PhD Thesis

Analysis of Structure and Function of the Intrinsically Disordered Dehydrin ERD14

Bianka Ildikó DR. SZALAINÉ ÁGOSTON

Supervisors

Prof. Dr. András PERCZEL

and

Dr. Péter TOMPA

Chemistry Doctoral School, Doctoral Program of

Synthetic Chemistry, Materials Science and Biomolecular Chemistry

Head of the Doctoral School: Prof. Dr. György INZELT

Head of the Doctoral Program: Prof. Dr. András PERCZEL

Department of Organic Chemistry, Institute of Chemistry, Eötvös Loránd University

and

Laboratory of Intrinsically Disordered Proteins, Institute of Enzymology,

Research Centre for Natural Sciences, Hungarian Academy of Sciences

Budapest, 2013

To my family.

Nem akkor nyered el jutalmad, ha belekezdesz valamibe,

hanem egyedül és csak is akkor, ha kitartasz mellette.

Sienai Szent Katalin

You will not get rewarded at the time you engage into something,

but only if you endure and hold on to it.

Saint Catherine of Siena

Contents

Acknowledgements

First of all I would like to thank my supervisors, *Dr. Péter Tompa* and *Dr. András Perczel* for their precious guidance and support at all stages of my work, as well as their openness to collaboration, without which the aims of this work could not have been envisioned, and the presented results could not have been reached.

I also thank all my colleagues within both research groups for their everyday assistance and a lot of valuable discussions, especially *Dr. Dénes Kovács*, who originally started to work with ERD14.

Furthermore, the financial support of the Foundation for the Hungarian Peptide and Protein Research in form of a scholarship, as well as the support of the Access to Research Infrastructures activity in the 6th Framework Programme of the EC (Contract # RII3-026145, EU-NMR) in the form of transnational access to high-field NMR spectrometers are gratefully acknowledged.

Last, but not least, I would like to express my special gratitude towards my family, especially my husband *Tamás*, for both his patience and everyday encouragement, and my daughter *Zsófia*, as well as her grandparents, who assisted a lot through their tireless babysitting during the seemingly endless times of writing.

Abbreviations and Acronyms

ABA	Abscisic Acid
Abs_{600}	Absorbance/optical density at 600 nm and 200 µL volume in a 96 well plate
AFM	Atomic Force Microscopy
ATP	Adenosine-5'-triphosphate
BMRB	Biological Magnetic Resonance Bank (www.bmrb.wisc.edu)
CARA	Computer Aided Resonance Assignment program
CD	Circular Dichroism (Spectroscopy)
CFU	Colony Forming Units
ChP	Charge-Peptide (Segment)
COSY	Correlation Spectroscopy
CS	Chemical Shift
CSI	Chemical Shift Index
%CV	Relative Cell Viability, given in percent
DisProt	Database of Protein Disorder (http://www.disprot.org)
DLS	Dynamic Light Scattering
DSC	Differential Scanning Calorimetry
DSS	4,4-Dimethyl-4-silapentane-1-sulfonic acid
EDTA	Ethylenediaminetetraacetic acid
EM	Electron Microscopy
EPR	Electron Paramagnetic Resonance
ERD14	Early Response to Dehydration 14
ESI	Electrospray Ionization
ESR	Electron Spin Resonance
FCS	Fluorescence Correlation Spectroscopy
FRET	Fluorescence Resonance Energy Transfer
FT-IR	Fourier-Transform Infrared (Spectroscopy)
GST	Glutathione S-Transferase
hetNOE	Heteronuclear ^{1}H-^{15}N NOE
HSP	Heat Shock Protein
HSQC	Heteronuclear Single Quantum Coherence
IDP/IDR	Intrinsically Disordered Protein/Region

IPTG	Isopropyl β-D-1-thiogalactopyranoside
LEA	Late Embryogenesis Abundant
MES	2-(*N*-Morpholino)ethanesulfonic acid
MLP	Muscular LIM protein
MOPS	3-Morpholinopropane-1-sulfonic acid
MoRFs	Molecular Recognition Features
MS	Mass Spectrometry
ncSPC	Neighbor Corrected Structural Propensity Calculator
NMD	Nonsense-Mediated Decay
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
O/N	Over-Night (typically a bacterial growth of about 16 hours)
OD_{600}	Optical Density at 600 nm and 1 cm path length
PDB	Protein Data Bank (www.rcsb.org)
PPII helix	Polyproline II helix
PRE	Paramagnetic Relaxation Enhancement
PSEs	Preformed Structural Elements
RDCs	Residual Dipolar Couplings
refDB	Re-referenced Protein Chemical shift Database (http://refdb.wishartlab.com)
ROA	Raman Optical Activity
ROS	Reactive Oxygen Species
RT	Room Temperature
SAXS	Small Angle X-Ray Scattering
SCS	Secondary Chemical Shift
SD	Standard Deviation
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate – Polyacrylamide Gel Electrophoresis
SSP	Secondary Structure Propensities calculator
TFE	Trifluoroethylene
TRIS	Tris(hydroxymethyl)aminomethane
TROSY	Transverse Relaxation Optimized Spectroscopy
UniProtKB	UniProt Knowledgebase (http://www.uniprot.org)
UV	Ultraviolet
WT	Wild Type (i.e., a non-mutated protein)

1. Scientific Background

1.1. Dehydrins – a subclass of plant stress proteins

Dehydrins belong to the class of Late Embryogenesis Abundant (LEA) proteins. These were first found to be present in cotton (*Gossypium hirsutum*) seeds, accumulating at a late stage in embryogenesis, just before final desiccation of the seeds (Dure et al., 1981). Since then, increasingly more LEA proteins have been detected in many other plants, and found to be expressed not only during seed formation. They also appear in high amounts in various other plant tissues under water-related stress conditions, such as cold, drought, or high salinity (Ingram and Bartels, 1996; Cuming, 1999; Thomashow, 1999). Although much effort has been made to understand their role in stress tolerance, as overviewed in section 1.2, the molecular function of most LEA proteins is still unclear (Tunnacliffe and Wise, 2007; Hundertmark and Hincha, 2008).

1.1.1. Classification of LEA proteins

LEA proteins have been classified into 3 major (group 1-3) and several minor subgroups according to similarities within their sequences, mostly in the form of short (about five to twenty amino acids long), conserved sequence motifs, the so-called segments (Dure et al., 1989; Bray, 1993, 1994; Cuming, 1999). Otherwise, LEA proteins typically have a lot of variation within their sequences (even within the subgroups), while preserving their overall hydrophilic nature, which results from a high content of charged and polar residues, and a rather low content of hydrophobic residues (Tunnacliffe and Wise, 2007; Hundertmark and Hincha, 2008). This kind of amino acid distribution is typical for intrinsically disordered proteins (IDPs), as discussed in detail in section 1.4, and is similar to that observed for a group of proteins called hydrophillins (this definition also includes a high glycine content, as present in some, but not all LEA proteins; Garay-Arroyo et al., 2000).

1.1.2. General characteristics of dehydrins, a subclass of LEA proteins

Proteins from group 2 of the LEA proteins are commonly called dehydrins because of an early proposed function in plants surviving drought stress (though nowadays it is known that also LEA proteins from other groups have similar functions; Close et al., 1989; Tunnacliffe and Wise, 2007). In contrast to groups 1 and 3 of LEA proteins, of which homologues have

also been found in bacteria, and certain invertebrates, such as nematodes and rotifers (Tunnacliffe and Wise, 2007), dehydrins can be found exclusively in plants. In the plant kingdom they are ubiquitously expressed in all seed plants, as well as in mosses and lycopods (Lüttge et al., 2011, p. 290).

Conserved segments of dehydrins

Typical conserved sequence motifs of dehydrins are the K-segment (EKKGIMDKIKEK-LPG), Y-segment (TDEYGNP), S-segment (SDSSSSSSS) (Close et al., 1989; Close, 1996) and the ChP-segment (charge-peptide-segment, also called the Lys-rich segment, in which several negative residues, like glutamic and aspartic acids, cluster before a group of positive residues, mostly a stretch of consecutive lysines, e.g. EEGEDGEKKKKEKKKK) (Mouillon et al., 2006), as shown for the dehydrins of *Arabidopsis thaliana (A. thaliana)* in figure 1 (*A. thaliana* is also called the thale cress, or mouse-ear cress, and is a small flowering plant typically used as a model organism in molecular biology). Sometimes, a rather loosely defined ϕ-segment is also mentioned, which is a less conserved region, rich in polar amino acids (Campbell and Close, 1997).

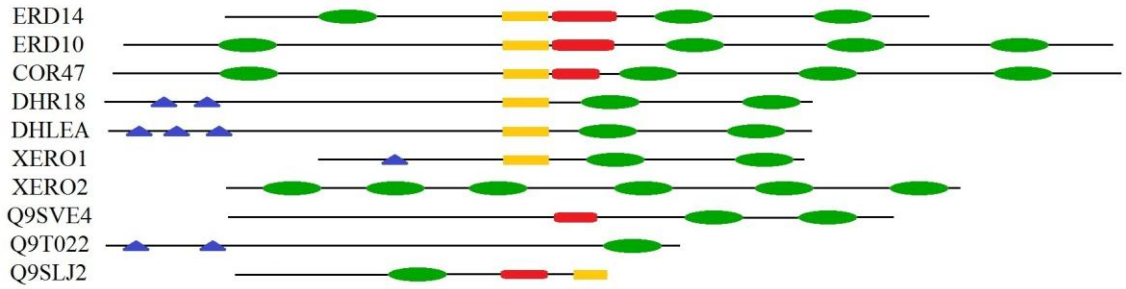

Figure 1. Schematic representation of the ten dehydrins found in A. thaliana with their characteristic segments highlighted: K-segment (green ovals), Y-segment (blue triangles), S-segment (yellow rectangles), and ChP-segment (red boxes). Non-conserved regions are marked with a black line, which is proportional in length to the number of amino acids it represents.

By definition, every dehydrin contains one or more K-segment(s), combined with a varied composition and number of the other segments (Dure, 1993; Close, 1996). Dehydrins are often classified according to the presence and order of the main segments within the sequence (denoted as: $Y_nS_nK_n$) (Close, 1996), though no obvious relation of this classification to the function of the included proteins has been described so far (Campbell and Close, 1997; Nylander et al., 2001; Lüttge et al., 2011, p. 292). It is noted, however, that this classification does not take into account the presence or absence of the ChP-segment.

In computational analyses comparing sequences to functions and expression patterns known so far, the presence or absence of Y-segments (independently from other segments) was suggested for a classification into two subgroups 2a and 2b, respectively (Wise, 2003; Wise and Tunnacliffe, 2004; Tunnacliffe and Wise, 2007). According to this differentiation, proteins with Y-segments – named group 2a – are preferentially expressed late in embryogenesis (as originally defined for LEA proteins), while those without the Y-segment are more associated with cold tolerance and named group 2b. It was also noted, that dehydrins from group 2a are typically neutral or basic in overall charge, while proteins from group 2b are mostly acidic (Tunnacliffe and Wise, 2007). In addition, at least for dehydrins in *A. thaliana*, it can be observed that the ChP-segment is present only in dehydrins that do not have a Y-segment (i.e., group 2b proteins, compare also with figure 1).

Prevalence and expression of dehydrins

Typically, each plant species accommodates several different dehydrins, though different species accommodate different numbers of dehydrins, e.g. there are ten in *A. thaliana* (as listed in figure 1), eight in rice and thirteen in barley, but only three in poplar and three in grapevine (Hundertmark and Hincha, 2008). Some of these dehydrins are upregulated upon abiotic stress (e.g. six out of ten in *A. thaliana*), though expression intensities vary depending on both the protein and the type of stress. Similar expression patterns were found upon treatment of plant leaves with the plant hormone abscisic acid (ABA). Therefore, the stress-induced expression of dehydrins (and of other LEA proteins) is thought to be ABA-dependent (Hundertmark and Hincha, 2008). ABA is also known to regulate the maturation process in seeds and the onset of dormancy (Lüttge et al., 2011, p. 275). However, ABA-independent dehydrins have also been reported (Zhang, Ervin, et al., 2006).

Structural features of dehydrins

As already suggested by predictions, and verified experimentally for several examples (see further below), dehydrins lack a fixed, three-dimensional structure, and thus belong to the class of IDPs (introduced in detail in section 1.4). Most investigations so far included the analysis of dehydrins by Circular Dichroism (CD) spectroscopy, Differential Scanning Calorimetry (DSC), 1D NMR or wide-line NMR (Nuclear Magnetic Resonance) spectroscopy. According to these analyses dehydrins were shown to contain a high degree of disorder, with a minor (0.7–5%) inclusion of α-helices (Lisse et al., 1996; Koag et al., 2003; Soulages et al., 2003; Mouillon et al., 2006; Kovács, Kalmár, et al., 2008; Bokor et al., 2005; Tompa et

al., 2006). However, an increase in helicity was found in several cases upon increasing concentrations of structure promoting molecules, such as trifluoroethylene (TFE, Mouillon et al., 2006). Furthermore, from the analysis of the temperature dependence of CD spectra, a small content in polyproline II (PPII) helices was also suggested for some dehydrins (10-20%, which is lost upon higher temperatures, Soulages et al., 2003; Mouillon et al., 2006).

As suggested by CD spectroscopy investigations, the structural state of dehydrins is susceptible to their actual environment. Therefore, it was proposed that they might be switched on (into their functional state) by binding-induced conformational changes, as is known to be the case for several intrinsically disordered proteins, that undergo a so-called folding-upon-binding process (Lüttge et al., 2011, p. 294). However, dehydrins ERD10 and COR 47 have been shown to retain their intrinsically disordered state (Mouillon et al., 2008) even under severe molecular crowding conditions (for more details see also sections 1.4 and 1.7). Therefore, alternatively to the folding-upon-binding model, the possibility of specific interactions of the conserved segments with their targets as the basis of function has also been suggested. In this model, segments may act as "beads on a string" for specific recognition, interaction with other proteins and membranes, or intermolecular scaffolding (Lüttge et al., 2011, p. 296; Mouillon et al., 2006, 2008).

To date, an in depth analysis providing residue-level information on detailed structural traits and residual structural elements has been performed only on dehydrin 1a from *V. riparia* (as well as a part thereof, BMRB accession numbers 16445 and 16450; Findlater and Graether, 2009) and on ERD14 from *A. thaliana* (present work, Szalainé Ágoston et al., 2011). For a detailed comparison of the results of these two analyses see section 5.1.

1.1.3. Dehydrin ERD14

The dehydrin of *A. thaliana* investigated within the present work, ERD14 (Early Response to Dehydration 14, UniProtKB accession number P42763), contains three K-, one S- and one ChP-segment, as marked within its amino acid sequence shown in figure 2. It should be noted, however, that the first of the three K-segments (denoted Ka in this work) shows only 30% of sequence similarity to the classical K-segment, and is therefore not always denoted as such. Thus ERD14 is often classified to be an SK_2 dehydrin (though classification as SK_3 would be justifiable as well), or as a dehydrin belonging to group 2b (acidic dehydrins).

Similarly to other dehydrins, ERD14 was shown previously by limited proteolysis, heat stability, 1D ^{1}H NMR and CD spectroscopy (Kovács, Kalmár, et al., 2008) to be intrinsically

disordered with a minor α-helical content and/or helical preference. ERD14 was also shown to bind to acidic phospholipid vesicles *in vitro* and to the plasma membrane *in vivo*. The *in vitro* interaction is electrostatic in nature, affecting membranes only peripherally via their phospholipid head groups (rather than immersing into the membrane, Kovács, Kalmár, et al., 2008). Lipid binding of dehydrins typically is ascribed to their K-segments, as these can – at least in theory – form amphipathic α-helices (see also section 1.2).

```
1       10        20        30        40        50        60
MAEEIKNVPEQEVPKVATEESSAEVTDRGLFDFLGKKKDETKPEETPIASEFEQKVHISE
                                 Ka*
        70        80        90       100       110       120
PEPEVKHESLLEKLHRSDSSSSSSSEEEGSDGEKRKKKKEKKKPTTEVEVKEEEKKGFME
                  S                 ChP                   Kb
       130       140       150       160       170       180
KLKEKLPGHKKPEDGSAVAAAPVVVPPPVEEAHPVEKKGILEKIKEKLPGYHPKTTVEEE
                                                Kc
  185
KKDKE
```

*Figure 2. Sequence of ERD14 with its conserved regions highlighted: K-segment (green), S-segment (yellow), and ChP-segment (red). *Segment Ka shows only 30% of sequence similarity to the usual K-segment, and therefore is not always denoted as such.*

Large-scale proteomics analyses have shown that ERD14 is present *in vivo* in chloroplasts (Kleffmann et al., 2004; Zybailov et al., 2008), is attached to membranes (Benschop et al., 2007; Nikolovski et al., 2012), and is also found in the cytosol (Ito et al., 2011). Furthermore, *in silico* predictions suggest nuclear localization (using PredictNLS, Cokol et al., 2000), and DNA binding (using PredictNLS and/or LOCtree, Nair and Rost, 2005). It is noted, that both the nuclear localization signal and the DNA binding sequence recognized by PredictNLS are part of (or overlapping with) the ChP-segment. Interestingly, two other dehydrins (from group 2a, see section 1.2) have been experimentally shown to localize upon phosphorylation to the nucleus – these however are different from ERD14 in that they do not contain the ChP-segment with the predicted nuclear localization signal.

ERD14 is phosphorylated at several sites *in vitro* as well as *in vivo*, which significantly increases its calcium-binding capacity, probably through its S-segment (Alsheikh et al., 2003, 2005). Interestingly, another phosphorylation site (S59) also seems to be functionally important, as ABA treatment resulted in a significant decrease in its phosphorylation state *in vivo* (Kline et al., 2010).

1.2. The molecular function of dehydrins

Several molecular functions of dehydrins have already been proposed, which could contribute to plant stress tolerance. Suggested molecular functions of dehydrins include, e.g., chaperone function, membrane stabilization, water buffering, radical scavenging and metal binding. Many of these have been investigated either at the molecular level *in vitro* or through the *in vivo* observation of phenotypes of stress resistance in transgenic experiments. As outlined in detail below, several results relate to the one or other dehydrin, but the lack of systematic analyses precludes sound generalizations, as different results are often obtained for different proteins, plant species and/or experimental setups. It has also been suggested that dehydrins might not have one general function (common to all dehydrins) but instead have specialized individual functions, probably explaining why functional investigations give a variety of results for different dehydrins.

1.2.1. Stress protection *in vivo*: phenotypic investigations

The effect of overexpression in living transgenic plants has been investigated for several dehydrins and host organisms, resulting in many cases in improved stress resistance. In the few cases known, a knockout of dehydrins resulted in the opposite effect, as demonstrated by studies detailed below (Lüttge et al., 2011, p. 293; Ochoa-Alfaro et al., 2012).

Stress protection in plants

Improved stress resistance in transgenic plants, such as tobacco, cucumber, rice and *A. thaliana*, has been observed for several dehydrins. Examples include cold tolerance improved through the citrus dehydrin CuCOR19 (Hara et al., 2003), the potato dehydrin DHN24 (Yin et al., 2006), the rhododendron dehydrin RcDhn5 (Peng et al., 2008), as well as freezing tolerance induced by paddle cactus DHN1 (shown in figure 3 A, Ochoa-Alfaro et al., 2012) or by the simultaneous overexpression of (endogenous) RAB18 and COR47 or ERD10(LTI29) and LTI30 dehydrin pairs in *A. thaliana* (Puhakainen et al., 2004). Drought and salt resistance was improved by a wheat dehydrin (Cheng et al., 2002), and osmotic stress tolerance by maize DNH1/Rab17 (Riera et al., 2004) and wheat DHN5 dehydrins (Brini et al., 2007).

Intriguingly, negative examples (where no change in stress tolerance was observed) have also been reported, e.g., no effect on drought tolerance was observed for tobacco plants expressing two dehydrins from a resurrection plant (Iturriaga et al., 1992), nor for *A. thaliana*

plants simultaneously overexpressing (endogenous) RAB18 and COR47 or ERD10(LTI29) and LTI30 dehydrin pairs (Puhakainen et al., 2004).

An overview of the experiments performed until 2007 concludes, that transgenic plants with a dehydrin from group 2b all seem to have improved chilling and freezing tolerance, while group 2a proteins usually confer little or no such effect (Tunnacliffe and Wise, 2007). Results since 2007 seem to confirm this thesis, as the overexpression of further group 2b proteins improved cold stress tolerance to their host plants, e.g BjDHN2/BjDHN3 from mustard (Xu, Zhang, Guan, et al., 2008) or RcDhn5 from rhododendron (Peng et al., 2008).

On the other hand, the absence (knockout) of the dehydrin DHN1 from *Vigna unguiculata* impaired chilling tolerance of their host plants (Ismail et al., 1999). Similarly, the independent knockout of two dehydrins of the moss *Physcomitrella patens* reduced in both cases the ability of the plants to recover after salt or osmolyte stress (Saavedra et al., 2006; Ruibal et al., 2012), as shown in figure 3B.

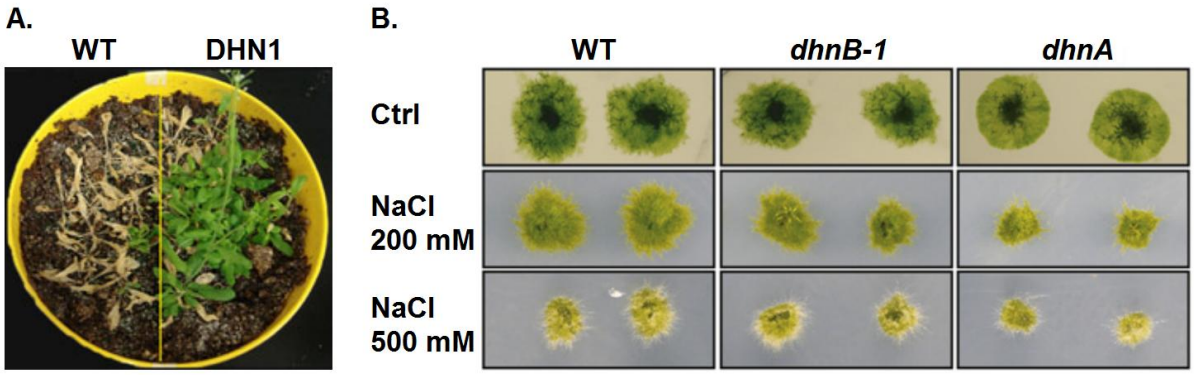

Figure 3. Exemplary results from stress tolerance studies of transgenic plants. A. Control and DHN1-overexpressing A. thaliana plants twenty-one days after freezing stress, showing increased stress tolerance in the presence of the dehydrin (Ochoa-Alfaro et al., 2012). B. Control and two dehydrin-knockout mutants of the moss P. patens under normal and salt stress conditions, showing reduced growth after salt stress in the absence of dehydrins (Ruibal et al., 2012).

Stress protection in other model organisms

To exclude secondary effects of other dehydrins (other than the protein under investigation) present in plants, the protective effect of dehydrins on model organisms, such as *Saccharomyces cerevisiae* (baker's yeast) or *Escherichia coli*, was also investigated. A protective effect of dehydrins was shown in several cases, as discussed below, though many proteins conferred protection to their host cells in only one of the stress effects tested.

Furthermore, for two dehydrins also a growth inhibitory effect on *E. coli* cells has been reported.

For example, the overexpression of the tomato dehydrin LE4 (now called TAS14) in *S. cerevisiae* resulted in increased freezing and KCl tolerance of host cells, but provided no protection against NaCl and sorbitol (Zhang et al., 2000). In comparison, other LEA proteins conferred *S. cerevisiae* an increased freezing tolerance (group 3 and 4 proteins) and NaCl tolerance (group 1, 3 and 4 proteins), while in none of the tested cases (group 3 and 4 proteins) any protective effect against sorbitol was observed (Imai et al., 1996; Honjoh et al., 1999; Swire-Clark and Marcotte, 1999; Zhang et al., 2000).

The overexpression of soybean dehydrin ZLDE-2 in *E. coli* cells resulted in a shortened osmoadaptive lag period upon NaCl challenge, while no such protective effect was observed against KCl or cold (4°C) treatment (Lan et al., 2005). In comparison, group 1 and 3 LEA proteins were effective in all three, NaCl, KCl and cold stress treatments (Lan et al., 2005; Liu and Zheng, 2005).

Furthermore, dehydrins ERD10 and COR47 from *A. thaliana*, the two closest relatives to ERD14, have been reported to inhibit *E. coli* cell growth (Campos et al., 2006). Based on the effect of deletion mutants of ERD10, the cationic nature of the ChP-segment ("K-stutter") and the K-segment of the dehydrins was suggested to be responsible for this antibiotic effect. However, in these experiments, not only the full ERD10 sequence, but also three shorter sequence parts were expressed, and the presence of these additional peptides also correlated with the growth inhibitory effect (Campos et al., 2006).

1.2.2. The molecular function of dehydrins: *in vitro* investigations

Several possible molecular functions of dehydrins have been suggested, mainly connected to either chaperone function, i.e. the protection of proteins and stabilization of membranes, or to metal binding and radical scavenging functions, as outlined below. The relevance of phosphorylation, DNA binding and localization in dehydrin function will be discussed as well.

Chaperone functions: the involvement in protein and membrane stabilization

The possible chaperone function of dehydrins has been investigated in several *in vitro* chaperone assays, in which the effect on either thermostability (cryoprotection, heat protection) or chemical stability of model substrates such as lactate dehydrogenase, lysozyme

or citrate synthase was investigated (Hanin et al., 2011, table 1). Potent chaperone activity has been demonstrated in this way for several dehydrins, such as ERD10 and ERD14 from *A. thaliana*, P80/DHN5 from barley, PCA60 from peach, DHN5 from wheat or CuCOR19 from citrus (Kovács, Kalmár, et al., 2008; Kovács, Ágoston, et al., 2008; Bravo et al., 2003; Wisniewski et al., 1999; Brini et al., 2010; Hara et al., 2001).

In addition, a nematode group 3 LEA protein reduced the propensity of polyglutamine and polyalanine expansion proteins to aggregate *in vivo* (in a human cell line, Chakrabortee et al., 2007). This finding is especially relevant from a medical point of view, as the aggregation of polyglutamine and polyalanine expansion proteins has been associated with neuro-degenerative diseases.

In the context of chaperone function, it is noteworthy that the ChP-segment of dehydrins has high sequence similarity to several known chaperones, e.g. to the linker region of HSP90 proteins (Mouillon et al., 2006), or to the so-called "KEKE-motif" found in chaperonins (Realini et al., 1994). This similarity to different chaperones might implicate the importance of the ChP-segment in the chaperone function of dehydrins, which however has not been investigated so far.

With respect to the molecular mode of action of intrinsically disordered chaperones (for a more detailed discussion see section 1.5), they could act indirectly as a hydration buffer, through entropy transfer or as molecular shields (Tompa and Csermely, 2004; Wise and Tunnacliffe, 2004; Goyal et al., 2005; for a more detailed discussion see section 1.5). Dehydrins could act as molecular shields either as disordered coils freely in solution or associated (via their segments as binding regions) to proteins or membrane surfaces. For example in the latter case the dehydrins could prevent membrane fusion and collapse by blocking the close approach of bilayers during the shrinking of cells due to water loss. This is closely related to the space filling role that has been suggested as early as 1953: of dehydrins being "plasticisers or mechanical buffers in the cell" (Siminovitch et al., 1953).

Membrane binding and stabilization might also be important in dehydrin function, as upon cold stress an accumulation of dehydrins near membrane surfaces (Steponkus et al., 1998; Danyluk et al., 1998; Ruibal et al., 2012), and the localization in membrane-rich areas within particular cellular compartments (Asghar et al., 1994; Egerton-Warburton et al., 1997) have been observed. Furthermore, some dehydrins have been shown *in vitro* to bind to different types of lipid or phospholipid vesicles (including ERD14; Kovács, Kalmár, et al., 2008)

which induces their folding (e.g. increased helicity of maize DHN1, Koag et al., 2003). Hereby, mainly K-segments are proposed to be involved, as they may form amphipathic α-helices similar to known membrane binding elements. However, an isolated K-segment (short peptide on its own) appears to be fully disordered in solution (with a helical propensity of below 1%), and attains even in 90% TFE (which strongly promotes the formation of helices) only about 10% α-helical structure (Mouillon et al., 2006). This draws into question whether an amphipathic helix formed by the K-segment can truly be made responsible for membrane association.

Inherent to the typical overall amino acid composition of dehydrins – containing a high number of charged amino acids – is that they can bind high amounts of water and possibly act as hydration buffers. In this sense they might withhold water within the cell even under stress conditions, allowing sufficient water activity for other proteins to retain their function (Rinne et al., 1999), as well as slowing down the rate of water loss during dehydration (McCubbin et al., 1985; Cuming, 1999; Garay-Arroyo et al., 2000). According to a more detailed concept, a shift in dehydrin structure might take place as desiccation proceeds from a fully-hydrated random coil state to a final collapsed structure enriched in α-helices or β-sheets, while water is gradually released (Mouillon et al., 2006).

Metal binding and radical scavenging activities

Another hypothesis for dehydrin function is that of protection against metal toxicity and oxidation, either by sequestering specific metal ions and radicals (Dure, 1993; Danyluk et al., 1998) or by acting as "ion-sinks" (i.e., to generally bind solute ions, rather than specific ones, in order to keep ion concentrations low even during dehydration, Tompa et al., 2006).

According to different *in vitro* investigations, some dehydrins bind metal ions, such as calcium (in some cases dependent on phosphorylation, see further below), or other bi- or trivalent ions such as iron, copper, zinc and nickel (Svensson et al., 2000; Kruger et al., 2002; Hara et al., 2005). In this regard, it was shown for the dehydrin CuCOR15 that its His-rich region was responsible for copper binding (however, this protein did not bind magnesium or calcium, Hara et al., 2005).

Also, an *in vivo* upregulation of some dehydrins after exposure to selenium (Pelah and Cohen, 2005) or heavy metals such as mercury, cadmium, zinc and copper (Zhang, Li, et al., 2006) has been shown – though in the latter case the same gene did not respond to drought stress or ABA, as dehydrins otherwise typically do. On the other hand, overexpression of two

dehydrins (from mustard, observed to be upregulated upon heavy metal stress) enhanced heavy metal tolerance of transgenic tobacco plants (Xu, Zhang, Wei, et al., 2008). Furthermore, *in vivo* antioxidant properties have been reported for a citrus dehydrin (in transgenic tobacco plants, Hara et al., 2003) which has *in vitro* scavenging activity for hydroxyl and peroxyl radicals (Hara et al., 2004).

