Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Structure and function of human heat shock transcription factor 1

Tobias Franziskus Neudegger

aus

München, Deutschland

2016

Erklärung

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. F. Ulrich Hartl betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 04.08.2016 Tobias Franziskus Neudegger

Dissertation eingereicht am 26.08.2016

1. Gutachter Prof. Dr. F. Ulrich Hartl

2. Gutachter Prof. Dr. Andreas Ladurner

Mündliche Prüfung am 25.10.2016

Acknowledgement

I am very thankful to my supervisor Prof. Dr. F. Ulrich Hartl for giving me the opportunity to work on this interesting project. His guidance and the inspiring and communicative atmosphere in his department contributed greatly to this work.

This work would have not been possible without the excellent supervision of Dr. habil Andreas Bracher. His advice in theoretical backgrounds as well as in experimental design were truly vital for my scientific and professional progress. His calm and encouraging way of dealing with technical challenges and his way of interpersonal communication contributed extraordinarily to my experience.

I thank Dr. habil Andreas Bracher, Dr. Jacob Verghese, Dr. Manajit Hayer-Hartl and Prof. Dr. F. Ulrich Hartl for their fruitful collaboration on the structural characterization of Hsf1.

I am grateful for the help of the department's staff members: Albert Ries, Emmanuel Burghardt, Nadine Wischnewski, Romy Lange, Anastasia Jungclaus, Evelyn Frey-Royston and Darija Pompino. Further, I want to thank the members of the MPI and crystallization core facilities: Elisabeth Weyher-Stingl, Karina Valer and Sabine Pleyer. Thanks to Dr. Claire Basquin for help with fluorescence polarization experiments. I am thankful to Dr. Javaid Y. Bhat and Dr. Roman Körner for help with the mass spectrometry analysis.

Thanks to all people in the department for the friendly and open-minded atmosphere – especially the "Rubisco Office" - for making my time at the institute so pleasant. I also want to thank Leonie Mönkemeyer and Zhuo Li for valuable tipps during the initial phase. Thanks to Stefan Juhas and Martin Hennemann for technical support.

Great thanks to David Balchin, Amit Gupta, Timm Hassemer, Thomas Hauser, Christian Löw, Martin Müller, Leonhard Popilka and Robert Wilson for friendship and support.

I am grateful to my family for continuous support and for creating my interest in natural sciences. I am grateful to Sabine for her support, patience and love during my course of studies.

Contents

1 Summary

In response to protein conformational stress, eukaryotic cells turn on a specific cellular response program, termed the heat shock response. The major coordinator of this stress response in the cytosol is heat shock transcription factor 1 (Hsf1). Hsf1 binds to the promoters of stress protective genes such as those encoding heat shock proteins (Hsp), thereby selectively inducing their expression. Under non-stress conditions Hsf1 populates a latent, inactive conformation. Stress induces activation of Hsf1 via oligomerization and a variety of post-translational modifications. Considering the role of Hsf1 in safeguarding protein homeostasis, and its potency in supporting tumor growth, there is high interest in targeting Hsf1 with either activating or inhibiting small molecule compounds, respectively. Despite being intensively studied for decades, structural data on Hsf1 remained sparse.

In this study, we applied a variety of techniques to analyze both the overall domain organization and atomic structure of Hsf1. We analyzed the oligomeric state of human Hsf1 using chemical crosslinking, native mass spectrometry and multi-angle light scattering, and propose a model of its three dimensional domain arrangement based on chemical crosslinking coupled to mass spectrometry. We obtained insight into the atomic details of DNA-recognition by human Hsf1 using X-ray crystallography. Crystal structures of the Hsf1 paralog Skn7 from the thermophilic fungus *Chaetomium thermophilum* (Ct) provided insights into the sequence specificity of DNA-binding. We solved crystal structures of the DNA-binding domain (DBD) of human Hsf1 both in complex with a heat shock element (HSE) motif found in promoters of Hsp, and in complex with the motif found in nuclear stress bodies (SatIII). The crystal structures of CtSkn7 were solved in complex with its cognate binding element Sln1 responsive element (SSRE) as well as the alternative binding motif HSE.

In addition to the Hsf-type DBD, crystal structures of the coiled-coil domain of the *Chaetomium* protein were solved, and provided the basis for a homology model of the human oligomerization domain.

By combining the structural data with mutational analysis, we are able to propose a structural model for the human Hsf1-DNA complex in which the Hsf1 trimer embraces the DNA double helix, with the DBDs on the opposite side of the DNA relative to the coiled-coil domain. Furthermore, our model suggests how adjacent DBDs interact at the DNA, and how the DNA-distal surface and loop region can interact with Hsf1 binding proteins such as factors of the transcription associated machinery.

2 Introduction

All living cells, the basic structural and functional units of life, need to adapt to their specific habitat in order to survive. Cells have to cope with a variety of adverse environmental conditions which challenge the integrity of their cellular protein machinery. It is essential that protein homeostasis (proteostasis) - a balanced state of all protein components - is carefully maintained in the face of acute and chronic proteotoxic stress thereby preventing disintegration of the cellular proteome. Imbalance in the complex quality control network of proteostasis results in the activation of universal and highly conserved stress response pathways such as the cytosolic stress response (Morimoto 2008; Anckar & Sistonen 2011; Kim et al. 2013). This signaling pathway transiently upregulates so-called heat shock proteins (Hsps), including many molecular chaperones, in an attempt to restore proteostasis.

2.1 Proteostasis

The term proteostasis refers to a balanced and healthy state of the cellular proteome (Balch et al. 2008). To achieve and maintain proteostasis in the crowded cellular environment, a complex, integrated quality control network has evolved (Zimmerman & Trach 1991; Powers et al. 2009). Organisms invest a large amount of resources into this network comprising several hundreds of proteins (~1300 proteins in human cells) to perform protein biogenesis (~400 proteins), conformational maintenance (~300 proteins) and degradation (~700 proteins) (Hipp et al. 2014).

While molecular chaperones and their respective co-chaperones assist in the *de novo* folding, refolding, or disaggregation of proteins, the ubiquitin-proteasome system (UPS) and the autophagy machinery perform the proteolytic degradation of irreversible misfolded proteins and protein aggregates (Kim et al. 2013). Multiple interconnected stress-inducible signaling pathways adjust the proteostasis network in response to specific forms of cellular stress, including the cytosolic heat shock response (HSR) (Morimoto 2008), the unfolded protein response (UPR) (Ron & Walter 2007) in the endoplasmic reticulum, and the mitochondrial UPR (Haynes & Ron 2010). The network is connected to signaling pathways in

response to oxidative stress, starvation and longevity and is coupled to the translational capacity of the ribosomes (Hartl et al. 2011).

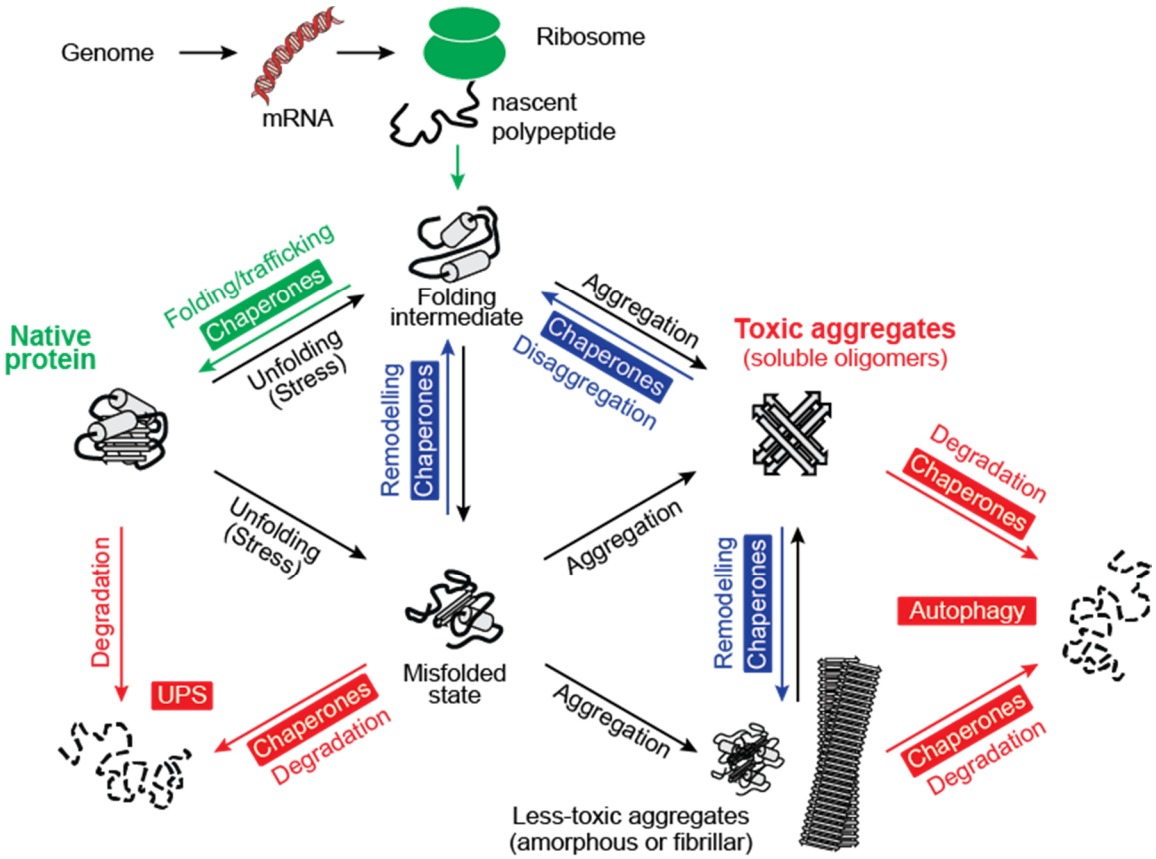

Figure 1 Cellular proteostasis network. Genomic information is transcribed into mRNA and subsequently translated into a polypeptide chain. This polypeptide chain folds with or without the help of molecular chaperones to its native state (green). Remodeling of misfolded states of a protein fold is performed by molecular chaperones (blue). Degradation of proteins is performed via several pathways (red). Adapted by permission from Elsevier: Trends Cell Biol. (Hipp et al. 2014), copyright 2014

2.2 Protein folding in the cell

2.2.1 Protein folding – and aggregation

Proteins are synthesized as linear chains of amino acids connected by peptide bonds. To function, however, most proteins must fold into a specific three dimensional topology. The fold of a small protein or single domain (<~100 amino acids) is typically encoded alone by its amino acid sequence, as observed by Anfinsen in seminal *in vitro* refolding experiments (Anfinsen 1973). During folding,

hydrophobic amino acids collapse to a hydrophobic core, thereby limiting the accessible conformations for the remaining parts of the domain (Dinner et al. 2000). Subsequently, the search for the native conformation is aided by additional energetically favorable interactions such as hydrogen bonds, ionic and Van der Waals interactions. It is currently imagined that each protein explores a unique funnel-shaped potential energy landscape during its folding process (Figure 2) (Dill & Chan 1997).

However, larger domains (>~100 amino acids) and proteins with more complex topologies often fold slowly or inefficiently (Hartl et al. 2011; Balchin et al. 2016). Furthermore, proteins are only marginally stable in their native state and are prone to unfolding, and - in more severe cases - aggregation (Pace et al. 1981), (Figure 2). Protein aggregates are nonfunctional interactions between protein molecules, which typically result from the exposure of hydrophobic patches on non-native proteins (Dobson 2003). Aggregates sequester functional proteins, and can also induce “gain of function” toxic effects (described in more detail in paragraph 2.6) (Kim et al. 2013).

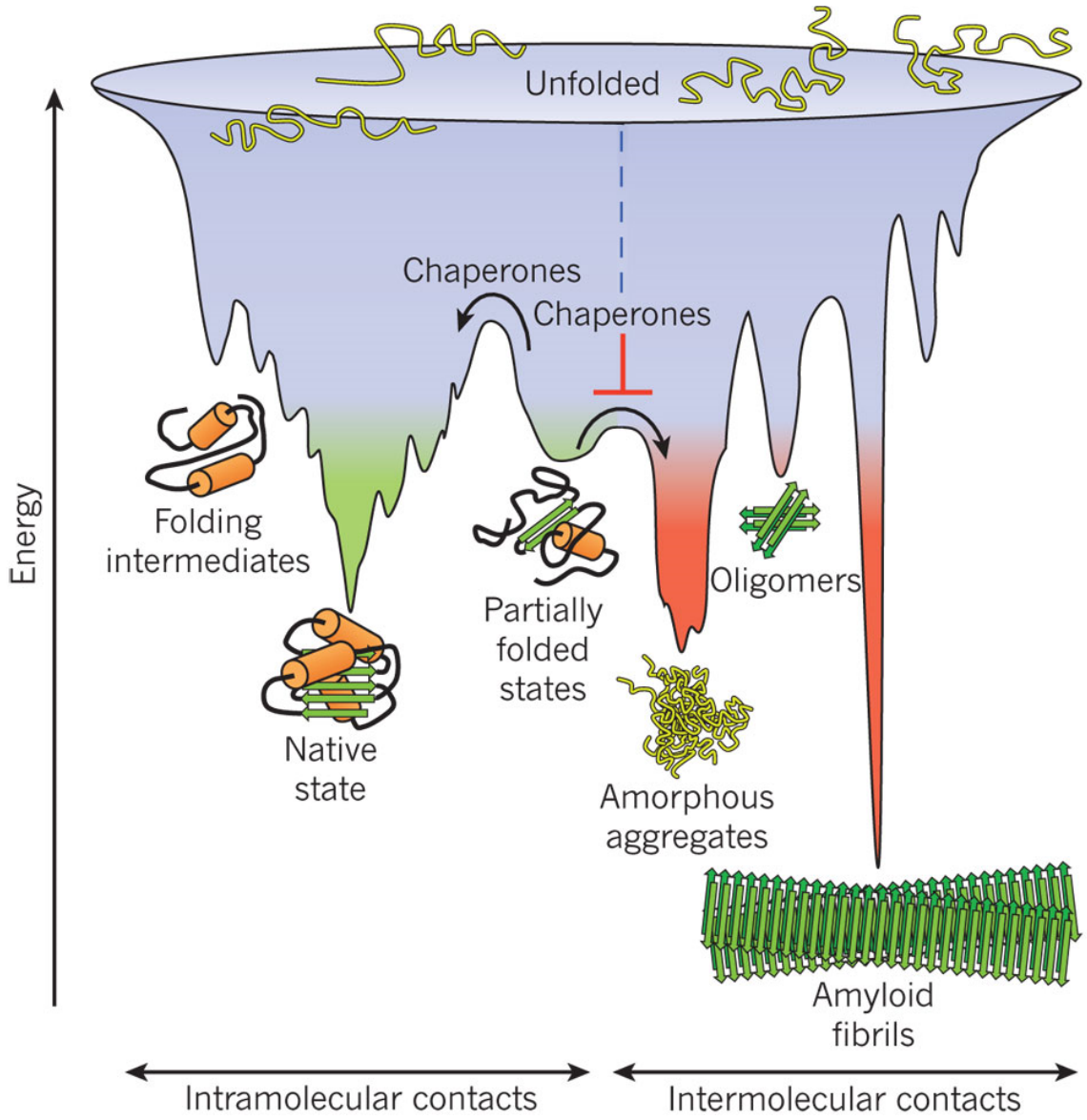

Figure 2 Funnel shaped free-energy landscape of protein folding and aggregation. Scheme of the multiple conformations of polypeptide chains on their conformational search towards their native state (green). Molecular chaperones may assist polypeptide chains at crossing free-energy barriers to reach their native conformations. The formation of amorphous aggregates, toxic oligomers and highly ordered amyloid fibrils (red) results from intermolecular contacts between folding intermediates or misfolded proteins, which are normally prevented by molecular chaperones. Reprinted by permission from Macmillan Publishers Ltd: Nature (Hartl et al. 2011), copyright 2011

2.2.2 The concept of molecular chaperones

To facilitate folding and prevent aggregation, cells express a specific subset of folding helper proteins – the molecular chaperones (Figure 1).The molecular chaperone concept was generally proposed for the first time in 1987 (Ellis 1987),

and chaperones have since been defined as “proteins that interact with or aid the folding or assembly of other proteins without being part of the target protein final structure” (Ostermann et al. 1989; Kim et al. 2013).

An increasing number of studies has shown that the chaperone machinery not only contributes to the initial folding of proteins, but is involved in maintaining a protein’s functional fold, as well as facilitating the disaggregation and degradation of misfolded proteins (Balch et al. 2008; Labbadia & R. Morimoto 2015; Hartl & Hayer-Hartl 2009).

2.2.3 Classes of molecular chaperones

Molecular chaperones cover a variety of functions and interact with proteins through their entire life time – from synthesis at the ribosome until their degradation at the proteasome or other pathways (Figure 1 and Figure 3) (Hartl et al. 2011). Because their expression is increased in challenging situations such as heat stress – which also lead to their discovery (paragraph 2.4.2) - chaperones are often referred to as heat shock proteins (Hsps) (Tissières et al. 1974).

Historically, molecular chaperones have been classified according to their monomeric molecular weight in kDa. Several classes have been discovered: small Hsp (<30 kDa), Hsp40, Hsp70, Hsp90, Hsp100. Additionally, chaperonins (Hsp60s) which are large, cylindrical molecular complexes (800-1000 kDa) forming an internal folding chamber, and ribosome associated chaperones have been described (Kim et al. 2013).

In all three kingdoms of life the chaperone pathways contributing to protein biogenesis and the *de novo* folding of proteins are organized according to the same principle (Balchin et al. 2016), (Figure 3). As the growing peptide chain emerges from the ribosome, it interacts with ribosome-associated chaperones for stabilization and co-translational initialization of the folding process. In case *de novo* folding requires the help of additional chaperones, the peptide chain is transferred to downstream chaperones such as the cytosolic Hsp70/Hsp40 and Hsp60 or Hsp90 protein families in an attempt to complete the folding process in a posttranslational manner (Hartl et al. 2011; Langer et al. 1992). In the following

section different classes of chaperones are described in brief, in the order they would interact with a nascent chain emerging from the ribosome.

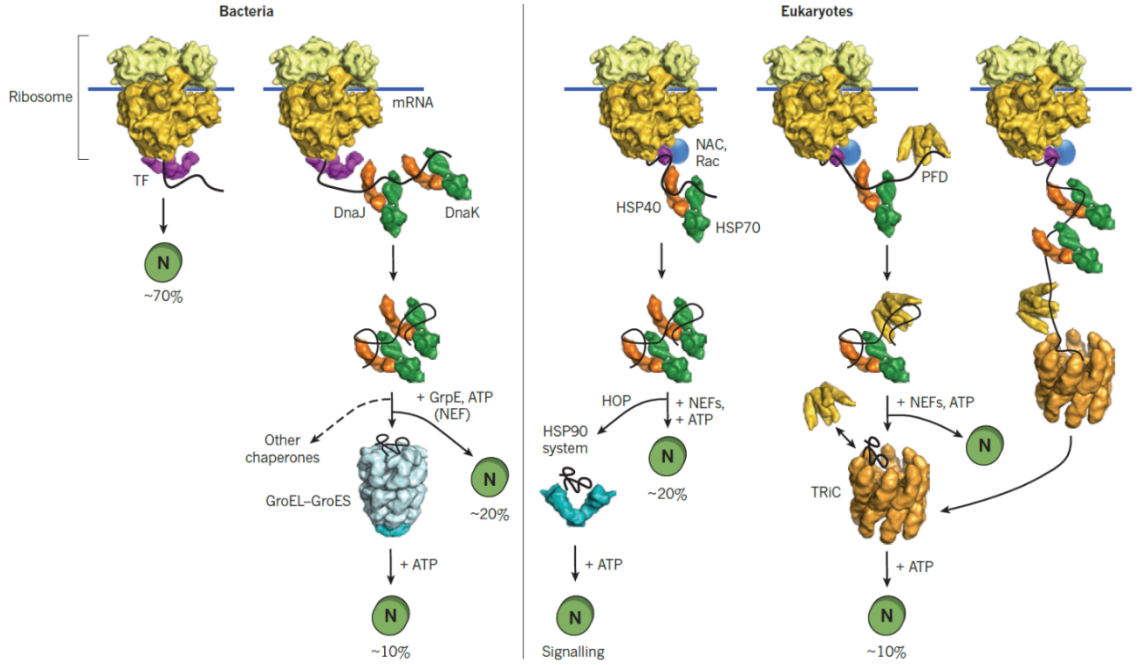

Figure 3 De novo protein folding pathways in eubacteria and eukaryotes. Molecular chaperones interact with folding substrates in a conserved, hierarchical manner: The emerging peptide chain is initially bound by ribosome-associated chaperones for stabilization and folding initialization, and can be transferred to downstream chaperone subsets (Hsp70, Hsp60 (GroEL or TRIC), Hsp90) for completion of the folding process if required. Reprinted by permission from Macmillan Publishers Ltd: Nature (Hartl et al. 2011), copyright 2011

2.2.3.1 Ribosome associated chaperones

Proteins are synthesized as linear polypeptides on ribosomes, where they are already able to form α-helices or small domains (e.g. zinc finger domains) in the ribosome exit tunnel (Nilsson et al. 2015; Holtkamp et al. 2015). During translation, the nascent polypeptide exposes hydrophobic segments to the surrounding environment. In order to avoid aggregation at this point, several ribosome-associated chaperones have evolved in prokaryotes and eukaryotes (Figure 3). At the bacterial ribosome exit tunnel a single chaperone termed trigger factor (~48 kDa, three domains) binds to highly hydrophobic and positively charged stretches, whereas a subset of specialized chaperones associates with translating ribosomes in eukaryotes (Ferbitz et al. 2004). In mammals, the protruding nascent chain is bound by the nascent polypeptide-associated complex (NAC) and the

(mammalian) ribosome-associated complex ((m)RAC) when exposing hydrophobic stretches and disordered fragments. mRAC is a specialized Hsp70-Hsp40 complex (Hsp70L1 and MPP11) which binds directly to the ribosome (Otto et al. 2005). NAC is a heterodimer comprising an α-subunit and a β-subunit and also binds directly to the 60S ribosome via the β-subunit (Wang et al. 1995; Preissler & Deuerling 2012). Thus, a fraction of proteins and emerging domains in multi-domain proteins completes folding on the ribosome. The ribosome associated machinery hands over the remaining substrates to downstream chaperone pathways involving Hsp70s, Hsp90 and the chaperonins (Hartl et al. 2011; Scior & Deuerling 2014).

2.2.3.2 The Hsp70 system

Hsp70s are probably the most versatile chaperone subset (Calloni et al. 2012). They are involved in protein folding, refolding, protein transport and protein degradation pathways. The human Hsp70 family comprises 13 proteins which can be expressed constitutively (e.g. Hsc70) or inducibly (e.g. Hsp70-1) in response to environmental stimuli (Radons 2016).

Hsp70 proteins (DnaK in bacteria) consist of two domains: a nucleotide binding domain (NBD) and a substrate binding domain (SBD) which are connected via a conserved linker (Kim et al. 2013). The substrate binding and release cycle is allosterically regulated (Zuiderweg et al. 2013). The SBD binds substrate peptides in the open state, which is maintained by ATP binding to the NBD. As soon as ATP is hydrolyzed to ADP (accelerated by Hsp40s (bacteria: DnaJ)) in the NBD, allosteric rearrangement forces the SBD into its closed state, whereby the bound peptide is clamped between the β-sandwich domain and an α-helical lid domain (Kampinga & Craig 2010). Subsequently, nucleotide exchange factors (NEF) stimulate the release of ADP from the NBD, resetting the SBD to the open state and releasing the bound substrate (Figure 3) (Mayer & Bukau 2005).

Substrate proteins which cannot acquire their native conformation spontaneously, or with the help of ribosome associated chaperones or the Hsp70/Hsp40 system, are transferred to downstream chaperone machineries like Hsp90 or the chaperonin (Langer et al. 1992; Hartl et al. 2011).

2.2.3.3 Chaperonins in Pro- and Eukaryotes

Chaperonins are large homo- or hetero oligomeric double ring complexes (7-9 subunits per ring) which form a cavity within each ring for the encapsulation of substrate proteins (Hartl & Hayer-Hartl 2009). Chaperonins are classified as either group I or group II. Chaperonins of bacterial origin belong to class I, and archaeal and eukaryotic cytosolic chaperonins to group II (Lopez et al. 2015). Chaperonin group I subunits have a molecular weight of 60 kDa, and form heptameric rings (~800 kDa). GroEL, the chaperonin of bacteria, is the best studied homolog. Non-native substrates are thought to be delivered to GroEL mainly by the Hsp70/Hsp40 system (Figure 3) (Langer et al. 1992). Substrates bind to the apical domains of GroEL in the apo state. ATP binding triggers conformational rearrangements enabling binding of GroES, the heptameric lid-shaped co-chaperone of GroEL (Hayer-Hartl et al. 2016). GroES binding displaces the substrate into the GroEL-GroES cavity, which is then free to fold for the time required by the chaperonin to hydrolyze bound ATP. The release of GroES, ADP and substrate is allosterically triggered by ATP binding to the opposite ring. GroEL does not only act as passive cage that prevents aggregation, but actively accelerates folding of several model substrate proteins (Gupta et al. 2014; Georgescauld et al. 2014). Group II chaperonins include the thermosome in archaea and TRiC/CCT (tailless complex polypeptide-1 (TCP-1) ring complex/ chaperonin-containing TCP-1) in eukaryotes. TRiC is responsible for folding approximately 10% of the eukaryotic proteome, including the abundant cytoskeletal proteins, actin and tubulin (Hartl et al. 2011). TRiC is a hetero-oligomeric ~1 MDa double-ring complex made up of 8 different subunits per ring. Each ring has a built-in, iris-like lid which closes upon ATP hydrolysis to encapsulate the substrate. Despite its physiological significance, how TRiC recognizes and folds its substrates remains poorly understood (Lopez et al. 2015; Hayer-Hartl et al. 2016).

2.2.3.4 The Hsp90 system

Hsp90 chaperones play an important role in the maturation and conformational maintenance of many signaling proteins in the cell (Figure 3). The human Hsp90 family comprises 17 proteins which can be expressed constitutively (e.g. HSP90AB1) or inducibly (e.g. HSP90AA1) upon environmental stimuli (Chen et

al. 2005). Hsp90 is a flexible dimer whereby each subunit is comprised of an N-terminal ATPase domain, a substrate-binding middle domain and C-terminal dimerization domain. Hsp90 undergoes an ATP dependent cycle which is characterized by an open, nucleotide-free state and a closed, ATP bound conformation (Rehn & Buchner 2015). Hsp90 and its reaction cycle are tightly regulated by various co-chaperones. HOP and Cdc37 stabilize the open, substrate binding conformation of Hsp90, whereas Aha1 stimulates ATP hydrolysis and formation of the closed state (Li et al. 2012). The closed state is stabilized by the co-chaperone p23. Substrate folding presumably occurs in the closed state of Hsp90, however the detailed mechanism of substrate binding and folding is not clear yet (Kim et al. 2013).

2.3 Protein degradation systems in the cell

2.3.1 UPS and autophagy in protein degradation

Proteins that cannot be folded or refolded by the machinery of molecular chaperones have to be removed from the cellular proteome in order to avoid interference with functionally intact proteins. Three major pathways have been discovered to do so: the ubiquitin-proteasome system, chaperone mediated autophagy and macroautophagy (Figure 4), (Ciechanover & Kwon 2015).

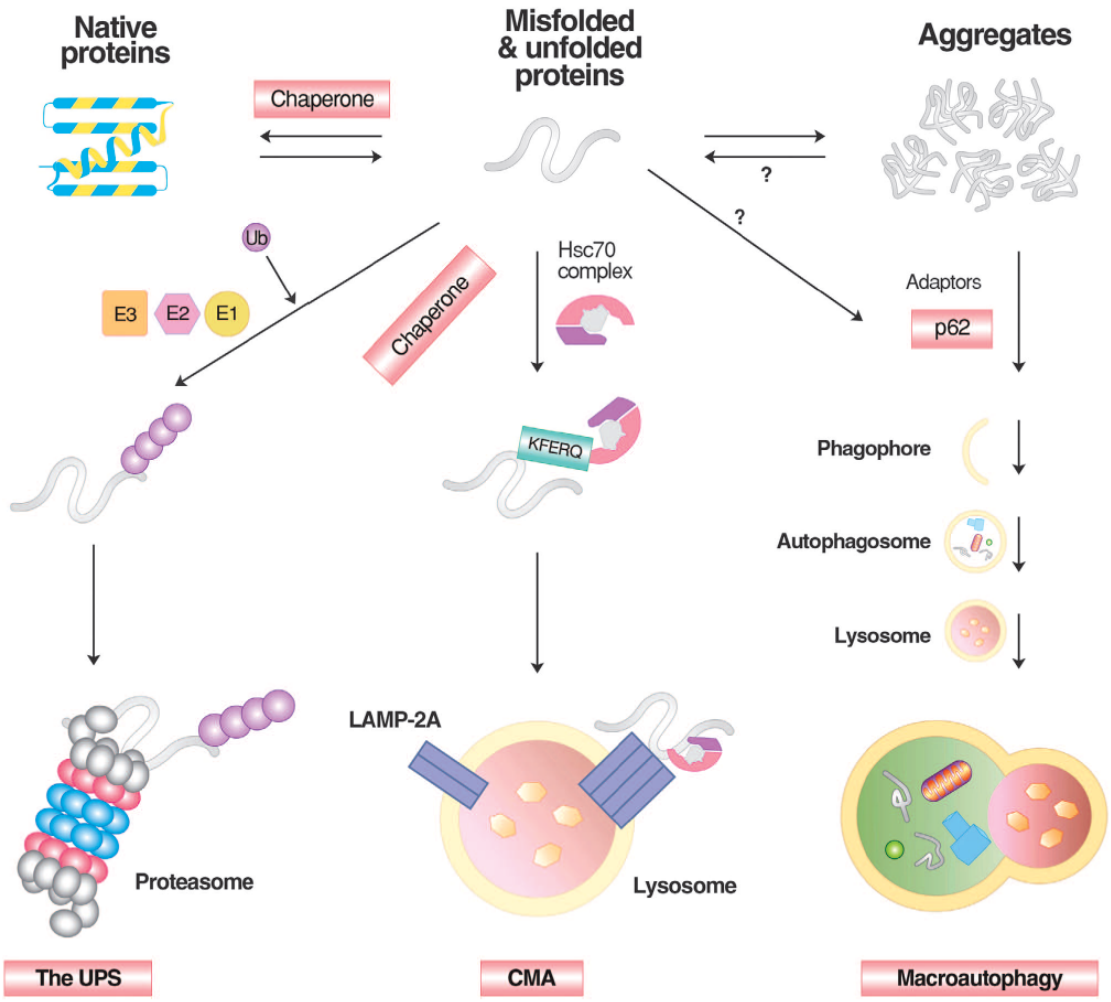

Figure 4 Protein degradation pathways. Proteins can be degraded via the ubiquitin proteasome pathway (UPS), chaperone mediated autophagy (CMA) and macroautophagy depending on their nature. In CMA, substrates exposing the KFERQ motif are bound by Hsc70. The substrates are subsequently delivered to the lysosomal membrane, translocated to the lumen, and degraded by lysosomal hydrolases. Aggregates are directed to macroautophagy. Misfolded protein substrates of macroautophagy are recognized by molecular chaperones such as Hsc70, ubiquitinated, and delivered to the autophagic adaptor p62, and subsequently delivered to autophagic membranes for lysosomal degradation. Reprinted from (Ciechanover & Kwon 2015)

Monomeric unfolded proteins with a high turnover are usually degraded by the ubiquitin-proteasome system (UPS). The terminally misfolded substrate is recognized by chaperones (e.g. Hsc70, Bcl-2–associated athanogene (BAG1)) and ubiquitin ligases (e.g. C-terminus of HSC70-Interacting Protein (CHIP)) if hydrophobic residues are exposed or abnormal folds (e.g. mislinked side chains) are detected (McDonough & Patterson 2003). Subsequently, a cascade of ubiquitin ligases (from E1 to E3) transfers ubiquitin molecules to the substrate

protein, which encode a signal for degradation if the ubiquitins are linked via Lys48 and the chain is at least four units long. The poly-ubiquitinated substrate is subsequently recognized by the regulatory subunit of the proteasome, the 19S particle, and finally shredded in the 20S core particle into short peptides (Ciechanover & Kwon 2015).

A second pathway, chaperone mediated autophagy, is a selective proteolytic pathway which recognizes misfolded proteins carrying a target sequence (Kaushik & Cuervo 2012). The constitutive version of Hsp70 (Hsc70) recognizes the target sequence KFERQ, which is only surface exposed in the non-native state (~30 % of cytosolic proteins carry the sequence motif). Next, Hsc70 delivers the misfolded substrate to the lysosome for import via LAMP-2A (lysosomal membrane-associated protein 2A) and subsequent lysosomal degradation by hydrolases.

An alternative autophagy pathway, termed macroautophagy, describes a process whereby larger aggregates and whole organelles are targeted to autophagosomes, which later fuse with lysosomes for degradation (Ciechanover & Kwon 2015). Targeting of ubiquitinated and aggregated proteins to phagophores, the precursors of autophagosomes, is achieved via the autophagosomal adaptor p62 which binds ubiquitin chains and aggregates via different domains.

2.4 Stress-responsive pathways

2.4.1 Challenges to proteostasis

Under steady state conditions, the proteostasis machinery ensures correct folding and timely degradation of cellular proteins. However, if the cell experiences proteotoxic stress – like the presence of heavy metals, heat stress, aging, non-functional mutant proteins or oxidative stress – the proteostasis machinery has to have the capacity to buffer the effects (Figure 5), (Balchin et al. 2016).

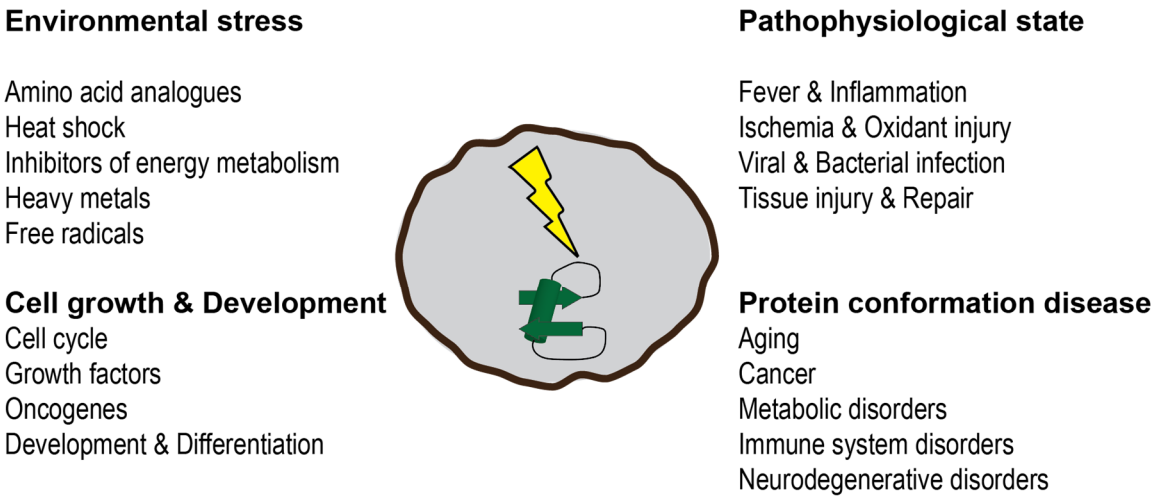

Figure 5 Challenges to the proteostatic equilibrium. A variety of stresses challenges the capacity of the proteostasis network by formation of an increasing amount of non-native proteins. Stress inducing conditions are grouped into four major categories: Three classes of environmental and physiological stresses (environmental stress, pathophysiological state, protein conformation disease) and a fourth class that summarizes intrinsic stimuli during cell growth and development. Adapted from (Morimoto 2008).

To conserve cellular resources, the proteostasis machinery is not adjusted to its highest functional capacity under normal conditions (Morimoto 2008). However, cells must be able to react and adapt rapidly to the changing environment. Cells have therefore evolved several stress response pathways which are compartment-specific (Kourtis & Tavernarakis 2011): the cytosolic stress response, and the unfolded protein response of the endoplasmic reticulum and mitochondria. Since not only proteins can be damaged by different stressors, a nuclear DNA damage response has also evolved.

2.4.2 Cytosolic stress response and Hsf1

Upon proteotoxic stress, every living cell of all three kingdoms of life responds with an evolutionarily conserved defense program to protect itself (Lindquist 1986). This transcriptional response was first described in the 1960s when the geneticist Ritossa discovered that polytene chromosomes in salvary glands of *Drosophila melanogaster* showed a different puffing pattern (corresponding to transcriptional activity) after he had accidently shifted the larvae to higher temperature (Ritossa 1962).

In the following decades the genes which are activated during the heat shock response were identified and termed heat shock protein (Hsp) genes (Tissières et al. 1974). As described above, many of these proteins function as molecular chaperones and are also constitutively expressed as abundant proteins to maintain proteostasis during both stress and normal conditions (paragraph 2.2.2.). Since proteins are the cell's primary molecular machines, their maintenance is of high importance and therefore it is not surprising that the expression of Hsp genes is tightly regulated.

In eukaryotes, the genes encoding components of the cytosolic heat-shock response are under the control of promoters containing characteristic conserved heat shock elements (HSE), consisting of inverted nGAAn sequence motifs (Pelham 1982). The multidomain transcription factors which bind to these promoters harboring HSEs were termed heat-shock transcription factors (Hsf) (Parker & Topol 1984; Wu 1984; Wu 1995).

Prokaryotes such as *E. coli* developed an alternative mechanism for the inducible expression of Hsp, involving a RNA polymerase II subunit termed σ^{32} (Neidhardt & VanBogelen 1981). σ^{32} is increasingly translated during heat stress and directs Pol II to its target genes (Nonaka et al. 2006).

Whereas in yeast, invertebrates, nematodes and fruit flies a single Hsf (Hsf1) is responsible for inducing the heat shock stress response, vertebrates possess a whole family of homologous Hsfs (Akerfelt et al. 2010). Humans express at least three different Hsf - Hsf1, Hsf2 and Hsf4 – which fulfill different tasks. Hsf2 has shown to play a critical role in brain development, spermatogenesis and basal chaperone transcription, but similar to Hsf4 – which has been shown to contribute to development and basal chaperone transcription – is much less well characterized than Hsf1, the master regulator of the heat-shock response (HSR). Hsf1 is the vertebrate functional counterpart to the single Hsf in invertebrates. It is the only stress inducible regulator in the family and its function cannot be compensated by Hsf2 or Hsf4 (Anckar & Sistonen 2011).

Hsf1 is the best-characterized member of the vertebrate Hsf family and is ubiquitously expressed in human body tissues (Uhlén et al. 2015). It has been reported to be primarily located in the nucleus, with the ability to shuttle between

nucleus and cytoplasm (Mercier et al. 1999). During non-stress conditions, Hsf1 assumes a monomeric state, which is inactive and does not bind to HSEs. Hsf1 shifts to its active, oligomeric (probably trimeric) form upon proteotoxic stress (Figure 6) (Sarge et al. 1993).

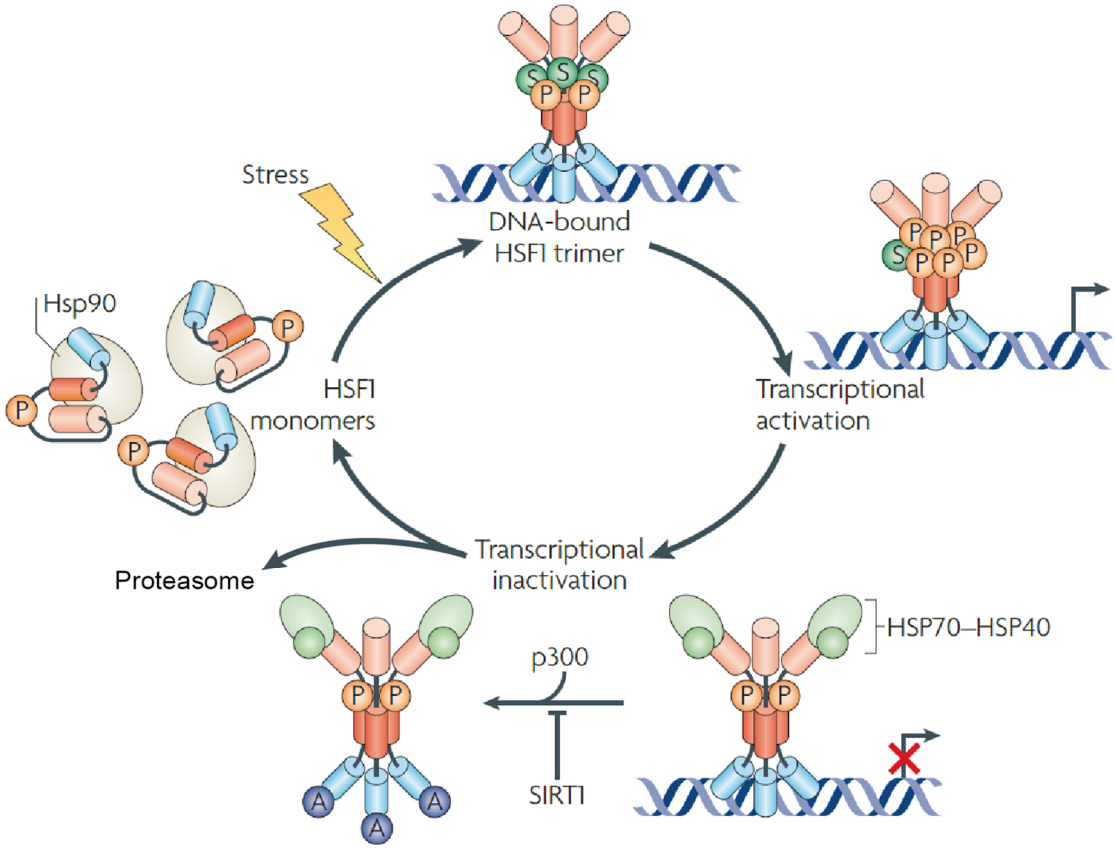

Figure 6 Current Hsf1 activation model. Hsf1's monomeric state is stabilized by molecular chaperones in the monomeric state. Upon proteotoxic stress, Hsf1 trimerizes, accumulates in the nucleus and binds to HSE in promoters of target genes and activates transcription. Activity is additionally regulated by PTMs. Hsf1 is deactivated by further PTMs and dissociates or is degraded. Adapted by permission from Macmillan Publishers Ltd: Nat. Rev. Mol. Cell Bio.(Akerfelt et al. 2010), copyright 2010

A variety of chaperones such as TRiC, Hsp70 and Hsp90 bind Hsf1 during non-stress conditions and thereby stabilize the monomeric state (Neef et al. 2014). Upon stress, Hsf1 is thought to be displaced from its stabilizing chaperones by increasing amounts of unfolded or misfolded proteins, and is able to adopt its active, oligomeric form (Zou et al. 1998; Neef et al. 2014). The Hsf1 oligomer binds cooperatively and with high affinity to HSE-containing promoters of target genes (Xiao et al. 1991). Although initially thought to be only involved in regulating the

heat shock response, recent studies have discovered that Hsf1 is involved in a diverse signaling network of ~1500 genes ranging from protein folding to cell cycle regulation (Trinklein et al. 2004; Page et al. 2006).

Hsf1 undergoes extensive post-translational modification such as phosphorylation, acetylation and phosphorylation-dependent sumoylation (Xu et al. 2012). Whereas the role of phosphorylation in Hsf1 function has largely remained enigmatic, differential acetylation appears to regulate Hsf1 stability and promoter binding (Budzyński et al. 2015; Raychaudhuri et al. 2014).

Besides HSEs, Hsf1 binds to repeats of Satellite III DNA upon heat shock, giving rise to nuclear stress bodies (Grady et al. 1992). These Satellite III repeats are found in the central region on chromosome 9. This region has been shown to form “satellite” bands when human DNA is separated in a density gradient, indicating a different base composition compared to normal DNA. It has been found that this region is comprised of long repeats of a 5 bp motif, GGAAT – similar to those found in HSEs. These sites have been shown to be the origin of non-coding RNA transcripts of unknown function (Biamonti & Vourc’h 2010).

2.4.3 Organelle-specific stress response of the endoplasmic reticulum

Prokaryotes do not possess membrane-enclosed organelles and therefore have only one type of stress response to protect the proteins of their cytoplasm. Eukaryotic cells are subdivided into different, membrane-enclosed organelles such as the nucleus, endoplasmic reticulum, Golgi apparatus and mitochondria - each harboring a different set of proteins. Therefore eukaryotes do not only have the cytosolic stress response (as described above) but have also developed organelle-specific stress responsive pathways (Kourtis & Tavernarakis 2011). The endoplasmic reticulum (ER) is an important compartment in which secretory proteins are folded, post-translationally modified and undergo quality control before their delivery to other membrane compartments. Eukaryotic cells have developed three branches of a response to unfolded and misfolded proteins in the ER – termed the unfolded protein response (UPR). The UPR comprises the IRE1 (inositol requiring enzyme 1), the PERK (double-stranded RNA-activated protein

kinase-like ER kinase) and the ATF6 (activating transcription factor 6) mediated pathway (Walter & Ron 2011). The three branches are partially redundant, and are activated upon distinct stress signals.

