrate protein by crosslinking mass spectrometry. *Manuscript*.

Papers I, II, and III were reproduced with permission from John Wiley and Sons.

Contributions by the authors of the papers in this thesis:

- Paper I: EÅ and CE designed the research, EÅ performed experiments, except nanoESI-MS, which was performed by JAA and CVR, WL constructed the homology model, and EÅ and CE analyzed the data and wrote the paper.
- Paper II: EÅ, HH and CE initiated the project, WL, EÅ, PP, KC, and DE performed experiments, PJBK analyzed the data, and PJBK, WL, HH, and CE interpreted the data and wrote the paper.
- Paper III: WL and CE designed the research, WL and GR performed experiments, CAGS wrote the data analysis program, WL analyzed the data, and WL and CE wrote the paper.
- Paper IV: CAGS, WL and CE initiated the project, CAGS, WL, SK, AW, RPW, CM, GR, and CE performed experiments and analyzed data, and CS, WL and CE wrote the paper.
- Paper V: WL and CE designed the research, WL, GR, KB, SK, and CE performed the experiments and analyzed the data, and WL and CE wrote the paper.

Other publications by the author of this thesis, not included in the thesis:

- Drew, D., Slotboom, D.J., Friso, G., Reda, T., Genevaux, P., Rapp, M., Meindl-Beinker, N.M., Lambert, W., Lerch, M., Daley, D.O., van Wijk, K.J., Hirst, J., Kunji, E., and de Gier, J.W. (2005). A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. *Protein Sci* 14, 2011-2017.
- Rutten, L., Geurtsen, J., Lambert, W., Smolenaers, J.J.M., Bonvin, A.M., de Haan, A., van der Ley, P., Egmond, M.R., Gros, P., and Tommassen, J. (2006). Crystal structure and catalytic mechanism of the LPS 3-O-deacylase PagL from *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* 103, 7071-7076.
- Luoto, S., Lambert, W., Blomqvist, A., and Emanuelsson, C. (2008). The identification of allergen proteins in sugar beet (*Beta vulgaris*) pollen causing occupational allergy in greenhouses. *Clin Mol Allergy* 6, 7-16.

# Abbreviations

AP	affinity purification
ATP	adenosine triphosphate
BS <sup>3</sup>	bis(sulfosuccinimidylsuberate)
CID	collision induced dissociation
CS	citrate synthase
CXMS	crosslinking mass spectrometry
DTSSP	3,3'-dithiobis(sulfosuccinimidylpropionate)
DTT	dithiothreitol
EM	electron microscopy
EPR	electron paramagnetic resonance
ESI	electrospray ionization
FRET	fluorescence resonance energy transfer
FTICR	Fourier-transform ion cyclotron resonance
IDR	intrinsically disordered region
IM	ion mobility
LC	liquid chromatography
LIT	linear ion trap
LTQ	linear trap quadrupole
MALDI	matrix assisted laser desorption ionization
MALS	multiangle light scattering
MDH	malate dehydrogenase
MS	mass spectrometry
MS3D	mass spectrometry based structural biology approaches
NMR	nucleic magnetic resonance

Q	quadrupole
QIT	quadrupole ion trap
SANS	small-angle neutron scattering
SAXS	small-angle X-ray scattering
SCX	strong cation exchange chromatography
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size-exclusion chromatography
sHsp	small heat shock protein
sHsps	small heat shock proteins
TEM	transmission electron microscopy
TIS	timed ion selector
TOF	time of flight



# Contents

Abstract .....	3
List of papers .....	4
Abbreviations .....	6
Contents .....	8
1 Introduction .....	10
2 Small heat shock proteins .....	12
2.1 Dimeric building blocks .....	13
2.2 Oligomeric structure .....	16
2.3 Dynamic subunit exchange .....	19
2.4 Chaperone activity and substrate interactions .....	20
2.5 Hsp21 in <i>Arabidopsis thaliana</i> chloroplasts .....	23
3 Crosslinking mass spectrometry .....	25
3.1 Complexity of the crosslinked peptide mixture .....	26
3.2 Chemical crosslinking reagents .....	27
3.3 Mass spectrometry instrumentation .....	29
3.4 Data analysis – identifying crosslinks .....	32
3.5 Interpretation of detected crosslinks .....	33
3.6 Complementary structural biology techniques .....	34
4 Crosslinking studies on Hsp21 (this work) .....	37
4.1 Oligomeric structure of Hsp21 .....	38
4.2 The crosslinked Hsp21 dodecamer .....	39
4.3 Crosslinks within Hsp21 .....	40
4.4 Intra- and inter-monomeric Hsp21 crosslinks .....	42
4.5 Crosslinks within $\alpha$ B-crystallin .....	44

4.6	Hsp21 chaperone function.....	45
4.7	Crosslinks between Hsp21 and model substrate proteins.....	45
4.8	Crosslinking mass spectrometry to probe protein-protein interactions	47
5	Concluding remarks.....	49
6	Future perspectives .....	51
7	About this thesis .....	53
7.1	Popular scientific summary in English .....	53
7.2	Populärvetenskaplig sammanfattning på svenska .....	54
7.3	Populairwetenschappelijke samenvatting in het Nederlands .....	55
	Acknowledgements .....	56
	References .....	58

# 1 Introduction

In the crowded environment of the cell, protein aggregation is a major threat to cell survival. Proteins have evolved to be only marginally stable, so partially unfolded proteins being prone to aggregation form a continuous risk (Chiti and Dobson, 2009). The nascent chains of proteins being translated by the ribosome have yet to fold into their native state (Selmer and Liljas, 2008), and even proteins that are already folded are continuously in equilibrium with their partially unfolded forms. Stress conditions shift this equilibrium in favour of the unfolded proteins. To prevent damage to the cell by toxic unfolded protein aggregates, evolution has provided cells with an extensive protein quality control system, in which main players are the molecular chaperones (Hartl et al., 2011).

Molecular chaperones are proteins that assist other proteins in arriving at and maintaining their native (folded) states, thereby preventing unfolded proteins from aggregating. Some chaperones assist newly synthesized proteins in folding, whereas others are specialized in refolding partially unfolded or misfolded proteins, for which energy in the form of ATP (adenosine triphosphate) is normally required (Mayer, 2010). Still other chaperones play an important role in the clearance of unfolded proteins, by directing them to the degradation machinery, or allowing controlled aggregation (Dougan et al., 2002; Tyedmers et al., 2010), both of which are also important components of the protein quality control system. Many chaperone proteins are called heat shock proteins, because their expression was found to be transcriptionally inducible by heat stress. However, other stresses can also induce heat shock protein expression, while at the same time many heat shock proteins are housekeeping proteins that are constitutively expressed.

The small heat shock proteins (sHsps) are a class of molecular chaperones that bind unfolding proteins, but cannot actively refold them and do not require ATP. Instead, sHsps hold their client proteins in a refolding competent state, preventing them from aggregating (Basha et al., 2011; Nakamoto and Vigh, 2007). ‘Small’ refers to their monomeric size of between 12 and 42 kDa, but many sHsps can form large oligomers with up to 40-50 subunits (McHaourab et al., 2009). Compared to some intensively studied chaperones like those belonging to the Hsp60 and Hsp70 families (Mayer, 2010), the mechanism of sHsps is still poorly understood. This is mainly attributed to their enormous diversity and heterogeneity, in terms of oligomeric structure and substrate binding, making them technically challenging to study (Eyles and Gierasch, 2010).

In the work described in this thesis, crosslinking mass spectrometry has been used to study small heat shock proteins. This technique combines the chemical crosslinking of proteins and protein complexes with the mass spectrometric detection of crosslinked peptides that result from enzymatic digestion. By analyzing which amino acid residues of the peptides have crosslinked, interactions within and between proteins can be characterized (Leitner et al., 2010; Singh et al., 2010; Sinz, 2006). Even though crosslinking mass spectrometry is becoming an established technique in the field of structural biology, especially in combination with other techniques (Rappsilber, 2011; Stengel et al., 2012), there are still challenges that need to be overcome until the method can really live up to its promises.

The main objective of the work described in this thesis is to use crosslinking mass spectrometry to characterize interactions within small heat shock proteins and with their substrate proteins. The focus is on the chloroplastic sHsp Hsp21 from *Arabidopsis thaliana*, and in some cases, the human sHsp  $\alpha$ B-crystallin has been investigated in parallel. Throughout the thesis, a crosslinking mass spectrometry workflow is developed to use on an offline nano-LC-MALDI-TOF/TOF mass spectrometry platform. In paper I, the sHsp Hsp21 and the model substrate citrate synthase (CS) are crosslinked with the crosslinker DTSSP, and analyzed by MALDI-TOF. In paper II, information from crosslinks detected within Hsp21 is combined with negative stain electron microscopy (EM) and homology modelling to gain insight into the subunit organization of Hsp21. The disulfide bridge containing crosslinker DTSSP is problematic to use in the presence of too many free cysteine residues, which is addressed in paper III. Here, and in paper IV, the workflow with offline nano-LC, MALDI-TOF/TOF mass spectrometry, and data analysis with the in-house developed program FINDX, is developed and optimized. Important improvements include the use of isotope labelled crosslinkers DTSSP and BS<sup>3</sup>, the addition of nano-LC separation of peptides before mass spectrometry, and fine-tuning of the data analysis. The crosslinker BS<sup>3</sup> does not have a disulfide bridge, which explains its preferred use over DTSSP for proteins containing free cysteine residues. Finally, in paper V, the optimized crosslinking mass spectrometry workflow is used to investigate the transient interaction between Hsp21 and the model substrate protein malate dehydrogenase (MDH).

## 2 Small heat shock proteins

Small heat shock proteins form an evolutionary ancient superfamily (de Jong et al., 1998), as reflected by their presence in almost all organisms (Haslbeck et al., 2005). They can be strongly induced by heat stress, hence their name, but it has become clear that many sHsps are being constitutively expressed in all kinds of cell types (Narberhaus, 2002). The most well-known sHsps are the  $\alpha$ -crystallins  $\alpha$ A- and  $\alpha$ B-crystallin, which are major structural proteins in the vertebrate eye lens.  $\alpha$ A- and  $\alpha$ B-crystallin are also called HspB4 and HspB5, according to the more systematic nomenclature guidelines for human heat shock proteins (Kampinga et al., 2009). Originally, the  $\alpha$ -crystallins were exclusively related to lens function, until they were also found to be expressed in other tissues. Since  $\alpha$ -crystallins and other sHsps were first shown to be able to act as molecular chaperones (Horwitz, 1992; Jakob et al., 1993), overwhelming evidence has accumulated that sHsps are ubiquitous key players in the protein quality control system of cells (Eyles and Gierasch, 2010). Today, it is speculated that the function of sHsps may not even be confined to chaperone activity alone (Basha et al., 2011).

Several mutations in small heat shock proteins have been linked to inherited diseases, including cataract, cardiac and skeletal myopathies, and neuropathies (Sun and MacRae, 2005). In Alzheimer's disease, Parkinson's disease and multiple sclerosis, sHsps have been found to be associated with protein aggregates in neurons (Basha et al., 2011). Some cancer cells display an altered sHsp expression pattern, which is typical for stressed cells. In heat shocked cells of *L. interrogans* (a human pathogen), sHsps were among the most upregulated proteins (>10x) (Beck et al., 2009). Undoubtedly, small heat shock proteins are crucial for cell survival, especially under stress conditions, but exactly how they fulfil their protective role remains elusive. The heterogeneity and polydispersity of many sHsps have been problematic for traditional structural biology approaches, but in recent years, hybrid techniques have proven very helpful to elucidate small heat shock protein (sHsp) structures. Below, the structure of sHsps will be described in more detail.

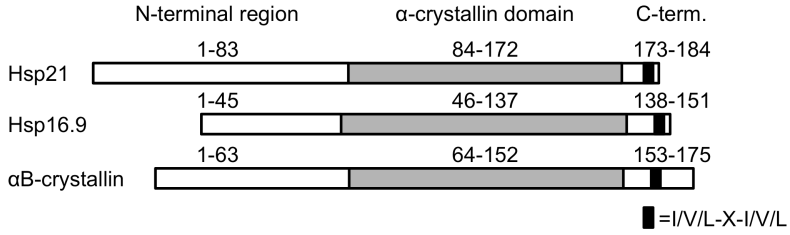
## 2.1 Dimeric building blocks

The  $\alpha$ -crystallin domain, which is named after the human, best known sHsps  $\alpha$ A- and  $\alpha$ B-crystallin, characterizes all members of the sHsp superfamily (Poulain et al., 2010). This central region is around 90 amino acid residues and forms an immunoglobulin domain that consists of 7  $\beta$ -strands forming a  $\beta$ -sandwich. The N-terminal region is very diverse among sHsps, both in sequence and in length. C-terminal to the central  $\alpha$ -crystallin domain is the C-terminal extension, which is also very diverse, except for a conserved I/V/L-X-I/V/L motif that is essential for oligomer formation. A sequence alignment including sHsps for which the structure has been determined, illustrates the sequence diversity but at the same time the conservation of the main structural elements in the sHsp superfamily (figure 1).

High-resolution structures of sHsps almost all reveal dimeric building blocks consisting of two  $\alpha$ -crystallin domains, coming from two different monomers (the sHsp Tsp36 from tapeworm being an exception, which contains two divergent sHsp repeats within one monomer (Stamler et al., 2005)). The dimeric building blocks form larger oligomeric species in most cases. While the  $\alpha$ -crystallin domain always forms the same 7-stranded  $\beta$ -sandwich, two different architectures of the dimeric building block can be distinguished among sHsp structures, as determined by crystallography, solid-state nucleic magnetic resonance (NMR), small-angle X-ray scattering (SAXS), or a combination.

In the crystal structures of Hsp16.5 from *M. jannaschii* (Kim et al., 1998), Hsp16.9 from *T. aestivum* (Van Montfort et al., 2001b), HspA from *X. citri* pv. *citri* (Hilario et al., 2011), and Hsp14.0 from *S. tokodaii* (Takeda et al., 2011), two  $\alpha$ -crystallin domains bind by  $\beta$ -strand exchange (figure 2A,C). The  $\beta$ -strands of  $\alpha$ -crystallin domain are conventionally named  $\beta$ 2- $\beta$ 9, after the first sHsp crystal structure solved (Kim et al., 1998). The strand named  $\beta$ 6 is located on a loop extending away from the  $\beta$ -sandwich of one monomer, and participates in  $\beta$ -sheet formation by binding strand  $\beta$ 7 of the other monomer. Strand  $\beta$ 6 of the other monomer interacts again with  $\beta$ 7 of the first monomer.

A



B

```

AtHsp21  --MQDQRENSIDV-VQQGQKGNQ-GSSVEKRPQQLTMDVSPFGLLDPLSPMRTMRQML 56
TaHsp21  SAAQENRDNNSVDVQVSQAQNAQNAGNQGNVQRRPR-RAGFDISPFGLVDPMSPMRTMRQML 59
AtHsp18.1 -----MSLIPSIFGGRRSNVFDPF-----SQDLW 24
TaHsp16.9 -----MSIV-----RRTNVFDPF-----ADLWA 18
BtHspB4   -----MDIAIQHPWFKR---TLGPFY-PSRLFDQFFGEGFL 32
HsHspB5   -----MDIAIHHPWIRR---PFFPFHSPSRLFDQFFGEGHLL 33
RnHspB6   -----MEIRVPVQPSWLRASAPLPGFSFPGRLFDQRFEGGLL 38
MjHsp16.5 -----MFGRDPF-----DSLFL 11

```

```

AtHsp21  DTMDRMFEDTMPVSGRNRGGSGVSEIR---APWDIKEEHEIKMRFDMPLGSKEDVKIS 112
TaHsp21  DTMDRLFDDAVGFPTTR-RSPAARARRR---MPWDIMEDEKEVKMRFDMPLGSRREEVRVM 114
AtHsp18.1 DPFEFFFTSSALANASTARDVAAFN---ARVDWKETPEAHVFKADLPGLKKEEVKWE 80
TaHsp16.9 DPFDTFRSIVPAISGGGSE--TAAFAN---ARMDWKETPEAHVFKADLPGVKKEEVKVE 72
BtHspB4   EYDLLPFLSSTISPYYR--QSLFRT--VLDSGISEVRSDRDKFVFLDVKHFSPEDLTVK 88
HsHspB5   ESDLFP-TSTSLSPFYLRPPSFLRAPSWFDTLGSEMRLEKDRFSVNLDVKHFSPEELKVK 92
RnHspB6   EAELASLCPAAIAPYYLRAPSVALP-----TAQVPTDPGYFSVLLDVKHFSPEELISVK 91
MjHsp16.5 ERMFKEFFATPMTGTTMIQSSTGIQISGKGFMPISIEGDQHIKVIAWLPVGNKEDIILN 71

```

```

AtHsp21  VED-NVLVIKGEQKKEDS-----DDSWS--GRSVSSYGTRLQLPDCNCEKDKIKAEELKN 162
TaHsp21  VED-DALVIRGEHKKKEAGEGQEGGDGWWK--ERSVSSYDMRLALPDECDKSOVRAELKN 171
AtHsp18.1 VEDKNVLQISGERSKENEE-----KNDKWHRVERASGKFMRRFRLPENAKMEEVKATMEN 135
TaHsp16.9 VEDGNVLVVGERTKEKED-----KNDKWHRVERSSGKFMRRFRLLEDAKVEEVKAGLEN 127
BtHspB4   VQE-DFVEITHGKHNERQDD-----HGYISREFHRRYRLPSNVDAQSALSCSLSA 135
HsHspB5   VLG-DVIEVHGKHEERQDE-----HGFISREFHRRYRIPADVPLTITSSLSL 139
RnHspB6   VVG-DHVEVHARHEERPDE-----HGFIAREFHRRYRLPPGVDDPAAVTSALSP 138
MjHsp16.5 AVG-DTLETRAKRSPLMITE-----SERIYSEIPEEEIYRTIKLPATVKEENASAKPEN 126

```

```

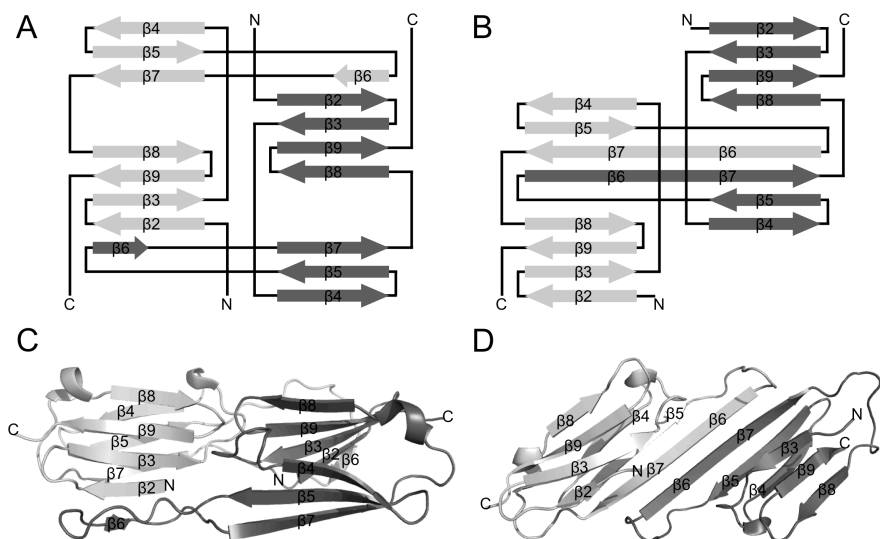
AtHsp21  -GVLFITIPKT-----KVER-KVIDVQIQ----- 184
TaHsp21  -GVLLVSPKR-----ETER-KVIDVQVQ----- 193
AtHsp18.1 -GVLTVVVPKA-----PEKKPQVKSIDISGAN----- 161
TaHsp16.9 -GVLTVTVPKA-----EVKKPEVKAIQISG----- 151
BtHspB4   DGMLTFSGPKIPSGVDAGHSERAI PVSREEKP--SSAPSS 173
HsHspB5   DGVLTVNGPRK----QVSGPERTIPITREEKPAVTAAPKK 175
RnHspB6   EGVLSIQATPA-----SAQASLPSPPAK----- 162
MjHsp16.5 -GVLSVILPKA-----ESSIKKGINIE----- 147

```

**Figure 1.** Primary structure of sHsps. A: Schematic representation of the N-terminal region (white), the  $\alpha$ -crystallin domain (gray), and the C-terminal extension (white) containing the I/V/L-X-I/V/L motif of Hsp21 from *A. thaliana*, Hsp16.9 from *T. aestivum*, and human  $\alpha$ B-crystallin. B. Sequence alignment of AtHsp21 (Hsp21 from *A. thaliana*, UniProtKB ID P31170), TaHsp21 (Hsp21 from *T. aestivum*, UniProtKB ID Q00445), AtHsp18.1 (Hsp18.1 from *A. thaliana*, UniProtKB ID P19037), TaHsp16.9 (Hsp16.9 from *T. aestivum*, UniProtKB ID Q41560), BtHspB4 (bovine  $\alpha$ A-crystallin, UniProtKB ID P02470), HsHspB5 (human  $\alpha$ B-crystallin (HspB5), UniProtKB ID P02511), RnHspB6 (HspB6 from *R. norvegicus*, UniProtKB ID P97541), and MjHsp16.5 (Hsp16.5 from *M. jannaschii*, UniProtKB ID Q57733). The sequences were aligned using ClustalW2 (Larkin et al., 2007). Identical residues are denoted by ‘\*’, conserved substitutions by ‘:’, and semi-conserved substitutions by ‘.’ under the alignment. The  $\alpha$ -crystallin domain is gray-shaded and the  $\beta$ -strands are indicated above the alignment. For TaHsp16.9, BtHspB4, HsHspB5, RnHspB6, and MjHsp16.5 the  $\beta$ -strand-forming residues according to the crystal structures (PDB ID’s 1GME, 3L1E, 2WJ7, 2WJ5, and 1SHS, resp.) are underlined (adapted from (Basha et al., 2011)). The highly conserved I/V/L-X-I/V/L motif is pinpointed in bold.