Hydroxyl radicals and singlet oxygen (so-called reactive oxygen species, ROS) are generated in plant cells by catalytic metals, can accumulate during water deficit and are highly reactive towards organic molecules. Therefore, dehydrins with antioxidant properties are supposed to act either directly by scavenging ROS or indirectly by sequestering metal ions that generate ROS. Therefore, combined with the membrane binding capacity mentioned previously, dehydrins might also protect membranes against lipid peroxidation (Hara et al., 2003, 2004).

Other possible functions: DNA binding

Dehydrins have been suggested to have a possible function connected to DNA binding, possibly mediated by their ChP-segment. In fact, DNA binding has been shown *in vitro* for at least one dehydrin, CuCOR15, where its His-rich domain and ChP-segment were identified as being involved in DNA binding (Hara et al., 2009). However, the functional significance of this observation is unknown so far.

Possible regulation of function: Phosphorylation of dehydrins

Dehydrins can be phosphorylated *in vivo* as well as *in vitro*, which involves no structural changes (ERD10/LTI29 and COR47 proteins, Mouillon et al., 2008), but leads in some cases to an increased calcium-binding activity, with the S-segment serving as the main phosphorylation site (e.g. ERD10, ERD14 and COR47; Alsheikh et al., 2003, 2005). However, also metal ion binding without phosphorylation has been demonstrated (Hara et al., 2005), and RAB18 was shown not to bind Ca^{2+} at all, independent of its phosphorylation state (Alsheikh et al., 2005). It is noted, however, that the calcium-binding ability of some dehydrins can indicate both a function as a (phosphorylation dependent) Ca^{2+}-buffer and e.g. a chaperone-like activity that is phosphorylation-activated (and Ca^{2+}-dependent, Ochoa-Alfaro et al., 2012). For two group 2a dehydrins an altered localization upon phosphorylation has been shown (from cytosolic to nuclear, Jensen et al., 1998; Mehta et al., 2009), which may have a functional importance or serve in the activation of the protein.

All these results imply that phosphorylation might be a critical step in the function and/or localization of some – though not all – dehydrins.

Suggested functions for specific segments

The segments of dehydrins have been thought to fulfill some specific functional roles, as these are the only conserved sequence regions within these proteins. They have been repeatedly proposed to be recognition motifs, like "beads on a string", driving interaction with other cellular components (Mouillon et al., 2006), or alternatively, these sequence motifs might also allow a function to be performed upon the dehydrin, e.g. the localization to a particular cell or tissue compartment, or a specific processing event such as activation or degradation, for instance through phosphorylation (Tunnacliffe and Wise, 2007). Table 1 summarizes the main proposed functions of specific conserved segments of dehydrins, as inferred from the *in vitro* and *in vivo* experiments described above.

Region	**Proposed function(s)**
K-segment	amphipathic helix formation; membrane binding and stabilization; protein binding; chaperone effect; antibacterial effect
S-segment	phosphorylation resulting in calcium binding and/or nuclear targeting
ChP-segment	chaperone effect; phosphorylation; nuclear targeting; antibacterial effect
Y-segment	connected to expression in seeds late in embryogenesis
His-rich regions	binding of calcium and other metal ions

Table 1. Overview of proposed functions for the characteristic regions of dehydrins. In contrast to the dehydrin segments, His-rich regions are not strictly conserved in sequence, and therefore not denoted a dehydrin segment, but regions with accumulated histidines appear within several dehydrins.

1.2.3. Conclusions from the functional investigations of dehydrins

Although a lot of examples of different *in vitro* and *in vivo* functions of dehydrins are known, the general mode of function of dehydrins at the molecular level is still unknown, and leaves the question of generality of their function (versus individual functions) open. Therefore, every dehydrin still has to be evaluated as a case on its own, and investigated in detail, as performed for ERD14 in the present work.

Moreover, it should be noted, that not only dehydrins and LEA proteins play a role in defense or repair mechanisms upon abiotic stress effects of plants: there are also other hydrophilic proteins, small heat shock proteins, sugars (e.g. trehalose and sucrose) and large protein complexes such as expansins and aquaporins involved (Lüttge et al., 2011, chapters 15.11, 15.13 and 16.4). Therefore, abiotic stress effects, and the possible defense mechanisms involved and observed, have to be considered in conjunction with this overall picture of stress

reaction. For example, heat shock proteins– just as the dehydrin ERD14 – are also expected to confer dehydration tolerance by chaperone function (through the prevention of denaturation and/or aggregation of proteins) as well as by their protective effect on membranes.

1.3. The cellular effect of stress conditions: model system *E. coli*

In order to investigate the effect of dehydrins within a living organism (the plant), often the stress resistance of transgenic plants is studied, in which the target dehydrin gene is either silenced or activated for overexpression (see section 1.2 above for examples). Alternatively, single-celled microorganisms such as *S. cerevisiae* (baker's yeast) or *E. coli* have been used, as overexpression of target proteins in these model organisms can be more easily regulated and contain no dehydrins or other LEA proteins, thus interference by other dehydrins can be excluded. However, the testing of stress resistance of yeast or bacteria requires different experimental setups from those used for plants (which is quite heterogeneous in itself, though; Lüttge et al., 2011).

As in the present work *E. coli* was used as a model system for *in vivo* investigations, previously used experimental conditions for other LEA proteins in *E. coli* will be reviewed below to provide a basis for comparison to the methods and the results presented in this work.

Experimental conditions testing the stress tolerance conferred by LEA proteins

The effect of LEA proteins on cold stress tolerance of *E. coli* cells has been tested using a 24 hour incubation period at 4°C (a shortening of the lag phase of growth was observed, Lan et al., 2005). For the investigation of salt stress tolerance, agar-plates were supplemented with 500 mM NaCl or KCl in one case (increased colony numbers were observed, Liu and Zheng, 2005), while plates and liquid media with 800-1000 mM NaCl or 700-800 mM KCl were used in the other case (a shortening of the lag phase of growth as well as an increase in colony numbers on plates were observed for some of the proteins, see also section 1.2, Lan et al., 2005). Osmotic stress (dehydration) was tested using 1100 mM sorbitol (though no protective effect could be observed, Liu and Zheng, 2005). As a control, an induced empty vector was used in both works.

For the quantitation of the protective effect of the LEA proteins under investigation, either bacterial growth was followed in liquid media (observing the length of lag phase and/or the growth rate) or the number of resulting colonies on a solid medium – containing salt or osmolyte – was counted (also called the "CFU method", "Colony Forming Units" method).

Possible further stress conditions on *E. coli* cells

The viability of *E. coli* cells under several different stress conditions (without the presence of dehydrins) was investigated in several earlier analyses. Survival rates of *E. coli* cells after freezing-thawing cycles using different freezing velocities (e.g. 5-30,000 °C/min., Dumont et al., 2004) and different osmotic stress treatments at several temperatures (e.g. 5-11 M glycerol and 4-37°C, Mille et al., 2002; or 50% sucrose at 4°C using different dilution rates, Bayer, 1967) have been published, and guided the choice of experimental conditions within the present work, as described in section 4.2.3. Similar analyses on yeast have also been reported (e.g. up to 2.5 M NaCl or 4 M glycerol, Morris et al., 1986). In all these examples the CFU method was used to determine the viability of cells.

1.4. Intrinsically disordered proteins: IDPs

Proteins or regions of proteins, which lack a single well-defined structure under physiological conditions, but sample a wide conformational space of several interconverting structures, are called intrinsically disordered proteins or regions (IDPs/IDRs, sometimes also called natively unstructured). As the existence and relevance of IDPs has been realized only recently, a detailed overview of the relevance, function and known structural features of IDPs is given below.

1.4.1. History and relevance of IDPs

The first conceptual papers in the field of IDPs were published between 1999 and 2002 (Wright and Dyson, 1999; Uversky et al., 2000; Dunker et al., 2001; Tompa, 2002), though deviations from the structure-function paradigm have been apparent but neglected and/or ignored before (Tompa, 2012). Contrary to the general view (that has still been dogmatically evident little more than ten years ago) that a stable well-defined three-dimensional structure is an essential prerequisite of protein function (the so-called structure-function paradigm, typically illustrated by the key-and-lock principle of selective enzyme function, see figure 4), nowadays a wide range of evidence shows that IDPs are able to function despite their lack of structure (Tompa, 2012).

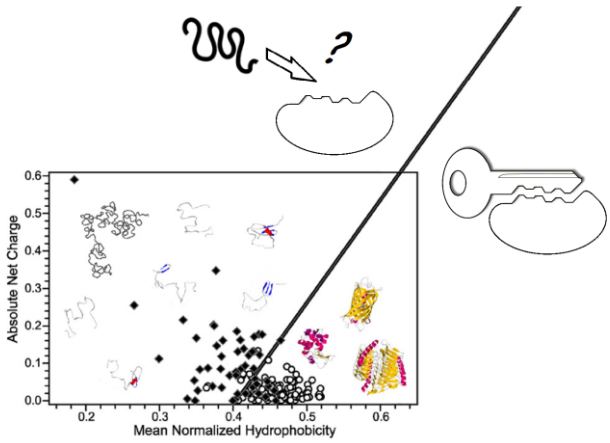

Figure 4. Charge-hydrophobicity plot, as used for the distinction of intrinsically disordered (left, black diamonds) and globular proteins (right, open circles). Several examples of possible structures are depicted within the plot (Uversky, 2011b). Next to the plot a schematic representation of the key-and-lock principle (explaining selective enzyme function and protein recognition) is shown (right), as well as a schematic indication of an intrinsically disordered "key" defying the classical structure-function paradigm.

Moreover, the functioning of several IDPs is possible only through their observed structural heterogeneity, as discussed further below.

It has to be emphasized, of course, that the structure-function paradigm has extreme explanatory power, as the about 80 000 protein structures (as of January 2013) deposited in the Protein Data Bank, PDB (www.rcsb.org, Berman et al., 2000), give ample insight into their functions and functional details. But the paradigm has to be expanded to include also the function and variable structures of IDPs (Tompa, 2012).

IDPs constitute a substantially different class of proteins from globular ("normal", ordered) ones, outstanding already in their primary sequence, and hence general properties such as net charge and hydrophobicity (as can be seen e.g. in the charge-hydrophobicity plot (Uversky et al., 2000) shown in figure 4). Per a more detailed analysis, IDPs are typically depleted in order-promoting amino acids such as Cys, Trp, Tyr, Ile, Phe, Val, and Leu and enriched in disorder-promoting amino acids such as Glu, Pro, Gln, Ser, Arg, Lys and Met (Radivojac et al., 2007). Based on this difference in amino acid composition – and later on also based on other sequence-based characteristics – many predictors of disorder have been developed, such as e.g. PONDR© (Romero et al., 2001) or IUPred (Dosztányi et al., 2005a, 2005b). It is to be noted that the most accurate prediction can be reached when results from several predictors are combined and/or compared – on this account meta-predictors combining the results of primary predictors have been developed (see e.g. Radivojac et al., 2007 for a list of available predictors and meta-predictors).

Using such predictors, full genomes/proteomes have been screened for the prevalence of disorder in different organisms. The results indicate that about 10-35% of all prokaryotic and 15-45% of all eukaryotic proteins contain long disordered regions (spanning more than 30 residues), with an overall tendency of higher disorder ratios within more complex organisms (Tompa, 2012). Therefore, intrinsic disorder of proteins is not an exception – though the number of experimentally verified IDPs is still limited, see below – but rather a still largely undiscovered field within protein research.

As for the functional importance of IDPs, proteomics analyses also revealed, that disordered proteins or regions are especially abundant within processes correlated with signal transduction, regulation of transcription and protein-protein interactions (also including chaperones, as described in section 1.5), as well as in proteins involved in diseases (Kovács and Tompa, 2012). In contrast little or almost no disorder is found within functional classes

associated with catalysis, biosynthesis, metabolism, transport (e.g. electron transport, sugar transport), or trans-membrane proteins (Tompa, 2012; Uversky, 2011b).

1.4.2. Experimental evidence and functional classification of IDPs

More and more proteins are not just predicted, but also experimentally shown to be intrinsically disordered. As of the release date 10/15/2012, approximately 1500 regions in more than 600 partially or entirely disordered proteins have been deposited into DisProt, the "Database of Protein Disorder" (www.disprot.org, Sickmeier et al., 2007). Our knowledge about the structure and function of IDPs is ever growing. Nevertheless, it is still a challenge to understand the exact mode of function of IDPs on a molecular level (i.e., to expand the structure-function paradigm to include the dynamic behavior of IDPs), and intensive research is done to elucidate details about concrete examples to be able to make generalizations. It seems to be clear that the function of IDPs stems either directly from their disordered state (entropic chains) or from molecular recognition, i.e., the selective binding of partner molecules, often accompanied by an induced folding process (disorder-to-order transition, folding-upon-binding process; Tompa, 2002, 2012). In certain cases the binding of IDPs without folding has also been observed when the IDP preserved a significant level of disorder even in its bound state, a phenomenon termed fuzziness (Tompa and Fuxreiter, 2008). Continuing the main distinction mentioned above, IDPs have been categorized into the following functional classes: entropic chains, display sites, chaperones, effectors, assemblers and scavengers (Tompa, 2005), as well as the extra class of prion proteins (Tompa, 2009, chapter 12), as depicted in figure 5.

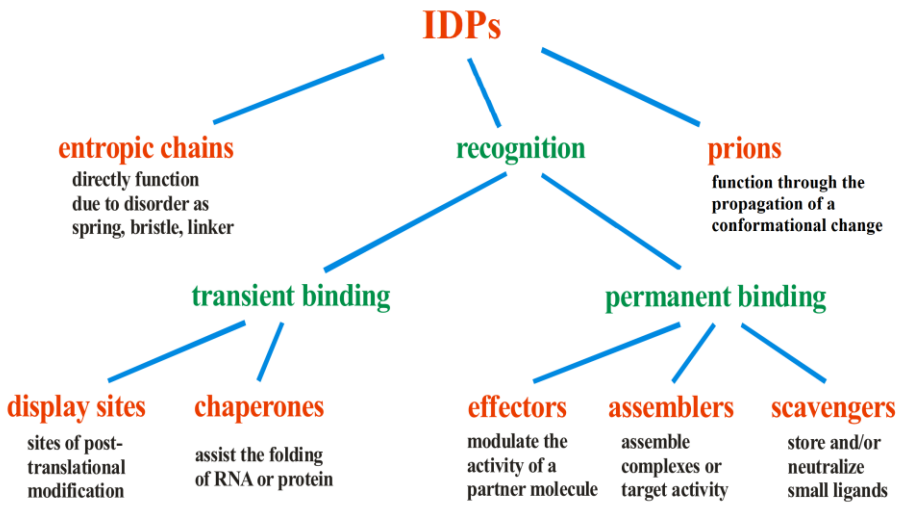

Figure 5. Overview of the seven functional classes of IDPs (adapted from Tompa, 2005). A short description of each of the functions is given in the text.

Entropic chain functions of IDPs arise directly from their disordered state, meaning that they either define the appropriate spatial separation of ordered domains (linkers) and of cell structures (entropic bristles, e.g. the MAP2 projection domain defining spacing in the cytoskeleton), or provide a spring-like elasticity to proteins, cellular structures or tissues, i.e. exert a force against mechanical stretching (entropic springs, e.g. titin in muscles).

Molecular functions arising from the recognition of partner molecules, with which transient binding is established, can be display sites, i.e. sites of posttranslational modifications within disordered regions (as for these – next to a given primary sequence motif – structural exposure and flexibility of the modified site and its environment are required), or molecular chaperones that assist in the folding of RNA or other protein molecules (see in detail in section 1.5).

Further recognition functions of IDPs involve the permanent binding of partner molecules, such as in the case of effectors (activating or inhibiting their partner through permanent binding), assemblers and scavengers. Assemblers recruit and localize different binding partners in close proximity for an effective interaction and/or complex formation process, while scavengers are able to bind a large number of small organic compounds or ions (either for disposal, e.g. of tannins by salivary proteins, or for temporary storage and a later release upon the need of the organism, e.g. Ca^{2+} binding by calsequestrin).

Finally, prion functions involve the self-sustaining propagation of a conformational change, assuming that the prion is a protein that can adopt two distinct stable structural states, one of which is converted to the other through catalysis by molecules that already have acquired the second conformation. This conformational change then either leads to a functional advantage (e.g. in memory formation) or to adverse effects leading to deadly neurodegenerative diseases collectively termed prion diseases (e.g. the Creutzfeldt-Jakob disease, CJD, in humans or the bovine spongiform encephalopathy, BSE, in cattle).

1.4.3. Structural features of IDPs

From a structural point of view IDPs do not represent a homogeneous class of proteins. To differentiate between three or four different structural states the “Protein Trinity” or “Protein Quartet” models were proposed, which include the random coil, (pre-molten globule), molten globule, and ordered states of proteins (Dunker and Obradovic, 2001; Uversky, 2002). However, it should be noted that no strict boundary between these states can be drawn, but proteins are rather positioned along a structural continuum spanning these four possibilities.

By now, several specific techniques have been developed for the structural characterization of IDPs, as overviewed in section 1.6. Several IDPs and IDRs have been shown – while being overall dynamic and flexible – to contain different amounts of residual structure, with partially folded secondary structural elements and/or transient tertiary contacts. As the constant dynamic movement of IDPs does not fit into the classical description of protein structures (using a single set of atomic coordinates), new descriptions have been developed using molecular dynamics simulations to best describe experimental results with a large, but limited (typically 50-100) set of structures. These structural ensembles characterize the structural space that a given IDP samples over time (e.g. Tompa, 2012; Terakawa and Takada, 2011) as depicted for the N-terminal domain of tetrameric p53 in complex with DNA in figure 6.

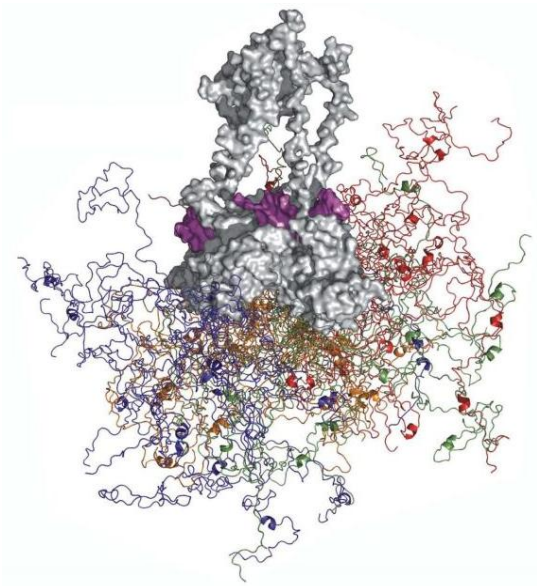

Figure 6. Disorder in p53. The structural ensemble of the N-terminal domain of tetrameric p53 is shown as several thin traces of the polypeptide chain (20 traces per monomer to represent the structural space sampled by the domain). The ordered tetramerization and DNA binding domains of p53 (light gray), as well as the bound DNA (dark gray) are shown in space filling model, while the C-terminal domain is not shown for clarity reasons (Wells et al., 2008; and cover picture Tompa, 2009).

Residual secondary structural elements observed within these structural ensembles are believed to be of possible functional importance, as so-called preformed structural elements (PSEs, Fuxreiter et al., 2004), or molecular recognition features (MoRFs, Mohan et al., 2006), through which binding and eventual folding of the IDP to its partner protein is initiated (though not all residual secondary structural elements take part in binding, e.g. Uversky, 2010). Folding of the disordered protein can take place either before (conformational selection

of a preformed structural element) or after binding (induced folding, where the unbound form indicates a structural preference towards the bound form), (Onitsuka et al., 2008). Furthermore, the presence of PSEs/MoRFs might also be exploited for drug development, as small molecules – designed to resemble the PSE/MoRF motifs – can compete with the binding of the given IDP protein, and thereby inhibit its function. The effectiveness of this approach has been demonstrated by the success of nutlins, which inhibit p53-MDM2 interaction and reactivate the p53 pathway in cancer cells (Cheng et al., 2006; Vassilev et al., 2004).

The presence of several such interaction motifs along the sequence can bring further functional benefits to IDPs, enabling stronger and/or more specific binding (when binding to the partner molecule involves several motifs) or assembler functions through binding of different motifs to different partners, hereby bringing them in proximity for complex formation, as observed for UPF proteins involved in the complex assembly during nonsense-mediated decay (NMD, Clerici et al., 2009). Furthermore, dynamics and mobility of IDPs also enable their adaptation to several unrelated binding partners, a property called binding promiscuity or moonlighting (Tompa et al., 2005; Tompa, 2012). In this process, IDPs can adopt different structures upon binding to different substrates, e.g. residues 367–391 of p53 are able to bind to seven different partners (Schreiber and Keating, 2011), as shown for four interactions in figure 7.

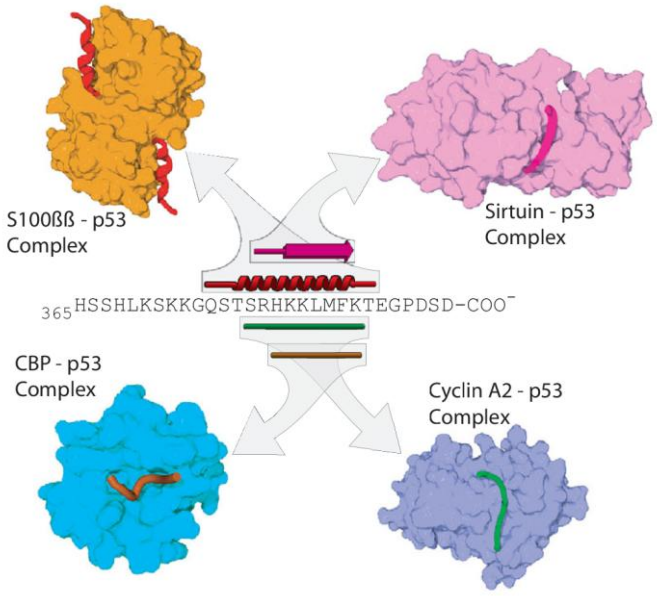

Figure 7. Four different conformations of the same sequence part of p53 in its bound states in complex with four different partners. Primary, secondary and quaternary structures are depicted (Oldfield et al., 2008).

1.4.4. IDPs under physiological conditions within the cell

The question is also open whether the disordered state observed for IDPs *in vitro* is also their true functional state within the cell (*in vivo*), or if structural disorder is an artifact of the given experimental conditions. This latter view assumes that disordered proteins in the cell fold either due to the presence of their physiological binding partner and/or through the general high concentration of other proteins and solutes within the cell, termed macromolecular crowding. Crowding promotes protein folding (for globular proteins) and assembly of complexes under cellular conditions by favoring compact conformations over extended ones due to the limited space available for macromolecules (excluded volume effects, Tompa, 2009, chapter 8.1; see also section 1.7 in the present work). For several IDPs, however, it has been shown that molecular crowding promotes the folding of local secondary structural elements, but does not induce an overall folding of the proteins (Mouillon et al., 2008; Szász et al., 2011; Cino et al., 2012). A similar conclusion was drawn by direct measurements within living cells by in-cell NMR (for details see section 1.7). Still, the observed examples of crowding-promoted local folding are important observations, indicating that preformed structural elements may significant *in vivo* (see above, and Tompa, 2009, chapter 8.2).

Furthermore, IDPs might diffuse faster under crowded conditions than globular proteins of comparable size (though the opposite is true in dilute solutions), which may be related to the high number of intrinsically disordered signaling proteins (enabling faster signal transduction, Wang et al., 2012). Faster diffusion can also contribute to the observed effect that generally IDPs are able to bind their partner molecules faster than ordered proteins (Huang and Liu, 2009).

The question of the physiological lifetime of unbound IDPs within live cells has also been raised, given their observed extreme proteolytic sensitivity *in vitro* and the ability of the 20S proteasome to degrade unfolded proteins without ubiquitination in the cytosol in a process also called “degradation by default” (see also Tompa, 2012). A large-scale high-throughput analysis, however, showed that the *in vivo* half-lives of 3750 yeast proteins only weakly correlated with predicted protein disorder, indicating that IDPs can have very long lifetimes *in vivo* (Belle et al., 2006; Tompa, 2009, chapter 8.4).

1.5. Structural disorder in chaperones

Chaperones are proteins that can either assist the folding of their substrates (either RNA or other proteins, or both) or prevent the unfolding, inactivation and aggregation of a previously properly folded substrate under denaturing conditions. Certain chaperones may also actively dissociate already formed aggregates (Uversky, 2011a). As some dehydrins, such as also ERD14, have been shown to have chaperone function – despite and/or because of their intrinsic disorder – the proposed mechanisms of chaperone action of IDPs are outlined below.

For a long time, chaperone function was generally thought to be structurally and functionally demanding and to strictly require a fine-tuned and complex three-dimensional structure. Especially protein chaperones have been considered to be large protein machines, using ATP (Adenosine-5'-triphosphate) energy to rescue their partly folded or misfolded substrates from kinetic folding traps, i.e. local minima along their folding pathway, thus enabling their proper folding or refolding. The best known such chaperone machines are the chaperonin GroEL-GroES, and the heat-shock proteins HSP90 and HSP70 (Kovács and Tompa, 2012). However, in both protein and RNA chaperones, a high amount of intrinsic disorder can be found (comparable to the percentage of disorder found in regulatory and cell signaling proteins, Tompa and Csermely, 2004), including short and long disordered stretches within large globular chaperones, and also fully disordered chaperones such as α-synuclein, β-synuclein, α-casein and several LEA proteins and dehydrins (Kim et al., 2002; Park et al., 2002; Bertoncini et al., 2007; Bhattacharyya and Das, 1999; and section 1.2 in this work).

In several cases, the functional significance of disorder in chaperone function has been shown experimentally (for a detailed review, see Uversky, 2011a). Furthermore, in several cases a reduction of flexibility within the disordered region of largely ordered chaperones during chaperone action could be observed, e.g. in the chaperonin cpn60 (Gorovits and Horowitz, 1995) as well as in α-crystallin (Lindner et al., 1998), which has contributed to developing a mechanistic model of chaperone action by protein disorder.

Mechanistic model of ID chaperone function

This complex mechanistic model proposed by Tompa and Csermely (2004) is called the entropy-transfer model. The model involves several mechanistic components, as depicted schematically in figure 8. The first component is that disordered regions often contain different recognition elements and offer a unique versatility in recognizing different binding

partners, and ensuring the wide substrate-specificity needed (or expected) for chaperones. This can be especially relevant in the case of Janus chaperones that have both, RNA and protein chaperone function. Specific sequences may also be able to recognize partially denatured proteins through their exposed hydrophobic patches. As a second component, disordered segments (once bound to their substrate) can provide a solubilizing effect and thus prevent the aggregation of misfolded substrates. This may be due to their hydrophilic nature and/or through an entropic exclusion effect (also termed molecular shield function). A further element of the model is the entropy transfer, a local disorder-to-order transition (loss of entropy) of the chaperone that energetically contributes to the concomitant unfolding (increase in entropy) of the misfolded substrate, thereby rescuing the latter from a local kinetic folding trap. Furthermore, binding of different parts of the disordered chaperone can keep partly folded structural elements of the substrate at close proximity, to speed up the folding process. Lastly, the binding of disordered regions can be of transient nature, enabling repeated cycles of contact and release of the misfolded substrate as well as ensuring the final release of the properly folded substrate (Tompa and Csermely, 2004; Kovács and Tompa, 2012).

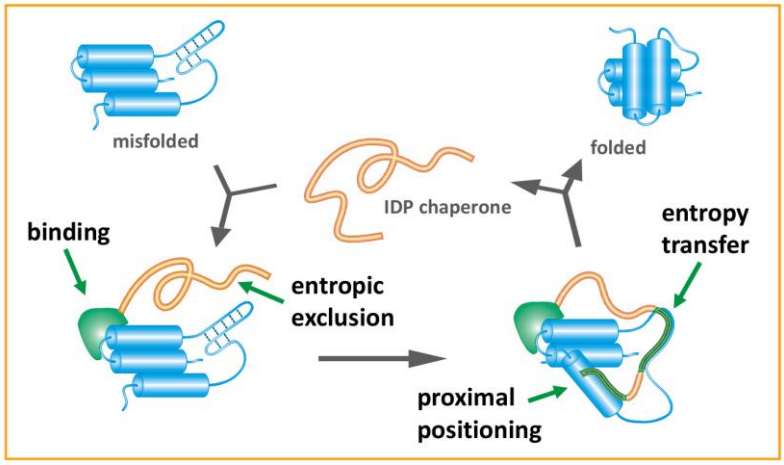

Figure 8. Scheme and main components of the entropy transfer model, which has been suggested to be a possible mechanism of action of intrinsically disordered chaperones (adapted from Tompa and Csermely, 2004).

It is noted, that these distinct functional modes may not all be active at the same time – for some proteins only the one or other action might be important for function. In all, it is to be emphasized that this model can not only explain the contribution of disordered segments to the chaperone action of ordered (and mostly ATP-dependent) chaperones, but also the mechanism of ATP-independent chaperone action of fully disordered proteins.

1.6. Possibilities for the structural characterization of IDPs

1.6.1. General overview

Several techniques are available for the identification and structural analysis of IDPs, as has been described in detail (Uversky and Longhi, 2010; Tompa, 2009). Here only a short overview is given to account for the multiple possibilities and the information content that can be gained through the different techniques. Main aspects of the techniques used within this work are outlined in more detail.

Methods for the structural analysis of IDPs will be categorized into the following groups (in analogy to Tompa, 2009): indicative techniques (for recognition or verification of the overall disordered state of a protein), hydrodynamic techniques (that assess diffusion, size and shape of the proteins in solution), spectroscopic techniques (assessing the presence and ratio of secondary structural elements and/or the local molecular environment of a chromophore), and single-molecule techniques (investigating one molecule at a time – in contrast to other techniques that observe an average over all conformational states present in the sample). A special case – and therefore a category on its own – is NMR spectroscopy, the most powerful technique for IDPs, which is unique in enabling the assessment of a variety of structural and dynamic features of a protein at atomic resolution.

Indicative techniques

Indicative techniques can verify the presence of disorder, though generally without providing detailed structural insight. They include the assessment of resistance of proteins to heat or chemical denaturation, DSC, SDS-PAGE (Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis) mobility, proteolytic sensitivity, the effect of chemical cross-linking, and H/D exchange.

IDPs are known to resist heat and chemical denaturation, i.e. they stay soluble even after boiling, and can therefore be identified and/or purified using this feature. Similarly, the disordered state of proteins can be detected by DSC, where the typical phase transition of the heat-induced cooperative unfolding of globular proteins is not observed in case of IDPs.

The SDS-PAGE mobility of IDPs has been shown to be typically less than normally expected for a globular protein of similar size. This effect is a result of their special amino acid composition, due to which they bind less SDS, resulting in an apparent molecular weight 1.2-1.8 times higher than expected (Tompa, 2002). Furthermore, the absence or presence of

(residual) structure can be tested in a special form of 2D-electrophoresis (native PAGE in the first dimension combined with urea-denatured PAGE in the second dimension, as explained in detail in Tompa, 2009, chapter 7.2; and Csizmók et al., 2006), in which separation of IDPs from partially and fully folded proteins can be achieved.