IRE1 mediates the best-studied branch of the UPR (Walter & Ron 2011). IRE1 is a bifunctional transmembrane kinase/endoribonuclease which contains an ER luminal domain that is able to detect unfolded proteins. Upon an increase in the amounts of unfolded proteins in the ER, IRE1 oligomerizes and the cytoplasmic domain phosphorylates itself, activating IRE1. Active IRE1 splices the mRNA of the UPR-specific transcription factor XBP1 into the alternative product XBP1s (Yoshida et al. 2001). XBP1s acts as transcription factor for proteins involved in translation, protein folding and degradation as well as lipid synthesis. ER-associated protein degradation (ERAD) is executed by retro-translocation into the cytosol followed by proteasomal degradation (Vembar & Brodsky 2008).

The second branch of the system involves PERK, an ER-resident transmembrane kinase which oligomerizes upon exposure to unfolded proteins in the ER (Ron & Walter 2007). Oligomeric PERK phosphorylates itself and eIF2α, which induces the translation of the transcription factor ATF4. ATF4's target genes include genes for protein folding, antioxidant response proteins, apoptosis regulating genes and transcription factors.

The third branch of the UPR is mediated via ATF6. ATF6 is a transcription factor, which harbors a transmembrane domain (Ron & Walter 2007). It is inserted into the ER membrane and has an ER luminal part, which is able to sense unfolded proteins in the ER. Upon activation, ATF6 is transported to the Golgi apparatus, where it is cleaved by two proteases. The N-terminus of ATF6 (ATF6(N)) is released as a transcription factor, which activates protective target genes in the nucleus.

2.5 Gene regulation mediated by transcription factors

When the atomic structure of the DNA-double helix was solved and the short-lived messenger was discovered to be RNA (Watson & Crick 1953; Weiss & Gladstone 1959), an intriguing question remained: how is gene activity regulated? How do proteins interact with DNA and the genome? Today, substantial progress has

been made on the details of these interactions. Large groups of proteins interact with DNA for genome maintenance, replication and transcription. Thousands of proteins – in humans around 10% of the proteome - have been identified to be DNA interacting proteins, which act as transcription factors (Vaquerizas et al. 2009). Transcription factors (TF) are proteins which bind DNA in a sequence-specific manner and direct transcription initiation to specific promoters, but do not have enzymatic function or belong to the RNA polymerase II (Pol II) initiation complex (Vaquerizas et al. 2009). TF carry out a great variety of functions in many cellular processes such as metabolism, response to stimuli and development.

2.5.1 Role of transcription factors in transcription initiation

2.5.1.1 Initiation of transcription requires formation of the preinitiation complex

To initiate the transcription of genomic DNA to mRNA, a specific TF binds to a DNA-sequence motif upstream (e.g. 150 bp) of the transcription start site (TSS) of its target genes (Dynan & Tjian 1983). Accessibility of these upstream regions – termed promoters – is regulated by the chromatin structure (Guertin & Lis 2010). After binding to DNA, the TF trans-activation domain recruits other proteins such as general coactivators and adaptor proteins, which facilitate the recruitment of general transcription factors for the assembly of the preinitiation complex of transcription (PIC) for RNA Polymerase II (Pol II) (Sainsbury et al. 2015). Gene regulation is highly complex and tightly regulated by a great variety of proteins and processes. Not only specific TF, general TF and co-activators and the Pol II play an important role, but also chromatin structure and histone exchange (Venkatesh & Workman 2015). An additional layer of complexity is added by a myriad of post-translational modifications on proteins directly involved in transcription and chromatin structure. Therefore only an outline of transcription regulation is described in the following.

In the first step leading to the PIC, TATA-box binding protein (TBP) which is part of the general transcription factor TFIID complex (several TBP-associated factors (TAF)) binds the TATA-box within in the promoter (Figure 7), (Sainsbury et al. 2015).

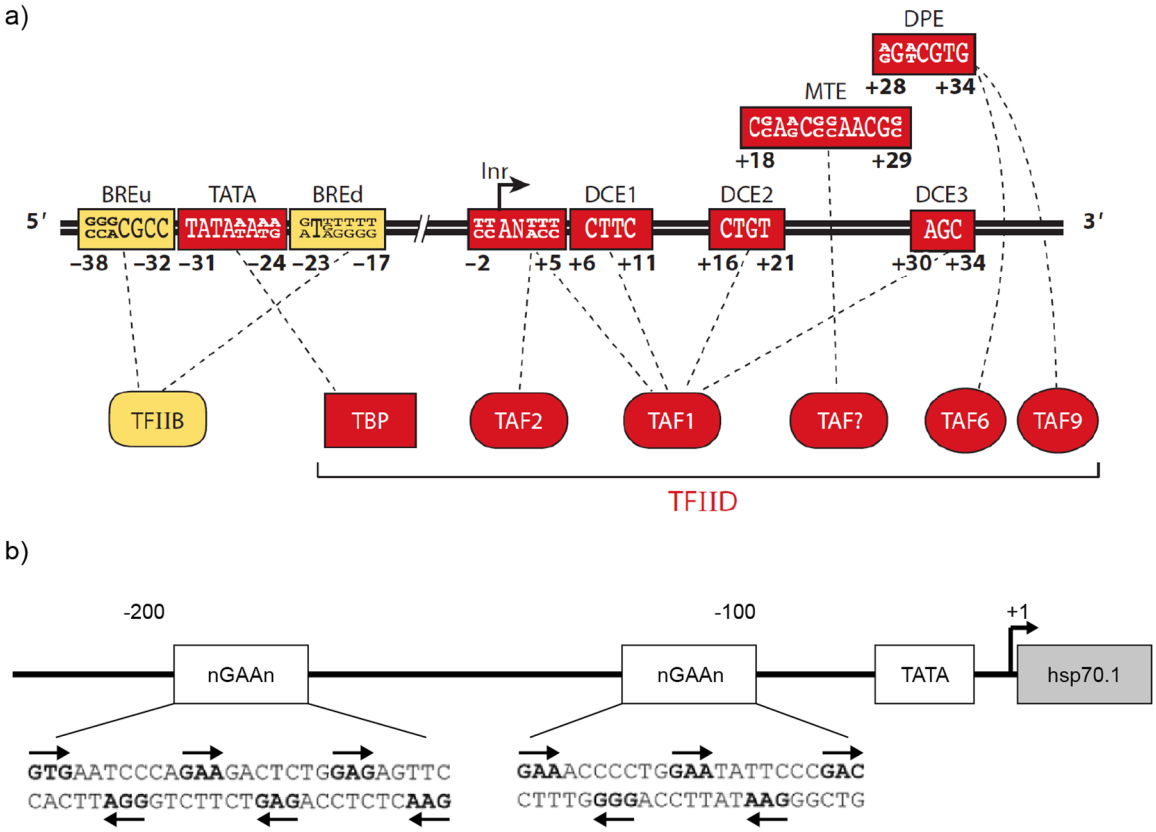

Figure 7 Eukaryotic promoter architecture. a) General architecture of eukaryotic promoters with binding sites for general transcription factors. Details can be found in the text. b) Architecture of the human hsp70.1 promoter with distal and proximal HSE motifs. These are followed by the generic elements of eukaryotic promoters such as TATA-box and the transcription start site (+1). Adapted from (Decker & Hinton 2013; Hietakangas & Sistonen 2006)

TBP inserts into the minor groove of the DNA causing a bend of ~90 degrees (Figure 8). Subsequently, TFIIA and TFIIB are recruited to the complex. TFIIA stabilizes the TBP-DNA complex and can stimulate transcription. TFIIB binds to the complex and to TFIIB recognition elements (BRE) which are up- and downstream of the TATA-box. TFIIB is required to recruit and activate Pol II. Subsequently, Pol II bound to TFIIF binds to the preinitiation complex. Upon binding of TFIIE and TFIIH, which has an ATP-dependent helicase activity, the preinitiation complex is complete and transcription can initiate under consumption of ATP and nucleotide triphosphates (NTP) (Sainsbury et al. 2015).

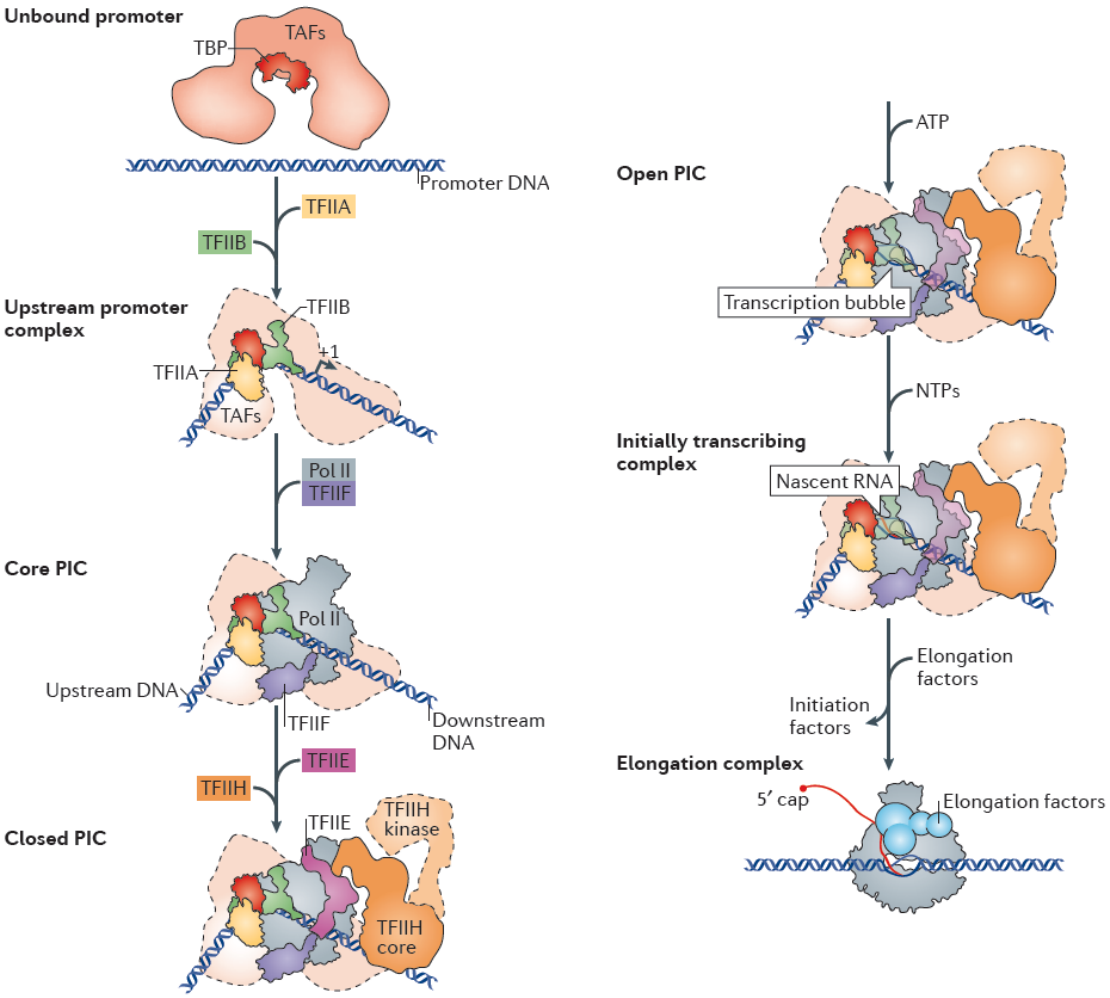

Figure 8 Transcription initiation by general transcription factors, the PIC and RNA Polymerase II. Shown is the canonical model for stepwise pre-initiation complex (PIC) assembly from general transcription factors (various colors) and RNA polymerase II (Pol II; grey) on promoter DNA. The intermediate complexes are labelled. Details can be found in the text. Adapted by permission from Macmillan Publishers Ltd: Nat. Rev. Mol. Cell Bio. (Sainsbury et al. 2015), copyright 2015

2.5.1.2 Productive elongation is dependent on elongation factors

Transcription elongation requires another set of proteins, elongation factors, to avoid pausing of Pol II in promoter-proximal regions (Jonkers & Lis 2015). In most cases, Pol II pauses immediately after initiation after transcribing ~25-50 bp, due to the influence of DRB sensitivity-inducing factor (DSIF) and negative elongation factor (NELF). For elongation, positive transcription elongation factor b (P-TEFb (which has a Cdk9 subunit)) has to be recruited (e.g. by a TF such as Hsf1) which

phosphorylates DSIF, NELF and the Pol II C-terminal domain. Thus, DSIF and NELF dissociate from Pol II and the Pol II continues transcription (Figure 9).

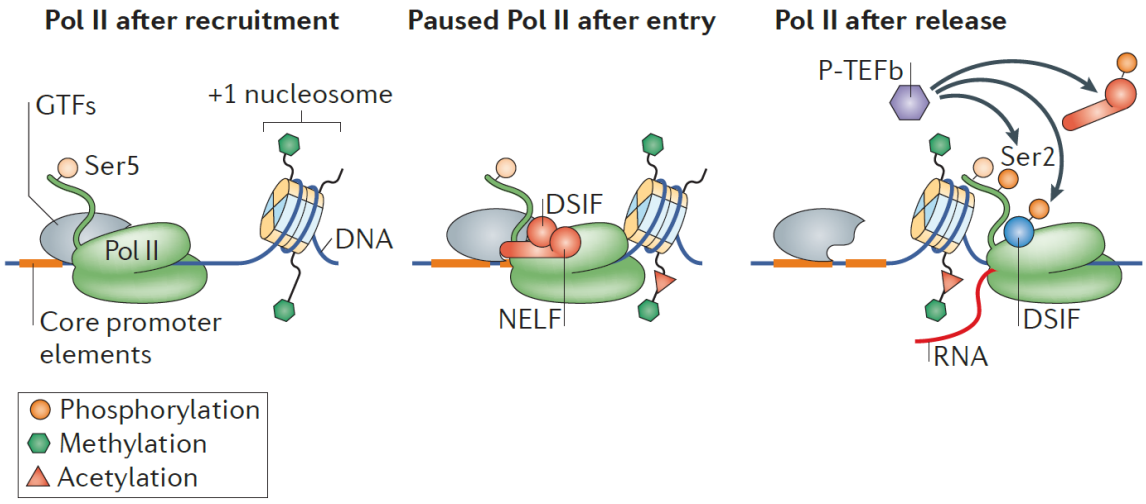

Figure 9 Pol II pausing and release in transcription. Pol II requires elongation factors for continued transcription. Pol II pausing is caused by negative elongation factor (NELF) and DRB-sensitivity-inducing factor (DSIF), with contribution from core promoter elements and the +1 nucleosome. Positive transcription elongation factor-b (P-TEFb) releases paused Pol II by phosphorylating NELF, DSIF and the carboxy-terminal domain (CTD) of Pol II. Adapted by permission from Macmillan Publishers Ltd: Nat. Rev. Mol. Cell Bio. (Jonkers & Lis 2015), copyright 2015

2.5.1.3 Trans-activation domains of TFs act as recruitment agents for general transcription factors

Specific transcription factors have been shown to directly recruit TFIID and other general transcription or elongation factors (Sainsbury et al. 2015). A specific type of TF trans-activation domain defined by a pattern of nine amino acids (9aaTAD), which is also present in Hsf1 (amino acid 412-420), was shown to interact with the general transcriptional coactivator TAF9, the acetyltransferase and adaptor family p300-CBP and MED15, a subunit of the mediator complex (Piskacek 2009). The “mediator of transcription” complex is a large, ~20 subunit protein complex that interacts with activators/repressors of transcription and regulates the assembly of the preinitiation complex (Allen & Taatjes 2015). Furthermore, successful binding of general co-factors can only occur if the promoter is not tightly wrapped around the histones on a nucleosome. Therefore the histone acetyltransferase function of e.g. CBP or p300 can be essential to acetylate histones and therefore make the

promoter accessible to co-factors of the preinitiation complex (Jonkers & Lis 2015).

2.5.2 Common structural motifs for specific transcription factors

TF do not only read out the base sequence of the major groove of B-DNA, which offers a sequence specific hydrogen bond donor/acceptor or other group for interaction with the TF, but also utilize the base readout of the minor groove or global or local DNA shape (Figure 10) (Rohs et al. 2010).

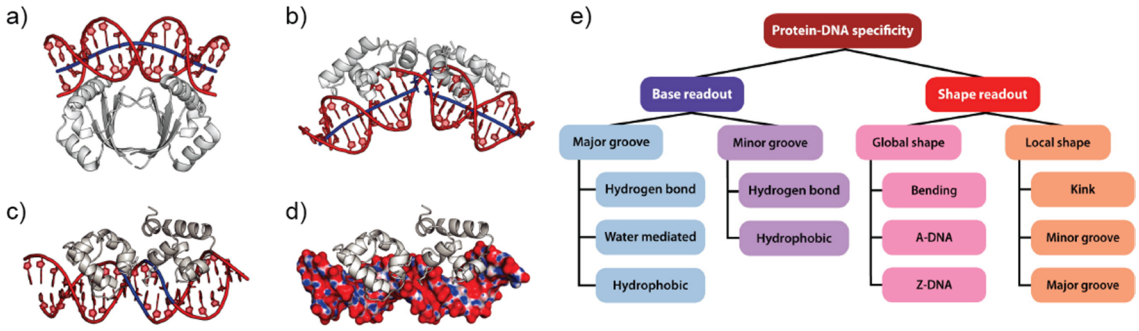

Figure 10 Transcription factors read base composition and/or shape of DNA double helix. a) The viral protein HPV-18 E2 causes the DNA helix to bend, b) Lac repressor causes a kink in the DNA, Phage 434 repressor probes conformations (c)) and electrostatic potential (d)) of the DNA minor groove. Adapted from (Rohs et al. 2010).

The database of Structural Classification of Proteins (SCOP) reports about ~70 SCOP super families for DNA-binding proteins based on atomic structures solved so far (Andreeva et al. 2007; Fox et al. 2014). Amongst these, Cys2-His2-Zincfinger and Helix-turn-helix DNA-binding domains (DBD) are the most common three dimensional classes (Figure 11). Together with Helix-loop-helix DNA-binding domains, they comprise about 80 % of the determined repertoire (Rohs et al. 2010; Vaquerizas et al. 2009).

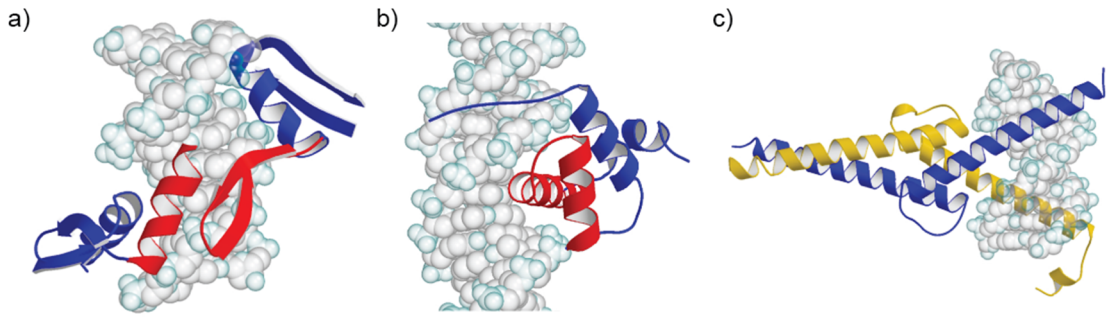

Figure 11 The three most common folds of DNA-binding domains. a) Zinc finger (1AAY), b) Homeodomain (Helix-turn-helix) (1FJL), c) Helix-loop-helix (1AM9). Adapted from (Luscombe et al. 2000)

The Zinc finger DNA-binding domain consists of a short α-helix which is inserted into the major groove of the DNA and reads out the base composition, and a two-stranded antiparallel β-sheet (Luscombe et al. 2000). The name derives from a Zn^{2+} ion which is coordinated by Histidine and Cysteine residues between the α-helix and the β-sheets. The Zinc finger is defined by its Cys-His motif (e.g. Cys2-His2, Cys4, Cys6).

Helix-loop-helix domains consist of a short α-helix which reads out the base sequence in the major groove of the DNA, and a longer α-helix which is able to form a dimeric leucine zipper. The two helices are connected via a loop.

Helix-turn-helix (HTH) structures are of the most frequently observed structural motifs for DNA-binding domains (Rohs et al. 2010). The helix-turn-helix fold can be subclassified depending on the number of helices present (2-4) and other structural features such as a large incorporated wing-loop (paragraph 2.5.3.1). The main feature of this particular fold is two helices which are orientated perpendicular to each other, whereby the first α-helix is inserted into the major groove of B-DNA for a base specific read out (recognition helix), and the second exhibits generic contacts to the DNA stabilizing the orientation of the first helix (Brennan & Matthews 1989). The two helices are connected via a short turn. The best-studied example for HTH domains are homeodomain DNA-binding domains which consist of three α-helices.

2.5.3 Structural details of the Heat shock transcription factor family and Hsf1

Human Hsf1 is a 57 kDa multidomain protein (Rabindran et al. 1991). It comprises a DNA-binding domain (DBD), a coiled-coil oligomerization domain, a trans-activation regulator domain (RD) and a transactivation domain (TAD) (Figure 12).

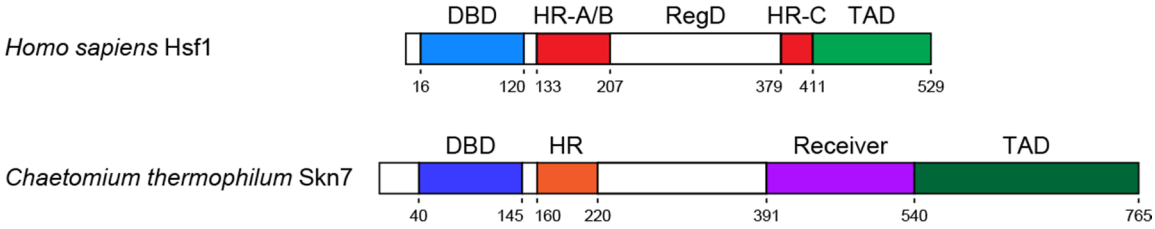

Figure 12 Domain architecture of human Hsf1 and the fungal paralog *C. thermophilum* Skn7.

2.5.3.1 The Hsf DNA-binding domain

The DNA-binding domain of heat-shock factors belongs to the family of HTH domains (Harrison et al. 1994). It comprises three major α-helices and a short, four-stranded β-sheet whereby the connector between β-sheet 3 and β-sheet 4 is a long loop termed “wing” (Figure 13).

Hsf HTH domains therefore belong into the subfamily of winged helix-turn-helix (wHTH) motifs (Harrison et al. 1994). In many wHTH domains, the wing contacts the DNA and reads out the base composition of the minor groove, whereas Hsf-type wHTH domains do not use the wing for protein-DNA contacts, but rather protein-protein contacts (Littlefield & Nelson 1999).

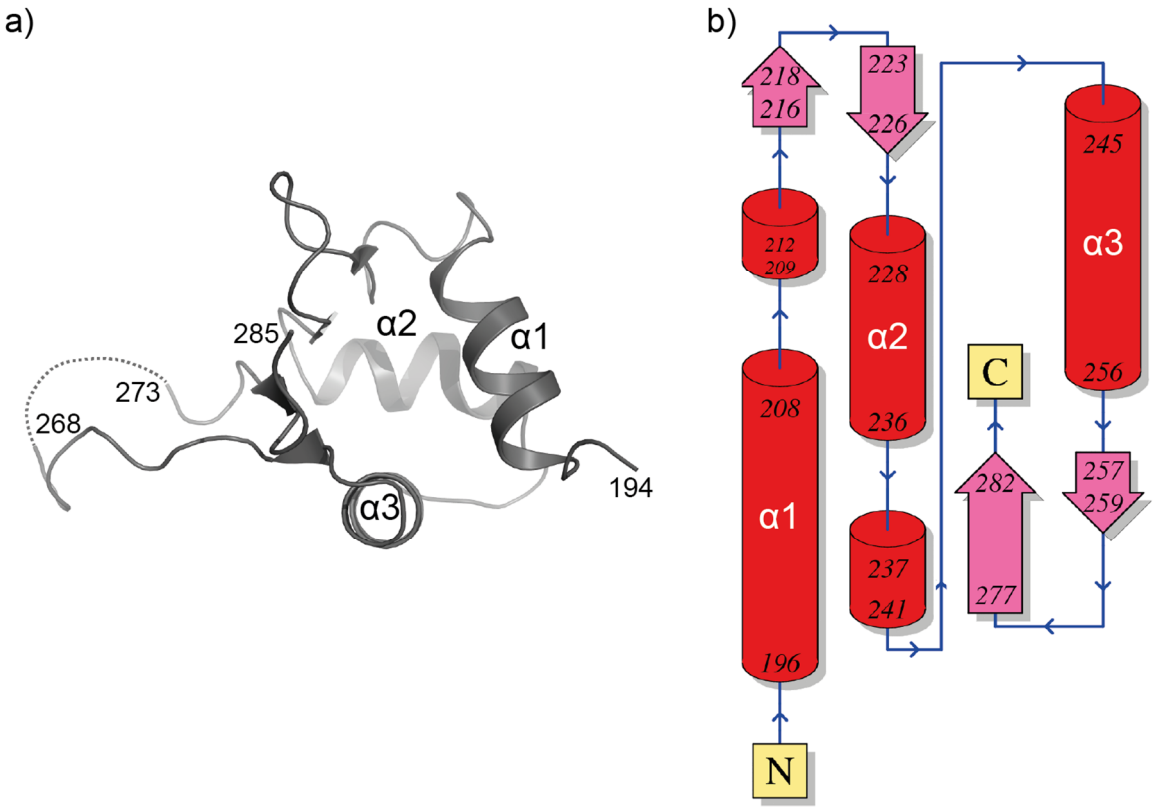

Figure 13 Domain topology of Hsf-type DBDs. a) First published Hsf1 DBD crystal structure from *Kluveromyces lactis* (including residues 194-282, PDB 2HTS).(Harrison et al. 1994) α-helix 3 inserts into the major groove of the DNA double helix. b) Domain topology diagram of *K. lactis* Hsf1 DBD (2HTS).(de Beer et al. 2014) Hsf DBD C-terminus is missing (aa 282-299).

Structural details of Hsf1 DBDs have been resolved using nuclear magnetic resonance spectroscopy (NMR) and macromolecular protein X-ray crystallography. Structures for isolated Hsf DBDs have been published from the fungus *Kluveromyces lactis* (Figure 13), *Drosophila melanogaster* and *Homo sapiens* (Liu et al. 2011; Harrison et al. 1994; Vuister et al. 1994). Furthermore the structure of a complex between a two times nGAAn repeat with a part of the DBD of *K. lactis* is known (Littlefield & Nelson 1999). The DBD of Hsf1 binds specifically to the sequence motif nGAAn (or its inverted repeat nTTCn), to which the recognition helix (α-helix 3 in Hsf DBDs) forms contacts in the major groove of the DNA double helix. Heat-shock elements (HSE) typically consist of 2-6 such repeats. The affinity of a monomer binding to one repeat is therefore several orders of magnitude lower than oligomer binding to the an array in a cooperative manner (affinity monomer: ~0.6 µM vs oligomer: 12 nM) (Xiao et al. 1991; Jaeger et al. 2014). Furthermore, it has been shown that the wing-loop plays a role in

DBD-DBD interactions with adjacent inverted repeats and is required for full activity in yeast Hsf1 (Cicero et al. 2001). So far the role of the DBD C-terminal region in DNA-binding was not clear, although it was shown to be essential in yeast (Flick et al. 1994). None of the previously published Hsf DBD crystal structures included the DBD C-terminus (aa 282-299 in *K. lactis* Hsf1). The DBD is connected to the adjacent coiled-coil heptad repeat domain (HR) by a ~15 amino acid linker (Flick et al. 1994).

2.5.3.2 Heptad repeat domains regulate Hsf1 oligomerization

Oligomerization of Hsf family proteins is thought to be achieved by assembly of an α-helical coiled-coil domain consisting of several arrays of heptad repeats (HR) (Sorger & Nelson 1989; Peteranderl & Nelson 1992). These HR are composed of seven amino acids usually denoted a-b-c-d-e-f-g - whereby amino acids a and d are hydrophobic (such as Leucine), forming the core of the helix bundle. They do so via a “knobs-into-holes” pattern and therefore the resulting structures have been alternatively termed “leucine zipper”. Amino acids in positions e and g point into the solvent and interact with e and g from their opposing strand, which determines the specificity of oligomer formation (Mason & Arndt 2004; Krylov & Vinson 2001). The HR domain of human Hsf1 stretches out over ~75 amino acids and is divided into the subdomains HR-A (133-175) and HR-B (180-207) (Neef et al. 2013). Whereas the Hsf1’s from lower eukaryotes such as yeast and mammalian Hsf4 only contain HR-A/B, Hsf1’s from metazoan organisms additionally include a C-terminal leucine zipper domain termed HR-C which suppresses spontaneous oligomer formation (Rabindran et al. 1993). Mutational studies have shown that in human Hsf1 all three HR domains are required to keep Hsf1 in a monomeric state (Zuo et al. 1994).

2.5.3.3 Regulation of Hsf1’s trans-activating activity

Upon binding to the promoters of its target genes, oligomeric HSF1 facilitates transcription via its C-terminal trans-activation domain. As described above, TADs of TF are thought to recruit general transcription factors, adaptors and acetyltransferases to enable the formation of the preinitiation complex of transcription (PIC) for RNA-polymerase II. Alternatively, TF (such as Hsf1) can trigger the of release paused Pol II (Rougvie & Lis 1988). The TAD of mouse Hsf1

has been described as intrinsically unstructured domain which adopts limited structure upon temperature shift, acidic pH and other stimuli (Pattaramanon et al. 2007).

The TAD of Hsf1 is under the control of the regulatory domain (RD), an unstructured segment. The RD forms the central region of the Hsf1 protein sequence and has been described to have the capacity to sense stress intrinsically as well as being controlled by differential post-translational modifications (PTM) (Anckar & Sistonen 2011). However, the code by which these PTMS regulate Hsf1 function is not yet clear (Xu et al. 2012; Knauf et al. 1996; Budzyński et al. 2015).

2.5.4 Oxidative and cell wall stress response in fungi

Fungi contain a paralog of Hsf1, Skn7, which is involved in the response to cell wall stress (Raitt et al. 2000). In contrast to Hsf1, Skn7 contains a phosphor receiver domain, an element of a phosphorelay signaling system (Figure 12).

Bacteria, protozoa and fungi often employ histidine kinase-based phosphorelays in response to certain stimuli (Figure 14). One of the best studied systems is the Sln1 phosphorelay, in which the membrane associated fungal kinase Sln1 senses cell wall stress and transmits the signal to the response regulator Skn7 (Fassler & West 2011). However, the Hsf paralog Skn7 contributes as a transcription factor not only to the response to cell wall stress but also to oxidative stress. Skn7 interacts with the transcription factor Yap1 to coordinate the oxidative stress response in budding yeast (Morgan et al. 1997; Lee et al. 1999).

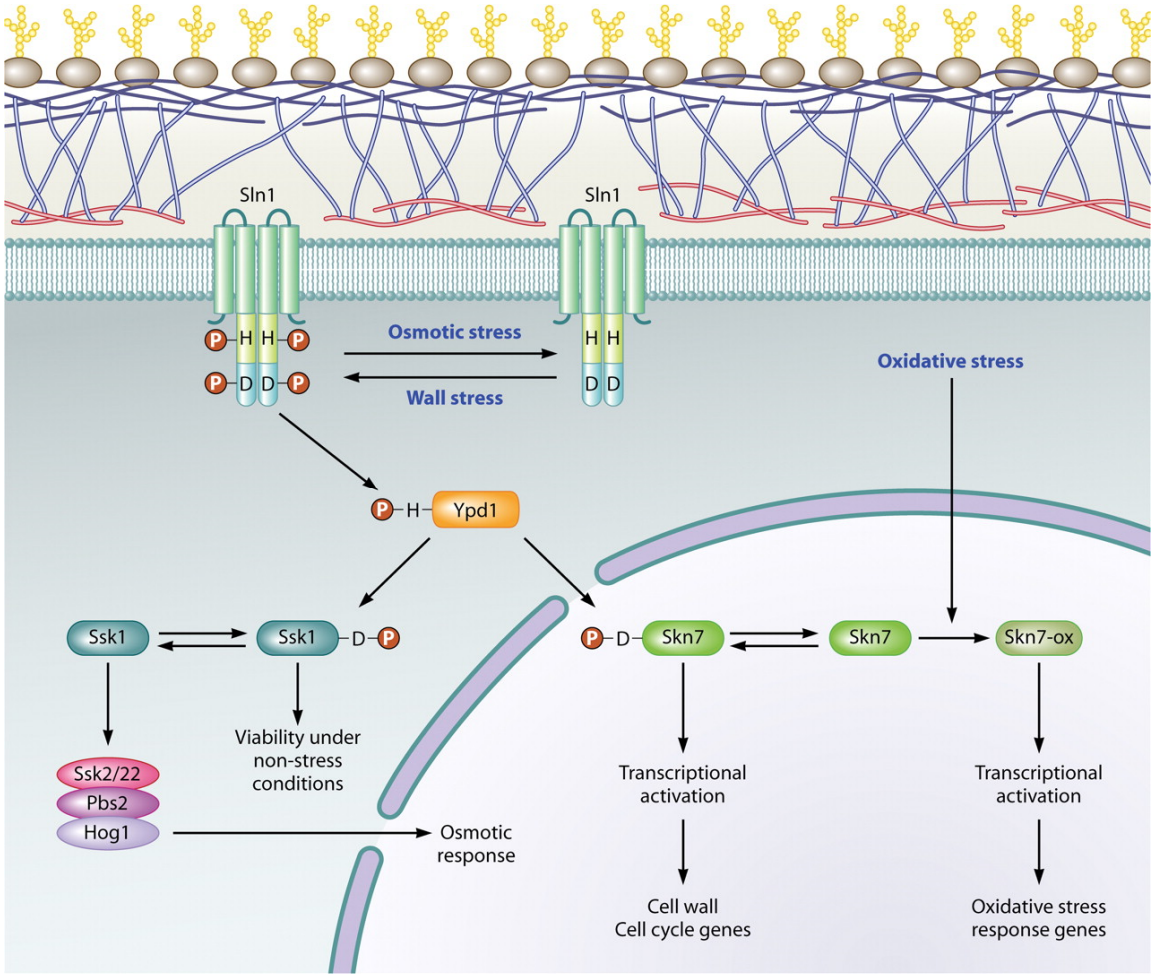

Figure 14 Cell wall and oxidative stress response in fungi mediated by Sln1 and Skn7. Cell wall stress is sensed by the membrane associated kinase Sln1 and the response mediated via a phosphorelay to the transcription factor Skn7. In addition Skn7 can contribute to respond to oxidative stress. (Ssk1, response regulator, Hog1 pathway is involved in osmoregulation). Reproduced from (Fassler & West 2011)

Via Skn7, oxidative stress induces expression of molecular chaperones such as Ssa1p (Hsp70) and Hsp82p (Hsp90) (Lee et al. 1999; Raitt et al. 2000). The primary sequence specificity of Skn7 is towards Sln1-responsive elements (SSRE) which are clearly distinct from HSEs, as shown using the model promotor of OCH1, encoding a α-1,6 mannosyl transferase, which has two such motifs divided by a 11 bp spacer (Li et al. 2002). In addition, Skn7 binding to HSE on the TRX2 (thioredoxin) promotor of *S. cerevisiae* has been reported (Morgan et al. 1997).

2.5.4.1 Function of the receiver domain of Skn7

The DBD and coiled-coil regions of Skn7 are close in sequence to those of metazoan Hsf1. However, the regulatory segment of Skn7 comprises a receiver domain which resembles a specific fold found in phospho-relay pathways (Figure 12), (Raitt et al. 2000). The receiver domain is phosphorylated at the conserved Asp427 residue in response to cell wall stress, thereby activating Skn7 (Li et al. 1998). Modification of Asp427 does not seem to play a role in the oxidative stress response (Morgan et al. 1997). PTMs in form of Serine-phosphorylation and acetylation of Skn7 have also been reported (He et al. 2009; Gómez-Pastor et al. 2013). Furthermore, a direct interaction betweenSkn7 and Hsf1 has been reported in yeast (Raitt et al. 2000).

2.6 Role of Hsf1 in proteostasis in aging and disease

Cellular life as we know it is strictly dependent on a functional protein network. Each eukaryotic cell produces ~10000 different proteins, which are present in high total concentrations up to 400 mg/ml in the cytosol (Hartl et al. 2011). Therefore, each cell invests a great amount of metabolic energy to maintain its molecular machinery in a functional state. As described above, the functional network maintaining proteins is the proteostasis network, comprising the machineries for translation, protein folding and protein degradation (Figure 1). This machinery is tuned to work efficiently under normal conditions, but has to be adjusted under conditions which increase the amount of misfolded proteins in the cell (Morimoto 2008).

Several human disease states have been described in which protein deposits are found in the intra or extra cellular space, consisting of amorphous protein aggregates or highly ordered fibrils (Ross & Poirier 2004; Labbadia & R. Morimoto 2015). These diseases were also termed proteopathies (Carrell & Lomas 1997; Walker & LeVine 2000). Some are loss-of-function diseases, such as cystic fibrosis, in which a specific mutation causes misfolding and inefficient trafficking of the membrane protein Cystic Fibrosis Transmembrane Conductance Regulator (CFTR). Gain of toxic function diseases have been described, whereby proteins with increased aggregation propensity cause proteotoxicity and induce

malfunction of the proteostasis network. The latter diseases include Alzheimer's disease (extracellular plaques of Aβ-peptides and intracellular tangles of Tau protein), Parkinson's disease (intracellular α-synuclein aggregates (Lewy Bodies)), Huntington's disease (intracellular aggregates of huntingtin with polyQ expansion) and Amyotrophic lateral sclerosis (ALS) (including intracellular aggregates of superoxide dismutase 1 (SOD1) and TAR DNA-binding protein 43 kDa (TDP-43)) (Ross & Poirier 2004).

Intriguingly, most of the gain-of-toxic function disorders manifest later in life and sometimes the disease onset can be clearly correlated to the level of underlying proteostasis challenge (Hipp et al. 2014). A classic example for this relation is Huntington's disease. Huntington's disease is caused by a mutationally expanded poly-glutamine region in exon 1 of huntingtin, a protein of unknown function. When the polyQ region exceeds the length of 37 glutamines, the individual will likely show the disease phenotype at some point in its life. The length of the polyQ repeat determines the age of onset of the disease. Long repeats cause early onset whereas shorter repeats shift the age of onset to higher age. *In vitro* experiments showed that longer polyQ repeats also show increased aggregation propensity (Scherzinger et al. 1999). Therefore it was concluded that expansion of the polyQ tract increases the aggregate load in the cell and is inversely correlated with the age of onset of the disease (Gusella & MacDonald 2006). Proteostasis network capacity seems to decline with sexual maturity and aging (Labbadia & R. I. Morimoto 2015; Ben-Zvi et al. 2009).

Several approaches to pharmacologically target disease states by influencing the proteostasis network have been proposed. First, one could try to stabilize aggregation-prone variants of proteins with small molecule drugs, tailored to the specific disease-causing protein. Therefore the proteostasis network's load decreases. This has been proved to work for the retinol-binding serum protein transthyretin (TTR) which causes familial amyloid polyneuropathy – a disease state in which mutations in TTR cause its aggregation, leading to neurodegeneration and cardiomyopathy (Bulawa et al. 2012). Second, adjusting the proteostasis network and boosting its capacities by so-called "proteostasis regulators" could be of use for a variety of protein proteopathies (Balch et al. 2008). Since the proteostasis network is built around synthesis, maintenance and

degradation of proteins, each branch could be addressed in isolation or combination.

Boosting the folding and refolding capacity of the proteostasis network by overexpression of molecular chaperones has been shown to be protective for cells in the case of α-synuclein aggregation and polyQ aggregation (Auluck et al. 2002; Warrick et al. 1999). Therefore activation of Hsf1, the master regulator of the cytosolic stress response and main inducer of molecular chaperones, is considered a promising strategy (Neef et al. 2010).

Moreover, several studies observed interaction of Hsf1 with the insulin signaling pathway in *C. elegans* (Hsu et al. 2003; Anckar & Sistonen 2011). In these studies, overexpression of Hsf1 increased lifespan of the nematode about 40 %. However, daf16, the FOXO transcription factor homologue of *C. elegans*, is required for the insulin pathway as well.

On the other hand, overexpression of Hsf1 has been linked to cancer progression and poor prognosis in cancer patients (Santagata et al. 2011; Fang et al. 2012). Consistent with these findings, a landmark study found a remarkable resistance to carcinogenesis in Hsf1 knock-out mice (Dai et al. 2007). Recent studies also observed a multifaceted transcription program linked to Hsf1 (Mendillo et al. 2012). Tumor tissues are characterized by a high protein metabolism, which requires an up-regulated translation and protein folding machinery (Dai & Sampson 2015). However, not only proteins involved in proteostasis, but also other cellular processes such as cell cycle and signaling pathways are differentially regulated. It is therefore not surprising that Hsf1 has the capacity to reprogram not only tumor cells but also the tumor surrounding fibroblasts to an increased malignant phenotype of the tumor (Scherz-Shouval et al. 2014).

Considering the protective role of Hsf1 in proteostasis failure and its tumor-driving role in cancer, there is substantial interest in targeting Hsf1 with either activating or inhibiting small molecule compounds (Au et al. 2009; Whitesell & Lindquist 2009; Neef et al. 2010; Calamini et al. 2011).

3 Aim of the study

Although discovered more than 30 years ago and functionally well characterized, structural details of heat shock transcription factors have largely remained enigmatic (Wu 1995). Only structural data for isolated DBDs of the Hsf homologs from *Kluveromyces lactis, Drosophila melanogaster* and *Homo sapiens* (Liu et al. 2011; Harrison et al. 1994; Vuister et al. 1994) and of a protein-DNA complex of a two-foldnGAAn repeat with a part of the DBD of *K. lactis* had been published (Littlefield & Nelson 1999). The aim of this study was therefore to elucidate the structures of functional complexes of human Hsf1 using X-ray crystallography as the core technique. Furthermore, domain topology and dynamics of Hsf1 were of interest, with the focus on the intramolecular interaction of the heptad repeat domains. Structural details of Hsf1 at atomic resolution might facilitate structure-guided drug design for cancer and proteopathies.

4 Materials and Methods

4.1 Enzymes and Chemicals

4.1.1 Enzymes

Supplier	Enzyme or chemical
New England Biolabs (Frankfurt am Main, Germany)	Restriction enzymes
Promega (Mannheim, Germany)	Pfu polymerase
	T4 ligase
Roche (Basel, Switzerland)	Proteinase K (Enzyme)

4.1.2 Chemicals

Supplier	Enzyme or chemical
Biomol (Hamburg, Germany)	HEPES
Biozym (Hessisch Oldendorf, Germany)	Biozym LE Agarose
Creative Molecules Inc.	H_{12}/D_{12}-DSS
Difco (Heidelberg, Germany)	Bacto Agar
	Bacto Tryptone
	Bacto peptone
	Bacto Yeast Extract
Fermentas (St. Leon-Rot, Germany)	GeneRuler 1kb
	PageRuler Prestained Protein Ladder
Fluka (Deisenhofen, Germany)	Polyethylene glycols of different molecular weights
	Sodium cacodylate
GE Healthcare (München, Germany)	Chloramphenicol

	MES
Invitrogen (Karlsruhe, Germany)	pProEx-HtA, pProEx-HtB
	SYBR Safe DNA gel strain
Metabion (Martinsried, Germany)	dNTP
	Oligonucleotides
	Primers
Promega (Mannheim, Germany)	Wizard Plus SV Miniprep DNA Purification System
	Wizard SV Gel and PCR Clean-Up System
Qiagen (Hilden, Germany)	Qiagen Plasmid Midi Kit
Roche (Basel, Switzerland)	DTT
	EDTA-free Complete protease inhibitor cocktail
Roth (Karlsruhe, Germany)	Ampicillin
	IPTG
Serva (Heidelberg, Germany)	Acrylamide-Bis
	Coomassie Blue R250
	PMSF
	SDS
Sigma Aldrich (Steinheim, Germany)	Poly(dI-dC)
	DMTMM
Thermo Fisher Scientific (Waltham, USA)	EGS

4.2 Strains

Supplier	**Strain**
Novagen (Darmstadt, Germany)	*E. coli* DH5α

Stratagene (Heidelberg, Germany)	*E. coli* BL21(DE3) CodonPlus-RIL
	E. coli BL21 Rosetta (DE3)

4.3 Instruments

Supplier	**Product**
Avestin (Ottawa, Canada)	Emulsiflex EF-C5/C502290
Beckman Coulter (Krefeld, Germany)	Centrifuges (GS-6R, Allegra-6R, Avanti J-25 with rotors JKA 10.500 and JA 25.50, J6-MI with rotor JS 4.2)
Biometra (Göttingen, Germany)	PCR thermocycler
Bio-Rad (München, Germany)	ChemiDoc XRS
	Horizontal agarose gel electrophoresis
	Mini Protean II electrophoresis cell
	Power Pac 300
	Micro Bio-Spin 6 columns
Eppendorf (Hamburg, Germany)	Centrifuges (5415D and 5417R)
	Pipettes (2, 10, 20, 100, 200, 1000 µl)
	Thermomixer comfort
Fisher Scientific (Schwerte, Germany)	Accumet Basic pH meter
Fujifilm (Minato, Japan)	Image reader FLA-2000
GE-Healthcare (München, Germany)	Äkta FPLC
	Äkta Purifier/Explorer
	Ettan LC
	Pre-packed chromatography columns as following: HiPrep

	26/10 Desalting, HiLoad 26/60 Superdex 200, Mono S HR 10/10
Hampton Research (Aliso Viejo, USA)	Mounted cryo loops
	Siliconized glass square cover slides (22 mm)
	VDX Plate with sealant
Mascom (Bremen, Germany)	nano-ESI pipettes
Mettler Toledo (Gießen, Germany)	Balances (AG285, PB602)
Millipore (Schwalbach, Germany)	Centriprep concentrators (3000, 10.000, 30.000 and 50.000 Da MWCO)
New Brunswick Scientific (now Eppendorf)	Innova 44 incubator / shaker
PEQLAB (Erlangen, Germany)	Nanodrop 1000
Roth (Karlsruhe, Germany)	ZelluTrans dialysis membrane
Scientific Industries, Inc. (Bohemia/NY, USA)	Vortex-Genie 2
Tecan (Männedorf, Schweiz)	GENios Pro plate reader
Thermo Fisher Scientific (Waltham, USA)	Easy-nLC 1000 UPLC system
	Q-Exactive Orbitrap mass spectrometer
	Novex NuPAGE SDS-PA Gels
Waters Corporation (Manchester, UK)	Synapt G2-Si Q-TOF Mass spectrometer

4.4 Media and Buffers

2YT-Medium	80 g Bacto peptone 50 g Bacto yeast extract 25 g NaCl 7.5 mL 5 M NaOH Add water to 5 L volume, stir vigorously and steam-autoclave bottles

LB-Medium:	50 g Bacto peptone 25 g Bacto yeast extract 50 g NaCl Add water to 5 L volume, stir vigorously and steam-autoclave bottles

Buffers were filtered through a 0.22 µm pore size filter and degassed. Milli-Q water was used.