Structures of vertebrate sHsps show a different mode of dimer formation. There is no long loop bearing strand  $\beta$ 6, but instead  $\beta$ 6 forms an extended  $\beta$ -strand together with  $\beta$ 7, called  $\beta$ 6+7 (figure 2B,D). Strand  $\beta$ 6+7 of one monomer binds to strand  $\beta$ 6+7 of the other monomer in anti-parallel fashion, forming an extended  $\beta$ -sheet made up of two monomers. This monomer-monomer interface was first characterized in human  $\alpha$ A-crystallin (HspB4) and Hsp27 (HspB1) by spin labelling and electron paramagnetic resonance (EPR) studies (Berengian et al., 1997; Berengian et al., 1999; McHaourab et al., 1997), and later also by crystallography for  $\alpha$ A-crystallin (HspB4),  $\alpha$ B-crystallin (HspB5), and Hsp20 (HspB6) (Bagneris et al., 2009; Clark et al., 2011; Laganowsky et al., 2010; Laganowsky and Eisenberg, 2010). In the crystal structure of human Hsp27, a different interface, between strands  $\beta$ 4 and  $\beta$ 7, holds together the  $\alpha$ -crystallin domain of monomers, but this appears to be a crystal form only – the same protein construct in solution was shown to form a dimer held together by the interface between two  $\beta$ 6+7 strands, as determined by SAXS (Baranova et al., 2011). A study on human  $\alpha$ B-crystallin, combining solid-state NMR and SAXS, revealed the same dimer architecture (Jehle et al., 2010).



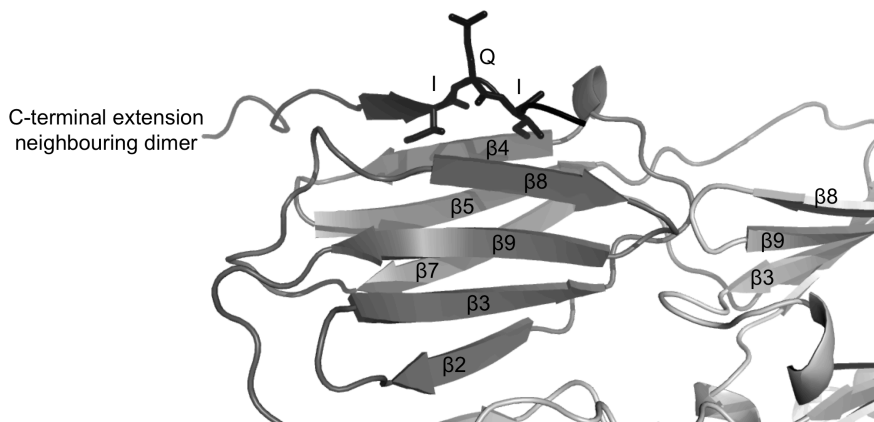


**Figure 2.** Architecture of the dimeric building block. A: Schematic representation of how the  $\beta$ -strand-exchange of  $\beta 6$  between two monomers stabilizes the dimeric building block. This topology of the  $\beta$ -strands can be found in for example Hsp16.9 from *T. aestivum*. B: Schematic representation of how two extended  $\beta$ -strands ( $\beta 6+7$  from each monomer) interact with each other to stabilize the dimeric building block. This topology of the  $\beta$ -strands can be found in vertebrate sHsps such as human  $\alpha$ B-crystallin. C: The  $\alpha$ -crystallin domain of Hsp16.9 from *T. aestivum* (PDB ID 1GME, residues 46-137 of chains A and B). D: The  $\alpha$ -crystallin domain of human  $\alpha$ B-crystallin (PDB ID 2WJ7, residues 69-153 of chain A and residues 78-149 of chain B). Figures 2C and D were prepared with PyMOL ([www.pymol.org](http://www.pymol.org)) (DeLano, 2002).

## 2.2 Oligomeric structure

Most of the small heat shock proteins characterized so far, form large oligomers. Several interactions hold the dimeric building blocks of these oligomers together, of which the interaction involving the conserved I/V/L-X-I/V/L motif on the C-terminal extension is best defined. Interactions involving the N-terminal region are also believed to be important for oligomer formation, but in most of the crystal structures mentioned in the previous section, the N-terminal region is unstructured, or missing from the start because a truncated protein was used for crystallization. This implies, together with other experiments such as hydrogen/deuterium exchange (Cheng et al., 2008) and solution-state NMR (Jehle et al., 2011), that the N-terminal region is intrinsically disordered.

The conserved I/V/L-X-I/V/L motif binds a hydrophobic groove created by strands  $\beta 4$  and  $\beta 8$  on a neighbouring dimer, strands  $\beta 4$  and  $\beta 8$  being the ‘edge’ of the  $\beta$ -sandwich formed by the  $\alpha$ -crystallin domain (figure 3). This interaction was first observed in the first sHsp crystal structure of Hsp16.5 from *M. jannaschii* (Kim et al., 1998), and has since been found in all other sHsps that were structurally characterized, albeit with two opposite orientations, which is possible because of the motif’s palindromic nature (Laganowsky et al., 2010).

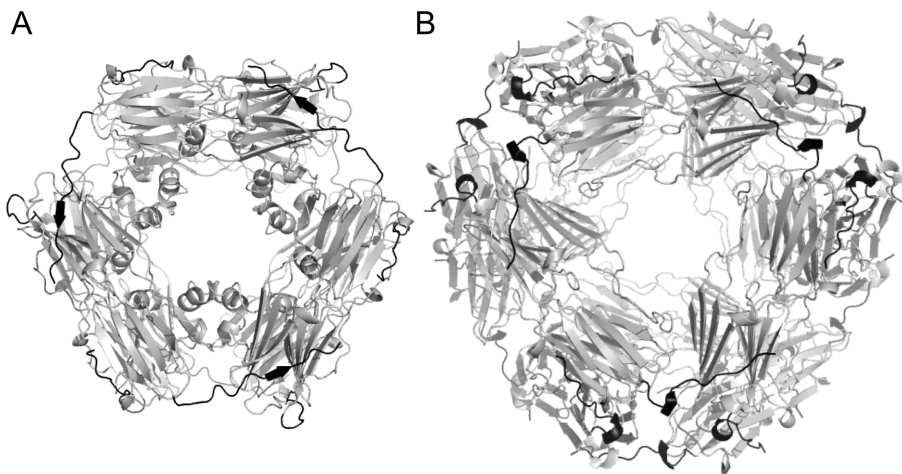


**Figure 3.** The I/V/L-X-I/V/L motif on the C-terminal extension interacts with a hydrophobic groove formed by  $\beta$ -strands 4 and 8. A dimer (gray) from Hsp16.9 from *T. aestivum* is shown together with the C-terminal extension (residues 138-151) of a neighbouring dimer (black) (PDB ID 1GME). The figure was prepared with PyMOL ([www.pymol.org](http://www.pymol.org)) (DeLano, 2002).

The flexibility of the part of the C-terminal extension between the  $\alpha$ -crystallin domain and the I/V/L-X-I/V/L motif, allows endless variation on how the dimeric building blocks can be oriented with respect to each other. This is reflected in the large variation of oligomeric states of sHsps. Not only do different sHsps differ in their oligomeric states, but for some sHsps, a single protein exists as an ensemble of differently sized oligomers. As protein crystallography intrinsically selects for a single or a few protein orientations, alternative techniques such as electron microscopy (EM), SAXS, and more recently native mass spectrometry (MS), have been very valuable to study especially the highly polydisperse sHsps. Several examples of both homogeneous and polydisperse sHsps will be discussed below.

Examples of homogeneous sHsp oligomers are Hsp16.5 from *M. jannaschii* and Hsp16.9 from *T. aestivum*, which are the only two oligomers that have been successfully crystallized in their native oligomeric state (Kim et al., 1998; Van Montfort et al., 2001b) (figure 4). Hsp16.5 is a 24-mer with the dimeric building

blocks forming the edges of an octahedron. The C-terminal extensions are wrapping around the octahedron on the outside, whereas the N-terminal regions appear to be located on the inside, but they are unstructured. Hsp16.9 is a dodecamer (12-mer), formed by two stacked hexamers each of which is made up of three dimers. In this structure, the C-terminal extensions are also located on the outside the two hexameric rings. Six of the N-terminal regions are unstructured, but the other six are structured on the inside of the dodecamer and connect dimers from the two different rings. So in this case, the N-terminal region forms important oligomeric contacts, in addition to the established contacts of the C-terminal extension binding the hydrophobic groove on the next dimer. Another dodecameric sHsp, Acr1 from *M. tuberculosis*, was characterized by cryo- and negative stain EM and shown to form a tetrahedron, with the dimers on its six edges (Kennaway et al., 2005).



**Figure 4.** sHsp oligomeric structure. Despite their different sizes, sHsp oligomers are built up of similar dimeric building blocks (gray) that are connected by C-terminal extensions (black) containing the I/V/L-X-I/V/L motif that can bind a neighbouring dimer. Shown are two sHsp crystal structures, both of which were crystallized in their native oligomeric state. A: Hsp16.9 from *T. aestivum* (12-mer) (PDB ID 1GME). B: Hsp16.5 from *M. jannaschii* (24-mer) (PDB ID 1SHS). Figures 4A and B were prepared with PyMOL ([www.pymol.org](http://www.pymol.org)) (DeLano, 2002).

Many sHsps do not form homogeneous oligomers, but instead form a wide range of differently sized oligomers. Even Hsp16.5 from *M. jannaschii* can be induced to form novel structural ensembles by genetically engineering short flexible peptide insertions into the wild-type protein, as shown by cryo-EM and EPR (Shi et al., 2006). Human  $\alpha$ B-crystallin was recently recorded to span a range of anywhere between 12 and 48 monomers (Baldwin et al., 2011c). Given this enormous

polydispersity, it is not surprising that  $\alpha$ B-crystallin and other sHsps have always frustrated crystallization attempts of their full-lengths forms. Almost all of the available sHsp crystal structures are structures of truncated protein constructs, and do not represent the native oligomeric state (Basha et al., 2011).

EM, SAXS, and native MS have been more successful than crystallography in characterizing sHsp oligomeric structure(s). A novel strategy was developed to separate mixed populations of structurally different single particles from cryo-EM micrographs (White et al., 2004), by which different assembly forms of Hsp26 from *S. cerevisiae* could be distinguished (White et al., 2006). Interestingly, only 24-mers of Hsp26 were characterized in this study, whereas the same protein was later shown to occupy a range of oligomeric states by native MS (Benesch et al., 2010), even though the 24-meric form is most abundant at room temperature. The oligomeric structure of human  $\alpha$ B-crystallin has also been studied by several different techniques. Based on single particle negative stain EM, a 24-mer was reconstructed (Peschek et al., 2009). Two studies combining solid-state NMR, SAXS, single particle negative stain EM, and computational methods also report on a 24-meric model for  $\alpha$ B-crystallin, including atomic coordinates, and propose a mechanism for multimerization into higher-order oligomers (Jehle et al., 2010; Jehle et al., 2011). However, real insight into the polydispersity and the distribution of the different oligomeric structures of  $\alpha$ B-crystallin has come from a combination of NMR, native MS, ion-mobility (IM), and EM experiments, which together were able to provide structural models for the most abundant oligomers (24-, 26-, and 28-mers) and to suggest routes of conversion between the different oligomeric states (Baldwin et al., 2011b). The development of the MS methodology that allows nano-electrospray ionization (ESI) and subsequent mass analysis of intact protein complexes has been crucial to these results (Benesch et al., 2007), and this technique has also proven to be useful to study the dynamics of sHsp oligomers (Sharon and Robinson, 2007), as described in the next section.

## 2.3 Dynamic subunit exchange

A characteristic feature of sHsps is their dynamic subunit exchange, which means that subunits are continuously dissociating from and associating with the larger oligomers. This process is important to maintain polydispersity by allowing conversion into different oligomeric states, but homogeneous sHsp oligomers are also highly dynamic. A method based on fluorescence resonance energy transfer (FRET) was developed to monitor the subunit exchange of  $\alpha$ A-crystallin (Bova et al., 1997), and used to show that  $\alpha$ A-crystallin,  $\alpha$ B-crystallin, and Hsp27 can reversibly form heterooligomers (Bova et al., 2000). Several other sHsps have also been shown to be capable of forming heterooligomers when mixed with each other, even though it is unknown to what extent heterooligomer formation is

relevant *in vivo*. Two sHsps from different species, Hsp18.1 from *P. sativum* (pea) and Hsp16.9 *T. aestivum* (wheat) can form heterooligomers, and their subunit exchange was measured by nano-ESI MS (or ‘native MS’) (Sobott et al., 2002).

Important developments in nano-ESI MS technology allowed real-time monitoring of subunit exchange between two related sHsps from *A. thaliana*, Hsp17.6 and Hsp18.1 (Painter et al., 2008). Notably, dimers were concluded to be the exchanging subunits in this case, whereas the heterooligomers composed of odd numbers of sHsps, implying monomeric subunit exchange, were observed when mixing Hsp18.1 from *P. sativum* and Hsp16.9 *T. aestivum* (Sobott et al., 2002). The dynamic subunit exchange of sHsps is considerably faster (0.04 to 0.40/min) than that of other proteins ( $4.5 \times 10^{-4}$ /min) for which similar measurements have been carried out, and the exchange rate is increased with increasing temperature, suggesting these dynamics to be essential for sHsp function (Basha et al., 2011). Upon temperature increase, Hsp18.1 from *P. sativum* shifts from being almost exclusively dodecameric to forming monomers, dimers, and larger oligomers. In presence of a substrate protein, a wide range of differently sized sHsp-substrate complexes with different stoichiometries form, showing that the dynamics and polydispersity of sHsps are integral to chaperone function (Stengel et al., 2010). A combination of NMR and nano-ESI MS experiments on  $\alpha$ B-crystallin, showed that dissociation events of the C-terminal-extension-to-hydrophobic-groove interaction of  $\alpha$ B-crystallin govern the dynamics of the polydisperse ensemble of  $\alpha$ B-crystallin oligomers (Baldwin et al., 2011a).

## 2.4 Chaperone activity and substrate interactions

Classically, the chaperone activity of sHsps is assayed *in vitro* using model substrate proteins that are known to unfold upon temperature increase. In presence of sHsps, the model substrate proteins remain soluble upon temperature increase, whereas in absence of sHsps, they unfold and start to aggregate. Protein aggregation can be measured by light scattering or analyzed by centrifugation and subsequent gel electrophoresis (Basha et al., 2011). Or, protein deactivation as a consequence of unfolding can be monitored with a protein activity assay. Typically used thermo-sensitive model substrate proteins are citrate synthase (CS), luciferase, and malate dehydrogenase (MDH). To study the sHsp interaction with a predominantly folded protein (in contrast to the heat-stressed unfolded proteins mentioned above), fluorescently labelled T4 lysozyme variants with slightly destabilizing mutations have been especially designed (McHaourab et al., 2009).

An *in vivo* assay for chaperone activity was developed using a  $\Delta$ ClpB1 strain of *Synechocystis*, where thermotolerance strongly depends on the presence of functional Hsp16.6 – the only sHsp gene in this organism. By replacing wild-type Hsp16.6 with mutations affecting oligomerization *in vitro*, the effect of these

mutations could also be studied *in vivo*, linking changes in oligomerization ability to chaperone activity (Giese and Vierling, 2002, 2004). In another study, Hsp16.6 mutations were screened for, that reduced the ability of the protein to provide thermotolerance *in vivo*, and a selection of these was overexpressed and purified, to study their oligomeric structure and *in vitro* chaperone activity (Giese et al., 2005). Mutations in the  $\alpha$ -crystallin domain compromised both oligomerization and chaperone activity *in vitro*, whereas mutations in the N-terminal region did not have an effect *in vitro*, emphasizing the importance of *in vivo* studies.

Attempts to identify endogenous substrate proteins indicate that proteins of a wide variety of cellular functions can be protected by sHsps in bacteria and yeast (Basha et al., 2004; Haslbeck et al., 2004). It remains to be determined whether this general protection is common for all sHsps, or whether some sHsps may have more specific substrates than others. More specific interactions have been reported for mammalian sHsps (Van Montfort et al., 2001a; Vos et al., 2008), such as the interactions of  $\alpha$ B-crystallin and Hsp27 with cytoskeletal components (During et al., 2007; Ghosh et al., 2007; Goldfarb et al., 2008).

The molecular details of the interaction between sHsps and substrate proteins are still largely unknown (Basha et al., 2011; Haslbeck et al., 2005; McHaourab et al., 2009). Unfolding proteins generally have increasingly exposed hydrophobic regions, which leads to the logical hypothesis that sHsps bind these regions to maintain solubility. It is generally assumed that sHsp hydrophobic surfaces buried in the oligomers, can become available for substrate binding (Haslbeck et al., 2005; McHaourab et al., 2009; Van Montfort et al., 2001a). This is supported by studies showing the relationship between oligomerization and chaperone activity (Benesch et al., 2008; Giese and Vierling, 2002, 2004). However, it is debated whether monomeric, dimeric, or oligomeric species are responsible for substrate recognition. Because both the N-terminal region and the C-terminal extension are known to be flexible (described in section 2.2), oligomer dissociation may not be required for substrate binding. In addition, the dynamic subunit exchange of the oligomers (as described in section 2.3) supports continuous exposure potential substrate binding regions without dissociation into stable suboligomeric species (Basha et al., 2011). The emerging picture is that sHsp chaperone activity is based on a set of delicate equilibria between sHsp oligomers, sHsp suboligomeric species, native substrate, and unfolding substrate (figure 5).