A further, indirect measure of structural disorder is proteolytic sensitivity (as a random coil chain is much more accessible to proteases than a globular protein) as well as the missing effectiveness of chemical cross-linking (due to the typical absence of tertiary contacts within an IDP). The rate of hydrogen-to-deuterium (H/D) exchange of backbone amides, which can be measured by FT-IR (Fourier-Transform Infrared) spectroscopy, MS (Mass Spectrometry) or NMR spectroscopy, is also a measure of solvent accessibility (and therefore of the compactness and structure of proteins). In certain cases, H/D exchange can also provide residue-level information on the presence or absence of structure and disorder.

Hydrodynamic techniques

Hydrodynamic techniques are able to estimate the size and shape of hydrated proteins, either directly or through assessing their diffusion characteristics. Available methods include gel filtration, SAXS (Small Angle X-Ray Scattering), analytical ultracentrifugation, DLS (Dynamic Light Scattering) and FCS (Fluorescence Correlation Spectroscopy), as well as certain MS and NMR methods.

Gel filtration (also termed size exclusion chromatography) is a technique that separates proteins by their hydrodynamic size, which can then be compared to the corresponding size of globular standard proteins, resulting in an estimate of different hydrodynamic parameters and/or of the apparent molecular weight of the investigated protein. Thus the estimated apparent molecular weight of IDPs is typically four to six times higher than its calculated weight (for a random coil like behavior; less is expected for a molten-globule-like IDP, Tompa, 2009, chapter 4.1).

SAXS measures the intensity of X-rays diffracted by proteins in a solution, thereby providing information on their overall size and shape distribution. From the scattered intensity the radius of gyration can be obtained and a kind of low-resolution structure (or low-resolution structural ensemble) can be calculated. Furthermore, the presence of residual structure in IDPs can be detected, and molten-globule, pre-molten-globule, and random-coil-like IDPs distinguished.

Analytical ultracentrifugation can be used to determine the size and shape asymmetry of IDPs (measuring sedimentation velocity) or the molecular weight of proteins (shape-independently, through sedimentation equilibrium measurement). Similarly, next to its molecular weight, the topology (compaction, size and shape) of a protein can be estimated by MS, evaluating either the charge distributions after electrospray ionization (ESI) or the ion mobility during flight and separation in a drift gas such as helium (Šamalíková et al., 2010). Although the MS analysis of proteins includes their evaporation into the gas phase, the conformational state, size and shape determined – at least within the ESI technique – are characteristic of the hydrated state of the proteins.

Hydrodynamic parameters of proteins are also often calculated from their respective diffusion characteristics, as determined by dynamic light scattering, fluorescence correlation spectroscopy, or pulsed field gradient NMR.

A further NMR methodology to assess the hydrodynamic properties of proteins is wide-line NMR, by which the amount of water within the hydrate layer of proteins is determined (IDPs have by far larger hydrate layers, than globular proteins of similar size).

Spectroscopic techniques

Most spectroscopic methods provide insight into more structural details of IDPs, such as ratios of (residual) secondary structural elements within the protein or information on the local environment of an intrinsic or introduced chromophore. Available techniques include two quite special cases: X-ray crystallography and NMR (as discussed in detail below), as well as several other techniques such as: FT-IR, CD, ROA (Raman Optical Activity), fluorescence, and EPR/ESR (Electron Paramagnetic Resonance or Electron Spin Resonance) spectroscopy.

X-ray crystallography is a very powerful technique for the structure determination of properly folded proteins, but its use is very limited for IDPs, as fully or largely disordered proteins often fail to crystallize. Even for proteins that do crystallize, disordered regions often only appear as segments that lack electron density, without any insight into their structure and dynamics. However, X-ray crystallography has played a significant role in the recognition of IDPs/IDRs, providing first evidence about the coincidence of function and disorder.

FT-IR, ROA, and CD spectroscopy measure the absorption and optical activity of a protein at different wavelengths, thereby providing information on the presence and relative ratios of

secondary structural elements within proteins. Most often (far UV) CD spectroscopy is used for the estimation of the secondary structure content of both ordered and disordered proteins.

Near UV (ultraviolet) CD spectroscopy, on the other hand, measures the optical activity of aromatic side-chains of the protein, and provides information about their local molecular environment, i.e. their solvent exposure, which is also probed by fluorescence spectroscopy. Similarly, the local environment and mobility around a paramagnetic center (naturally bound paramagnetic metal ion or an introduced small organic free radical) is sampled by EPR/ESR.

A widely employed fluorescence methodology is FRET (Fluorescence Resonance Energy Transfer), in which fluorescence activity of a fluorophore (donor) is passed on to another fluorophore (acceptor) in a distance-related manner. This method can be used to measure intra- or inter-protein distances, providing a measure of local tertiary contacts, structural compactness, and flexibility.

Single-molecule techniques

Single-molecule techniques are able to sample the behavior and characteristics of a single protein molecule at a time, enabling the assessment of single conformations of IDPs, instead of observing the average characteristics of the structural ensemble present in solution.

Due to its high sensitivity, fluorescence (e.g. FRET), can also be used to study single molecules and their interactions. Other possibilities for the observation of single molecules are force spectroscopy techniques, that either assess the shape of the molecules (e.g. Electron Microscopy, EM), or the force needed for the mechanical unfolding of a protein, which is proportional to the amount of structure formed previously (Atomic Force Microscopy, AFM).

1.6.2. Nuclear Magnetic Resonance Spectroscopy

NMR spectroscopy is the most powerful technique for IDPs, as it is able to assess a variety of structural and dynamic features of proteins at atomic (or at least residue-level) resolution.

NMR is based on the observation that within an external magnetic field the intrinsic magnetic moment of NMR-active nuclei is split into two (or more) non-degenerate energy levels, where the magnitude of splitting is measured and depends – among others – on the local chemical environment of the nucleus. Therefore, the main measured quantity of NMR, the chemical shift, provides information on local environment, such as the primary and secondary structure in case of backbone atoms. Other frequently measured quantities in NMR include relaxation data (R_1, R_2, and hetNOE values, as detailed below, that give information

about local dynamics and overall motions of the protein), NOE intensities (Nuclear Overhauser Effect, containing information on the average distance in space between relatively near nuclei), PRE ratios (Paramagnetic Relaxation Enhancement, containing information on long-range distances within proteins), and RDCs (Residual Dipolar Couplings, providing information on relative angles of chemical bonds within ordered proteins, or protein domains). Detailed descriptions of these quantities and the methods of their measurement with actual examples can be found in a recent book chapter (Mulder et al., 2010) or several excellent reviews in the field (e.g. Dyson and Wright, 2004). The key to all of this information is resonance assignment, i.e. the process of identifying which observed signal (or: "resonance") belongs to which particular atom within which particular residue.

First indications of disorder

As a first approximation, already a simple 1D or unassigned 2D (typically a so-called HSQC, heteronuclear single quantum coherence) spectrum can give a clue about the general structural state of a protein, as shown in figure 9. In case of disorder, the chemical shifts of a given type of atom (typically the backbone amide hydrogens termed within NMR amide "protons") are quite similar for every residue, as their local chemical environment is averaged due to their flexible and hydrated character in solution.

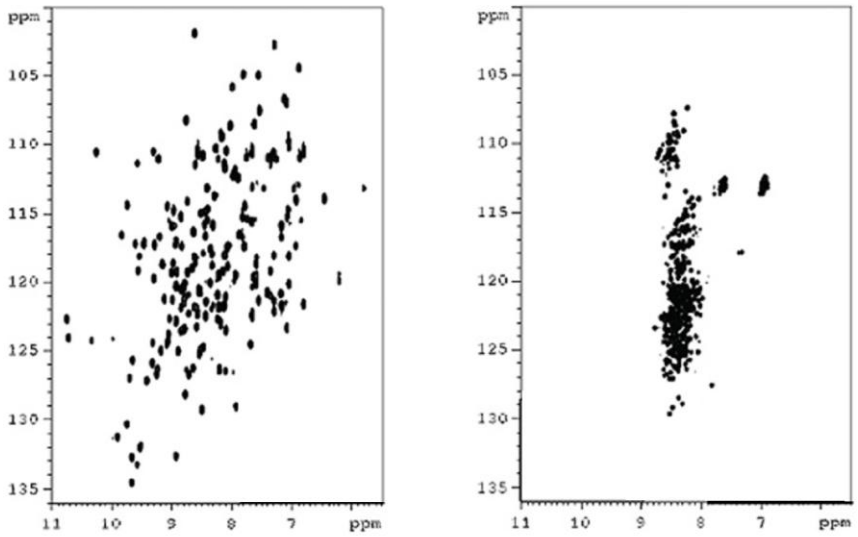

Figure 9. 2D ^{15}N-HSQC spectra of an ordered (cyclophilin A, left) and an intrinsically disordered protein (tau protein, right) to exemplify the difference in peak dispersion of amide protons in a folded versus a random coil environment (adapted from Lippens et al., 2012).

In contrast, within a globular, folded domain manifold local environments can be found (from different conformations within secondary structural elements through hydrogen bonds and hydrophobic centers to effective solvation on the protein surface) and therefore chemical shifts spread over a wide range of values, i.e. show broad "peak dispersion", as indicated in figure 9.

Resonance assignment

In a straight-forward and widely used method of resonance assignment several sets of standard, well-established three-dimensional NMR spectra on doubly-labeled (i.e., ^{15}N and ^{13}C isotope labeled) proteins are measured. These enable the identification of resonances of neighboring residues, as described for ERD14 in section 4.1.1. Other possibilities include the measurement of spectra of higher-dimensionality, spectra especially designed for IDPs (see below) and/or the use of computer-assisted automated resonance assignment programs (e.g. Lukin et al., 1997; Fiorito et al., 2006; Gossert et al., 2011; Alipanahi et al., 2011).

Spectra types designed especially for IDPs typically involve higher dimensionality (e.g. Bermel et al., 2012), alternative detection modes (detecting ^{13}C nuclei, e.g. Bermel et al., 2006; or $^{1}H^{\alpha}$ nuclei, e.g. Mäntylahti et al., 2011), sensitivity enhancement (for more rapid acquisition, e.g. using paramagnetic relaxation enhancement, Theillet et al., 2011), or connect more than two neighboring residues to facilitate the assignment process, e.g. the HNN-type spectra used in the present work (Chatterjee et al., 2005; Löhr et al., 2000; Löhr and Rüterjans, 1998; see also description in section 4.1.1).

It is noted, that in standard protein NMR experiments, proline resonances are typically missing, as they lack the amide proton usually used in signal detection. For folded proteins this fact is mostly ignored, but cannot be neglected for IDPs with a relatively high content of proline residues. Therefore techniques providing information on proline residues are of special importance in the NMR of IDPs. Such techniques include some of the HNN spectra, as well as the ^{13}C- and $^{1}H^{\alpha}$-detected experiments mentioned above.

Structural information based on chemical shifts

Following successful assignment, the structural information inherent to the chemical shifts can be assessed first. The evaluation of chemical shifts in terms of proteins is based on the observation that resonances from residues within the two main secondary structural elements found in proteins deviate in opposite directions from the average chemical shift of each amino

acid (Dalgarno et al., 1983). This deviation is called the secondary chemical shift (SCS) of an atom, and it is an average value over all sampled conformational states of the IDP; it therefore serves as an indicator of the amount of residual secondary structure at the given site (i.e., the percentage of molecules in that specific secondary structure at a given time point or the percentage of time a given molecule spends in that specific secondary structural state).

Several sets of reference ("random coil") chemical shift values have been published, and the choice of reference values is known to influence the evaluation of secondary chemical shifts (Wishart and Nip, 1998). This is especially relevant in the case of disordered proteins or regions whose chemical shift values are close to those of a random coil. Nevertheless, this issue has only recently been addressed in the literature (e.g. K. Tamiola, B. Acar, and F.A.A. Mulder, 2010b; Kjaergaard et al., 2011), as the in depth investigation of IDPs has moved more and more into focus within structural research. At the beginning of the present work, however, there was still no such comparison available. Therefore a detailed discussion and comparison of reference chemical shifts has been performed and is given in section 4.1.2 (including also recent publications).

Information on protein dynamics based on relaxation measurements

For an assessment of the flexibility and dynamics of a protein, several kinds of relaxation data (in the first instance R_1, R_2, $R_{1\rho}$ and hetNOE values) can be acquired, which give residue-specific information on local dynamics as well as on overall motions of the protein. R_1 and R_2 stand for the longitudinal (spin-lattice) and transversal (spin-spin) relaxation rates (i.e., the decay rates with which the respective magnetization components are lost or "dephased"), while hetNOE denotes heteronuclear ^{15}N-{^{1}H} NOE values. Low R_1, R_2 and hetNOE values are indicative of relatively high local flexibility, while high values indicate ordered (rigid) structure, as demonstrated in figure 10. All relaxation values are dependent on temperature (as is the dynamics of proteins) and field strength. For ordered proteins, several other parameters connected to local and global dynamics can also be calculated from these raw data, including the order parameter S^2 (cf. figure 10); however, these are either not applicable or only of little use for IDPs and will therefore not be discussed herein.

R_2 and $R_{1\rho}$ values are identical in most cases, though an observed difference can indicate the presence of two or more chemically exchanging states, e.g. the folded and unfolded states of a protein in continuous equilibrium. Multiple other NMR techniques can also be used for

determining further details of protein dynamics, but will not be listed or explained here (for an excellent review in the field see e.g. Palmer, 2004).

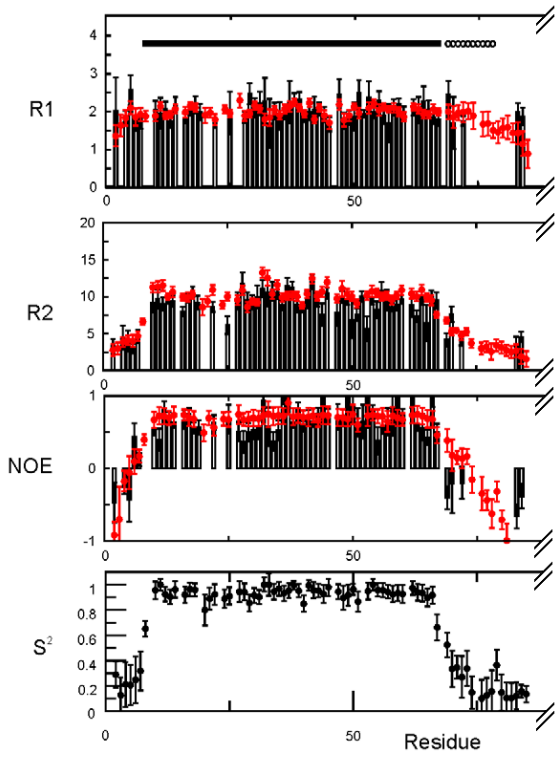

Figure 10. Example of NMR relaxation values of the ordered versus disordered domains of a protein. The higher R_1, R_2, hetNOE and S^2 values of the ordered LIM1 domain (marked by a straight bar at the top) and the lower values of the neighboring intrinsically disordered glycine-rich domain (marked by open circles at the top) of muscular LIM protein (MLP) at 295 K can clearly be distinguished. Black bars show relaxation parameters measured on the full MLP protein, while red dots indicate the very similar values measured on only the part of the protein shown herein (i.e. without the LIM2 domain, Schallus et al., 2009).

1.7. Structure and dynamics in a living cell: in-cell NMR

In-cell NMR is a unique technique that provides direct information on structure and dynamics, as well as on interactions and the function of proteins within the living cell (Tompa, 2009, chapter 8.3; Ito and Selenko, 2010). It is based on the observation of a selectively isotope labeled target protein – in front of the complex, but unlabeled and therefore basically "NMR-invisible" cellular background – by NMR. Main aspects and major challenges of this technique will be reviewed below and involve the introduction of the appropriately labeled target protein into an unlabeled, living cell, as well as possible problems of cell death or protein leakage during the experiment. Issues concerning reproducibility, protein concentration, and the sensitivity of the measurement also require special attention.

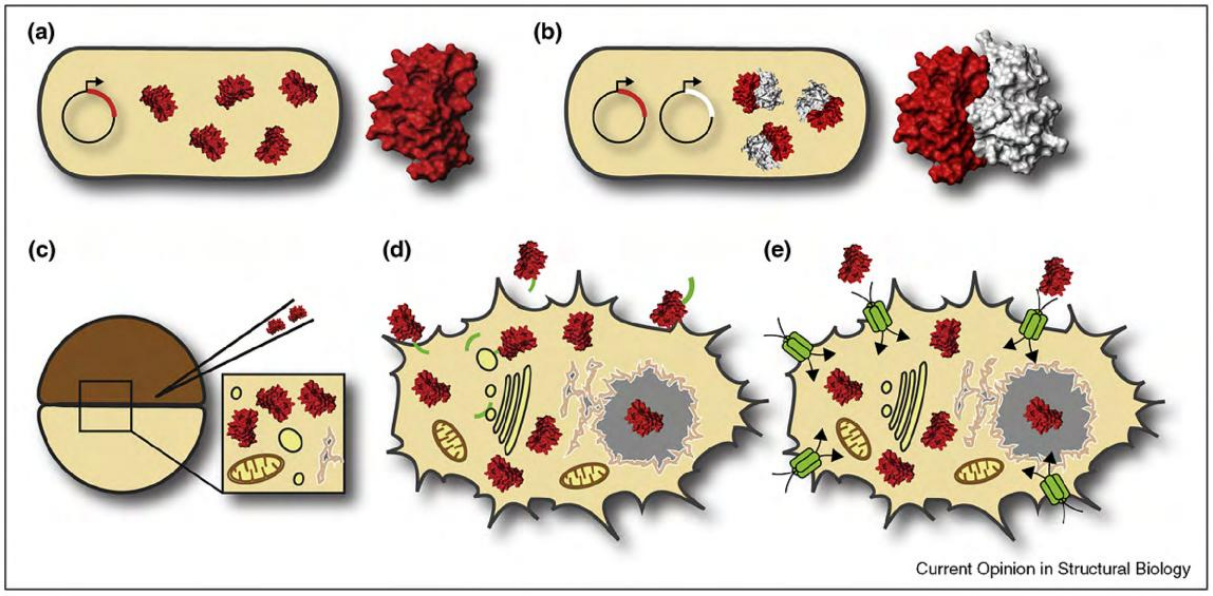

Figure 11. Overview of possibilities of delivery of selectively isotope labeled proteins into living cells for in-cell NMR. Possibilities include the direct overexpression of the target protein within E. coli cells using labeled growth media during induction (a), the coexpression of labeled target protein and its unlabeled binding partner in a dual-expression system in E. coli (b), the microinjection of a purified protein solution into large cells such as Xenopus laevis oocytes (c), or by delivery of purified protein solutions into mammalian cells using either a fused cell penetrating peptide tag (d, which is then cleaved off inside the cell) or pore-forming agents such as cell-permeabilizing toxins (e) (Ito and Selenko, 2010).

Selective isotope labeling for in-cell NMR

As shown schematically in figure 11, selective isotope labeling can be achieved by one of five methods (Ito and Selenko, 2010). The most straight forward method is direct over-expression of the target protein within *E. coli* cells (a and b). Here, selective labeling of the target protein is achieved by using unlabeled growth media for cell propagation before induction (and also during induction of the interaction partner in a dual-expression system in case b), combined with an exchange to labeled growth medium for the time of induction of the target protein (Serber et al., 2001; Burz et al., 2006b, 2006a).

Alternatively, an appropriately labeled and previously purified protein solution can be introduced into target cells by one of three approaches exploited so far. Microinjection can be used for large cells such as *Xenopus laevis* oocytes (c, e.g. Selenko et al., 2006; Sakai et al., 2006; Bodart et al., 2008; Thongwichian and Selenko, 2012). For the delivery of labeled protein into mammalian cells, either a fused or attached cell penetrating peptide tag (d, which is then cleaved off inside the cell, e.g. Inomata et al., 2009; Bekei, Rose, Herzig, Dose, et al., 2012) or pore-forming agents such as cell-permeabilizing toxins (e, e.g. Ogino et al., 2009; Bekei, Rose, Herzig, and Selenko, 2012) can be used.

Other challenges of in-cell NMR

Challenges of in-cell NMR mainly stem from introducing the selectively labeled target protein into the cell, as reviewed above, but also from the relative insensitivity of NMR, due to which relatively high protein concentrations are needed which may far exceed physiological levels (Ito and Selenko, 2010). However, in case of dehydrins, such as ERD14 investigated in the present work, a high protein concentration is rather close to physiological, because dehydrins are expressed in high levels upon stress treatment of plants.

Another critical issue is the need for rigorous control experiments to check/prove the in-cell status of the target protein(s), as in some cases the leaking of labeled target protein out of the cells has been observed (and two papers had to be retracted for this reason: Bryant et al., 2005, 2006; see also the retraction: Pielak, 2007). Therefore, sedimentation of the cells after measurement and the detection of potential leakage signals in the supernatant (by SDS-PAGE as well as by NMR) is important for in-cell NMR. If encountered, the leaking of cells can be corrected by alginate encapsulation (Li et al., 2008).

Furthermore, especially in the case of eukaryotic cells, the stability (i.e. integrity and viability) of cells during long measurement times, as well as the reproducibility of in-cell

spectra (i.e. eventual unexpected changes due to variations in the physicochemical nature of the intracellular environment) have to be addressed (Bodart et al., 2008).

Interpretation of in-cell NMR spectra

When comparing in-cell spectra to spectra of the same protein measured in dilute buffer, attention has to be paid to use a reference solution of a pH comparable to that inside the cell. This is especially relevant in case of *E. coli* cells, since the *E. coli* cytosolic pH values have been reported to be higher than for eukaryotic cells, i.e. between pH 7.6-7.8 (Slonczewski et al., 1981; Zilberstein et al., 1984).

The disappearance of in-cell NMR signals has been interpreted to be caused either by the high viscosity (or molecular crowding, see below) inside the cell or by an interaction of the target protein with large cellular partners (e.g. Bodart et al., 2008). Molecular crowding is expected to be an unspecific effect based on a simple competition for space due to the high concentration of other proteins and solutes within the cell (see also section 1.4, and Tompa, 2009, chapter 8.1). To distinguish between these two scenarios, molecular crowding is typically mimicked using model compounds and the resulting spectra are compared to the in-cell spectra for the differentiation of unspecific *versus* specific effects.

Possible model conditions for molecular crowding mimicry include high concentrations of either naturally occurring small molecules (e.g. sugars), average size model proteins (e.g. BSA), or large polymers such as dextran or ficoll 70 – at concentrations up to about 400 g/L (Tompa, 2009; in accordance with estimated overall solute concentrations within the cell, Zimmerman and Minton, 1993). In such *in vitro* experiments (without in-cell NMR measurements), two dehydrins were shown to retain their intrinsically disordered state even under severe crowding conditions (ERD10 and COR47, Mouillon et al., 2008). Similarly, other IDPs also stayed disordered under crowding conditions, with an eventual slight increase in secondary structure propensity at distinct sites (compare to the concept of PSEs, as discussed in sections 1.4.3 and 1.4.4, and Cino et al., 2012).

According to a recent review (Ito and Selenko, 2010), in-cell NMR measurements in *E. coli* have already been used for determining the structure of folded proteins (Sakakibara et al., 2009; Ikeya et al., 2010), demonstrating the disorder of an IDP, α-synuclein (McNulty et al., 2006; Croke et al., 2008), and measuring its dynamics in comparison to a folded protein (Li et al., 2008). Moreover, in case of another IDP, tau protein, both functional binding and *in vivo* phosphorylation could be detected within *X. laevis* oocytes (Bodart et al., 2008).

2. Aims

The goal of the present work was to investigate the dehydrin ERD14 from a structural and functional point of view, under *in vitro* as well as *in vivo* conditions.

Concluding from the literature available at the time of the beginning of the work, we hypothesized that the disordered state of the protein is relevant in fulfilling its function as a passively protective chaperone, while its conserved segments might play a role in binding to other proteins or membranes. To test this hypothesis the present work aimed to characterize in detail the disordered state of the protein (both *in vitro* and *in vivo*), to verify its chaperone function under *in vivo* conditions and to eventually determine its functionally important sequence regions.

For the structural characterization of ERD14, a thorough structural analysis by NMR under *in vitro* conditions was planned, including the expression of single (^{15}N) and double (^{15}N-^{13}C) labeled, purified ERD14 for the measurement, assignment, and structural as well as dynamic characterization of the protein. The sparse literature about the applicability of general interpretation methods of chemical shifts to IDPs (available at the time of beginning of the work) indicated a need for a detailed comparison of the relevant possibilities and the validation of their use on ERD14. In order to assess the structural state of ERD14 under *in vivo* conditions, in-cell NMR experiments were also planned, consisting of the preparation and measurement of samples of single (^{15}N) labeled ERD14 within living *E. coli* cells.

The functional analysis aimed to verify the known *in vitro* chaperone function of ERD14 also under *in vivo* conditions. To accomplish this, a novel *in vivo* chaperone assay was designed, using *E. coli* cells as an easy-to-handle, living model environment. The implementation of several different stress treatments (comparable to those abiotic stress effects that invoke dehydrin response in plants) was planned.

Furthermore, the preparation and investigation of several deletion mutants (lacking the conserved segments hypothesized to be important in function) was also planned for their comparison of structural and functional traits with respect to the wild type protein.

3. Materials and Methods

3.1. Buffers and basic procedures

NZYM growth medium:

10 g/L NZ-Amin, 5 g/L Yeast, 85.5 mM NaCl, 8 mM $MgSO_4$, 50 mg/L carbenicillin, pH = 7.5.

NZYM-agar plates:

15 g agar powder dissolved in 1 L of hot NZYM medium, with the antibiotics (carbenicillin) added only after the solution cooled down to below 40°C.

M9 minimal medium:

32 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 0.1 mM $CaCl_2$, 1 mM $MgSO_4$, 50 mg/L carbenicillin, 10 mL/L vitamin mix from Sigma-Aldrich (optional), 10 mL/L 100X trace elements solution (optional); autoclaved and then supplemented with 1 g/L $^{15}NH_4Cl$ (added as a powder), 2-4 g/L uniformly ^{13}C-labeled or unlabeled glucose·H_2O (added as a powder).

NMR-M9 buffer:

32 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.6 mM NaCl, 0.1 mM $CaCl_2$, 1 mM $MgSO_4$, 50 mg/L carbenicillin, 4 g/L unlabeled glucose·H_2O (added as a powder).

100X trace elements solution:

17 mM EDTA, 3 mM $FeCl_3$, 0.6 mM $ZnCl_2$, 0.08 mM $CuCl_2$, 0.04 mM $CoCl_2$, 0.16 mM H_3BO_3, 0.007 mM $MnCl_2$.

Lysis buffer:

50 mM TRIS, 150 mM NaCl, 1 mM EDTA, 0.4 mM DTE, pH 7.5; + Complete™ ETDA-free Protease Inhibitor Cocktail Tablet (Roche) just before use

Buffer “A12”:

50 mM TRIS, pH 7.5.

Buffer “B1”:

50 mM TRIS, 1 M NaCl, pH 7.5.

Buffer “A13”:

50 mM Na_2HPO_4, 150 mM NaCl, pH 7.5.

Preparation of competent cells:

Cells were grown O/N (over-night) from the appropriate glycerol stock, inoculated into 50 mL of fresh NZYM medium (using a 100x dilution), and grown at 37°C. At an OD_{600} value of about 0.4, cells were harvested by mild centrifugation (2700xg, 6 minutes, 4°C), gently resuspended by swirling with 5 mL of sterile buffer I (50 mM $MnCl_2 \cdot 4H_2O$, 13.5 mM $CaCl_2$, 30 mM $KCOOCH_3$, 13% glycerol, pH 5.8), and placed on ice for 20 minutes. Then cells were collected again by mild centrifugation (as above), gently resuspended by swirling with 1 mL of sterile buffer II (10 mM RbCl, 100 mM $CaCl_2$, 10 mM MOPS, 13% glycerol, pH 6.8) and placed on ice for at least 20 minutes. Aliquots of 100 μL were used within 2-3 hours or frozen in liquid nitrogen and stored at -70°C.

Transformation:

E. coli cells were transformed by adding 1 μL of DNA solution (miniprep) to 100 μL of competent cells. After 30 minutes on ice, cells were heat-shocked for 45 seconds at 42°C, cooled on ice for 2 minutes, and incubated with 200 μL of antibiotic-free NZYM medium at 37°C for one hour. Finally, cells were either plated onto NZYM-agar plates or inoculated into 5 mL of fresh NZYM medium and grown O/N.

Cell growth:

Cells were grown in liquid medium (NZYM) at 37°C while shaking at 250 rpm shaking, unless stated otherwise. Cells on agar plates (NZYM-agar) were always grown O/N at 37°C, without shaking.

3.2. Expression and purification of wild type ERD14

Plasmid Construct

For all experiments a previously prepared pET22b/ERD14 construct was used (Kovács, Kalmár, et al., 2008). For this construct, the ERD14 gene was amplified from the original

A. thaliana genome, the intron removed and the gene cloned into the pET22b vector from Novagen, using NdeI and NotI cloning sites. Additionally, a stop codon was introduced at the end of the sequence, so that the expressed protein would not contain any additional tags and/or amino acids, but only the sequence of ERD14 (UniProtKB entry number P42763, entry name ERD14_ARATH). DNA sequencing confirmed the correct sequence of ERD14.

Unlabeled Expression

For expression of unlabeled protein, the ERD14 construct was transformed into BL21 (DE3) Star cells, grown O/N in NZYM growth medium and frozen to -70°C with 10% glycerol to form a bacterial stock. For each expression, cells from this stock were grown O/N, then inoculated into fresh NZYM medium (using 100x dilution) and grown at 37°C. At an OD_{600} value of about 0.6 protein expression was induced with IPTG at a final concentration of 0.5 mM. After 3 hours of induction (at 37°C) cells were harvested by centrifugation (3400xg, 20 minutes, 4°C), resuspended in 20 mL lysis buffer / 500 mL growth medium and frozen until the day of purification (Kovács, Kalmár, et al., 2008).