Buffer A	25 mM MES-NaOH pH 6.5 300 mM NaCl 10% glycerol 3 mM betaME
Buffer B	25 mM MES-NaOH pH 6.5 30 mM NaCl 10% glycerol 3 mM betaME
Buffer C	10 mM MES-KOH pH 6.5 100 mM KCl
Buffer D	25 mM HEPES-NaOH pH 7.4 300 mM NaCl 10% glycerol 3 mM betaME
Buffer E	25 mM HEPES-NaOH pH 7.4 150 mM NaCl 10% glycerol 1 mM DTT
Buffer F	25 mM Tris-HCl pH 8.5 300 mM NaCl 10% glycerol 3 mM betaME
Buffer G	25 mM HEPES-NaOH pH 7.4 150 mM NaCl 10% glycerol 3 mM betaME
Coomassie destaining solution:	10 % (v/v) ethanol 10 % (v/v) acetic acid
Coomassie staining solution:	0.1 % (w/v) Serva Coomassie Blue R250 40 % (v/v) ethanol 10 % (v/v) acetic acid

DNA annealing buffer	10 mM HEPES-NaOH pH 7.0 100 mM NaCl 2.5 mM MgCl2
DNA-loading buffer (6x):	0.25 % (w/v) bromophenol blue 0.25 % (w/v) xylene cyanol FF 40 % (w/v) sucrose in H2O
SDS-loading buffer (2x)	100 mM Tris-HCl pH 6.8 4 % (w/v) SDS 200 mM DTT 20 % (v/v) glycerol 0.2 % (w/v) bromphenolblue
SDS-running buffer (10x)	250 mM Tris 2.5 M glycine 1 % (w/v) SDS
Native running buffer (10x)	250 mM Tris-HCl pH 7.5 2.5 M glycine 5 % glycerol
TAE Buffer	40 mM Tris 20 mM acetic acid 1 mM EDTA

4.5 Molecular Biology Methods

4.5.1 Agarose gel electrophoresis

DNA was analyzed using agarose gel electrophoresis with gels containing 0.8 % (w/v) agarose dissolved in 1x TAE buffer and mixed with SYBR-Safe DNA gel stain (Life-technologies; 0.1 µL/mL agarose gel). Samples containing 1X DNA-loading buffer (NEB) were loaded into the gel and separated in 1x TAE buffer at a constant voltage of 80 V for 15 – 20 min.

4.5.2 Preparation of double stranded DNA oligonucleotides

DNA-oligonucleotides for EMSA and fluorescence polarization measurements were obtained at HPLC-purified grade from Metabion. Single stranded unlabeled DNA was dissolved in DNA-annealing buffer to a concentration of 5 mM. Solutions

of complimentary strands were mixed to a final concentration of 2.5 mM and annealed by slow-cooling in a metal heating block (~40 °C h^{-1}) from 95 °C to RT.

Fluorescently labeled DNA-oligonucleotides were synthesized with a 6-carboxyfluorescein (6-FAM) fluorescence label at the 5' end (Metabion). Single stranded DNA was dissolved at 100 µM concentration in DNA-annealing buffer and annealed as described above. Double stranded DNA oligonucleotides were stored at -20 °C.

4.5.3 Polymerase chain reaction (PCR)

Amplification of DNA constructs were performed using Pfu polymerase (Promega). Each reaction was set up as follows:

Component	Amount
Template	1 µL Miniprep
Primer 1 and 2	0.5 µM each
dNTPs	200 µM
Pfu-Buffer	1x
Pfu-Polymerase	3 U
Total volume:	50 µL

PCR annealing temperatures were chosen according to the melting temperatures of the primers (in most cases approximately 60 °C). PCR extension times varied based on target length and the elongation speed of the Pfu-polymerase. The program was started with an initial 5 min denaturation step at 98 °C. The first step of the cycle was 30 sec at 96 °C for denaturation of elongated DNA. This was followed by 30 sec at 53 °C for primer annealing. An optimal extension temperature of 72 °C was set for elongation using Pfu-polymerase, assuming approximately 1000 bp per 1 min as elongation velocity. This cycle was repeated for 30-35 times followed by a final elongation step of 5 min at 72 °C and subsequent cooling to 4 °C till further processing.

4.5.4 Cloning strategies and molecular cloning

Synthetic DNA for CtSkn7(35-652+685-765) and the HsHsf1-CtSkn7 chimera (HsHsf1(1-132)-CtSkn7(160-220)-HsHsf1(194-529)) was purchased at GeneArt (Life technologies). The missing CtSkn7 sequences in CtSkn7(35-652+685-765) were highly repetitive and not accessible to the DNA synthesis protocol of the company. The DNA sequence of CtSkn7 was codon-optimized for expression in

E. coli whereas the DNA sequence of the HsHsf1-CtSkn7 chimera was codon-optimized for expression in human cell lines. Domain borders were determined by multiple sequence alignment using Clustal Omega and secondary structure prediction with Jpred (Drozdetskiy et al. 2015).

The HsHsf1 DBD construct, residues 1-120, was PCR-amplified from pCMV-HsfHsf1-Flag (Plasmid#1932, Addgene), and cloned into pProEx-HtB (Life technologies) using EcoRI and HindIII restriction sites. The resulting construct can be expressed as a TEV protease-cleavable, N-terminal His6-tagged fusion protein in *Escherichia coli*. As a consequence of the cloning strategy, HsHsf1(1-120) contained the additional N-terminal sequence GAMGSGILRGG after His6-tag cleavage. The expression constructs CtSkn7(40-143), CtSkn7(40-220), CtSkn7(160-209), CtSkn7(160-220), HsHsf1-CtSkn7 chimera and HsHsf1 point mutants were PCR-amplified and cloned into the pHUE plasmid for expression as N-terminally His6-tagged ubiquitin (Ub) fusion proteins using SacII and HindIII restriction sites. This tag is readily cleavable using the de-ubiquitinating enzyme Usp2 to produce pure untagged protein (Catanzariti et al. 2004). The His_6-SUMO-HsHsf1 expression strain was a generous gift of Dr. N. Hentze and Dr. M. P. Mayer (Heidelberg).

For the generation of complementary ends for DNA recombination, DNA was digested with the respective restriction enzymes. Digests were performed in the recommended reaction buffers and in presence of BSA for 1 h at 37°C. For plasmid backbone preparation, plasmid were digested for up to 24 h at 37°C to avoid traces of uncut vector.

Component	Amount
H_20	10.5 µL
DNA	15 µL
Restriction enzyme 1 and 2	0.5 – 1 µL each
10x Reaction buffer	3 µL
BSA	0.5 µL
Total volume	30 µL

Ligations were performed using 100 ng plasmid backbone and 50 fmol (~33 ng for 1 kb) insert in 20 µL total volume with 1 µL T4 ligase, and supplemented with 2 µL 10x T4 reaction buffer (Promega). The reaction was incubated at 16°C for at least 1.5 h before transformation.

4.5.5 Site-directed mutagenesis

Overlapping primers were designed to include the mutated DNA base triplet within approximately 30 bp of complementary sequence to the target DNA. Subsequently, a standard PCR with an annealing temperature of 50-53 °C and an elongation time to allow whole plasmid amplification (e.b. for pHUE with 500 bp insert: 10 min) was performed. Source plasmid DNA background was digested by addition of DpnI for 1.5 h at 37ºC. Sequencing data was aligned and analyzed with Multalign (Corpet 1988).

4.5.6 Competent *E. coli* cells

For the preparation of CaCl-chemically competent *E. coli* cells, a culture of the respective strain was grown to mid-log phase (A_{600} = 0.5) at 37 ºC in 1 L of LB medium. The flask was subsequently chilled on ice for 10 minutes before pelleting the cells for 30 min, 4000 rpm, at 4 ºC. In the final step, cells were resuspended gently in 20 mL ice-cold 0.1 M $CaCl_2$, 15 % glycerol and incubated on ice for 20 min. Finally, the cells were aliquoted as 100 µL fractions into chilled (-20 ºC) sterile Eppendorf tubes and frozen in liquid nitrogen. Cells were stored at -80 ºC.

4.5.7 *E. coli* cell transformation

For transformation 100 µL CaCl-competent *E. coli* cells were thawed on ice and mixed with 1 µL plasmid DNA (approximately 100 ng) or 10 µL of a ligation reaction. After incubation on ice for 30 minutes, cells were heat-shocked for 5 minutes at 37°C and then cooled on ice. Next, 900 µL LB-Medium was added and cells were incubated 1-2h at 37°C, 450 rpm in a shaker. Finally, 50 µL of pelleted cells (4000 rpm, 2 min) were re-suspended in 50 µL LB and plated on LB agar plates containing the respective selective antibiotic.

4.6 Protein biochemistry

4.6.1 Protein analytical methods

4.6.1.1 Determination of protein concentration

Measurement of protein concentration at 280 nm was performed using a Nanodrop device (Thermo Scientific). Protein solution (2 µL) was applied to the Nanodrop to obtain the absorbance and the concentration calculated using the molecular weight and calculated extinction coefficient (calculated from the sequence using ExPASy ProtParam). DNA contaminations in purifications of Hsf family proteins containing a DBD could be estimated by monitoring the absorption at 260 nm and 280 nm.

4.6.1.2 Protein quantification and SDS-PAGE

Components for 10 mL (2 gels)	**Resolving gel: 15% acrylamide**	**Stacking gel: 5% acrylamide**
H_2O	2.3 mL	6.8 mL
30% acrylamide / 0.8% bis-AA	5.0 mL	1.7 mL
1.5 M Tris (pH 8.8)	2.5 mL	-
1.0 M Tris (pH 6.8)	-	1.25 mL
10% SDS	100 µL	100 µL
10% ammonium persulfate	100 µL	100 µL
TEMED	4 µL	10 µL

Protein content and purified proteins were analyzed by discontinuous SDS-PAGE (Tris-glycine sodium dodecylsulfate polyacrylamide gel electrophoresis). This method separates proteins primarily according to their molecular weight (Laemmli 1970). For a perfectly even resolving-to-stacking gel boundary, the resolving gel casts were covered with a layer of isopropanol immediately after casting. Protein samples were mixed with 2x or 4x SDS-loading dye, boiled at 95°C for 5 minutes for SDS-PAGE analysis. Electrophoresis was performed for 40 min in SDS-running buffer at a constant voltage of 220 V. After electrophoresis, gels were stained with Coomassie staining solution and destained with Coomassie destaining solution on a shaker. Images of the gels were taken with a V750 Pro gel documentation system (Epson). As molecular weight marker, PageRuler (prestained protein ladder) (Thermo Scientific) was used.

4.6.1.3 Protein production and purification

All protein purification steps were performed at 4 °C and lysis buffer was supplemented with 1 mM Phenylmethylsulfonylfluorid (PMSF) unless noted otherwise. Protein concentrations in the final preparations were determined as described in 4.6.1.1. The protein mass of the final product was estimated by SDS-PAGE and confirmed by mass spectrometry.

HsHsf1(1-120) – *E. coli* Bl21-CodonPlus-RIL cells were transformed with pProEx-HtB-HsHsf1(1-120), for expression as a N-terminal His_6-tagged fusion protein, His_6-HsHsf1(1-120). The resulting strain was grown at 37 °C using flasks of LB medium containing 0.1 mg mL^{-1} ampicillin and 0.034 mg mL^{-1} chloramphenicol. After reaching an OD_{600nm} of 0.6, the cultures were shifted to 18 °C. After 45 min, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.2 mM, and the culture continued for 16-20 h. Cells were harvested by centrifugation at 4200 rpm (5020 x g) for 25 min using a JS4.2 rotor (Beckman). The cell pellet from 1 L of culture was resuspended in 25 mL ice-cold buffer A (25 mM MES-NaOH pH 6.5, 300 mM NaCl, 10 % (v/v) glycerol, 3 mM β-mercaptoethanol (β-ME))) supplemented with 1 mg mL^{-1} lysozyme, and the cell suspension passed three times through an ice-cooled Emulsoflex C5 French press (Avestin). Cell debris was removed by centrifugation at 20,000 rpm (48,400 x g) for 30 min using a JA25.50 rotor (Beckman). The clear supernatant was applied to a Ni^{2+}-nitrilotriacetic acid (NTA) Superflow (Qiagen) column (3 mL for protein from 1 L culture) using a FPLC. The eluate fractions were analyzed by SDS-PAGE, followed by Coomassie staining. The fractions containing the His_6-tagged protein were combined and His_6-TEV protease was added at a protein mass ratio of 1:100 together with 3 mM β-ME and 0.5 mM EDTA (final concentrations), followed by incubation overnight at 4 °C. Subsequently the buffer was exchanged against buffer B (25 mM MES-NaOH pH 6.5, 30 mM NaCl, 10 % (v/v) glycerol, 3 mM β-mercaptoethanol (β-ME)),) using a HiPrep-desalting column (GE Healthcare). In the following step, HsHsf1(1-120) was applied to cation exchange chromatography on Source-S (GE Healthcare) equilibrated in buffer B. The protein was eluted with a linear salt gradient to 500 mM NaCl. Size-exclusion chromatography on a Superdex-200 (GE Healthcare) column equilibrated with buffer C (10 mM MES pH 6.5, 100 mM KCl) was used as the final purification step.

HsHsf1(1-120) was subsequently concentrated to 18 mg mL^{-1} concentration by ultrafiltration using a Vivaspin (3 kDa cut-off) centrifugal device (Vivascience, Hannover). After snap-freezing in liquid nitrogen, protein aliquots were stored at -80 °C.

HsHsf1-GCN4 – *E. coli* Rosetta 2 (DE3) cells were transformed with the respective pHUE plasmid for expression as N-terminal His_6-ubiquitin-tagged fusion protein. The resulting strain was grown in shaking flask culture at 37 °C using 2YT medium containing 0.1 mg mL^{-1} ampicillin and 0.034 mg mL^{-1} chloramphenicol. At an OD_{600nm} of 0.6, the cultures were shifted to 18 °C. After 45 min, IPTG was added to a final concentration of 0.2 mM, and the culture continued for 16-20 h. Cells were harvested by centrifugation at 4200 rpm (5020 x g) for 25 min using a JS4.2 rotor (Beckman). The cell pellet from 1 L culture was resuspended in 25 mL ice-cold buffer A containing 1 mg mL^{-1} lysozyme and 0.05 % Triton X-100. The cell suspension was then passed three times through an ice-cooled Emulsoflex C5 French press. Cell debris was removed by centrifugation at 20,000 rpm (48,400 x g) for 30 min using a JA25.50 rotor. The supernatant was applied to Ni^{2+}- NTA beads (~5 mL CV for protein from 1 L culture). The eluate fractions were analyzed by SDS-PAGE, followed by Coomassie staining. The fractions containing the His_6-Ub-fusion protein were combined, His6-Usp2 protease added at a protein mass ratio of 1:100 together with 3 mM β-ME (final concentration) and incubated over night at 4°C. Subsequently the buffer was exchanged against buffer B using a HiPrep-Desalting column (GE Healthcare). In the following step, the constructs were applied to cation exchange chromatography on Source-S (GE Healthcare) equilibrated with buffer B. Subsequently, the proteins were eluted by a linear salt gradient to 500 mM NaCl. Size-exclusion chromatography on Superdex-200 (GE Healthcare) equilibrated with buffer D was used as the final purification step. The purified proteins were subsequently concentrated up to ~44 mg mL^{-1} by ultrafiltration using a 10 kDa cut-off. After snap-freezing in liquid nitrogen, protein aliquots were stored at -80 °C.

Human Hsf1 - This protocol was adapted from the Mayer lab procedure.(Hentze et al. 2016). *E. coli* Bl21 CodonPlus-RIL cells were transformed with the respective plasmids for expression as N-terminal His_6-Sumo-tagged fusion protein or N-terminal His_6-Ubiquitin-tagged fusion protein in case of HsHsf1 K80Q, K116Q and

K118Q. The resulting strains were grown in flask cultures at 37 °C using 2YT medium containing 0.1 mg ml^{-1} ampicillin and 0.034 mg ml^{-1} chloramphenicol. At an OD_{600nm} of 0.6, the cultures were shifted to 20 °C. After 45 min, IPTG was added to a final concentration of 0.2 mM, and the culture continued for 2 h. Cells were harvested by centrifugation at 4200 rpm (5020 g) for 25 min using a JS4.2 rotor. Cells from 1 L culture were resuspended in 25 mL buffer G, 1 mg mL^{-1} lysozyme and Complete protease inhibitor cocktail (Roche) was added, subsequently the cell suspension passed three times through an ice-chilled french press. Cell debris was removed by centrifugation as above. The clear supernatant was applied to a Ni^{2+}-TED gravity flow column (Macherey Nagel) containing 0.125 g dry material per pellet of 1 L expression culture. The column was washed with 100 CV buffer G supplied with 1 M NaCl. Bound His_6-Sumo tagged protein was eluted with 10 CV 250 mM imidazole in buffer G. The eluate fractions were analyzed by the Bradford assay. Protein-containing fractions were merged and His_6-Ulp1 protease added at a protein mass ratio of 1:20, together with 3 mM β-ME (final concentration) and incubated at 4 °C for 2–3 h. The final purification step was performed with a Superdex 200 26/60 column equilibrated with buffer E. Protein-containing fractions were pooled and concentrated to 10 µM using a Vivaspin (30 kDa cut-off) centrifugal device (Vivascience).

CtSkn7 truncation constructs CtSkn7(40-143), CtSkn7 (40-209) and CtSkn7(40-220) – *E. coli* Rosetta 2 (DE3) cells were transformed with the respective pHUE plasmids for expression as N-terminal His_6-ubiquitin-tagged fusion proteins. The resulting strain was grown in shaking flask culture at 37 °C using 2YT medium containing 0.1 mg mL^{-1} ampicillin and 0.034 mg mL^{-1} chloramphenicol. At an OD_{600nm} of 0.6, the cultures were shifted to 18 °C. After 45 min, IPTG was added to a final concentration of 0.2 mM, and continued incubation for 16-20 h. Cells were harvested by centrifugation at 4200 rpm (5020 x g) for 25 min using a JS4.2 rotor (Beckman). The cell pellet from 1 L of culture was resuspended in 25 mL ice-cold buffer A supplied with 1 mg mL^{-1} lysozyme, and the cell suspension passed three times through an ice-cooled Emulsoflex C5 French press. Cell debris was removed by centrifugation at 20,000 rpm (48,400 x g) for 30 min using a JA25.50 rotor. The clear supernatant was applied to a Ni^{2+}- NTA Superflow column (~3 mL CV for cells from 1 L culture). The eluate fractions were analyzed by SDS-PAGE,

followed by Coomassie staining. The fractions containing the His_6-Ub-fusion protein were combined, and His6-Usp2 protease added at a protein mass ratio of 1:100 together with 3 mM β-ME (final concentration) and incubated over night at 4°C. Subsequently the buffer was exchanged against buffer B using a HiPrep-Desalting column (GE Healthcare). In the following step, the constructs were applied to cation exchange chromatography on Source-S (GE Healthcare) equilibrated with buffer B. Subsequently, the proteins were eluted by a linear salt gradient to 500 mM NaCl. Size-exclusion chromatography on Superdex-200 (GE Healthcare) equilibrated with buffer C (and 5-10 % glycerol in case of CtSkn7(40-209) and CtSkn7(40-220)) was used as the final purification step. The purified proteins were concentrated to 20 mg mL^{-1}, 32 mg mL^{-1} and 33 mg mL^{-1} concentration, respectively, by ultrafiltration using a 3 kDa cut-off. After snap-freezing in liquid nitrogen, protein aliquots were stored at -80 °C.

CtSkn7(35-652+685-765) – Expression and purification were as described for the other CtSkn7 truncation constructs, except that size-exclusion chromatography on Superdex-200 (GE Healthcare) equilibrated with 25 mM HEPES-NaOH, 150 mM NaCl, 10 % glycerol was used as final purification step. The purified protein was concentrated to 1.84 mg mL^{-1} by ultrafiltration using a 10 kDa cut-off.

CtSkn7 truncation constructs CtSkn7(160-209) and CtSkn7(160-220) – Expression and purification were performed as described for the other CtSkn7 truncation constructs up to the cleavage of the expression tag. However, instead of buffer A, buffer F supplied with imidazole was used to develop the Ni^{2+}-NTA resin. After cleavage, residual His_6-tagged protein and His_6-Usp2 protease were removed by affinity chromatography on Ni^{2+}-NTA resin. Size-exclusion chromatography on a Superdex-200 (GE Healthcare) column equilibrated with 10 mM Tris-HCl pH 8.5, 100 mM KCl was used as final purification step. The final protein concentrations were 24 and 11 mg mL^{-1}, respectively.

Hsf1-CtSkn7-chimera – This protocol was adapted from the procedure of the Thiele lab. (Jaeger et al. 2014) *E. coli* Bl21 CodonPlus-RIL cells were transformed with the respective pHUE plasmid for expression as N-terminal His_6-Ub-tagged fusion protein. The resulting strain was grown in a shaking flask culture at 37°C using 2YT medium containing 0.1 mg mL^{-1} ampicillin and 0.034 mg mL^{-1} chloramphenicol. At an OD_{600nm} of 0.6, the cultures were shifted to 20°C. After 45

min, IPTG was added to a final concentration of 0.2 mM, and the culture continued for 2 h. Cells were harvested by centrifugation at 4200 rpm (5000 x g) for 25 min using a JS4.2 rotor. Cells from 1 L culture were resuspended in 25 mL buffer D (25 mM Hepes-NaOH pH 7.4, 300 mM NaCl, 10 % glycerol and 3 mM β-ME), supplied with 20 mM imidazole, 1 mg mL^{-1} lysozyme and Complete protease inhibitor cocktail, and the cell suspension passed three times through an ice-cooled french press. Cell debris was removed by centrifugation as above. The clear supernatant was applied to a Ni^{2+}-NTA FPLC column of 1 mL CV. The column was washed with 5 CV of buffer D containing 40 mM imidazole, and 2 CV of buffer D containing 40 mM imidazole, 5 mM ATP, 20 mM $MgCl_2$. Bound His_6-Ub tagged protein was eluted with 10 CV buffer D containing 250 mM imidazole. Protein-containing fractions were merged and His_6-Usp2 protease added at a protein mass ratio of 1:50 together with 3 mM β-ME (final concentration) and incubated at 4°C for 2–3 h. The final purification step was performed on a Superdex-200 column equilibrated with buffer E (25 mM HEPES pH 7.4, 150 mM NaCl, 10 % glycerol, 1 mM DTT). Protein-containing fractions were pooled and concentrated to ~10 µM using a Vivaspin (30 kDa cut-off) centrifugal device (Vivascience).

4.6.1.4 Circular dichroism spectroscopy (CD)

Secondary structure content and thermal stability of purified protein constructs were assessed by CD-spectroscopy. Proteins were desalted into DTT-free buffer or buffer freshly supplemented with DTT using the Bio-Spin6 chromatography column (Bio-Rad) according to manufacturer's instructions. Protein content was adjusted to a final concentration of 0.1-0.2 mg/ml and the CD spectra were recorded on a JASCO 715 CD-spectrometer (average of 4 passes, 250 nm to 195 nm, 0.1 nm data pitch, 4-90 °C). CD-spectra were fitted with the CONTIN algorithm on the SMP56 base dataset to estimate the secondary structure content (Greenfield 2006).

4.6.1.5 Small angle X-ray scattering (SAXS)

SAXS measurements were performed at beamline BM29 of the European Synchrotron Radiation Source (ESRF), Grenoble, France. Protein samples in the respective storage buffer were exposed for 1 s to the X-ray beam at three different

concentrations. Scattering data from ten repeats were averaged and buffer background signal was subtracted. The protein scattering data were processed with Primus and the radii of gyration were determined using the Guinier approximation. For calculating the density distribution function scattering curves were fitted with Gnom (Petoukhov et al. 2012).

4.6.1.6 Multi-angle light scattering (MALS)

Protein samples (~60 to 120 µg) were analyzed using static and dynamic light scattering by MALS coupled to size exclusion chromatography. This was performed in buffer E at 20 °C. The SEC column was connected to an UV absorbance detector, a MALS detector and a refractive index detector on a system from Wyatt Technology. Bovine serum albumin (Thermo) was used as the calibration standard.

The SEC-MALS experiments were performed by Dr. Manajit Hayer-Hartl (MPI of Biochemistry, Martinsried, Germany).

4.6.1.7 Limited proteolysis with Proteinase K

Limited proteolysis was performed to probe the purified protein construct for unstructured regions. A concentration series of the unspecific protease Proteinase K was prepared and added to the sample (final concentrations: 0 – 50 µg/ml Proteinase K, target protein at 11 µM). The reactions were incubated for 1 h on ice and stopped by addition of PMSF to a final concentration of 2 mM. Samples were analyzed by SDS-PAGE and mass spectrometry, or frozen in liquid nitrogen and stored at -80 °C until further use. The mass spectrometry measurements were performed by Dr. Roman Körner (MPI of Biochemistry, Martinsried, Germany).

4.6.1.8 Native mass spectrometry

Purified CtSkn7(40-220) and CtSkn7(160-220) at a concentration of ~30 µM were transferred into 100 mM ammonium acetate pH 7.5-8, using Micro Bio-Spin 6 chromatography columns (BioRad). HsHsf1 at a concentration of 45 µM was transferred directly, or heat shocked for 5 min at 36 or 42 °C and subsequently transferred into 100 mM ammonium acetate pH 7.5-8 as above. Native MS analysis was performed on quadrupole IM time-of-flight hybrid mass spectrometer

with a Z-spray nano-ESI source (Synapt G2-Si, Waters) in positive ion mode, using gold-platted 10 µm nano-ESI pipettes as capillaries. Capillary and sample cone voltages were optimized to 1-1.6 kV and 100-150 V, respectively. Cesium iodide (30 mg/ml) in 1:1 acetonitrile:water was used for spectra calibration.

The native MS experiments were performed by Dr. Javaid Y. Bhat (MPI of Biochemistry, Martinsried, Germany).

4.6.1.9 EGS crosslinking

Protein samples were diluted into buffer E to a final concentration of 9.4 µM. Ethylene glycol bis(succinimidyl succinate) (EGS) was added to the desired final concentration (0, 0.1, 0.2, 0.5, 1 mM), using a 10 mM EGS stock solution in dimethyl sulfoxide (DMSO), and incubated for 45 min at 15 °C. Reactions were quenched with 30 mM glycine for 15 min and analyzed by SDS-PAGE on a 4–12 % polyacrylamide gradient gel followed by Coomassie-staining.

4.6.1.10 Crosslinking coupled to mass spectrometry

Latent HsHsf1 or HsHsf1 heat shocked at 42 ºC for 10 min was incubated at 10 µM in 25 mM HEPES pH 7.4, 150 mM NaCl, 10 % glycerol and 1 mM DTT with 1 or 2 mM of a 1:1 isotropic mixture of [H_{12}]- and [D_{12}]-disuccinimidylsuberate (H_{12}/D_{12}-DSS), for 30 min at RT. Reactions were quenched using NH_4HCO_3 (150 mM final concentration). Crosslinking with 4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMTMM) was performed using 10 µM HsHsf1 and 20 mM crosslinker in 25 mM HEPES pH 7.4, 150 mM NaCl, 10% glycerol,1 mM DTT and 20 mM residual DMSO for 15 or 45 min at RT. The reactions were quenched by buffer exchange using Micro Bio-Spin 6 chromatography columns. Both crosslinkers were solubilized in 100% DMSO. Crosslinked complexes were separated on 4-12% Bis-Tris Novex NuPAGE SDS-PAGE and visualized by Coomassie staining. Excision of gel bands followed by reduction, alkylation and in-gel digestion and desalting was performed using standard protocols. Desalted peptides were dissolved in 5% formic acid and analyzed on an Easy-nLC 1000 UPLC system (Thermo) connected to a Q-Exactive Orbitrap mass spectrometer (Thermo). Peptides were separated on home-made nano spray-columns (ID 75 µm, 20 cm long with 8 µm tip opening, New Objective) containing 1.9 µm C18

beads (Reprosil-Pur C18-AQ, Dr. Maisch GmbH) on a 60 min gradient using 0.2% formic acid in water and 0.2% formic acid in acetonitrile. The column was loaded with sample by the Thermo Easy-nLC 1000 autosampler without a trap column, at a flow rate of 0.5 µl per min. The UPLC flow rate during sample analysis was 0.2 µL per min. MS/MS analysis was performed using standard data-dependent mode settings of alternate 1 high resolution (resolution =70,000) MS1 scan (*m/z* of 400-1750) followed by 10 MS2 scans (resolution =17,500) of 10 most intense ions, of charge states 3 or above, run on Xcalibur software (Thermo).

For data analysis, Thermo Xcalibur.raw files were converted to .mfg (Mascot generic file) format, using Proteome Discoverer 1.4 (Thermo). To identify crosslinked peptides, the .mgf files were analyzed on StavroX 3.1.19 (Götze et al. 2015) with following parameters: MS1 tolerance = 3 ppm, MS2 tolerance = 0.8 Da, missed cleavages for K and R = 3 and 1, respectively, signal to noise ratio ≥2, fixed and variable modifications = Cys-carbamidomethylation and Met-oxidation, respectively. The potential crosslinking sites considered for DSS were lysine and N-termini. All the identified linked peptides were validated manually for b- and y-ion assignment and were only included in the final list if precursors of the respective peptides contained doublets with a mass difference of 12.0753 Da (mass difference between DSS-H_{12} and DSS-D_{12}).

The mass spectrometry measurements and majority of crosslink identifications were performed by Dr. Javaid Y. Bhat (MPI of Biochemistry, Martinsried, Germany).

Crosslinks were visualized for analysis with XVis and Xlink Analyzer (Grimm et al. 2015; Kosinski et al. 2015).

4.6.1.11 Electrophoretic mobility shift assay (EMSA)

EMSAs with CtSkn7 were performed with 15–20 µM protein and 200 nM fluorescently labelled DNA-oligonucleotides in buffer E supplemented with 250 µg mL^{-1} Poly(dI-dC) (Sigma) in a total volume of 20 µL. Prior to electrophoresis, samples were incubated at RT for 20-30 min.

EMSAs with HsHsf1 were performed with 4 µM protein and 200 nM fluorescently labelled DNA-oligonucleotides in buffer E supplemented with 250 µg mL^{-1} Poly(dI-

dC) (Sigma) in a total volume of 20 µL. Samples were incubated at 42 ºC (heat shock) or on ice for 20 min.

After adding 0.75 µL DNA-loading dye, all samples were separated for 2 h on 3 % agarose gels in Tris-Acetate-EDTA buffer at 50 V in the same apparatus as used for DNA gels (or on 5 % polyacrylamide native gel supplied with 5 % glycerol in native running buffer at 150 V for ~90 min). The electrophoresis chamber was ice-cooled and protected from light with aluminum foil. The gels were recorded with an Image reader FLA-2000 (Fujifilm, Minato) and analyzed with the Aida Image analyzer software package (Raytest).

For dissociation constant determination, the band intensity of the protein-DNA complex and the free DNA was quantified and plotted. The resulting graph was fitted using the Hill equation in Origin (OriginLab) whereby n was fixed (n=1) to avoid over interpretation regarding cooperativity since the exact stoichiometry of the complex is unknown.

$$y = START + (END - START) * \frac{x^n}{k^n + x^n}$$

Equation 1

4.6.1.12 Fluoresence polarization measurement

For fluorescence polarization measurements, 5 nM of fluorescently labled DNA-oligonucleotides were incubated with a concentration series of HsHsf1-CtSkn7-Chimera in buffer E supplied with 1 mM $MgCl_2$. The total volume per sample was 50 µL. The titration series was performed in FIA plates (F-Form, non-binding; Greiner bio-one), incubated for 20 min and fluorescence polarization was read out by a GENios Pro plate reader (Tecan) with a wavelength of 485 nm for excitation and 535 nm for emission. Values were corrected for blank measurement and fit will the Equation 1.

4.6.2 Protein crystallization and structure determination

4.6.2.1 Crystallization

HsHsf1(1-120)–HSE complex – Formation of the HsHsf1(1-120)–HSE complex was initiated by mixing the protein solution at 24 mg mL^{-1} with ds-DNA at a volume

ratio of 2.34:1, resulting in final concentrations of 1.25 mM protein and 0.75 mM DNA-duplex, respectively. After adding DTT to a final concentration of 20 mM, the mixture was incubated for 30 min on ice. Initial crystals in space group $P22_12_1$ were obtained by the sitting drop vapor diffusion method using the PEGs Suite (Qiagen, Hilden) crystallization screen at 4°C. Crystallization drops were set up by mixing 100 nL sample with 100 nL reservoir using the robotics setup of the MPIB Crystallization Facility. The final crystals grew in a hanging-drop vapor diffusion setup at 20°C after equilibrating a mixture of 1 µL HsHsf1(1-120)–HSE complex and 1 µL reservoir containing 0.2 M Ammonium sulfate and 21 % PEG-3350 as precipitant. For cryoprotection, crystals were dipped for 20 s into 25 % PEG-3350 containing 10 % glycerol and flash-frozen in liquid nitrogen.

HsHsf1(1-120)–SatIII complex – The HsHsf1(1-120)–SatIII complex was formed by mixing the protein solution of 24 mg mL^{-1} with ds-DNA, as described above. After adding DTT to a final concentration of 20 mM, the mixture was incubated for 30 min on ice. Crystals in space group *C*2 were obtained by the sitting drop vapor diffusion method using the PEGs Suite (Qiagen) crystallization screen at 4°C by mixing 100 nL sample with 100 nL reservoir using the robotics setup of the MPIB Crystallization Facility. The best crystals grew using a reservoir containing 0.09 M MgAc and 31 % PEG-4000 as precipitant. For cryoprotection, crystals were dipped for 20 s into 34 % PEG-4000 containing 5 % glycerol and flash-cooled in liquid nitrogen.

CtSkn7(40-143)–HSE complex – The complex was formed by mixing a 20.15 mg mL^{-1} protein solution with ds-DNA in a volume ratio of 2.60:1 resulting in final concentrations of 1.16 mM CtSkn7(40-143) and 0.69 mM HSE duplex. The solution was then incubated on ice for 30 min. Crystals of CtSkn7(40-143)–HSE in space group $C222_1$ were obtained using the PEGs Suite (Qiagen) crystallization screen at 18°C using the previously mentioned robotics setup. The final crystals grew in hanging-drop vapor diffusion setup at 20°C after equilibrating a mixture of 1 µL CtSkn7(40-143)–HSE complex and 1 µL reservoir containing 0.18 M tri-ammonium citrate and 11 % PEG-3350 as precipitant. The crystals were transferred stepwise into cryosolution containing 17 % PEG-3350 and 15 % glycerol.

CtSkn7(40-143)–SSRE complex – The complex was formed using the same protocol as for CtSkn7(40-143)–HSE. Initial crystals of the CtSkn7(40-143)–SSRE complex in space group *C*2 were obtained using the PEGs Suite (Qiagen) crystallization screen at 18°C at the 100+100 nL scale. The final crystal grew in hanging-drop vapor diffusion setup at 20°C by mixing 1 µL CtSkn7(40-143)–SSRE complex with 1 µL precipitant containing 0.1 M Hepes-NaOH pH 7.5 and 25 % PEG-2000 MME. For cryoprotection, crystals were soaked in 30 % PEG-2000 MME containing 5 % glycerol.

CtSkn7(160-209) – Initial crystals of CtSkn7(160-209) were obtained using the pH Clear I Suite (Qiagen) crystallization screen at 18°C at 200 nL scale. The final crystals in space groups $C222_1$ and $P2_12_12_1$ grew in hanging-drop setup at a protein concentration of 37 mg mL^{-1} using 0.1 M Hepes-NaOH pH 7.0 and 20 % PEG-6000 or 21 % PEG-6000 as precipitant, respectively. For cryoprotection, the crystals were transferred stepwise into cryosolution containing 25 % PEG -6000 and 10 % glycerol.

CtSkn7(160-220) – Crystals of CtSkn7(160-220) in space group $P2_1$ were obtained using the Index crystallization screen (Hampton Research) at 20°C at 200 nL scale at a protein concentration of 82 mg mL^{-1}. The conditions were 0.1 M Bis-Tris pH 5.5, 0.2 M $MgCl_2$ and 25 % PEG-3350 as precipitant. The crystals were directly transferred into liquid nitrogen.

4.6.2.2 Structure solution, refinement and analysis

X-ray diffraction data were collected at the Macromolecular Crystallography (MX) beamlines ID23-2, ID29 and ID30A at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. All data were integrated and scaled with XDS (Kabsch 2010). Pointless and Scala (Evans 2006), Aimless (Evans & Murshudov 2013) and Truncate (French & Wilson 1978) were applied to convert the data to CCP4 format (Collaborative Computational Project 1994).

CtSkn7(160-209) crystal form I – Phases for this dataset were obtained by sulfur-single-wavelength anomalous diffraction (SAD) using SHELX-CDE, as implemented in HKL2MAP (Sheldrick 2010). Subsequently, the model was almost entirely auto-built using ARP/wARP (Langer et al. 2008) due the resolution of the

native dataset of 1.03 Å. The model of the coiled-coil domain CtSkn7(160-209) contains a half-tetramer.

CtSkn7(160-209) crystal form II – A CtSkn7(160-209) half-tetramer was found in the asymmetric unit by molecular replacement using Molrep (Vagin & Isupov 2001). Electron density for an additional CtSkn7(160-209) trimer could be observed in the resulting map of this partial structure solution and the missing model was built manually (Emsley & Cowtan 2004).

CtSkn7(160-220) – This structure was solved by molecular replacement using the CtSkn7(160-209) half-tetramer structure as search model in Molrep (Vagin & Isupov 2001). The asymmetric unit contains a CtSkn7(160-220) tetramer.

DBD-DNA co-crystal structures – The dataset of the CtSkn7(40-143)–HSE co-crystal was solved using molecular replacement with Molrep (Vagin & Isupov 2001). A search template using the *Kluyveromyces lactis* Hsf1-HSE crystal structure (PDB code 3HTS) was applied. The refined model of the CtSkn7(40-143)–HSE complex was used as search template for the other DBD-DNA complex structures. The asymmetric units of HsHsf1(1-120)–HSE and CtSkn7(40-143)–HSE contain one DBD and a single strand of DNA. The unit cells of CtSkn7(40-143)–SSRE and HsHsf1(1-120)–SatIII contain the entire biological assembly.

All models were refined with Refmac5 (Murshudov et al. 1997). Coot was employed for manual model building in alternating building-refinement cycles with Refmac5. Residues facing solvent channels with disordered side chains were modeled as alanines.

Coordinates were aligned with Lsqkab, Lsqman or the alignment algorithm of Pymol (Kleywegt 1999). Homology models were calculated with Modeller (Eswar et al. 2007). Modeller, using the ModLoop server or the Chimera interface to Modeller, was used to calculate the Linker regions between DBD and HR-A/B of HsHsf1 (Fiser & Sali 2003; Pettersen et al. 2004; Eswar et al. 2007). Energy minimization was done using the algorithm implemented in Chimera. Figures were generated and rendered with Pymol (http://www.pymol.org) and ESPript (Robert & Gouet 2014).

5 Results

5.1 Characterization of human Hsf1

In order to explore structural features of human Hsf1 (HsHsf1), we expressed full-length HsHsf1 in *E. coli*. Purified HsHsf1 protein was analyzed regarding its oligomerization state and three dimensional domain topology. Furthermore, its DNA-binding properties were verified.

5.1.1 Purification of HsHsf1

Structural analysis of proteins requires a highly pure and homogenous sample. Therefore we established a purification procedure harnessing HsHsf1 expressed as His_6-SUMO fusion protein in *E. coli*. The protein was purified using a protocol with only two chromatographic steps to shorten the purification process since large unfolded regions in the protein render it sensitive to degradation by proteolysis. Human Hsf1 was captured from lysate on a Ni^{2+}-Tris(carboxymethyl)ethylene diamine (TED) column and washed extensively with a high-salt buffer to remove contaminating proteins and bound bacterial DNA. The His_6-SUMO tag of the protein was cleaved off after elution for 2-3 hours, and the mixture was directly loaded onto a gel filtration Superdex 200 column. The peak fractions were concentrated up to 10 µM HsHsf1 and stored at -80 ºC. The yield was 0.3 mg HsHsf1 per 1 L of bacterial culture. The protocol mainly yields the monomeric form as indicated by the retention volume in SEC (Figure 15). In order to prevent oligomerization of human Hsf1, several parameters were carefully controlled during purification. The temperature of expression was critically required to be 20 ºC or lower, to slow down the protein synthesis rate. Furthermore, the protein showed a strong tendency to oligomerize, if the concentration of the purified protein exceeded 10 µM or the glycerol concentration of the buffer solution was below 10 % (v/v). Apart from its large content of unstructured parts, oligomeric heterogeneity render full-length HsHsf1 not amenable for standard protein crystallography.

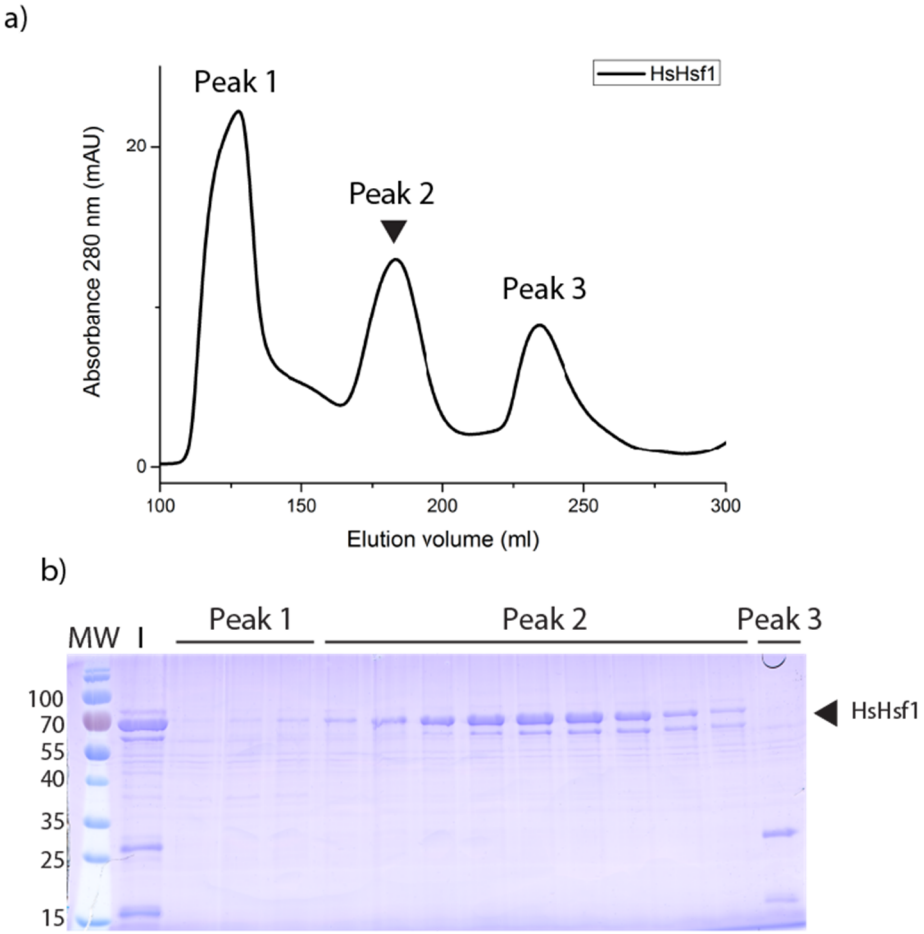

Figure 15 Purification of HsHsf1 expressed in *E. coli*. a) Elution profile of HsHsf1 during Superdex 200 26/60 SEC. Absorbance of light at a wavelength of 280 nm is shown. Peak 1 corresponds to high molecular weight impurities, Peak 2 to monomeric HsHsf1, Peak 3 to Ulp1 protease and cleaved His_6-SUMO tag. b) SDS-PAGE analysis of the protein content in the elution fractions corresponding to in the major absorbance peaks of the Superdex 200 column separation in a). Coomassie-stained 12 % polyacrylamide gel; I, Input; MW, Molecular weight marker in kDa.