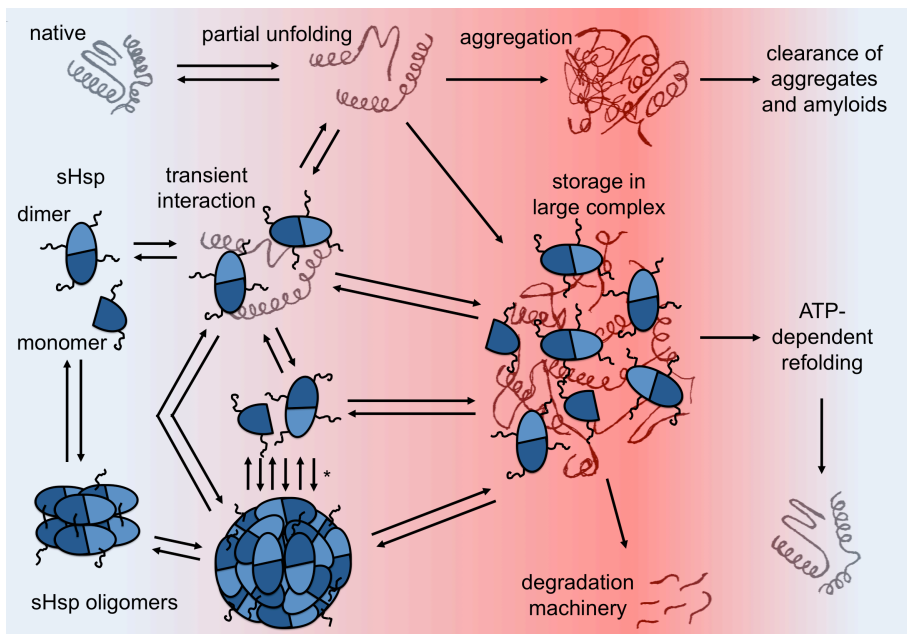
Apart from whether substrate binding occurs via monomers, dimers or oligomers, or a mixture of these, sHsps recognize substrates at an early stage of unfolding (McHaourab et al., 2009), initially interacting with them reversibly (Ahrman et al., 2007b) (paper I, this thesis). Upon prolonged denaturing conditions, large heterogeneous sHsp-substrate complexes form, keeping the partially unfolded substrate protein soluble. Once in these large complexes, the substrate proteins cannot refold without the help of ATP-dependent chaperones, but ATP-dependent

refolding is much more efficient in presence of sHsps (Mogk et al., 2003a; Mogk et al., 2003b), even after aggregation (Ratajczak et al., 2009). The substrate proteins appear to be firmly bound within the complexes, whereas (a subpopulation of) sHsp subunits remain dynamic (Friedrich et al., 2004), probably enhancing the solubility of the complex. The polydisperse sHsp-substrate complexes can be observed by size-exclusion chromatography (SEC), and have been studied by EM, but their heterogeneity has prevented any conclusions to be drawn on their structure. Recently, it was shown by nano-ESI MS that Hsp18.1 from *P. sativum* and luciferase build up to 300 different complexes during incubation at increased temperature, varying in both size and stoichiometry (Stengel et al., 2010). This finding clearly links the dynamics and polydispersity of sHsps with their chaperone activity. Considering the polydispersity of one species of sHsp by itself (Baldwin et al., 2011c), and then the polydispersity of just one sHsp and one substrate protein, it can be imagined how an armoury of several sHsps in the cell is capable of protecting such a wide variety of proteins (Stengel et al., 2010).

Given the heterogeneity of the complexes described above, it is no surprise that no one particular sHsp binding site for substrate proteins has been identified. In analogy to the polydisperse oligomeric ensemble of an sHsp alone, the sHsp-substrate complexes are also presumed to be dynamic, at least when it concerns sHsp subunits, allowing quick adaptation to a changing 'environment' of (partially) unfolded proteins (Stengel et al., 2010). This is supported by the same hydrogen/deuterium exchange levels of sHsps in absence or presence of substrate protein MDH, which indicates that the flexibility of especially the N-terminal and C-terminal regions is retained even within sHsp-substrate complexes (Cheng et al., 2008). However, in a study where a photo-activatable crosslinker was introduced on different positions of the sHsp Hsp18.1, crosslinking between the N-terminal region and the substrate was much more frequently observed than crosslinking with the  $\alpha$ -crystallin or C-terminal extension (Jaya et al., 2009), suggesting the N-terminal region to be responsible for substrate binding. By constructing chimera's of the previously well-characterized sHsps Hsp18.1 from *P. sativum* and Hsp16.9 *T. aestivum*, it was shown that the N-terminal region is important for chaperone activity towards the substrate proteins CS and luciferase, but that for MDH, both the N-terminal region and the  $\alpha$ -crystallin domain are important (Basha et al., 2006).

The chaperone activity and substrate binding modes may also vary for different sHsps, as confirmed by a study comparing two classes of plant sHsps (Basha et al., 2010). Additionally, different mechanisms of activation have been found for different sHsps. Vertebrate sHsps can be phosphorylated, thereby effecting oligomerization, which is probably regulating activity (Ecroyd et al., 2007; Shashidharamurthy et al., 2005). The yeast sHsp Hsp26 has a 'middle domain' before the  $\alpha$ -crystallin domain that acts as a thermosensor, and induces a switch to

a high-affinity oligomeric conformation upon temperature increase (Franzmann et al., 2008), as well as changed dynamics (Benesch et al., 2010). However, this domain is not general in all sHsps. Obviously, since their success very early on in evolution, sHsps have evolved into a divergent superfamily, with different members employing a variety of strategies for activation, substrate binding and substrate protection. Yet, general to all sHsps seems to be their ability to act as sensors capable of adapting to very subtle changes in the protein stability environment in the cell (McHaourab et al., 2009), thereby forming a first line of defence against stress.



**Figure 5.** Possible model for sHsp chaperone function. The background shading refers to the degree of stress (inducing unfolding of proteins in the cell), with red for much stress and light-blue for little stress. The asterisk (\*) highlights the increased subunit exchange of the large sHsp oligomers, resulting in the increased exposure of substrate binding sites. The figure was in part adapted from (Lindner et al., 2001) and (Basha et al., 2011).

## 2.5 Hsp21 in *Arabidopsis thaliana* chloroplasts

The sHsps as part of the protein homeostasis network in cells have a particularly pronounced role in plants. During heat stress in plants, sHsps have higher



expression levels than for example Hsp70, which often is the most abundant heat shock protein in other heat stressed eukaryotes (Waters et al., 1996). Plants also have among the highest number of genes encoding sHsps. Whereas the genomes of for example *E. coli* and *S. cerevisiae* only encode two sHsps (IbpA and B, and Hsp26 and Hsp42, respectively), *A. thaliana* has 19 genes encoding sHsps (Scharf et al., 2001; Siddique et al., 2008). Currently, 11 subfamilies of plant sHsps have been recognized, of which 6 are cytosolic, and 5 are organelle-localized: one in peroxisomes, one in the endoplasmic reticulum, one in chloroplasts, and two in mitochondria (Waters et al., 2008). Organelle-localized sHsps are almost unique to plants, with just one known exception of a mitochondrial sHsp in *D. melanogaster* (Basha et al., 2011). The chloroplast sHsps most likely evolved via gene duplication from a nuclear-encoded cytosolic sHsp, and not via gene transfer from the chloroplast endosymbiont (Waters and Vierling, 1999).

The focus of this thesis is Hsp21 (AtHsp25.3), which is the (only) chloroplast-localized sHsp in *A. thaliana*, and has orthologs in all higher plants. Transgenic *A. thaliana* plants that constitutively overexpress Hsp21 were shown to be more resistant to heat stress (Harndahl et al., 1999). Characteristic of Hsp21 is its relatively long N-terminal region, which contains a highly conserved methionine-rich domain, structure-predicted to form an amphipathic  $\alpha$ -helix. When these methionines are oxidized, the chaperone activity of Hsp21 is lost (Harndahl et al., 2001), but can be restored by the protein peptide methionine sulfoxide reductase (Gustavsson et al., 2002). This may be a mechanism to regulate Hsp21 activity, as well as a way to scavenge reactive oxygen species (Sundby et al., 2005).

The chaperone activity and substrate interaction of Hsp21 have previously been investigated using the model substrates citrate synthase (CS) and malate dehydrogenase. Hsp21 can protect substrate protein up to its own weight, and possibly even more (unpublished results), affirming its status as a powerful chaperone. A peptide array screen covering the sequence of porcine CS revealed strongest binding to the most N-terminal peptide of CS, which is part of a domain that is missing in CS from thermophilic archaea (Ahrman et al., 2007a). Interactions between Hsp21 and CS have also been characterized by crosslinking mass spectrometry (Ahrman et al., 2007b) (paper I, this thesis). Crosslinking mass spectrometry will be further exploited in the work described in this thesis to investigate interactions within and between Hsp21 and substrate proteins.

# 3 Crosslinking mass spectrometry

The use of crosslinking reagents in protein research is well established and was already reported in the 1970's, to study the topology of *E. coli* ribosomes (Clegg and Hayes, 1974). The 1990's were marked by the beginning of the era of mass spectrometry-based proteomics, powered by technical advances in protein mass spectrometry (Aebersold and Mann, 2003). During the last decade, an approach combining the strengths of both crosslinking and mass spectrometry has emerged, called crosslinking mass spectrometry (also called CXMS or MS3D), the first studies based on this technology being reported in the beginning of this century (Young et al., 2000). Important developments of specialized crosslinking reagents, mass spectrometry instrumentation, and software for crosslinking data analysis have allowed this method to mature into a useful tool in structural biology (Stengel et al., 2011).

Within the field of crosslinking mass spectrometry, two general workflows can be distinguished (Singh et al., 2010; Sinz, 2006). The 'top-down' approach involves the mass spectrometric analysis of intact crosslinked proteins or protein complexes, which are further interrogated by fragmentation within the mass spectrometer. In the work described in this thesis, the alternative 'bottom-up' approach has been used, which involves the enzymatic digestion into peptides, and will be discussed in detail here. The principles of bottom-up crosslinking mass spectrometry are: to crosslink amino acid residues within or between proteins, to enzymatically digest the proteins into peptides, and to look for crosslinked peptides by mass spectrometry. Deducing which amino acid residues have crosslinked on the identified peptides, yields structural information about the protein or protein complexes, because the known length of the crosslinker places a constraint on the distance between the residues (Lee, 2008; Leitner et al., 2010; Sinz, 2006).

Despite the apparent straightforwardness of this approach, many technical challenges have not yet been overcome, and the number of studies where crosslinking mass spectrometry has resolved major biological issues is still very limited (Leitner et al., 2010; Singh et al., 2010; Sinz, 2010). The main reason for this is the low abundance of 'structurally informative' crosslinks, making them difficult to detect. Strategies to tackle this and related problems can be directed at different stages of the crosslinking mass spectrometry workflow.

## 3.1 Complexity of the crosslinked peptide mixture

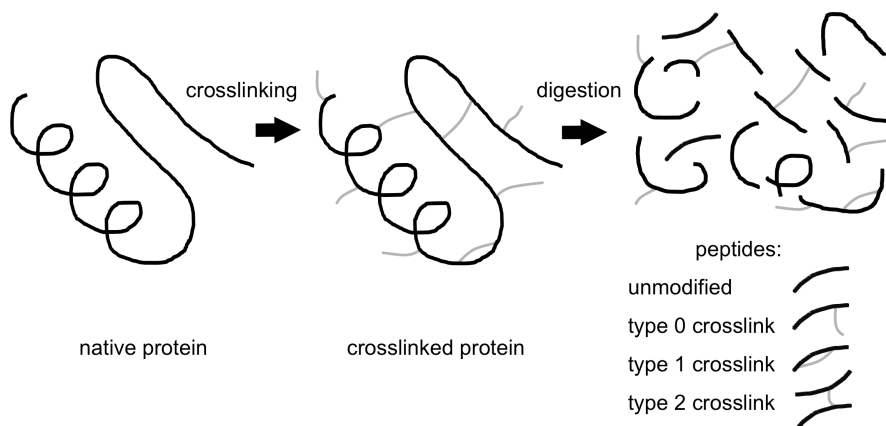
Under typical conditions in a crosslinking experiment, the bulk of the targeted amino acid residues may not be modified at all by the crosslinking reagent. Those residues that do react with the reagent, form differently modified peptides after enzymatic digestion. The most informative type of compound is formed when the crosslinking reagent reacts with two amino acid residues that are in close spatial proximity in the protein structure, but are far away from each other in the protein sequence, or are located on two different protein subunits/proteins (an interpeptide crosslink). However, completely unmodified peptides, peptides where the crosslinking reagent has reacted with the same peptide on both ends (an intrapeptide crosslink), or has reacted with a peptide on one end and a water molecule on the other end (a dead-end crosslink), are much more abundant in the mixture to be analyzed with mass spectrometry. In addition, peptides with multiple crosslinker modifications can be present in the mixture, such as three peptides covalently linked by two crosslinkers.

According to the commonly accepted nomenclature, dead-end crosslinks, intrapeptide crosslinks, and interpeptide crosslinks, are called type 0, type 1, and type 2 crosslinks, respectively (Schilling et al., 2003). Peptides with multiple crosslinker modifications can also be classified according to this system, but they are generally less abundant and often not considered during data analysis, also because their mass spectrometric identification is even more difficult than that of type 2 crosslinks. Figure 6 illustrates which main compounds are expected after crosslinking and subsequently digesting a single protein.

Especially when the starting material consists of several proteins, the resulting peptide sample after crosslinking and digestion is very complex. To reduce the sample complexity and increase the chances of detecting structurally informative type 2 crosslinks, these crosslinks can be enriched using different strategies. One strategy is to separate the peptide mixture by strong cation exchange chromatography (SCX) (Fritzsche et al., 2012; Leitner et al., 2012). Compared to unmodified peptides, and type 0 and type 1 crosslinks, type 2 crosslinks elute at longer retention times because they are more highly charged in solution due to a higher number of protonation sites, especially after digestion with an enzyme cleaving at basic residues, such as trypsin. Another enrichment strategy is the use of crosslinking reagents with an incorporated affinity tag, by which crosslinks (of all types) can be enriched with affinity chromatography.

Before mass spectrometric analysis, the peptide sample can also be separated by reversed phase liquid chromatography (LC). This separation is not necessarily a means of crosslink enrichment, but the separation of the complex mixture reduces the complexity per analyzed fraction, which drastically improves the quality of the mass spectrometry data. Moreover, crosslink enrichment is in fact achieved to some extent, as type 2 crosslinks are typically larger and thus more hydrophobic

than the other peptides, so they tend to elute later from the reversed phase column. During the work described in this thesis, nano-LC (nano referring to the low flow rates) with direct elution onto MALDI-targets was used to separate peptide samples.



**Figure 6.** Crosslinking mass spectrometry sample preparation. A protein sample is crosslinked and subsequently digested into peptides. The main resulting peptide species are unmodified peptides, type 0 crosslinks (only one end of the crosslinking reagent has reacted with an amino acid residue), type 1 crosslinks (both ends of the crosslinking reagent are connected to the same peptide), and type 2 crosslinks (the crosslinking reagent is connecting two different peptides).

## 3.2 Chemical crosslinking reagents

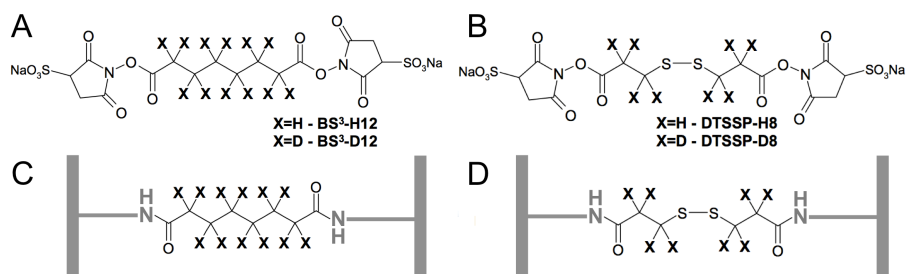
Today, a broad range of crosslinking reagents is available, some of which have been especially designed for crosslinking mass spectrometry experiments. The reagents vary in their linker length and type of reactive groups, but also in the presence of all sorts of other chemical groups, including cleavable groups, affinity groups, and functional groups for mass spectrometry, which for example produce reporter ions upon MSMS fragmentation (Petrotchenko and Borchers, 2010; Sinz, 2006). Crosslinking reagents can also often be isotope-labelled, because this is helpful for the mass spectrometric identification of peptide crosslinks (Muller et al., 2001). Not all reagents have been widely used yet, as only a subset is commercially available, and some are quite laborious to synthesise.

The reactive groups determine which amino acid residues are targeted for crosslinking. Commonly used crosslinking reagents react with amine- (lysine residues and the protein N-terminus) or sulfhydryl (cysteine residues) groups. Homobifunctional crosslinking reagents have two of the same reactive groups on

either end, whereas heterobifunctional crosslinking reagents have two different reactive groups. There are also trifunctional reagents that can crosslink three amino acid residues or where the third group is for example an affinity group.

The length of the linker region is important to consider. A longer crosslinker will yield more crosslinks, but these crosslinks will be less informative, because the resulting distance constraints for the protein structure are less stringent. Zero-length crosslinkers are compounds that mediate covalent linking between amino acid residues without leaving behind a linker region. Such crosslinks are highly informative, but the yield will be lower. Zero-length crosslinkers are often photo-reactive, meaning that UV-light exposure triggers the reaction. Such reagents are non-selective, which positively affects the crosslink yield. However, this is a great disadvantage for mass spectrometric analysis, because possible crosslinker modifications must be considered on all the amino acid residues, making it very difficult to identify crosslinks.

The most popular reagents in crosslinking mass spectrometry are the amine-reactive *N*-hydroxysuccinimidyl and sulfosuccinimidyl esters, the latter being more water-soluble. Advantages of these reagents are their high reactivity and the relatively high prevalence of lysine residues in proteins (Leitner et al., 2010). Disadvantages may be the competing hydrolysis reaction in aqueous solutions, as well as unwanted reactions with serine and threonine hydroxyl groups, and with contaminant ammonium ions in the buffer solution (Sinz, 2006; Swaim et al., 2004). The two crosslinking reagents used in the work described in this thesis are bis(sulfosuccinimidylsuberate) and 3,3'-dithiobis(sulfosuccinimidylpropionate) (BS<sup>3</sup> and DTSSP, respectively) (figure 7). Both of these are commercially available, including their isotope-labelled forms, where the hydrogen atoms on the alkyl chain have been substituted by deuterium atoms.



**Figure 7.** Commonly used lysine-specific crosslinking reagents and their desired reaction products. A: Isotope-labelled BS<sup>3</sup>. B: Isotope-labelled DTSSP. Both compounds are commercially available as 1:1 mixtures of unlabelled and isotope-labelled reagent (Creative Molecules Inc., Victoria, Canada). C: Two peptides crosslinked by BS<sup>3</sup>. D: Two peptides crosslinked by DTSSP.

The difference between DTSSP and BS<sup>3</sup> is the disulfide bridge in DTSSP. The disulfide bridge makes the crosslinker cleavable, which is helpful during identification by mass spectrometry. Upon reduction of the disulfide bond, the peak representing the crosslink should disappear from the MS spectrum, and the two peptides that were crosslinked in a type 2 crosslink should appear in the MS spectrum as two individual peaks representing each peptide modified with a reduced (half of a) crosslinker (Bennett et al., 2000). Collision induced dissociation (CID) fragmentation of type 2 DTSSP crosslinks gives typical 66 Da doublets in the MSMS spectrum, because of the asymmetric fragmentation around the disulfide bond (King et al., 2008). The crosslinking efficiency of DTSSP can be easily assessed with gel filtration or SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis), by comparing samples that have or have not been treated with a reducing agent such as dithiothreitol (DTT) (Bennett et al., 2000). Unfortunately, the disulfide bridge in DTSSP is not only advantageous, because thiol-exchange can lead to disulfide bond scrambling, which can result in false positive type 2 crosslinks. This problem is addressed in paper III in this thesis.

### 3.3 Mass spectrometry instrumentation

A mass spectrometer measures the mass-to-charge ratio ( $m/z$ ) of ions in the gas phase. To be able to do this, the instrument consists of the following three basic modules: an ion source, a mass analyzer, and a detector. The ionization of large biomolecules such as peptides and proteins can be achieved by electrospray ionization (ESI) (Fenn et al., 1989) or matrix-assisted laser desorption ionization (MALDI) (Karas and Hillenkamp, 1988). The inventors of both technologies have been awarded the Nobel Prize, as these soft ionization techniques dramatically expanded the possibilities to study biomolecules by mass spectrometry.

There are four types of mass analyzers typically used for protein research and proteomics: the quadrupole (Q), ion trap (quadrupole ion trap, QIT; linear ion trap, LIT or LTQ) time-of-flight (TOF), and Fourier-transform ion cyclotron resonance (FTICR) mass analyzers (Han et al., 2008). The mass analyzer type determines important analytical properties such as the mass accuracy, mass resolution, sensitivity, and speed of the instrument. Most mass spectrometers are hybrid instruments combining the capabilities of the different analyzers. In this way, proteins or peptides can first be ionized and analyzed intact (MS), and then dissociated into fragments that can also be analyzed (tandem MS or MSMS). Ion trapping instruments can even perform multiple steps of analysis (MS<sup>n</sup>). The type of detector is dependent on the type of mass analyzer.