Labeled Expression

For expression and purification of single (^{15}N) and double (^{15}N-^{13}C) labeled protein, the ERD14 construct was transformed into BL21 (DE3) Star pLysS cells, grown O/N in NZYM growth medium and frozen to -70°C with 10% glycerol to form a bacterial stock. For each expression, cells from this stock were grown O/N, inoculated into fresh NZYM medium (using 100x dilution), and grown at 37°C. At an OD_{600} value of about 0.4-0.5 cells were collected by gentle centrifugation under sterile conditions (2700xg, 20 minutes, RT), resuspended and grown further in fresh M9 minimal medium containing labeled ammonium chloride and labeled or unlabeled glucose as the only nitrogen and carbon source (using a quarter of the initial volume, according to the protocol of Marley et al., 2001). Protein expression was induced one hour later with IPTG at a final concentration of 0.8 mM. After O/N induction (at 37°C) cells were harvested by centrifugation (3400xg, 20 minutes, 4°C), resuspended in 40 mL lysis buffer / 500 mL M9 minimal medium, and frozen until the day of purification (Szalainé Ágoston et al., 2011).

Purification

The frozen cells were thawed, and their cell walls were disrupted by sonication on ice for 15 seconds, repeated 6 times with 45 seconds breaks in between. Cell debris was removed by centrifugation (3400xg, 20 minutes, 4°C), and proteins were purified from the supernatant.

Labeled and unlabeled ERD14 was purified in a 4-step procedure. First – as ERD14 is known to be heat stable – the supernatant of cell lysis was placed in a boiling water bath for 10 minutes, and all aggregated proteins were removed by centrifugation (1st step: 3400xg, 20 minutes, 4°C; 2nd step: 24000xg, 40 minutes, 4°C), and filtration (0,22 μm pore size syringe filter). As a preparation for ion-exchange chromatography, lysis buffer was exchanged to the salt-free chromatography buffer "A12" by using a GE HighPrep™ 26/10 Desalting column. Next, anion-exchange chromatography was performed on a GE MonoQ 5/50 GL column. After sample injection a washing step with buffer "A12" was performed, followed either by a linear gradient of 0-70% buffer "B1" (containing 1 M NaCl) in 25 column volumes or a one-step elution with 22% buffer "B1", corresponding to 220 mM NaCl. Finally, as the last step of purification, 5 mL aliquots of the protein solution were injected onto a semi-preparative sized gel filtration column (HiLoad™ 16/60 Superdex™ 200 prep grade from GE) and eluted isocratically with buffer "A13". Fractions were collected and purity was checked by SDS-PAGE gel electrophoresis. The final three steps (desalting, ion exchange and gel filtration) were performed on an ÄKTAExplorer 100 FPLC system from GE (formerly Amersham Pharmacia Biotech) under cooled conditions (about 12°C).

Finally, the ERD14-containing fractions were pooled, dialyzed against distilled water (using a dialysis membrane with a molecular weight cut-off of 3500 Da) and lyophilized. The final purity of the protein sample was assessed from the respective SDS- and native PAGE electropherograms, as shown in the results section 4.1.5, in figure 28. The final yield was about 14-18 mg/L full medium for unlabeled and 7-12mg/L minimal medium for labeled ERD14.

3.3. Preparation of mutant ERD14 proteins

Mutagenesis for deletion mutants

Conserved amino acid sequence regions – segments –from wild-type ERD14 were deleted one by one. The respective mutants (basic mutants) were named after the deleted region: conserved segments Ka (E24-P43), Kb (E113-D134), Kc (E156-G170), S (L74-S85) and

ChP (E87-K103) were deleted, as is shown in figure 13. Deletion mutants were prepared by PCR directly from the circular plasmid containing the WT ERD14 gene. Primers were designed so that they aligned with the same sequence regions on opposite DNA strands and already incorporated the desired mutation/deletion (figure 12 a). A short PCR program containing 18 cycles was used with *PfuTurbo*® DNA polymerase enzyme according to the following protocol:

PCR reaction mix: V = 50 µL

33 µL sterile water, 5 µL 10x buffer for *PfuTurbo*® polymerase, 2.5 µL forward primer (5 µM), 2.5 µL reverse primer (5 µM), 1 µL dNTP-mix (10 mM each dNTP), 5 µL template DNA (10x diluted miniprep), 1 µL *PfuTurbo*® DNA polymerase.

PCR program:

94°C/2 min – [94°C/30 sec – 57°C/60 sec – 72°C/4.5 min]$_{18x}$ – 72°C/10 min.

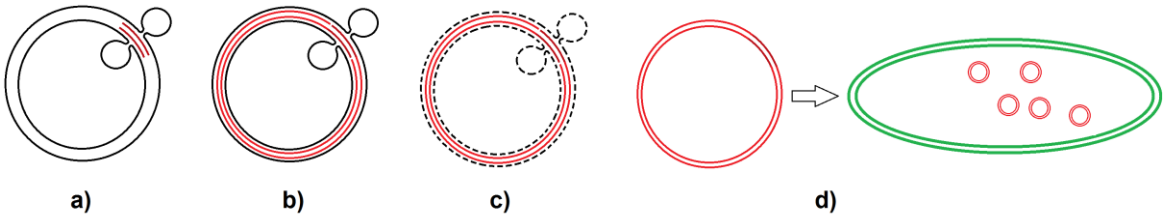

Figure 12. Concept of Mutagenesis (PCR of a plasmid with modified primers): a) Mutagenic primers align to the template DNA. The primers align to sequence parts before and after the region to be deleted, hereby causing a bulge in the original template, which will not be replicated. b) PCR reaction produces mutated, non-methylated DNA. It is noted that the polymerase used for this reaction stops replication when it has reached double stranded DNA (the annealed primer), that is, it fulfills exactly one round of replication around the template in each PCR cycle. In this manner the full, functional plasmid is replicated, together with the desired mutation. c) Methylated template DNA is selectively digested by DpnI enzyme. The mutated DNA (PCR product) differs from the original DNA template strand in not being methylated (E. coli cells produce methylated DNA, while PCR does not). Using this difference, pure target DNA (non-methylated PCR product) can be gained by selective digestion of methylated DNA by the enzyme DpnI. d) Target DNA is amplified by transformation into E. coli cells and O/N growth. It is noted, that at this point the plasmid (PCR product) transformed into the cells is not cyclic, but linear with blunt ends. This is corrected for automatically by the bacterium.

After the reaction 10 units of *DpnI* enzyme were added to 50 µL of PCR product and incubated for 1 h at 37°C (compare figure 12 b and c). 10 µL of this mixture were directly transformed into competent XL1-Blue cells and grown O/N on NZYM-agar plates (figure 12 d).

Resulting colonies were checked for the effectiveness of the mutation by PCR by using flanking primers and detection of the modified DNA fragment size on a high percentage (2%) agarose gel. Selected (positive) colonies were grown O/N in liquid NZYM medium, and the incorporated plasmids were extracted using NucleoSpin® Plasmid DNA purification kit of Macherey-Nagel ("miniprep"). The correct DNA sequence was verified by DNA sequencing, and the constructs were stored at -20°C. The double deletion mutants Kab, Kbc, Kac and the triple deletion mutant K3 (Ka, Kb and Kc deleted at the same time) were prepared following the same procedure, in two or three consecutive steps, respectively. See figure 13 for a sequence comparison of the different mutants.

```
CLUSTAL O(1.1.0) multiple sequence alignment

wtERD14     MAEEIKNVPEQEVPKVATEESSAEVTDRGLFDFLGKKKDETKPEETPIASEFEQKVHISEPE
ERD14Ka     MAEEIKNVPEQEVPKVATEESSA--------------------EETPIASEFEQKVHISEPE
ERD14Kb     MAEEIKNVPEQEVPKVATEESSAEVTDRGLFDFLGKKKDETKPEETPIASEFEQKVHISEPE
ERD14Kc     MAEEIKNVPEQEVPKVATEESSAEVTDRGLFDFLGKKKDETKPEETPIASEFEQKVHISEPE
ERD14S      MAEEIKNVPEQEVPKVATEESSAEVTDRGLFDFLGKKKDETKPEETPIASEFEQKVHISEPE
ERD14ChP    MAEEIKNVPEQEVPKVATEESSAEVTDRGLFDFLGKKKDETKPEETPIASEFEQKVHISEPE
ERD14K3     MAEEIKNVPEQEVPKVATEESSA--------------------EETPIASEFEQKVHISEPE
            ************************                     ******************

wtERD14     PEVKHESLLEKLHRSDSSSSSSEEEGSDGEKRKKKKEKKKPTTEVEVKEEEKKGFMEKLKE
ERD14Ka     PEVKHESLLEKLHRSDSSSSSSEEEGSDGEKRKKKKEKKKPTTEVEVKEEEKKGFMEKLKE
ERD14Kb     PEVKHESLLEKLHRSDSSSSSSEEEGSDGEKRKKKKEKKKPTTEVEVKE------------
ERD14Kc     PEVKHESLLEKLHRSDSSSSSSEEEGSDGEKRKKKKEKKKPTTEVEVKEEEKKGFMEKLKE
ERD14S      PEVKHESLLEK------------EEEGSDGEKRKKKKEKKKPTTEVEVKEEEKKGFMEKLKE
ERD14ChP    PEVKHESLLEKLHRSDSSSSSSE-----------------PTTEVEVKEEEKKGFMEKLKE
ERD14K3     PEVKHESLLEKLHRSDSSSSSSEEEGSDGEKRKKKKEKKKPTTEVEVKE------------
            ***********            *                 *********

wtERD14     KLPGHKKPEDGSAVAAAPVVVPPPVEEAHPVEKKGILEKIKEKLPGYHPKTTVEEEKKDKE
ERD14Ka     KLPGHKKPEDGSAVAAAPVVVPPPVEEAHPVEKKGILEKIKEKLPGYHPKTTVEEEKKDKE
ERD14Kb     ----------GSAVAAAPVVVPPPVEEAHPVEKKGILEKIKEKLPGYHPKTTVEEEKKDKE
ERD14Kc     KLPGHKKPEDGSAVAAAPVVVPPPVEEAHPV---------------YHPKTTVEEEKKDKE
ERD14S      KLPGHKKPEDGSAVAAAPVVVPPPVEEAHPVEKKGILEKIKEKLPGYHPKTTVEEEKKDKE
ERD14ChP    KLPGHKKPEDGSAVAAAPVVVPPPVEEAHPVEKKGILEKIKEKLPGYHPKTTVEEEKKDKE
ERD14K3     ----------GSAVAAAPVVVPPPVEEAHPV---------------YHPKTTVEEEKKDKE
                     **********************               ***************
```

Figure 13. Sequence of mutants versus WT ERD14 (multiple alignment using Clustal Omega) of Ka, Kb, Kc, S, ChP (basic mutants), as well as mutant K3. The conserved segments are highlighted in colors: K-segments (green), S-segment (yellow), and ChP-segment (red). It is noted, that the first K-segment (Ka) shows only about 30% of sequence similarity to the typical sequence of K-segments and therefore is not always designated as such.

Protein expression and purification

Mutants were expressed and purified as described for wild type ERD14 ("WT ERD14"), except that gel filtration (for purification purposes) was performed only where this was necessary (ERD14 Ka, Kb and S mutants). The purity of the purified protein samples (basic mutants) was assessed from the respective SDS- and native PAGE electropherograms shown in the results section 4.1.5 (figure 28).

3.4. General Analysis of Disorder

3.4.1. Gel electrophoresis

Wild type ERD14 and its basic mutants were compared by SDS-PAGE (Sodium dodecyl sulfate – polyacrylamide gel electrophoresis) as well as native gel electrophoresis.

SDS-PAGE was run under standard conditions on a 4% upper and 12.5% lower gel and stained with Coomassie Brilliant Blue R250 dye solution. Samples (15 µL) were boiled for 3 minutes with 5 µL sample cocktail prior to loading the gel. As a reference, 10 µL of Protein Marker Broad Range (2-212 kDa) from New England BioLabs was applied next to the samples.

Native gel electrophoresis was performed in a similar manner as SDS-PAGE, but by omitting SDS from the gel as well as from the electrophoresis solution. Samples of 15 µL were mixed with 5 µL of SDS-free sample cocktail, and were not (!) boiled prior to loading the gel. Protein markers were not used for native gel electrophoresis, as migration length is not only dependent on protein size, and compactness, but also on its charge state. Instead, the mutants were compared to each other and to the wild type protein.

3.4.2. CD spectroscopy

CD spectra of wild type ERD14 and its basic mutants were acquired under the following three conditions, all at 8 µM protein concentration:

100 mM Na_2HPO_4, 100 mM NaCl buffer at pH 6.2
100 mM Na_2HPO_4, 100 mM NaCl buffer at pH 8.8
100 mM Na_2HPO_4 buffer at pH 6.2

Measurements were performed on a JASCO J-720 spectropolarimeter at 25°C. Spectra were recorded from 200 to 250 nm at a resolution of 1 nm and a scanning speed of

50 nm/min. Values below 200 nm could not be measured, as absorption was too high to be measured reliably. In all cases three consecutively measured spectra were averaged, and background corrected by subtraction of the empty buffer spectrum. Molar ellipticity [$deg \cdot M^{-1} \cdot m^{-1}$] was calculated by dividing the measured ellipticity value [deg] by the protein concentration [mol/L] and the path length [m].

3.4.3. Size exclusion chromatography

Size exclusion chromatography was used to evaluate and compare the hydrodynamic properties of wild type ERD14 and its basic mutants. Protein samples (0.1 mg/mL, 100 μL) were applied one-by-one to an analytical sized gel filtration column (Superdex^{TM} 200 10/300 GL) and isocratically eluted with buffer "A13" at a flow rate of 0.5 mL/min. Elution volumes were compared to elution volumes of a variety of standard globular proteins (alcohol dehydrogenase, bovine serum albumin, ovalbumin, chymotrypsinogen, ribonuclease A and N-acetyl tryptophan amide – each at 0.2 mg/mL concentration in 100 μL). The effective hydrodynamic protein size of ERD14 proteins was estimated using a linear fit to the elution volumes of the standard proteins plotted against the logarithm of their molecular weights.

3.5. NMR experiments performed on purified, wild type ERD14

The following NMR experiments were performed on purified, wild type ERD14 in order to obtain detailed structural (from secondary chemical shifts) and dynamic (from NMR relaxation behavior) information on ERD14. For both, backbone assignment is essential, hence the first step as described below.

3.5.1. Experiments for backbone assignment

Lyophilized, ^{15}N,^{13}C-labeled WT ERD14 was dissolved in 330 μL of 10 mM MES buffer containing 10% D_2O and 1 mM DSS, and the pH was adjusted to 6.54. The final protein concentration was about 1 mM. Main measurements for resonance assignment were performed at 288 K on a Bruker Avance 800 MHz spectrometer (equipped with a CryoProbe) and consisted of a ^{1}H-^{13}C HSQC spectrum, of TROSY-based ^{1}H-^{15}N HSQC, HNCO, HN(CA)CO, HNCACB, HN(CO)CACB spectra and of three HNN-type spectra: TROSY-(H)N(CA)NH (straight-through, Chatterjee et al., 2005), TROSY-HN(CA)N (out&back, Löhr et al., 2000) and TROSY-HNN-COSY (out&back, Löhr and Rüterjans, 1998, performed on the ^{15}N-labeled sample used in the relaxation experiments described below). For the measurements of H^{α} shifts a ^{13}C-detected H(CA)CON experiment was

performed on a Bruker Avance 600 MHz spectrometer on the same ^{15}N,^{13}C-labeled sample about 1 month after the other experiments. Although the sample was stored at 4°C, changes in its ^{1}H-^{15}N TROSY-HSQC spectrum (measured as a control) were observed, as discussed in "Assignment of H^{α} shifts" in section 4.1.1.

Spectra were processed directly at the spectrometer (with Topspin software), referenced to internal DSS and analyzed using CARA, the Computer Aided Resonance Assignment program (Keller, 2004).

Secondary Chemical Shifts were calculated using several different sets of reference chemical shifts, and results were compared as described in section 4.1.2. Finally, the most appropriate set of random coil chemical shifts was determined to be from Kjaergaard et al. (Kjaergaard et al., 2011; Kjaergaard and Poulsen, 2011), but for comparison also SCS values calculated with random coil chemical shifts from Tamiola et al. (K. Tamiola, B. Acar, and F.A.A. Mulder, 2010b) and De Simone et al. (De Simone et al., 2009) are presented.

3.5.2. Relaxation experiments

Lyophilized, ^{15}N-labeled WT ERD14 was dissolved in 330 µL of 10 mM MES buffer containing 10% D_2O and 1 mM DSS, and the pH was adjusted to 6.54. The final protein concentration was about 1 mM. Relaxation measurements (R_1, R_2, $R_{1\rho}$ and ^{15}N-{^{1}H} NOE) were performed on Bruker Avance 600 MHz spectrometers with and without a CryoProbe. All spectra were collected at 288 K (the temperature was exactly calibrated with 99.8% deuterated (d_4-) methanol to a chemical shift difference of 1.639 ppm) and referenced to DSS.

For the determination of relaxation rates, resonance heights were extracted and exponentials were fitted to the decaying signals using Sparky (Goddard and Kneller). The ^{15}N-{^{1}H} NOE (hetNOE) ratios were calculated using signal intensities from NOE and reference spectra (I_{NOE}/I_{ref}). For error estimation within R_2 and $R_{1\rho}$ experiments, two data points were each measured twice, while for R_1 and hetNOE data the whole experiment was repeated.

3.5.3. Data deposition

The chemical shift assignments and all measured relaxation parameters of ERD14 in 10 mM MES buffer (containing 10% D_2O) at pH 6.54 and 288 K were deposited in the Biological Magnetic Resonance Data Bank (BMRB; http://www.bmrb.wisc.edu) under the BMRB accession number 16876.

3.6. In-cell NMR experiments and controls

The following in-cell NMR experiments were performed together with the corresponding control experiments in order to test for the structural state of ERD14 within the living cell.

3.6.1. Preparation of cells and acquisition of in-cell spectra

Cells were grown from the frozen ERD14/BL21 (DE3) Star pLysS stock over the day at 37°C (about 8 h, no visible growth yet), then cells were inoculated into fresh NZYM medium (100x dilution) and further grown O/N at 30°C to reach an OD_{600} of about 0.6-0.8 in the morning. Cells were collected by gentle centrifugation (under sterile conditions, 2700xg, 20 minutes, RT) and resuspended in fresh ^{15}N-labeled M9 minimal medium (without trace elements; using a quarter of the initial volume of full medium). After one hour of additional growth at 37°C, the first sample was taken ("-I") and the rest of the cells induced with IPTG at a final concentration of 1.6 mM. After three hours of induction at 37°C, another sample ("+I") was taken.

For both samples, 100 mL of bacterial culture were pelleted by gentle centrifugation (under sterile conditions, 3000xg, 20 minutes, 20°C), resuspended in 1.5 mL of NMR-M9 (M9 minimal medium without trace elements and without ^{15}N-labeled ammonium chloride, see section 3.1), followed by another gentle centrifugation step and resuspension in 800 µL of NMR-M9. Finally, 100 µL of D_2O were added to 1 mL of this very dense cell suspension and inserted into a standard NMR tube (figure 14). For both samples 1D (watergate), as well as 2D (^{15}N-HSQC) in-cell NMR spectra were acquired on a Bruker DRX 500 MHz spectrometer at 277 K. HSQC spectra were recorded with a resolution of 1024 and 256 measured time points and spectral widths of 4.691 +/- 5 ppm and 118.49 +/- 11.5 ppm in the ^{1}H and ^{15}N dimensions, respectively, with 4/8 (no induction) and 4/16 (induced) dummy/regular scans.

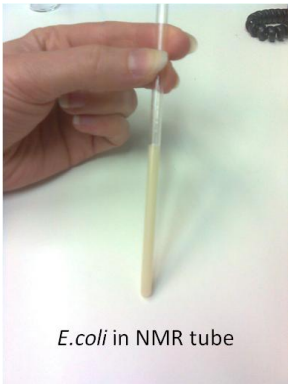

Figure 14. Photo of E. coli cells in the NMR tube just after the in cell NMR measurement.

3.6.2. Control experiments

All control experiments were performed on the same Bruker DRX 500 MHz spectrometer using the same parameters as were used for the in-cell samples, unless stated otherwise. All spectra were referenced to internal DSS.

For the most important control experiment ("supernatant", cf. section 1.7) the cell suspension was recollected from the NMR tube right after the in-cell measurements, and the cells were pelleted by strong centrifugation (13000xg, 2 minutes, RT). DSS was added, and the spectrum of the supernatant was acquired in exactly the same way as for the in-cell spectra, so as to give evidence of the true in-cell status of the measured protein (no leaking of the cells).

As further proof of the presence of the protein within the cell and its absence from the supernatant, at all stages of the in-cell experiment (in-cell spectra and supernatant spectra, before as well as after the measurements) gel electrophoresis samples were collected and SDS-PAGE was performed (as described in section 3.4.1). Samples were collected either by centrifugation of 1 mL of the bacterial culture (parallel to the in-cell samples) or by centrifugation of 20 µL of the bacterial in-cell sample after measurement. The cell pellet was dissolved in an amount of sample cocktail proportional to the amount of cells (OD_{600}x100 µL of sample cocktail, then application of 10 µL onto the gel). For supernatant samples, 18 µL of the supernatant were mixed with 6 µL of sample cocktail (the whole amount was applied to the gel).

In order to get an appropriate reference spectrum to the relatively high pH conditions inside an *E. coli* cell (about pH 7.6-7.8, see section 1.7), a pH titration of purified, ^{15}N-labeled WT ERD14 was performed. Spectra were acquired at 277 K at the following pH values: 6.58, 6.94, 7.29, 7.42, 7.57, 7.70. Next, to facilitate the correct assignment transfer, spectra were acquired at pH 6.58 at different temperatures: 288 K, 282.5 K and 277 K (see the "Assignment transfer to other experimental conditions" part in results section 4.1.1).

As a third control, to mimic unspecific crowding effects within the cell, different amounts of dextran (from Leuconostoc ssp, Fluka, MW ~70 000 Da) were added to a sample of purified WT ERD14 in NMR-M9 buffer at pH 7.79 and 277 K. The same sample was used as in the temperature and pH titration, but about 1 month after the other experiments, which resulted in the appearance of several minor peaks, but these did not further disturb the evaluation of the dextran titration. Spectra were recorded the same way as in-cell spectra at final

dextran concentrations of 0, 150, 300 and 400 g/L. A small pH shift of 7.79 to 7.63 was observed (but not corrected for).

3.7. Measurements of the *in vivo* chaperone effect of ERD14 (method development)

An *in vivo* chaperone assay was developed and performed as described below, to test for the supposed chaperone effect of ERD14 within the living cell.

3.7.1. Measurement of cell density

For the assessment of cell density, the absorbance (optical density) at 600 nm was measured in transparent 96-well plates in 200 µL sample volume, after 15 seconds of rigorous shaking, in a Synergy Mx monochromator-based multi-mode microplate reader (BioTek) – instead of the usual manual, one-by-one measurement in 1 cm cuvettes (which is referred to as the OD_{600} value). Each cell culture solution was measured directly as well as in several appropriate dilutions in order to stay in the linear range of the absorption measurement. A straight proportional fit to all values within the linear range and the slope (or calculated value at no dilution) was evaluated to be the absorbance value at 600 nm (Abs_{600}). It is noted, that this value is proportional to the OD_{600} value generally used to measure cell density (or cell growth), with the main difference being, that in this measurement – as it is performed in a 96-well plate with 200 µL sample volume – the path length is shorter than in the usual 1 cm cuvette used for OD_{600} measurements. The benefit of measuring the Abs_{600} value instead of the OD_{600} value is the easier and faster handling of high numbers of parallel samples (in the 96-well plate) and the more reliable absorbance values (because of the row of dilutions measured for each sample and the constant control of being in the linear range of the absorbance measurement).

Based on our observations, the two equivalent values can be converted by using the following empiric relation:

$$OD_{600} = 1{,}995 \times Abs_{600}$$

3.7.2. Overall procedure: measurement of the *in-vivo* chaperone effect

The chaperone effect of ERD14 within live cells was investigated by assessing its effect on cell viability under stress conditions. To this end, viability of ERD14-containing cells was

measured with and without different stress treatments, and compared to the viability of control cells (without ERD14), as described below.

Unless stated otherwise, an ERD14 construct (wt or mutant, within a pET22b plasmid) and an empty pET22b plasmid (control) were transformed into competent BL21 (DE3) Star cells and grown O/N in NZYM medium at 37°C. O/N cultures were used for a maximum of 7 days. On the day of measurement, three parallel cultures were grown from each sample (each 100x dilution into 5 mL of fresh NZYM medium) for 2 hours at 37°C, and then protein expression was induced with IPTG at a final concentration of 0.5 mM, for exactly 3 hours at 37°C. From each growth two samples were taken: one that underwent stress treatment ("stressed", see below in section 3.7.3), and another that was handled similarly to the stressed sample but without the actual stress effect ("non-stressed"). Additionally, for each sample, the cell density (absorbance at 600 nm, i.e. the Abs_{600} value, as described above) was measured.

After the stress treatment, the amount of viable cells in each sample ("stressed" as well as "non-stressed") was measured, as described in detail below in section 3.7.4.

The relative cell viability (%CV) after any stress treatment was calculated as the amount of viable cells in each "stressed" sample ($n_{stressed}$) relative to the amount of viable cells in the respective "non-stressed" sample ($n_{non\text{-}stressed}$) in percent:

$$\%CV = 100 * n_{stressed} / n_{non\text{-}stressed}$$

3.7.3. Applied stress conditions

Stress conditions were chosen to best mimic those abiotic stress conditions, which originally induce the expression of dehydrins (and of ERD14) within the plant, while cell viability of *E. coli* cells is reduced during these stress conditions by a measurable amount.

Freezing stress

Freezing stress was applied to cells by immersing 550 µL of bacterial cell culture in cryotubes into liquid nitrogen, followed by thawing in a 37°C water bath. In the testing phase, freezing time was varied between 2-6 minutes, and the effect of repeated freezing-thawing cycles (2 minutes freezing, 5 minutes thawing, repeated 1-2 times) was measured.

The final protocol involved 2 minutes of freezing, and 5 minutes of thawing for the stressed cells, while non-stressed cells were left at RT. Both stressed and non-stressed cells were then diluted 1:10 into fresh NZYM medium to a final volume of 4 mL, left 15 minutes

to rest at RT (relaxation time), and then further processed for the detection of cell viability, as described below in section 3.7.4.

As an additional test, freezing stress tolerance of control and ERD14-containing cells was measured immediately after diluting to different cell densities (Abs_{600} values between 0.2-0.7 for the control, and 0.5 for ERD14-containing cells), i.e. cells were grown as usual, diluted to the given values using fresh NZYM medium, stressed, and finally the cell viability was determined as described in section 3.7.4. Also, freezing stress tolerance of ERD14-containing cells was measured after different growth times, in order to test for relative cell viability values at different cell densities without the dilution step. Cells were grown for 0, 1 or 2 hours before induction and for another 2 hours thereafter. Final cell densities were measured (Abs_{600} values between 0.1-0.8), and the test results were compared.

High salt stress

High salt stress was applied to cells for 15 minutes, followed by a 1:100 dilution and a 15 minutes relaxation phase. Two salt concentrations were tested for effectiveness (2 M and 4 M NaCl). Given salt concentrations are final concentrations active during stress phase. Cells were handled at room temperature during all steps.

In the final protocol NaCl-enriched NZYM medium (enriched to a NaCl concentration of 4 M) was added in a 1:1 ratio to the cell suspensions to finally give an active salt concentration (effective stress) of 2 M NaCl. For the non-stressed cells, regular NZYM medium was used. Cells were stressed for 15 minutes, then diluted in one step (1:100 ratio) into fresh NZYM medium and left 15 minutes to rest (relaxation phase). Then the suspensions were further processed for the detection of cell viability, as described below in section 3.7.4.

As an additional test, high salt stress tolerance of control cells was measured at different cell densities, i.e. cells were grown for 0.5, 1, 1.5, 2 and 2.5 hours before induction and for another 3 hours thereafter. Final cell densities were measured (Abs_{600} values between 0.1-0.9) and the test results were compared.

Dehydration stress

Dehydration stress was applied to cells through the osmotic pressure of glycerol. Cells were diluted at a ratio of 1:20 into glycerol/H_2O solutions, stressed for given amounts of time, followed by a 1:10 dilution into fresh NZYM medium, and a 15 minutes relaxation phase.

Then the cell suspensions were processed for the detection of cell viability, as described below in section 3.7.4.

Three different stressing times (20, 30 and 45 minutes) and two different glycerol concentrations (6 and 7 M, corresponding to 30 and 40 MPa osmotic pressure, respectively) were tested for effectiveness, while the same "stressing" times and physiological osmotic pressure (0,5 M glycerol) were used for non-stressed cells. Cells were handled at room temperature during all steps.

3.7.4. Detection of cell viability

As an effect of the introduced stress treatment some of the cells are expected to die, while others (just a small number for the control cells) should stay viable. It is assumed, that the level of ATP (Adenosine-5'-triphosphate) within a viable cell is constant, therefore the number of surviving (viable) cells is expected to be proportional to the amount of ATP present within the cells. By the use of the BacTiter-Glo^{TM} Microbial Cell Viability Assay (from Promega) the amount of ATP present can be easily detected in a quantitative way based on the ATP-dependent luminescence activity of the included luciferase enzyme.

However, due to the stress treatment some of the cells are expected to die, and ATP from these cells might be released into extracellular space, which could interfere with the subsequent cell viability measurement. Therefore, to remove any disturbing chemicals, such as the mentioned ATP or high amounts of remaining stressors (e.g. NaCl or glycerol), cells were collected by gentle centrifugation (2700xg, 6 minutes, RT) and then resuspended in fresh NZYM medium before the actual measurement. Following the freezing stress treatment cells were resuspended in the same volume of NZYM medium as they have been before centrifugation. However, to concentrate the samples following the high salt and dehydration stress treatments, in these cases less NZYM medium was used for resuspension (1/10th of the initial volume before centrifugation).

For each sample, 100 µL of the resulting cell suspension were mixed with 100 µL of BacTiter-Glo^{TM} reagent in wells of a black colored 96-well plate, and luminescence was measured on a Synergy Mx monochromator-based multi-mode microplate reader (BioTek) in a time-dependent manner at 30°C. Luminescence (lum) values measured at different time points were averaged (2.5, 5 and 7.5 minutes after addition of the reagent, unless stated otherwise), and background-corrected by subtracting the luminescence value measured for 100 µL of fresh, sterile NZYM medium with 100 µL of BacTiter-Glo^{TM} reagent. Finally, as lu-

minescence intensity was assumed to be directly proportional to the number of viable cells in the sample, luminescence values of stressed cells ($n_{stressed}$) were directly related to luminescence values of non-stressed cells ($n_{non\text{-}stressed}$), to calculate the relative cell viability %CV, as described above in section 3.7.2.

3.7.5. Measurement of the effects of long-term stress conditions on cell growth

As an alternative to the newly developed *in vivo* chaperone assay described above, the effect of different stress conditions on cell growth was assessed in the presence and absence of ERD14.