5.1.2 Purified HsHsf1 forms trimers upon heat shock

Because the retention volume in SEC only provides a rough estimate of the oligomeric state, we applied native mass spectrometry (native MS) to receive additional information about the oligomeric state in solution. This technique can provide information about native protein complexes such as stoichiometry and composition. Analysis of the preparation of human Hsf1 revealed only ionization states of monomeric HsHsf1 (Figure 16). However, when the sample was heated to 36 ºC for 5 minutes prior injection, a trimeric species was also detected. This result is consistent with the current model of Hsf1 (Anckar & Sistonen 2011).

Interestingly when the sample was heated to 42 ºC for 5 minutes prior to measurement, trimeric as well as a low amount of pentameric species were detected (Figure 16). Occurrence of a pentameric species is intriguing since the proximal promoter of HSP70.1 contains 5 inverted nGAAn repeats (Figure 7). The residual HsHsf1 monomer signal may be explained by dissociation in the gas phase.

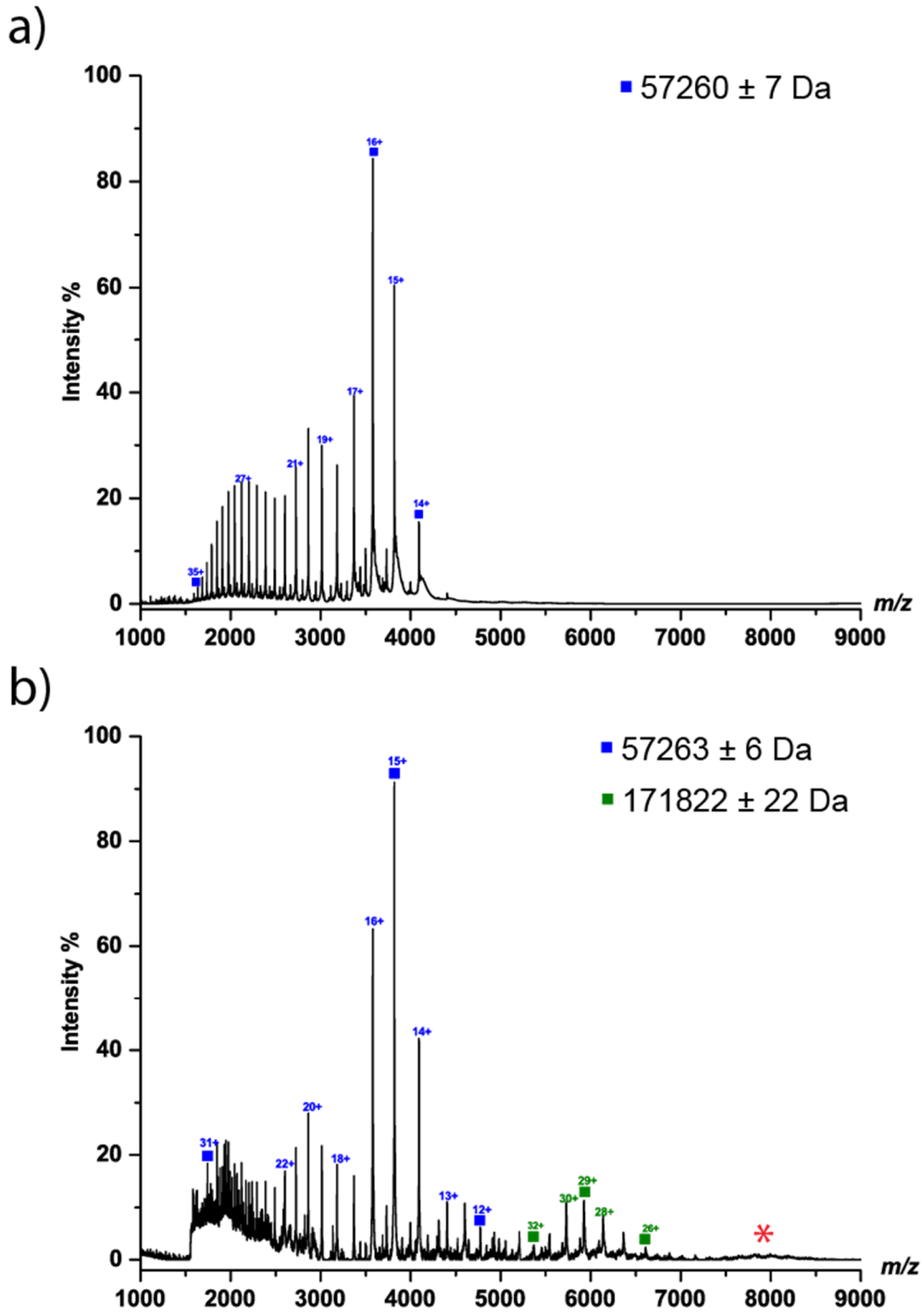

Figure 16 Native mass spectrometry analysis of HsHsf1 in the latent (a) and activated (b) state after 5 min incubation at 42 ºC. Symbols indicate charge-state distributions; charge states are shown for some peaks. Analysis reveals a temperature-dependent intrinsic oligomerization capacity of HsHsf1. Different oligomeric species are indicated by different colors (– Monomer, – Trimer, * Pentamer). The calculated mass around the m/z values of the respective protein complexes and the accuracy of mass values calculated from the different m/z peaks were 57260±7 Da for the monomer in the latent state, 57263±6 Da for the monomer and 171822±22 Da for the trimer in the activated state. MS analysis was performed by Dr. Javaid Y. Bhat.

5.1.3 Analysis of intra- and inter-molecular interactions in the domain topology of HsHsf1

Next, we were curious to understand how the monomeric and oligomeric state are maintained and which domains interact with each other in order to do so. Previous studies revealed that the trimerization of metazoan Hsf1 is suppressed by the C-terminal leucine zipper domain HR-C (Rabindran et al. 1993; Farkas et al. 1998). It was hypothesized that HR-C binds back to HR-A/B by formation of a coiled-coil structure by careful analysis of different mutant and truncation constructs. Zou and colleagues were able to show that all three coiled-coil domains (HR-A, HR-B and HR-C) are required to stabilize the monomeric state (Zuo et al. 1995). Their results were consistent with a working model of the HR-A, B and C forming an intramolecular antiparallel triple-stranded coiled-coil. Nevertheless biochemical data showing a direct interaction between HR-A/B and HR-C had been missing.

Chemical crosslinking combined with mass spectrometry (CXMS) has the potential to reveal the molecular architecture of proteins and protein complexes. We applied this technique to purified HsHsf1 using the bifunctional compound disuccinimidylsuberate (DSS) and the heterobifunctional compound DMTMM (4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride) as crosslinking agents. DSS crosslinks primary amines (lysine sidechains, amino terminals) that are less than 30 Å apart, as shown by the Cα positions of DSS-crosslinked lysine residues in model proteins with a known structure (Leitner et al. 2013).

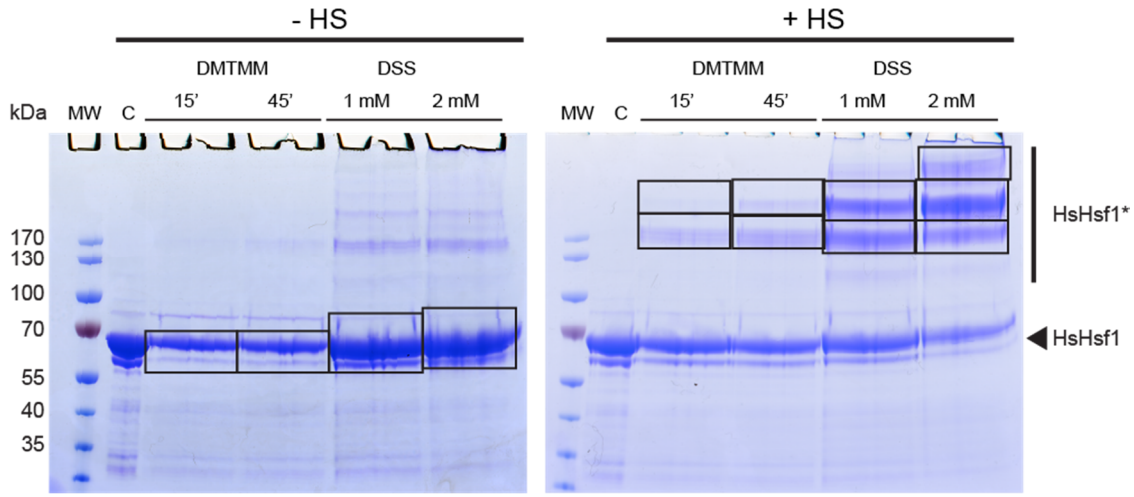

Figure 17 SDS-PAGE analysis of HsHsf1 crosslinking experiment. Both crosslinkers were applied under two different conditions (15 min or 45 min for DMTMM and, 1 mM or 2 mM for DSS).

Samples were analyzed by SDS-PAGE followed by Coomassie staining. MW, Molecular weight marker; C, non-crosslinked HsHsf1 control. For the no heat-shock condition (-HS), only bands at the height of monomeric HsHsf1 were analyzed. For the heat-shock condition (+HS), only shifted bands corresponding to intermolecular crosslinked HsHsf1 (HsHsf1*) were analyzed. The mass spectrometry measurement and majority of crosslink identification were performed by Dr. Javaid Y. Bhat.

For identification of crosslinked peptides using mass spectrometry (MS) a 1:1 isotopic mixture of light (H_{12}) and heavy (D_{12}) DSS was used. These peptides show a MS-MS doublet with a mass difference of 12.0753 Da in the spectra (Leitner et al. 2013). The compound DMTMM activates carboxyl groups (Asp and Glu) for the reaction with a primary amine (Leitner et al. 2014). The spacer distance is thus zero.

Purified latent or heat-shocked HsHsf1 was incubated in the presence of each compound in order to chemically crosslink interacting segments. After quenching, the samples were applied to SDS-PAGE in order to separate crosslinked products by size (Figure 17). For the analysis of monomeric HsHsf1, only gel bands which had not shifted in size were cut out and analyzed by trypsin digestion followed by MS analysis of peptide fragments. Thus, only intramolecular crosslinks were included in the analysis. Spurious gel bands corresponding to intermolecular crosslinked HsHsf1 were not analyzed. Crosslinks were assigned as robust if found in two experimental conditions and in at least 60 % of analyzed gel bands (Figure 17, Table 1, Table 2, and all identified crosslinks in Table 5). The MS-based identification procedure for crosslinked peptides was adjusted to a false discovery rate of ~5%.

We found crosslinks between Lys80-Lys206, Lys80-Lys126 and Lys126-Lys139 residues which are within the region of DBD and HR-A/B, respectively. These crosslinks are consistent with HR-A forming an anti-parallel arrangement with HR-B in monomeric Hsf1. Otherwise Lys80 and Lys206 would be far apart. Other crosslinks (Met1-Lys184, Lys80-Lys206, Lys80-Lys298, Lys118-Lys524) suggested that the DBD folds back towards the coiled-coil domains and contacts the regulatory domain. Nevertheless, our crosslink data did not indicate a direct interaction between HR-A/B and HR-C, but reproducibly revealed crosslinks which

are consistent with the previously suggested compact state of monomeric Hsf1 (Figure 18).

Next, we analyzed HsHsf1 activated by a 10 min heat shock at 42 ºC using DSS-crosslinking. For heat-shocked HsHsf1, only shifted gel bands, corresponding to oligomeric HsHsf1, were analyzed. Identified crosslinks were classified as intermolecular, if the crosslinked peptides showed overlapping sequences. Analysis of these gel bands revealed robust intermolecular DSS crosslinks (Lys131-Lys131, Lys178-Lys184, Lys206-Lys206, Lys298-Lys298), which are consistent with formation of a parallel coiled-coil helix bundle involving the HR-A/B domains of individual Hsf1 proteins. The crosslink between Lys298-Lys524 suggested that the TAD is positioned towards the RD and DBD.

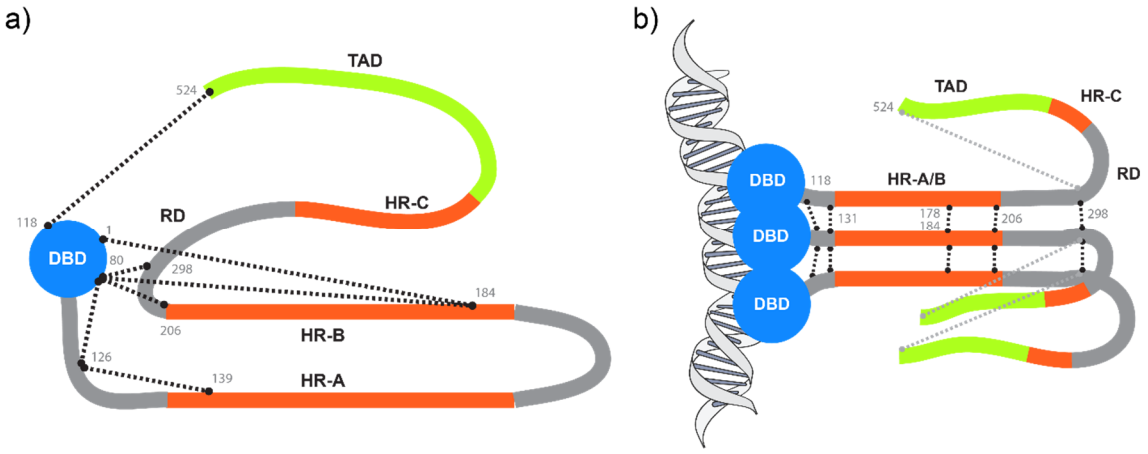

Figure 18 Probing intra and intermolecular contacts of HsHsf1 by chemical crosslinking coupled to mass spectrometry for the latent (a) and activated state (b). Identified crosslinks are depicted by dotted lines between different sections/domains of HsHsf1. Residues involved in particular crosslink are indicated by their respective residue number. Only a selection of robust DSS crosslinks are depicted for clarity. Crosslinks were assigned as robust if found in two experimental conditions and in at least 60% of analyzed gel bands (for gel see Figure 17).

DMTMM crosslinking was less efficient (Figure 17) and the majority of crosslinks were found within the DBD (e.g. Met1 to Glu55/Glu109/Glu113) (Figure 17, Table 2 and Table 6 (all identified crosslinks)).

Table 1 Robust DSS crosslinks

No heat shock	
Pos1	Pos2
1	118
1	184
80	118
80	126
80	139
80	184
80	206
80	298
116	298
118	139
118	188
118	298
118	524
126	139
188	298

Heat shock		
Pos1	Pos2	Inter SU
1	59	-
1	62	-
1	91	-
1	116	-
1	118	-
1	126	-
1	131	-
1	148	-
1	206	-
1	298	-
59	91	-
80	118	-
116	118	-
116	131	-
118	126	Y
118	131	-
126	126	Y
126	131	Y
178	184	Y
184	298	-
188	206	-
206	206	Y
206	208	-
206	298	-
208	298	-
298	298	Y
298	524	-

Table 2 Robust DMTMM crosslinks

No heat shock	
Pos1	Pos2
1	55
1	109
1	113
131	135
206	339

Heat shock		
Pos1	Pos2	Inter SU
1	55	-
1	109	-
1	113	-
1	128	-
55	91	-
126	128	Y

5.1.4 HsHsf1 binding to HSE DNA

Next, we tested whether the purified recombinant HsHsf1 had the same DNA-binding characteristics as authentic HsHSf1. Monomeric HsHsf1 does not exhibit stable binding to a trimeric HSE, whereas trimers do. Therefore electrophoretic mobility shift assays were performed (Figure 19). The purified, mostly monomeric HsHsf1 exhibited a smear, suggesting the formation of transient complexes. Upon heating to 42 ºC, a time-dependent development of a sharp band with decreased electrophoretic mobility was observed, while the band for unbound DNA diminished. Incubation of HsHsf1 at 42 °C for 5 min resulted in maximal activation of the protein as measured by EMSA and was consistent with our findings in native MS. HsHsf1 is a temperature and concentration sensitive protein which likely predisposes the protein to spontaneous trimerization. This was displayed, to some extent, by HsHsf1 through the formation of DNA-binding oligomers in the non-heat shock condition.

The observed temperature dependent binding of HSE-DNA is consistent with previous findings for endogenous HsHsf1 (Sarge et al. 1993).

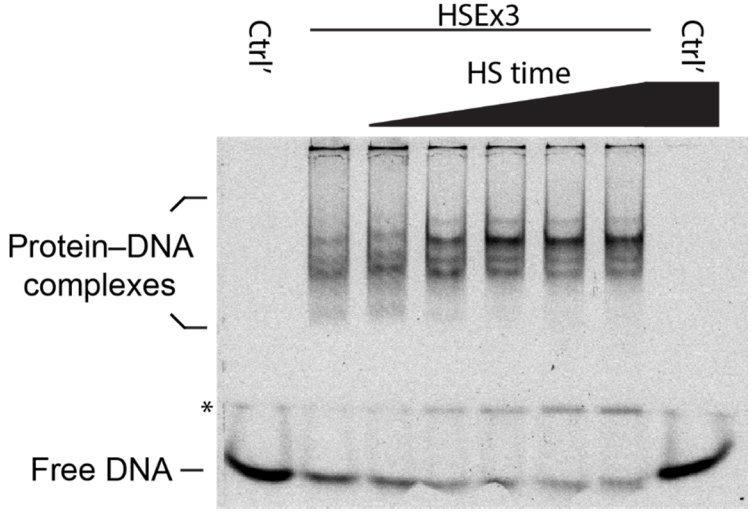

Figure 19 HsHsf1 binds strongly to HSE DNA-oligonucleotides upon heat shock induced oligomerization in EMSA. A 5% polyacrylamide gel shows the electrophoretic migration pattern of HsHsf1-DNA complexes, measured by fluorescence detection of fluorescein labeled HSE-DNA. HsHsf1 was activated by heat shock (HS) at 42 ºC (30 s, 2 min, 5 min, 10 min, 20 min). Control lanes (Ctrl') represent the migration of labelled DNA in absence of HsHsf1. Free DNA and DNA bound by protein complexes are indicated, * putative HsHsf1 degradation product associated with DNA.

5.2 Strategies for structural characterization of human Hsf1 using X-ray crystallography

Studying protein structure by X-ray crystallography critically depends on the ability of the target proteins to form a highly ordered protein crystal lattice. Hsf1 proteins, especially its human homolog HsHsf1 contain high amounts of unstructured regions (60 %), which are incompatible with forming an ordered crystal lattice (Anckar & Sistonen 2011; Pattaramanon et al. 2007; Drozdetskiy et al. 2015). Therefore a series of fusion, truncation and mutant proteins which excluded poorly structured parts of the protein was generated in the precursor study (Neudegger 2012). However, this strategy has not led to successful crystallization, since all constructs comprising the HR-A/B domain showed an uncontrollable oligomerization behavior resulting in heterogeneous preparations. In order to achieve a homogenous preparation, subsequently several strategies were tested, which are described below.

5.2.1 Stabilization of truncated HsHsf1 by an engineered GCN4 zipper

Because the studied truncation constructs of human Hsf1 showed inevitable oligomerization behavior, we fused HsHsf1 amino acids 1-192 with an engineered trimeric GCN4 (triGCN4) leucine zipper for purification and analysis (Figure 20) (Hernandez Alvarez et al. 2008). The GCN4 fusion partner was designed to facilitate structure determination of coiled-coil proteins.

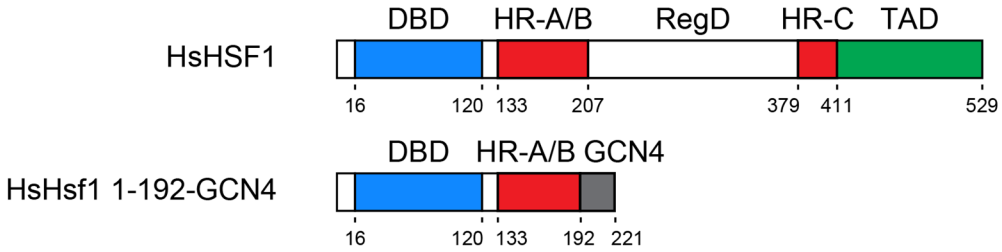

Figure 20 Multidomain protein constructs of HsHsf1 for functional and structural analysis. Fusion of unstable coiled-coil constructs to engineered GCN4 zipper has proven to increase the stability of purified proteins for structural studies (Hernandez Alvarez et al. 2008).

This construct (HsHsf1-GCN4) was expressed as a soluble protein in *E. coli* that could be purified to homogeneity using IMAC, cation exchange chromatography and SEC (Figure 21). The protein was in a likely oligomeric state as judged by retention volume and could be concentrated beyond 10 µM. Interestingly, an HsHsf1 coiled-coil construct triGCN4 fusion (GCN4-HsHsf1(145-192)-GCN4) was insoluble when expressed in *E. coli* and refolding attempts failed (data not shown). The truncation construct HsHsf1(133-193) was partly soluble in *E. coil* lysate, but precipitated quantitatively upon IMAC elution.

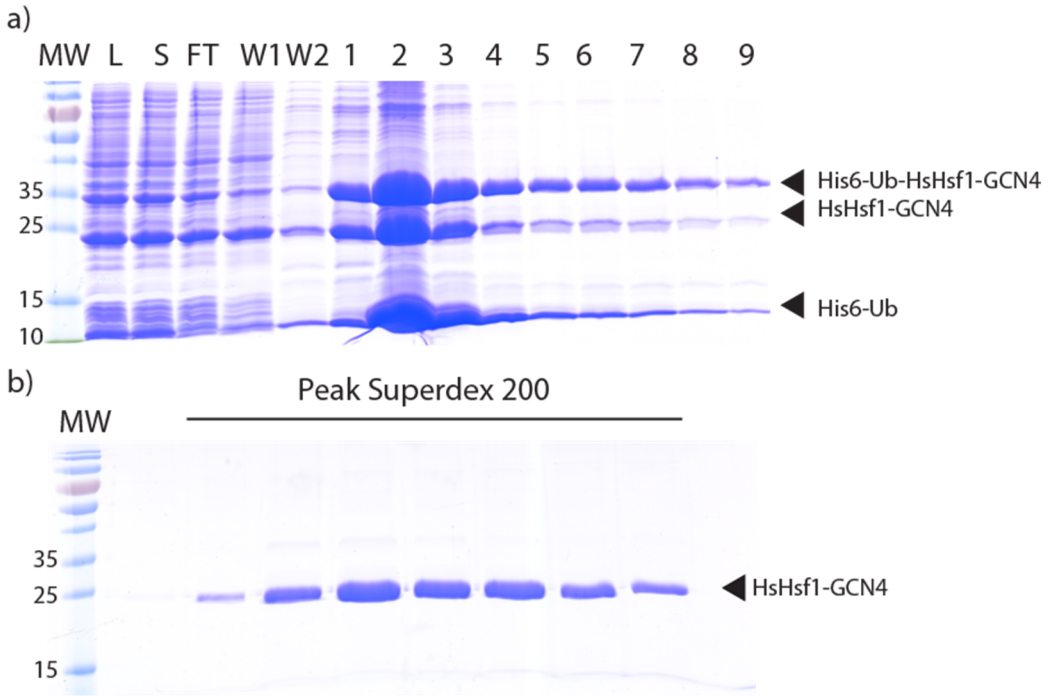

Figure 21 Purification of a HsHsf1-GCN4 fusion protein. a) SDS-PAGE analysis of the total lysate (L), the soluble fraction (S), and fractions from IMAC. Flow through (FT), wash fractions (W1, W2) and elution fractions 1-6; MW, Molecular weight marker in kDa. b) The protein content in the major peak of the size exclusion chromatography step on Superdex 200 column was analyzed by SDS-PAGE followed by Coomassie staining.

5.2.1.1 Oligomeric state of HsHsf1-GCN4

In order to characterize the oligomerization state of the purified HsHsf1-GCN4 fusion protein in more detail, we probed it initially by Ethylene glycol bis[succinimidylsuccinate] (EGS) crosslinking. EGS is a homobifunctional crosslinker which reacts with the N-termini of proteins and the ε-amino group of

lysine and covalently links them via a 16.1 Å spacer arm. Thus the crosslinking product shifts to higher molecular weight in denaturing polyacrylamide gels.

Depending on the EGS concentration, multiple bands for cross-linked HsHsf1-GCN4 were observed. The band patterns suggest the existence of a trimer or tetramer, with a propensity to associate to larger oligomers (Figure 22).

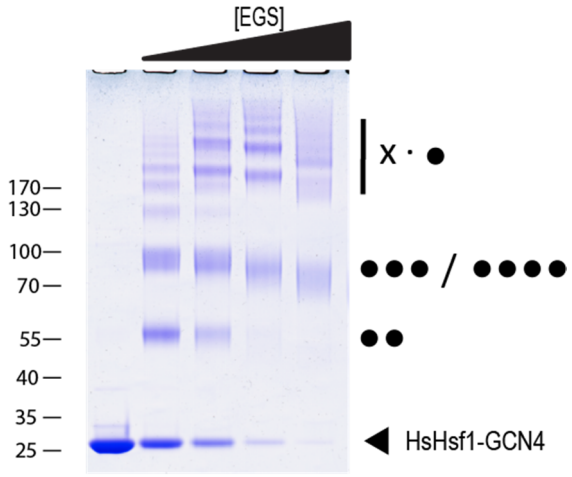

Figure 22 EGS crosslinking of HsHsf1-GCN4. SDS-PAGE analysis of EGS crosslinking experiments of HsHsf1 1-192-GCN4. The purified protein was incubated with increasing amounts of the cross-linker EGS, quenched and subsequently analyzed by SDS-PAGE. Stoichiometry of crosslinked oligomers is indicated by the number of dots ●. Molecular weight marker in kDa is indicated on the left.

Whereas EGS crosslinking depends on the accessibility of lysine residues, size exclusion chromatography (SEC) coupled to multi-angle light scattering (MALS) offers absolute determination of the average molecular mass in solution. SEC-MALS revealed that HsHsf1-GCN4 was homogeneous and probably tetrameric (~110 kDa, theoretical mass monomer: 25.4 kDa) (Figure 23).

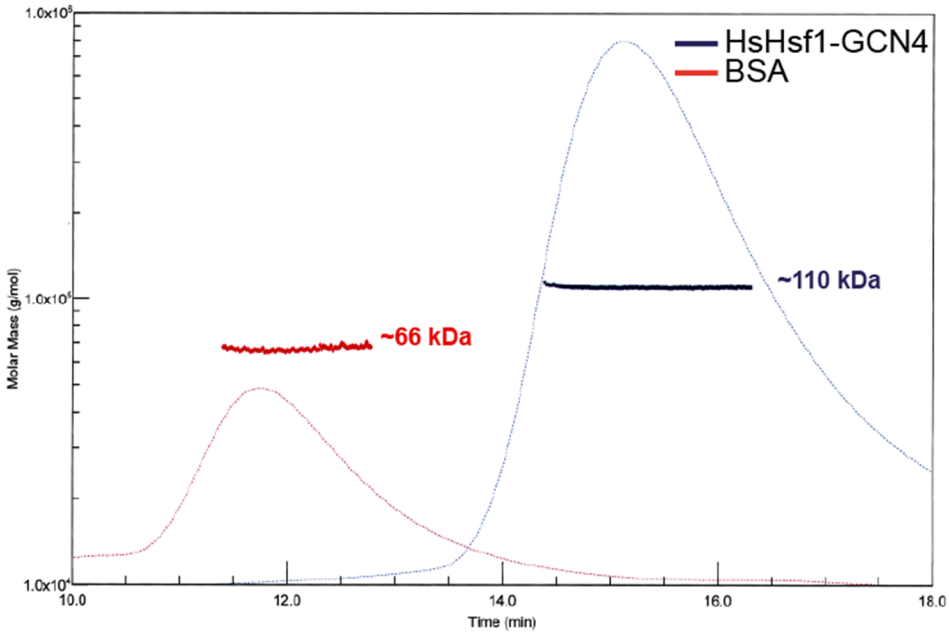

Figure 23 MALS coupled to SEC of purified HsHsf1-GCN4. A chromatogram with UV absorbance and MW is shown. Horizontal lines across the peaks indicate molar mass and homogeneity. Calculated molar masses are indicated. Construct shows homogenous behavior and the approximate MW of a tetramer. Experiment was performed by Dr. Manajit Hayer-Hartl.

5.2.1.2 Crystallization of HsHsf1-GCN4

Since the biochemical characterization of the fusion construct revealed a defined oligomeric state and a homogenous sample preparation, we set up crystallization trials with the purified protein to gain insights into HsHsf1 structural organization. Needle-shaped micro-crystals were obtained, which diffracted at ESRF beamline ID23-2 to a maximum of 9 Å resolution. Attempts to grow larger, better diffracting crystals failed (in absence and presence of HSE DNA-oligonucleotides). Thus experiments with HsHsf1-GCN4 were stopped (Figure 24).

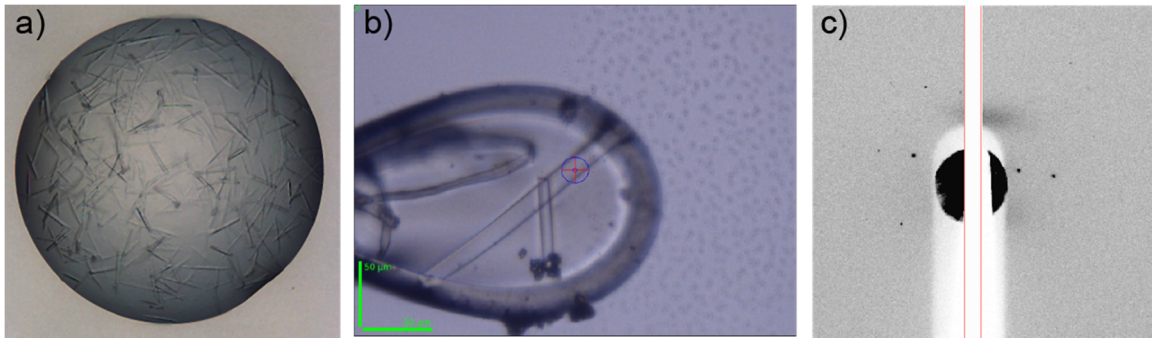

Figure 24 Crystallization and diffraction of crystals of HsHsf1-GCN4. Crystals showed needle shape (a, b) and exhibited low X-ray diffraction (c). For size estimation: crystal drop in a) is composed of 200 nL liquid. b) shows mounted crystals, the blue circle indicates X-ray beam size and position.

5.2.2 Characterization of HsHsf1 DBD

Since all HsHsf1 constructs involving the HR-A/B coiled-coil domain proved to be highly heterogeneous, we focused our structural studies on the isolated DBD from HsHsf1. As second, independent approach we structurally characterized an Hsf homolog from a thermophilic eukaryote. For structural characterization of the HR-A/B domain, we used an Hsf family protein from a thermophilic eukaryote as described below (paragraph 5.3).

5.2.2.1 Purification of HsHsf1 DBD

Sequence alignments of Hsf family proteins showed a conserved region of ~110 residues at the N-termini of the proteins. Thus as a putative DBD construct, HsHsf1 residues 1-120 were cloned into the vector pProEx-HtB and expressed as a TEV protease-cleavable, N-terminal His_6-tagged fusion protein in *E. coli*. As a consequence of the cloning strategy, HsHsf1(1-120) contained the additional N-terminal sequence GAMGSGILRGG after TEV protease treatment. In a first purification step, the fusion protein was enriched by immobilized metal affinity chromatography (IMAC) using a Ni^{2+} chelating agarose. After cleavage of the affinity capture tag by TEV protease, the construct was further purified by cation exchange chromatography on a SourceS column. As for HsHsf1-GCN4 and other DBD containing constructs, this step removed bound bacterial DNA and enhanced purity (paragraphs 5.2.1). Finally, size exclusion chromatography on a Superdex

200 column was applied as a polishing step. The resultant material was of apparent homogeneity (Figure 25).

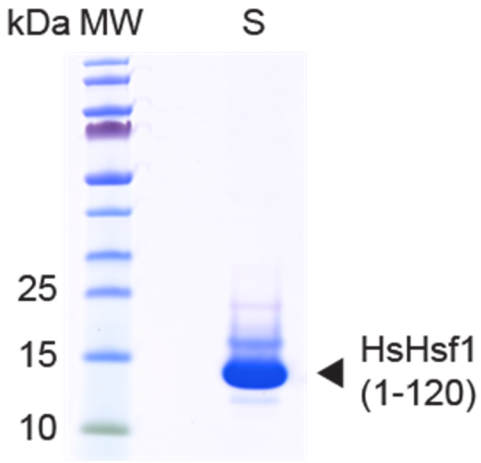

Figure 25 Coomassie-stained SDS-PAGE gel with concentrated sample of HsHsf1 DBD (HsHsf1(1-120)) purification. MW, Molecular weight marker; S, Sample.

5.2.2.2 Crystallization of HsHsf1 DNA-binding domain complexes

Transcription factors are frequently amenable to structural characterization by X-ray crystallography only when bound to their cognate DNA. Therefore we focused our crystallization efforts on complexes between the HsHsf1 DBD and DNA constructs. We obtained crystals of the HsHsf1 DBD in complex with a DNA oligonucleotide resembling the binding motifs within promoters of Hsf1 target genes (HSE ggTTCtaGAAcc (2x nGAAn in tail-to-tail orientation)). Crystal screening was performed using 200 nL droplets with the PEGs Suite (Qiagen) and an automated robotics setup. Subsequently the reservoir composition of the initial condition was refined to 0.2 M Ammonium sulfate and 21 % PEG-3350 (Figure 26 a). In addition we were able to co-crystallize HsHsf1 with a DNA oligonucleotide resembling the binding motif in nuclear stress bodies (SatIII (cGGAATGGAATg)). The initial crystals of the DBD-SatIII complex were found using the same commercial screen. The composition of the precipitant was 0.09 M MgAc and 31 % PEG-4000 (Figure 26 b).

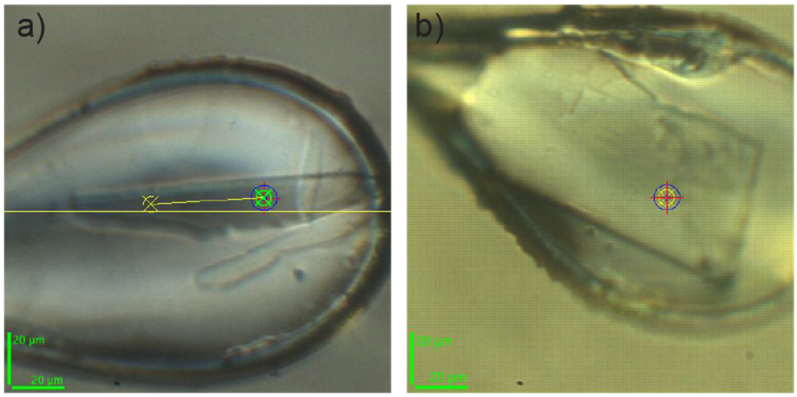

Figure 26 Crystals of HsHsf1 DBD in complex with HSE and SatIII containing DNA-oligonucleotides, respectively. Shown are the mounted crystals which were used to obtain the datasets of the HsHsf1 DBD-HSE (a) and SatIII (b) complex structures. The blue circles have 10 µm diameter and indicate X-ray beam size and position.

5.2.2.3 Structural analysis of human Hsf1 DBD

5.2.2.3.1 Features of HsHsf1 DBD and HsHsf1 DBD-DNA contacts

The co-crystals diffracted X-rays to 2.91 Å and 2.55 Å resolution (Table 3), respectively. We were able to solve the structures of HsHsf1 DBD-HSE and DBD-SatIII by molecular replacement. While in the DBD-HSE structure the asymmetric unit (a.u.) contained one DBD chain and one strand of the DNA molecule and the biological assembly is completed by two-fold crystal symmetry, the HsHsf1 DBD-SatIII a.u. contained the complete biological assembly.

Table 3 Data collection and refinement statistics for HsHsf1 DBD-DNA complexes

	HsDBD–HSE	HsDBD–SatIII
Data collection		
Space group	$P22_12_1$	$C2$
Cell dimensions		
a, b, c (Å)	37.44, 39.82, 94.38	104.81, 112.68, 39.87
α, β, γ (°)	90, 90, 90	90, 92.56, 90
Wavelength (Å)		
Resolution (Å)	47.19 – 2.91 (3.09 – 2.91)[a]	39.83 – 2.55 (2.66 – 2.55)
R_{merge}	0.138 (0.821)	0.130 (0.669)
$I/\sigma I$	9.0 (2.1)	9.3 (2.2)
Completeness (%)	96.9 (82.7)	99.6 (97.2)
Redundancy	4.9 (4.8)	4.4 (4.1)
Refinement		
Resolution (Å)	30 – 2.91	30 – 2.55
No. reflections	3137	14354
R_{work} / R_{free}	0.223 / 0.267	0.190 / 0.234
Number of atoms		
Protein	810	1669
DNA/Mg^{2+}	243	488
Water	2	108
B-factors		
Protein	65.38	34.69
DNA/Mg^{2+}	51.37	24.39
Water	35.84	27.75
R.m.s. deviations		
Bond length (Å)	0.005	0.011
Bond angles (°)	0.883	1.397

The three crystallographically independent copies of the HsHsf1 DBD showed little deviation from each other and were essentially identical (Cα residual mean square deviation (r.m.s.d.) of 0.570-0.658 Å). The comparison to previously solved Hsf1 DBD structures showed similar results (r.m.s.d. of 1.488-1.952Å for 2LDU, the apo HsHsf1 DBD NMR structure, and r.m.s.d. of 1.223 Å for 3HTS, the DBD of *K. lactis* in complex with HSE DNA (Littlefield & Nelson 1999)). The fold of the DBD had a winged-helix-turn-helix topology (Figure 27 a). Previous crystal structures of Hsf1 DBDs in isolation or in complex with DNA were determined using constructs which were ~20 residues shorter at the C-terminus. These residues have been shown to

be critical for DNA binding in yeast (Flick et al. 1994). The construct we used for our structural studies additionally included this part (aa 101-120) (Figure 27 a, b).

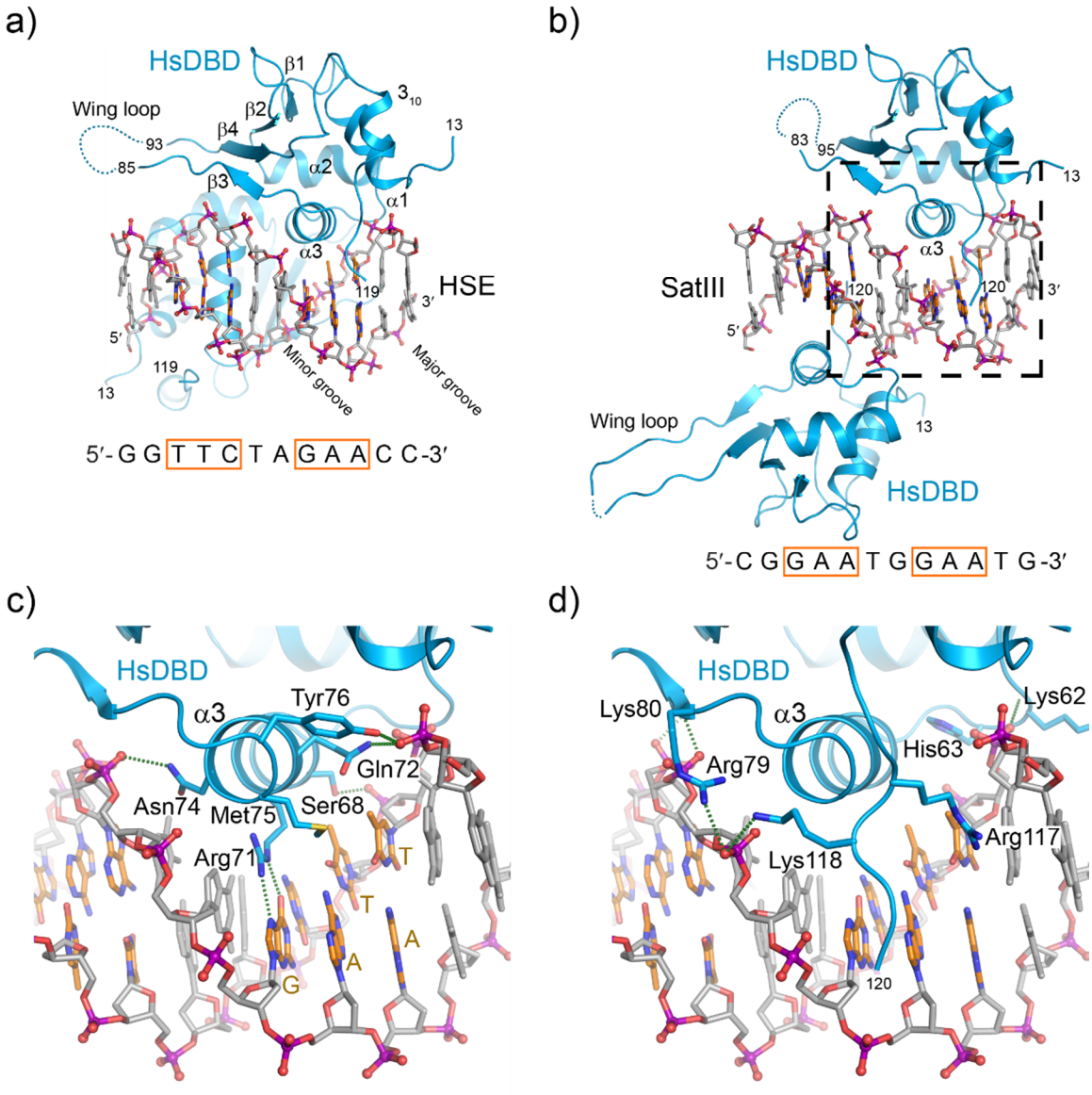

Figure 27 Crystal structures of HsHsf1 DBD-HSE complexes reveal the structural basis for HSE recognition and role of C-terminus. a) Crystal structures of the HsHsf1 DBD (HsDBD)-HSE complex, and b), the HsHsf1 DBD-SatIII complex. Protein is depicted in ribbon representation; the copies in the DBD-HSE complex are symmetry related. DNA is shown in stick representation with the phosphate backbone and the GAA motifs highlighted. Peptide chain termini are indicated. DNA sequence is shown below the models. c) and d) Detailed view of the DNA-contacts of the recognition helix 3 and the DBD C-terminus. Side chains of critical residues are shown in stick representation with hydrogen bonds represented by dotted lines. The C-terminal tail of the DBD is omitted for clarity in c).

This additional part formed a short 3_{10} helix and continued into the linker to the coiled-coil domain. The structure of this C-terminal part is stabilized by hydrophobic interactions of Phe104, Leu113 and Ile115 with the hydrophobic core of the domain. In addition, the backbone amide of Lys118 probably forms a hydrogen bond with the carbonyl group in the backbone of Met75. The wing of the DBD does not make any contacts with the DNA, but reaches out to adjacent DBD units. In the crystal structures, the wing loops were only partially structured, showing a similar behavior to the wing of KlHsf1 in the DNA complex structure (Littlefield & Nelson 1999). α-Helix 3, which serves as so called recognition helix of the helix-turn-helix fold is inserted into the major groove of the DNA double strand (Figure 27 c). An array of side chains of the highly conserved residues Ser68, Gln72, Asn74, Met75 and Tyr76 position α-helix 3 in the groove by formation of hydrogen bonds with the phosphate backbone of the DNA double helix. A pair of positively charged side chains in the N- and C-terminal connectors of the recognition helix (Lys62, His63 and Arg79, Lys80) intensify the interaction by hydrogen bonding and electrostatic interactions. When the domain is correctly positioned above the HSE element GAA, Arg71 reads out the Hoogsteen face of the guanine group by hydrogen bonding with its guanidine group. Both adenine groups do not form contacts with the DBD, but the complementary thymines do. Furthermore an essential ionic DBD-DNA contact is found at β-sheet 3 where the amine group of Lys80 forms a salt bridge to the phosphate backbone of the DNA adjacent to the previous HSE binding motif in the strand. In the C-terminal part of the domain, basic residues enhance the DBD-DNA affinity (Figure 27 d). Arg117 probably makes contacts in the major groove or towards the phosphate backbone, whereas Lys118 forms a salt bridge with the negatively charged DNA backbone.

Cys36 and Cys103 have also been implied to regulate Hsf1 function by disulfide bond formation (Ahn 2003). In our structures these Cysteine residues point into the solvent and thus are, in principle, accessible to covalent bond formation.

5.2.2.3.2 HsHsf1 exhibits DBD-DBD contacts

The two HsHsf1 DBDs bound to a two times nGAAn repeat motif in tail-to-tail orientation showed an intermolecular contact involving α-helix 2 of each subunit. However, in our crystal structure the closest distance between the helices (4.95

Å) is beyond van der Waals distance (4.6 Å), which points towards a small bend of the DNA under physiological conditions to realize better packing (Figure 28 a, b). Electrostatic interactions between Lys54 and Asp95 might be the driving force for the contact. Additional several polar, but uncharged, side chains are involved in stabilizing the interface - probably via van der Waals contacts. At the closest point of approach, Gly50 might contact its opposing Gly50 in a way that water molecules are excluded from the interface.