Examples of classic ESI mass spectrometry instruments commonly used in protein research and proteomics are the Q-Q-Q, Q-Q-LIT, Q-TOF, Q-Q-TOF, and FTICR

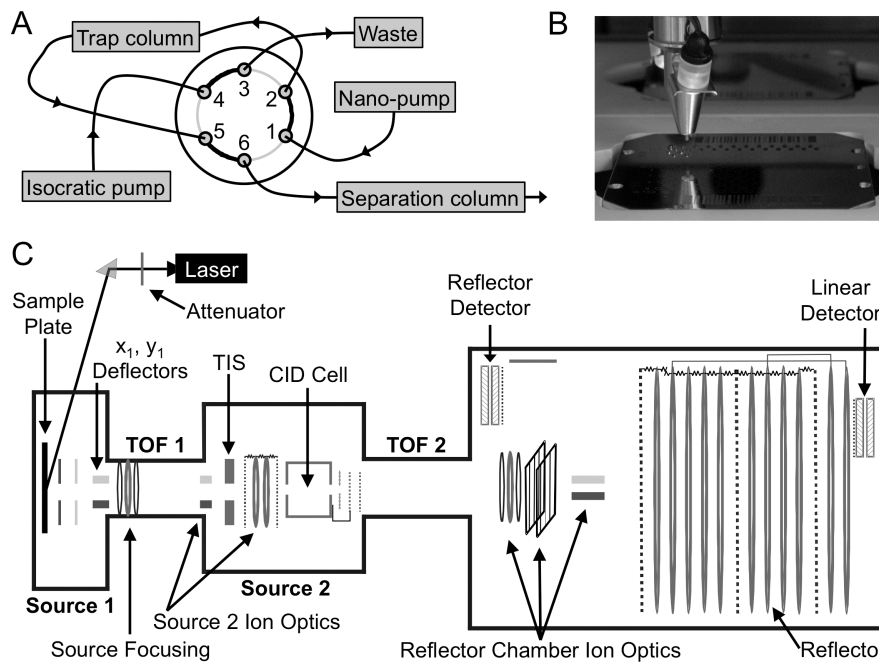
instruments, whereas MALDI ionization is typically combined with a TOF-TOF mass analyzer (Domon and Aebersold, 2006). The introduction of a new type of mass analyzer, the orbitrap, has led to the recent development of the LTQ-Orbitrap instruments, which have an unprecedented analytical performance (Han et al., 2008).

For most of the mass spectrometry experiments described in this thesis, a MALDI-TOF/TOF instrument was used (figure 8C). For peptide ionization with MALDI, a peptide sample is usually co-crystallized with the matrix compound  $\alpha$ -cyano-4-hydroxycinnamic acid on a MALDI-target. In the vacuum of the instrument, laser light pulses produce clouds of mainly matrix ions and analyte molecules. Within a cloud, the charge of the matrix ions gets transferred to the analyte, producing analyte ions which are mostly +1 charged. The ions are extracted into the flight tube for TOF or TOF/TOF analysis. Because the ions in the cloud compete for charge from the matrix ions, a sample containing more different analytes will yield lower intensities in the mass spectrum, a phenomenon called ion suppression. To reduce the ion suppression effect, peptide samples can be separated by reversed phase LC (figure 8A) into very small volume fractions, which are directly spotted onto a 192 well MALDI-target (figure 8B).

During MS analysis with a MALDI-TOF/TOF instrument, ionized peptides with a broad range of masses (as defined by the user) enter the flight tube and reach the detector, resulting in an MS spectrum. For an MSMS spectrum, ionized peptides are again produced from the same sample on the MALDI-target, but this time just one (peptide) precursor ion from the original MS spectrum is selected by the timed ion selector (TIS) for fragmentation in the first part of the flight tube and MSMS analysis in the second part of the flight tube. Because the sample is preserved in dry form on the MALDI-target, MSMS analyses do not need to follow MS analysis directly. MSMS (or MS) analyses can also be repeated on the same sample on the MALDI-target (provided there is enough material left on the target). The nano-LC system and the mass spectrometer do not need to be directly coupled either, so the complete instrument setup is also being referred to as offline-LC-MALDI-TOF/TOF.

MS fragmentation in a MALDI-TOF/TOF instrument is mediated by collision induced dissociation (CID), which causes peptides to preferentially fragment at their peptide bonds. Most of the times, there is just one fragmentation event per peptide molecule, so the resulting fragments are pieces of the peptide from either end, and the difference in mass between them corresponds to one or several amino acid residues, thus providing sequence information about the peptide. However, fragmentation is not limited to the peptide bond, and it is normally not possible to ‘read off’ the peptide sequence or parts of it manually from the MSMS spectrum. Automated identification of peptides by programs such as Mascot (Perkins et al.,

1999) and SEQUEST (Yates et al., 1995) is based on an empirically determined set of fragmentation rules (Paizs and Suhai, 2005).



**Figure 8.** Offline-LC-MALDI-TOF/TOF. A: Reversed phase nano-LC to separate a peptide sample. After initial binding of peptides to the trap column, they are subsequently separated on the reversed phase separation column by gradient elution at low flow rates delivered by the nanopump. B: The fractions resulting from LC separation are directly collected on a MALDI target. C: Schematic representation of a MALDI-TOF/TOF instrument, the 4700 Proteomics Analyzer from former Applied Biosystems (currently AB Sciex).

Concerning the mass spectrometry of crosslinked peptide samples, the interpretation of MSMS spectra of type 2 crosslinks is even more difficult. Fragment series from four ‘ends’ of this type of molecule can exist in the same spectrum, and there are additional fragments and in some cases rearrangements in the gas phase that are caused by the crosslinker (King et al., 2008; Santos et al., 2011). Moreover, high quality MSMS spectra are often difficult to obtain, because the intensity of the precursor crosslink peak in the MS spectrum is usually very low compared to unmodified peptide peaks. The low intensity of crosslink peaks can be largely attributed to their low abundance in the sample, but the ionization of crosslinked peptides may also be less efficient than that of unmodified peptides.



## 3.4 Data analysis – identifying crosslinks

Most of the recent progress in the field of crosslinking mass spectrometry has been driven by the development of software that can handle mass spectrometric data from crosslinking experiments. Whereas intelligent software for the mass spectrometric identification of non-crosslinked peptide samples was developed more than a decade ago (Perkins et al., 1999; Yates et al., 1995), widely applicable software for the analysis of crosslinking mass spectrometry data is only just emerging (Fabris and Yu, 2010; Gotze et al., 2012; Mayne and Patterson, 2011; Rasmussen et al., 2011). The challenge of the identification of in particular type 2 crosslinks lies in the vast theoretical search space that needs to be considered (Rinner et al., 2008). Whereas the number of theoretical unmodified peptides only linearly increases with the length of the sequences of the proteins digested, the combinatorial increase of the number of theoretically crosslinked peptides from a crosslinked protein sample is much faster; *i.e.* so fast, that the number of theoretically possible crosslink combinations from 30 crosslinked proteins, has the same order of magnitude as a theoretical tryptic digest of 100,000 proteins (the proteome of an organism for example) (Panchaud et al., 2010).

Most of the available programs were designed for specific experimental methods, and some of them require non-crosslinked control samples, isotope-labelled proteins or isotope-labelled crosslinking reagents (Mayne and Patterson, 2011). Other important differences between programs include whether the crosslink identification is based on MS and/or MSMS data, whether the crosslinked residues within the crosslink can be identified, and whether identified crosslinks receive some kind of score, allowing evaluation of the identification. Especially for very complex samples, for which a vast search space needs to be considered, sophisticated scoring functions are a prerequisite for the confident identification of crosslinks, and MSMS data need to be the basis for identification. Even for simple systems with only a few proteins, unambiguous assignment of a candidate crosslink peak in an MS spectrum to one particular combination of two peptides that got crosslinked is often not possible, because the mass matches to several combinations of peptides. The frequency of ambiguity in matching candidate masses is of course decreased when the data can be analyzed with lower tolerance settings, indicating the benefit of mass spectrometry data with high accuracy.

In the work described in this thesis, crosslinking experiments were conducted with at most 3 different proteins. The software FINDX was developed to support crosslink identification in these low-complexity systems, in a workflow with offline-LC-MALDI-TOF/TOF (papers III, IV). Crosslink identification by FINDX is based on initially matching experimental masses from MS data to theoretically possible crosslinks for the user-defined proteins, followed by validation of the crosslinks by the analysis of MSMS data, as described in more detail in the next chapter.

### 3.5 Interpretation of detected crosslinks

After the identification of crosslinks from crosslinked protein samples, the crosslinks should be placed in the context of the protein structure or protein-protein interactions. In case of homooligomeric complexes, it is not possible to know for a given type 2 crosslink whether it is linking two peptides from the same subunit (intra-monomeric), or from two different subunits (inter-monomeric). This information can be obtained from a MIX experiment with mixed heterooligomers containing 50% unlabeled and 50%  $^{15}\text{N}$ -isotope-labelled protein (Taverner et al., 2002). An intra-monomeric crosslink will yield two mass peaks (corresponding to the  $^{14}\text{N}$ - $^{14}\text{N}$  and the  $^{15}\text{N}$ - $^{15}\text{N}$  isotopic forms), whereas an inter-monomeric crosslink will yield four mass peaks (corresponding to the  $^{14}\text{N}$ - $^{14}\text{N}$ ,  $^{14}\text{N}$ - $^{15}\text{N}$ ,  $^{15}\text{N}$ - $^{14}\text{N}$ , and  $^{15}\text{N}$ - $^{15}\text{N}$  isotopic forms).  $^{15}\text{N}$ -isotope-labelling is also useful in determining which crosslink match is correct in case of several candidate alternatives with the same or similar mass, because the different candidate alternatives may not have the same  $^{15}\text{N}$ -isotope-labelled mass.

Another potential issue when interpreting identified crosslinks is that the detected crosslinks do not necessarily reflect the majority of the population of protein molecules in the sample. Unless specifically designed to be so, mass spectrometry experiments are not quantitative, and the detection of an analyte means that it was present in the sample, but the absence of a signal does not necessarily mean that the analyte was absent in the sample. There are many factors that determine whether a crosslink is successfully identified, given its presence in the protein sample. First of all, enzymatic digestion of the crosslinked protein has to go to completion, which may be more difficult for crosslinked proteins than for native proteins. Secondly, the crosslink to be detected has to be MS-compatible– not too small and not too large, and ionisable (in this thesis by MALDI). Thirdly, it should give high quality MS and MSMS-data, and finally, it should be identified by the software. From crosslinking mass spectrometry studies on proteins with known structures, it is known that the number of detected crosslinks lies far below the number of expected crosslinks based on the proximity of crosslinkable residues in the protein structure (Leitner et al., 2010).

Despite this ‘under-sampling’ of crosslinks that currently still is inherent to the method, some impressive studies have already shown that it is possible to obtain new insights by crosslinking mass spectrometry. Examples are studies on the Ndc80 complex (Ciferri et al., 2008; Maiolica et al., 2007), the GroEL-GroES system (Trnka and Burlingame, 2010), and RNA polymerase I (Chen et al., 2010). Another complex, AKAP79, was investigated by a combination of native mass spectrometry and crosslinking mass spectrometry (Gold et al., 2011), convincingly demonstrating the strength of crosslinking mass spectrometry in hybrid structural biology approaches.

## 3.6 Complementary structural biology techniques

The major advantages of crosslinking mass spectrometry include its speed, sensitivity, and compatibility with large complexes. In addition, the proteins investigated are in solution, and under native conditions. However, the structural information from crosslinks is of rather low resolution (depending on the length of the crosslinker), and the yields are typically quite low so far. Therefore, the method is well suited to combine with other structural biology techniques to study proteins, some of which will be briefly mentioned here.

Protein X-ray crystallography is superior to most other structural biology methods in terms of resolution – atomic coordinates are the final result of this method (Drenth, 2007). However, diffracting crystals needed for structure determination can be difficult to obtain and generally require large amounts of highly purified protein. Once at the stage of diffracting crystals, structure determination can still not be guaranteed to be successful. Inherent to the method, is the perfect homogeneity of the determined protein structure, because crystals are made up of ‘endlessly’ repeating units. Therefore, heterogeneous samples (as more often the case for larger proteins/protein complexes) are less suitable for crystallography, even though a homogeneous subpopulation of the samples could be crystallized and structurally determined.

In the field of protein NMR, NMR spectroscopy can be used to deduce structural information based on the magnetic properties of atomic nuclei, but also protein dynamics can be studied by this method (Berg, 2012). NMR studies can provide atomic resolution, but is even more than X-ray crystallography limited to smaller proteins. To study protein-protein interactions, isotope-labelling with  $^2\text{D}$ ,  $^{13}\text{C}$  and/or  $^{15}\text{N}$  is usually required. EPR spectroscopy is similar to NMR in its basic physical concepts, but is based on the spin properties of unpaired electrons. Unpaired electrons are not normally present in stable biomolecules such as proteins, but can be introduced by non-reactive radical reagents that have been specifically designed for the purpose of spin-labelling biological samples.

Electron microscopy is the only microscopy technique that can approach the resolution needed to see details of proteins and protein complexes (Zhou, 2008). To obtain a high resolution, single particle analysis is applied to transmission electron microscopy (TEM) images of tens of thousands of protein particles. For the selection of particles, they need to have sufficient contrast, which is more easily obtained for larger protein complexes. In cryo-electron microscopy, the particles are suspended in vitreous ice at very low temperatures, which keeps the proteins in their native state in solution, and limits radiation damage. Even better contrast is obtained by negative staining, but the stain may effect the native protein conformation. After taking images of tens of thousands of protein molecules deposited on carbon-coated grids, particles are manually or automatically selected for further image processing. The particles are categorized into different groups,

representing different views (orientations) of the molecule. The higher the symmetry of the particle (as for instance in the case of homooligomeric protein complexes), the fewer groups of different views are needed to characterize the particle. The different views together are used to reconstruct a 3D model of the particle. In paper II in this thesis, single particle negative stain EM has been used to reconstruct a 3D model of Hsp21. Whereas single particle EM is traditionally used to study a homogeneous population of particles, advancements in image processing have allowed the separation of different particles (such as the same protein in different conformations) in a heterogeneous population (White et al., 2004).

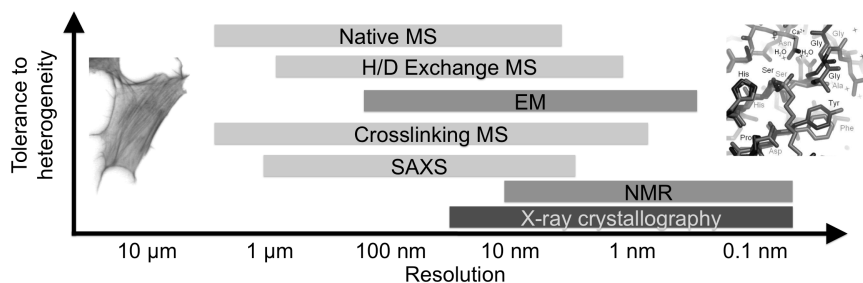
Multiangle light scattering (MALS) experiments can provide the molar mass and the average size of particles in solution, whereas measuring how particles scatter X-rays (SAXS) or neutrons (SANS) at small angles provides structural information of a resolution in the nanometer range (Putnam et al., 2007). The shape or even the overall fold can be determined by SAXS for particles of between 5 and 25 nm, but this requires a very homogeneous sample.

In absence of structural information of atomic resolution, models of protein structures can be constructed based on sequence homology to proteins with a known structure, as sequence homology usually means that the protein fold has been conserved throughout evolution. Homology modelling based protein structures can subsequently be fit into lower-resolution models from for instance EM or SAXS studies. Especially large protein complexes consisting of many subunits, are very difficult to study by X-ray crystallography, but much better suited to study by EM or SAXS. High-resolution structures of the subunits, obtained by crystallography or homology modelling, can subsequently be docked into the model of the complex (as for instance in paper II). Distance constraints on how the subunits connect within the complex form important complementary information, and they can be provided by for example crosslinking mass spectrometry experiments.

MS-based structural biology techniques include crosslinking mass spectrometry, but other strategies have also recently developed into high potential methods for structural studies of proteins (Stengel et al., 2012). As already briefly touched upon in the sections about the oligomeric structure and dynamics of small heat shock proteins, the mass spectrometry of intact protein assemblies, ‘nano-ESI MS’, or ‘native MS’, allows studying protein complexes intact in the mass spectrometer (Heck, 2008; Hernandez and Robinson, 2007). Adapted and optimized parameters and conditions in typically ESI-Q-TOF instruments can prevent the dissociation of non-covalently bound subunits and help to retain the native protein complex structure in the instrument. After initial mass measurements of the intact complex, controlled dissociation can be induced by collisional activation of the complex, to interrogate subunit connectivity, in analogy to the MSMS fragmentation of peptides. Different pathways of complex-

into-subunits dissociation reveal the topology of the complex. The possibilities to study intact protein complexes with native MS, are further extended by coupling ion mobility separation (IM) to MS. A particular strength of native MS is its use to heterogeneous systems; unlike NMR, EM, and SAXS, which measure average properties of all molecules, native MS can be exploited to characterize polydisperse ensembles such as  $\alpha$ B-crystallin (Baldwin et al., 2011b), as mentioned the previous chapter about small heat shock proteins.

In addition to crosslinking mass spectrometry and native MS, affinity purification (AP) and subsequent ‘classic’ mass spectrometric analysis of the purified components form a powerful tool to study protein complexes and protein-protein interactions (Stengel et al., 2012), even though topological information is very difficult to deduce. Another application of mass spectrometry to structural biology, is by measuring hydrogen/deuterium exchange patterns of peptides. Deuterated peptides indicate solvent exposed regions in the protein, because backbone hydrogen atoms involved in hydrogen bonding within the protein do not get deuterated. Figure 9 compares some of the structural biology techniques mentioned here with respect to resolution and sample requirements.



**Figure 9.** A rough comparison of different structural biology techniques in terms of obtained resolution and protein sample requirements. The y-axis represents the need for a heterogeneous sample. The grayscale of the bars with the different techniques gives an indication of how much protein is needed, with darkest gray for most protein needed. The figure was adapted from [http://en.wikipedia.org/wiki/Biological\\_small-angle\\_scattering](http://en.wikipedia.org/wiki/Biological_small-angle_scattering).

# 4 Crosslinking studies on Hsp21 (this work)

The biggest secret of the small heat shock protein chaperones is how they molecularly they protect their substrate proteins from aggregation. Despite our increased understanding of the general structural features, the polydispersity, and dynamic nature of sHsps, the exact mechanism of substrate recognition remains poorly understood. This is why the crosslinking mass spectrometry studies on the small heat shock protein Hsp21 described in this thesis were initiated, with the final aim of characterizing the interaction of Hsp21 with model substrate proteins. Crosslinking mass spectrometry as an emerging technique in structural biology still suffers from some technical issues, but the low amounts of sample required, high tolerance to heterogeneous samples, and the (potential) speed of analysis were all reasons to select this approach to investigate the Hsp21-substrate interaction.

In an initial trial, the crosslinking reagent DTSSP was used to crosslink Hsp21 and the model substrate protein CS. Crosslinks could be identified with a straightforward approach using MALDI-TOF MS without prior peptide sample separation by nano-LC (paper I). With detected crosslinks within Hsp21 and between Hsp21 and the model substrate protein, the need for a structural model of Hsp21 for the interpretation of crosslinks led to homology modelling and single particle negative stain EM studies on Hsp21 (papers I and II). The Hsp21-Hsp21 crosslinks identified in the first study (paper I) all fitted the proposed structure model of Hsp21 (paper II). To improve the yields and speed up the data analysis of the crosslinking mass spectrometry approach, dedicated software was developed to handle mass spectrometry data from nano-LC separated samples, and was further optimized for crosslinking with isotope-labelled BS<sup>3</sup> (paper IV). The reason for the preferred use of BS<sup>3</sup> over the disulfide-bond containing reagent DTSSP is that DTSSP crosslinks can potentially scramble after digestion, leading to false positives, depending on the presence of free cysteine residues in the protein sample. This problem was addressed in paper III, where both Hsp21 and the mammalian sHsp  $\alpha$ B-crystallin were crosslinked, in presence or absence of model substrate proteins. Especially MDH has many cysteine residues inducing scrambling, so the optimized crosslinking mass spectrometry workflow with the reagent BS<sup>3</sup> was applied to the interaction between Hsp21 and MDH (paper V). The results of papers I-V will be discussed in more detail in the following sections.