Effect of induction on cell growth

Induction is a special form of stress effect for *E. coli* cells, which in many cases results in reduced cell growth by focusing all efforts of the cell on the expression of the target protein. The effect of induction on cell growth was measured and compared for control, wild type protein and the K3 mutant of ERD14.

The ERD14 and control (empty pET22b vector) plasmids were transformed into competent BL21 (DE3) Star cells and grown O/N in NZYM medium at 37°C. The following day, three parallel cultures were grown from each O/N culture (each at 100x dilution into 5 mL of fresh NZYM medium) for 2 hours at 37°C, then protein expression was induced with IPTG at a final concentration of 0.5 mM, and cells were grown further at 37°C. Cell growth was detected every hour by measuring cell density (Abs600 value, as described in section 3.7.1).

Effect of long-term salt exposure on cell growth

The effect of long-term salt exposure – at much lower levels, than used in the short-term stress analysis – was measured by following cell growth in NZYM medium of slightly elevated salt concentration (after induction).

Three parallel cultures were grown from O/N cultures (the second day after transformation) of cells containing either an ERD14 construct or an empty pGEX5x plasmid (each 100x dilution into 5 mL of fresh NZYM medium) for 2 hours at 37°C. Then protein expression was induced with IPTG at a final concentration of 0.5 mM. Cells were diluted 1:20 into fresh NZYM medium of either elevated or normal salt concentration (to reach a final concentration of 171.2 or 85.6 mM NaCl, respectively). Cell growth was followed by the measurement of Abs600 values in a Synergy Mx monochromator-based multi-mode micro-

plate reader (BioTek) during continuous rigorous shaking and incubation at 37°C within the plate reader (using cell cultures of 200 µL volume in a 96-well plate, closed against evaporation with a self-adhesive film (BrandTech Polyester Sealing Film 781391), that was perforated at four points per well in order to enable the air exchange and oxygen supply that is needed for bacterial growth).

Finally, salt tolerance during growth was calculated as the ratio of growth rates with and without salt treatment and expressed in percent values. Growth rates were determined from the time needed to reach an Abs_{600} value of 0.55, as the time to reach the same optical density is proportional to the growth rate of the bacteria at the given conditions for cell suspensions of equal concentration at the first time point (as in this case). This kind of evaluation was observed to be more robust than the determination of growth rates from a linear fit to the exponential phase of growth.

Selection of the appropriate assay reference for cell growth measurements

As discussed in section 4.2.5, cells containing only the empty pET22b plasmid (that were used as a control in all other experiments) do not grow further upon induction. Therefore, for the observation of the effect of salt on the cell growth after induction, another reference had to be introduced. For this purpose, GST-containing cells (glutathione S-transferase, expressed from an empty pGEX5x vector) were used, as pGEX5x-containing cells continue exponential growth even after induction. Furthermore, GST is a globular protein that is expressed upon induction, but not expected to interfere with the actual measurements (i.e. is not expected to have a protective effect on the cells).

4. Results and Discussion

4.1. Structural features of ERD14

Purified, wild type ERD14 was investigated in order to obtain detailed structural (from secondary chemical shifts) and dynamic (from NMR relaxation behavior) information. In-cell NMR of wild type ERD14 was performed in order to obtain information on the structural state within the living cell. For all these, backbone assignment was essential, and therefore the first step, as described below.

Furthermore, the basic mutants of ERD14 were subjected to a general analysis of disorder and their overall structural properties were compared to those of the wild type protein.

4.1.1. The basis for NMR: Resonance assignment

The HSQC spectrum of ERD14 (figure 15) contains all general features typical for intrinsically disordered proteins, such as the narrow signal dispersion in the ^{1}H dimension and the relatively narrow and intense signals compared to the size of the protein. Also, a high amount of signal overlap can be observed, especially in the central region of the spectrum.

Resonance assignment was finally achieved by a combined method using (i)-(i-1) connectivities from standard 3D assignment spectra (HNCO-HN(CA)CO and HN(CO)CACB-HNCACB spectra pairs) combined with (i+1)-(i)-(i-1) connectivities from three HNN-type spectra, namely the (H)N(CA)NH, HN(CA)N and HNN-COSY spectra.

Starting points

Starting points for the assignment process were the C-terminal residue (characteristic H^N, N and C^0 chemical shifts, because of the negative charge at the carboxyl group (deprotonated at pH 6.5) at the end of the protein backbone), residues A, G, S and T (characteristic N, C^α, C^β and C^0 chemical shifts), glycines and residues neighboring G or P (characteristic pattern in the (H)N(CA)NH spectrum, see further below), as well as combinations of the above mentioned. Table 2 lists examples of starting points used within the assignment of ERD14. It should be noted that the amino acid cysteine has characteristic chemical shifts as well, but is not contained in ERD14.

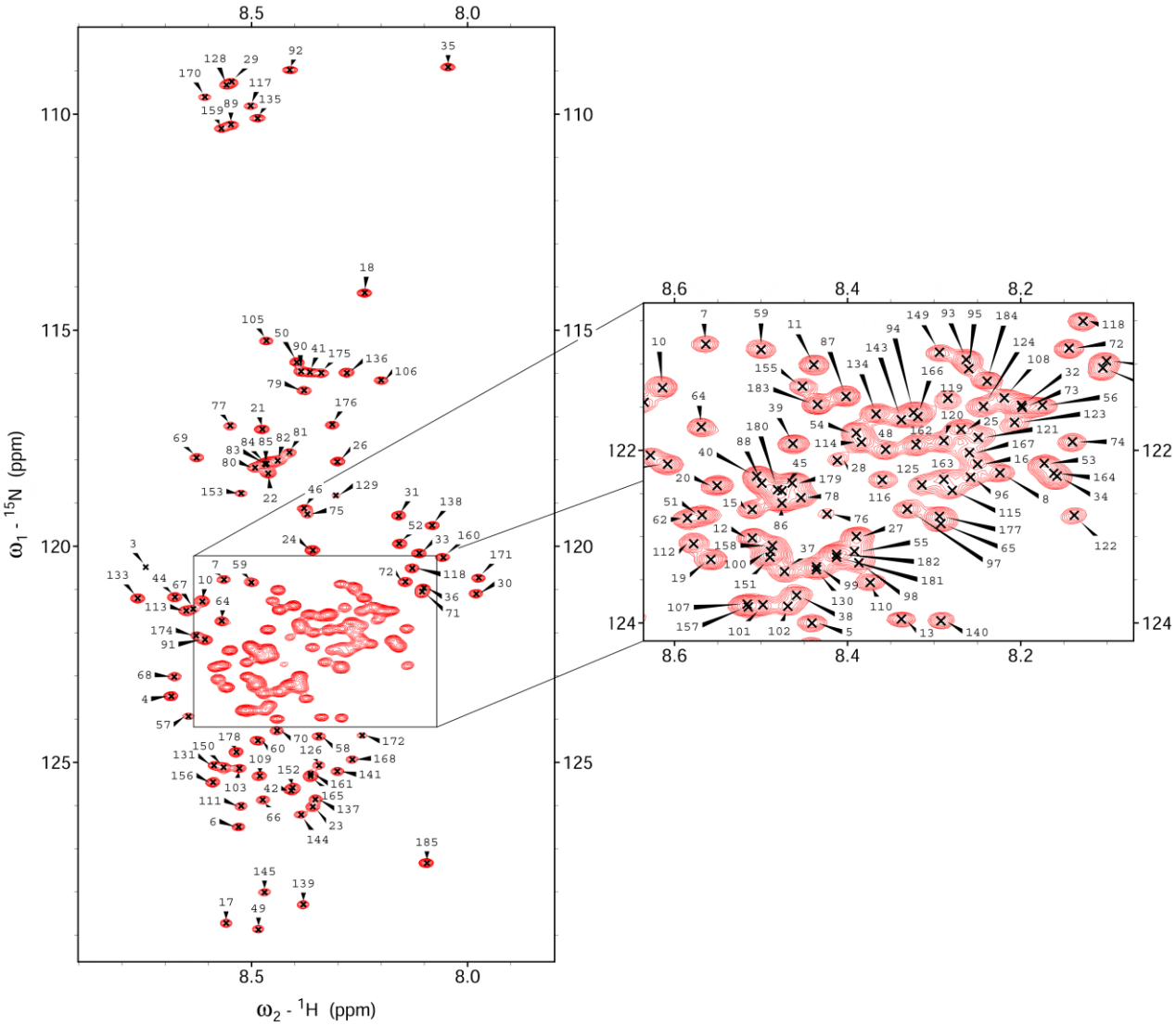

Figure 15. Assigned HSQC spectrum of ERD14 in 10 mM MES buffer, pH 6.54, 288 K.

C-term.	S/T next to S/T	S/T next to A	S/T next to G	G next to P	other special neighbors of P
E-185	20-ESSAE-24 103-KPTTE-107 173-PKTTV-177	16-VATE-19 47-PIASE-51	88-EGSDGE-93 134-DGSAV-138	127-PGHKKP-132 169-PGYHP-173	43-PEETP-47 61-PEP-63 151-EAHP-154 139-AAAP-142

Table 2. Examples of starting points used for the assignment of ERD14. Amino acids with characteristic chemical shifts are marked in colors. Glycine, as well as residues next to glycines or prolines also give characteristic patterns in the (H)N(CA)NH spectrum.

Proceeding backbone assignment

An example for the proceeding assignment along the sequence of ERD14 is depicted in figure 16 and figure 17, showing representative slices of all assigned spectra with horizontal

lines connecting signals belonging to the same residue and atom, respectively. The strips of the spectral pairs for standard 3D assignment in figure 16 (HNCO-HN(CA)CO and HN(CO)CACB-HNCACB) are shown interleaved, so that for each strip resonances of both spectra can easily be compared to each other. The spectra HNCO and HN(CO)CACB – colored yellow in the background – show only the C^{O} and C^{α}-C^{β} resonances of the preceding amino acid (i-1), respectively, while the HN(CA)CO and HNCACB spectra – white background – show both, resonances of the preceding (i-1) and of the present (i) amino acid.

Figure 17 demonstrates for P47, how the use of HNN-COSY and HN(CA)N spectra results in a continuous backbone assignment even with proline residues (blue circles). In this manner, N chemical shifts of 15 out of 16 proline residues were determined, and the correct assignments of the neighboring fragments were confirmed. As such, the only breakpoint in the final backbone assignment of ERD14 is the sequence 146-PPP-148, where none of the afore mentioned experiments could give information on the middle proline, P147.

Figure 17 also shows that the three HNN-type spectra have similar, but not quite equal informational content. In the upper two spectra – (H)N(CA)NH and HNN-COSY – diagonal peaks are positive, and cross-peaks are negative, however in the (H)N(CA)NH spectrum this pattern is disturbed in a specific manner at and next to glycine residues (see in more detail below and in figure 18). However, this is on the one hand of great value within the assignment process, but on the other hand this spectrum does not show proline resonances (but also note that spectral width could be set more tightly, so resolution is higher in the vertical dimension). In the third (lower) spectrum – HN(CA)N – all peaks are positive. Here, the noise levels compared to the average cross-peak intensities are higher than in the HNN-COSY spectrum. Nevertheless the spectrum turned out to be of great use, where HNN-COSY cross-peak resonances were very weak, e.g. in the case of E51 (last strip in figure 17).

As mentioned above, in the (H)N(CA)NH spectrum specific positive/negative patterns are expected for the diagonal- and cross-peaks of glycines and neighboring residues, as well as residues next to prolines (Chatterjee et al., 2005). However, in the case of two residues within ERD14 – L30 and I160, both preceded by a glycine – the actual measured resonances do not match the expected pattern, as shown in figure 18 (blue circles). Other neighbors of glycines within ERD14 exhibit the theoretical pattern, as shown e.g. for Y171 (first strip, yellow circle). A possible explanation of the observed deviation could be that these amino acids have unusual coupling values; however, this hypothesis was not examined any further.

While the assignment process was straight-forward within some regions of ERD14 (e.g. the sequence shown in figure 16), other parts required great care to resolve overlapping resonances of similar and repetitive sequences, such as a repetitive part of the S-segment (7 consecutive serines, S79-S85) and the regions of high similarity shown below:

```
 36      KKKDE              40
 93   EKRKKKKEKKK          103 - part of the ChP-segment
178   EEEKKDKE             185
111 KEEEKKGFMEKLKEKLPG    128 - part of the second K-segment (Kb)
156    EKKGILEKIKEKLPG    170 - part of the third K-segment (Kc)
```

The combined use of all measured spectra and great care taken during the assignment process enabled the unambiguous assignment of almost all H^N, N, C^α, C^β and C^0 resonances.

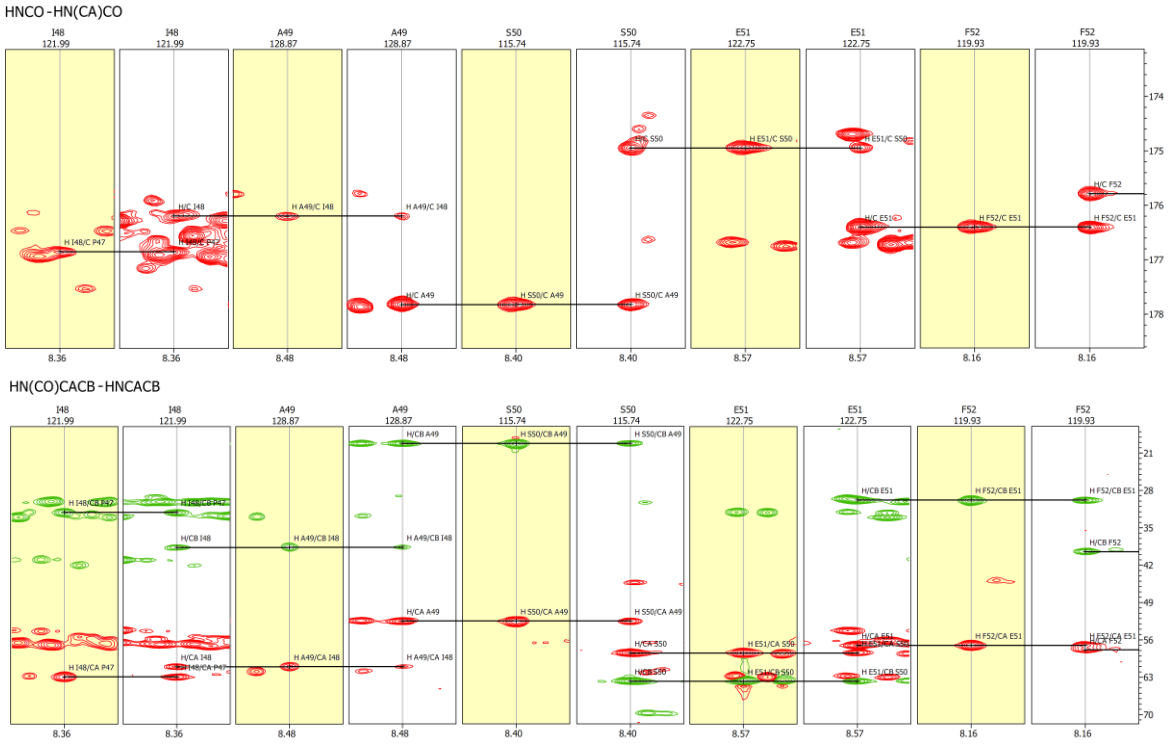

Figure 16. Example for proceeding resonance assignment through (i)-(i-1) connectivities observed in the HNCO-HN(CA)CO and HN(CO)CACB-HNCACB spectra pairs for a region of ERD14: 48-IASEF-52. The region is preceded by a proline. Horizontal lines connect signals belonging to the same residue and atom in the different strips/spectra. The strips of each spectrum pair are shown interleaved, so that for each strip resonances of both spectra can easily be compared to each other. The spectra HNCO and HN(CO)CACB – yellow background – show only the C^O and C^α-C^β resonances of the preceding amino acid (i-1), respectively, while the HN(CA)CO and HNCACB spectra – white background – show both, resonances of the preceding (i-1) and of the present (i) amino acid.

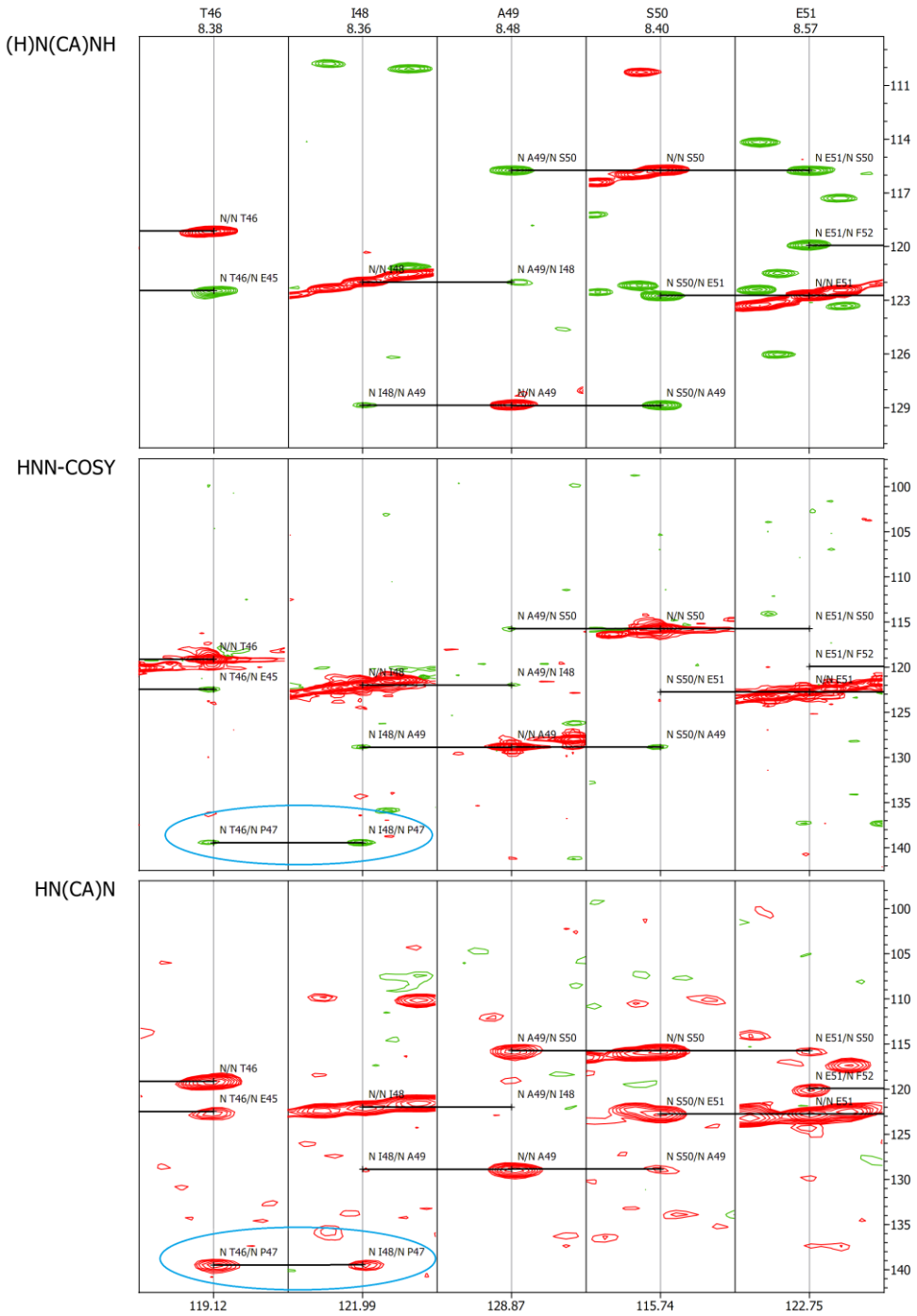

Figure 17. Example for proceeding resonance assignment through (i+1)-(i)-(i-1) connectivities observed in the HNN-type spectra for a region of ERD14: 46-TPIASE-51. Horizontal lines connect signals belonging to the same residue in the different strips. The example clearly shows, how through the HN(CA)N and HNN-COSY spectra the assignment can be bridged over proline residues: cross-peaks for P47 can be seen in both neighboring strips, T46 and I48, respectively (blue circles). In contrast, in the (H)N(CA)NH spectrum no cross-peaks for prolines are expected (and therefore spectral width are set narrower in this dimension).

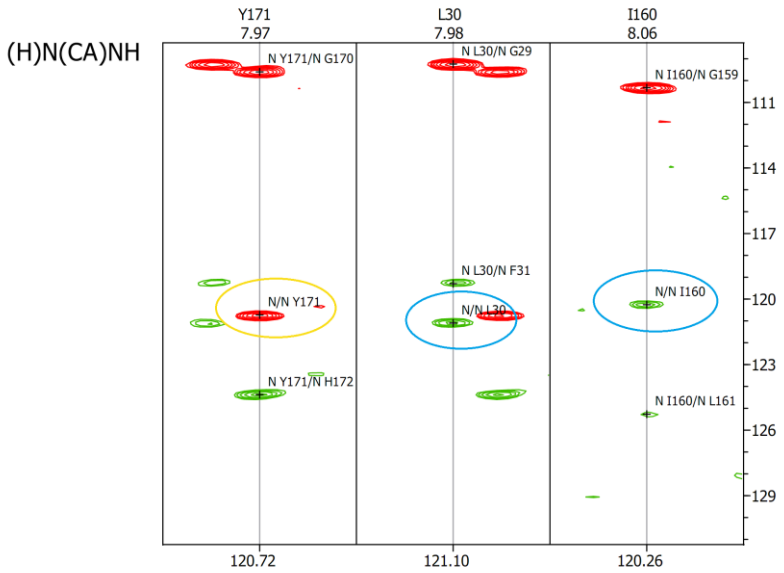

Figure 18. Expected and deviating resonance patterns in strips of the (H)N(CA)NH spectrum for residues following glycine. Within ERD14, 7 out of 9 residues (after glycine) conform to the expected positive/negative pattern, as shown here for Y171 (first strip; note especially the positive diagonal peak in the yellow circle). However, the observed pattern of resonances in the strips of L30 and I160 (second and third strip, respectively) deviates from the expected, as the corresponding diagonal peaks are negative (blue circles).

Assignment of H^{α} shifts

For the assignment of H^{α} shifts a ^{13}C-detected H(CA)CON experiment was recorded. This experiment acquires chemical shifts of H^{α} and C^{O} nuclei at position (i-1) and N at position (i). The advantage of ^{13}C-detection (on C^{O} in this case) is that proline residues are also detected, and therefore H^{α} chemical shifts of residues preceding prolines are also gained.

Unfortunately, as the H(CA)CON spectrum was acquired on the same double-labeled sample one month after the other experiments, changes in the spectra of the protein could be observed. Peaks of residues P132-A137 and V143-V145 have shifted in the ^{15}N-HSQC spectrum possibly due to spontaneous isoaspartate-formation of D134 (Aswad et al., 2000), so H^{α} shifts of these residues had to be excluded from further analysis.

In addition, resonances of fully overlapping peaks could not be resolved unambiguously, as no other, comparable source of information was available in this regard. However, partial overlap could be resolved in some cases by an exact comparison of line shapes of the overlapping residues.

It should be noted, that with the use of the H(CA)CON spectrum CO and N resonances of P146 and P147 could also be assigned, but these were excluded from further analysis together with the regions mentioned above, due to being potentially involved in the chemical shift changes observed for the preceding amino acids V143-V145.

Final assignment result

Finally, for almost all residues (180 out of 185) H^N, N, C^α, C^β and C^0 chemical shifts could be assigned, including all but two proline residues, P146 and P147, as they are both followed by another proline. The first two residues could not be detected, and E3 was too weak to assign more than its H^N and N shifts. Furthermore, 149 H^α shifts could be assigned.

The final H^N-N assignment is shown in figure 15, while a full list of annotated resonances (at 288 K and pH 6.5) can be found in addendum 6.1.

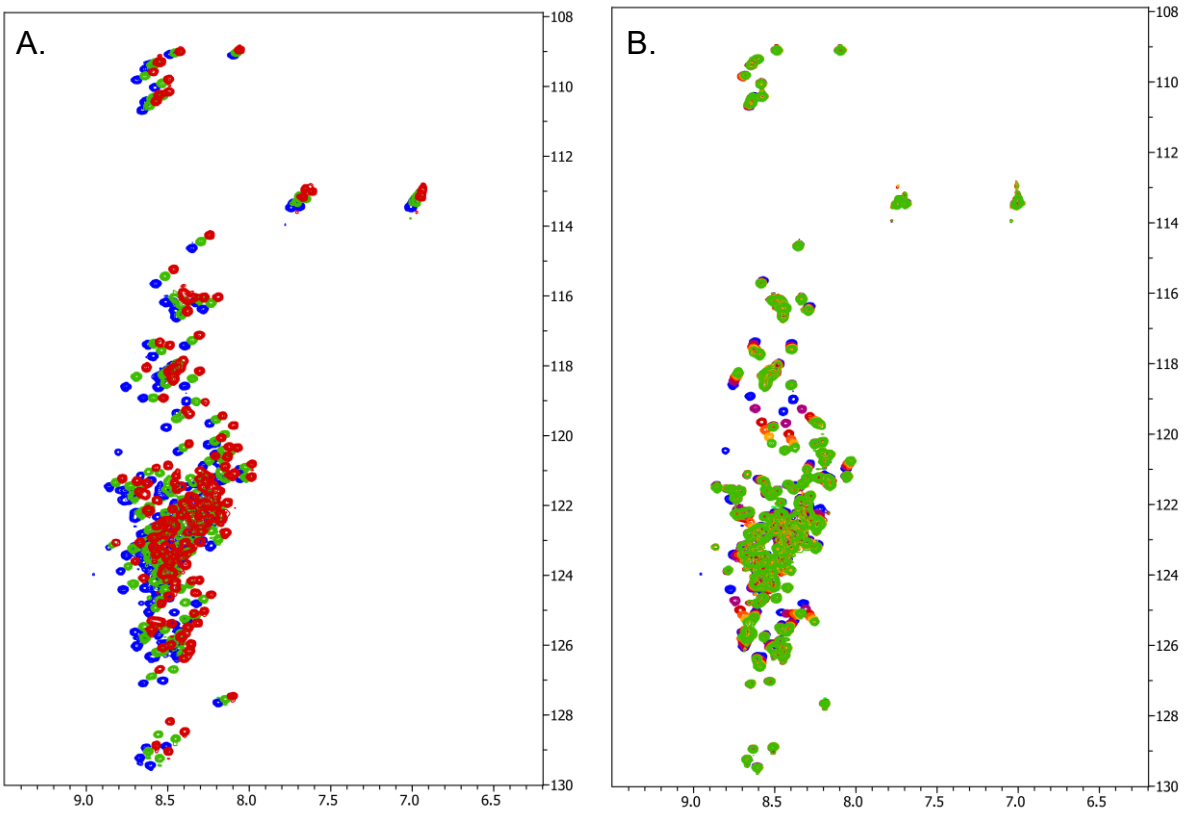

Figure 19. Temperature and pH dependence of the HSQC spectrum of WT ERD14. A. Overlay of spectra measured at constant pH (6.58), at different temperatures: 277 K (blue), 282.5 (green) and 288 (red). B. Overlay of spectra measured at constant temperature (277 K), at different pH values: 6.58 (blue), 6.94 (violet), 7.29 (dark red), 7.42 (orange), 7.57 (gold), 7.70 (green).

Assignment transfer to other experimental conditions

Finally, the H^N-N assignment was first transferred at a constant pH of 6.5 from 288 K over 282.5 K to 277 K and then transferred at a constant temperature of 277 K from pH 6.5 to pH 7.7, as shown by the overlay of HSQC spectra in figure 19. The HSQC spectrum at 277 K and pH 7.7 was used as a reference for the in-cell spectra (see also section 4.1.4).

4.1.2. Evaluation of chemical shifts: choosing the correct reference

Several "random coil" chemical shift reference sets for secondary chemical shift calculations have been used on the assigned resonances of ERD14 and will be compared herein. Table 3 gives an overview of all used reference sets, their origin and comments on their usefulness on ERD14, i.e. in general: on intrinsically disordered proteins measured under near native conditions.

The random coil chemical shifts published by the Wüthrich group in 1978-79 (Richarz and Wüthrich, 1978; Bundi and Wüthrich, 1979) represent the measured chemical shifts of all 20 amino acids (X) within the peptide GGXA. These and similar short peptides are not expected to form any specific secondary structure, enabling their use in the determination of random coil chemical shifts, as was done in several other studies thereafter. Used with ERD14, resulting secondary chemical shifts contain several outliers and are overall "noisy", indicating that the chemical shift of a nucleus is not only determined by its secondary structure type and the identity of the amino acid it is located in, but it is also influenced by the identity of the sequentially neighboring amino acids (especially proline, see comment "XP correction needed"), as well as the proximity of charges at the side chains (charged residues) and at the termini of the peptides (and further effects, as discussed below). Further analyses therefore mostly used peptides with capped termini and took into account the identity of the neighboring amino acids. Proline sterically restricts the free random coil movement of its preceding residue, and therefore it has a rather big effect on its chemical shift compared to other neighboring residues, hereby causing huge outliers within the SCS values, as shown in figure 20 A. This effect is typically accounted for by the use of correction terms for residues followed by proline (XP correction). The use of reference chemical shift sets with neighbor correction for all kinds of neighboring residues (not only proline) – e.g. as for N chemical shifts (YX correction, Braun et al., 1994) or for chemical shifts of all backbone nuclei (ABXCD correction, Schwarzinger et al., 2001) – generally results in less "noise" within the SCS values (not shown). Within recent publications (De Simone et al., 2009; K. Tamiola, B.

Acar, and F.A.A. Mulder, 2010b), neighbor correction terms have been included even in the calculation of statistically derived random coil chemical shifts. However, in SCS values calculated with random coil values from De Simone et al. (where weighted AXB neighbor corrections are used), still small outliers can be observed for amino acids preceding proline, indicating that the presented correction values might underestimate the effect of prolines.

Other effects that can disturb chemical shift analysis of proteins are the effects of pH and of temperature. A difference in pH between the analyzed experimental values and the used reference set of chemical shifts results in (more or less strong) outliers for the charged amino acids Glu, Asp and His, as is shown in figure 20 B. This effect is most obvious in the present example with the reference random coil values from Schwarzinger et al. that were measured at pH 2.3 (Schwarzinger et al., 2000). This pH was chosen with the legitimate goal of being comparable to chemical shifts of acid-denatured globular proteins, but is not appropriate when compared to chemical shifts measured at near-neutral pH values. For a correct pH correction, some reference chemical shift sets include chemical shift values for the protonated, as well as for the deprotonated forms of the charged amino acids, based on pH titrations, together with the respective pK_a values (e.g. Braun et al. and Kjaergaard et al.).