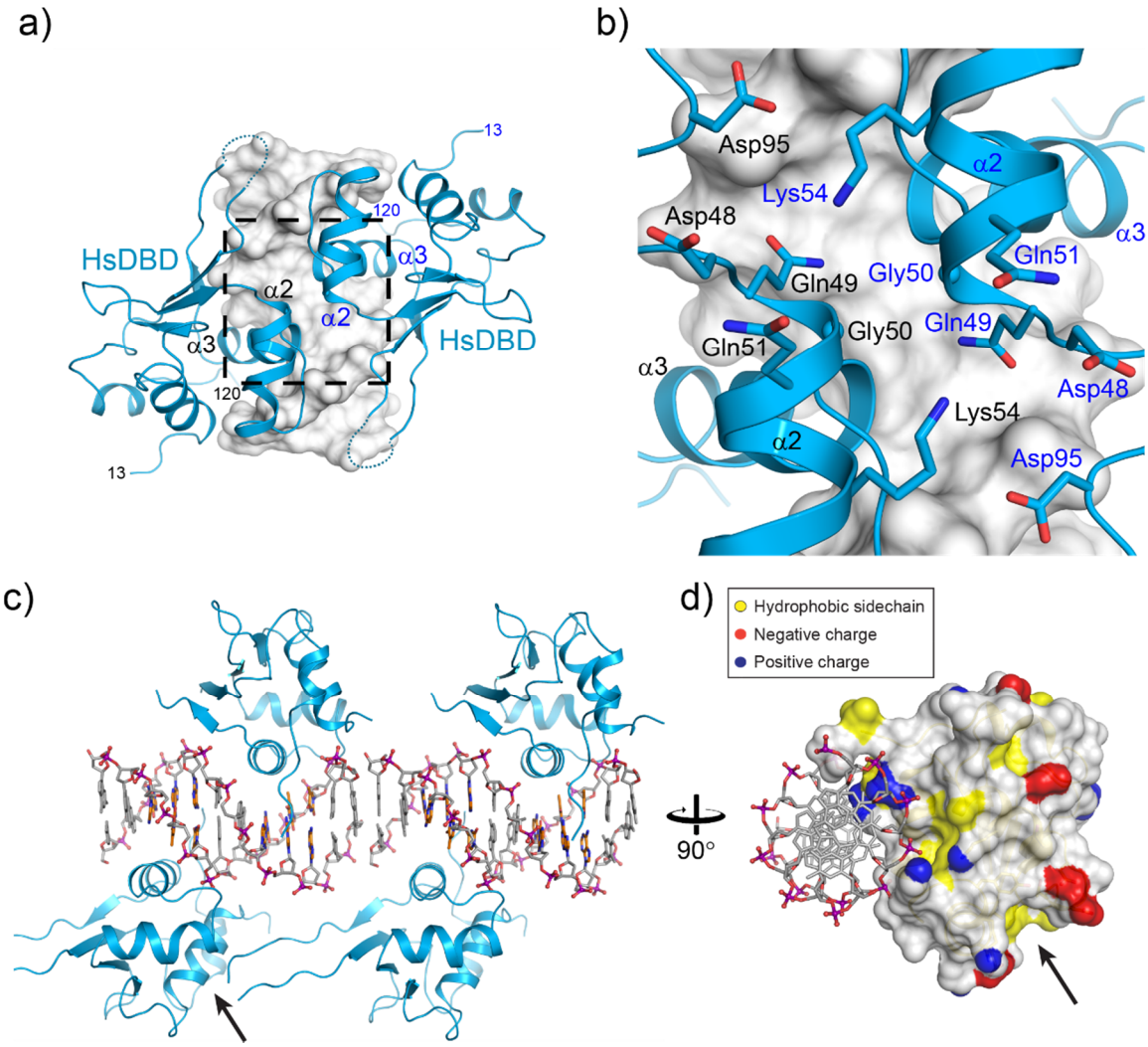

Figure 28 HsHsf1 DBD inter subunit contacts are mediated via two different interfaces. a) and b) Interface between two adjacent HsHsf1 DBDs bound to the tail-to-tail HSE motif. c) The wing in the HsHsf1-SatIII complex structure points towards a hydrophobic groove between helix 1 and 2 in the crystal lattice. Protein is shown in ribbon representation, DNA is shown in surface representation or stick representation with the phosphate backbone highlighted. d) Surface representation with indication of hydrophobic and charged side chains of HsHsf1 DBD reveals hydrophobic groove between helix 1 and 2.

In the crystal lattices of the DBD-HSE and DBD-SatIII complexes the DNA duplexes formed linear stacks stabilized by π-staking of Watson-Crick base pairs resembling a continuous DNA duplex (Figure 28 c). This arrangement potentially reveals insights into other DBD-DBD interactions along the DNA duplex. Contacts between DBDs which are bound to every second nGAAn motif on the DNA double helix have previously been proposed to be mediated by the wing-loop (Littlefield & Nelson 1999). In the crystal lattice of the DBD-HSE complex no contacts between DBDs in head-to-head orientation were observed.

As mentioned above, the wing loop tips were disordered. The least disorder in this region was observed in the DBD-SatIII complex. The complex structure of HsHsf1 DBD-SatIII therefore might provide a model for the orientation of the loop between two adjacent DBDs (Figure 28 c, d). However, the interface is separated by two additional base pairs and turned out of register in the crystal lattice. The wing loop (which is almost fully structured in this crystal structure) appears to reach out towards a hydrophobic groove located between α-helix 1 and α-helix 2 of the adjacent DBD. This shallow groove is defined by Leu34 and Val56 and framed by several charged and polar residues (Asp28, Asp30, Thr31, Gln51, and Lys59). However, Tyr60 seems to sterically block access to the hydrophobic groove (Figure 28 d).

5.2.2.3.3 Role of PTMs in DBD-DNA interactions

Previous studies have shown that Hsf1 undergoes extensive post-translational modifications (PTM) to fine-tune its regulation (Xu et al. 2012). These PTMs comprise phosphorylations, acetylations, and sumoylations. It was observed that specific lysines of HsHsf1 in the DBD are acetylated upon attenuation and that these acetylations contribute to the regulation of Hsf1's DNA-binding capabilities (Westerheide et al. 2009). A previous study conducted in our lab found that K80, K116 and K118 are acetylation sites, and that K80 and K118 are increasingly acetylated with enduring stress, leading to Hsf1 deactivation (Raychaudhuri et al. 2014). Nuclear stress body (nSB) formation is abolished in HeLa cells expressing acetylation mimicking mutants (K80Q, K118Q) of HsHsf1. With the help of our HsHsf1 DBD-DNA complex structures we are able to explain these findings (Figure 27). As discussed above, the ionic interactions of positively charged Lys80

and Lys118 with the negatively charged phosphate backbone of the DNA are predicted to be abolished by charge neutralization upon acetylation. We tested this hypothesis using HsHsf1 with the site-specific mutation K80Q, K116Q and K118Q, respectively. When the purified HsHsf1 mutant proteins were assayed by EMSA using a DNA oligonucleotide containing three inverted nGAAn repeats, K116Q and K118Q showed a slight decrease in DNA-binding affinity, whereas K80Q showed virtually no capacity for DNA-binding under non-heat shock conditions (Figure 29).

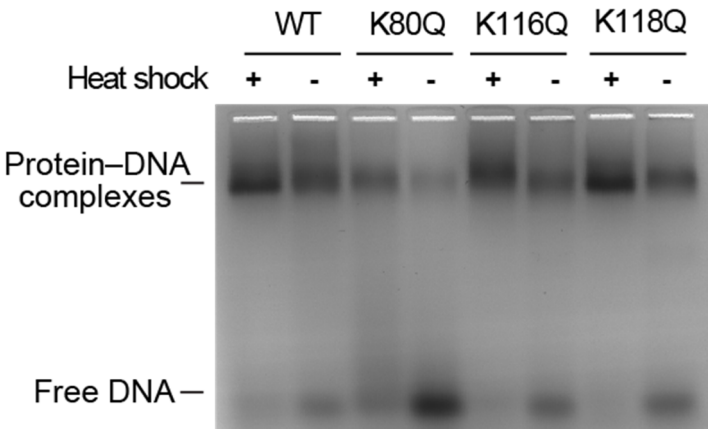

Figure 29 Acetylation of conserved lysines in the HsHsf1 DBD contributes to DNA-binding regulation. Mixtures of protein and fluorescent HSE-DNA were incubated with or without heat treatment and analyzed by agarose gel electrophoresis. The fluorescence signal in the gel is shown. Relative DNA-binding behavior of HsHsf1 WT and the acetylation mimicking mutants K80Q, K116Q and K118Q HsHsf1 were monitored by EMSA. Free DNA and DNA bound to protein complexes are indicated.

5.3 Characterization of the Hsf1 paralog Skn7 from *C. thermophilum*

Working with homologous proteins and protein complexes from thermophilic organisms is a popular strategy to elucidate the structure of molecular machines (Amlacher et al. 2011). These homologs often appear virtually “frozen” under the conditions of structure determination, whereas their dynamics is key to function under physiological conditions. The protein machinery of the thermophilic fungus *Chaetomium thermophilum* is adapted to a growth temperature range of 26-61 ºC, and thus expected to be less dynamic than mesophilic counterparts at room temperature (Amlacher et al. 2011; la Touche 1948). The genome of *C.*

thermophilum contains two sequence homologs of Hsf1: CTHT_0005280 (Uniprot ID: G0RY37) and CTHT_0048700 (Uniprot ID: G0SB31). By comparison with the respective *S. cerevisiae* sequences we concluded that CTHT_0005280 represents the ortholog of yeast Hsf1, whereas CTHT_0048700 represents the paralog Skn7 (Figure 12). However, *C. thermophilum* Skn7 (CtSkn7) shows a higher sequence homology to human Hsf1 in the conserved N-terminal region (DBD and HR-A/B) than the functional homolog. The predicted putative linker region between the DBD and the putative coiled-coil domain of CtSkn7 has approximately the same length as in HsHsf1 (linker between DBD and HR: CtSkn7: ~13 residues, HsHsf1: ~15 residues, CtHsf1: ~45 residues) (see Appendix, Figure 64). Therefore we focused on CtSkn7 to elucidate structural features of Hsf proteins which were not accessible in HsHsf1.

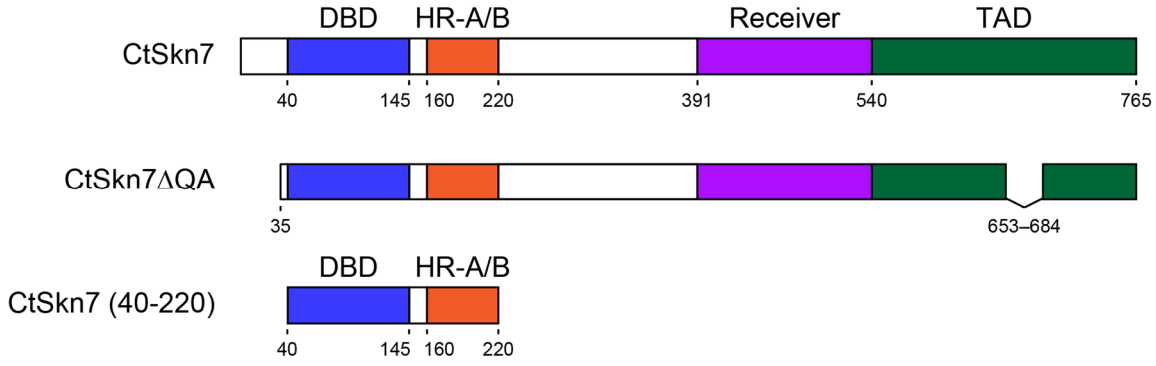

Figure 30 Multidomain protein constructs of CtSkn7 for functional and structural analysis. Protein domain graphs for multidomain constructs which were analyzed.

5.3.1 Purification of CtSkn7 for *in vitro* analysis

We began our analysis with a synthetic, nearly full-length construct of CtSkn7 (CtSkn7ΔQA; residues 35-765, excluding residues 1-34 and 653-684 which included low complexity glutamine and alanine stretches, respectively) and truncation constructs containing only the DBD and HR as judged by computational secondary structure prediction and multiple sequence alignment (CtSkn7(40-209) and CtSkn7(40-220)) (Figure 30).

All CtSkn7 constructs were cloned into the bacterial expression vector pHUE for expression as N-terminal His_6-ubiquitin fusion proteins in *E. coli* (Catanzariti et al.

2004). Expression as ubiquitin-fusion proteins in *E. coli* proved to enhance solubility in previous projects (Neudegger 2012; Hauser et al. 2015).

For CtSkn7 constructs containing the DBD, the purification was carried out in the same way as for HsHsf1(1-120) (paragraph 5.2.2.1). IMAC, followed by expression tag cleavage and cation exchange chromatography was used. As a further purification step, size exclusion chromatography was performed using a Superdex 200 column (Figure 31). The final yield of the purifications showed great variation between the constructs. Per L culture up to 10 mg for CtSkn7(40-220) and CtSkn7ΔQA, and up to 60 mg for CtSkn7(40-209) were obtained. The applied procedures resulted in apparent homogeneous products as visualized on Coomassie-stained SDS polyacrylamide electrophoresis gels. Proteins were concentrated and stored at -80 ºC.

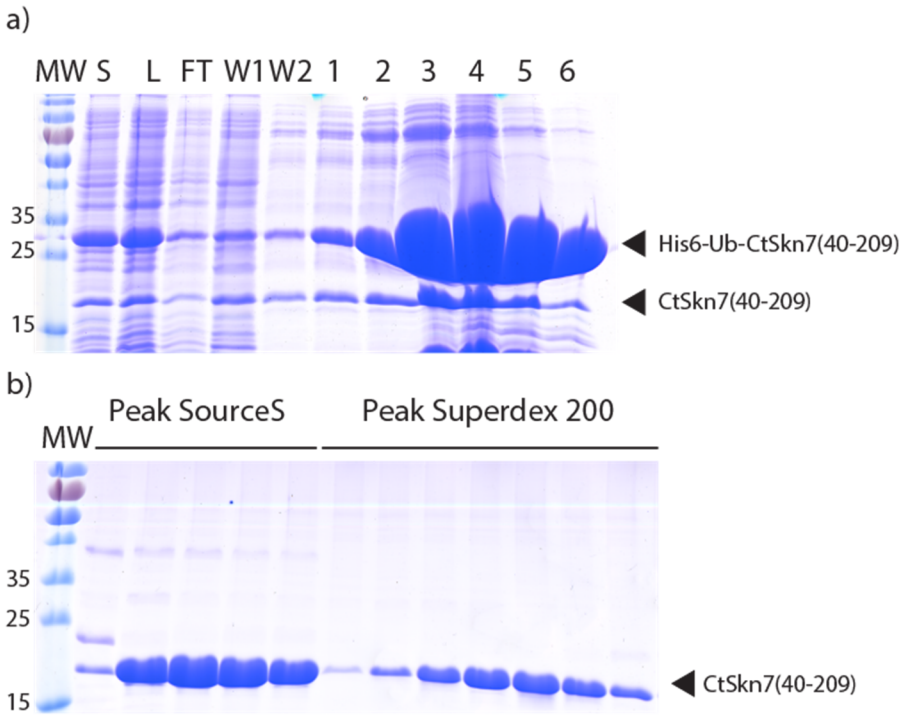

Figure 31 SDS-PAGE analysis of a representative CtSkn7(40-209) purification. a) SDS-PAGE analysis of the total lysate (L), the soluble fraction (S), and fractions from IMAC. FT, Flow through; W1, W2, wash fractions and elution fractions 1-6. MW, Molecular weight marker in kDa. b) SDS-PAGE analysis of the protein content in the major peaks of the ion exchange step on Source S column and on size exclusion Superdex 200 column.

5.3.2 CtSkn7 binds HSE and SSRE DNA sequence motifs

The Skn7 homolog of *S. cerevisiae* has been reported to interact both with HSEs (inverted repeats of the nGAAn motifs) as well as SSREs of the consensus sequence nGGCnnGSCn (S designates G or C) (Badis et al. 2008). In order to functionally characterize CtSkn7, we went on to study the DNA sequence motif preference of our constructs by EMSA analysis (Figure 32). Both, CtSkn7ΔQA and also the C-terminal truncated construct CtSkn7(40-220) formed complexes with HSE as well as SSRE motifs, whereas the control DNA oligonucleotide showed no retention by either protein.

Thus, our CtSkn7 constructs showed the expected DNA sequence preferences. The minimal constructs (CtSkn7(40-209) and CtSkn7(40-220) exhibited no difference in their DNA-binding properties when compared to CtSkn7ΔQA (Figure 32).

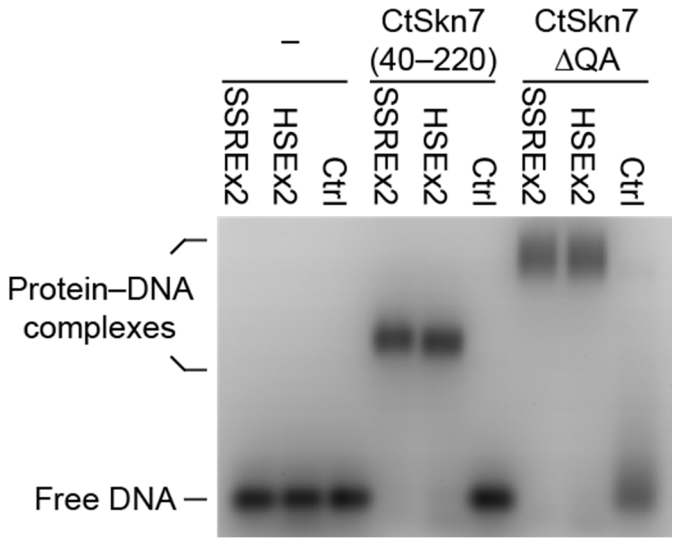

Figure 32 CtSkn7ΔQA and CtSkn7(40-220) bind to HSE and SSRE DNA sequence motifs in EMSA. Purified protein constructs were incubated with fluorescently tagged DNA-oligonucleotides. Complexes and free DNA were then separated from each other using agarose gel electrophoresis. The fluorescence signal in the gel is shown and free DNA and DNA bound by protein complexes are indicated. A representative gel of 3 independent experiments is shown. Ctrl, control DNA.

Next we strived to get information on whether CtSkn7 had any preference for either of the two sequence motifs. In order to determine the dissociation constants (K_D), we titrated protein against fluorescently labelled DNA and monitored the fluorescent band pattern by EMSA (Figure 33). DNA oligonucleotides containing

four nGAAn or four nGGCn continuous repeats showed a K_D of 3.27±0.93 µM and 1.23±0.22 µM, respectively. The discontinuous SSRE found in the promoter of OCH1, which contains two double nGGCn repeats separated by an 11 bp spacer, showed a K_D of 3.24±0.96 µM. Control DNA showed a dissociation constant of 20.47±6.42 µM. This is surprising since most transcription factors show affinities to their cognate recognition motif in the nano and pico molar range (Strauch 2001). Therefore the measured values indicate, that CtSkn7 possibly requires another interaction partner at the DNA for joint DNA-binding to increase its affinity (see paragraph 2.5.4). Nevertheless, CtSkn7 seems to slightly prefer SSREs over HSEs in our assay.

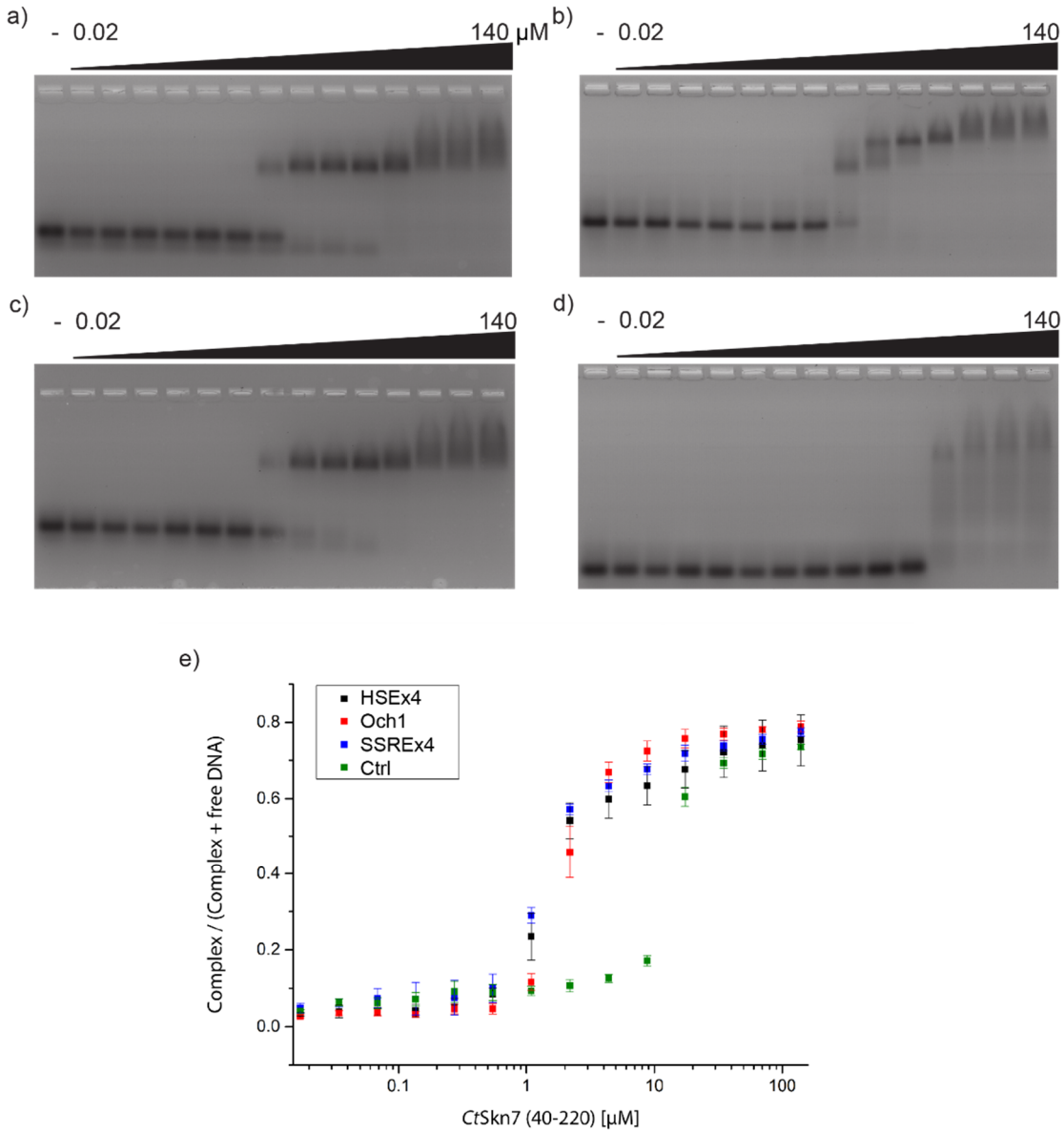

Figure 33 Quantitative analysis of DNA-binding affinities of CtSkn7 in EMSA reveal high dissociation constant of CtSkn7 to HSE and SSRE motifs. CtSkn7(40-220) titration series against a) HSEx4, b) OCH1, c) SSREx4 and d) Ctrl DNA-oligonucleotide with fluorescein label. The fluorescence signal in the gel is shown. The band intensity of the protein-DNA complex and the free DNA was quantified (e). Error bars represent s.d. values from three independent experiments.

5.3.3 Flexible connection between DBD and HR in CtSkn7

Flexible joint regions in proteins are thought to impair three dimensional ordering of proteins in crystal lattices. Such flexible regions are typically more accessible to proteolytic degradation compared to stably structured parts (Koth et al. 2003). Therefore, we subjected CtSkn7(40-220) in presence and absence of HSE DNA

to limited Proteinase K digest to identify unstructured regions. Separation of the digestion reaction on SDS-PAGE revealed three major fragment bands (Figure 34). Mass spec identified these bands as residues 40-132, 40-150 and 40-209. According to these values the predicted linker (aa 144-160) as well as the HR C-terminus (aa 209-220) seemed to be highly protease-sensitive, indicating that the DBD and HR domain are connected via a highly flexible linker. This finding is corroborated by secondary structure prediction using JPred, and earlier work on Hsf proteins (Drozdetskiy et al. 2015; Flick et al. 1994).

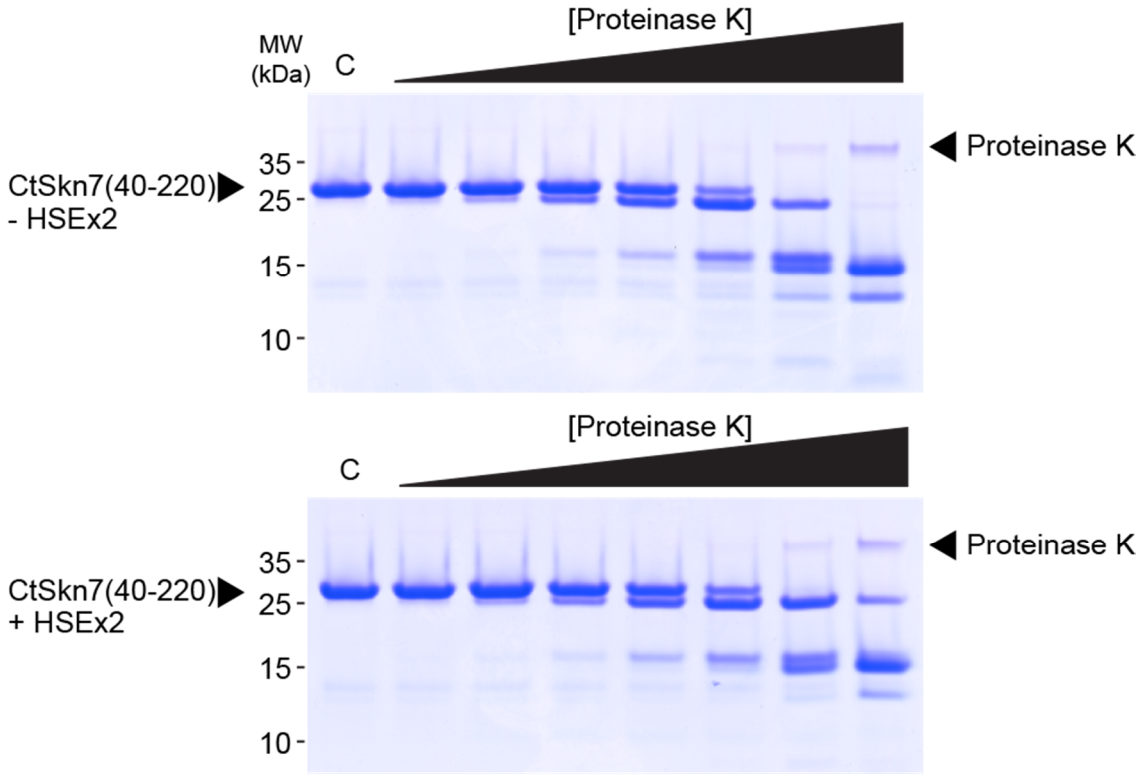

Figure 34 Probing CtSkn7(40-220) for unstructured parts by limited Proteinase K proteolysis in the presence (top) or absence (bottom) of cognate DNA. CtSkn7(40-220) was incubated in absence or presence of HSE DNA with Proteinase K concentration series (0 – 50 µg/ml Proteinase K). The resulting digestion products were analyzed on Coomassie-stained SDS-PAGE gels. MW, Molecular weight marker.

5.3.3.1 SAXS analysis of CtSkn7

Information about the size, domain organization, and shape of macromolecules and macromolecular complexes can be obtained by small-angle X-ray scattering

(SAXS), which gives low-resolution information about their shape in solution (Figure 35). Radial scattering intensity distributions were recorded for a concentration series of CtSkn7(40-209).

Linear fit of the Guinier region of the scattering curve suggested a radius of gyration (Rg) for CtSkn7(40-209) of ~4 nm (Figure 35 a). Fitting the X-ray scattering curve with the indirect Fourier transformation program GNOM provided us with a density distribution function with a maximum diameter of ~14 nm. An approximate molecular weight between a dimeric and trimeric species (39.88 - 53.18 kDa (theoretical MW monomer: 20.0 kDa)) was estimated from the decay of scattering intensity at high angles with the program POROD, suggesting a dimer or trimer. The density distribution graph suggested a dumbbell-shape for the particles, suggestive of two domains (Figure 35 c). Consistently, the Kratky plot suggests a flexible linkage, as indicated by the steady increase of q^2 with q (Figure 35 b).

Therefore we concluded the presence of two stably folded domains (DBD and HR) which are connected by a flexible linker, consistent with the limited proteolysis experiment.

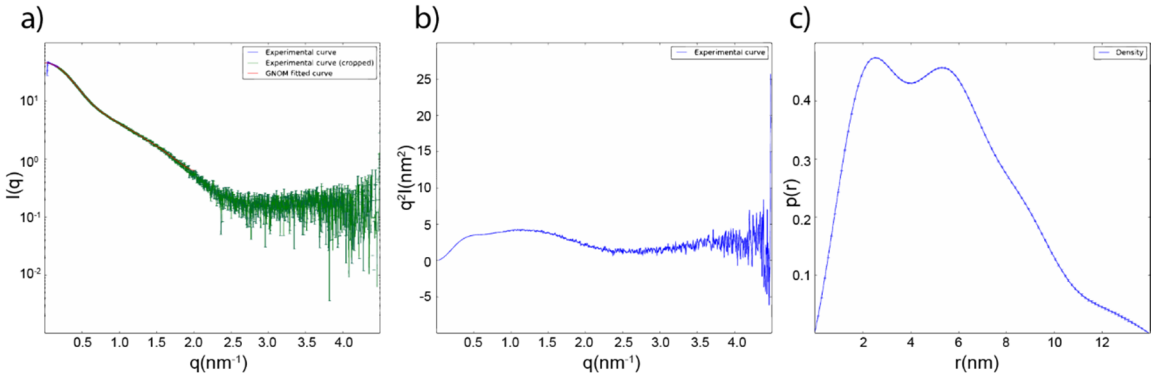

Figure 35 Small angle X-ray scattering analysis of CtSkn7(40-209). Exemplary scattering curve. a) X-ray scattering curve shows absence of aggregation at low scattering angles, b) Kratky plot, c) Density distribution graph of GNOM suggests two domain architecture and reveals maximum extension of 14 nm.

5.3.4 Analysis of isolated CtSkn7 domains

Thus we decided to create single-domain constructs, the DBD (CtSkn7(40-143)) and the HR (CtSkn7(160-209) and CtSkn7(160-220)). The proteins were soluble after expression in *E. coli.* The purification of the DBD followed the generic

protocol for HsHsf1(1-120), and was slightly modified for the purification of the HR constructs (paragraph 4.6.1.3). Interestingly, a coiled-coil construct similar to CtSkn7(160-220) including residues 240-310 from CtHsf1 was insoluble when expressed in *E. coli* (data not shown).

The applied procedures resulted in protein of high purity and apparent homogeneity as visualized on Coomassie-stained SDS-PAGE gels (Figure 36). Proteins were concentrated and stored at -80 ºC.

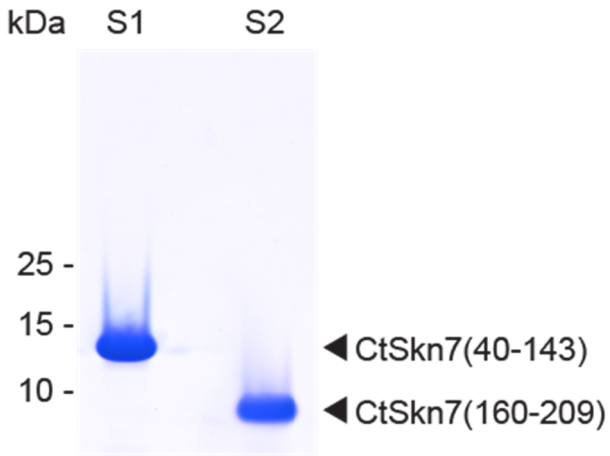

Figure 36 Coomassie-stained SDS-PAGE gel with concentrated samples of purifications of CtSkn7 domains (DBD (40-143) and HR (160-209)). Molecular weight marker (MW), Sample 1 (CtSkn7(40-143)) and 2 (CtSkn7(160-209)) (S1 and S2).

5.3.5 CtSkn7 forms oligomers via coiled-coil domain

Subsequently, our initial insights into the oligomeric nature of CtSkn7 using SAXS (paragraph 5.3.3.1) were complemented by multiple experimental approaches.

First, we used Ethylene glycol bis[succinimidylsuccinate] (EGS) crosslinking followed by SDS-PAGE analysis. EGS covalently crosslinks primary amines, resulting in complexes which run at a higher molecular weight in denaturing polyacrylamide gels. SDS-PAGE analysis showed, dependent on the amount of added EGS, a ladder of bands, indicative of crosslinked dimers, trimers and few tetramers of CtSkn7(40-220) (Figure 37 a). A very similar banding pattern was found for the HR domain construct of CtSkn7(160-220) (Figure 37 b).

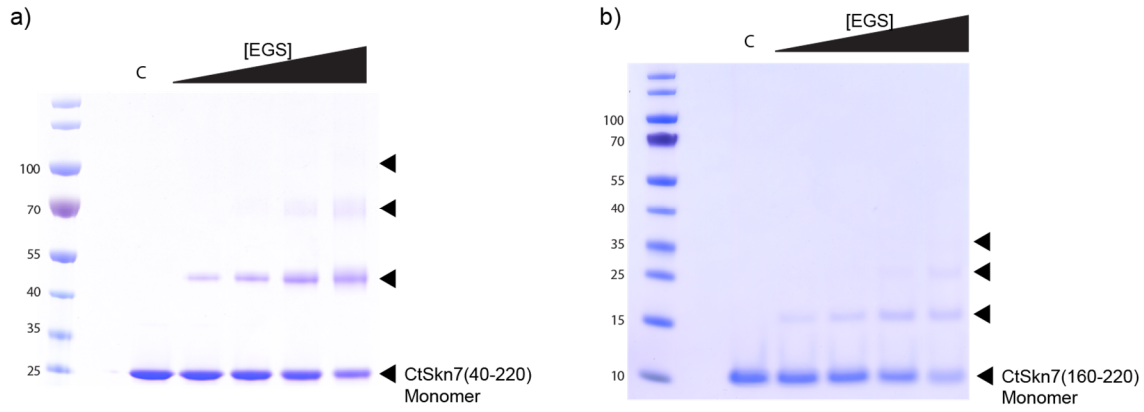

Figure 37 EGS crosslinking of CtSkn7(40-220) and CtSkn7(160-220). SDS-PAGE analysis of EGS crosslinking experiments of CtSkn7(40-220) (a) and CtSkn7(160-220) (b). Purified protein was incubated with increasing amounts of EGS. After quenching the crosslinking reaction, samples were analyzed by SDS-PAGE and Coomassie staining. Molecular weight marker in kDa is indicated on the left. Crosslinked oligomers are indicated by arrow heads.

To get independent information on the oligomeric state of the CtSkn7 truncation constructs, we applied native mass spectrometry (native MS). The spectra for construct CtSkn7(160-220) showed m/z ratios consistent with monomers and dimers (Figure 38 a). For CtSkn7(40-220), native MS indicated the presence of monomers, dimers, trimers and tetramers in solution, whereby monomers and lower oligomeric states seemed to be more abundant (Figure 38 b). The monomeric peaks in both analyses might derive from dissociation of oligomers in the gas phase. Nevertheless, native MS strictly depends on the solubility and ionization properties of the measured protein in an ammonium acetate buffer and thus is not quantitative. Therefore, some in-solution complex states might not have been detected by this method. Native MS analyses were performed by Dr. Javaid Y. Bhat.

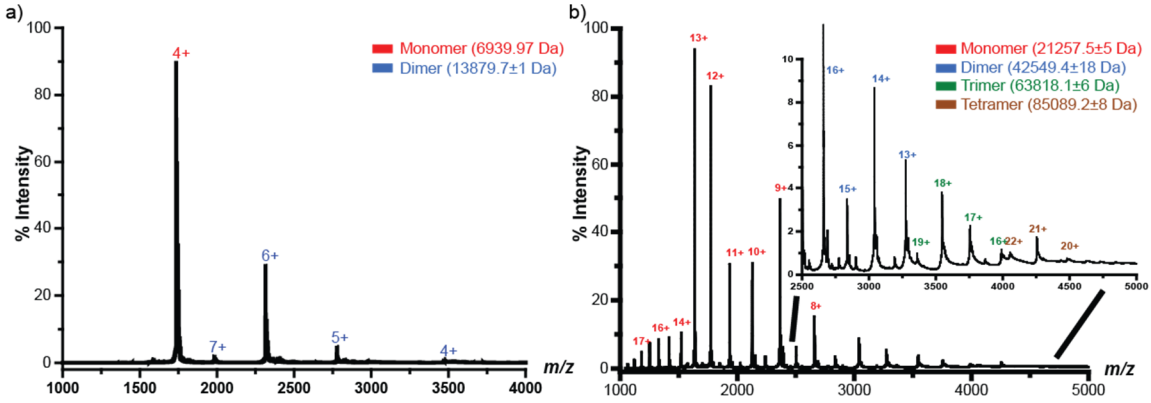

Figure 38 Native mass spectrometry analysis of CtSkn7(160-220) (a) and CtSkn7(40-220) (b). Native mass spectra are shown. Insert in b) shows magnification of the high m/z area. Different oligomeric species are indicated by different color labels. Charge states are shown for selected peaks. The calculated mass around the m/z values of the respective protein complexes and the accuracy of mass values calculated from the different m/z peaks are indicated. MS analysis was performed by Dr. Javaid Y. Bhat.

Size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) offers an alternative approach to determine the molecular mass of particles in solution.

CtSkn7(40-220) formed a symmetrical narrow peak in the SEC chromatogram indicating a homogenous behavior (Figure 39 a). CtSkn7(40-220) (theoretical mass = 21.3 kDa) showed a molecular mass of ~63 kDa, which corresponds to a trimer. The coiled-coil domain CtSkn7(160-209) showed a molecular weight of 14.5 kDa indicating a mixture of probably dimers and trimers in solution (CtSkn7(160-209), theoretical subunit mass = 5.7 kDa) (Figure 39 a). As expected, the Hsf-type DBD in CtSkn7(40-143) (theoretical mass = 12.6 kDa), showed monomeric behavior (~13.7 kDa) and thus probably does not contribute to oligomerization (Figure 39 b). SEC-MALS experiments were performed by Dr. Manajit Hayer-Hartl.

Together, the results from SAXS (paragraph 5.3.3.1), EGS crosslinking, native MS and SEC-MALS suggested that CtSkn7 adopts an oligomeric state defined by an equilibrium between a dimeric, trimeric and tetrameric species.

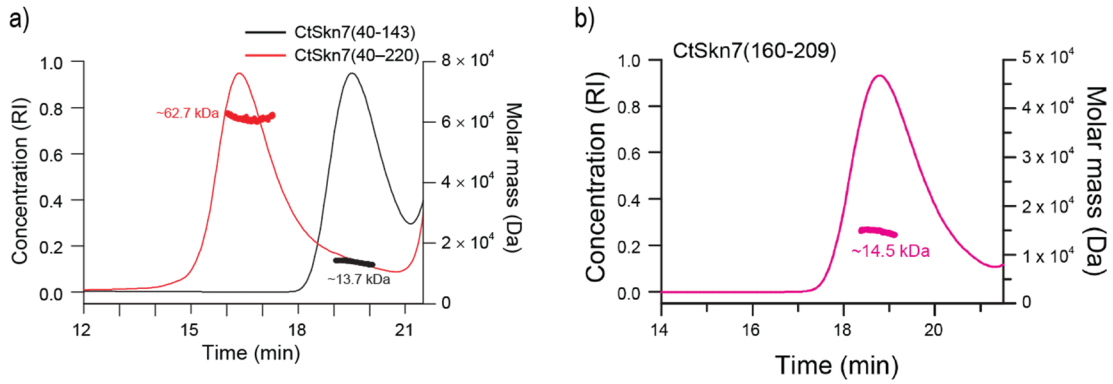

Figure 39 Multi-angle light scattering analysis of CtSkn7 truncation constructs reveals oligomerization via the coiled-coil domain. Horizontal lines across the peaks indicate molar mass and homogeneity. Calculated molar masses are indicated. a) SEC-MALS analysis of CtSkn7(40-220) and CtSkn7(40-143). CtSkn7(40-143) shows the apparent MW of a monomer. CtSkn7(40-220) reveals MW corresponding to a trimer. b) Coiled-coil construct CtSkn7(160-209) reveals MW corresponding to a dimer - trimer mixture. RI, refractive index. Experiments were performed by Dr. Manajit Hayer-Hartl.

5.3.6 Thermal stability of CtSkn7 constructs

To assess the secondary structure content of CtSkn7 we performed CD spectroscopy. CD spectroscopy provides information about the secondary structure and folding properties of proteins by differential absorption of circularly polarized light (Greenfield 2006). The spectra of all three constructs (CtSkn7(40-220), CtSkn7(160-220) and CtSkn7(160-209)) showed strong minima at 208 nm and 222 nm and a maximum at 195 nm, indicative of a high amount of α-helical secondary structure content. The spectrum for CtSkn7(40-220) deviated only slightly from those of the HR-only constructs, indicating a minor influence of other secondary structure elements.

The high α-helical secondary structure content and the absence of other secondary structure elements shown by the Skn7 HR domain is consistent with the formation of a putative coiled-coil structure.

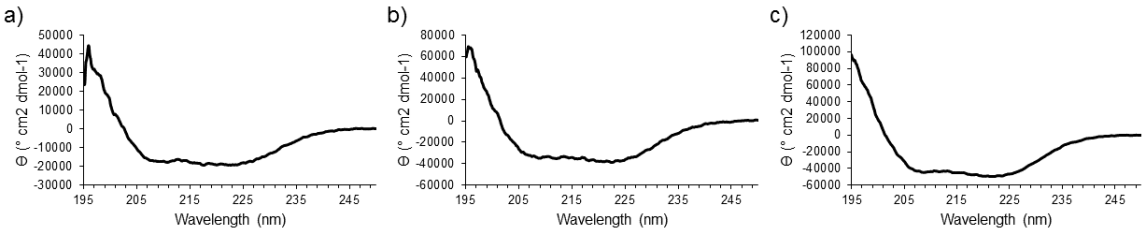

Figure 40 CD spectroscopic analysis of CtSkn7 truncation constructs. CD spectra of CtSkn7 constructs a) CtSkn7(40-220), b) CtSkn7(160-220) and c) CtSkn7(160-209) are shown. Molar ellipticity versus wavelength is shown. A strong maximum at 195 nm and two strong minima at 208 and 222 nm are indicative of α-helical structures. A minimum at 218 nm and maximum at 195 nm suggest ß-sheet structures.

In order to probe the thermal stability of the Skn7 constructs we used CD spectroscopy at a wavelength of 222 nm, indicative of α-helix content. The melting curves of the constructs CtSkn7(40-209), CtSkn7(160-220) and CtSkn7(160-209) were sigmoidal, indicative of a cooperative unfolding mechanism (Figure 41). We found melting points of 47.0±0.04 ºC for CtSkn7(40-220), 63.4±0.03 ºC for CtSkn7(160-220) and 61.9±0.06 ºC for CtSkn7(160-209). The melting point for the coiled-coil domains is not considerably higher than the melting point for the GCN4 zipper (61 ºC, (Ciani et al. 2010)) from the mesophilic organism *S. cerevisiae*, but still consistent with *C. thermophilum's* adaptation to optimal growth temperature of 45 ºC (la Touche 1948). Therefore, our findings suggest that CtSkn7 is constitutively oligomeric at this temperature. Whereas unfolding of the coiled-coil constructs was reversible, the CD signal of the native state was not recovered upon slow cooling with the construct involving DBD, Linker and coiled-coil.

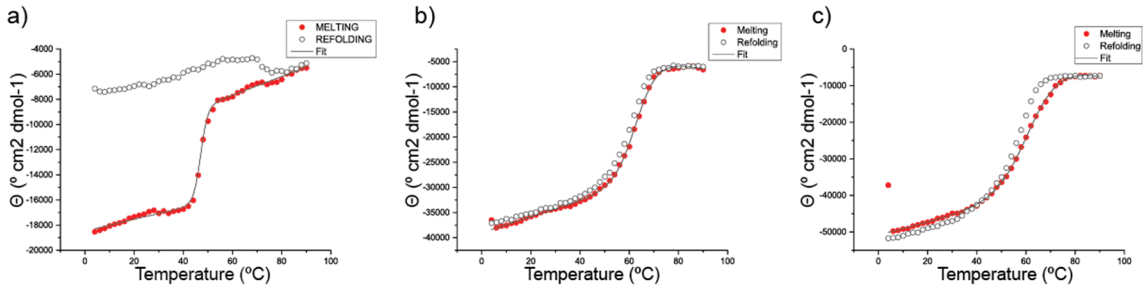

Figure 41 Thermal unfolding of of CtSkn7 protein constructs as monitored by CD spectroscopy. Thermal melting curves of a) CtSkn7(40-220), b) CtSkn7(160-220), c) CtSkn7(160-209) are shown. The molar ellipticity at 222 nm wavelength during slow heating is shown. The red dots represent the signal during heating, the white circles during cooling. The line represent the fit of the melting curve for melting point determination using Jasco Software.

5.3.7 Crystallization of *Chaetomium thermophilum* Skn7

Having highly pure, well-characterized CtSkn7 protein constructs in hand we continued with crystallization trials. The experiments were set up in the MPI Crystallization Facility. First, we studied the crystallization of a construct of CtSkn7 which included the DBD and the HR connected by its linker, CtSkn7(40-209), in the absence and presence of HSE DNA-oligonucleotides. Crystals in the shape of ~100 µm long needles grew in presence of HSE DNA. Representative pictures of crystals are shown in Figure 42. Attempts to grow larger crystals failed. The crystals however exhibited poor X-ray diffraction to a resolution of ~10 Å similar to HsHsf1-GCN4.

Following this, we decided to continue with the separate structural analysis of the DBD (CtSkn7 residues 40-143) and the HR domain constructs (CtSkn7 residues 160-209, and 160-220).