## 4.1 Oligomeric structure of Hsp21

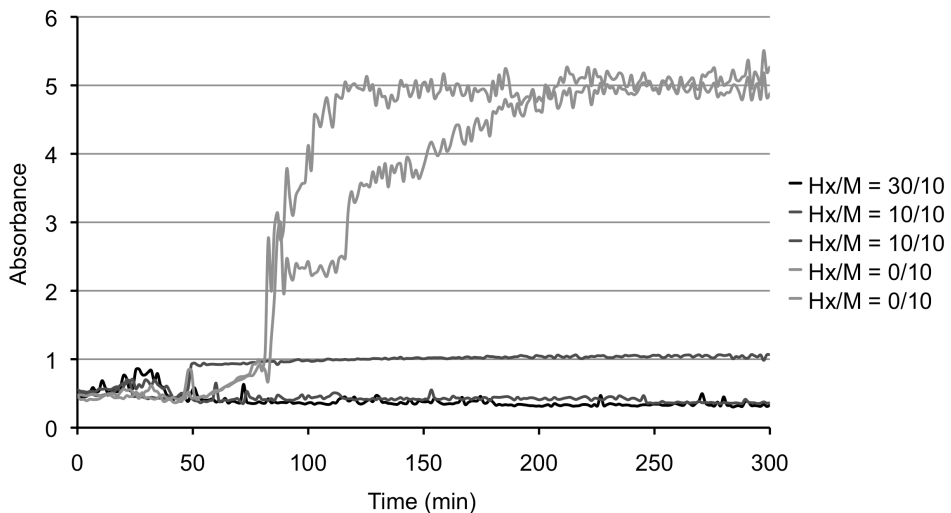
To characterize the structure of Hsp21, a combination of homology modelling, native MS, single particle negative stain EM, and crosslinking mass spectrometry resulted in a structure model of the Hsp21 dodecamer (papers I and II). Sequence homology analyses predict that all plant sHsps, including Hsp21 from *A. thaliana*, are built up of dimeric building blocks with architectures as in wheat Hsp16.9. Of the available sHsp crystal structures, wheat Hsp16.9 also has the highest sequence similarity to Hsp21. Therefore, an homology model of the Hsp21 dimer was constructed based on the wheat Hsp16.9 dimer (paper II, p. 293, figure 2).

The oligomeric state of Hsp21 was determined to be dodecameric by native MS (paper I, p. 1466, figure 1). Crosslinking Hsp21 by DTSSP also captures a dodecamer, as revealed by SDS-PAGE (paper I, p. 1467, figures 3 and 4). By negative stain EM, single particles of both native and crosslinked Hsp21 could be visualized, which matched a dodecameric state in size (paper II, p. 294, figure 3). Because the samples for both native MS and negative stain EM were prepared at room temperature, it cannot be excluded that higher oligomeric species form at increased temperatures. Wheat Hsp16.9, which was crystallized in its native oligomeric state (Van Montfort et al., 2001b), is dodecameric, and so is another plant sHsp, Hsp18.1 from pea. However, Hsp18.1 was shown by native MS to change from a homogeneous dodecamer to heterogeneous higher oligomers upon raising the temperature (Stengel et al., 2010). Hsp21 samples crosslinked at 25°C and 47°C were analyzed by SDS-PAGE, in absence and in presence of substrate (paper I, p. 1467, figures 3 and 4). In some lanes, extra material appears to not have entered the gel for the 47°C samples, which can be interpreted as higher oligomers. Interestingly, the presence of substrate at 47°C seems to have the largest effect on the oligomeric state of Hsp21. The band corresponding to a dodecamer disappears under these conditions (paper I, p. 1467, figure 4).

The structure of oligomeric Hsp21 was further characterized by single particle averaging of negative stain EM images of Hsp21 particles, resulting in an image reconstruction of the dodecamer (paper II, p. 295, figure 4). The dimeric subunits are organized in rings of three dimers, two of which are stacked to form the dodecamer. Despite the low resolution, an atomic model based on homology modelling to Hsp16.9 was docked into the density map, which required rotating the two rings with respect to each other (paper II, p. 296, figure 6). Even though the particles selected for the image reconstruction were of homogeneous size, the rotated conformation could also indicate some degree of heterogeneity, considering the current understanding of sHsp dynamics. However, the resolution limits further conclusions on the atomic structure of the dodecamer. Yet, the current structure model of Hsp21 has been very helpful for the interpretation of crosslinking mass spectrometry data (see below).

## 4.2 The crosslinked Hsp21 dodecamer

Three-dimensional images based on single particle negative stain EM were constructed for both the native and the crosslinked dodecamer, revealing a much more compact structure of the crosslinked dodecamer than the native conformation (paper II, figure 5, p. 296). The image reconstruction of crosslinked Hsp21 has extra density on the inside of the two rings, which probably represents crosslinked N-terminal regions. The image reconstruction of the expanded native Hsp21 even lacks connectivity for one end of each of the six dimers, as if every other monomer is ‘protruding out’, so the atomic model of the Hsp21 dodecamer was docked into the crosslinked dodecamer. The protrusions of the native conformation may also reflect some degree of heterogeneity of the particles, considering the reported dynamic subunit exchange of sHsps (section 2.3). Although the crosslinked dodecamer cannot dissociate (as shown by SDS-PAGE in paper I, figure 3, p. 1467), it can still suppress thermo-induced aggregation of MDH, when assayed by light-scattering measurements of MDH aggregation (figure 10). Because very few crosslinks are required for keeping together the dodecamer, at least some of the 12 flexible N-terminal regions and the 12 flexible C-terminal extensions, as well as other sites suggested to potentially be substrate-binding (Van Montfort et al., 2001b), could still be available for partially unfolded substrate.



**Figure 10.** Crosslinked Hsp21 protects MDH from aggregation. Light-scattering assay analogous to figure 1 in paper V (p. 3). The numbers in the legend refer to  $\mu\text{M}$  concentrations of crosslinked Hsp21 (Hx) and MDH (M). At  $45^\circ\text{C}$ , MDH alone (0/10 – light gray) starts aggregating after 1 hour. At molar ratios of Hx/M of 1 and higher (10/10 – dark gray and 30/10 – black), MDH aggregation is almost fully suppressed.



### 4.3 Crosslinks within Hsp21

Crosslinking experiments on Hsp21, where crosslinks were identified on the peptide level by MALDI-TOF/TOF mass spectrometry, are central to this thesis. Therefore, it seems appropriate to give an overview of all crosslinks identified within Hsp21 in papers I, III, IV, and V (table 1). This table actually reflects the problem of reproducibility with crosslinking mass spectrometry as a technique. Out of the 35 listed Hsp21-Hsp21 crosslinks only six were identified in all five cases. Replicates crosslinked and processed with the same protocol for nano-LC and MALDI-TOF/TOF, do not result in the same lists of identified crosslinks, as illustrated by the variation in the number of observations shown in table 2 of paper V (p. 6). The values shown in this particular column refer to the number of observations out of 17 similar datasets in total, of which eight datasets are reproductions of basically the same original sample.

The crosslinks reported in papers III and IV were all confirmed by MSMS fragmentation spectra. The crosslinks identified in papers I and V based on MS data only, were subjected to more selective criteria. In paper I, where crosslinking mass spectrometry was first implemented, the lack of high-quality MSMS data can be attributed to the fact that the samples were not separated by nano-LC prior to mass spectrometric analysis. This results in a very high sample complexity per acquired MS spectrum, which causes low peak intensities due to ion suppression, and which makes it difficult to exclusively select one precursor ion for MSMS fragmentation. In paper V, high-quality MSMS spectra could only be obtained for some of the Hsp21-Hsp21 crosslinks (data not shown). Because all samples in paper V were separated by nano-LC, ion suppression was much less of a problem here, compared to in paper I. However, still no good MSMS data could be obtained for all Hsp21-Hsp21 crosslinks in the samples containing both Hsp21 and MDH, because the peak intensity in the MS spectra of many of the crosslinks was very low, despite the efforts of optimizing both nano-LC separation and the data analysis software, as described in paper IV.

Both the low reproducibility and the difficulties of obtaining high-quality MSMS data for Hsp21-Hsp21 crosslinks in presence of another protein indicate that the abundance of many of the crosslinks is low compared to the limit of detection of the crosslinking mass spectrometry workflow used. So, the absence of peaks matching crosslinks in the MS spectra does not necessarily mean that this crosslink is absent in the sample, but that it could simply not be detected. Actually, when a certain crosslink was detected in for example 4 samples out of 17 by automatic data analysis, it was often the case that the corresponding peak was also observed in other samples by manual inspection of the MS-spectra, but of too low intensity to pass the filters set for peak picking.

**Table 1.** Hsp21-Hsp21 crosslinks identified in papers I, III, IV, and V.

Mass crosslink (Da)		Matched sequence		Residue		Paper				
BS <sup>3</sup>	DTSSP	1	2	1	2	<sup>a</sup> I	III	IV	<sup>b</sup> V-	V+
1491.668	1527.581	M1-R5	M1-R5	1	1	1	1	1	1	1
2385.169	2421.082	M1-R5	G19-R32	1	27	1	1	1	0	1
2438.166	2474.079	M1-R5	A84-K96	1	89	1	1	1	1	1
2014.933	2050.845	M1-R5	E90-R98	1	96	0	1	1	0	0
2180.037	2215.950	M1-R5	F99-K110	1	106	0	1	1	1	1
2485.297	2521.210	M1-R5	I111-K125	1	121	1	1	1	0	0
2095.899	2131.812	M1-R5	K126-R136	1	126	0	1	1	1	1
1758.942	1794.855	M1-R5	D154-K161	1	155	0	0	0	0	1
1515.821	1551.733	M1-R5	I156-K161	1	157	1	1	1	1	1
2357.290	2393.203	M1-R5	A158-K171	1	161	0	1	0	0	0
1446.738	1482.650	M1-R5	T172-R176	1	173	1	1	1	1	1
1574.832	1610.745	M1-R5	T172-K177	1	173	0	0	0	0	1
1756.927	1792.840	M1-R5	K177-Q184	1	177	1	1	1	1	1
2141.140	2177.053	M1-R5	V174-Q184	1	177	0	0	1	0	0
3128.582	3164.495	E6-K27	T172-R176	18	173	0	1	1	1	1
2297.196	2333.109	G19-R32	G122-K126	27	125	0	1	0	0	1
2211.185	2247.098	G19-R32	D154-K157	27	155	0	0	0	0	1
2340.238	2376.151	G19-R32	T172-R176	27	173	0	1	0	0	1
3042.397	3078.310	A84-K96	K126-R136	89	126	0	1	1	0	0
3591.819	3627.732	A84-R98	L145-K157	89	153	0	1	0	0	0
2393.235	2429.148	A84-K96	T172-R176	89	173	0	0	1	1	1
2619.160	2655.073	E90-R98	K126-R136	96	126	0	0	1	0	0
2880.554	2916.467	E90-R98	A158-K171	96	161	0	1	0	0	0
1970.002	2005.915	E90-R98	T172-R176	96	173	0	0	0	0	1
2098.097	2134.010	E90-R98	T172-K177	96	173	0	0	0	0	1
2092.060	2127.973	F99-K110	G122-K126	106	125	0	0	1	0	0
2833.543	2869.455	F99-K110	N162-K173	106	171	0	0	0	0	1
2440.366	2476.279	I111-K125	T172-R176	121	173	0	1	1	1	1
2379.205	2415.118	G122-K126	M97-K110	125	106	0	0	1	0	1
2269.320	2305.233	G122-K126	A158-K171	125	161	0	0	1	0	0
1668.954	1704.867	G122-K126	K177-Q184	125	177	0	0	0	1	1
2050.968	2086.881	K126-R136	T172-R176	126	173	0	1	0	0	0
1272.753	1308.666	D154-K157	T172-R176	155	173	0	1	1	1	1
1400.848	1436.761	D154-K157	T172-K177	155	173	0	0	0	0	1
1470.890	1506.802	I156-K161	T172-R176	157	173	1	1	1	1	1

<sup>a</sup> Crosslinks identified in paper I were only included here if also identified elsewhere, because of the quite different workflow in paper I. <sup>b</sup> Crosslinks identified in paper V were categorized into crosslinks found in absence (-) and in presence (+) of the model substrate protein MDH.

Even though an undetected crosslink may not always mean the absence of this crosslink in the sample, all crosslinks that were detected and convincingly identified certainly indicate the presence of those crosslinks in the sample. Accordingly, all Hsp21-Hsp21 crosslinks identified match the proposed dodecameric structure model, at least the part that was modelled, *i.e.* the  $\alpha$ -crystallin domain and the C-terminal extension. Based on the crosslinker length, the allowed distance between two C $\alpha$  atoms of lysine residues involved in a crosslink is around 20 Å, and this was found to be the case for the lysine pairs of the identified crosslinks. This is illustrated in paper II with the Hsp21-Hsp21 crosslinks in table I (p. 297), which were mapped into the structure in figure 7 (p. 298). In paper III, the crosslinks in table II (p.1687) were also shown to fit the structure (figure 4, p. 1688), as well as in paper IV, where the crosslinks in table 1 (p. 4) were mapped into the structure in figure 4 (p. 6). Some Hsp21-Hsp21 crosslinks in paper V had not previously been detected, but for those that involve lysine residues in the  $\alpha$ -crystallin domain and the C-terminal extension, the distance of the lysine residues in the structure was also within the length of the crosslinker used (not shown).

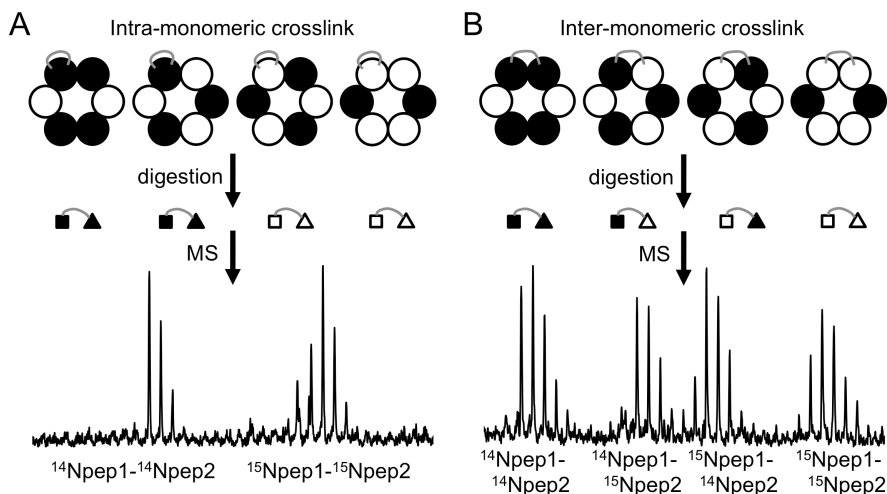
Many of the identified Hsp21-Hsp21 crosslinks involve residues 1, 18, or 27, which are in the N-terminal region of Hsp21. Because of the predicted disorder and the low sequence homology to the wheat Hsp16.9 template of the N-terminal region, this region was not included in the structure model. Therefore, the crosslinks involving the N-terminal region could not be mapped into the structure. Since the N-terminal region is very flexible and most probably intrinsically disordered (Uversky et al., 2009), it can be imagined to reach just about any lysine residue of Hsp21, on the same monomer, or on another monomer in the dodecamer. Whether an identified Hsp21-Hsp21 crosslink is a crosslink within the same monomer or between monomers, cannot be concluded from the data, unless, different monomers carry different isotope labels. To facilitate the interpretation of Hsp21-Hsp21 crosslinks, some experiments were conducted with <sup>15</sup>N-isotope-labelled Hsp21, as described in the next section.

## 4.4 Intra- and inter-monomeric Hsp21 crosslinks

Even if a mass in the MS spectrum unambiguously matches a crosslink involving two particular peptides, it is not always possible to know exactly which lysine residues in the structure of Hsp21 are the ones crosslinked. The reason for this is that Hsp21 is a homooligomer of 12 monomers, so a crosslink can either involve two lysine residues in the same monomer, or in two different monomers. Most of the Hsp21-Hsp21 crosslinks involving the  $\alpha$ -crystallin domain and the C-terminal extension, that were directly mapped into the structure, could be mapped within the Hsp21 dimeric building block, and most of the times it was obvious whether

this was within one monomer or between the two monomers of the dimer. However, for crosslinks involving residues 1, 18, or 27 of the N-terminal region, it was not at all obvious whether these would be crosslinks within or between monomers, so a crosslinking experiment with a mixture of unlabelled and  $^{15}\text{N}$ -isotope-labelled Hsp21 was conducted to distinguish the two.

Because of the dynamic subunit exchange of sHsp oligomers, heterooligomers containing both unlabelled and  $^{15}\text{N}$ -isotope-labelled Hsp21 monomers form after some time upon mixing unlabelled and  $^{15}\text{N}$ -isotope-labelled Hsp21 preparations. After one hour at room temperature, unlabelled and  $^{15}\text{N}$ -isotope-labelled Hsp21 monomers had not completely mixed yet (based on the  $^{14}\text{N}/^{15}\text{N}$  isotope pattern described below; data not shown), but one hour at  $37^\circ\text{C}$  was enough to complete heterooligomer formation. When adding the crosslinking reagent to completely mixed heterooligomers, crosslinks will form within and between unlabelled and  $^{15}\text{N}$ -isotope-labelled Hsp21 monomers. Now, a hypothetical crosslink between peptide 1 (pep1) and peptide 2 (pep2), can be determined to be intra- or inter-monomeric, because it will give rise to either 2 or 4 mass peaks, respectively (figure 11).



**Figure 11.** Heterooligomers of  $^{14}\text{N}$ - and  $^{15}\text{N}$ -labelled Hsp21 to distinguish intra- and inter-monomeric crosslinks. For clarity, Hsp21 oligomers are represented by 6 subunits. Only a selection of possible heterooligomers is shown. A: An intra-monomeric crosslink between peptide 1 (pep1, squares) and peptide 2 (pep2, triangles) will yield 2 mass peaks:  $^{14}\text{N}$ -pep1- $^{14}\text{N}$ -pep2 and  $^{15}\text{N}$ -pep1- $^{15}\text{N}$ -pep2. B: An inter-monomeric crosslink between these peptides will yield 4 mass peaks:  $^{14}\text{N}$ -pep1- $^{14}\text{N}$ -pep2,  $^{14}\text{N}$ -pep1- $^{15}\text{N}$ -pep2,  $^{15}\text{N}$ -pep1- $^{14}\text{N}$ -pep2, and the  $^{15}\text{N}$ -pep1- $^{15}\text{N}$ -pep2 crosslink.

Examples of the MS spectra of an intra- and inter-monomeric Hsp21-Hsp21 crosslink are given in paper IV (figure 3, p. 5). In this way, it was possible to determine for most of the identified Hsp21-Hsp21 crosslinks in paper IV whether they are intra- or inter-monomeric (paper IV, table 1, p. 4, last column). Most of the Hsp21-Hsp21 crosslinks involving the  $\alpha$ -crystallin domain and the C-terminal extension were intra-monomeric. A crosslink that connects two monomers of the dimeric building block was inter-monomeric, according to its  $^{14}\text{N}/^{15}\text{N}$  isotope pattern. A frequently observed crosslink between lysine residues 157 and 173 (mass 1470.890 Da with BS<sup>3</sup> and 1506.802 Da with DTSSP), was in paper IV shown to be intra-monomeric by its  $^{14}\text{N}/^{15}\text{N}$  isotope pattern, even though it was previously interpreted as an inter-monomeric crosslink (paper I, figure 7, p. 1473).

Most of the Hsp21-Hsp21 crosslinks involving the N-terminal region were inter-monomeric, based on their  $^{14}\text{N}/^{15}\text{N}$  isotope patterns. It seems plausible that the highly flexible N-terminal arm is just as likely to crosslink to a residue on the same monomer as on a different monomer, also because all lysine residues other than 1, 18, and 27 are far away in primary sequence. There are also inter-monomeric crosslinks between two N-terminal regions, which fits well to the observed extra density in the EM image reconstruction of the crosslinked Hsp21 dodecamer.

## 4.5 Crosslinks within $\alpha$ B-crystallin

In paper III, crosslinks within another sHsp,  $\alpha$ B-crystallin, were identified and mapped into the structure (table II, figure 4). The available crystal structures of  $\alpha$ B-crystallin do not include atomic coordinates for the N-terminal region and the C-terminal extension, so only crosslinks involving residues in the  $\alpha$ -crystallin domain could be mapped. The distance between the lysine pairs of the identified crosslinks that were mapped was within 20 Å.