The effect of a temperature difference between experimental and reference chemical shifts is less dominant, and does not result in sequence-specific outliers, but rather in a general bias of the SCS values. This can be accounted for using temperature coefficients (determined e.g. by Merutka et al., 1995; or by Kjaergaard et al., 2011). Temperature correction is especially important when not only the fully formed secondary structure is estimated from the sign of SCSs, but also the extent of partially formed secondary structural elements within disordered regions from the sign *and* value of SCSs.

Furthermore, the basis of any SCS analysis is the careful and correct referencing of the measured chemical shifts, including the reference chemical shifts. In the present analysis all chemical shifts values are referenced either directly or indirectly to DSS or for older sets of reference chemical shifts corrected to be equivalent to DSS calibration.

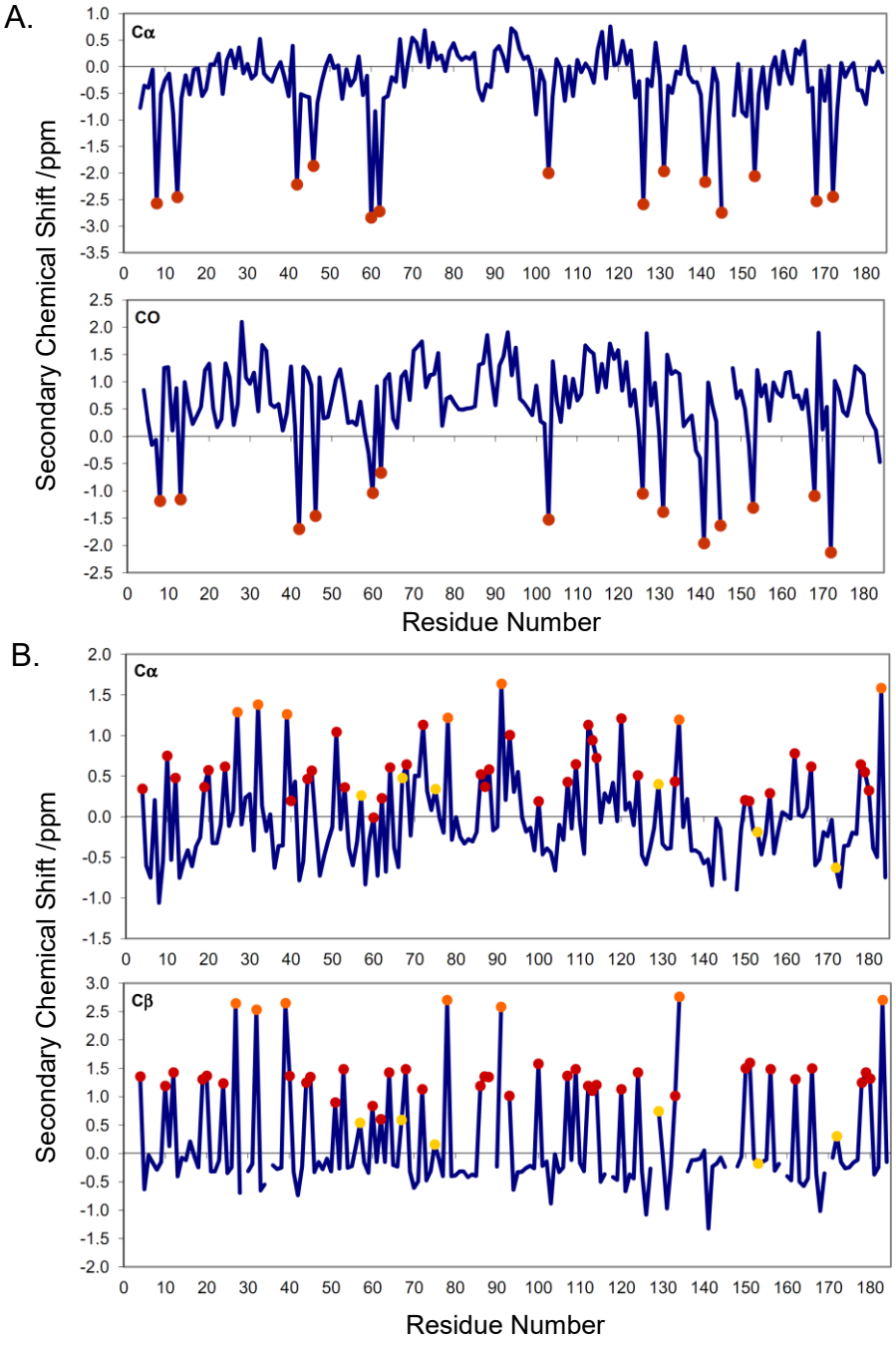

Figure 20. Typical examples of outliers within the SCS values of ERD14 because of the use of non-adequate sets of random coil chemical shifts. A) Missing neighbor correction: C^{α} and C^{O} SCS using reference chemical shifts from Richarz and Wüthrich. Residues preceding prolines are marked in red. Additionally to the presence of systematic outliers, the bias of C^{O} SCS can be seen. B) Great difference in pH: C^{α} and C^{β} SCS using reference chemical shifts from Schwarzinger et al. Charged residues: Glu, Asp and His are marked in red, orange and yellow, respectively.

Year	Publication	Available spins and corrections	Experimental conditions and/or source of data	Remarks on use with the present chemical shift data of ERD14
(1978) (1979) (1994)	Richarz and Wüthrich Bundi and Wüthrich Braun et al.	CA, CB, CO H, HA N with YX corr.	(H)-G-G-X-A-(OH) pH 7.0 308 K = 35°C	- CB, CO: biased - overall "noisy" - CA, CB, CO, HA: XP corrections needed
(1995)	Wishart et al.	N, H, HA, CA, CB, CO XP corr. for all	Ac-G-G-X-Y-G-G-NH_2 pH 5.1 1M Urea, 99.9% D_2O 298 K=25°C	- N: biased - CB, HA: Glu outliers (small difference)
(1995)	Merutka et al.	H, HA	(H)-G-G-X-G-G-(OH) pH 5.0 T dependence: 278-318 K=5-45°C various TFE conc.	- H: biased - HA: XP correction needed, and small bias
(1997)	Plaxco et al.	H, HA	Ac-G-G-X-Y-G-G-NH_2 pH 5.0 0-8 M guanidine.HCl 293 K=20°C	- HA: XP correction needed
(1997)	Lukin et al.	N, CA, CB, CO	Statistics from coil states of proteins from BMRB and other selected proteins	- CA, CB, CO: XP corrections needed - N: biased
(2000) (2001)	Schwarzinger et al. Schwarzinger et al.	N, H, HA CA, CB, CO ABXCD corr.	Ac-G-G-X-Y-G-G-NH_2 pH 2.3 8 M Urea 293 K=20°C	- CA, CB, CO: Glu, Asp and His give strong outliers
(2002)	Wang and Jardetzky	N, H, HA CA, CB, CO	Statistics from random coil regions of proteins in BMRB and PDB	- N, H: biased - CA, CB, CO, HA: XP corrections needed - HA: His gives outliers
(2003)	Zhang et al.	N, H, HA CA, CB, CO	Statistics from refDB, average of chemical shifts in coil regions	- overall "noisy" - XP corrections needed - N, H: biased
(2009)	De Simone et al.	N, H, HA CA, CB, CO AXB weighted correction	Statistics from loop regions of proteins in BMRB and PDB including nearest neighbor effects	- CA, CB, CO, HA: XP corrections needed (small difference) - H, HA: a bit biased

Table 3, to be continued

Year	Publication	Available spins and corrections	Experimental conditions and/or source of data	Remarks on use with the present chemical shift data of ERD14
(2010)	Tamiola et al. (ncIDP library)	N, H, HA CA, CB, CO AXB corr.	Statistics from preselected IDPs by SVD analysis	- HA: biased
(2011)	Kjaergaard and Poulsen	N, H, HA CA, CB, CO ABXCD corr.	Ac-Q-Q-X-Q-Q-NH_2 pH 6.5 (pH titration for D,E,H) 278 K=5°C -> basic reference set	- CO, N: Gly outliers (small difference)
(2011)	Kjaergaard et al.		Ac-G-G-X-G-G-NH_2 pH 6.5 1 M Urea 278-318 K = 5-45°C -> T coefficients -> neighbor corr. for G	

Table 3. Table comparing random coil chemical shift reference sets. For each reference set, the year and first author of publication, available spins and correction factors, source of data and/or experimental conditions (measured peptides, pH, buffer composition (denaturant concentrations), temperature; or for statistical reference sets: basis of the statistical analysis), as well as remarks on the usefulness with ERD14 (i.e. in general with intrinsically disordered proteins measured under near native conditions) are listed.

At the time of publication of the assignment results of ERD14, the most appropriate reference for chemical shift analysis seemed to be the random coil chemical shifts from Wishart et al. with XP correction for all nuclei, combined with the YX correction factors for amide nitrogens from Braun et al. (which also corrected for the observed bias of N secondary chemical shifts). However, newer results, as the chemical shift libraries from De Simone et al., Tamiola et al. (except for the biased H^{α} shifts) and especially from Kjaergaard et al. (combined values, as recommended in Kjaergaard and Poulsen, 2011) published specifically for the intended use with intrinsically disordered proteins, do give a better match, and are used and presented in the following chapters within this work (see below in section 4.1.3).

It is noted that the DeVise visualization application of the BMRB database does not use the most appropriate set of reference chemical shifts (i.e. not appropriate for this protein and these conditions) for the calculation of secondary chemical shifts, nor for the calculation of secondary chemical indices. As such, the indication of the structural features described in this work cannot be seen on the automated visualizations presented by DeVise.

Conclusions from the comparison of random coil chemical shifts

The best reference set of random coil chemical shifts for any chemical shift analysis is the one that is closest in pH to the experimental conditions (especially big differences can be expected and observed for charged amino acids near their pKa values: 4.1 for Asp, 4.5 for Glu and 6.8-6.9 for His; Kjaergaard et al., 2011; Kjaergaard and Poulsen, 2011) and that is corrected for neighboring effects, especially nitrogen chemical shifts and for residues preceding prolines. Furthermore, correct referencing and correction of temperature differences is needed to avoid biased SCS values. But even when taking into account all of these major effects, biased values and outliers might be inherent to reference chemical shifts, as demonstrated e.g. for residues preceding prolines within SCS values calculated with reference chemical shifts from De Simone et al. or for H^{α}-SCS values calculated with reference chemical shifts from Tamiola et al.

The use of the most appropriate set of random coil chemical shifts is especially important for intrinsically disordered proteins, as often minor differences in populations of secondary structural elements are inferred from minor chemical shift changes, and neighbor or temperature effects can have the same magnitude – or be even bigger – as changes arising from the investigated transient secondary structural effects, and therefore might interfere with the interpretation of the data (as can be seen e.g. in figure 23 in section 4.1.3 below).

4.1.3. Structural features of purified ERD14

Relaxation behavior

Relaxation parameters R_1, R_2, $R_{1\rho}$ and hetNOE determined for ERD14 at pH 6.5, 288 K and 600 MHz are shown in figure 21 and listed in addendum 6.2. The measured values overall correspond to values of an intrinsically disordered protein. However, five distinct regions of relatively restricted motion (elevated R_2, R_1, $R_{1\rho}$ and hetNOE values) can be observed, which means, that amino acids in these regions are somewhat less flexible and spend at least part of their time in a motionally restricted structural element. Taking into account only residues in a window of 3 amino acids where at least 30% of the measured relaxation parameters (R_1, R_2 and hetNOE; $R_{1\rho}$ is excluded, as considered to be equivalent to R_2 values) have values higher than the average plus one standard deviation, the following five regions can be defined to experience relatively restricted motion: 1: R28-K37; 2: E68-H75; 3: G92-K96; 4: E109-E124; 5: K163-K165 (indicated by gray bars in figure 21).

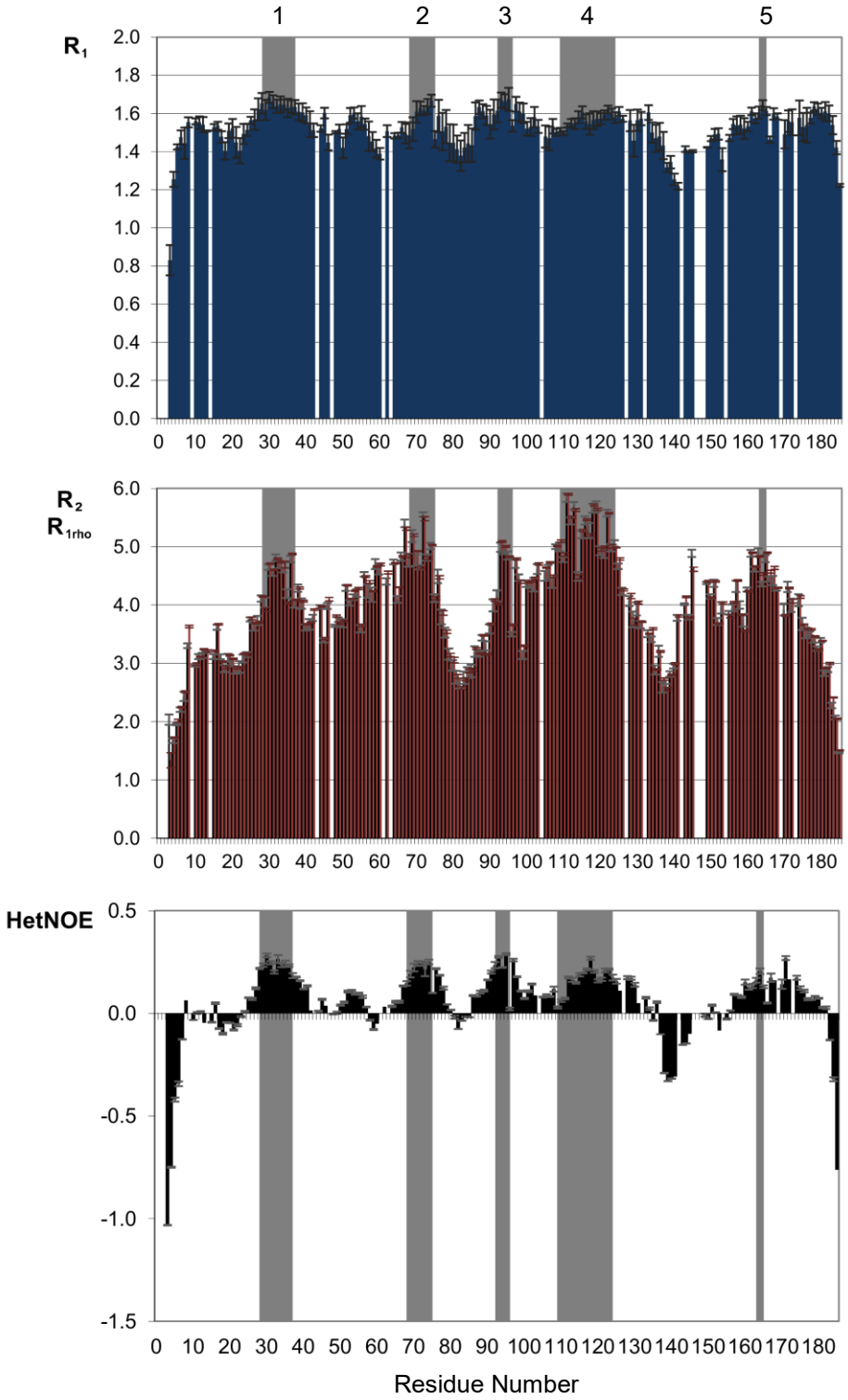

Figure 21. Relaxation values of ERD14 at 288K vs. sequence. Gray bars indicate regions with relatively restricted motion within the protein, while the green box highlights an especially mobile region with hydrophobic character.

The hydrophobic region around amino acids A137-V145 shows lower relaxation values and therefore is more flexible than other parts of the protein. This latter result is surprising, as

hydrophobic regions are usually expected to form structure more readily than highly charged regions, but in this case show almost as high mobility as the termini of the polypeptide chain.

Last but not least, from the equivalence of R_2 and $R_{1\rho}$ values it can be inferred, that ERD14 undergoes no form of chemical (nor conformational) exchange on the NMR timescale.

Secondary chemical shifts

Secondary chemical shifts (SCS) calculated with the three most appropriate sets of reference chemical shifts (Kjaergaard et al., 2011; Kjaergaard and Poulsen, 2011; K. Tamiola, B. Acar, and F.A.A. Mulder, 2010b; De Simone et al., 2009) are depicted in figure 22.

The figure shows that although SCS vary somewhat according to the different sets of reference chemical shifts used (even though all three reference sets seem to be appropriate for ERD14 and the used experimental conditions, as discussed above in section 4.1.2), all give SCS values near zero, i.e. confirm the intrinsic disorder of ERD14 under the given conditions. However, a closer look at the data shows that within the regions of restricted mobility (gray bars in the figure, inferred from the relaxation measurements as described above) C^{α}- and C^{O}-SCS are slightly higher (positive values), while C^{β}-, H^{α}-, H^{N}-, and N-SCS values are slightly lower (negative values) than the average within other parts of the protein, indicating partial α-helical tendency for these regions. These trends can be seen from all three sets of reference chemical shifts used, irrespective of the variance between the three datasets, which moreover seems to decrease within these regions. H^{α}-SCS calculated with reference chemical shifts from Tamiola et al. display an overall bias towards more positive values, but even here a lowering of the SCS values within the restricted regions can be observed.

For folded proteins, often the chemical shift index (CSI) is determined (Wishart et al., 1992; Wishart and Sykes, 1994) from the sign of the SCS values to identify secondary structural elements within the sequence. Using this approach for the present data set (Random Coil Index online calculator; Berjanskii and Wishart, 2007) determines ERD14 to be of random coil structure along its entire length. In this case, as for disordered proteins in general, a more sophisticated evaluation of even small deviations of the SCS values is needed. For secondary structural elements that are only partially formed – i.e. partially populated – a more quantitative characterization can be performed, which takes into account not only the sign, but also the extent of the SCS values.

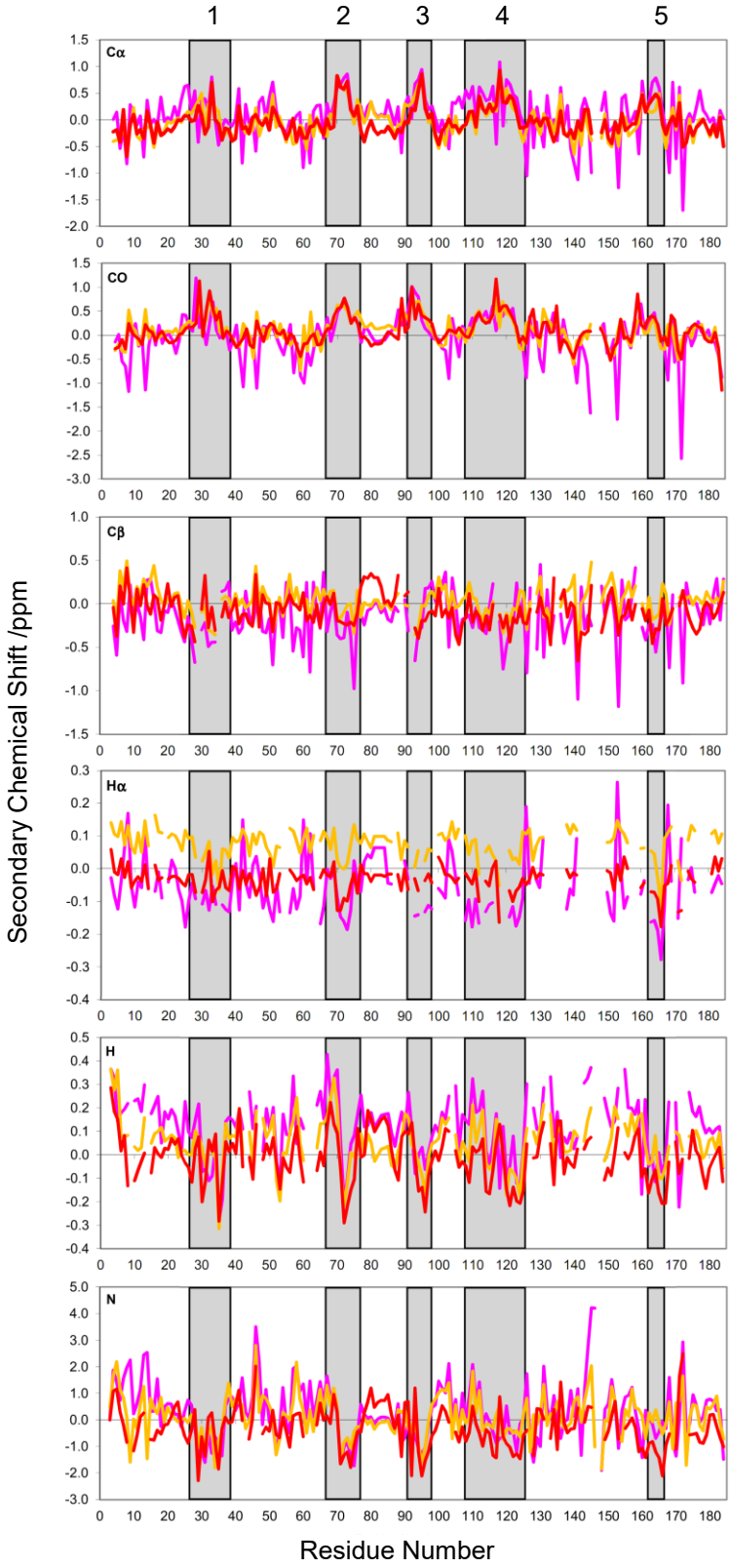

Figure 22. Secondary chemical shifts of ERD14 for all measured nuclei using the three most adequate sets of reference chemical shifts: De Simone et al. (pink), Tamiola et al. (gold) and Kjaergaard et al. (red). Grey bars indicate regions of restricted mobility as inferred from the NMR relaxation experiments. H^{α}-SCS calculated with reference chemical shift values from Tamiola et al. are outliers, due to a systematic bias in the reference chemical shifts.

The results of two such quantification procedures with basically different approaches, δ2D (Camilloni et al., 2012) and the neighbor corrected Structural Propensity Calculator, ncSPC (manuscript: K. Tamiola, B. Acar, and F.A.A. Mulder, 2010a), are presented in a graphical form in figure 23 A and B, respectively.

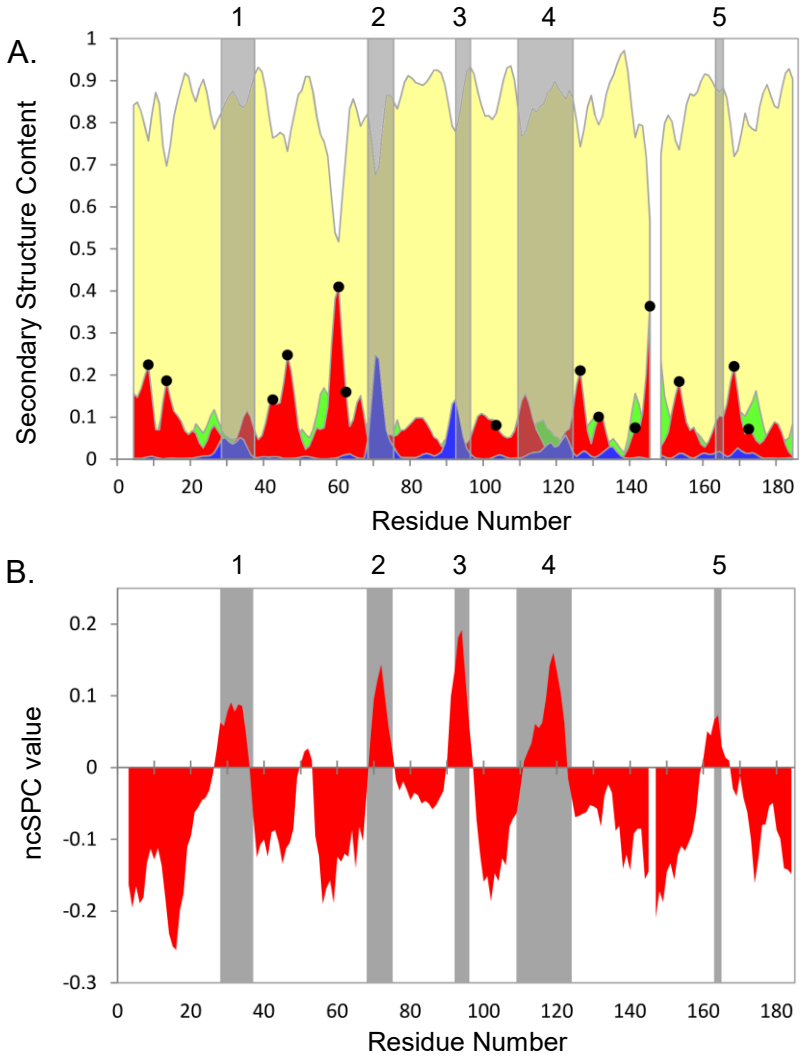

Figure 23. Structural propensities for ERD14 as inferred from a combination of the measured chemical shifts of several nuclei using the δ2D method (A) and the ncSPC method (B). In (A) estimated secondary structure populations are depicted in blue for α-helices, red for β-sheets, green for PPIIs and yellow for random coils, while gray boxes indicate regions of restricted flexibility as inferred from the NMR relaxation experiments. Black dots mark residues preceding prolines, where β-sheet populations are overestimated due to an underestimation of XP corrections within SCS calculation (as discussed in section 4.1.2). In (B) positive and negative values represent estimated secondary structure populations for helical and sheet conformations, respectively.

The δ2D method is trained on a database of known structures and predicts relative probabilities for secondary structural populations for α-helices (blue), β-sheets (red), polyproline II regions (PPII, green) and random coils (yellow). It is noted that the δ2D method uses the same reference chemical shifts for random coil values as published by De Simone et al. (derived SCS values shown in pink within figure 22), where due to an underestimation of the neighbor effect of prolines, residues preceding prolines give small, but distinct outliers. For these residues, β-sheet populations are overestimated by the δ2D algorithm (marked by black dots in the figure). This again demonstrates how important it is to use an appropriate set of random coil chemical shifts for IDPs and how important neighbor corrections terms are in this sense (compare section 4.1.2). Furthermore, H^N and H^α chemical shifts were omitted from the analysis to avoid a further systematic overestimation of extended structures arising from the bias inherent to these SCS values (see table 3).

In contrast, the ncSPC analysis, is a neighbor corrected variant of the Secondary Structure Propensity (SSP) method (Marsh et al., 2006) that calculates estimated secondary structural populations for α-helical (positive values) and β-sheet (negative values) regions from a linear combination of weighted SCS values related to the expected (tabulated, amino acid dependent) SCS values for the fully formed structural elements. For this calculation only C^α, C^β and H^α chemical shifts – the most indicative for secondary structural elements – were used (as suggested by the authors of SSP) with reference chemical shifts from Tamiola et al.

Within the regions of ERD14 determined to be of restricted mobility (gray bars within the figure), both methods show α-helical tendencies to be stronger than in other parts of the protein. While the ncSPC method clearly indicates helical populations for all five regions of restricted mobility, according to the δ2D algorithm helical tendency is present in, but less evident for regions 1, 4 and 5.

Region	Average		Maximum	
	δ2D	ncSPC	δ2D	ncSPC
1	4%	7%	5%	9%
2	13%	8%	25%	14%
3	8%	14%	14%	19%
4	3%	8%	6%	16%
5	2%	6%	2%	7%

Table 4. Average and maximal helical tendencies for the five motionally restricted regions of ERD14 as predicted by the δ2D and ncSPC algorithms.

Averages and maxima of helical tendencies within the highlighted regions according to both prediction methods are given in table 4. Both methods agree on regions 2 and 3 to have stronger helical tendency than regions 1, 4 and 5. Taking into account that errors of such methods are estimated to be between 5-10% on a per residue basis (e.g. 8% for δ2D, Camilloni et al., 2012), the extent of helical tendency predicted by both methods is similar for all five regions.

Implication of the structural features of ERD14

The presented in depth analysis of ERD14 in buffer under near-native conditions on the one hand confirms the fully disordered state of ERD14, while on the other hand it reveals five regions within the sequence of ERD14 that have relatively restricted mobility and a slight preference for helical conformations. Although average helicity in these regions was determined to be only on the magnitude of 2-14%, these conformational preferences indicate possible binding regions of ERD14, as they might represent so-called structurally preformed elements, which are known to exist within intrinsically disordered proteins and are proposed to facilitate the recognition and binding of partner molecules (see section 1.4.3).

Furthermore, three of the five regions of relatively restricted mobility and helical tendency (regions 1, 3 and 5) overlap fully with the three conserved K-segments of ERD14, while the other two regions (2 and 4) show partial overlap with the conserved S- and ChP-segments, respectively, as demonstrated in table 5. This coincidence – although only partial for two of the regions – further affirms the importance of both, the conserved regions and the preformed structural elements.

Regions of restricted mobility and helical tendency	1: R28-K37	2: E68-H75	3: G92-K96	4: E109-E124	5: K163-K165
Conserved segments	K-segment: E24-P43	S-segment: L74-S85	ChP-segment: E87-K103	K-segment: E113-D134	K-segment: E156-G170

Table 5. Borders of the five regions of relatively restricted mobility and helical tendency within the sequence of ERD14 compared to the conserved segments of the protein. Full overlap with the segments is observed for regions 1, 3 and 5, while the overlap is only partial for regions 2 and 4.

4.1.4. Structural features within the cell: in-cell NMR

Figure 24 A) and B) show the NMR spectra of living *E. coli* cells before and after induction of protein expression, i.e. in the absence and presence of ERD14, respectively. Both spectra are shown as an overlay of the in-cell NMR spectrum (with cells) with the corresponding control spectrum of the supernatant after centrifugation (without cells). The spectra show that ERD14 is expressed in measurable amounts upon induction and is present within the cells only, i.e. there is no leakage of target protein into outercell space, which is also confirmed by the SDS-PAGE of the control samples shown in 0. Uninduced and supernatant spectra only contain peaks of ^{15}N-containing metabolites of the cell. Also, no disturbing background peaks from other proteins can be seen, i.e. while also other proteins are present within the cell, only the target protein ERD14 contains the ^{15}N-labels detected by NMR.