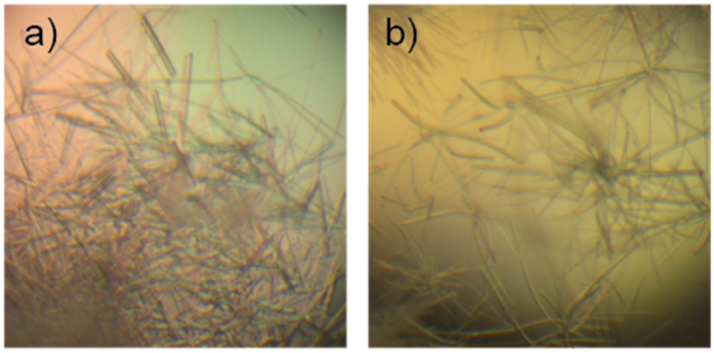

Figure 42 Crystallization of CtSkn7(40-209) in complex with HSE DNA-oligonucleotide. Representative crystals of constructs of CtSkn7 including the DBD, the linker and the coiled-coil domain resulted in needle shaped crystals.

We anticipated that complexation of the DBD of CtSkn7 with cognate DNA would rigidify the protein. For cognate DNA, we tested the sequences of HSE (HSEx2 (ggTTCtaGAAcc)) and SSRE (SSRE (ATTTGGCTGGGCC)). Initial crystals of the CtSkn7 DBD in complex with HSE (ggTTCtaGAAcc) were obtained using the commercial crystallization screen PEGS suite (Qiagen) and a robotic setup producing 200 nL droplets (Figure 43 a, b). The precipitant composition was subsequently refined to 0.18 M tri-ammonium citrate and 11 % PEG-3350.

The initial conditions for the CtSkn7 DBD-SSRE (ATTTGGCTGGGCC, (Badis et al. 2008)) complex crystals were also obtained using the PEGS suite and were

refined to 0.1 M HEPES-NaOH pH 7.5 and 25 % PEG-2000 MME as precipitant (Figure 43 c).

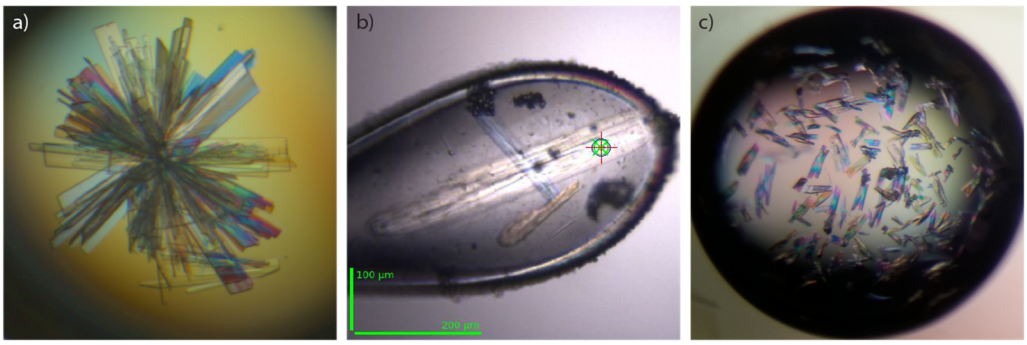

Figure 43 Crystallization of CtSkn7 DBD with HSE and SSRE containing DNA-oligonucleotides. a) and b) showing representative crystals of CtSkn7 DBD in complex with HSE DNA, c) shows representative crystals of CtSkn7 DBD-SSRE complex. The crystals grew in large clusters. Each rectangle represents an individual plate-shaped crystal. b) shows a cryo-cooled crystal mounted for the X-ray diffraction experiment at the ESRF.

To determine CtSkn7's coiled-coil oligomerization domain, two slightly different constructs comprising the amino acid range 160-209 and 160-220 (Figure 44) were used. The construct borders were assigned in accordance with sequence conservation to human HR-A/B, secondary structure predictions using JPRED, and limited proteolysis experiments (paragraph 5.3.3). Initial crystallization conditions for CtSkn7(160-209) were obtained using the pH Clear I Suite (Qiagen). Subsequent analysis showed the presence of two distinct crystal forms. Both crystal forms grew with 0.1 M HEPES-NaOH pH 7.0 and PEG-6000 (20 or 21 %) as precipitant (Figure 44). Crystallization of the construct CtSkn7(160-220) was initially achieved using the Index crystallization screen (Hampton Research). The final condition was refined to 0.1 M Bis-Tris pH 5.5, 0.2 M $MgCl_2$ and 25 % PEG-3350 as precipitant.

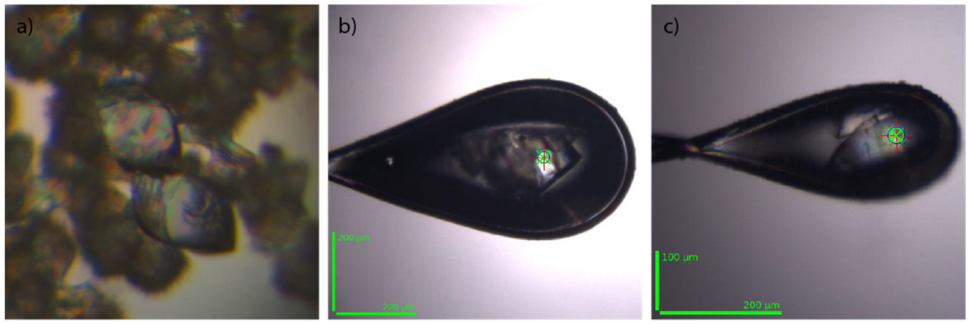

Figure 44 Crystallization of CtSkn7(160-209). a) and b) show the crystals which led to the structure of CtSkn7(160-209) crystal form II. c) shows crystals of CtSkn7(160-209) of crystal form I.

5.3.7.1 Structure solution and refinement of CtSkn7

The structures of the CtSkn7 DBD-DNA complexes were solved by molecular replacement with the existing *K. lactis* Hsf-DBD-HSE complex structure as search model. In contrast, we used single-wavelength anomalous diffraction from the sulfur atoms (S-SAD) in Cysteine and Methionine residues for the phasing of the initial CtSkn7(160-209) structure.

Table 4 provides an overview of the data collection and model refinement statistics of the CtSkn7 domain construct crystal forms.

As an initial search model for the CtSkn7(40-143)-HSE complex we used the *Kluyveromyces lactis* Hsf1-HSE crystal structure (PDB code 3HTS (Littlefield & Nelson 1999)) and solved the structure with the program Molrep. After refinement of the CtSkn7 DBD-HSE complex structure, we used this model as search template for our other DBD-DNA complexes using Molrep. Whereas the asymmetric unit in the DBD-HSE structure contained one DBD chain and one strand of the DNA molecule, the CtSkn7 DBD-SSRE contained the complete biological assembly featuring two DBD molecules bound to one DNA duplex.

The first crystal form of CtSkn7(160-209) (crystal form I) was solved by sulfur-single-wavelength anomalous diffraction (S-SAD), using a wavelength of ~1.7 Å to maximize the anomalous signal. The positions of the sulfur atoms were identified by direct methods using SHELXD, and after several cycles of density modification SHELXE was able to build an initial molecular model. Subsequently, an almost complete model was auto-built by ARP/wARP using native data up to a resolution of 1.03 Å. The asymmetric unit contained two chains of a crystallographic tetramer. This asymmetric unit was used as search model in Molrep for solving CtSkn7(160-209) crystal form II and the CtSkn7(160-220) crystal structure. CtSkn7(160-209) crystal form II contained a half-tetramer and a trimer. The trimer found in crystal form II was manually built using Coot. The asymmetric unit of CtSkn7(160-220) crystals contained a full tetramer. All models were refined by iterative cycles of model building using Coot and refinement using Refmac 5.

Table 4 Data collection and refinement statistics for CtSkn7 domain structures

	CtDBD–HSE	CtDBD–SSRE	S-SAD	CtSkn7(160–209)-I	CtSkn7(160–209)-II	CtSkn7(160–220)
Data collection						
Space group	$C222_1$	$C2$	$C222_1$	$C222_1$	$P2_12_12$	$P2_1$
Cell dimensions						
a, b, c (Å)	39.77, 78.40, 90.08	73.73, 43.33, 88.78	24.81, 151.47, 47.74	24.82, 151.54, 47.75	47.00, 61.03, 75.79	50.73, 40.20, 52.78
α, β, γ (°)	90, 90, 90	90, 92.78, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 118.6, 90
			Remote			
Wavelength (Å)			1.70074			
Resolution (Å)	45.04 – 2.35 (2.48 – 2.35)	44.34 – 2.39 (2.48 – 2.39)	47.74 – 1.95 (2.0 – 1.95)	40.39 – 1.03 (1.05 – 1.03)	47.53 – 1.7 (1.79 – 1.7)	44.55 – 1.9 (1.94 – 1.9)
R_{merge}	0.045 (0.958)	0.092 (0.793)	0.047 (0.769)	0.066 (1.103)	0.104 (0.808)	0.069 (0.599)
$I/\sigma I$	17.4 (1.8)	10.5 (1.7)	44.0 (2.9)	17.1 (1.5)	10.5 (2.2)	7.2 (1.7)
Completeness (%)	98.9 (99.2)	99.3 (94.0)	99.4 (94.5)	98.7 (74.1)	99.7 (98.3)	97.8 (96.6)
Redundancy	4.7 (4.8)	5.1 (4.5)	44.8 (22.7)	11.4 (5.4)	8.8 (9.1)	2.8 (2.8)
Refinement						
Resolution (Å)	30 – 2.35	30 – 2.4	–	30 – 1.03	30 – 1.7	30 – 1.9
No. reflections	5791	10639	–	42643	23400	13834
R_{work} / R_{free}	0.215 / 0.250	0.228 / 0.289	–	0.165 / 0.199	0.209 / 0.285	0.241 / 0.295
Number of atoms						
Protein	816	1671	–	849	1967	1850
DNA/Mg^{2+}	243	527	–	–	–	–
Water	22	39	–	127	121	77
B-factors						
Protein	76.73	51.55	–	21.18	34.89	46.95
DNA/Mg^{2+}	56.12	51.82	–	–	–	–
Water	61.42	41.28	–	32.49	38.00	44.47
R.m.s. deviations						
Bond length (Å)	0.006	0.006	–	0.022	0.020	0.019
Bond angles (°)	1.044	1.101	–	1.853	1.981	1.757

[a] Values in parenthesis for highest-resolution shell

5.3.8 Structural analysis of CtSkn7

5.3.8.1 Analysis of the CtSkn7 DBD-HSE complex structure

5.3.8.1.1 DBD-DNA and DBD-DBD contacts of CtSkn7 DBD

The crystal structure of the DBD of CtSkn7 in complex with the HSE DNA revealed a topology very similar to the corresponding HsHsf1 DBD complex (r.m.s.d. 0.798 HsHsf1 DBD-HSE versus CtSkn7 DBD-HSE) (Figure 45 a). The C-terminal part of CtSkn7, residues 125-143, is structured as in HsHsf1, demonstrating that this conserved region constitutes an integral part of the Hsf family DBD. The recognition helix α3 is inserted into the major groove of the DNA double helix and held in position by a mostly invariant array of polar and charged residues (Ser93, Gln97, Asn99, Tyr101; Lys87 and His88, His104 and K105) (Figure 45 c). Arg96 corresponds to the key residue Arg71 in HsHsf1, forming identical contacts to the guanine group in the nGAAn motif. C-terminally located positively charged residues (Arg142 and Lys143) contact the DNA backbone, as seen for HsHsf1 DBD (Figure 45 d).

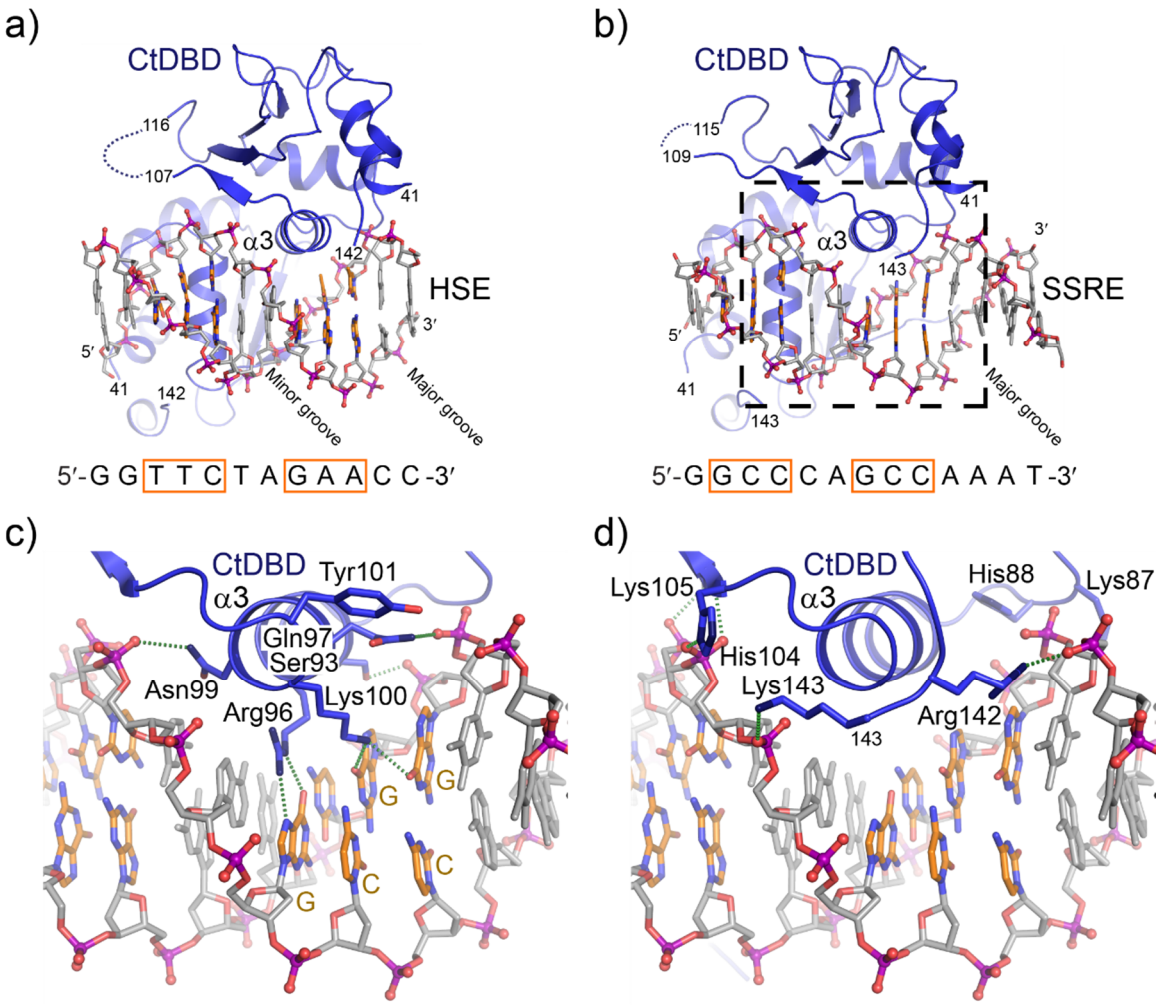

Figure 45 CtSkn7 DBD-DNA complexes reveal structural basis of HSE/SSRE recognition. a) Crystal structure of CtSkn7 DBD (CtDBD)-HSE complex. Crystal structure of CtSkn7 DBD-HSE complex reveals a tail-to-tail orientation of bound DBDs. b) Crystal structure of CtSkn7 DBD-SSRE complex. CtSkn7 DBD-SSRE head-to-tail complex shows a tail-to-tail orientation of bound DBDs. c), d) Detailed views on CtSkn7 DBD SSRE DNA interactions. c) and d) reveal contacts of the recognition helix α3 and the DBD C-terminus which are similar to the ones found for HsHsf1. DNA double helix is shown in stick representation with the phosphate backbone, and the GAA, GGC and GCC sequence motifs highlighted. K100 plays important role in not distinguishing SSRE from HSE. The C-terminal tail of the DBD is omitted for clarity in c).

5.3.8.1.2 Functional role of DBD C-terminus is conserved

Our crystal structures of the DBD of HsHsf1 and CtSkn7 in complex with DNA revealed that positively charged amino acids at the DBD C-terminus interact with the DNA phosphate backbone. In HsHsf1, Lys118 is critical for DNA-binding (Raychaudhuri et al. 2014). To test if this dependence is conserved, Lys143, which corresponds to Lys118 in HsHsf1 was mutated in CtSkn7 to Glutamine for charge

removal. Substitution of the positive charged Lys143 of CtSkn7 with neutral Glutamine resulted in reduced DNA-binding affinity (Figure 46).

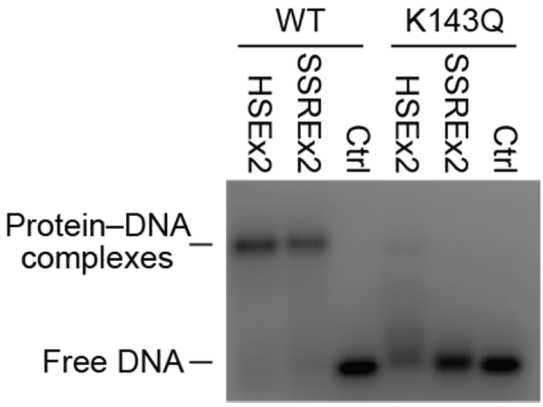

Figure 46 Relative DNA-binding of CtSkn7 WT and charge inverting mutant K143Q (corresponds to K118 in HsHsf1). Shown is a fluorogram of an EMSA assay. Purified CtSkn7(40-220) protein was incubated with fluorescently tagged DNA-oligonucleotides. Complexes and free DNA were then separated on an agarose gel. Free DNA and DNA bound by protein complexes are indicated. Ctrl, control DNA.

5.3.8.2 Analysis of the CtSkn7 DBD-SSRE complex structure

Surprisingly, the complex of the CtSkn7 DBD with the asymmetric SSRE motif exhibited the same topology as that of the CtSkn7 DBD-HSE structure (Figure 45 a, b). We had expected a linear arrangement of the CtSkn7 DBDs, similar to the HsHsf1 DBD-SatIII complex (Figure 27 b). The forward contacts to the nGCCn motif are realized by Lys100 contacting the guanine bases in the complimentary strand (nGGCn) (Figure 45 c). In the reverse contact to the nGGCn motif, Lys100 contacts the second guanine. These structural findings are consistent with the observation that the second base of the second motif can be G or C such as nGGCnnGSCn which inverts the second motif (S representing C or G ambiguity, (Badis et al. 2008)).

5.3.8.2.1 Single amino acid change determines sequence specificity of CtSkn7

Lys100 is conserved in Skn7 sequences, whereas Hsf1 sequences have methionine in the corresponding position, as an amino acid sequence alignment of Hsf1 sequences against Skn7 sequences showed (Figure 47 a). Other residues in helix α3 are highly conserved. We therefore tested for the effect of substituting K100 in CtSkn7 with methionine. We were able to observe a switch in the binding

specificity of CtSkn7, as the mutant K100M lost its ability to bind to the SSRE motif but gained specificity for HSE (Figure 47 b).

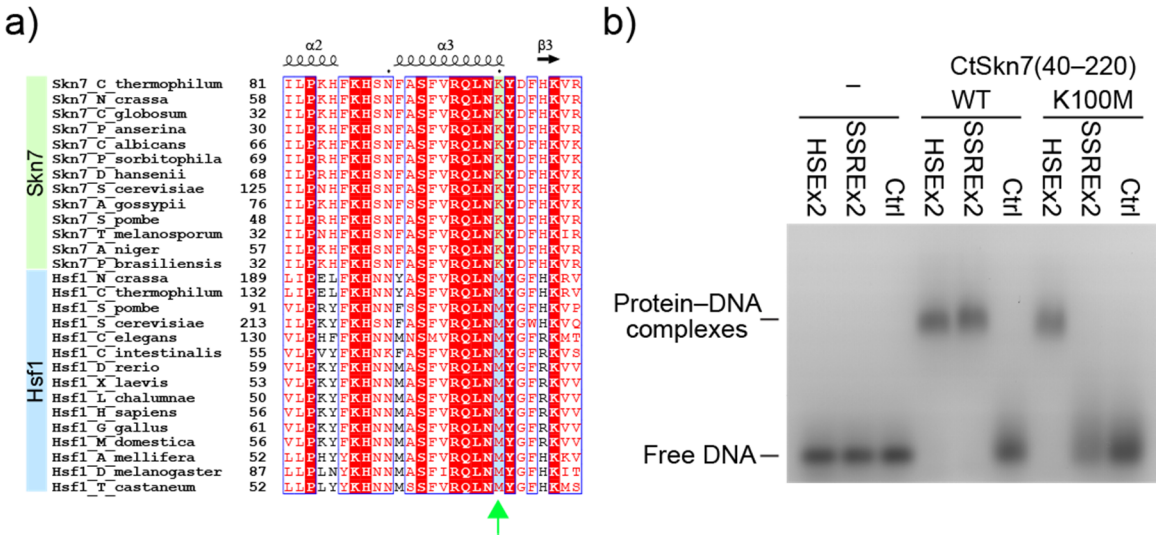

Figure 47 A single amino acid change shifts DNA sequence specificity of Hsf-type DBDs. a) Excerpt from a multiple sequence alignment of Skn7 and Hsf1 from different species. The systematic amino acid substitution is indicated by a green arrow. b) EMSA of WT and K100M mutant of CtSkn7(40-220) shows different binding preferences. The fluorescence signal in the agarose gel is shown. A representative gel of 3 independent experiments is shown. WT, wild type; Ctrl, control DNA.

5.3.8.3 Coiled-coil domain of CtSkn7 shows classic leucine zipper

We were able to solve the first structures of a heat shock factor family protein coiled-coil domain (Figure 48). The constructs CtSkn7(160-209) and CtSkn7(160-220) crystallized primarily as a tetramer, presumably due to the high protein concentrations present in the crystallization drop. The protein chains in crystal form I of CtSkn7(160-209) formed a tetrameric arrangement, whereby the asymmetric subunit contained two adjacent chains which were linked by a disulfide bond formed by Cys185. Of note, no reducing agent was present during the crystallization process. Crystal form II contained a tetrameric coiled-coil with disulfide bridges as well as a trimeric arrangement with reduced Cysteines. CtSkn7(160-220) crystallized as tetramer with reduced cysteines. The hydrophobic residues arranged in the hydrophobic heptad repeat pattern (a and d) were engaged in typical coiled-coil interactions with “knobs-into-holes” packing, forming the hydrophobic core of the coiled-coil topology. Additionally, the residues peripheral to the interaction interface (e and g) were engaged in stabilizing

interactions - e.g. hydrogen bonds between Glu198 and Gln203 from adjacent chains.

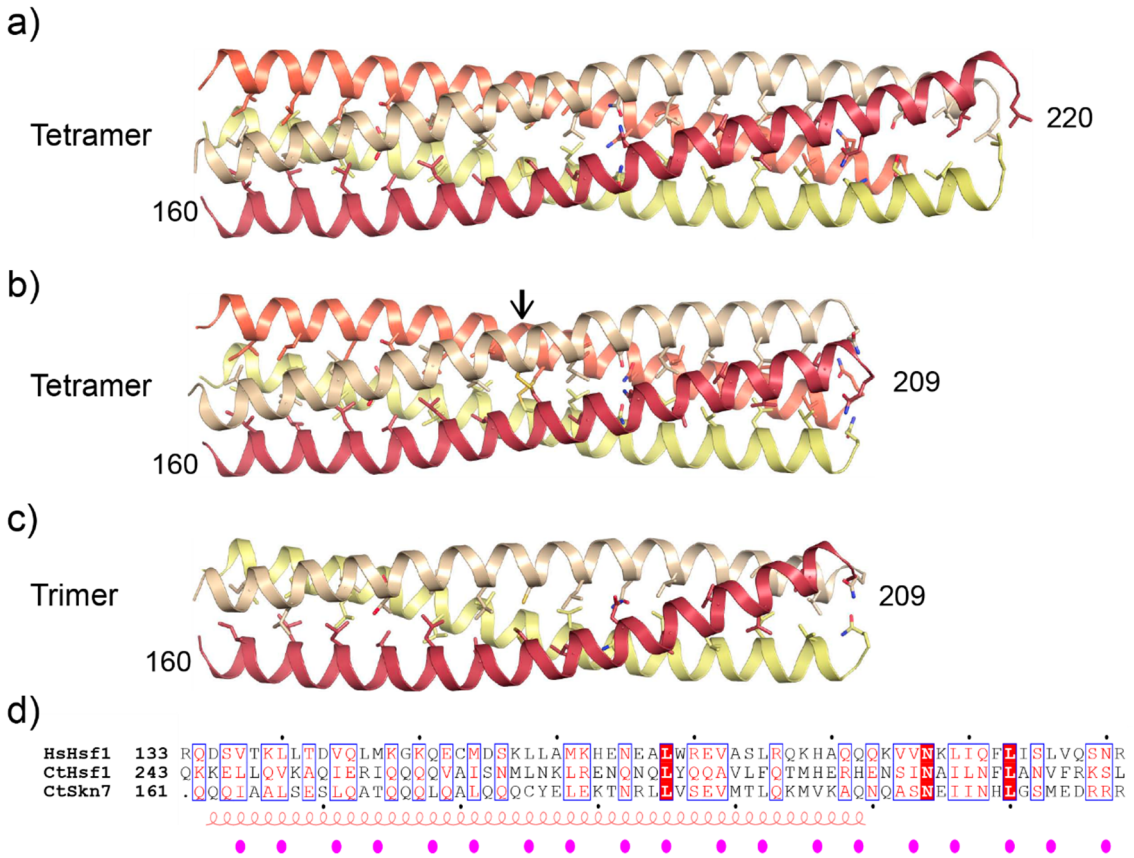

Figure 48 Crystals of CtSkn7 coiled-coil domains reveal trimeric and tetrameric structures. a) Crystal structure of CtSkn7(160-220) tetramer. b) Crystal structure of CtSkn7(160-209) tetramer. c) Crystal structure of CtSkn7(160-209) trimer. Side chains of hydrophobic-layer residues are shown in stick representation. Residues engage in typical coiled-coil interactions. d) Multiple sequence alignment of HsHsf1, CtHsf1 and CtSkn7 heptad repeat sequences. Hydrophobic heptad repeat pattern is indicated with pink dots below. Arrow indicates a disulfide bridge in CtSkn7(160-209).

Despite excellent resolution limits (1.03 Å, 1.7 Å and 1.9 Å), the R-factors of the crystallographic models stayed behind the average values for these particular resolution ranges (Table 4, (Read et al. 2011)). This might be explained by thermal movement of the chain ends, which are difficult to model, or by partial misorientation of coiled-coil bundles by 180 degrees in the crystal lattice.

5.3.9 Structure based functional analysis of CtSkn7

5.3.9.1 HsHsf1-CtSkn7 chimera shows constitutively active phenotype

To check whether the CtSkn7 heptad repeat domain is functional within human Hsf1, we replaced HsHsf1 HR (aa 133-194) with the CtSkn7 HR (160-220), which has a sequence identity of ~20% (Figure 49). Interestingly, the resulting chimera HsHsf1(CtHR) showed constitutive DNA-binding and trimerization *in vitro* (Figure 50).

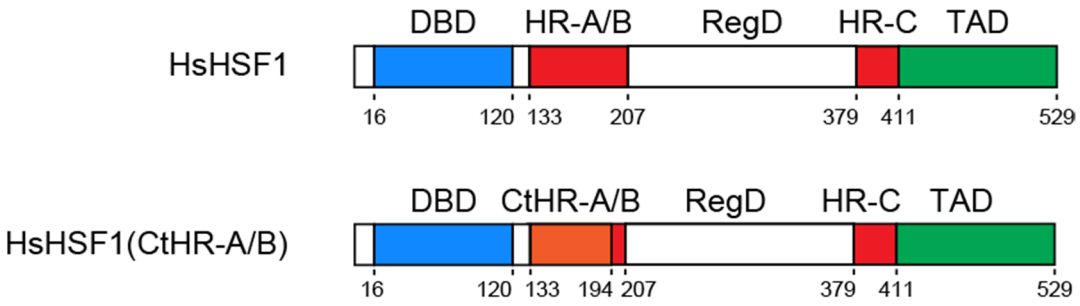

Figure 49 Schematic of the domain organization of HsHsf1 and the HsHsf1–CtSkn7 chimera, HsHsf1(CtHR-A/B). HsHsf1 HR is replaced with the respective sequence of CtSkn7 HR.

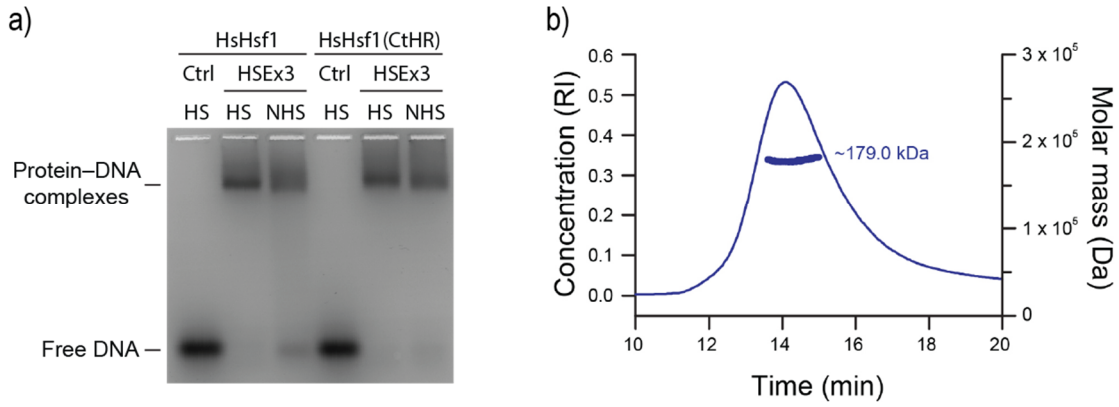

Figure 50 HsHsf1(CtHR) chimera shows constitutive activation in EMSA and constitutive trimerization in MALS. a) Purified HsHsf1-CtSkn7HR chimera shows constitutive DNA-binding in EMSA. The fluorescence signal in the gel is shown. b) HsHsf1(CtHR) is a constitutive trimer. Horizontal lines across the peaks indicate molar mass and homogeneity. Calculated molar masses are indicated. MALS was performed by Dr. Manajit Hayer-Hartl.

In vivo data confirmed that the Chimera is able to functionally substitute wild type Hsf1 in HeLa cells, showing a constitutively activated phenotype (Neudegger et al. 2016).

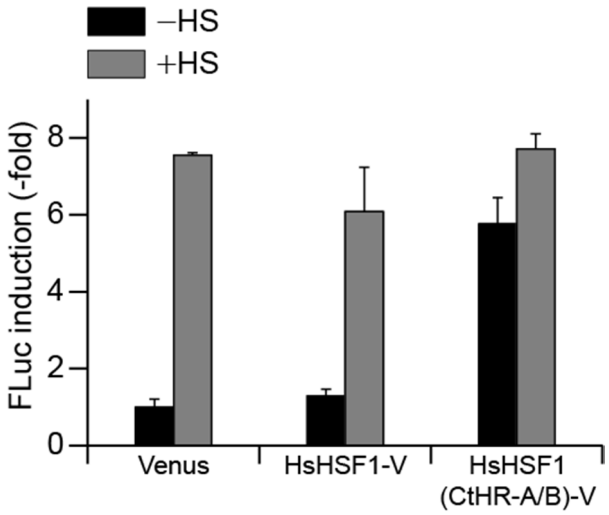

Figure 51 Induction of firefly luciferase (Fluc) expression under control of the Hsp70 promoter by different Hsf1 constructs. Shown is the relative to constitutive expression of Renilla luciferase under the CMV promoter in HeLa cells transfected with Venus (V) fusion proteins HsHsf1-V, HsHsf1(CtHR-A/B)-V or Venus alone as a control as described in (Neudegger et al. 2016). Luciferase activities were monitored with and without heat shock (±HS). Induction levels were normalized to Venus alone without heat shock. Error bars represent s.d. values from three independent experiments. Experiment was performed by Dr. Jacob Verghese.

The dissociation constants measured for the constitutively oligomeric chimera were values in the nano molar range (34.5±1.9 nM for HSEx3, 45.7±2.3 nM for HSEx4) and closely matched the literature value for a constitutively oligomerized Hsf1 mutant and HSEx3 (35±7.1 nM (Jaeger et al. 2014)) (Figure 52). The binding data was fitted with the Hill equation model (Equation 1), n was not fixed to 1 and suggested a cooperative effect with 1.48±0.13 for n (Reduced Chi-Sqr if n=1: 2.98, for n=1.48±0.13: 0.79).

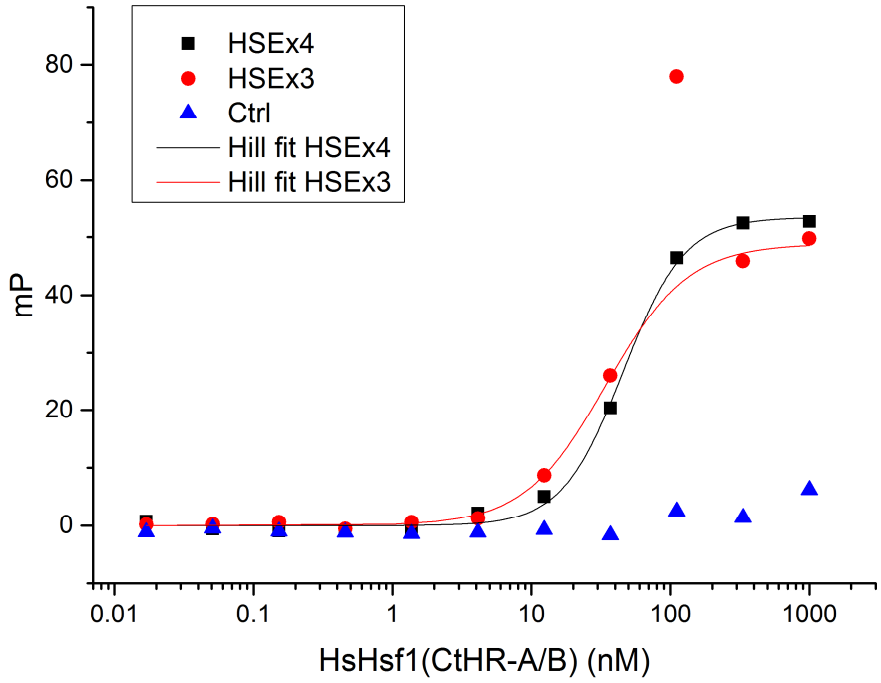

Figure 52 HsHsf1(CtHR) Chimera binds to fluorescently labeled HSE DNA-oligonucleotides with nanomolar affinity. Fluorescence anisotropy measurement is shown. Purified HsHsf1-CtSkn7HR chimera was titrated to fluorescein labeled DNA and the data fitted with the Hill equation model (Equation 1). The analysis resulted in a K_D in the nanomolar range (34.5±1.9 nM for HSEx3, 45.7±2.3 nM for HSEx4).

6 Discussion

Hsf1 was discovered more than thirty years ago, yet the structural organization and functional features underlying its molecular functions have remained unclear (Parker & Topol 1984; Wu 1984; Anckar & Sistonen 2011). In this study we used numerous approaches to gain insight into the mechanism of activation and oligomerization of human Hsf1. Our crystal structures demonstrate that previous Hsf family protein DBD structures missed a constitutive C-terminal region, which forms vital contacts with the phosphate backbone in DNA complexes. In addition, we solved the first structure of an Hsf coiled-coil oligomerization domain.

6.1 Comparison of the atomic structures of the DBDs in human and fungal Hsf1

Prior to our study, the only Hsf1 family protein-DNA complex structure available was from the yeast *K. lactis* (KlHsf1) (Littlefield & Nelson 1999). A comparison of the crystal structures of the DBD fragment of KlHsf1 with HsHsf1 showed only slight deviations (r.m.s.d. of 1.223 Å for PDB 3HTS) of the peptide backbone conformation (Figure 53). However, the protein-protein interface of the DBDs bound to the tail-to-tail motif revealed substantial differences (Figure 53).

KlHsf1 DBDs interact with each other in a manner similar to that found in β-sheets, with the carbonyl and amide groups stabilizing an antiparallel interaction of the wing loop and the linker between α-helix 2 and 3. In contrast, HsHsf1 DBDs do not engage in β-sheet contacts but exclude water molecules from the contact site on α-helix 2 – a feature that is often found in functional protein interfaces (Nero et al. 2014). Sequence conservation of the residues in helix α2 suggests that similar protein-protein contacts are present in all Hsf1 orthologs of animals.

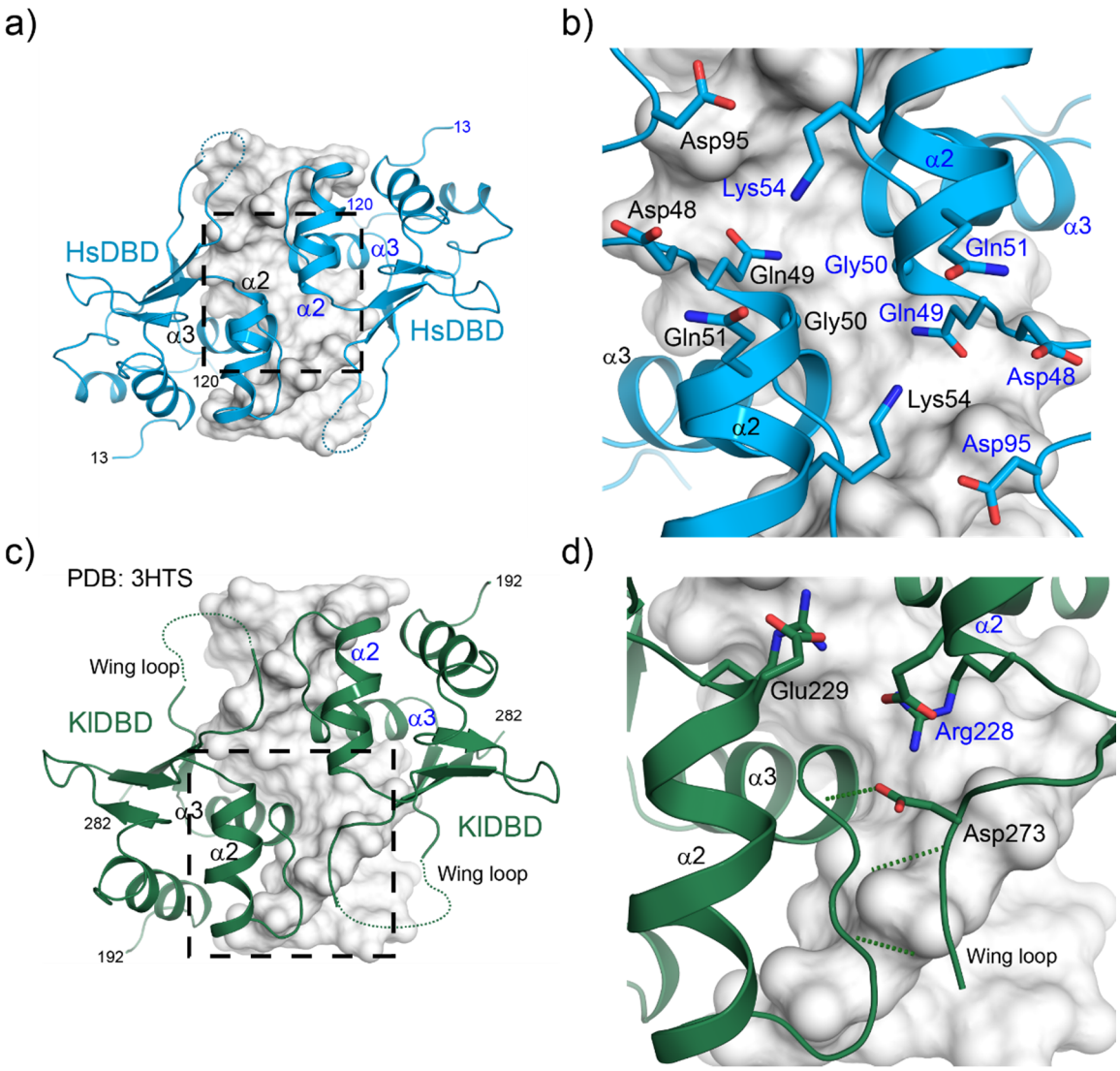

Figure 53 Differences of the DBD-DBD interface of adjacent DBDs on double stranded DNA between KlHsf1 and HsHsf1. a) and b) DBD-DBD interface of HsHsf1 DBDs on tail-to-tail HSE motif. c) and d) show the same region of KlHsf1 DBD-DBD interface on tail-to-tail HSE motif. Protein is shown in ribbon representation, DNA is shown in surface representation.

6.2 Comparison of the crystal structures of the DBDs from human Hsf1 and Hsf2

The recent HsHsf2 DBD-DNA complex structures by Jaeger and colleagues provided evidence that Hsf1 and Hsf2 have the same DNA motif preference based on the high level of conservation of the DNA-proximal residues, but are regulated differently via their divergent distal surface and wing-loops (Figure 54) (Jaeger et al. 2016). Using *in vitro* sumoylation experiments with chimeric DBDs, which had the wing-loops swapped between Hsf1 and Hsf2, they showed that the wing-loop

sequence origin dictates the regulatory modification of the wing tip lysine of the protein.

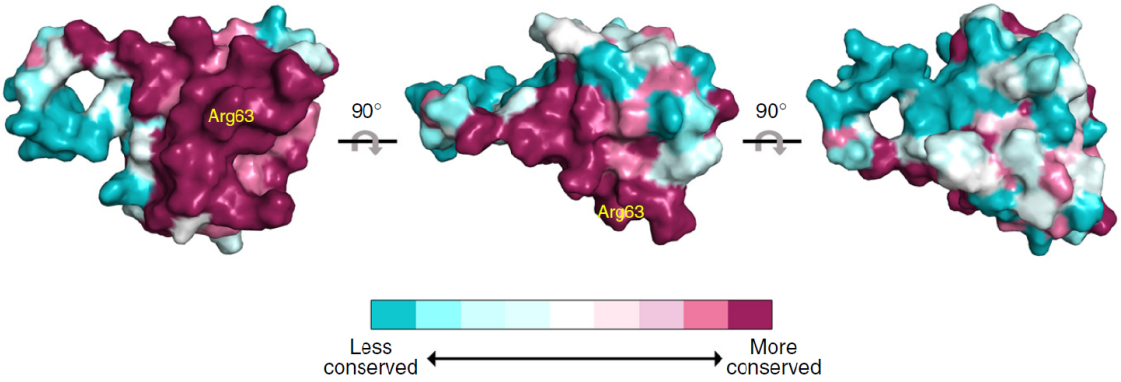

Figure 54 HsHsf1 and HsHsf2 differ in their DBD distal surface and wing-loop. Conservation of surface residues from HsHsf1 to HsHsf2 DBD mapped onto the HsHsf2 crystal structure. HsHsf2 Arg63 corresponds to Arg71 in HsHsf1 which reads out the Hoogsteen face of the guanine group of the DNA sequence nGAAn. Adapted by permission from Macmillan Publishers Ltd: Nat. Struct. Mol. Biol. (Jaeger et al. 2016), copyright 2016

When we compared the crystal structures of the DBD of HsHsf2 with HsHsf1 minor deviations were found in terms of the peptide backbone conformation (r.m.s.d. of 2.213 Å for PDB datasets 5D5U and 5D8K) (Figure 55).

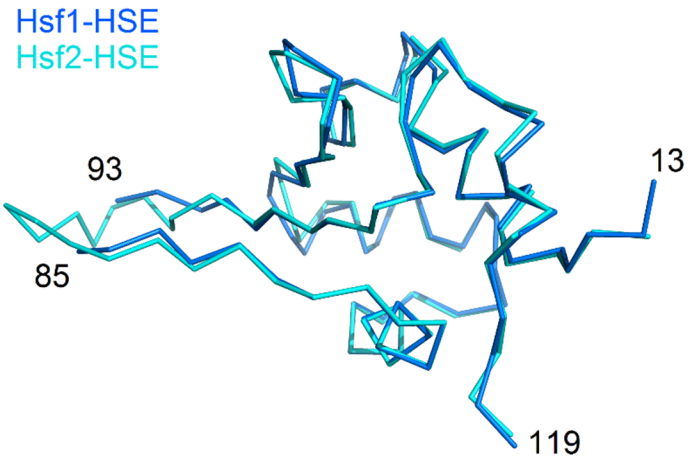

Figure 55 HsHsf1 and HsHsf2 differ slightly in their peptide backbone conformation. Superposition of the DBD structure in the HsHsf1 DBD-HSE complex and the DBD in the HsHsf2 DBD-HSE complex is shown. Displayed are Cα traces of the PDB datasets 5D5U and 5D8K (Jaeger et al. 2016).

The residues of the recognition helix α3 are highly conserved between HsHsf1 and HsHsf2 (Figure 27 and Figure 56). The sidechains of the conserved HsHsf2 residues Asn66, Gln64, Tyr68 and Ser60 (corresponding to Asn74, Gln72, Tyr76

and Ser68 in HsHsf1) position α-helix 3 in the major groove of the DNA double helix by hydrogen bond formation with the phosphate backbone of the DNA. Two pairs of positively charged side chains (Lys54, His55 and Arg71, Lys72) in the N- and C-terminal connectors of the recognition helix stabilize the interaction by hydrogen bonding and electrostatic interactions with the DNA phosphate backbone. If helix α3 is correctly positioned, HsHsf2 Arg63 (Arg71 in HsHsf1) reads out the Hoogsteen face of the guanine base of the nGAAn motif. At the C-terminal tail of the HsHsf2 DBD, Arg109 is positioned to contact the major groove or the phosphate backbone, whereas Lys110 forms a salt bridge with the negatively charged DNA backbone.