When using the crosslinking reagent DTSSP,  $\alpha$ B-crystallin has the advantage that it does not contain any cysteine residues. DTSSP contains a disulfide bond, and the presence of free cysteine residues can induce scrambling of crosslinks after digestion, as described in paper III. For  $\alpha$ B-crystallin no scrambling of identified crosslinks was seen in the isotope patterns, whereas for Hsp21 with one cysteine residue, some of the identified crosslinks could in part or completely be ascribed to disulfide bond scrambling. The proportion of false positive crosslinks caused by disulfide bond scrambling was highest for samples including MDH, which has 8 cysteine residues. Because MDH was selected as model substrate to investigate the Hsp21-substrate interaction in paper V, the crosslinking reagent BS<sup>3</sup> (without a disulfide bond) was used in papers IV and V.

## 4.6 Hsp21 chaperone function

The chaperone activity of Hsp21 towards the model substrate proteins CS and MDH was investigated in papers I and V. In absence of Hsp21, both CS and MDH aggregated when incubated at 47°C for 20 minutes, and after centrifugation the aggregated protein was recovered in the pellet (paper I, figure 2, p. 1466). In presence of 12 times molar excess Hsp21, no protein was found in the pellets, so Hsp21 had protected CS and MDH from aggregation. The Hsp21 chaperone activity towards MDH was also determined by light-scattering in paper V (figure 1, p. 3), and similar data were recorded for Hsp21 chaperone activity towards CS (not shown).

For relatively short incubation times at aggregation-inducing temperatures, the interaction between Hsp21 and CS is reversible (paper V, figure 2, p. 3). This supports a model of sHsp chaperone based on transient interactions that take place as a result of subtle and initial changes in the conformation of CS. Even under conditions where CS would not aggregate (at 25°C), there appears to be an interaction between Hsp21 and CS, because the presence of equimolar amounts of CS changed the appearance Hsp21 on SDS-PAGE (paper I, figure 4, p. 1467). In presence of equimolar amounts of CS at 47°C, the Hsp21 oligomers disappeared completely, and instead, the crosslinker captured larger complexes (presumably consisting of Hsp21 and CS), which do not enter the gel.

The crosslinking mass spectrometry workflow based on crosslinking with BS<sup>3</sup>, nano-LC MALDI-TOF/TOF mass spectrometry, and data analysis with the program FINDX, as described in papers III and IV, was developed to study the transient interactions between Hsp21 and model substrate proteins. In paper V, Hsp21 was crosslinked to MDH under conditions where Hsp21 protects MDH from aggregation. Paper I describes how Hsp21 was crosslinked to CS in the initial trials to explore the possibilities of crosslinking mass spectrometry to probe the Hsp21-substrate interaction. The identified crosslinks between Hsp21 and the model substrate proteins in papers I and V will be discussed in more detail below.

## 4.7 Crosslinks between Hsp21 and model substrate proteins

In paper V, Hsp21 was crosslinked to a model substrate protein, and analyzed with a workflow that was in several aspects improved compared to the initial trials described in paper I. The Hsp21-substrate crosslinks suggested in paper I will therefore not be discussed as much as those identified in paper V. Crosslinking Hsp21 and CS with DTSSP (paper I), as well as crosslinking Hsp21 and MDH with BS<sup>3</sup> (paper V), resulted in higher molecular weight bands on SDS-PAGE (paper I, figure 4, p. 1467 and paper V, figure 4, p. 5). The conditions under which

Hsp21 was crosslinked to the model substrate are the same: Hsp21 and the model substrate were incubated at a temperature that induces aggregation of the model substrate in absence of Hsp21, and for only a short time during which the interaction between them is still reversible.

In paper I, the first attempts were made to apply crosslinking mass spectrometry to explore Hsp21-substrate interactions. Here, non-isotope-labelled DTSSP was used, excluding to possibility to monitor potential disulfide scrambling (as described in paper III). In addition, the complex tryptic digests of crosslinked samples were analyzed without prior fractionation and only MS data were acquired. The data were analyzed with GPMW (Peri et al., 2001). Even though all peaks present in mass spectra of samples without DTSSP and with DTT were subtracted from the peak lists (paper I, figure 5, p. 1468), not all possible matches to experimental masses were taken into consideration when the assigning crosslinks to experimental masses.

The factors mentioned above explain why some of the suggested crosslinks in paper I (named ‘dipeptides’ in paper I) are most likely false positives. However, many of the Hsp21-Hsp21 crosslinks in paper I (table 2) are valid and were also confirmed in the other papers, as shown in table 1 in section 4.3. Concerning Hsp21-CS crosslinks, it is difficult to say which and how many of these represent true crosslinks that were formed during the interaction between Hsp21 and CS. Regardless of this, a general interpretation of these data is that both the interacting residues of Hsp21 and CS are quite much spread out over the both molecules (paper I, table 3, p. 1471 and figure 8, p. 1474). Taking into account previously reported data on the Hsp21-CS interaction by peptide array screening, which indicated preferential binding of Hsp21 to the thermo-sensitive stalk of CS (Ahrman et al., 2007a), it is interesting to note that this part of CS is exclusively involved in crosslinks to the N-terminal residue M1 of Hsp21 (paper I, figure 8, p. 1474).

In paper V, 10 crosslinks between Hsp21 and the model substrate MDH were identified (table 1, p. 6). These crosslinks involve only 3 residues of Hsp21: M1, K27, and K173. Two of these, M1 and K173 are also overrepresented in Hsp21-Hsp21 crosslinks, and as discussed in more detail in paper V, crosslinks involving these residues may rather reflect favourable properties of these residues (amine reactivity) and the peptides containing them (compatibility with MS detection by MALDI), than preferential crosslinking of MDH to these residues. On the contrary, K27 is not particularly abundant in Hsp21-Hsp21 crosslinks, so its presence in the relatively few Hsp21-MDH crosslinks identified may well indicate preferential interaction of MDH to the N-terminal region of Hsp21.

The abundance of the Hsp21 residues involved in crosslinks is extensively analyzed in paper V, and this analysis also considers the Hsp21-Hsp21 crosslinks identified in papers III and IV (paper V, figure 6, p. 8). From this analysis, it is

clear that the different residues involved in crosslinks display a general pattern of how often they are observed. This pattern is presumably caused by two technical factors, namely: the reactivity of the residue in the crosslinking reaction, and the compatibility of the peptide containing the residue with MALDI-TOF mass spectrometric detection. These factors were also considered in the interpretation of the crosslinking mass spectrometry data for the Hsp21-substrate interaction.

As discussed in paper V, crosslinks involving Hsp21 residue K27 probably reflect an interaction of the N-terminal region of Hsp21 to partially unfolding MDH. This finding is in agreement with other studies reporting on the importance of the N-terminal region of sHsps for chaperone activity (Giese et al., 2005) and also substrate binding (Jaya et al., 2009; Sharma et al., 2000). The crosslinks of the thermosensitive part of CS to the Hsp21 N-terminus (paper I, table 3, p. 1471 and figure 8, p. 1474) also support that the N-terminal region is substrate binding, even though residue M1 is a particularly reactive residue in general (paper V, figure 6, p. 8). The N-terminal region of Hsp21 appears to be an intrinsically disordered region (IDR) (Uversky et al., 2009), since it is extremely quickly degraded by limited proteolysis (unpublished data). The disorder, and the large number of conserved methionine residues that could confer promiscuous binding sites as in SRP54 and calmodulin (Snijder et al., 2011), would make the N-terminal region of Hsp21 suitable to accommodate all sorts of different unfolding substrate proteins *in vivo*.

## 4.8 Crosslinking mass spectrometry to probe protein-protein interactions

In this work, crosslinking mass spectrometry was used to study Hsp21 and the transient interaction between Hsp21 and substrate proteins. Throughout the papers in this thesis, the crosslinking mass spectrometry workflow was developed to finally apply it to the Hsp21-substrate system in its optimized form (paper V). Still, the yields of identified crosslinks are very low, especially concerning crosslinks between Hsp21 and the substrate protein (paper V, table 1, p. 6). As mentioned before in section 4.3, this is most likely not due to the lack of crosslinks in the sample, but rather to the low sensitivity of the method developed thus far, as also recognized by others working in this field (Leitner et al., 2010; Singh et al., 2010).

The mass spectrometric analyses of the tryptic digests of crosslinked samples discussed in this thesis, as well as those not discussed, clearly show a pattern in the abundance of the different types of analytes in the digests. Unmodified peptides are the most abundant, closely followed by peptides with a dead-end crosslink modification (type 0). After this, intra-peptide crosslinks (type 1) follow in their abundance in the sample. Thereafter, inter-peptide crosslinks (type 2) within the



same monomer are still fairly easy to detect. Following these, inter-peptide crosslinks (type 2) between subunits of a stable complex can be detected (such as in the example shown in paper V, figure 3, p. 4). Finally, inter-peptide crosslinks (type 2) between two proteins transiently interacting are by far the least abundant. The difficulty detecting these can partly be attributed to their low number of copies in the sample. However, because they can be observed with low intensity and low reproducibility, it can be concluded that the main reasons for the difficulties in detecting them in all samples, should be sought in the detection method.

The most promising development in the field of crosslinking mass spectrometry, are methods to enrich type 2 crosslinks. Recently, cation ion exchange chromatography of the tryptic digests of crosslinked samples has been shown to improve the yields of type 2 crosslinks (Fritzsche et al., 2012; Leitner et al., 2012). Other separation techniques, such as size exclusion chromatography, may also be helpful to enrich the relatively large type 2 crosslinks. In addition, the development of specialized crosslinking reagents with incorporated affinity tags for enrichment is an intensive area of research (Petrotchenko and Borchers, 2010). Efficient enrichment of type 2 crosslinks may also improve the yields of crosslinks for the Hsp21-substrate interaction investigated here.

Regardless of that general enrichment of type 2 crosslinks can improve the overall yields, it is still a problem for the interpretation of identified crosslinks that the different residues involved in the crosslinks, cannot be compared in an unbiased way. As mentioned before, the two main reasons for this are the differences in the reactivity of crosslinking residues, and the differences in the compatibility with MS detection of the tryptic peptides containing the crosslinked residues. Basically, this problem is inherent to mass spectrometry, and has led to more quantitative methods in the field of mass spectrometry based proteomics (Domon and Aebersold, 2006). In crosslinking mass spectrometry, the availability of isotope-labelled crosslinking reagents already sets the stage for the more general implementation of quantitative methods. Additionally, for simple systems involving a limited number of proteins, the proteins themselves could be isotope-labelled.

Concerning the Hsp21-substrate interaction under investigation, crosslinking mass spectrometry strategies are promising to extend our current understanding of this system. When the yields of type 2 crosslinks of interest can be improved, reproducibility should also be improved and quantitative methods can be used to design experiments comparing the Hsp21-substrate interaction under different conditions, such as temperature.

## 5 Concluding remarks

The motivation for initiating the work described in this thesis, is the limited understanding of how an important class of chaperones, the small heat shock proteins, can prevent other proteins from aggregation. Their presence in all organisms, and their predominance among proteins upregulated in the stress response, clearly mark their importance in protecting the cell from potentially harmful aggregates under stress conditions. Despite their presumably essential role in maintaining proteostasis, very little is known about the molecular details of the interaction between sHsps and substrate, or ‘client’, proteins.

The system in focus in this thesis is the interaction between the sHsp Hsp21 and two model substrate proteins. Hsp21 from *A. thaliana* is important for plant stress resistance, and its distinguishing features compared to other sHsps include an unusually large N-terminal region containing several highly conserved methionine residues. The interactions between Hsp21 and the thermo-sensitive model substrate proteins CS and MDH were explored by crosslinking mass spectrometry. The quest for the ultimate crosslinking mass spectrometry workflow to study these interactions, and the actual investigations concerning these interactions, led to a number of conclusions, which are briefly summarized here.

It was found that Hsp21 is dodecameric, and that it can protect CS from aggregation. To characterize the Hsp21-CS interaction, the advantages of the crosslinking reagent DTSSP were exploited to identify Hsp21-Hsp21 and Hsp21-CS crosslinks with a straight-forward approach using MALDI-TOF MS without prior sample fractionation (paper I).

A structure model of Hsp21 was generated by combining homology modelling and single particle negative stain EM. Image reconstructions of native and crosslinked Hsp21 based on single particle averaging revealed a stack of two hexameric rings, which are slightly rotated with respect to each other. The structure model of Hsp21 was used to verify that the identified Hsp21-Hsp21 crosslinks are mappable into the structure (paper II).

By using isotope-labelled DTSSP and inspecting the isotope-pattern for a diagnostic ‘scrambling’ peak, it was possible to assess whether an identified crosslink is a true crosslink reflecting a true interaction from before the sample was digested, or whether it is purely the result of disulfide bond scrambling. The

data analysis program FINDX for crosslinking mass spectrometry data was developed to process data from DTSSP crosslinked samples (paper III).

The crosslinking mass spectrometry workflow and the data analysis program FINDX were further optimized to handle data from more than one crosslinked protein, and with the crosslinking reagent BS<sup>3</sup>, which is similar to DTSSP, but without the thiol-cleavable disulfide bond. The LC MALDI-TOF/TOF platform, in which LC, MALDI-TOF MS, and MALDI-TOF/TOF MSMS are decoupled, allows quick detection of candidate crosslinks, which limits the amount of MSMS-spectra that needs to be acquired for the validation of the crosslinks (paper IV).

The transient Hsp21-substrate interactions that protect model substrate proteins from temperature-induced aggregation were monitored by light-scattering and investigated by an optimized workflow with BS<sup>3</sup>, MALDI-TOF/TOF mass spectrometry and the data analysis program FINDX. The identified crosslinks point at an interaction between the N-terminal region of Hsp21 and the C-terminal part of the substrate protein MDH (paper V).

In conclusion, using the optimized crosslinking mass spectrometry workflow, crosslinks within the sHsp Hsp21 were extensively characterized, and despite the low yields of crosslinks between transiently interacting proteins, some crosslinks between Hsp21 and MDH were also identified. The identified Hsp21-substrate crosslinks indicate that the flexible N-terminal region of Hsp21 is responsible for substrate interaction.

The crosslinking mass spectrometry workflow developed throughout this thesis has provided new insights into the sHsp-substrate interaction, and into the Hsp21-model substrate specifically. The final conclusion, that the N-terminal region of Hsp21 is substrate-binding, can be reconciled with previous reports on the sHsp-substrate interaction, even though these are few and obtained with different techniques. The currently intensive developments within the field of crosslinking mass spectrometry will allow refinement of the results reported here. Specifically, it remains to be confirmed for Hsp21 whether the N-terminal region is exclusively responsible for the interaction with substrates. In addition, crosslinking mass spectrometry will be useful to examine the interactions of other types of sHsps with several different substrate proteins, because the mechanism of chaperone activity may not be general for all sHsps.

## 6 Future perspectives

This thesis reports on the extensive development of lysine-specific crosslinking and LC-MALDI-TOF/TOF mass spectrometry, to investigate interactions within small heat shock proteins, and ultimately the interactions between a small heat shock protein and model substrate proteins. Clearly, the Hsp21-substrate crosslinks described in paper V form just a glimpse of the interactions that can potentially be detected by crosslinking mass spectrometry, especially considering the promising developments in this field. Currently, the main limitation of crosslinking mass spectrometry as such is the low yield of informative type 2 crosslinks. Especially crosslinks between transiently interacting proteins are difficult to detect and would greatly benefit from the enrichment of type 2 crosslinks. With higher yields of type 2 crosslinks, the reproducibility between different experiments should also be improved. High yields and reproducibility would enable a variety of in-depth functional studies on Hsp21 and other small heat shock proteins.

The crosslinking mass spectrometry data so far support the view that the N-terminal region of Hsp21 interacts with an unfolding substrate protein, as has previously been suggested, without excluding the possibility of other interactions than those detected. Because the N-terminal region of Hsp21 only contains two lysine residues, three arginine-to-lysine mutations have been introduced to allow the detection of more different crosslinks involving the N-terminal domain. It is also important to extend crosslinking mass spectrometry studies of Hsp21 to interactions with substrate proteins other than MDH and CS, to explore and possibly understand the mechanism of action in more general terms.

To further investigate the sHsp-substrate interaction by crosslinking mass spectrometry, quantitative methods would allow comparative studies between different conditions, such as temperature and incubation time differences. Isotope-labelled crosslinking reagents are already available and could be used in a quantitative experiment, or recombinantly expressed protein can be isotope-labelled. With sufficient yields and reproducibility, the effect of two different temperatures for example could be compared within the same dataset. Importantly, this requires further development of software for the quantitative analysis of crosslinking mass spectrometry data.

So far, few studies have assessed the *in vivo* chaperone activity of sHsps. Very little is known about endogenous sHsp substrate proteins and the interaction of

sHsps with other components of the protein quality control systems. Large sHsp-substrate complexes presumably act as a storage place for partially unfolded proteins, which thereafter need to be refolded or degraded. This requires interactions with both the refolding and degradation machineries of the cell, and particularly the latter of these is an area that has received little attention. The *in vivo* sHsp-interactions with other proteins, including endogenous substrates, other chaperones, and protease components, could be studied using crosslinking mass spectrometry, as this method is becoming more robust. However, the experiences from the work described in this thesis emphasize the advantage of samples of limited complexity, so the analysis of cell lysates by crosslinking mass spectrometry to pinpoint interactions on the level of amino acid residues is most likely very challenging. A more classic approach to identify endogenous protein interaction partners is to use crosslinking in combination with affinity purification of the sHsp, and subsequent mass spectrometric identification of crosslinked proteins, but because the transient sHsp-substrate interaction may be very difficult to capture *in vivo* in this way.

As mentioned before, crosslinking mass spectrometry is especially suited as a technique complementary to other methods. Concerning the sHsp-substrate interaction, other mass spectrometry based methods such as native MS and H/D exchange seem attractive to study this dynamic and transiently interacting system. Another possibility to probe the sHsp-substrate interaction is to use NMR. This requires isotope-labelling of the protein, which for Hsp21 has already been accomplished. A combination of several techniques is most likely required to scrutinize how the small heat shock protein chaperones interact with and protect a wide range of different unfolding proteins under stress conditions in the cell.

# About this thesis

## 6.1 Popular scientific summary in English

The building blocks of all life on our planet are cells. Our own body, for instance, approximately consists of 10 trillion ( $10^{13}$ ) cells. Most bacteria consist of only one cell. What is inside the cells are biomolecules that govern life, such as proteins, sugars, lipids, and DNA. Cells are extremely packed, especially concerning the protein content, with protein concentrations approaching 300 mg/ml.

Proteins are linear molecules made up of different amino acids. For proteins to remain functional, this long amino acid chain needs to stay folded in its native three-dimensional fold. Different stress factors, such as heat, can cause proteins to partially unfold. Because it is so crowded in the cell, the risk of partially unfolded proteins is that they stick together and start to aggregate. This is also what happens when applying heat stress to an egg (with a protein concentration of about 100 mg/ml in the egg white): the proteins aggregate and the egg solidifies when frying or cooking it. In the cell, protein aggregation can disturb vital processes, and many different diseases are the result of too much protein aggregation (Alzheimer's and Parkinson's, and cataract for example).

Fortunately, cells are equipped with a special set of proteins called chaperones, whose task is to prevent other proteins from aggregating. However, it is still not well understood, on a molecular level, how these chaperones perform this task. Especially the small heat shock proteins, a subgroup of chaperones, have not been studied as much as some other chaperones.

The main topic of this thesis is how the small heat shock protein chaperone Hsp21 protects substrate proteins from aggregation. This question was addressed by combining crosslinking with mass spectrometry. A crosslinking reagent is a chemical that connects two amino acids. After cutting a crosslinked Hsp21-substrate mixture into small pieces, mass spectrometry can be used to detect the pieces that are connected by the crosslinker. The identification of these pieces reveals which amino acids of Hsp21 and the substrate protein were interacting with each other, when Hsp21 was protecting the substrate protein from aggregation. Experiments like these increase our understanding of how potentially harmful protein aggregation in the crowded environment of the cell is kept under control by chaperones.

## 6.2 Populärvetenskaplig sammanfattning på svenska

Byggstenarna av allt liv på denna planet är celler. Vår egen kropp till exempel består av ungefär 10 biljon ( $10^{13}$ ) celler. De flesta bakterier består av bara en cell. Inne i cellerna finns alla de biomolekyler som styr livet, såsom proteiner, sockerarter, lipider, och DNA. Celler är otroligt fullpackade, särskilt vad gäller proteiner, med protein koncentrationer av up till 300 mg/ml.