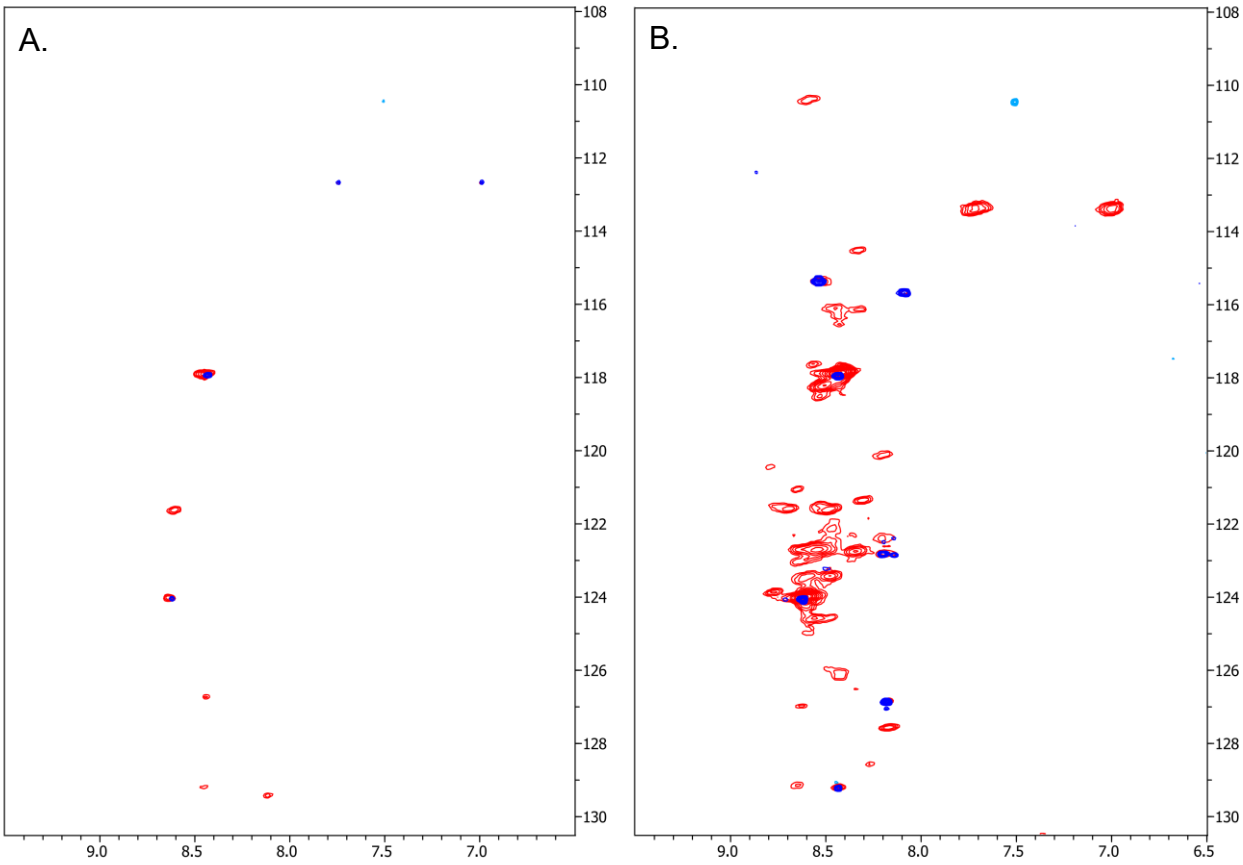

Figure 24. HSQC spectra of the in-cell samples of ERD14 and related control samples: A.) Overlay of the spectrum of the sample taken before induction (red) with the spectrum of its supernatant (blue). B.) Overlay of the spectrum of the sample taken after 3 hours of induction ("in-cell spectrum" of ERD14, red) with the spectrum of its supernatant (blue).

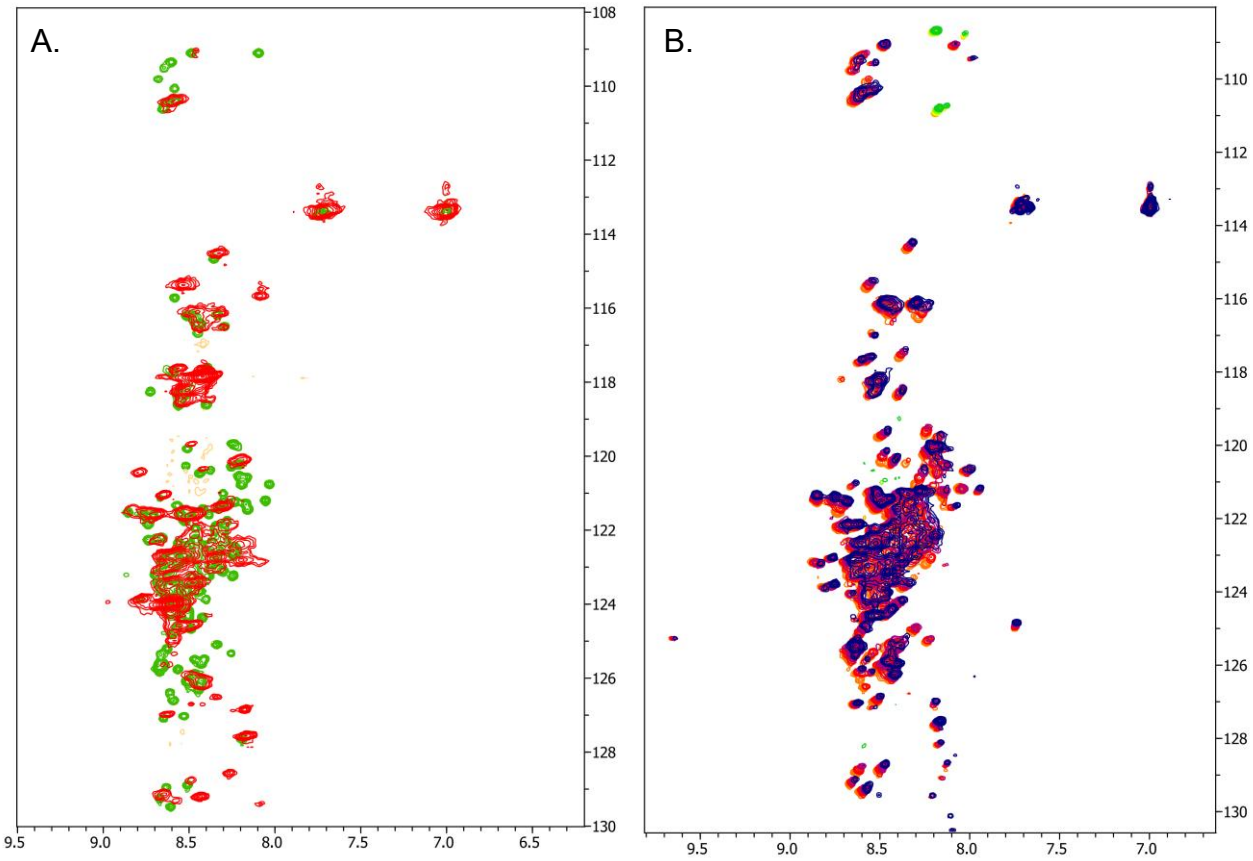

Figure 25. A.) Overlay of the in-cell spectrum of ERD14 (red) with the reference buffer spectrum of corresponding pH (green). B.) Overlay of the spectra of purified ERD14 at increasing dextran concentrations: 0 (orange), 150 (red), 300 (violet) and 400 g/L dextran (blue).

Figure 25 A shows the same in-cell spectrum as shown in figure 24 B, overlaid onto the buffer spectrum of purified ERD14 of corresponding pH (7.7, cf. sections 1.7, 3.6.2 and 4.1.1). The comparison of the in-cell spectrum with the corresponding buffer spectrum of ERD14 clearly shows that most of ERD14 is in the same intrinsically disordered state within the cell as in dilute buffer. However, a subset of the peaks is missing ("disappearing", i.e. presumably broadened beyond the detection limit) from the in-cell spectrum, while another couple of peaks appear shifted compared to the reference spectrum.

It is proposed that resonances in certain regions of ERD14 broaden out or disappear due to binding to slowly tumbling entities within the cell (either huge proteins or membranes). To test whether the disappearance of peaks can be attributed to a true, i.e. specific, binding event in the cell or can be explained rather by unspecific effects, HSQC spectra of purified ERD14 were measured also in the presence of 0-400 g/L dextran. Unspecific effects due to molecular crowding, as caused by the high number of proteins and other molecules within the cell are expected to be observable in the presence of high amounts of dextran as well.

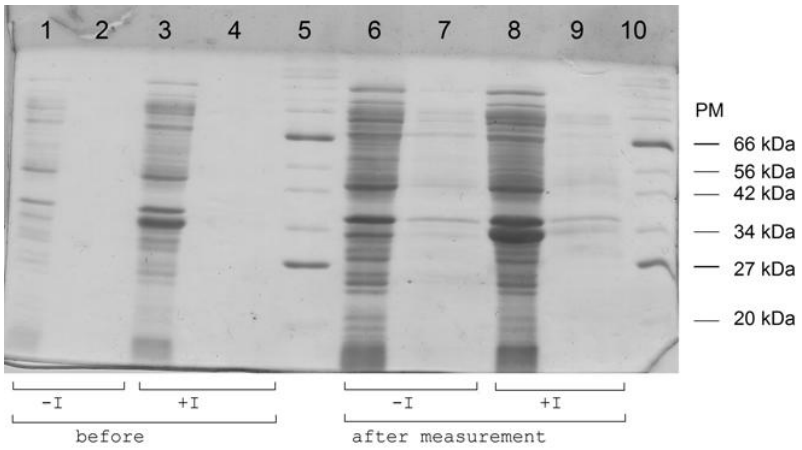

Figure 26. SDS-PAGE of control samples taken before (lanes 1-4) or after (lanes 6-9) the in-cell measurement. Lanes 1, 6: cells before induction, lanes 2, 7: supernatant before induction, lanes 3, 8: cells after 3h of induction, lanes 4, 9: supernatant after 3h of induction; lanes 5, 10: protein markers, with molecular weights as indicated.

As presented in figure 25 B, most of the resonances of ERD14 are observed even in the presence of 400 g/L dextran, therefore unspecific effects of molecular crowding can be excluded for almost all parts of the protein. The few exceptions (11 resonances, as listed in addendum 6.3) that are “disappearing” from the spectrum measured with dextran as well as from the in-cell spectrum are considered to be “unknown” for the interpretation of the in-cell data. A small shift of a couple of other resonances during dextran titration is probably due to a minor pH change that was not corrected for during the titration and was not further analyzed. Furthermore the presence of a set of minor peaks can be observed regardless of the dextran concentration, which resulted from using a rather old sample (about 1 month old) for the experiment. This set of small resonances did not disturb the present analysis, and was therefore ignored.

As the assignment of the corresponding reference spectrum (at pH 7.7, 277 K) is available, the peaks missing from or shifting in the in-cell spectrum and those certainly steady, can be mapped to the sequence of ERD14, as listed in a tabular form in addendum 6.3. Unfortunately no certain information can be obtained from a lot of resonances (marked as “unknown” in the list). This is due to the fact that peaks in the in-cell spectrum are somewhat broader than in the buffer spectrum (mainly due to the inherent inhomogeneity of the cell suspension), and an eventual shifting or disappearance of peaks within the crowded region of the spectrum cannot be determined.

Therefore, residue-specific information available for further evaluation is sparse. In order to be able to obtain at least general trends for certain regions within the protein, a "disappearance score" has been defined. Basic scores were set to "+1" for "disappearing", "0" for "unknown" or "shifted" and "-1" for "certainly steady" residues. Then the individual values were smoothed to give the final "disappearance score" as the running average over a window of five residues of the basic scores. A threshold of 0.2 has been set, in order to exclude effects that might affect only one specific residue rather than any longer stretch of amino acids within the sequence. In figure 27 this "disappearance score" is presented and compared to the regions of the protein known to be either conserved (colored bars) or established in this work to have a small but measurable preference for helicity (gray bars, for details see section 4.1.3).

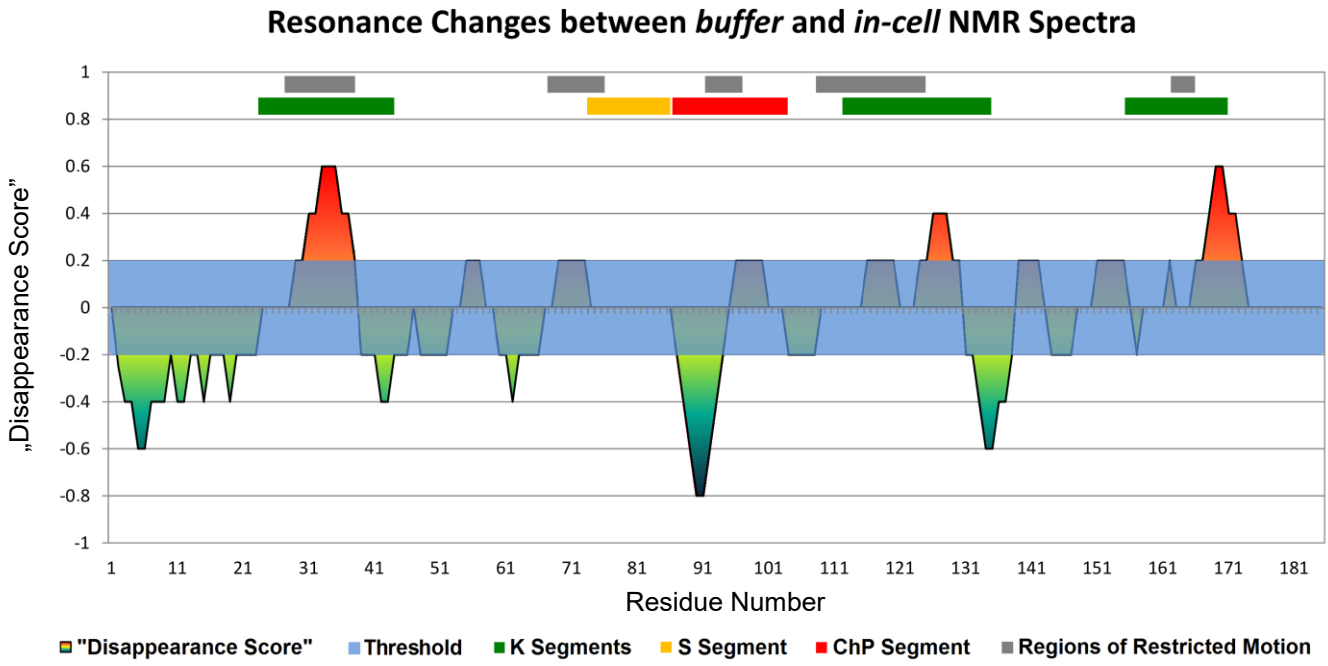

Figure 27. In-cell NMR data mapped to the sequence of ERD14. The "disappearance score" has been defined to be positive for regions of "disappearing" resonances and negative for regions of "certainly steady" residues (see text for details). Conserved regions of ERD14 are marked by colored bars at the top, while gray bars indicate regions of relatively restricted motion and helical tendency.

Interestingly, high disappearance scores are observed for regions of the protein near or within the K-segments, while a stretch of certainly steady peaks (low disappearance score) is observed e.g. within the ChP-segment. No information is available on the S-segment and the preceding region of partial helicity, as well as for long stretches in between the segments. Furthermore, it has to be emphasized, that the borders of these regions (of both, high and low disappearance scores) are borders to the "unknown" (where within the threshold), and the

possibility for neighboring stretches within the protein sequence to have similar tendencies is high. This is especially true for parts of the second and third K-segment, as here due to their high sequence similarity signal overlap is inherent to the sequence. In contrast, the first K-segment shows less sequence similarity to the other two K-segments, and therefore resonances of its residues – as well as the lack of these resonances – could be detected more readily for the whole segment.

Implications of the results of the in-cell NMR measurements

On the one hand, the in-cell NMR spectrum of ERD14 is very similar to the spectrum measured in buffer under near-native conditions, and therefore demonstrates that most parts of the protein adopt the same intrinsically disordered state, i.e. explore the same conformational space within the living cell, as determined in detail for the protein in buffer.

On the other hand, in certain regions within the protein sequence the resonances of ERD14 are broadened out beyond the detection limit. As mentioned above, it is proposed that the disappearance of these resonances in the in-cell spectra is due to the binding to slowly tumbling entities within the cell (e.g. huge proteins, partner molecules or membranes). As was shown in figure 27, at least three such binding regions exist within ERD14, which are located at least partly within the conserved K-segments. This finding implicates that the K-segments may be the binding regions of ERD14 to its partner molecule(s) and therefore may play a special role in the function of the protein, which underlines the importance of the conserved K-segments.

4.1.5. Structural features of mutant ERD14 proteins

The five basic mutants of ERD14 (Ka, Kb, Kc, S and ChP) were subjected to a general analysis of disorder and within each method compared to the wild type protein.

Figure 28 shows SDS-PAGE and native PAGE of purified WT ERD14 and its basic mutants. The apparent protein size of all mutants of ERD14, as estimated from SDS-PAGE (according to the protein marker, as given within the figure), is well above their calculated mass as expected for disordered proteins. The relative position (migration length) of the mutants compared to each other and to the wild type protein corresponds well to their relative calculated masses.

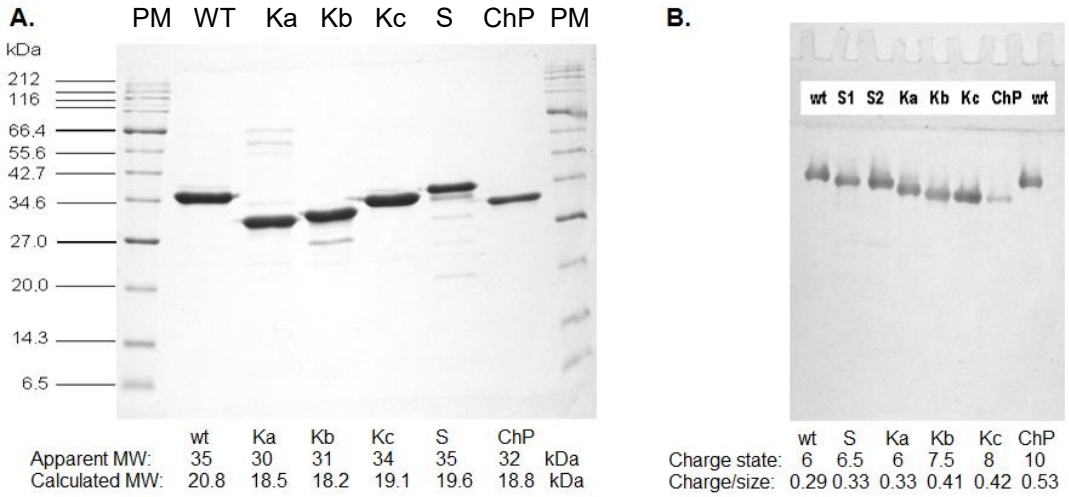

Figure 28. A) SDS-PAGE gel electrophoresis of the basic mutants of ERD14 and the wild type protein, with a protein marker of globular proteins of known size (PM) as reference. Apparent as well as calculated masses are given for each protein. B) Native gel electrophoresis of the basic mutants of ERD14 (from left to right: WT, S1, S2, Ka, Kb, Kc, ChP, and WT ERD14); S1 and S2 refer to two different preparations of the same mutant ERD14 S. The calculated charge states (overall net charge, which is negative in all cases), as well as charge/mass ratios are given for each protein.

Mobility within native gel electrophoresis, however, depends not only on the protein size (and amino acid composition, see section 1.6), but also on the charge state of the protein. As expected, the relative position (migration length) of the mutants compared to each other and to the wild type protein corresponds well to their relative charge state and size, as noted in the figure. Furthermore, it can be seen that neither dimeric, nor trimeric, nor oligomeric forms of the proteins are present.

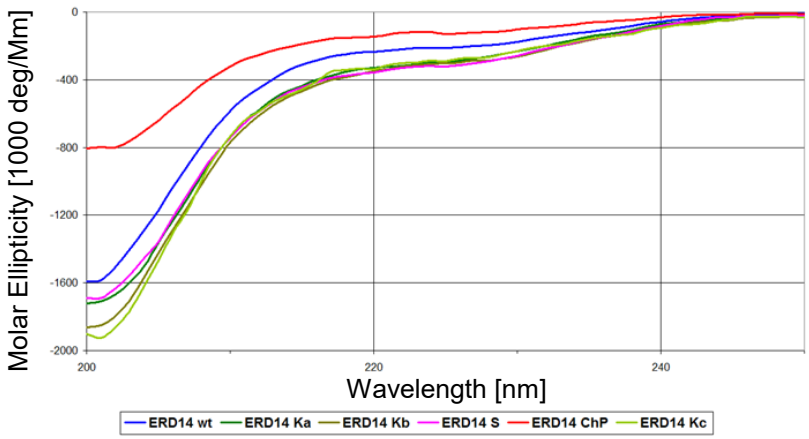

Figure 29. Comparison of CD spectra of WT ERD14 and its basic mutants (color code: WT, blue; Ka, Kb, Kc, green; S, pink; ChP, red).

An overlay of the CD spectra of wild type ERD14 and its mutants is presented in figure 29. The spectra all show a strong negative band at around 200 nm that is characteristic for disordered proteins. A small negative band (shoulder) at around 220 nm indicates the possible presence of small amounts of α-helices or slight α-helical tendency within parts of the protein.

Size exclusion chromatography experiments were performed in order to compare the apparent hydrodynamic radii of the ERD14 proteins. As derived from a comparison to a set of known globular proteins (standard mix), WT ERD14 and its mutants have hydrodynamic radii that correspond to globular proteins of 6-7 fold bigger molecular weights than their calculated mass, as is shown in figure 30. This is somewhat more than is generally expected for random coil like IDPs, for which the apparent molecular weight estimated from gel filtration is typically 4-6 times higher than calculated (while less is expected for a molten-globule- or pre-molten-globule-like IDP; see section 1.6.1).

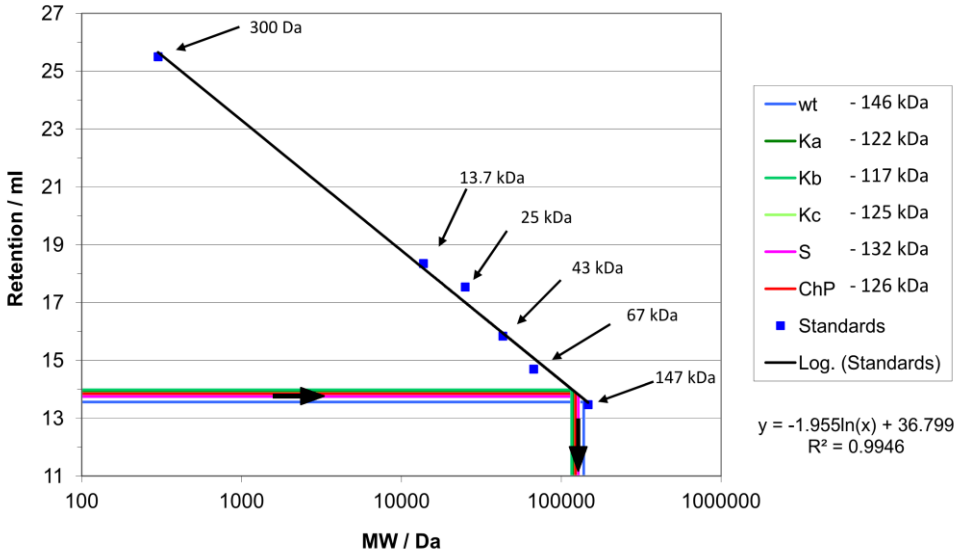

Figure 30. Size exclusion chromatography result of the basic mutants of ERD14 and the wild type protein. The retention of the proteins is plotted against a logarithmic scale of molecular weights. Apparent molecular weights are given in the figure legend and are calculated from the linear regression fitted to the values measured for the standard protein mix.

It can be concluded that all five basic mutants of ERD14 show the same – or at least very similar – general structural characteristics as the wild type protein, i.e. the deletions did not induce any observable general changes in the intrinsic disordered state of the protein.

4.2. Function of ERD14

It has already been shown – as discussed in detail in section 1.2 – that wild type ERD14 has potent chaperone activity *in vitro* (Kovács, Kalmár, et al., 2008; Kovács, Ágoston, et al., 2008). However, it has not been shown previously if this molecular chaperone activity is also present and/or relevant under *in vivo* conditions. Therefore, in the present work, an *in vivo* chaperone assay has been developed – with three different stress effects to observe – for the measurement of chaperone function within a living model system, as presented below.

4.2.1. Method development for the measurement of *in vivo* chaperone effects

As the true biological environment of dehydrins – the plant – is a rather complex organism, it is difficult to use for experimental purposes, especially if the investigations aim at obtaining specific functional information at a molecular level. Furthermore, investigations of dehydrins in plant cells are hindered also by the fact that several dehydrins (of supposedly equal or at least very similar function) are present in the *Arabidopsis thaliana* plant. Therefore, for the present analysis and method development of *in vivo* chaperone effects, instead of a living plant or living plant cells, *E. coli* was chosen as a model system, being a rather well-defined and easy-to-handle organism in which originally no dehydrins – nor other LEA proteins – are present (only the introduced protein under investigation, and therefore eventually disturbing secondary effects from other dehydrins can be excluded). Furthermore, for *E. coli* cells a lot of background information and complex scientific knowledge and experience are available. Especially the known technical background and easy performance of target DNA uptake and easily induced protein expression were determined to be of great advantage.

A major drawback of course is that although *E. coli* cells provide a well-testable living environment, they are different from plant cells, and there is a chance that specific interactions to plant proteins might play an important role but cannot occur – and therefore cannot be observed – in a foreign organism. However, ERD14, and other dehydrins, as well as intrinsically disordered chaperones in general, are supposed to function in a relatively unspecific mode of action, and hence are expected to have quite broad substrate specificity. Thus, dehydrins might be able to also find suitable substrates for their action in a foreign organism as well.

Therefore – at least as a first approach and in light of the present aims – *E. coli* was chosen as a model system for *in vivo* chaperone assays.

In order to be as near as possible to the physiological role of dehydrins, stress effects were chosen for the investigations that are known to result in elevated expression levels of ERD14 – and of dehydrins in general – within the plant, i.e. elevated salt concentration, cold and dehydration. These stress effects were adapted to the use upon bacterial cells and optimized in order to achieve a partial – but not total – loss of cell viability (see section 4.2.3), which is then expected to change due to the presence of a potent *in vivo* chaperone.

For the detection of cell viability of *E. coli* cells, in previous works samples of the cell suspensions were plated onto agar-plates, incubated O/N and the number of resulting colonies was counted (CFU method, see section 1.3). However, this process is quite laborious, especially for large numbers of different samples. Results are often imprecise, and the order of magnitude of cell numbers present in the sample solutions has to be known in advance (and eventually corrected for by dilution). Therefore, within the present analysis another method has been developed for the detection of cell viability based on the BacTiter-Glo™ Microbial Cell Viability Assay (from Promega, initially intended to measure either bacterial growth or inhibition thereof). With this novel method also high numbers of samples are handled easily using multiwell-plates and optical detection of the generated luminescent signal, hereby enabling high replica numbers of samples, where needed. Furthermore, the BacTiter-Glo™ Assay has a wide dynamic range (more than six orders of magnitude, according to the supplied technical bulletin; see also figure 32). As presented below, this method of detection of cell viability was tested and optimized for the present application.

All in all, a potent *in vivo* chaperone assay has been developed and optimized (for the experimental details of the final assay protocol see sections 3.7.2 - 3.7.4). Furthermore, the assay has been used to demonstrate the *in vivo* chaperone function of wild type ERD14, as well as the loss of this chaperone function upon directed mutagenesis.

4.2.2. Assay controls – for the correct use of the BacTiter-Glo™ reagent

Using the BacTiter-Glo™ Microbial Cell Viability Assay the number of viable bacterial cells can be determined based on the quantitation of the ATP (Adenosine-5'-triphosphate) present, which is an indicator for metabolically active cells. The formulation of the BacTiter-Glo™ reagent enables in one step the extraction of ATP from the cells and the generation of a "glow-type" luminescent signal, which is proportional to the amount of ATP present, using a

proprietary thermostable luciferase enzyme. The BacTiter-Glo™ Assay can be used in a multiwell-plate format for high numbers of samples and replicates (as indicated in the according Technical Bulletin from Promega).

1st control: linearity with ATP concentration

As the BacTiter-Glo™ reagent measures the amount of ATP present in the cells, as a first control it was tested if the measured value (luminescence) is truly proportional to ATP concentration and if it is linear within the range of interest.

A serial dilution of ATP between 100 pM and 1 µM was prepared in H_2O. Of each solution 100 µL were mixed with 100 µL of BacTiter-Glo™ reagent and luminescence was measured after several time points (1, 7, 10, 22 and 30 minutes).

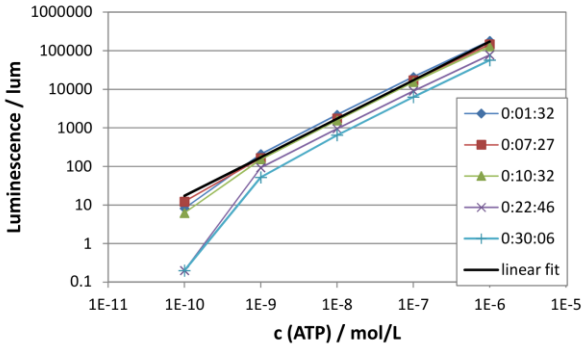

Figure 31. ATP calibration. Luminescence values were measured for a dilution series of ATP at different time points after the start of the reaction, as indicated. The black line indicates a linear fit to average values calculated from luminescence at time points 1, 7 and 10.

All luminescence values within the measured range were linear with ATP concentration (see figure 31). The typical decay of signal intensity with time (which is inherent to the method and due to the gradual hydrolysis of ATP by the luciferase enzyme) was observed to be moderate within the first 10 minutes, but significant at later time points. Thus, for the evaluation of further measurements, luminescence values measured 1-10 minutes after the start of the reaction were averaged.

2nd control: linearity with cell concentration

As a second control it was tested if the BacTiter-Glo™ reagent is capable of detecting ATP within living *E. coli* cells and if the measured value (luminescence) is linear with the number (concentration) of cells present, as indicated by the vendor.

BL21 (DE3) Star pLysS cells were grown O/N from the appropriate bacterial stock. The following day cells were inoculated into 5 mL of fresh NZYM medium (at 100x dilution) and grown (without induction) to an OD_{600} value of 0,85. From this cell suspension a serial dilution was generated in four consecutive steps of 1:10 dilution. All samples were measured as described above, luminescence values were averaged (1, 5 and 10 minutes after start of the reaction) and background-corrected by subtraction of the value measured for fresh NZYM medium without bacteria.