These conserved residues generally adopt identical conformations in HsHsf1 and HsHsf2. However, Arg71 in HsHsf2 showed a different sidechain conformation compared to its homologous residue, Arg79, in HsHsf1. The conformation of Arg71 in HsHsf2 allows contact to the phosphate backbone of the DNA with its ε-amino group (2.6 Å) whereas the closest backbone contact of HsHsf1 Arg79 is made by a η-amino group (3.1 Å). However, the function of both residues as mediators of generic DNA contacts is conserved.

These observations are consistent with the finding that the DNA-sequence preferences of HsHsf1 and HsHsf2 are indistinguishable (Vihervaara et al. 2013).

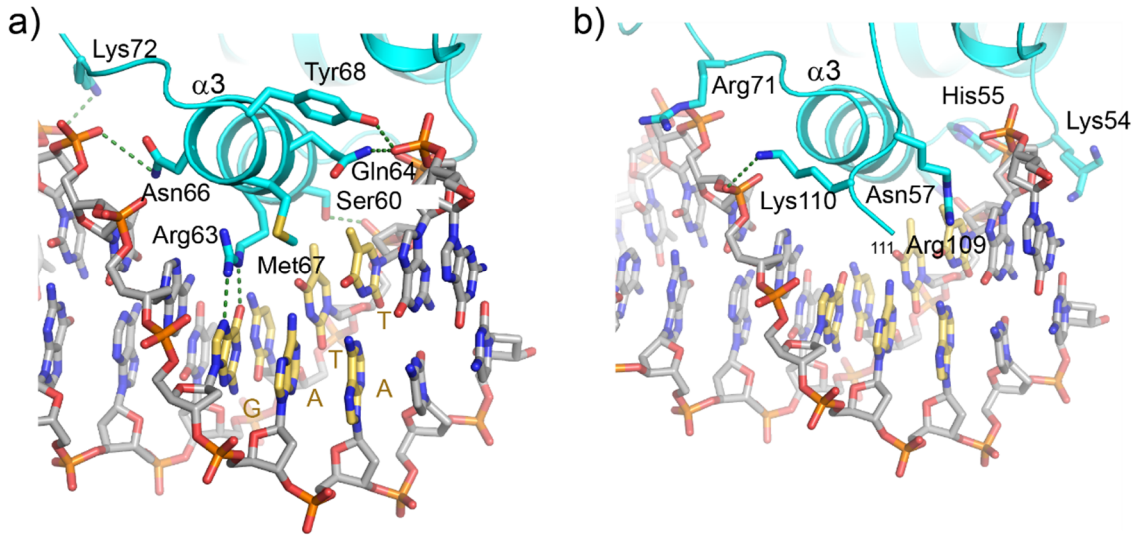

Figure 56 Crystal structure of HsHsf2 DBD-HSE complex reveals structural basis for HSE recognition and role of C-terminus. Detailed view of the DNA-contacts of the recognition helix α3 (a) and the DBD C-terminus (b). Protein is shown in ribbon representation, DNA is shown in stick representation with the phosphate backbone highlighted. Side chains of conserved residues

are shown in stick representation with hydrogen bonds represented by dotted lines. The C-terminal tail of the DBD is omitted for clarity in a) (see also **Figure 27** for comparison to HsHsf1). HsHsf2 residue Lys54 is depicted in two alternative conformations.

However, the protein-protein interface of DBDs bound to the tail-to-tail motifs revealed distinct modes of interaction (Figure 57 b, d). HsHsf2 DBDs interact with each other via the sidechain of Gln42 (corresponding to Gly50 in HsHsf1) in α-helix 2, which contacts the backbone carbonyl group of Gln42 (distance 3.0 Å) on the adjacent DBD. Furthermore, hydrogen bond formation between the sidechains of Asn56 and Asp86 (distance 2.8 Å) stabilizes the intermolecular interaction between the wing-loop and the linker between α-helix 2 and 3. Whereas the Glycine at position 50 in the HsHsf1 DBD allows a minimal backbone distance of 4.95 Å, Gln42 at the same position keeps the DBD backbones of HsHsf2 at a distance of 7.0 Å. As a consequence, each HsHsf2 DBD is shifted ~1.3 Å away from the interface when compared to HsHsf1. Interestingly, all secondary structure elements are shifted with the exception of the recognition helix α3 – its position relative to the DNA double helix matches in the HsHsf1- and HsHsf2-DNA complex structures.

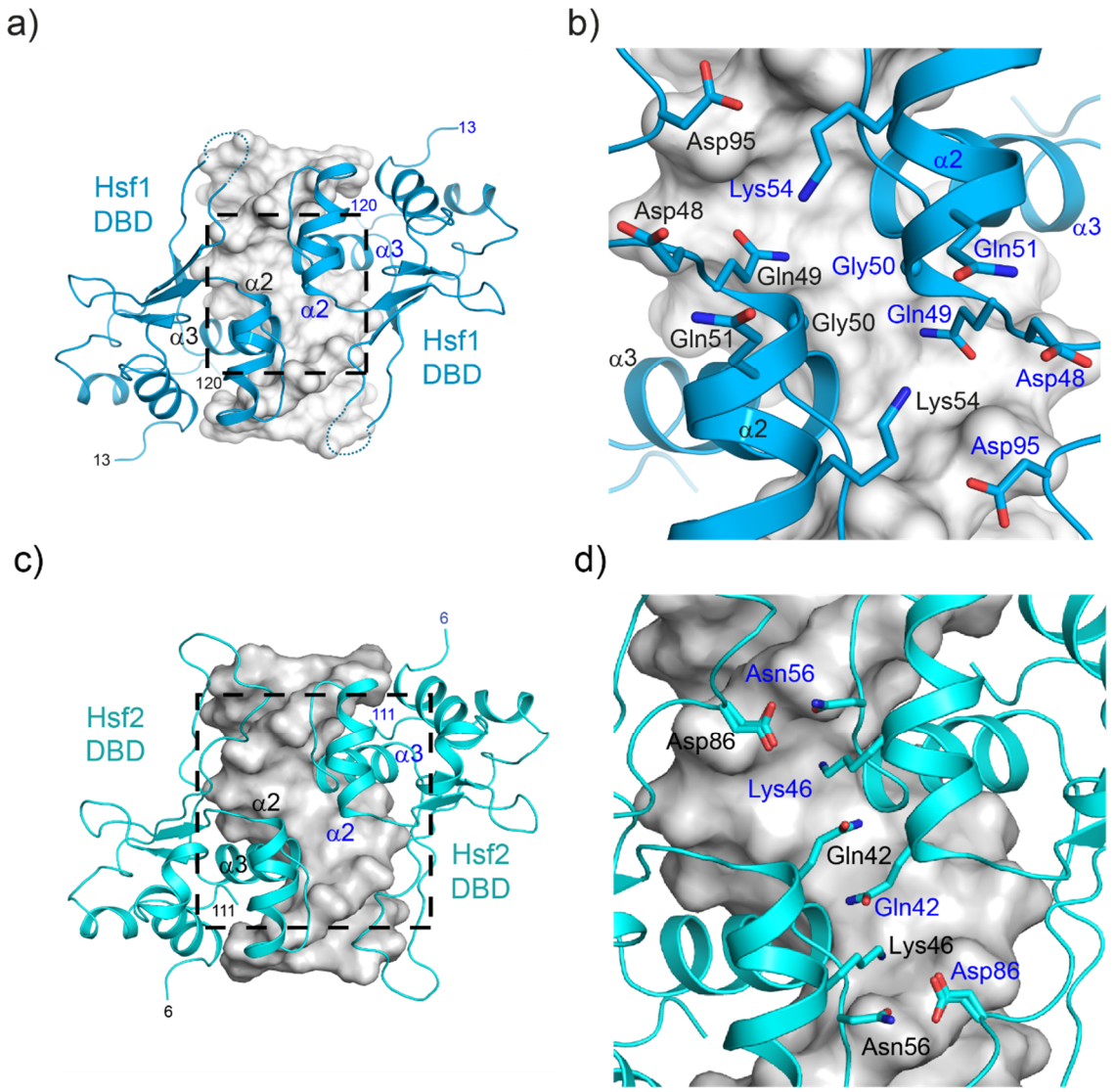

Figure 57 Comparison of the DBD-DBD interfaces between adjacent DBDs in the DNA complexes of HsHsf1 and HsHsf2. a) and b) Interface between HsHsf1 DBDs along a tail-to-tail HSE motif. c) and d) Same view of the HsHsf2 DBD-DBD interface along a tail-to-tail HSE motif (PDB 5D8K). HsHsf2 Asp86 is depicted in two alternative conformations. Protein is shown in ribbon representation, DNA is shown in surface representation. Side chains of residues within Van-der-Waals distance of the opposing monomer are shown in stick representation.

6.3 HsHsf1 and HsHsf2 in the formation of an Hsf1-Hsf2 hetero-trimer

Jaeger and colleagues also observed an interaction between Hsf1 and Hsf2 mediated by their coiled-coil domain in co-purification experiments (Figure 58) (Jaeger et al. 2016). These findings confirmed the model for hetero-oligomer formation (Sandqvist et al. 2009).

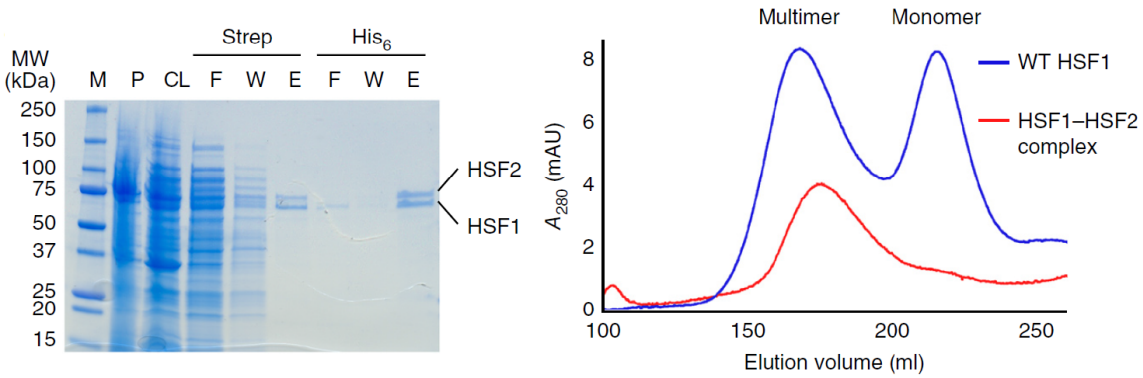

Figure 58 HsHsf1 and HsHsf2 form mixed complexes via their coiled-coil domain. HsHsf1 and HsHsf2 co-purify and form hetero complexes as monitored by SDS-PAGE analysis and size exclusion chromatography, respectively. Adapted by permission from Macmillan Publishers Ltd: Nat. Struct. Mol. Biol. (Jaeger et al. 2016), copyright 2016

In a hetero-trimer, two different kinds of protein-protein contacts along a three times nGAAn HSE motif could be mixed – tail-to-tail and head-to-tail (Figure 7 b). Conservation in the putative contact areas of a head-to-tail orientation – the wing tip and surface cleft between α-helix 1 and α-helix 2 – is high. Interestingly, the systematic sequence deviation is more pronounced in the tail-to-tail contact between HsHsf1 and HsHsf2.

Using the crystal structures, we created a model of the proposed interface between the DBD of HsHsf1 and the DBD of HsHsf2 in a hetero-trimer at the DNA in a tail-to-tail orientation (Figure 59).

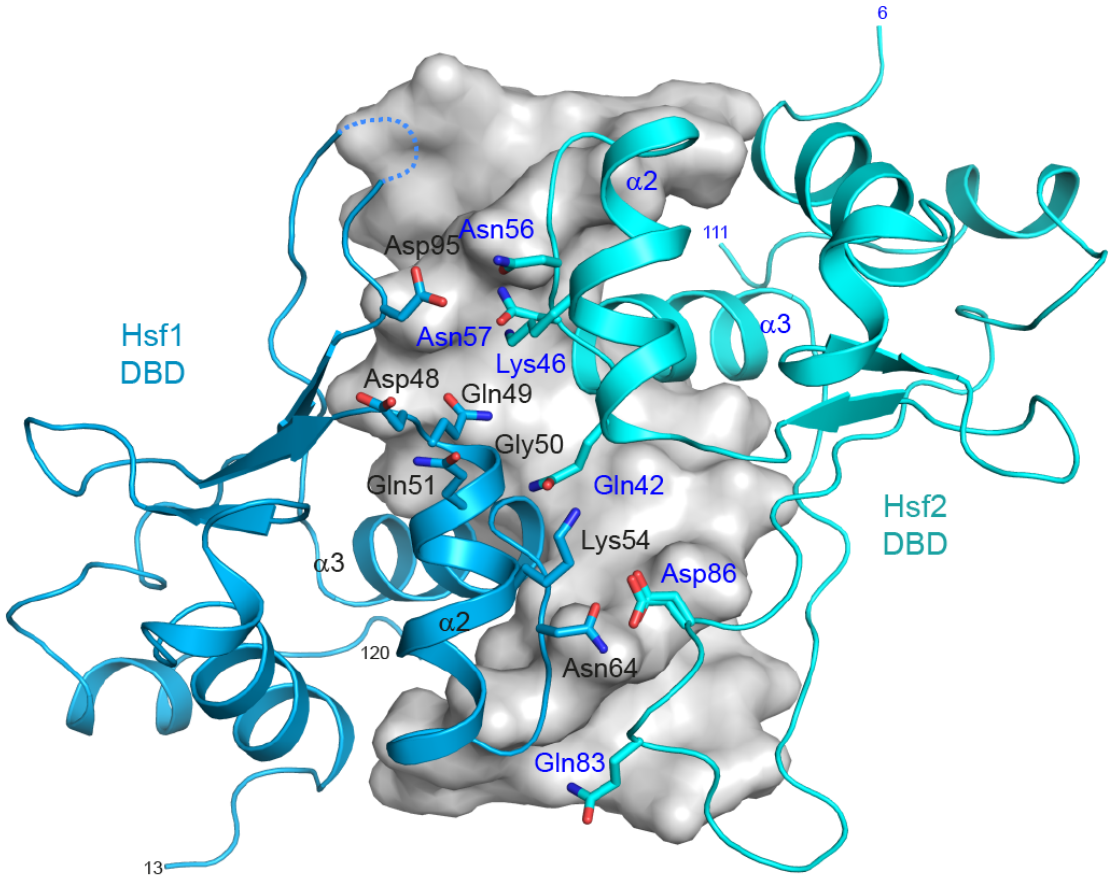

Figure 59 A model of a putative DBD-DBD interface in an HsHsf1 and HsHsf2 heterodimer on DNA double strand. HsHsf1 and HsHsf2 were aligned at HSE tail-to-tail DNA according to their crystal structures. Protein is shown in ribbon representation, DNA is shown in surface representation. Residues within a 5 Å distance are shown in stick representation. HsHsf2 Asp86 is depicted in two alternative conformations.

The interface does not bear obvious steric clashes or charge mismatches. Based on this model one might propose that the interaction is stabilized by hydrogen bonds and Van der Waals interactions between charged and polar amino acids in the wing-loops and α-helix 2 of the respective HsHsf1 and HsHsf2 DBDs. The following residues have the potential to be involved in these types of interactions and are within 5 Å of the neighboring DBD: HsHsf1 Asp48, Gln49, Gly50, Gln51, Lys54, Asn64, and Asp95; HsHsf2 Gln42, Lys46, Asn56, Gln83 and Asp86. Furthermore, ionic interactions between HsHsf1 Lys54 and HsHsf2 Asp86, as well as HsHsf2 Lys46 and HsHsf1 Asp95 could strongly stabilize this interaction. While these interactions may also stabilize a homomeric complex, these findings are consistent with a heterotrimeric model, in which such a Hsf1-Hsf2 interface was depicted in tail-to-tail orientation (Jaeger et al. 2016).

6.4 Specificity of helix-turn-helix DNA-binding domains

Systematic methionine (Hsf1) versus lysine (Skn7) substitution at the distal α-helix 3 relaxes the sequence selectivity to both HSE and SSRE motifs in the closely related Skn7 paralogs of Hsf1 (Figure 45). The sidechain flexibility of Lys100 in Skn7 allows for the recognition of both HSE and SSRE, where SSRE can be bound in two orientations, GGC and GCC (Figure 43 and (Neudegger et al. 2016)). This phenomenon has been previously observed for helix-turn-helix folds in homeodomains (Fraenkel et al. 1998; Treisman et al. 1989; Tucker-Kellogg et al. 1997). In the case of Engrailed, a homeodomain transcription factor involved in development, a single amino acid substitution to a lysine (Q50K in *D. melanogaster*) switches specificity towards guanine instead of thymine or adenine. Similar to Lys100 in CtSkn7, Lys50 of Engrailed makes contacts with the C6 oxygen of the guanine base via hydrogen bonding, driving specific binding to the motif TAATCC instead of TAATTA. Therefore, the systematic sequence deviation between Skn7 and Hsf1 represents another example of how a single amino acid substitution can determine sequence specificity for DNA-binding motifs.

6.5 Model of the active human Hsf1 on HSE DNA

Until a HsHsf1 coiled-coil domain structure becomes available, the structure of the coiled-coil domain of its paralog, Skn7, may serve as an informed model for the HsHsf1 HR-A/B domain structure. Based on the ~20% amino acid sequence identity between HsHsf1 and CtSkn7, we created a homology model of HsHsf1 HR-A using Modeller (Eswar et al. 2007). The model with the lowest Discrete Optimized Protein Energy (DOPE) of the 20 calculated was selected as a working model. Using this as a base, we were able to assemble a model for the conserved N-terminal (1-182) sequence of trimeric HsHsf1 on a 3-site HSE DNA double-strand consistent with our CXMS data (Figure 18 and Figure 60 a). For the linkers between the DBD and HR, several conformations were calculated using Modeller before manual selection of the final model. For the DBD-DNA contacts, a model of ideal B-DNA with three HSE sites was first created using Coot. This was then aligned with the HsHsf1 DBD - HSE crystal structures to position the three DBDs on their binding sites (Emsley & Cowtan 2004). The interaction of the wing-loop

with the adjacent DBD (head-to-tail interaction) was regularized by energy optimization of the whole model by Chimera for two rounds (Pettersen et al. 2004).

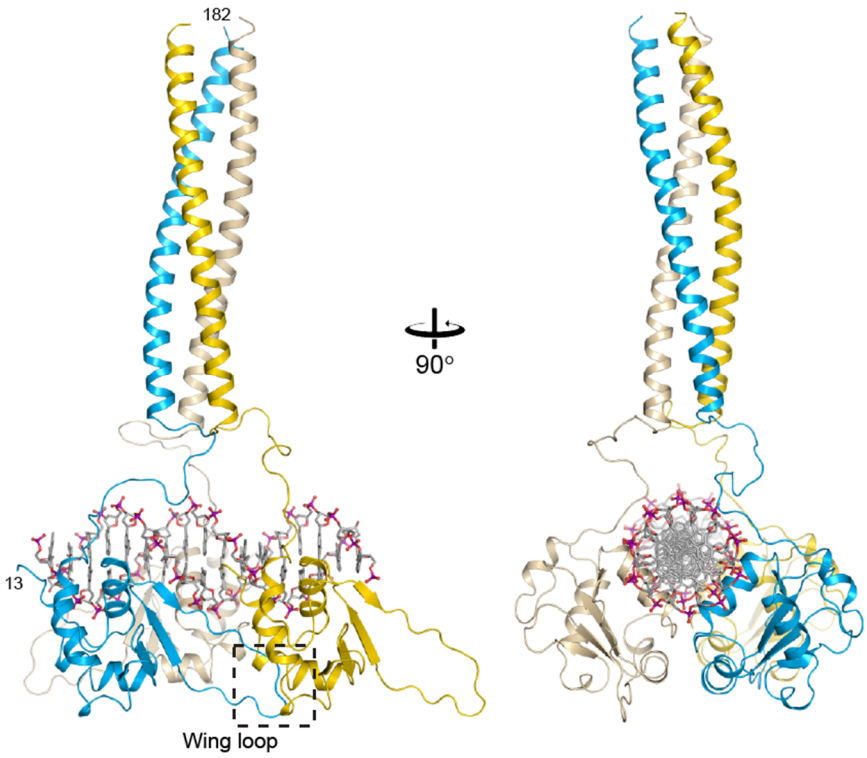

Figure 60 Model of N-terminal HsHsf1 in complex with HSE DNA-oligonucleotide. Model of N-terminal HsHsf1 with modelled linkers and homology model of HsHsf1 HR-A based on the crystal structure of CtSkn7(160-209). Protein is shown in ribbon representation, DNA is shown in stick representation.

Past models of trimeric Hsf1 placed the DBDs, coiled-coil domains and the remaining part of the protein on the same side of the DNA, due to the assumption that the DBD terminates ~20 residues earlier (Wu 1995; Anckar & Sistonen 2011). The crystal structures we obtained for the complete DBDs of HsHsf1 and CtSkn7 suggest a different mode: the DBD with the linker embraces the DNA and therefore sits on the opposite side of the DNA double helix. The basic residues at the C-terminal end of the DBD contribute to DNA affinity. This is consistent with the finding for the regulatory role for Lys118, which can be acetylated during post-translational modification. Mutants which mimic the PTM by inverting the charge (K1118Q) have been found to lose their DNA-binding capacity when monitored via the proxy of nuclear stress body formation (Raychaudhuri et al. 2014).

6.6 Binding of HsHsf1 to Satellite III repeats

The development of a model for trimeric HsHsf1 on Satellite III repeats proved to be more difficult. In Sat III repeats (GGAAT) the order and direction of individual nGAAn motifs is different from those found in HSEs. They consist of head-to-tail unidirectional arrays of nGAAn for long stretches. This abolishes the extended tail-to-tail contacts between adjacent DBDs since all DBDs have the same orientation along the DNA helix. Only the head-to-tail contacts with every second DBD C-terminus remain apart from the coiled-coil to support the contacts to DNA (Figure 61).

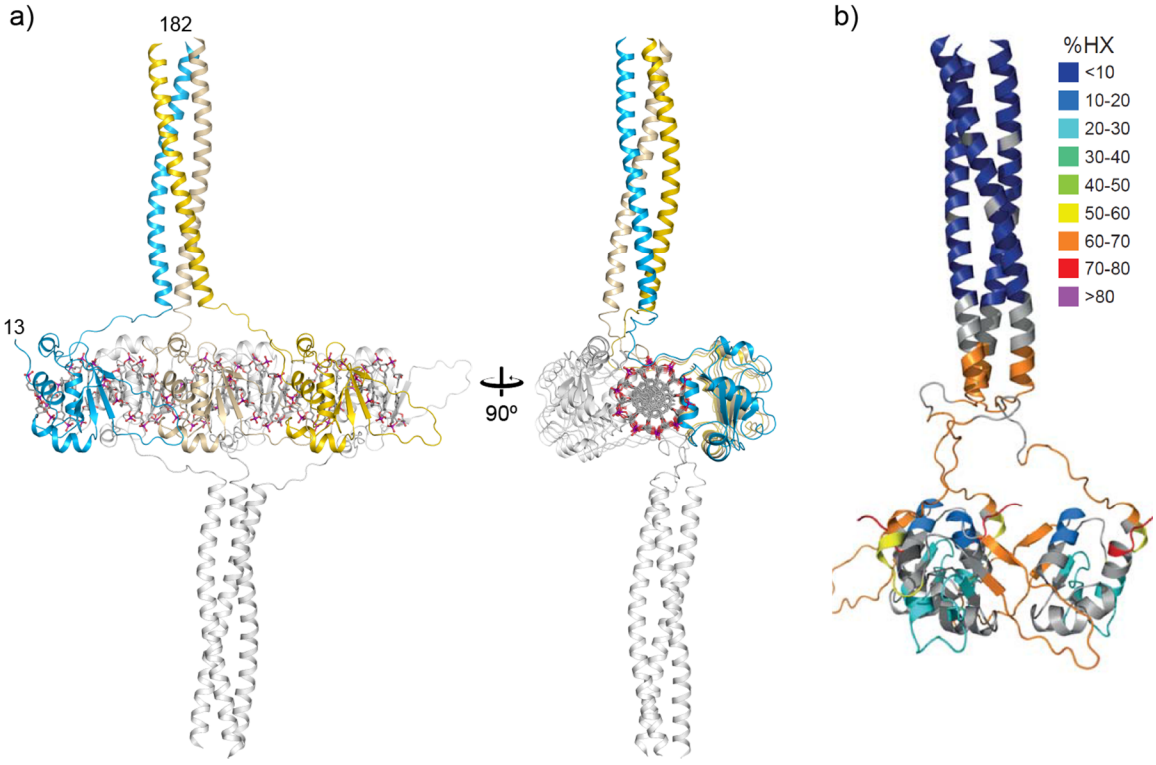

Figure 61 Binding of HsHsf1 to SatIII repeats requires fully extended linker conformation or partial unzipping of the coiled-coil. a) Model of N-terminal HsHsf1 bound to 6 motifs of SatIII repeats. Coiled-coil is homology model based on the crystal structure of CtSkn7 HR. b) Hydrogen/Deuterium exchange levels in oligomeric HsHsf1 mapped onto the model showing partial unprotection of the N-terminal coiled-coil domain consistent with partial unzipping. Figure b) adapted from (Hentze et al. 2016).

This linear arrangement would force two of the three linkers to adopt a fully extended conformation unless there is support from concave DNA bending, which would strain an opposing Hsf1 trimer. Another possibility would be that the N-terminus of the triple stranded coiled-coil partially unzips. Indeed, a study by

Hentze and colleagues found that the N-terminus (residues 130-140) of the coiled-coil domain is still partially unprotected in trimeric HsHsf1 against Hydrogen/Deuterium exchange (Hentze et al. 2016) (Figure 61). Hydrogen/Deuterium exchange (H/D exchange) coupled to mass spectrometry makes use of changes in the exchange rate of peptide backbone amide hydrogen atoms. The rate of exchange is influenced by solvent accessibility and hydrogen bonding of the respective hydrogen. Based on currently available data, the partial unzipping model therefore seems to be more plausible than the former. Partial unzipping might also enable pentameric HsHsf1 to bind to all five binding sites in the promoter of Hsp70.1 (see paragraph 5.1.2 and Figure 7).

6.7 Role of the wing-loop in protein-protein interfaces

The interaction between DBDs in head-to-tail motifs has been suggested to be mediated by the wing-loop (Littlefield & Nelson 1999). In previous studies, the Nelson group observed that yeast Hsf1 requires the wing-loop to exhibit its full activity and mutation of a conserved Glycine (Gly87 in HsHsf1) at the tip of the wing is detrimental to its function (Cicero et al. 2001; Hubl et al. 1994).

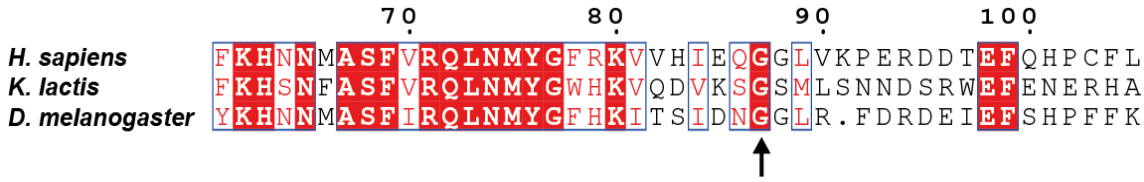

Figure 62 Excerpt of a sequence alignment of Hsf1 DBD amino acid sequences. Wing tip glycine (Gly87 in HsHsf1) is a conserved residue.

Based on our results and published observations, the wing loop might be involved in two modes of interaction. First, Replication protein A 1 (RPA1) of the hetero trimeric RPA complex requires the wing-loop of HsHsf1 for interaction (Fujimoto et al. 2012). RPA is a single-stranded DNA-binding complex (consistent of subunits RPA1-3) involved in DNA metabolism, recombination, repair and gene transcription (Wold 1997). Gly87 in the wing-loop of HsHsf1 proved to be essential for the physical interaction with RPA1. The DNA-bound Hsf1-RPA complex was shown to contribute to the opening of chromatin structure and transcription of

target genes by facilitating Pol II preloading and recruitment of the histone chaperone FACT (facilitates chromatin transcription) (Fujimoto et al. 2012).

Second, we observed the wing-loop reaching out to a hydrophobic surface patch at the groove between α-helix 1 and α-helix 2 of the adjacent DBD in in our crystal structure of the HsHsf1 DBD-SatIII complex (Figure 28 c). Our trimer model based on our crystal structure suggests a potential alternative mechanism in the DBD-DBD contacts, in which the extended wing-loop of DBD1 might be inserted into the groove between α-helix 1 and α-helix 2 of DBD3 (Figure 60). Either way, the wing-loop might moderately stabilize the interaction of adjacent DBDs on the DNA. Presumably, this effect only plays a role when at physiological HsHsf1 levels and not when over-expressed (Fujimoto et al. 2012).

The wing-loop of Hsf2 has been shown to contribute to HsHsf2 regulation via sumoylation at Lys82 but Hsf2 does not interact with RPA. Lysine sumoylation was not observed at HsHsf1 wing-loop residue Lys91 (Jaeger et al. 2016; Fujimoto et al. 2012; Hietakangas et al. 2003).

Inhibition or activation of Hsf1 has been proposed to be a promising approach for the treatment of cancer or proteopathies, respectively (Neef et al. 2011; Whitesell & Lindquist 2009). Several compounds which activate or inhibit the function of Hsf1 have been reported (e.g. Triptolide as inhibitor and Celastrol as inducer of the HSR). However, all compounds acting on the HSR have been discovered in cell-based assays and the mode of action has not been tested on isolated Hsf1 *in vitro* (Calamini et al. 2011; Au et al. 2009). Therefore most, if not all, discovered small molecules activating or inhibiting the HSR presumably act up- or downstream of Hsf1 and do not interact with Hsf1 itself. Thus, the binding of all compounds found to be active must be confirmed in *in vitro* experiments using purified HsHsf1 in future studies to rule out off target effects.

However, the putative protein-protein interfaces (PPI) mediated by the wing-loop in inter-HsHsf1 contacts or HsHsf1-RPA interactions might be an interesting target site for future drug development. Also, it is known that protein flexibility (as provided by a Gly backbone), such as induced fit mechanisms, enhance the drug ability of a given PPI (Teague 2003; Mangani 2013). Once candidate compounds are identified, rational drug design involving crystal structures of the respective

protein complexes with and without the compound might promote the development of a direct Hsf1 inhibitor.

6.8 An activation model for Hsf1

The overall topology of metazoan Hsf1 is still debated – e.g. which domains interact and how they interact. Of particular interest was the arrangement of the coiled-coil domains in the latent and active states (Wu 1995; Anckar & Sistonen 2011). Early studies had indicated, that the N-terminal coiled-coil domain comprises two subdomains – HR-A and HR-B. It was found that HR-A of yeast Hsf1 favors a trimeric state when studied in isolation, whereas HR-B showed the propensity to form higher oligomers (Peteranderl et al. 1999; Peteranderl & Nelson 1992). However, it was generally accepted that the regulatory C-terminal coiled-coil domain (HR-C) folds back onto the N-terminal heptad repeats to stabilize the monomeric state of the protein, whereas deletion of HR-C leads to constitutive oligomeric Hsf1 (Rabindran et al. 1993; Wu 1995). The Voellmy group introduced mutations in all three HR domains (HR-A, B and C) of HsHsf1, revealing that all three heptad repeat domains are required to maintain the monomeric state (Zuo et al. 1994; Zuo et al. 1995). Using GAL4-Hsf1 chimeric proteins, it was also shown that the regulatory domain (RD residues ~220-310), situated between the HR domains, was both able to repress and enhance the activity of HsHsf1 (Green et al. 1995).

The molecular events which happen during the transition from monomeric Hsf1 to oligomeric Hsf1, however, mainly remained an enigma. A study by Hentze and colleagues have recently provided insight into the mechanism (Hentze et al. 2016). With the help of H/D exchange coupled to MS, they were able to determine which regions are folded or unfolded at specific time points during the heat activated oligomerization process. Further, Hentze and colleagues proposed that HsHsf1 might exist in a monomer-dimer equilibrium in its latent state. However, the dimeric species was absent in our native MS analysis (paragraph 5.1.2, (Hentze et al. 2016)).

In light of our findings, we propose that HsHsf1 exists predominantly as monomer in its inactive state. The following domain topology for the HsHsf1 monomer is consistent with the data from CXMS (paragraph 5.1.3) and data from mutational

studies (Zuo et al. 1995; Zuo et al. 1994): The coiled-coil domains of Hsf1 are folded in a triple stranded, antiparallel coiled-coil involving HR-A, HR-B and HR-C (Figure 63). This arrangement also explains our experiments detecting crosslinks between the DBD and the end of HR-B such as Lys80-Lys206. The DBD is folded back onto the triple stranded coiled-coil, where it interacts with the RD and the TAD, which is orientated towards the DBD. This orientation is supported by additional crosslinks between Met1-Lys184, Lys80-Lys206, Lys80-Lys298, and Lys118-Lys524.

Transition of HsHsf1 to its active oligomeric state might be induced by a coupled process, dependent on concentration and temperature (Hentze et al. 2016). The nature of the oligomeric state was described as trimeric early on and received continuous support by many studies (Peteranderl & Nelson 1992; Sistonen et al. 1994; Wu 1995). However, the tendency of HsHsf1 to form higher oligomers had caused doubt towards the exclusive presence of a trimer (Raychaudhuri et al. 2014). By reviewing the data supportive for the trimeric state in combination with our new native MS data (suggesting the presence of HsHsf1 trimers and pentamers after heat activation), it seems plausible that the majority of oligomeric HsHsf1 is trimeric upon activation (paragraph 5.1.2). Upon HsHsf1's transition to its predominantly trimeric state, HsHsf1 subunits arrange in a parallel, elongated state from residues ~130 to ~300. This arrangement is consistent with our inter-subunit crosslinks Lys131-Lys131, Lys178-Lys184, Lys206-Lys206, and Lys298-Lys 298. The TAD remains orientated towards the RD (Lys298-Lys524) (Figure 63) (paragraphs 5.1.2, 5.1.3, (Sistonen et al. 1994)).

The crosslinks we identified within the HsHsf1 DBD connect residues with C_α-C_α distances of 6-37 Å for DSS and 7-31 Å for DMTMM crosslinks (paragraph 9.5). These values are consistent with previous studies which found that more than 90% of DSS crosslinks were within C_α-C_α distances 30 Å and more than 85% of DMTMM crosslinks were within 25 Å (Table 5, Table 6) (Leitner et al. 2014; Leitner et al. 2012).

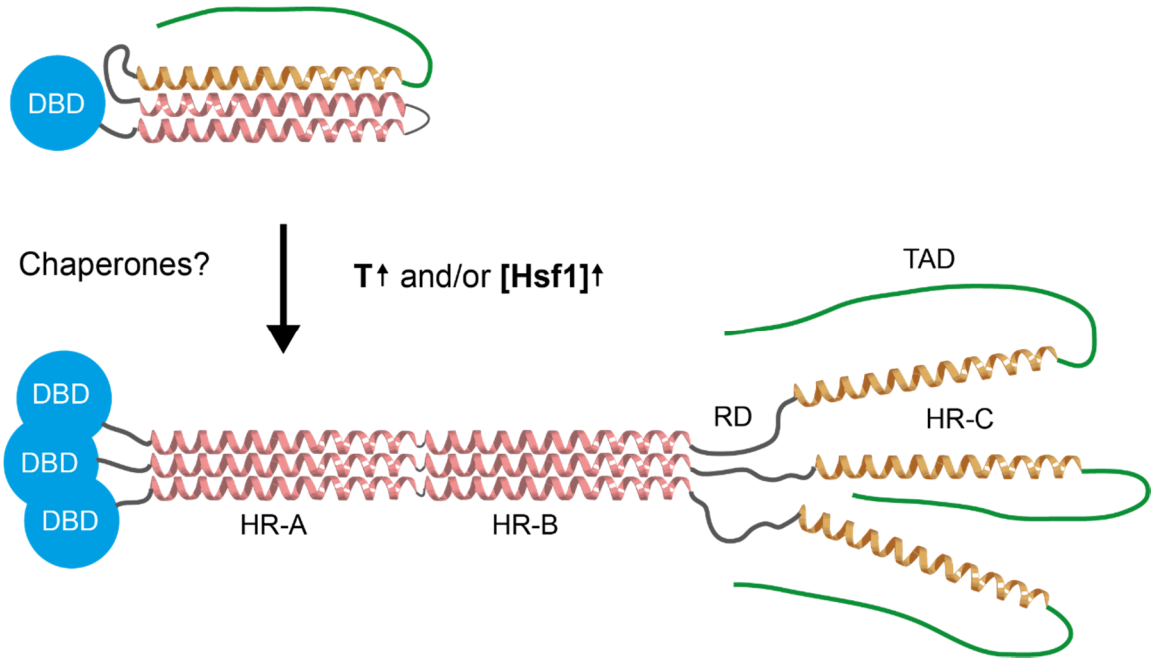

Figure 63 Activation model for human Hsf1. HsHsf1 adopts a compact conformation in its monomeric state, which unfolds into an elongated, parallel coiled-coil topology upon activation. Activation occurs under the influence of temperature, concentration and molecular chaperones.

6.9 The role of molecular chaperones in Hsf1 regulation

However, it should be noted that the results obtained in this study and the one of Hentze and colleagues were derived from HsHsf1 heterologously expressed in *E. coli* and thus lacking PTMs. Furthermore, a role of interacting proteins (e. g. molecular chaperones) within the cellular environment should be taken into consideration. To date, molecular chaperones have been mainly described as inhibitors in a feedback regulation pathway (Björk & Sistonen 2010).

Intriguingly, there are two studies, which suggest that molecular chaperones render HsHsf1 more activation-prone (Ahn et al. 2005; Hentze et al. 2016). Ahn and colleagues found that knockdown of Hsc70 expression decreased the ability of Hsf1 to induce expression of target genes, whereas overexpression of Hsc70 increased activation of Hsf1 – including the formation of higher amounts of oligomeric Hsf1. Hentze and colleagues found that the presence of Hsp90 seems to lower the midpoint of the activation temperature. Hsp90 also caused a change in the temperature-dependent activation curve towards a shallower slope, apparently lowering the cooperativity of the transition. According to their model, Hsp90 might act through destabilization and unfolding of HR-C while stabilizing

the dimeric state. Therefore, Hsp90 might facilitate Hsf1 oligomerization at lower concentrations in the coupled process – consistent with the low endogenous expression level of HsHsf1 (~100 nM) in human cells. Thus, the exact role of chaperones in the Hsf1 activation cycle awaits further exploration (Figure 63).

7 Abbreviations

Δ	Deletion
µL	Microliter
µM	Micromolar
CD	Circular dichroism
CtSkn7	*Chaetomium thermophilum* Skn7
DBD	DNA-binding domain
DNA	Desoxyribonucleic acid
EM	Electron microscopy
Fab	Fragment antigen binding
FACT	Facilitates chromatin transcription
FPLC	Fast protein liquid chromatography
H/D exchange	Hydrogen/Deuterium exchange (coupled to MS)
HeLa cells	Cervical cancer cell line derived from Henrietta Lacks
HEPES	N-(2-hydroxyethyl)piperacin-N'-2-ethanesulfonic acid
HIP	Hsc70-interacting protein
HOP	Hsp70/Hsp90-organizing protein
HR	Heptad repeat
Hsc70	Heat shock cognate protein 70
HSE	Heat shock element
HsHsf1	*Homo sapiens* Hsf1
HsHsf2	*Homo sapiens* Hsf2
Hsp	Heat shock protein
IMAC	Immobilized metal affinity chromatography
kDa	Kilodalton
mL	Milliliter

mM	Millimolar
Ni	Nickel
PAGE	Poly acrylamide gel electrophoresis
PCR	Polymerase chain reaction
RD	Regulatory domain
Rg	Radius of gyration
Rs	Stokes radius
SDS	Sodium dodecyl sulfate
SAT III	Satellite III (repeats)
SEC	Size exclusion chromatography
S-SAD	Sulfur-Single wavelength anomalous dispersion
TAD	Trans-activation domain
TAE	Tris-acetate EDTA
Tris	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet light

8 Bibliography

Ahn, S.G. et al., 2005. Heat-shock cognate 70 is required for the activation of heat-shock factor 1 in mammalian cells. *Biochem J*, 392(Pt 1), pp.145–152. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16050811\nhttp://www.ncbi.nlm.nih.gov/pmc/articles/PMC1317673/pdf/bj3920145.pdf.

Ahn, S.-G., 2003. Redox regulation of mammalian heat shock factor 1 is essential for Hsp gene activation and protection from stress. *Genes & Development*, 17(4), pp.516–528. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=195992&tool=pmcentrez&rendertype=abstract.

Akerfelt, M., Morimoto, R.I. & Sistonen, L., 2010. Heat shock factors: integrators of cell stress, development and lifespan. *Nature reviews. Molecular cell biology*, 11(8), pp.545–55. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3402356&tool=pmcentrez&rendertype=abstract.

Allen, B.L. & Taatjes, D.J., 2015. The Mediator complex: a central integrator of transcription. *Nature Reviews Molecular Cell Biology*, 16(3), pp.155–166. Available at: http://www.nature.com/doifinder/10.1038/nrm3951.

Amlacher, S. et al., 2011. Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile. *Cell*, 146(2), pp.277–289. Available at: http://dx.doi.org/10.1016/j.cell.2011.06.039.

Anckar, J. & Sistonen, L., 2011. Regulation of H <scp>SF</scp> 1 Function in the Heat Stress Response: Implications in Aging and Disease. *Annual Review of Biochemistry*, 80(1), pp.1089–1115. Available at: http://www.annualreviews.org/doi/abs/10.1146/annurev-biochem-060809-095203.

Andreeva, A. et al., 2007. Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Research*, 36(Database), pp.D419–D425. Available at: http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkm993.

Anfinsen, C.B., 1973. Principles that govern the folding of protein chains. *Science (New York, N.Y.)*, 181(4096), pp.223–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/4124164.

Au, Q. et al., 2009. Identification of Inhibitors of HSF1 Functional Activity by High-Content Target-Based Screening. *Journal of Biomolecular Screening*, 14(10), pp.1165–1175. Available at: http://jbx.sagepub.com/cgi/doi/10.1177/1087057109347472.

Auluck, P.K. et al., 2002. Chaperone suppression of alpha-synuclein toxicity in a Drosophila model for Parkinson's disease. *Science (New York, N.Y.)*, 295(5556), pp.865–868.

Badis, G. et al., 2008. A Library of Yeast Transcription Factor Motifs Reveals a Widespread Function for Rsc3 in Targeting Nucleosome Exclusion at Promoters. *Molecular Cell*, 32(6), pp.878–887. Available at: http://dx.doi.org/10.1016/j.molcel.2008.11.020.

Balch, W.E. et al., 2008. Adapting Proteostasis for Disease Intervention. , 319(February), pp.916–919.

Balchin, D., Hayer-Hartl, M. & Hartl, F.U., 2016. In vivo aspects of protein folding and quality control. *Science*, 353(6294), pp.aac4354–aac4354. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.aac4354.

de Beer, T.A.P. et al., 2014. PDBsum additions. *Nucleic acids research*, 42(Database issue), pp.D292–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24153109.

Ben-Zvi, A., Miller, E. a & Morimoto, R.I., 2009. Collapse of proteostasis represents an early molecular event in Caenorhabditis elegans aging. *Proceedings of the National Academy of Sciences of the United States of America*, 106(35), pp.14914–14919.

Biamonti, G. & Vourc'h, C., 2010. Nuclear stress bodies. *Cold Spring Harbor perspectives in biology*, 2(6), pp.1–13.

Björk, J.K. & Sistonen, L., 2010. Regulation of the members of the mammalian heat shock factor family. *FEBS Journal*, 277, pp.4126–4139.

Brennan, R.G. & Matthews, B.W., 1989. The helix-turn-helix DNA binding motif.

Journal of Biological Chemistry , 264 (4), pp.1903–1906. Available at: http://www.jbc.org/content/264/4/1903.short.

Budzyński, M. a. et al., 2015. Uncoupling Stress-Inducible Phosphorylation of Heat Shock Factor 1 from Its Activation. *Molecular and Cellular Biology*, 35(14), pp.2530–2540. Available at: http://mcb.asm.org/lookup/doi/10.1128/MCB.00816-14.

Bulawa, C.E. et al., 2012. Tafamidis, a potent and selective transthyretin kinetic stabilizer that inhibits the amyloid cascade. *Proceedings of the National Academy of Sciences*, 109(24), pp.9629–9634.

Calamini, B. et al., 2011. Small-molecule proteostasis regulators for protein conformational diseases. *Nature Chemical Biology*, 8(2), pp.185–196. Available at: http://dx.doi.org/10.1038/nchembio.763.

Calloni, G. et al., 2012. DnaK Functions as a Central Hub in the E. coli Chaperone Network. *Cell Reports*, 1(3), pp.251–264. Available at: http://linkinghub.elsevier.com/retrieve/pii/S2211124711000179.

Carrell, R.W. & Lomas, D.A., 1997. Conformational disease. *Lancet*, 350(9071), pp.134–138.

Catanzariti, A.-M. et al., 2004. An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein science : a publication of the Protein Society*, 13(5), pp.1331–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2286746&tool=pmcentrez&rendertype=abstract.