Proteiner är linjära molekyler som består av olika aminosyror. För att proteiner ska fungera, måste aminosyradjan vara veckad i en speciell tredimensionell form. Stressfaktorer, som till exempel värme, kan göra att proteinerna delvis förlorar denna veckning. Eftersom cellen är så fullpackad, så finns det en stor risk att sådana delvis oveckade proteiner klistrar fast i varandra och börjar klumpa ihop sig. En liknande process sker när ett ägg utsätts för stress i form av värme (äggvita har en proteinkoncentration av ungefär 100 mg/ml): proteinerna veckas ut och klumpar ihop sig, och man ser att ägget stelnar när man steker eller kokar det. I cellen kan ihop-klumpade proteiner orsaka allvarliga problem, och många olika sjukdomar är konsekvensen av att proteiner klumpat ihop sig (Alzheimers och Parkinsons sjukdom, och grå starr, till exempel).

Lyckligtvis finns det en speciell klass av proteiner i cellen som heter chaperoner, vars funktion är att hindra andra proteiner från att klumpa ihop sig. Det är emellertid inte helt utrett, på en molekylär nivå, hur det går till när chaperonerna gör detta. Särskilt vad gäller en mycket viktig undergrupp av chaperoner, som paradoxalt nog kallas '*small heat shock proteins*', finns det få undersökningar gjorda i jämförelse med andra chaperoner.

Fokus i denna avhandling är hur ett sådant '*small heat shock protein*', kallat Hsp21, skyddar andra proteiner från att klumpa ihop sig. För att studera detta har en kombination av kemisk koppling och masspektrometri använts. Kemisk koppling innebär att ett speciellt kemiskt reagens används för att parvis koppla ihop aminosyror i de olika proteinerna. Efter det klyvs proteinerna i små bitar. Två bitar som kopplats ihop kan sedan detekteras med hjälp av masspektrometri. Genom att identifiera vilka bitar och vilka aminosyror av proteinerna som kopplats, kan man studera vilka delar av proteinerna som var nära varandra och interagerade med varandra precis när man tillsatte det kemiska reagenset. På det sättet, kan information utvinnas gällande interaktionen mellan Hsp21 och det proteinet som skyddas av Hsp21. Experiment som dessa bidrar till vår förståelse av hur chaperonerna hela tiden skyddar de viktiga proteinerna i alla våra celler.

## 6.3 Populairwetenschappelijke samenvatting in het Nederlands

De bouwstenen van al het leven op onze planeet zijn cellen. Ons eigen lichaam bestaat bijvoorbeeld uit ongeveer 10 biljoen ( $10^{13}$ ) cellen. De meeste bacteriën bestaan maar uit één cel. In cellen zitten biomoleculen die verantwoordelijk zijn voor het leven, zoals eiwitten, suikers, vetten en DNA. Cellen zitten enorm volgepakt, vooral wat betreft eiwitten, met eiwitconcentraties van rond de 300 mg/ml.

Eiwitten zijn lineaire moleculen die zijn opgebouwd uit aminozuren. Om hun functie te kunnen uitoefenen, moeten de lange aminozuurketens gevouwen blijven in hun originele driedimensionale vorm. Stressfactoren, zoals hitte, kunnen ervoor zorgen dat eiwitten gedeeltelijk ontvouwen raken. Omdat de eiwitconcentratie in de cel zo hoog is, bestaat er voortdurend een risico dat deels ontvouwen eiwitten aan elkaar gaan plakken en gaan klonteren. Dit is ook wat er gebeurt als een ei wordt gebakken of gekookt (de eiwitconcentratie van ei is ongeveer 120 mg/ml): de eiwitten ontvouwen en klonteren, waardoor het ei stolt als het wordt verwarmd. In de cel kan eiwitklontering allerlei belangrijke processen verstoren. Eiwitklontering is dan ook de oorzaak van veel verschillende ziekten, zoals Alzheimer en Parkinson, en de oogaandoening grijze staar.

Gelukkig is er een speciale groep eiwitten in de cel die voorkomt dat andere eiwitten gaan klonteren. Deze groep beschermende eiwitten wordt chaperonne-eiwitten genoemd. Het is echter nog deels onduidelijk, op moleculair niveau, hoe dit precies in zijn werk gaat. De '*small heat shock proteins*' vormen een subgroep van de chaperonne-eiwitten, en een voorbeeld hiervan is het eiwit Hsp21.

Dit proefschrift gaat over hoe het chaperonne-eiwit Hsp21 kan voorkomen dat andere eiwitten gaan klonteren. Om deze vraag te beantwoorden, werd gebruik gemaakt van een combinatie van crosslinking en massaspectrometrie. Op het moment dat het chaperonne-eiwit Hsp21 ervoor zorgde dat een deels ontvouwen eiwit niet ging klonteren, werden Hsp21 en het ontvouwen eiwit aan elkaar gekoppeld met een crosslinker, een stofje dat verbindingen maakt tussen aminozuren van eiwitten. Vervolgens werd dit eiwitcomplex in kleine stukjes geknipt. De stukjes verbonden door de crosslinker werden opgespoord met behulp van massaspectrometrie, om te achterhalen welke aminozuren in contact met elkaar waren tijdens de chaperonne-activiteit van Hsp21. Door experimenten zoals deze begrijpen we beter hoe chaperonne-eiwitten werken, en hoe ze er dus voor zorgen dat onze cellen niet verstopt raken door eiwitklontering.



# Acknowledgements

This thesis, and the work described in it, could not have been finalized without the help and support of many people. Therefore, I wish to express my gratitude for all the assistance in various forms, and I would like to thank a few people in particular:

Cecilia Emanuelsson, ett stort TACK för excellent supervision och inexhaustible involvement and dedication. Det har varit ett stort nöje att få samarbeta så intensivt under de senaste åren.

All my scientific mentors over the years, who actually made me consider to start as a PhD student. Paul van Bergen en Henegouwen, Jan-Willem de Gier, en Piet Gros, bedankt voor jullie aanstekelijke enthousiasme voor de biomoleculaire wetenschappen. Lucy Rutten, jij hebt mij van dichtbij laten zien hoe leuk het is om met structuurbiologie bezig te zijn, bedankt hiervoor. Derek Logan, for convincing me to come to Lund in the first place. Folke Tjerneld, för ditt förtroende i mig genom att låta mig börja som doktorand hos dig, och för all hjälp genom åren. Det har varit trevligt att även dela intressen utöver biokemi.

All collaborators who contributed to the work described in this thesis. Emma Åhrman, tack för att du har påbörjat projektet och drog mig in i det, troligtvis omedvetet att det skulle bli ämnet av min avhandling sedan. Philip Koeck, Pasi Purhonen och Hans Hebert, för arbetet med EM, och för att jag har känt mig så välkommen i Stockholm. Det var trevliga utflykter från masspektrometri och jag har lärt mig åtminstone lite grann om EM. Christopher, för dina insatser vad gäller crosslinking ms data analys; ditt engagemang har varit ovärderligt. Sven, Gudrun, Cecilia M., Ragna, och Katja, för alla bidrag i de olika papper.

All other collaborators and colleagues, for their scientific input, inspiration and help in the lab. Sven Kjellström och Niklas Gustavsson, för att dela er expertis på masspektrometrin. Henrik Everberg, Johan Börjesson, Rikard Alm, och Emma Åhrman för all hjälp när jag bara precis börjat på avdelningen. Maria Agemark, Kristina Nordén, Hans-Olof Johansson, Morten Krogh och Susanne Luoto för trevliga samarbeten. Sofia Lindahl och Louise Hedskog, för att ni gjorde examensarbete hos mig; ni var superduktiga och hoppas ni lärt er lika mycket som jag lärde mig av er.

Everyone (former and present members) at the department of biochemistry and structural biology and CMPS, for the nice working environment, scientific

interaction and relaxing coffee breaks. Mina närmaste kollegor för att vi har kunnat dela våra dagliga bekymmer, och för att det har varit extra trevligt på jobbet: Katja, Ragna, Christopher, Gudrun, Cecilia M., Kristina, Andreas, Sinan, och Hans-Olof. Gert och Susanna för teknisk och administrativ support. Alla andra som engagerar sig för avdelningens väll-fungerandet och bra atmosfär!

All musicians and scientists whom I've had the pleasure of playing music and innebandy with; this has really been a wonderful part of my life in Lund, and has definitely had an impact on the work resulting in this thesis – after careful consideration I think it must have been a positive impact. Bland de olika musikaliska sammanhang jag har fått vara del av, vill jag särskild nämna Poco piu mossop, Sju sorters kakor, och Debussykvartetten här, samt LNK och akademiska kapellet. Tack för all härlig musik, skogspromenader, ostfondue, fester och vänskapen spelandet har fört med sig. Ett speciellt tack till Selma, Susanna och Roland. Här vill jag också nämna Dorota, för musiken och alltid en öppen dörr i Stockholm.

All friends and family, for their encouragement, interest and good times we spent together throughout the years. Bedankt voor al het bezoek uit Nederland dat we hebben mogen ontvangen in en nabij Lund, het was altijd erg gezellig. Ook als een tripje Zweden niet gelukt is, toch bedankt voor de betrokkenheid. Gelukkig is Wageningen 'dichterbij', want... 'alles' in Nederland is toch dichtbij? Elise, jou wil ik speciaal bedanken voor je onvoorwaardelijke vriendschap en betrokkenheid.

Ook mijn lieve familie wil ik bedanken voor alle steun en gezellige afleiding. Arne en Lianne, en Renske en Maarten, gelukkig zien we elkaar tegenwoordig weer wat vaker, ook al vonden jullie het ook prima om af en toe langs te komen in Lund. Binnenkort moeten we misschien wel afspreken op een heel ander continent. Schone zussen en hun partners en kinderen, dank voor jullie betrokkenheid, en ik vind het gezellig dat we nu wat dichterbij wonen. Ditzelfde geldt voor Henk en Ans, en bovendien wil ik jullie hierbij speciaal bedanken voor al jullie hulp, vooral wat betreft het oppassen op Linus en Annika; ik had me geen betere oplossing kunnen bedenken!

Lieve Niek en Jenneke, papa en mama, bedankt voor jullie vertrouwen in mij, en het creëren van alle randvoorwaarden voor de succesvolle afronding van dit proefschrift en mijn promotie. Jullie steun en belangstelling zijn van onschatbare waarde. Geweldig dat ik met de kindjes ook bij jullie terecht kon tijdens het schrijven.

Lieve Raymond, dankzij jou weet ik weer nieuwe energie op te doen als die even op is. Bedankt voor alle steun tijdens deze jaren van mijn promotieonderzoek, en fantastisch dat we samen ons avontuur in Zweden tot een succes hebben gemaakt. Jag älskar dig! Het allermooiste wat ons overkwam in Lund konden we gelukkig ook meenemen naar Nederland. Lieve Annika en Linus, nu is het boekje klaar, en dan komt mama weer.

# References

- Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* *422*, 198-207.
- Ahrman, E., Gustavsson, N., Hultschig, C., Boelens, W.C., and Emanuelsson, C.S. (2007a). Small heat shock proteins prevent aggregation of citrate synthase and bind to the N-terminal region which is absent in thermostable forms of citrate synthase. *Extremophiles* *11*, 659-666.
- Ahrman, E., Lambert, W., Aquilina, J.A., Robinson, C.V., and Emanuelsson, C.S. (2007b). Chemical cross-linking of the chloroplast localized small heat-shock protein, Hsp21, and the model substrate citrate synthase. *Protein Sci* *16*, 1464-1478.
- Bagneris, C., Bateman, O.A., Naylor, C.E., Cronin, N., Boelens, W.C., Keep, N.H., and Slingsby, C. (2009). Crystal structures of alpha-crystallin domain dimers of alphaB-crystallin and Hsp20. *J Mol Biol* *392*, 1242-1252.
- Baldwin, A.J., Hilton, G.R., Lioe, H., Bagneris, C., Benesch, J.L., and Kay, L.E. (2011a). Quaternary dynamics of alphaB-crystallin as a direct consequence of localised tertiary fluctuations in the C-terminus. *J Mol Biol* *413*, 310-320.
- Baldwin, A.J., Lioe, H., Hilton, G.R., Baker, L.A., Rubinstein, J.L., Kay, L.E., and Benesch, J.L. (2011b). The polydispersity of alphaB-crystallin is rationalized by an interconverting polyhedral architecture. *Structure* *19*, 1855-1863.
- Baldwin, A.J., Lioe, H., Robinson, C.V., Kay, L.E., and Benesch, J.L. (2011c). alphaB-crystallin polydispersity is a consequence of unbiased quaternary dynamics. *J Mol Biol* *413*, 297-309.
- Baranova, E.V., Weeks, S.D., Beelen, S., Bukach, O.V., Gusev, N.B., and Strelkov, S.V. (2011). Three-dimensional structure of alpha-crystallin domain dimers of human small heat shock proteins HSPB1 and HSPB6. *J Mol Biol* *411*, 110-122.
- Basha, E., Friedrich, K.L., and Vierling, E. (2006). The N-terminal arm of small heat shock proteins is important for both chaperone activity and substrate specificity. *J Biol Chem* *281*, 39943-39952.
- Basha, E., Jones, C., Wysocki, V., and Vierling, E. (2010). Mechanistic differences between two conserved classes of small heat shock proteins found in the plant cytosol. *J Biol Chem* *285*, 11489-11497.
- Basha, E., Lee, G.J., Breci, L.A., Hausrath, A.C., Buan, N.R., Giese, K.C., and Vierling, E. (2004). The identity of proteins associated with a small heat shock protein during heat

stress in vivo indicates that these chaperones protect a wide range of cellular functions. *J Biol Chem* 279, 7566-7575.

Basha, E., O'Neill, H., and Vierling, E. (2011). Small heat shock proteins and alpha-crystallins: dynamic proteins with flexible functions. *Trends Biochem Sci*.

Beck, M., Malmstrom, J.A., Lange, V., Schmidt, A., Deutsch, E.W., and Aebersold, R. (2009). Visual proteomics of the human pathogen *Leptospira interrogans*. *Nat Methods* 6, 817-823.

Benesch, J.L., Aquilina, J.A., Baldwin, A.J., Rekas, A., Stengel, F., Lindner, R.A., Basha, E., Devlin, G.L., Horwitz, J., Vierling, E., *et al.* (2010). The quaternary organization and dynamics of the molecular chaperone HSP26 are thermally regulated. *Chem Biol* 17, 1008-1017.

Benesch, J.L., Ayoub, M., Robinson, C.V., and Aquilina, J.A. (2008). Small heat shock protein activity is regulated by variable oligomeric substructure. *J Biol Chem* 283, 28513-28517.

Benesch, J.L., Ruotolo, B.T., Simmons, D.A., and Robinson, C.V. (2007). Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chem Rev* 107, 3544-3567.

Bennett, K.L., Kussmann, M., Bjork, P., Godzwon, M., Mikkelsen, M., Sorensen, P., and Roepstorff, P. (2000). Chemical cross-linking with thiol-cleavable reagents combined with differential mass spectrometric peptide mapping--a novel approach to assess intermolecular protein contacts. *Protein Sci* 9, 1503-1518.

Berengian, A.R., Bova, M.P., and McHaourab, H.S. (1997). Structure and function of the conserved domain in alphaA-crystallin. Site-directed spin labeling identifies a beta-strand located near a subunit interface. *Biochemistry* 36, 9951-9957.

Berengian, A.R., Parfenova, M., and McHaourab, H.S. (1999). Site-directed spin labeling study of subunit interactions in the alpha-crystallin domain of small heat-shock proteins. Comparison of the oligomer symmetry in alphaA-crystallin, HSP 27, and HSP 16.3. *J Biol Chem* 274, 6305-6314.

Berg, J.M.T., John L., 1948; Stryer, Lubert, 1975 (2012). *Biochemistry*, 7th edn (Palgrave Macmillan).

Bova, M.P., Ding, L.L., Horwitz, J., and Fung, B.K. (1997). Subunit exchange of alphaA-crystallin. *J Biol Chem* 272, 29511-29517.

Bova, M.P., McHaourab, H.S., Han, Y., and Fung, B.K. (2000). Subunit exchange of small heat shock proteins. Analysis of oligomer formation of alphaA-crystallin and Hsp27 by fluorescence resonance energy transfer and site-directed truncations. *J Biol Chem* 275, 1035-1042.

Chen, Z.A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., Rasmussen, M., Lariviere, L., Bukowski-Wills, J.C., Nilges, M., *et al.* (2010). Architecture of the RNA

polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *Embo J* 29, 717-726.

Cheng, G., Basha, E., Wysocki, V.H., and Vierling, E. (2008). Insights into small heat shock protein and substrate structure during chaperone action derived from hydrogen/deuterium exchange and mass spectrometry. *J Biol Chem* 283, 26634-26642.

Chiti, F., and Dobson, C.M. (2009). Amyloid formation by globular proteins under native conditions. *Nat Chem Biol* 5, 15-22.

Ciferri, C., Pasqualato, S., Screpanti, E., Varetti, G., Santaguida, S., Dos Reis, G., Maiolica, A., Polka, J., De Luca, J.G., De Wulf, P., *et al.* (2008). Implications for kinetochore-microtubule attachment from the structure of an engineered Ndc80 complex. *Cell* 133, 427-439.

Clark, A.R., Naylor, C.E., Bagneris, C., Keep, N.H., and Slingsby, C. (2011). Crystal structure of R120G disease mutant of human alphaB-crystallin domain dimer shows closure of a groove. *J Mol Biol* 408, 118-134.

Clegg, C., and Hayes, D. (1974). Identification of neighbouring proteins in the ribosomes of *Escherichia coli*. A topographical study with the cross-linking reagent dimethyl suberimidate. *Eur J Biochem* 42, 21-28.

de Jong, W.W., Caspers, G.J., and Leunissen, J.A. (1998). Genealogy of the alpha-crystallin--small heat-shock protein superfamily. *Int J Biol Macromol* 22, 151-162.

DeLano, W. (2002). The PyMOL Molecular Graphics System.

Domon, B., and Aebersold, R. (2006). Review - Mass spectrometry and protein analysis. *Science* 312, 212-217.

Dougan, D.A., Mogk, A., and Bukau, B. (2002). Protein folding and degradation in bacteria: to degrade or not to degrade? That is the question. *Cell Mol Life Sci* 59, 1607-1616.

Drenth, J. (2007). *Principles of Protein X-Ray Crystallography*, 3rd edn (Springer New York).

During, R.L., Gibson, B.G., Li, W., Bishai, E.A., Sidhu, G.S., Landry, J., and Southwick, F.S. (2007). Anthrax lethal toxin paralyzes actin-based motility by blocking Hsp27 phosphorylation. *Embo J* 26, 2240-2250.

Ecroyd, H., Meehan, S., Horwitz, J., Aquilina, J.A., Benesch, J.L., Robinson, C.V., Macphee, C.E., and Carver, J.A. (2007). Mimicking phosphorylation of alphaB-crystallin affects its chaperone activity. *Biochem J* 401, 129-141.

Eyles, S.J., and Gierasch, L.M. (2010). Nature's molecular sponges: small heat shock proteins grow into their chaperone roles. *Proc Natl Acad Sci U S A* 107, 2727-2728.

Fabris, D., and Yu, E.T. (2010). Elucidating the higher-order structure of biopolymers by structural probing and mass spectrometry: MS3D. *J Mass Spectrom* 45, 841-860.

- Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246, 64-71.
- Franzmann, T.M., Menhorn, P., Walter, S., and Buchner, J. (2008). Activation of the chaperone Hsp26 is controlled by the rearrangement of its thermosensor domain. *Mol Cell* 29, 207-216.
- Friedrich, K.L., Giese, K.C., Buan, N.R., and Vierling, E. (2004). Interactions between small heat shock protein subunits and substrate in small heat shock protein-substrate complexes. *J Biol Chem* 279, 1080-1089.
- Fritzsche, R., Ihling, C.H., Gotze, M., and Sinz, A. (2012). Optimizing the enrichment of cross-linked products for mass spectrometric protein analysis. *Rapid Commun Mass Spectrom* 26, 653-658.
- Ghosh, J.G., Houck, S.A., and Clark, J.I. (2007). Interactive sequences in the stress protein and molecular chaperone human alphaB crystallin recognize and modulate the assembly of filaments. *Int J Biochem Cell Biol* 39, 1804-1815.
- Giese, K.C., Basha, E., Catague, B.Y., and Vierling, E. (2005). Evidence for an essential function of the N terminus of a small heat shock protein in vivo, independent of in vitro chaperone activity. *Proc Natl Acad Sci U S A* 102, 18896-18901.
- Giese, K.C., and Vierling, E. (2002). Changes in oligomerization are essential for the chaperone activity of a small heat shock protein in vivo and in vitro. *J Biol Chem* 277, 46310-46318.
- Giese, K.C., and Vierling, E. (2004). Mutants in a small heat shock protein that affect the oligomeric state. Analysis and allele-specific suppression. *J Biol Chem* 279, 32674-32683.
- Gold, M.G., Stengel, F., Nygren, P.J., Weisbrod, C.R., Bruce, J.E., Robinson, C.V., Barford, D., and Scott, J.D. (2011). Architecture and dynamics of an A-kinase anchoring protein 79 (AKAP79) signaling complex. *Proc Natl Acad Sci U S A* 108, 6426-6431.
- Goldfarb, L.G., Olive, M., Vicart, P., and Goebel, H.H. (2008). Intermediate filament diseases: desminopathy. *Adv Exp Med Biol* 642, 131-164.
- Gotze, M., Pettelkau, J., Schaks, S., Bosse, K., Ihling, C.H., Krauth, F., Fritzsche, R., Kuhn, U., and Sinz, A. (2012). StavroX--a software for analyzing crosslinked products in protein interaction studies. *J Am Soc Mass Spectrom* 23, 76-87.
- Gustavsson, N., Kokke, B.P., Harndahl, U., Silow, M., Bechtold, U., Poghosyan, Z., Murphy, D., Boelens, W.C., and Sundby, C. (2002). A peptide methionine sulfoxide reductase highly expressed in photosynthetic tissue in *Arabidopsis thaliana* can protect the chaperone-like activity of a chloroplast-localized small heat shock protein. *Plant J* 29, 545-553.
- Han, X., Aslanian, A., and Yates, J.R., 3rd (2008). Mass spectrometry for proteomics. *Curr Opin Chem Biol* 12, 483-490.
- Harndahl, U., Hall, R.B., Osteryoung, K.W., Vierling, E., Bornman, J.F., and Sundby, C. (1999). The chloroplast small heat shock protein undergoes oxidation-dependent

conformational changes and may protect plants from oxidative stress. *Cell Stress Chaperones* 4, 129-138.

Harndahl, U., Kokke, B.P., Gustavsson, N., Linse, S., Berggren, K., Tjerneld, F., Boelens, W.C., and Sundby, C. (2001). The chaperone-like activity of a small heat shock protein is lost after sulfoxidation of conserved methionines in a surface-exposed amphipathic  $\alpha$ -helix. *Biochim Biophys Acta* 1545, 227-237.

Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature* 475, 324-332.

Haslbeck, M., Braun, N., Stromer, T., Richter, B., Model, N., Weinkauff, S., and Buchner, J. (2004). Hsp42 is the general small heat shock protein in the cytosol of *Saccharomyces cerevisiae*. *Embo J* 23, 638-649.

Haslbeck, M., Franzmann, T., Weinfurter, D., and Buchner, J. (2005). Some like it hot: the structure and function of small heat-shock proteins. *Nat Struct Mol Biol* 12, 842-846.

Heck, A.J. (2008). Native mass spectrometry: a bridge between interactomics and structural biology. *Nature Methods* 5, 927-933.

Hernandez, H., and Robinson, C.V. (2007). Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat Protoc* 2, 715-726.

Hilario, E., Martin, F.J., Bertolini, M.C., and Fan, L. (2011). Crystal structures of *Xanthomonas* small heat shock protein provide a structural basis for an active molecular chaperone oligomer. *J Mol Biol* 408, 74-86.

Horwitz, J. (1992).  $\alpha$ -crystallin can function as a molecular chaperone. *Proc Natl Acad Sci U S A* 89, 10449-10453.

Jakob, U., Gaestel, M., Engel, K., and Buchner, J. (1993). Small heat shock proteins are molecular chaperones. *J Biol Chem* 268, 1517-1520.

Jaya, N., Garcia, V., and Vierling, E. (2009). Substrate binding site flexibility of the small heat shock protein molecular chaperones. *Proc Natl Acad Sci U S A* 106, 15604-15609.

Jehle, S., Rajagopal, P., Bardiaux, B., Markovic, S., Kuhne, R., Stout, J.R., Higman, V.A., Kleivit, R.E., van Rossum, B.J., and Oschkinat, H. (2010). Solid-state NMR and SAXS studies provide a structural basis for the activation of  $\alpha$ HB-crystallin oligomers. *Nat Struct Mol Biol* 17, 1037-1042.

Jehle, S., Vollmar, B.S., Bardiaux, B., Dove, K.K., Rajagopal, P., Gonen, T., Oschkinat, H., and Kleivit, R.E. (2011). N-terminal domain of  $\alpha$ HB-crystallin provides a conformational switch for multimerization and structural heterogeneity. *Proc Natl Acad Sci U S A* 108, 6409-6414.

Kampinga, H.H., Hageman, J., Vos, M.J., Kubota, H., Tanguay, R.M., Bruford, E.A., Cheetham, M.E., Chen, B., and Hightower, L.E. (2009). Guidelines for the nomenclature of the human heat shock proteins. *Cell Stress Chaperones* 14, 105-111.

- Karas, M., and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* *60*, 2299-2301.
- Kennaway, C.K., Benesch, J.L., Gohlke, U., Wang, L., Robinson, C.V., Orlova, E.V., Saibil, H.R., and Keep, N.H. (2005). Dodecameric structure of the small heat shock protein Acr1 from *Mycobacterium tuberculosis*. *J Biol Chem* *280*, 33419-33425.
- Kim, K.K., Kim, R., and Kim, S.H. (1998). Crystal structure of a small heat-shock protein. *Nature* *394*, 595-599.
- King, G.J., Jones, A., Kobe, B., Huber, T., Mouradov, D., Hume, D.A., and Ross, I.L. (2008). Identification of disulfide-containing chemical cross-links in proteins using MALDI-TOF/TOF-mass spectrometry. *Anal Chem* *80*, 5036-5043.
- Laganowsky, A., Benesch, J.L., Landau, M., Ding, L., Sawaya, M.R., Cascio, D., Huang, Q., Robinson, C.V., Horwitz, J., and Eisenberg, D. (2010). Crystal structures of truncated alphaA and alphaB crystallins reveal structural mechanisms of polydispersity important for eye lens function. *Protein Sci* *19*, 1031-1043.
- Laganowsky, A., and Eisenberg, D. (2010). Non-3D domain swapped crystal structure of truncated zebrafish alphaA crystallin. *Protein Sci* *19*, 1978-1984.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* *23*, 2947-2948.
- Lee, Y.J. (2008). Mass spectrometric analysis of cross-linking sites for the structure of proteins and protein complexes. *Mol Biosyst* *4*, 816-823.
- Leitner, A., Reischl, R., Walzthoeni, T., Herzog, F., Bohn, S., Foerster, F., and Aebersold, R. (2012). Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Molecular & Cellular Proteomics*.
- Leitner, A., Walzthoeni, T., Kahraman, A., Herzog, F., Rinner, O., Beck, M., and Aebersold, R. (2010). Probing Native Protein Structures by Chemical Cross-linking, Mass Spectrometry, and Bioinformatics. *Molecular & Cellular Proteomics* *9*, 1634-1649.
- Lindner, R.A., Treweek, T.M., and Carver, J.A. (2001). The molecular chaperone alpha-crystallin is in kinetic competition with aggregation to stabilize a monomeric molten-globule form of alpha-lactalbumin. *Biochem J* *354*, 79-87.
- Maiolica, A., Cittaro, D., Borsotti, D., Sennels, L., Ciferri, C., Tarricone, C., Musacchio, A., and Rappsilber, J. (2007). Structural analysis of multiprotein complexes by cross-linking, mass spectrometry, and database searching. *Molecular & Cellular Proteomics* *6*, 2200-2211.
- Mayer, M.P. (2010). Gymnastics of molecular chaperones. *Mol Cell* *39*, 321-331.
- Mayne, S.L., and Patterson, H.G. (2011). Bioinformatics tools for the structural elucidation of multi-subunit protein complexes by mass spectrometric analysis of protein-protein cross-links. *Brief Bioinform* *12*, 660-671.



- McHaourab, H.S., Berengian, A.R., and Koteiche, H.A. (1997). Site-directed spin-labeling study of the structure and subunit interactions along a conserved sequence in the alpha-crystallin domain of heat-shock protein 27. Evidence of a conserved subunit interface. *Biochemistry* *36*, 14627-14634.
- McHaourab, H.S., Godar, J.A., and Stewart, P.L. (2009). Structure and mechanism of protein stability sensors: chaperone activity of small heat shock proteins. *Biochemistry* *48*, 3828-3837.
- Mogk, A., Deuerling, E., Vorderwulbecke, S., Vierling, E., and Bukau, B. (2003a). Small heat shock proteins, ClpB and the DnaK system form a functional triade in reversing protein aggregation. *Mol Microbiol* *50*, 585-595.
- Mogk, A., Schlieker, C., Friedrich, K.L., Schonfeld, H.J., Vierling, E., and Bukau, B. (2003b). Refolding of substrates bound to small Hsps relies on a disaggregation reaction mediated most efficiently by ClpB/DnaK. *J Biol Chem* *278*, 31033-31042.
- Muller, D.R., Schindler, P., Towbin, H., Wirth, U., Voshol, H., Hoving, S., and Steinmetz, M.O. (2001). Isotope tagged cross linking reagents. A new tool in mass spectrometric protein interaction analysis. *Analytical Chemistry* *73*, 1927-1934.
- Nakamoto, H., and Vigh, L. (2007). The small heat shock proteins and their clients. *Cell Mol Life Sci* *64*, 294-306.
- Narberhaus, F. (2002). Alpha-crystallin-type heat shock proteins: socializing minichaperones in the context of a multichaperone network. *Microbiol Mol Biol Rev* *66*, 64-93; table of contents.
- Painter, A.J., Jaya, N., Basha, E., Vierling, E., Robinson, C.V., and Benesch, J.L. (2008). Real-time monitoring of protein complexes reveals their quaternary organization and dynamics. *Chem Biol* *15*, 246-253.
- Paizs, B., and Suhai, S. (2005). Fragmentation pathways of protonated peptides. *Mass Spectrom Rev* *24*, 508-548.
- Panchaud, A., Singh, P., Shaffer, S.A., and Goodlett, D.R. (2010). xComb: A Cross-Linked Peptide Database Approach to Protein-Protein Interaction Analysis. *Journal of Proteome Research* *9*, 2508-2515.
- Peri, S., Steen, H., and Pandey, A. (2001). GPMAW--a software tool for analyzing proteins and peptides. *Trends Biochem Sci* *26*, 687-689.
- Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* *20*, 3551-3567.
- Peschek, J., Braun, N., Franzmann, T.M., Georgalis, Y., Haslbeck, M., Weinkauff, S., and Buchner, J. (2009). The eye lens chaperone alpha-crystallin forms defined globular assemblies. *Proc Natl Acad Sci U S A* *106*, 13272-13277.
- Petrochenko, E.V., and Borchers, C.H. (2010). Crosslinking combined with mass spectrometry for structural proteomics. *Mass Spectrom Rev* *29*, 862-876.

- Poulain, P., Gelly, J.C., and Flatters, D. (2010). Detection and architecture of small heat shock protein monomers. *PLoS One* 5, e9990.
- Putnam, C.D., Hammel, M., Hura, G.L., and Tainer, J.A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* 40, 191-285.
- Rappsilber, J. (2011). The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J Struct Biol* 173, 530-540.
- Rasmussen, M.I., Refsgaard, J.C., Peng, L., Houen, G., and Hojrup, P. (2011). CrossWork: software-assisted identification of cross-linked peptides. *J Proteomics* 74, 1871-1883.
- Ratajczak, E., Zietkiewicz, S., and Liberek, K. (2009). Distinct activities of Escherichia coli small heat shock proteins IbpA and IbpB promote efficient protein disaggregation. *J Mol Biol* 386, 178-189.
- Rinner, O., Seebacher, J., Walzthoeni, T., Mueller, L., Beck, M., Schmidt, A., Mueller, M., and Aebersold, R. (2008). Identification of cross-linked peptides from large sequence databases. *Nature Methods* 5, 315-318.
- Santos, L.F., Iglesias, A.H., and Gozzo, F.C. (2011). Fragmentation features of intermolecular cross-linked peptides using N-hydroxy-succinimide esters by MALDI- and ESI-MS/MS for use in structural proteomics. *J Mass Spectrom* 46, 742-750.
- Scharf, K.D., Siddique, M., and Vierling, E. (2001). The expanding family of Arabidopsis thaliana small heat stress proteins and a new family of proteins containing alpha-crystallin domains (Acid proteins). *Cell Stress Chaperones* 6, 225-237.
- Schilling, B., Row, R.H., Gibson, B.W., Guo, X., and Young, M.M. (2003). MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J Am Soc Mass Spectr* 14, 834-850.
- Selmer, M., and Liljas, A. (2008). Exit biology: battle for the nascent chain. *Structure* 16, 498-500.
- Sharma, K.K., Kumar, R.S., Kumar, G.S., and Quinn, P.T. (2000). Synthesis and characterization of a peptide identified as a functional element in alphaA-crystallin. *J Biol Chem* 275, 3767-3771.
- Sharon, M., and Robinson, C.V. (2007). The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu Rev Biochem* 76, 167-193.
- Shashidharamurthy, R., Koteiche, H.A., Dong, J., and McHaourab, H.S. (2005). Mechanism of chaperone function in small heat shock proteins: dissociation of the HSP27 oligomer is required for recognition and binding of destabilized T4 lysozyme. *J Biol Chem* 280, 5281-5289.
- Shi, J., Koteiche, H.A., McHaourab, H.S., and Stewart, P.L. (2006). Cryoelectron microscopy and EPR analysis of engineered symmetric and polydisperse Hsp16.5

assemblies reveals determinants of polydispersity and substrate binding. *J Biol Chem* 281, 40420-40428.

Siddique, M., Gernhard, S., von Koskull-Doring, P., Vierling, E., and Scharf, K.D. (2008). The plant sHSP superfamily: five new members in *Arabidopsis thaliana* with unexpected properties. *Cell Stress Chaperones* 13, 183-197.

Singh, P., Panchaud, A., and Goodlett, D.R. (2010). Chemical Cross-Linking and Mass Spectrometry As a Low-Resolution Protein Structure Determination Technique. *Analytical Chemistry* 82, 2636-2642.

Sinz, A. (2006). Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrometry Reviews* 25, 663-682.

Sinz, A. (2010). Investigation of protein-protein interactions in living cells by chemical crosslinking and mass spectrometry. *Anal Bioanal Chem* 397, 3433-3440.

Snijder, J., Rose, R.J., Raijmakers, R., and Heck, A.J. (2011). Site-specific methionine oxidation in calmodulin affects structural integrity and interaction with Ca<sup>2+</sup>/calmodulin-dependent protein kinase II. *J Struct Biol* 174, 187-195.

Sobott, F., Benesch, J.L., Vierling, E., and Robinson, C.V. (2002). Subunit exchange of multimeric protein complexes. Real-time monitoring of subunit exchange between small heat shock proteins by using electrospray mass spectrometry. *J Biol Chem* 277, 38921-38929.

Stamler, R., Kappe, G., Boelens, W., and Slingsby, C. (2005). Wrapping the alpha-crystallin domain fold in a chaperone assembly. *J Mol Biol* 353, 68-79.

Stengel, F., Aebersold, R., and Robinson, C.V. (2011). Joining forces: Integrating proteomics and crosslinking with the mass spectrometry of intact complexes. *Molecular & Cellular Proteomics*.

Stengel, F., Aebersold, R., and Robinson, C.V. (2012). Joining Forces: Integrating Proteomics and Cross-linking with the Mass Spectrometry of Intact Complexes. *Molecular & Cellular Proteomics* 11, R111 014027.

Stengel, F., Baldwin, A.J., Painter, A.J., Jaya, N., Basha, E., Kay, L.E., Vierling, E., Robinson, C.V., and Benesch, J.L. (2010). Quaternary dynamics and plasticity underlie small heat shock protein chaperone function. *Proc Natl Acad Sci U S A* 107, 2007-2012.

Sun, Y., and MacRae, T.H. (2005). The small heat shock proteins and their role in human disease. *Febs J* 272, 2613-2627.

Sundby, C., Harndahl, U., Gustavsson, N., Ahrman, E., and Murphy, D.J. (2005). Conserved methionines in chloroplasts. *Biochim Biophys Acta* 1703, 191-202.

Swaim, C.L., Smith, J.B., and Smith, D.L. (2004). Unexpected products from the reaction of the synthetic cross-linker 3,3'-dithiobis(sulfosuccinimidyl propionate), DTSSP with peptides. *J Am Soc Mass Spectr* 15, 736-749.

- Takeda, K., Hayashi, T., Abe, T., Hirano, Y., Hanazono, Y., Yohda, M., and Miki, K. (2011). Dimer structure and conformational variability in the N-terminal region of an archaeal small heat shock protein, StHsp14.0. *J Struct Biol* *174*, 92-99.
- Taverner, T., Hall, N.E., O'Hair, R.A., and Simpson, R.J. (2002). Characterization of an antagonist interleukin-6 dimer by stable isotope labeling, cross-linking, and mass spectrometry. *J Biol Chem* *277*, 46487-46492.
- Trnka, M.J., and Burlingame, A.L. (2010). Topographic studies of the GroEL-GroES chaperonin complex by chemical cross-linking using diformyl ethynylbenzene: the power of high resolution electron transfer dissociation for determination of both peptide sequences and their attachment sites. *Molecular & Cellular Proteomics* *9*, 2306-2317.
- Tyedmers, J., Mogk, A., and Bukau, B. (2010). Cellular strategies for controlling protein aggregation. *Nat Rev Mol Cell Biol* *11*, 777-788.
- Uversky, V.N., Oldfield, C.J., Midic, U., Xie, H., Xue, B., Vucetic, S., Iakoucheva, L.M., Obradovic, Z., and Dunker, A.K. (2009). Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *BMC Genomics* *10 Suppl 1*, S7.
- Van Montfort, R., Slingsby, C., and Vierling, E. (2001a). Structure and function of the small heat shock protein/alpha-crystallin family of molecular chaperones. *Adv Protein Chem* *59*, 105-156.
- Van Montfort, R.L., Basha, E., Friedrich, K.L., Slingsby, C., and Vierling, E. (2001b). Crystal structure and assembly of a eukaryotic small heat shock protein. *Nat Struct Biol* *8*, 1025-1030.
- Vos, M.J., Hageman, J., Carra, S., and Kampinga, H.H. (2008). Structural and functional diversities between members of the human HSPB, HSPH, HSPA, and DNAJ chaperone families. *Biochemistry* *47*, 7001-7011.
- Waters, E.R., Aebermann, B.D., and Sanders-Reed, Z. (2008). Comparative analysis of the small heat shock proteins in three angiosperm genomes identifies new subfamilies and reveals diverse evolutionary patterns. *Cell Stress Chaperones* *13*, 127-142.
- Waters, E.R., Lee, G.J., and Vierling, E. (1996). Evolution, structure and function of the small heat shock proteins in plants. *Journal of Experimental Botany* *47*, 325-338.
- Waters, E.R., and Vierling, E. (1999). Chloroplast small heat shock proteins: evidence for atypical evolution of an organelle-localized protein. *Proc Natl Acad Sci U S A* *96*, 14394-14399.
- White, H.E., Orlova, E.V., Chen, S., Wang, L., Ignatiou, A., Gowen, B., Stromer, T., Franzmann, T.M., Haslbeck, M., Buchner, J., *et al.* (2006). Multiple distinct assemblies reveal conformational flexibility in the small heat shock protein Hsp26. *Structure* *14*, 1197-1204.
- White, H.E., Saibil, H.R., Ignatiou, A., and Orlova, E.V. (2004). Recognition and separation of single particles with size variation by statistical analysis of their images. *J Mol Biol* *336*, 453-460.

Yates, J.R., 3rd, Eng, J.K., McCormack, A.L., and Schieltz, D. (1995). Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal Chem* 67, 1426-1436.

Young, M.M., Tang, N., Hempel, J.C., Oshiro, C.M., Taylor, E.W., Kuntz, I.D., Gibson, B.W., and Dollinger, G. (2000). High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc Natl Acad Sci U S A* 97, 5802-5806.

Zhou, Z.H. (2008). Towards atomic resolution structural determination by single-particle cryo-electron microscopy. *Curr Opin Struct Biol* 18, 218-228.