Chen, B. et al., 2005. The HSP90 family of genes in the human genome: Insights into their divergence and evolution. *Genomics*, 86(6), pp.627–637.

Ciani, B. et al., 2010. Molecular basis of coiled-coil oligomerization-state specificity. *Proceedings of the National Academy of Sciences of the United States of America*, 107(46), pp.19850–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21045134.

Cicero, M.P. et al., 2001. The wing in yeast heat shock transcription factor (HSF) DNA-binding domain is required for full activity. *Nucleic acids research*, 29(8), pp.1715–1723.

Ciechanover, A. & Kwon, Y.T., 2015. Degradation of misfolded proteins in neurodegenerative diseases : therapeutic targets and strategies. , 47(3), pp.e147–16. Available at: http://dx.doi.org/10.1038/emm.2014.117.

Collaborative Computational Project, N. 4, 1994. The CCP4 suite: programs for protein crystallography. *Acta crystallographica. Section D, Biological crystallography*, 50(Pt 5), pp.760–3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15299374.

Corpet, F., 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic acids research*, 16(22), pp.10881–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2849754.

Dai, C. et al., 2007. Heat Shock Factor 1 Is a Powerful Multifaceted Modifier of Carcinogenesis. *Cell*, 130(6), pp.1005–1018. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092867407009579.

Dai, C. & Sampson, S.B., 2015. HSF1 : Guardian of Proteostasis in Cancer. *Trends in Cell Biology*, xx, pp.1–12. Available at: http://dx.doi.org/10.1016/j.tcb.2015.10.011.

Decker, K.B. & Hinton, D.M., 2013. Transcription Regulation at the Core: Similarities Among Bacterial, Archaeal, and Eukaryotic RNA Polymerases. *Annual Review of Microbiology*, 67(1), pp.113–139. Available at: http://www.annualreviews.org/doi/abs/10.1146/annurev-micro-092412-155756.

Dill, K.A. & Chan, H.S., 1997. From Levinthal to pathways to funnels. *Nature structural biology*, 4(1), pp.10–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8989315.

Dinner, A.R. et al., 2000. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends in biochemical sciences*, 25(7), pp.331–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10871884.

Dobson, C.M., 2003. Protein folding and misfolding. *Nature*, 426(6968), pp.884–90. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14685248.

Drozdetskiy, A. et al., 2015. JPred4: a protein secondary structure prediction server. *Nucleic Acids Research*, 43(W1), pp.W389–W394. Available at:

http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkv332.

Dynan, W.S. & Tjian, R., 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell*, 35(1), pp.79–87.

Ellis, R., 1987. Proteins as molecular chaperones. *Nature*, 328(6129), pp.378–379.

Emsley, P. & Cowtan, K., 2004. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography*, 60(Pt 12 Pt 1), pp.2126–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15572765.

Eswar, N. et al., 2007. Comparative protein structure modeling using MODELLER. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]*, Chapter 2, p.Unit 2.9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18429317.

Evans, P., 2006. Scaling and assessment of data quality. *Acta crystallographica. Section D, Biological crystallography*, 62(Pt 1), pp.72–82. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16369096.

Evans, P.R. & Murshudov, G.N., 2013. How good are my data and what is the resolution? *Acta crystallographica. Section D, Biological crystallography*, 69(Pt 7), pp.1204–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23793146.

Fang, F., Chang, R. & Yang, L., 2012. Heat shock factor 1 promotes invasion and metastasis of hepatocellular carcinoma in vitro and in vivo. *Cancer*, 118(7), pp.1782–1794.

Farkas, T., Kutskova, Y. a & Zimarino, V., 1998. Intramolecular repression of mouse heat shock factor 1. *Molecular and cellular biology*, 18(2), pp.906–918.

Fassler, J.S. & West, A.H., 2011. Fungal Skn7 Stress Responses and Their Relationship to Virulence. *Eukaryotic Cell*, 10(2), pp.156–167. Available at: http://ec.asm.org/cgi/doi/10.1128/EC.00245-10.

Ferbitz, L. et al., 2004. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature*, 431(7008), pp.590–596.

Available at: http://www.nature.com/doifinder/10.1038/nature02899.

Fiser, A. & Sali, A., 2003. ModLoop: automated modeling of loops in protein structures. *Bioinformatics (Oxford, England)*, 19(18), pp.2500–1. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14668246.

Flick, K.E. et al., 1994. Yeast heat shock transcription factor contains a flexible linker between the DNA-binding and trimerization domains. Implications for DNA binding by trimeric proteins. *J Biol Chem*, 269(17), pp.12475–12481. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8175654.

Fox, N.K., Brenner, S.E. & Chandonia, J.-M., 2014. SCOPe: Structural Classification of Proteins—extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Research*, 42(D1), pp.D304–D309. Available at: http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkt1240.

Fraenkel, E. et al., 1998. Engrailed homeodomain-DNA complex at 2.2 A resolution: a detailed view of the interface and comparison with other engrailed structures. *Journal of molecular biology*, 284(2), pp.351–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9813123.

French, S. & Wilson, K., 1978. On the treatment of negative intensity observations. *Acta Crystallographica Section A*, 34(4), pp.517–525. Available at: http://scripts.iucr.org/cgi-bin/paper?S0567739478001114.

Fujimoto, M. et al., 2012. RPA assists HSF1 access to nucleosomal DNA by recruiting histone chaperone FACT. *Molecular Cell*, 48(2), pp.182–194. Available at: http://dx.doi.org/10.1016/j.molcel.2012.07.026.

Georgescauld, F. et al., 2014. GroEL/ES chaperonin modulates the mechanism and accelerates the rate of TIM-barrel domain folding. *Cell*, 157(4), pp.922–934. Available at: http://dx.doi.org/10.1016/j.cell.2014.03.038.

Gómez-Pastor, R. et al., 2013. Trx2p-dependent regulation of Saccharomyces cerevisiae oxidative stress response by the Skn7p transcription factor under respiring conditions. *PLoS ONE*, 8(12), pp.1–13.

Götze, M. et al., 2015. Automated assignment of MS/MS cleavable cross-links in protein 3D-structure analysis. *Journal of the American Society for Mass*

Spectrometry, 26(1), pp.83–97. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25261217.

Grady, D.L. et al., 1992. Highly conserved repetitive DNA sequences are present at human centromeres. *Proceedings of the National Academy of Sciences of the United States of America*, 89(5), pp.1695–1699.

Green, M. et al., 1995. A heat shock-responsive domain of human HSF1 that regulates transcription activation domain function. *Molecular and Cellular Biology*, 15(6), pp.3354–3362. Available at: http://mcb.asm.org/lookup/doi/10.1128/MCB.15.6.3354.

Greenfield, N.J., 2006. Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols*, 1(6), pp.2876–90.

Grimm, M. et al., 2015. xVis: a web server for the schematic visualization and interpretation of crosslink-derived spatial restraints. *Nucleic acids research*, 43(W1), pp.W362–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25956653.

Guertin, M.J. & Lis, J.T., 2010. Chromatin landscape dictates HSF binding to target DNA elements. *PLoS genetics*, 6(9), p.e1001114. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2936546&tool=pmcentrez&rendertype=abstract.

Gupta, A.J. et al., 2014. Active cage mechanism of chaperonin-assisted protein folding demonstrated at single-molecule level. *Journal of molecular biology*, 426(15), pp.2739–54. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24816391.

Gusella, J.F. & MacDonald, M.E., 2006. Huntington's disease: seeing the pathogenic process through a genetic lens. *Trends in Biochemical Sciences*, 31(9), pp.533–540.

Harrison, C.J., Bohm, a a & Nelson, H.C., 1994. Crystal structure of the DNA binding domain of the heat shock transcription factor. *Science*, 263(5144), pp.224–227. Available at: http://www.sciencemag.org/cgi/content/abstract/263/5144/224.

Hartl, F.U., Bracher, A. & Hayer-Hartl, M., 2011. Molecular chaperones in protein

folding and proteostasis. *Nature*, 475(7356), pp.324–332.

Hartl, F.U. & Hayer-Hartl, M., 2009. Converging concepts of protein folding in vitro and in vivo. *Nature Structural & Molecular Biology*, 16(6), pp.574–581. Available at: http://www.nature.com/doifinder/10.1038/nsmb.1591.

Hauser, T. et al., 2015. Structure and mechanism of the Rubisco-assembly chaperone Raf1. *Nature Structural & Molecular Biology*, 22(9), pp.720–728. Available at: http://www.nature.com/doifinder/10.1038/nsmb.3062.

Hayer-Hartl, M., Bracher, A. & Hartl, F.U., 2016. The GroEL–GroES Chaperonin Machine: A Nano-Cage for Protein Folding. *Trends in Biochemical Sciences*, 41(1), pp.62–76. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0968000415001401.

Haynes, C.M. & Ron, D., 2010. The mitochondrial UPR - protecting organelle protein homeostasis. *Journal of Cell Science*, 123(22), pp.3849–3855. Available at: http://jcs.biologists.org/cgi/doi/10.1242/jcs.075119.

He, X.J., Mulford, K.E. & Fassler, J.S., 2009. Oxidative stress function of the saccharomyces cerevisiae skn7 receiver domain. *Eukaryotic Cell*, 8(5), pp.768–778.

Hentze, N. et al., 2016. Molecular mechanism of thermosensory function of human heat shock transcription factor Hsf1. , (January).

Hernandez Alvarez, B. et al., 2008. A new expression system for protein crystallization using trimeric coiled-coil adaptors. *Protein Engineering, Design and Selection*, 21(1), pp.11–18.

Hietakangas, V. et al., 2003. Phosphorylation of serine 303 is a prerequisite for the stress-inducible SUMO modification of heat shock factor 1. *Molecular and cellular biology*, 23(8), pp.2953–68. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=152542&tool=pmcentrez&rendertype=abstract.

Hietakangas, V. & Sistonen, L., 2006. Regulation of the heat shock response by heat shock transcription factors. *Chaperones*, pp.1–34.

Hipp, M.S., Park, S.H. & Hartl, F.U., 2014. Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends in Cell Biology*, (Figure 2),

pp.1–9. Available at: http://dx.doi.org/10.1016/j.tcb.2014.05.003.

Holtkamp, W. et al., 2015. Cotranslational protein folding on the ribosome monitored in real time. *Science (New York, N.Y.)*, 350(6264), pp.1104–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26612953.

Hsu, A.-L., Murphy, C.T. & Kenyon, C., 2003. Regulation of Aging and Age-Related Disease by DAF-16 and Heat-Shock Factor. *Science*, 300(5622), pp.1142–1145. Available at: http://www.sciencemag.org/content/300/5622/1142.abstract.

Hubl, S.T., Owens, J.C. & Nelson, H.C.M., 1994. Mutational analysis of the DNA-binding domain of yeast heat shock transcription factor. *Nature Structural Biology*, 1(9), pp.615–620. Available at: http://www.nature.com/nsmb/journal/v1/n9/abs/nsb0994-615.html\nhttp://www.nature.com/doifinder/10.1038/nsb0994-615.

Jaeger, A.M. et al., 2014. Genomic heat shock element sequences drive cooperative human heat shock factor 1 DNA binding and selectivity. *The Journal of biological chemistry*, 289(44), pp.30459–69. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25204655.

Jaeger, A.M. et al., 2016. Structures of HSF2 reveal mechanisms for differential regulation of human heat-shock factors. *Nature Structural & Molecular Biology*, (August 2015). Available at: http://dx.doi.org/10.1038/nsmb.3150.

Jonkers, I. & Lis, J.T., 2015. Getting up to speed with transcription elongation by RNA polymerase II. *Nature reviews. Molecular cell biology*, 16(3), pp.167–177. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25693130.

Kabsch, W., 2010. XDS. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 2), pp.125–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20124692.

Kampinga, H.H. & Craig, E.A., 2010. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nature reviews. Molecular cell biology*, 11(8), pp.579–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20651708.

Kaushik, S. & Cuervo, A.M., 2012. Chaperone-mediated autophagy: A unique way

to enter the lysosome world. *Trends in Cell Biology*, 22(8), pp.407–417. Available at: http://dx.doi.org/10.1016/j.tcb.2012.05.006.

Kim, Y.E. et al., 2013. *Molecular Chaperone Functions in Protein Folding and Proteostasis*, Available at: http://www.annualreviews.org/doi/abs/10.1146/annurev-biochem-060208-092442.

Kleywegt, G.J., 1999. Experimental assessment of differences between related protein crystal structures. *Acta Crystallographica Section D Biological Crystallography*, 55(11), pp.1878–1884. Available at: http://scripts.iucr.org/cgi-bin/paper?S0907444999010495.

Knauf, U. et al., 1996. Repression of human heat shock factor 1 activity at control temperature by phosphorylation. *Genes & development*, 10(21), pp.2782–2793. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8946918.

Kosinski, J. et al., 2015. Xlink Analyzer : Software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *Journal of Structural Biology*, 189(3), pp.177–183. Available at: http://dx.doi.org/10.1016/j.jsb.2015.01.014.

Koth, C.M. et al., 2003. Use of Limited Proteolysis to Identify Protein Domains Suitable for Structural Analysis. In pp. 77–84. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0076687903680055.

Kourtis, N. & Tavernarakis, N., 2011. Cellular stress response pathways and ageing: intricate molecular relationships. *The EMBO Journal*, 30(13), pp.2520–2531. Available at: http://emboj.embopress.org/cgi/doi/10.1038/emboj.2011.162.

Krylov, D. & Vinson, C.R., 2001. Leucine zipper. *Encyclopedia of Life Sciences*, pp.1–7.

Labbadia, J. & Morimoto, R., 2015. The Biology of Proteostasis in Aging and Disease. *Annual Review of Biochemistry*, 84(1), pp.435–464. Available at: http://www.annualreviews.org/doi/abs/10.1146/annurev-biochem-060614-033955.

Labbadia, J. & Morimoto, R.I., 2015. Repression of the Heat Shock Response Is

a Programmed Event at the Onset of Reproduction. *Molecular Cell*, pp.1–12. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1097276515004992.

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(5259), pp.680–5.

Langer, G. et al., 2008. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature Protocols*, 3(7), pp.1171–1179. Available at: http://www.nature.com/doifinder/10.1038/nprot.2008.91.

Langer, T. et al., 1992. Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature*, 356(6371), pp.683–689. Available at: http://www.nature.com/doifinder/10.1038/356683a0.

Lee, J. et al., 1999. Yap1 and Skn7 control two specialized oxidative stress response regulons in yeast. *Journal of Biological Chemistry*, 274(23), pp.16040–16046.

Leitner, A. et al., 2014. Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 111(26), pp.9455–60. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4084482&tool=pmcentrez&rendertype=abstract.

Leitner, A. et al., 2012. The Molecular Architecture of the Eukaryotic Chaperonin TRiC/CCT. *Structure*, 20(5), pp.814–825. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0969212612001086.

Leitner, A., Walzthoeni, T. & Aebersold, R., 2013. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nature Protocols*, 9(1), pp.120–137. Available at: http://www.nature.com/doifinder/10.1038/nprot.2013.168.

Li, J., Soroka, J. & Buchner, J., 2012. The Hsp90 chaperone machinery: Conformational dynamics and regulation by co-chaperones. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1823(3), pp.624–635. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0167488911002618.

Li, S. et al., 2002. The eukaryotic two-component histidine kinase Sln1p regulates OCH1 via the transcription factor, Skn7p. *Molecular biology of the cell*, 13(2), pp.412–424.

Li, S. et al., 1998. The yeast histidine protein kinase, Sln1p, mediates phosphotransfer to two response regulators, Ssk1p and Skn7p. *The EMBO journal*, 17(23), pp.6952–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9843501.

Lindquist, S., 1986. THE HEAT-SHOCK RESPONSE.

Littlefield, O. & Nelson, H.C., 1999. A new use for the “wing” of the “winged” helix-turn-helix motif in the HSF-DNA cocrystal. *Nature structural biology*, 6(5), pp.464–470.

Liu, G. et al., 2011. Northeast Structural Genomics Consortium Target HR3023C. *To be Published*. Available at: http://www.rcsb.org/pdb/explore/explore.do?structureId=2LDU [Accessed January 19, 2016].

Lopez, T., Dalton, K. & Frydman, J., 2015. The Mechanism and Function of Group II Chaperonins. *Journal of Molecular Biology*, 427(18), pp.2919–2930. Available at: http://dx.doi.org/10.1016/j.jmb.2015.04.013.

Luscombe, N.M. et al., 2000. An overview of the structures of protein-DNA complexes. *Genome biology*, 1(1), p.REVIEWS001.

Mangani, S., 2013. *Disruption of Protein-Protein Interfaces* S. Mangani, ed., Berlin, Heidelberg: Springer Berlin Heidelberg. Available at: http://link.springer.com/10.1007/978-3-642-37999-4.

Mason, J.M. & Arndt, K.M., 2004. Coiled coil domains: Stability, specificity, and biological implications. *ChemBioChem*, 5(2), pp.170–176.

Mayer, M.P. & Bukau, B., 2005. Hsp70 chaperones: cellular functions and molecular mechanism. *Cellular and molecular life sciences : CMLS*, 62(6), pp.670–84. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15770419.

McDonough, H. & Patterson, C., 2003. CHIP: a link between the chaperone and proteasome systems. *Cell stress & chaperones*, 8(4), pp.303–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15115282.

Mendillo, M.L. et al., 2012. HSF1 drives a transcriptional program distinct from heat shock to support highly malignant human cancers. *Cell*, 150(3), pp.549–562.

Mercier, P.A., Winegarden, N.A. & Westwood, J.T., 1999. Human heat shock factor 1 is predominantly a nuclear protein before and after heat stress. *Journal of cell science*, 112 (Pt 1, pp.2765–74. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10413683.

Morgan, B. a et al., 1997. The Skn7 response regulator controls gene expression in the oxidative stress response of the budding yeast Saccharomyces cerevisiae. *The EMBO journal*, 16(5), pp.1035–1044.

Morimoto, R.I., 2008. Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. *Genes and Development*, 22(11), pp.1427–1438.

Murshudov, G.N., Vagin, A.A. & Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta crystallographica. Section D, Biological crystallography*, 53(Pt 3), pp.240–55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15299926.

Neef et al., 2014. A direct regulatory interaction between chaperonin TRiC and stress-responsive transcription factor HSF1. *Cell reports*, 9(3), pp.955–66. Available at: http://www.sciencedirect.com/science/article/pii/S2211124714008390.

Neef, D.W., Jaeger, a M. & Thiele, D.J., 2011. Heat shock transcription factor 1 as a therapeutic target in neurodegenerative diseases. *Nat Rev Drug Discov*, 10(12), pp.930–944. Available at: http://www.nature.com/nrd/journal/v10/n12/pdf/nrd3453.pdf.

Neef, D.W., Jaeger, A.M. & Thiele, D.J., 2013. Genetic selection for constitutively trimerized human HSF1 mutants identifies a role for coiled-coil motifs in DNA binding. *G3 (Bethesda, Md.)*, 3(8), pp.1315–24. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3737171&tool=pmcentrez&rendertype=abstract.

Neef, D.W., Turski, M.L. & Thiele, D.J., 2010. Modulation of heat shock transcription factor 1 as a therapeutic target for small molecule intervention in

neurodegenerative disease. *PLoS biology*, 8(1), p.e1000291. Available at: http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1000291.

Neidhardt, F.C. & VanBogelen, R.A., 1981. Positive regulatory gene for temperature-controlled proteins in Escherichia coli. *Biochemical and biophysical research communications*, 100(2), pp.894–900. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7023474.

Nero, T.L. et al., 2014. Oncogenic protein interfaces : small molecules , big challenges. *Nature Publishing Group*, 14(4), pp.248–262. Available at: http://dx.doi.org/10.1038/nrc3690.

Neudegger, T., 2012. *Structural and functional characterization of human Hsf1*. Technical University of Munich.

Neudegger, T. et al., 2016. Structure of human heat-shock transcription factor 1 in complex with DNA. *Nature Structural & Molecular Biology*, 23(2), pp.140–146. Available at: http://dx.doi.org/10.1038/nsmb.3149.

Nilsson, O.B. et al., 2015. Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell reports*, 12(10), pp.1533–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26321634.

Nonaka, G. et al., 2006. Regulon and promoter analysis of the. *Genes & Development*, pp.1776–1789.

Ostermann, J. et al., 1989. Protein folding in mitochondria requires complex formation with hsp60 and ATP hydrolysis. *Nature*, 341(6238), pp.125–130. Available at: http://www.nature.com/doifinder/10.1038/341125a0.

Otto, H. et al., 2005. The chaperones MPP11 and Hsp70L1 form the mammalian ribosome-associated complex. *Proceedings of the National Academy of Sciences*, 102(29), pp.10064–10069. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.0504400102.

Pace, C.N., Fisher, L.M. & Cupo, J.F., 1981. Globular protein stability: aspects of interest in protein turnover. *Acta biologica et medica Germanica*, 40(10-11), pp.1385–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6282021.

Page, T.J. et al., 2006. Genome-wide analysis of human HSF1 signaling reveals a transcriptional program linked to cellular adaptation and survival. *Molecular*

BioSystems, 2(12), p.627. Available at: http://xlink.rsc.org/?DOI=b606129j.

Parker, C.S. & Topol, J., 1984. A Drosophila RNA polymerase II transcription factor binds to the regulatory site of an hsp 70 gene. *Cell*, 37(1), pp.273–283.

Pattaramanon, N., Sangha, N. & Gafni, A., 2007. The carboxy-terminal domain of heat-shock factor 1 is largely unfolded but can be induced to collapse into a compact, partially structured state. *Biochemistry*, 46, pp.3405–3415.

Pelham, H.R., 1982. A regulatory upstream promoter element in the Drosophila hsp 70 heat-shock gene. *Cell*, 30(2), pp.517–528.

Peteranderl, R. et al., 1999. Biochemical and biophysical characterization of the trimerization domain from the heat shock transcription factor. *Biochemistry*, 38(12), pp.3559–69. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10090742.

Peteranderl, R. & Nelson, H.C., 1992. Trimerization of the heat shock transcription factor by a triple-stranded alpha-helical coiled-coil. *Biochemistry*, 31(48), pp.12272–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10090742.

Petoukhov, M. V. et al., 2012. New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography*, 45(2), pp.342–350. Available at: http://scripts.iucr.org/cgi-bin/paper?S0021889812007662.

Pettersen, E.F. et al., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13), pp.1605–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15264254.

Piskacek, M., 2009. Common Transactivation Motif 9aaTAD recruits multiple general co-activators TAF9, MED15, CBP and p300. *Nature Precedings*. Available at: http://precedings.nature.com/documents/3488/version/2.

Powers, E.T. et al., 2009. Biological and Chemical Approaches to Diseases of Proteostasis Deficiency. *Annual Review of Biochemistry*, 78(1), pp.959–991. Available at: http://www.annualreviews.org/doi/abs/10.1146/annurev.biochem.052308.114844.

Preissler, S. & Deuerling, E., 2012. Ribosome-associated chaperones as key

players in proteostasis. *Trends in biochemical sciences*, 37(7), pp.274–83. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22503700.

Rabindran, S.K. et al., 1991. Molecular cloning and expression of a human heat shock factor, HSF1. *Proceedings of the National Academy of Sciences of the United States of America*, 88(16), pp.6906–6910.

Rabindran, S.K. et al., 1993. Regulation of heat shock factor trimer formation: role of a conserved leucine zipper. *Science (New York, N.Y.)*, 259(5092), pp.230–234.

Radons, J., 2016. The human HSP70 family of chaperones: where do we stand? *Cell Stress and Chaperones*. Available at: http://link.springer.com/10.1007/s12192-016-0676-6.

Raitt, D.C. et al., 2000. The Skn7 response regulator of Saccharomyces cerevisiae interacts with Hsf1 in vivo and is required for the induction of heat shock genes by oxidative stress. *Molecular biology of the cell*, 11(7), pp.2335–2347.

Raychaudhuri, S. et al., 2014. Interplay of acetyltransferase EP300 and the proteasome system in regulating heat shock transcription factor 1. *Cell*, 156(5), pp.975–985.

Read, R.J. et al., 2011. A New Generation of Crystallographic Validation Tools for the Protein Data Bank. *Structure*, 19(10), pp.1395–1412. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0969212611002851.

Rehn, A.B. & Buchner, J., 2015. p23 and Aha1. In pp. 113–131. Available at: http://link.springer.com/10.1007/978-3-319-11731-7_6.

Ritossa, F., 1962. A new puffing pattern induced by temperature shock and DNP in drosophila. *Experientia*, 18(12), pp.571–573.

Robert, X. & Gouet, P., 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, 42(W1), pp.W320–W324. Available at: http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gku316.

Rohs, R. et al., 2010. Origins of Specificity in Protein-DNA Recognition. *Annual Review of Biochemistry*, 79(1), pp.233–269. Available at: http://www.annualreviews.org/doi/abs/10.1146/annurev-biochem-060408-

091030.

Ron, D. & Walter, P., 2007. Signal integration in the endoplasmic reticulum unfolded protein response. *Nat Rev Mol Cell Biol*, 8(7), pp.519–529. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17565364.

Ross, C. a & Poirier, M. a, 2004. Protein aggregation and neurodegenerative disease. *Nature Medicine*, 10(7), pp.S10–S17. Available at: http://www.nature.com/doifinder/10.1038/nm1066.

Rougvie, A.E. & Lis, J.T., 1988. The RNA polymerase II molecule at the 5′ end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. *Cell*, 54(6), pp.795–804. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092867488910872.

Sainsbury, S., Bernecky, C. & Cramer, P., 2015. Structural basis of transcription initiation by RNA polymerase II. *Nature reviews. Molecular cell biology*, 16(3), pp.129–143. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25693126.

Sandqvist, A. et al., 2009. Heterotrimerization of heat-shock factors 1 and 2 provides a transcriptional switch in response to distinct stimuli. *Molecular biology of the cell*, 20(5), pp.1340–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19129477.

Santagata, S. et al., 2011. High levels of nuclear heat-shock factor 1 (HSF1) are associated with poor prognosis in breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 108(45), pp.18378–18383. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3215027&tool=pmcentrez&rendertype=abstract.

Sarge, K.D., Murphy, S.P. & Morimoto, R.I., 1993. Activation of heat shock gene transcription by heat shock factor 1 involves oligomerization, acquisition of DNA-binding activity, and nuclear localization and can occur in the absence of stress. *Molecular and cellular biology*, 13(3), pp.1392–1407.

Scherzinger, E. et al., 1999. Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: Implications for Huntington's disease pathology. *Proceedings of the National Academy of Sciences*, 96(8), pp.4604–4609. Available at:

http://www.pnas.org/cgi/doi/10.1073/pnas.96.8.4604.

Scherz-Shouval, R. et al., 2014. The Reprogramming of Tumor Stroma by HSF1 Is a Potent Enabler of Malignancy. *Cell*, 158(3), pp.564–578. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0092867414008034.

Scior, A. & Deuerling, E., 2014. Functions of Ribosome-Associated Chaperones and their Interaction Network. , pp.27–35.

Sheldrick, G.M., 2010. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 4), pp.479–85. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20383001.

Sistonen, L., Sarge, K.D. & Morimoto, R.I., 1994. Human heat shock factors 1 and 2 are differentially activated and can synergistically induce hsp70 gene transcription. *Molecular and cellular biology*, 14(3), pp.2087–99. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=358569&tool=pmcentrez&rendertype=abstract.

Sorger, P.K. & Nelson, H.C., 1989. Trimerization of a yeast transcriptional activator via a coiled-coil motif. *Cell*, 59(5), pp.807–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2686840.

Strauch, M. a, 2001. Protein – DNA Complexes : Specific. *Encyclopedia of life sciences*, pp.1–7. Available at: www.els.net.

Teague, S.J., 2003. Implications of protein flexibility for drug discovery. *Nature Reviews Drug Discovery*, 2(7), pp.527–541. Available at: http://www.nature.com/doifinder/10.1038/nrd1129.

Tissières, A., Mitchell, H.K. & Tracy, U.M., 1974. Protein synthesis in salivary glands of Drosophila melanogaster: relation to chromosome puffs. *Journal of molecular biology*, 84(3), pp.389–98. Available at: http://www.ncbi.nlm.nih.gov/pubmed/4219221.

la Touche, G.J., 1948. A Chætomium-like Thermophile Fungus. *Nature*, 161(4087), pp.320–320. Available at: http://www.nature.com/doifinder/10.1038/161320a0.

Treisman, J. et al., 1989. A single amino acid can determine the DNA binding

specificity of homeodomain proteins. *Cell*, 59(3), pp.553–62. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2572327.

Trinklein, N.D. et al., 2004. The role of heat shock transcription factor 1 in the genome-wide regulation of the mammalian heat shock response. *Molecular biology of the cell*, 15(3), pp.1254–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14668476.

Tucker-Kellogg, L. et al., 1997. Engrailed (Gln50→Lys) homeodomain–DNA complex at 1.9 Å resolution: structural basis for enhanced affinity and altered specificity. *Structure*, 5(8), pp.1047–1054. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0969212697002566.

Uhlén, M. et al., 2015. Proteomics. Tissue-based map of the human proteome. *Science (New York, N.Y.)*, 347(6220), p.1260419. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25613900.

Vagin, A.A. & Isupov, M.N., 2001. Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta crystallographica. Section D, Biological crystallography*, 57(Pt 10), pp.1451–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11567159.

Vaquerizas, J.M. et al., 2009. A census of human transcription factors: function, expression and evolution. *Nature reviews. Genetics*, 10(4), pp.252–263.

Vembar, S.S. & Brodsky, J.L., 2008. One step at a time: endoplasmic reticulum-associated degradation. *Nature reviews. Molecular cell biology*, 9(12), pp.944–57. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19002207.

Venkatesh, S. & Workman, J.L., 2015. Histone exchange , chromatin structure and the regulation of transcription E i V. *Nature Publishing Group*, 16(3), pp.178–189. Available at: http://dx.doi.org/10.1038/nrm3941.

Vihervaara, A. et al., 2013. Transcriptional response to stress in the dynamic chromatin environment of cycling and mitotic cells. *Proceedings of the National Academy of Sciences*, 110(36), pp.E3388–E3397. Available at: http://www.pnas.org/lookup/doi/10.1073/pnas.1305275110.

Vuister, G.W. et al., 1994. Solution structure of the DNA-binding domain of

Drosophila heat shock transcription factor. *Nature structural biology*, 1(9), pp.605–14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7634100.

Walker, L.C. & LeVine, H., 2000. The cerebral proteopathies. *Neurobiology of Aging*, 21(4), pp.559–561. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0197458000001603.

Walter, P. & Ron, D., 2011. The Unfolded Protein Response: From Stress Pathway to Homeostatic Regulation. *Science*, 334(6059), pp.1081–1086.

Wang, S., Sakai, H. & Wiedmann, M., 1995. NAC covers ribosome-associated nascent chains thereby forming a protective environment for regions of nascent chains just emerging from the peptidyl transferase center. *The Journal of Cell Biology*, 130(3), pp.519–528. Available at: http://www.jcb.org/cgi/doi/10.1083/jcb.130.3.519.

Warrick, J.M. et al., 1999. Suppression of polyglutamine-mediated neurodegeneration in Drosophila by the molecular chaperone HSP70. *Nature genetics*, 23(4), pp.425–428.

Watson, J.D. & Crick, F.H.C., 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, 171(4356), pp.737–738. Available at: http://www.nature.com/doifinder/10.1038/171737a0.

Weiss, S.B. & Gladstone, L., 1959. A MAMMALIAN SYSTEM FOR THE INCORPORATION OF CYTIDINE TRIPHOSPHATE INTO RIBONUCLEIC ACID 1. *Journal of the American Chemical Society*, 81(15), pp.4118–4119. Available at: http://pubs.acs.org/doi/abs/10.1021/ja01524a087.

Westerheide, S.D. et al., 2009. Stress-Inducible Regulation of Heat Shock Factor 1 by the Deacetylase SIRT1. *Science*, 323(5917), pp.1063–1066. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.1165946.

Whitesell, L. & Lindquist, S., 2009. Inhibiting the transcription factor HSF1 as an anticancer strategy. *Expert opinion on therapeutic targets*, 13(4), pp.469–478.

Wold, M.S., 1997. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annual review of biochemistry*, 66, pp.61–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9242902.

Wu, C., 1995. HEAT SHOCK TRANSCRIPTION FACTORS : Structure and Regulation translocation , higher order assembly , and protein degradation (Gething &. *Annual Review of Cell and Developmental Biology*, 11, pp.441–469.

Wu, C., 1984. Two protein-binding sites in chromatin implicated in the activation of heat-shock genes. *Nature*, 309(5965), pp.229–34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6325944.

Xiao, H., Perisic, O. & Lis, J.T., 1991. Cooperative binding of Drosophila heat shock factor to arrays of a conserved 5 bp unit. *Cell*, 64(3), pp.585–593.

Xu, Y. et al., 2012. Post-Translational Modi fi cation of Human Heat Shock Factors and Their Functions: A Recent Update by Proteomic Approach. *Journal of proteome research*, 11, pp.2625–2634.

Yoshida, H. et al., 2001. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell*, 107(7), pp.881–91. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11779464.

Zimmerman, S.B. & Trach, S.O., 1991. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *Journal of molecular biology*, 222(3), pp.599–620. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1748995.

Zou, J. et al., 1998. Repression of heat shock transcription factor HSF1 activation by HSP90 (HSP90 complex) that forms a stress-sensitive complex with HSF1. *Cell*, 94(4), pp.471–480.

Zuiderweg, E.R.P. et al., 2013. Allostery in the Hsp70 chaperone proteins. *Topics in current chemistry*, 328, pp.99–153. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22576356.

Zuo, J. et al., 1994. Activation of the DNA-binding ability of human heat shock transcription factor 1 may involve the transition from an intramolecular to an intermolecular triple-stranded coiled-coil structure. *Molecular and cellular biology*, 14(11), pp.7557–68. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7935471.

Zuo, J., Rungger, D. & Voellmy, R., 1995. Multiple layers of regulation of human heat shock transcription factor 1. *Molecular and cellular biology*, 15(8), pp.4319–4330.

9 Appendix

9.1 Sequence alignment of the conserved region in Hsf proteins

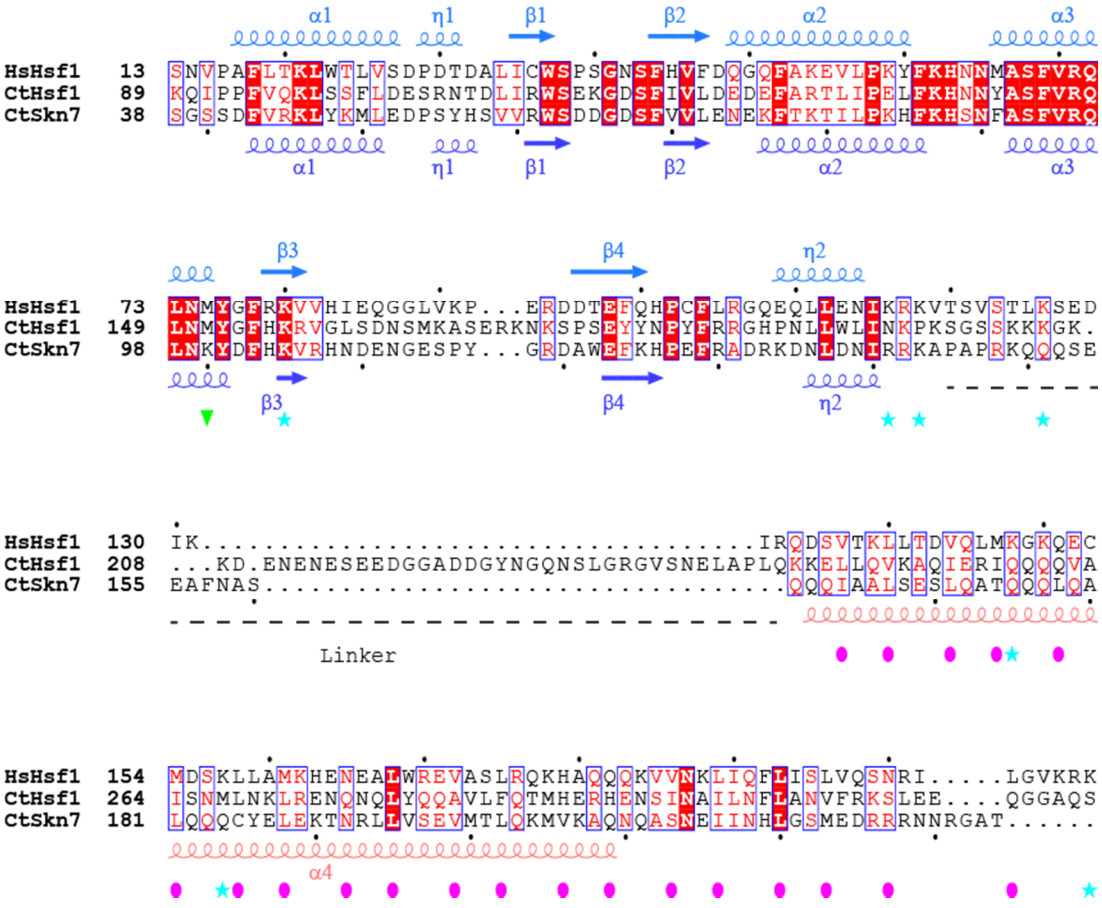

Figure 64 Multiple sequence alignment of N-terminal HsHsf1, CtHsf1, CtSkn7. Similar residues are shown in red, identical residues are shown in white using bold lettering on red background. Blue frames indicate homologous regions. Green arrowhead systematic Hsf1/Skn7 substitution. Asterisks denote acetylation sites in HsHSF1; pink ovals indicate hydrophobic layer residues in HsHSF1 coiled-coil.

9.2 Accession codes for coordinates in PDB

HsHsf1 DBD–HSE complex	5D5U
HsHsf1 DBD–SatIII complex	5D5V
CtSkn7 DBD–HSE complex	5D5W
CtSkn7 DBD–SSRE complex	5D5X

CtSkn7(160–209) crystal form I	5D5Y
CtSkn7(160–209) crystal form II	5D5Z
CtSkn7(160–220)	5D60

9.3 Primer for PCR

Alias	**Sequence**
pHUE_pProEXHtb_Combi_prefix_fwd	cgcGAATTCtCCGCGGtgga
pHUE_pProEXHtb_Combi_suffix_rev	CCCaagcttA
Hsf1_1-fwd	cgcGAATTCtCCGCGGtggaATGGATCTGCCCGTGGGCCC
Hsf1-529-rev	CCCaagcttAGGAGACAGTGGGGTCCTTGGC
Hsf1-120-rev	CCCaagcttAGGTCACTTTCCTCTTGATGTTCTCAAGGAGC
1-192-GCN4-rev	CCCaagcttAAAGCTTATTAAATCAGTTTTTTAATACGCGCAATTTCGTTTTC
Hsf1_133-fwd	cgcGAATTCtCCGCGGtggaCGCCAGGACAGCGTCACCAAG
Hsf1-182-rev	CCCaagcttATTGCTGGGCATGCTTCTGCCGAAG
CtSkn7_fwd_40-	cgcGAATTCtCCGCGGtggaAGCAGCGATTTTGTTCGTAAACTGTATAAAATGCT
CtSkn7-209-rev	CCCaagcttACTGTGCTTTAACCATTTTCTGCAGGGTCA
CtSkn7_160-fwd	cgcGAATTCtCCGCGGtggaAGCCAGCAGCAAATTGCAGCACTGA
CtSkn7-143-rev	CCCaagcttATTTACGACGAATGTTATCCAGATTATCTTTGCGAT
CtSkn7-209_elong-220-rev	CCCaagcttACAGATGGTTAATAATTTCGTTGCTCGCCTGGTTCTGTGCTTTAACCATTTTCTGCAGGGTCA
CtSkn7_35-fwd	cgcGAATTCtCCGCGGtggaGGTGGTAGCGGTAGCAGCGATTTT
CtSkn7-765-rev	CCCaagcttACTGAACAAAACCTGCAACGCCAACAC
HsHsf1_Chimera_1-fwd	cgcGAATTCtCCGCGGtggaATGGATCTGCCTGTGGGACCTGG
HsHsf1_Chimera_529-rev	CCCaagcttAGGACACGGTAGGATCTTTGGCCTTG
triGCN4-rev	CCCaagcttAAAGCTTATTAAATCAGTTTTTTAATACGCGCAATTTCGTTTTC
CtSkn7_K100M-fwd	AGCTTTGTTCGCCAGCTGAACatgTATGATTTTCATAAAGTGCGC
CtSkn7_K100M-rev	GCGCACTTTATGAAAATCATAcatGTTCAGCTGGCGAACAAAGCT
CtSkn7_K143Q-fwd	GATAACATTCGTCGTcagGCACCGGCACCGCGT
CtSkn7_K143Q-rev	ACGCGGTGCCGGTGCctgACGACGAATGTTATC

9.4 DNA-oligonucleotides for protein binding experiments

Alias	Sequence
SSREx2	cgacagGGCtaGCCacagacg
SSRE	**ATTTGGCTGGGCC**
SSREx4	cgacGGCagGCCtaGGCgaGCCgacg
OCH1pr	ccactATTTGGCCGGCCCaccgcgaaaagATTTGGCTGGGCCtcacc
SatIII	**cGGAATGGAATg**
HSE	**ggTTCtaGAAcc**
HSEx2	cgacagTTCtaGAAacagacg
HSEx3	cgacGAAagTTCtaGAAgacg
HSEx4	cgacGAAagTTCtaGAAgaTTCgacg
Ctrl	cgacagatacagatacagacg
Ctrl'	cgacagataggcatacagacg

DNA-oligonucleotides depicted in **bold** were used for crystallization and ordered without fluorescein tag.

9.5 Crosslinks identified by mass spectrometry

Crosslinks were assigned as robust if found in two experimental conditions and in at least 60% of analyzed gel bands (Figure 17).

9.5.1 DSS crosslinks in HsHsf1 identified by mass spectrometry

Table 5 DSS crosslinks identified in HsHsf1 by mass spectrometry.

No heat shock			
Pos1	**Pos2**	**Occurance 2/2 analyzed bands**	**Distance in crystal structure (Å)**
1	59	-	
1	62	-	
1	118	Y	
1	126	-	
1	131	-	
1	184	Y	
1	206	-	

Heat shock				
Pos1	**Pos2**	**Inter subunit**	**Occurance in 3/5 analyzed bands**	**Distance in crystal structure (Å)**
1	59	-	Y	
1	62	-	Y	
1	91	-	Y	
1	116	-	Y	
1	118	-	Y	
1	126	-	Y	
1	131	-	Y	

80	118	Y	20
80	126	Y	
80	131	-	
80	139	Y	
80	150	-	
80	184	Y	
80	206	Y	
80	298	Y	
91	184	-	
116	118	-	6
116	126	-	
116	139	-	
116	184	-	
116	298	Y	
118	131	-	
118	139	Y	
118	150	-	
118	184	-	
118	188	Y	
118	298	Y	
118	524	Y	
126	139	Y	
126	188	-	
131	206	-	
139	206	-	
150	298	-	
162	298	-	
184	298	-	
188	298	Y	
206	298	-	

1	148	-	Y	
1	206	-	Y	
1	208	-	-	
1	298	-	Y	
1	524	-	-	
59	91	-	Y	29
62	91	-	-	37
80	116	-	-	17
80	118	-	Y	20
80	206	-	-	
80	524	-	-	
116	118	-	Y	6
116	126	-	-	
116	131	-	Y	
118	118	Y	-	
118	126	Y	Y	
118	131	-	Y	
126	126	Y	Y	
126	131	Y	Y	
126	139	-	-	
131	131	Y	-	
131	148	-	-	
131	139	-	-	
131	206	-	-	
162	298	-	-	
178	184	Y	Y	
184	298	-	Y	
184	524	-	-	
188	206	-	Y	
188	208	-	-	
188	298	-	-	
206	206	Y	Y	
206	208	-	Y	
206	224	-	-	
206	298	-	Y	
208	208	Y	-	
208	298	-	Y	
224	298	-	-	
298	298	Y	Y	
298	524	-	Y	

9.5.2 DMTMM crosslinks in HsHsf1 identified by mass spectrometry

Table 6 DMTMM crosslinks identified in HsHsf1 by mass spectrometry

No heat shock			
Pos1	**Pos2**	**Occurance in 2/2 analyzed bands**	**Distance in crystal structure (Å)**
1	55	Y	
1	109	Y	
1	113	Y	
1	311	-	
2	116	-	
59	85	-	29
62	85	-	31
113	118	-	12
113	126	-	
131	135	Y	
148	339	-	
162	337	-	
171	178	-	
171	188	-	
184	339	-	
188	337	-	
188	339	-	
206	339	Y	
208	339	-	

Heat shock				
Pos1	**Pos2**	**Inter subunit**	**Occurance in 3/4 analyzed bands**	**Distance in crystal structure (Å)**
1	2	Y	-	
1	55	-	Y	
1	85	-	-	
1	109	-	Y	
1	113	-	Y	
1	128	-	Y	
2	91	-	-	
55	62	-	-	11
55	91	-	Y	25
91	93	-	-	7
118	128	-	-	
126	128	Y	Y	
126	129	Y	-	
128	131	Y	-	
129	131	Y	-	
162	164	Y	-	
162	166	Y	-	
162	525	-	-	
178	337	-	-	
178	339	-	-	
184	311	-	-	
184	312	-	-	
184	339	-	-	
188	339	-	-	
206	311	-	-	
206	322	-	-	
206	337	-	-	
206	339	-	-	
208	337	-	-	
208	339	-	-	
224	322	-	-	
298	322	-	-	
298	339	-	-	