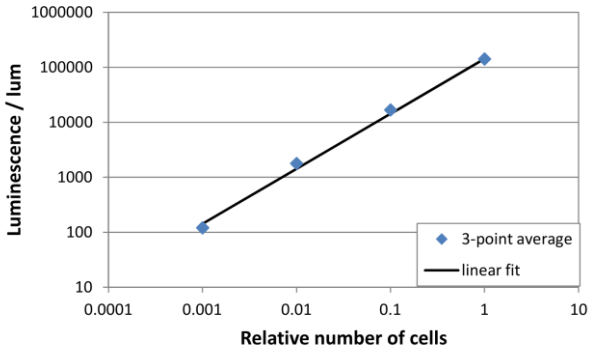

Figure 32. Calibration for E. coli cells. Luminescence values were measured for a dilution series of an E. coli cell suspension, and plotted against their relative cell concentration. Luminescence from the sample of 0.0001 relative cell number was below the detection limit, and is therefore not shown.

As shown in figure 32 all detected luminescence values were linear with concentration and on the same order of magnitude than in the dilution series of ATP. Luminescence from the last sample (4 x 1:10 dilution, 0.0001 relative cell number) was below the detection limit of the method, i.e. equal to the background luminescence measured for blank samples, and is therefore not shown on the graph. The typical decay of luminescence values with time was observed here as well, though at high bacterial concentration the first measured value (after 1 minute) was typically lower than expected (lower than the following value at 5 minutes), most probably due to the time needed for the full disruption of the bacterial cell wall and the release of all ATP. Thus, for the evaluation of further measurements, luminescence values were measured and averaged after 2.5, 5 and 7.5 minutes of reaction.

It is noted, though, that absolute luminescence values (and hence, absolute ATP levels) were observed to be quite different (meaning almost two orders of magnitude in luminescence) depending on the type of cells used, their growth phase and several other environmental factors, such as induction, nutrient composition (salt content), as well as the identity of the target protein that is expressed upon induction. But as in the presented *in vivo*

chaperone assay luminescence (ATP level) of cells is compared between parallel samples with and without stress treatment, their *relative* ATP levels can be expected to truly indicate *relative* cell viability, which is only modified by the applied stress effect.

3rd control: linearity with the amount of living cells

As within the *in vivo* chaperone assay a part of the cells is expected to die in consequence of the stress treatment applied, in the third control experiment it was tested if the assay set-up is capable of selectively detecting the amount of viable cells from a mixture of living and dead cells. In this test not only the effectivity of the BacTiter-Glo™ reagent was tested, but rather the whole experimental set-up of the *in vivo* chaperone assay, i.e. if the ATP released from the dead cells is effectively removed from the solution and if the final cell concentrations applied fall within the detectable (and linear) range of the BacTiter-Glo™ assay.

For this reason, a test measurement of the *in vivo* chaperone assay was performed as described in the section 3.7.2 ("overall procedure"). Control cells, i.e. cells containing the empty pET22b vector, were used. In place of the stress treatment cells were diluted with fresh NZYM medium to an Abs_{600} value of 0.5 and divided into two parts. One half was left at RT ("living cells"), while the other half was sonicated six times for 20 seconds ("dead cells"). Living and dead cells were mixed at different ratios (0, 25, 50, 75 and 100%) and the test was continued with the measurement of cell viability using the BacTiter-Glo™ reagent, as described in section 3.7.4 (centrifugation, resuspension and measurement). Finally, the measured luminescence values were plotted against the ratio of living cells (see figure 33).

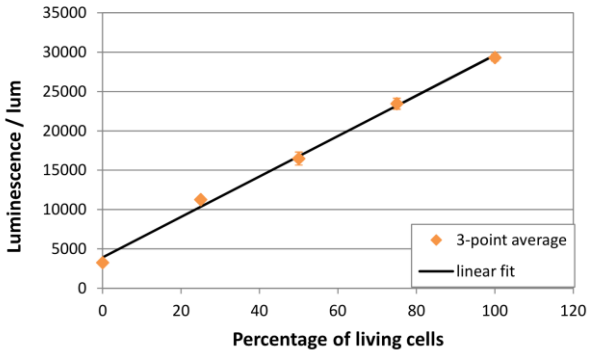

Figure 33. Calibration for a mixture of living and dead cells. Suspensions of live and dead E. coli cells were mixed at different ratios and luminescence was determined as described for the in vivo chaperone assay.

The observed luminescence values were linear with the relative amount of living cells, which proved that the assay (as presented in this work) is capable of selectively detecting the amount of viable (living) cells from a mixture of viable and dead cells.

4th control: effect of the presence of ERD14 on the assay result

It was found especially important to test whether the presence of ERD14 had an effect on the assay result (i.e. on the luminescence values measured), as ERD14 could possibly interfere with both, the luminescence reaction of luciferase or with the ATP-releasing disruption of the cell wall by the chemicals (and enzymes) included in the reagent.

The effect of wild type ERD14 on the assay result was therefore tested on a dilution series of ATP, as well as on non-stressed cells of different cell density (control and ERD14-containing cells grown for different time intervals). Either 5 µL of buffer (50 mM TRIS, 150 mM NaCl, pH 7.4) or 5 µL of WT ERD14 solution (1 mg/mL in the same buffer) were added to 95 µL of test solution (either ATP solution or cell suspension), mixed with an equal amount of BacTiter-Glo™ reagent and resulting luminescence was measured. Furthermore, luminescence values measured for control cells with or without added ERD14 were compared to values measured for cells expressing WT ERD14 internally.

As is shown in figure 34 A, the presence of added ERD14 does not influence the measurement of ATP concentration nor the measurement of cell viability of control cells.

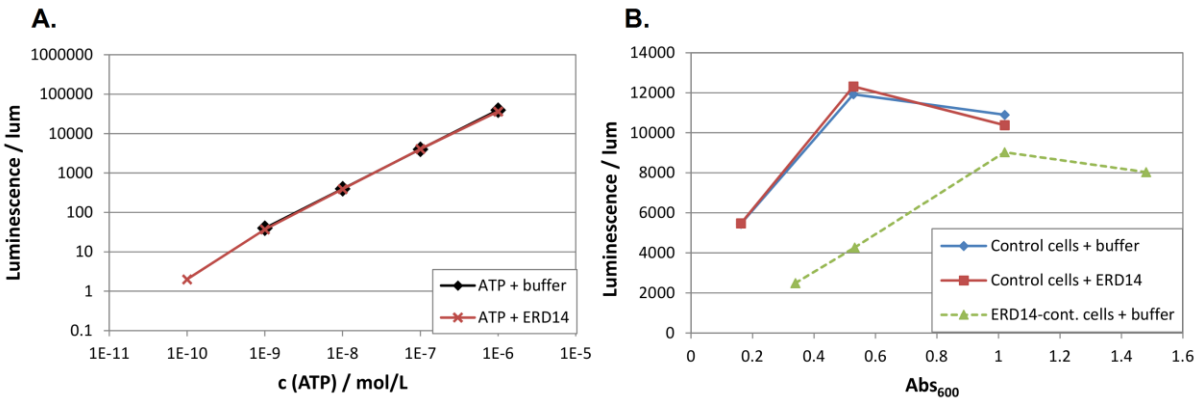

Figure 34. Test for the effect of the presence of ERD14 on the assay result. A. ATP dilution series with either buffer or purified ERD14 added. B. Control cells with and without external ERD14 compared to ERD14-expressing cells at different cell densities.

The effect of ERD14 expressed within the cells, however, is different from the effect of ERD14 added externally. If luminescence values measured for non-stressed control cells (containing only the empty vector) and ERD14-containing cells are compared, the latter

typically give luminescence values that are about 20-30% of the luminescence of the control cells, although ERD14 itself (when added externally) does not influence the measured luminescence values (as described above). It is concluded, that cells expressing ERD14 probably have lower ATP levels than the control cells, whose expression machinery is not used that extensively.

Furthermore it can be observed that although measured luminescence values are basically proportional to the concentration of the cells, luminescence values seem to reach a maximum at very high optical densities (around and above Abs_{600} = 1.0, corresponding to OD_{600} = 2.0). It is proposed that at very high bacterial concentrations the included components of the BacTiter-Glo™ reagent might not be able to fully disrupt the bacterial cell wall anymore, and therefore only a portion of all ATP is released and detected (within the time points measured and averaged).

Selection of the appropriate assay reference

As induction is a process that greatly affects cell function and energy distribution within the cell, it is expected to also greatly influence ATP levels, cell viability and stress resistance. Accordingly, the correct reference (negative control) for cell viability measurements has to be induced just like the cells under investigation, but without effectively producing any protein that could interfere with the measurements. Therefore, for the *in vivo* chaperone assay, cells containing the empty pET22b vector were chosen as a reference.

It is added that for the presented *in vivo* chaperone assay no adequate positive control was found. Typical positive control proteins, such as the globular chaperone HSP90 that has often been used within *in vitro* chaperone assays, cannot be used in this case, as no general protective function can be expected for HSP90 in these forms of abiotic stress effects and under such a variety of conditions (unlike intrinsically disordered chaperones, which are expected to function in form of a relatively unspecific mode of action).

4.2.3. Optimization of the applied stress conditions

Stress conditions were chosen to be as near as possible to stress effects that are known to result in elevated expression levels of dehydrins in plants. However, they were adapted for the use on bacterial cells (e.g. freezing instead of the lowered temperature mostly used with plants) and optimized in order to achieve a partial (but not total) loss of cell viability, which is

then expected to change due to the presence of a potent *in vivo* chaperone. Starting values for the optimization were gathered from the literature, as described in section 1.3.

Freezing stress

Freezing stress was applied to cells by freezing of the bacterial cell cultures in liquid nitrogen, followed by thawing in a water bath. As shown in figure 35, the effect of repeated freezing-thawing cycles (2 minutes freezing, 5 minutes thawing, repeated 1-2 times) was investigated. According to these test results and to give well measurable results, the final protocol was chosen to involve one freezing-thawing cycle for stressed cells, while during this time non-stressed cells were left at RT.

This method was tested on control cells and cells containing the wild type protein, as well as several mutants of ERD14 (data for mutants not shown). Significant differences in stress tolerance (relative cell viability, %CV) could be measured. However, as the measured %CV values were related to the actual cell density (Abs_{600}) present during the stress treatment, it became evident that cell viability after the freezing-thawing process was more dependent on the actual cell density rather than on the presence of any of the tested proteins within the cell. This was further analyzed in two systematic tests. As presented and compared in figure 36, control and ERD14-containing cells were measured at different cell densities, which were reached either through different dilutions of a dense cell suspension or through different growth times of several samples.

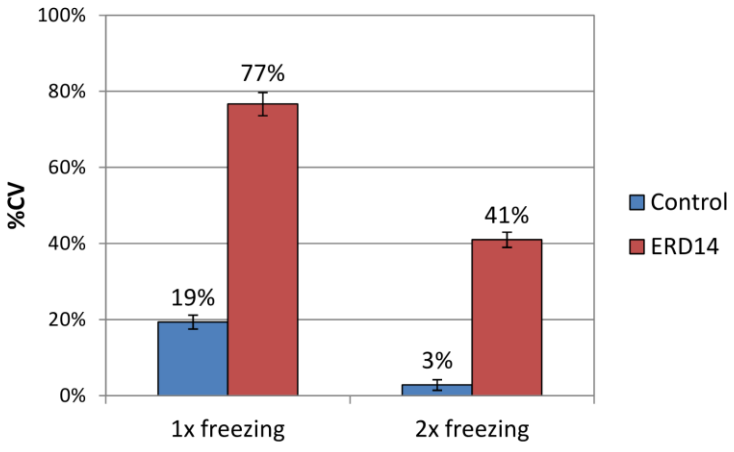

Figure 35. Relative cell viability (%CV) of control and ERD14-containing cells, comparing the effect of two different freezing stress treatments, i.e. of one and two freezing cycles, relative to unstressed cells.

Both systematic tests reveal that the observed %CV value strongly depends on the actual cell density during the applied stress effect. On the one hand this result means that freezing stress cannot be used to measure the effect of a potentially protective protein, as the influence of cell density is too strong, i.e. the effect of cell density on freezing tolerance is much bigger than (and possibly obscures) an eventual protective effect of the protein(s) under investigation. On the other hand, it is an interesting finding, that average cell density after equal times of growth was observed to be different for control, WT ERD14-containing, and mutant-containing cells, hence the significant differences in %CV values measured for different mutants of ERD14. This effect was further investigated, and the results are presented in more detail in sections 4.2.5 and 4.2.6.

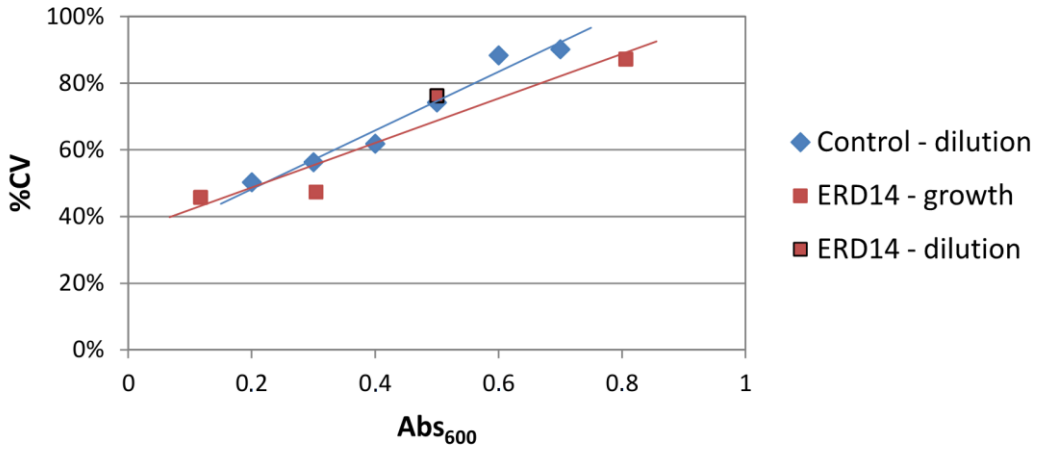

Figure 36. Relative cell viability (%CV) of control and ERD14-containing cells after freezing stress treatment in dependence of cell density (Abs_{600}). Different cell densities were achieved either by dilution from a more dense cell suspension or by different growth times without dilution, as indicated in the graph. Colored lines within the graph indicate linear fits to the measured values of either control or ERD14 cells, respectively.

High salt stress

The effect of high salt stress was tested using two different salt concentrations as is shown in figure 37. According to these test results, the effect of 2M NaCl stress was evaluated to give the best measurable %CV values, and was used for all the following measurements of high salt stress. It is noted that the salt concentrations used here are significantly higher than in other investigations of LEA proteins within *E. coli* cells (see section 1.3). However, in the present work, high salt stress is applied for a short time only, while in the other cases long-term effects of salt exposure were investigated.

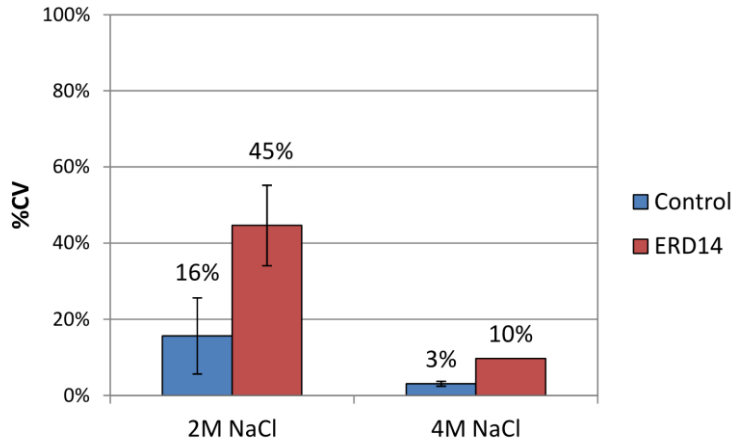

Figure 37. Relative cell viability (%CV) of control and ERD14-containing cells, comparing the effect of high salt stress treatments of different intensities, i.e. the effect of 2 M and 4 M effective salt concentration relative to unstressed cells.

As a further test, high salt stress tolerance of control cells was measured after different growth times, at different cell densities. As can be seen from figure 38, the measured relative cell viability shows no significant dependence on cell densities, which enables the use of this assay in the measurement of the protecting effect of proteins (indifferent of the growth rates of their host cells after induction).

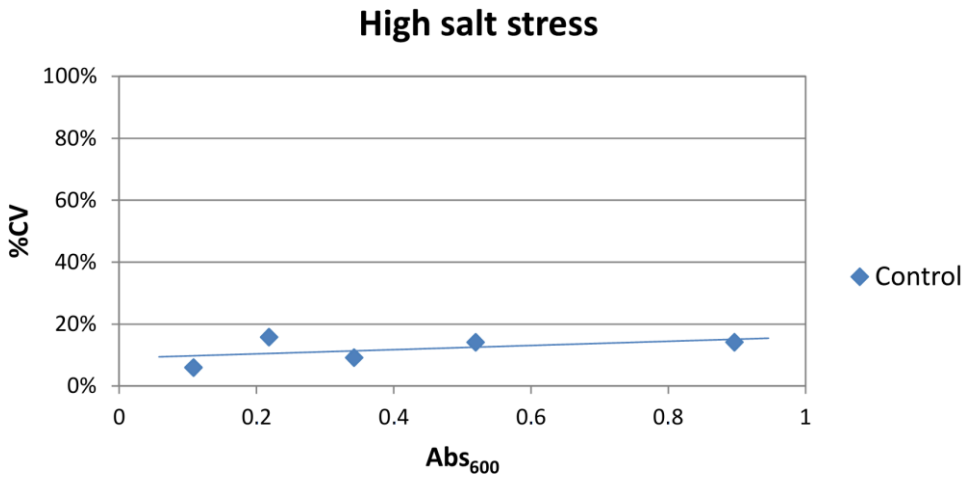

Figure 38. Relative cell viability (%CV) of control cells after high salt stress treatment in dependence of cell density (Abs_{600}). Different cell densities were achieved by different growth times without dilution. The line within the graph indicates a linear fit to the measured values.

Dehydration stress

Dehydration stress was applied to cells through the osmotic pressure of glycerol. Relative cell viability (%CV) was determined using glycerol/H_2O solutions of different concentrations and different effective stress times, as shown in figure 39. The time of stress treatment was varied as indicated, but did not significantly change the rate of cell survival.

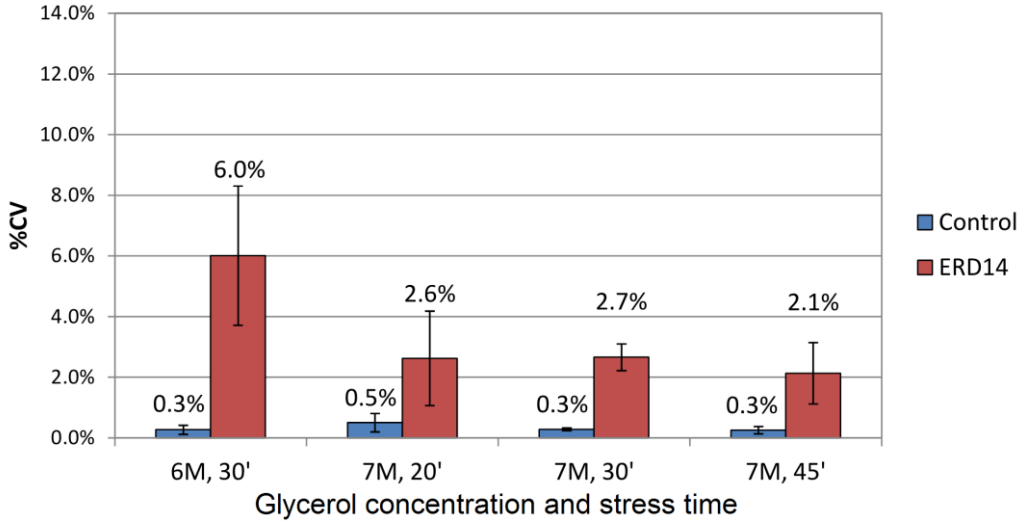

Figure 39. Relative cell viability (%CV) of control and ERD14-containing cells, comparing the effect of dehydration stress treatments of different intensities, i.e. the effect of 6 M and 7 M effective glycerol concentrations using several stressing times relative to unstressed cells.

At all conditions tested, the measured relative cell viability values were below 6%, which was considered to be too low for a reliable assessment of the *in vivo* chaperone function of proteins. Further refinement and the testing of lower glycerol concentrations would be necessary for this stress effect to be of use in the investigation of *in vivo* chaperone effects.

4.2.4. Effect of ERD14 on cell viability after stress treatment

After the *in vivo* chaperone assay was established and tested as described above, it was applied to measure the *in vivo* function of WT ERD14. As presented in figure 40, according to the results of all three *in vivo* chaperone assays – using freezing, high salt and dehydration stress treatments – WT ERD14 was shown to be a competent *in vivo* chaperone, i.e. it is able to at least partially protect its host cell from the adverse effects of three different kinds of abiotic stress.

The results from the freezing tolerance measurements are grayed out, in order to emphasize that the effect observed here is not only – or at least not directly – due to the presence of the ERD14 protein, but to a great extent depending on the actual cell density present during the

stress treatment. It is noted, however, that in an indirect manner, the presence of WT ERD14 is at least partially connected to the observed protective effect, as it is shown to cause growth rates to stay high even after induction (as discussed in section 4.2.5), hereby resulting in cell densities that are higher for ERD14-containing cells than for control cells after equal times of growth, which in turn causes survival rates in the *in vivo* chaperone assay to be high, as discussed above in section 4.2.3. Nevertheless, as far as freezing tolerance is concerned, it can be concluded, that this form of applied stress is not appropriate to be used to measure the *in vivo* chaperone effect of a potentially protective protein, as it is hardly possible to correct for different growth rates of different samples by using different incubation times. It is mentioned, that as a second choice, dilution to a given cell density might be used to correct for the observed effect of dependence on cell density, but this implicates other problems, such as the introduction of different quantities of new nutrients into the sample solutions, which is of unknown effect on cell viability and on ATP-levels within the cells.

In contrast, as salt stress tolerance shows no dependence on cell densities, it can be readily used to establish the *in vivo* protecting effect of possible chaperone proteins, even if growth rates of the respective host cells differ. Furthermore, this measurement can also be used for the comparative analysis of the effect of mutagenesis on the *in vivo* chaperone function of the protein of interest, as is presented in section 4.2.6 below.

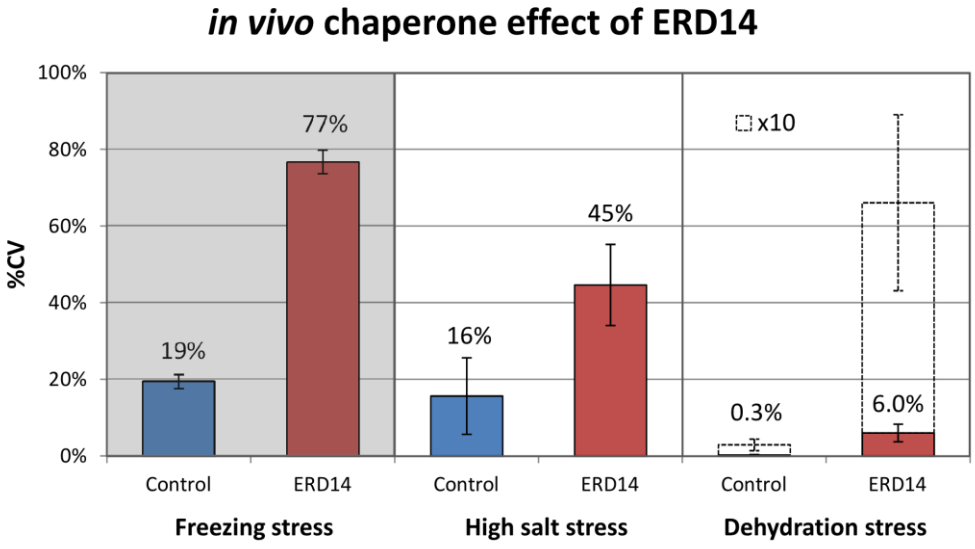

Figure 40. Relative cell viability (%CV) of control and ERD14-containing cells using three different stress treatments: freezing stress, high salt stress and dehydration stress. Results from the freezing stress measurement are grayed out due to the major influence of cell density on the assay result, as discussed in the text. For clarity, the small %CV values from the dehydration stress assay are presented also in a 10x-times magnification, as indicated in the graph.

As far as dehydration tolerance is concerned, the *in vivo* protective effect of WT ERD14 is further confirmed by the presented results. However, measured relative cell viability values (of both, control and WT ERD14-containing cells) are considered to be too low, i.e. the stress treatment to be too strong, for a reliable assessment of the *in vivo* chaperone effect, especially if possible subtle effects caused by mutagenesis are to be analyzed as well. Therefore, further testing and refinement of the dehydration tolerance assay are considered to be necessary before it can be used for more thorough investigations.

4.2.5. Effect of ERD14 on cell growth under stress conditions

As was observed during optimization of the measurement of freezing tolerance, the presence of WT ERD14 causes cell growth to continue excessively even after induction, while within control cells containing the empty pET22b plasmid cell growth is reduced to a minimum as a secondary effect of induction. This effect of induction on cell growth, as well as the effect of other long-term abiotic stresses (of slightly elevated salt concentration in this case), is investigated in more detail below.

Effect of induction on cell growth

Induction is a special form of stress effect for *E. coli* cells, which in many cases results in reduced cell growth by focusing all efforts of the cell on the expression of the target protein. The effect of induction on cell growth was measured and compared for control (empty pET22b plasmid) and WT ERD14-containing cells, as shown in figure 41.

It can be seen that in contrast to the control cells, which cease growing, the expressed WT ERD14 enables its host cells to continue in cell growth exponentially as before induction.

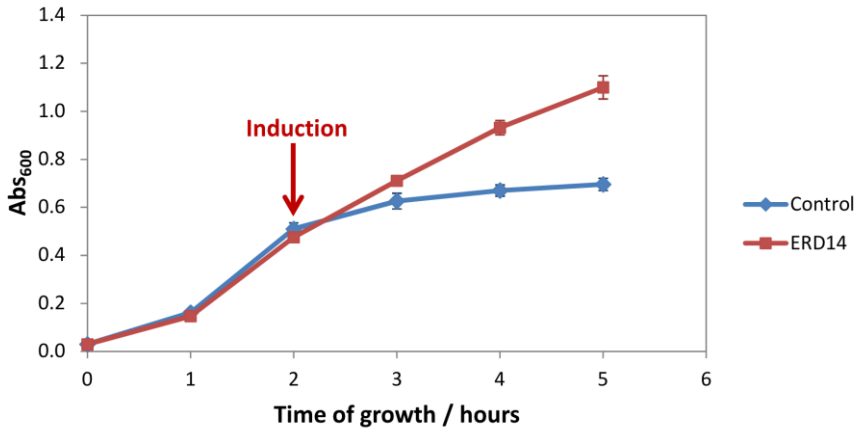

Figure 41. Cell growth after induction for cells containing a plasmid for WT ERD14 expression compared to cells containing the empty pET22b plasmid (control).

Effect of long-term salt exposure on cell growth

The effect of long-term salt exposure, at much lower levels than used in the (short-term) high salt stress analysis, was investigated by the observation of cell growth in NZYM medium of slightly elevated salt content, compared to growth in NZYM medium of normal salt content.

Figure 42 presents the evolution of average optical density (Abs_{600} values after induction) for cells expressing GST or WT ERD14, with or without enrichment of the medium with NaCl. In case of GST-containing cells (A), the growth rate within the exponential growth phase is retarded by the presence of salt, while the growth rate of ERD14-containing cells (B) is almost indifferent to the elevated salt content of the medium. For a quantitative evaluation of salt tolerance during cell growth, the ratio of growth rates with and without active salt stress was calculated for both proteins and is compared in part (C) of the figure. It is noted that the cell density at which the cells enter into the steady state phase of growth (plateau level) is different for media with or without NaCl enrichment. However, this difference was determined to be less indicative of salt tolerance than the growth rate within exponential growth phase and therefore, was not investigated further.

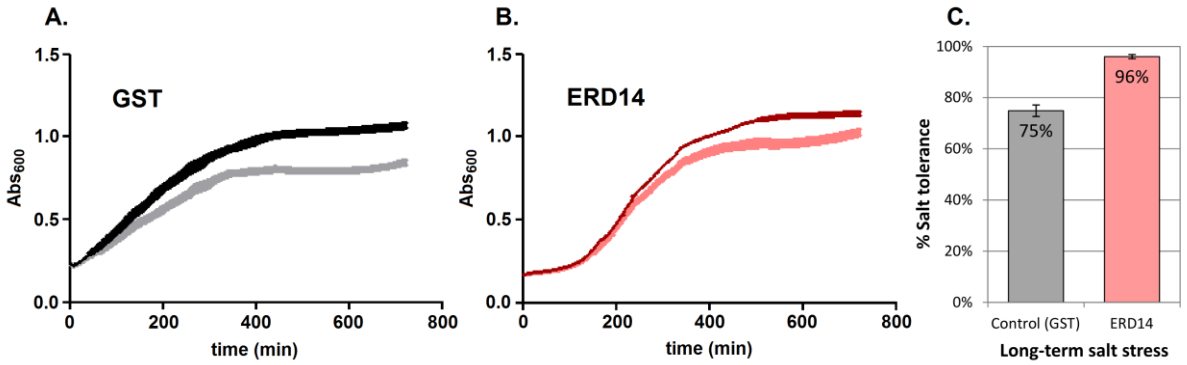

Figure 42. Cell growth of cells in media with (light colors) or without (dark colors) NaCl enrichment for cells expressing either GST (control, A) or WT ERD14 (B). Salt tolerance during cell growth was compared quantitatively as the ratio of growth rates with and without salt exposure in percent (C).

It can be concluded from the presented experiment that WT ERD14 is not only able to restore the adverse effects of short-term high salt stress (as discussed in section 4.2.4 above), but also provides potent protection from the growth-retarding effect of long-term salt exposure. Furthermore it is mentioned that this test can be used in general as a further confirmation of *in vivo* chaperone effects of proteins, with the restriction, however, to be only

applicable to proteins that enable their host cell to continue in cell growth even after induction, just as ERD14 (e.g. the test is not applicable to mutant ERD14 K3, see below).

4.2.6. Mapping function to sequence: measurement of the effect of mutagenesis on the *in vivo* chaperone function of ERD14

The *in vivo* chaperone function of WT ERD14 has been demonstrated in several different experiments, as presented and discussed above. In this section it is shown how the same *in vivo* chaperone assays and comparative growth experiments can be used to map this protective function to specific regions of the sequence of ERD14 by analyzing the effect of directed mutagenesis on the extent of the measured chaperone effects. This is presented herein on one of the ERD14 mutants, ERD14 K3, in which all three K-segments have been removed from the sequence of the protein.

ERD14 K3 was chosen for this comparative analysis as it was inferred from the structural investigations presented in section 4.1 that the K-segments might represent *in vivo* active binding sites of the protein. Therefore, the effect of the deletion of these regions from the protein is of major interest. Figure 43 shows the comparison of the *in vivo* chaperone effect upon high salt stress (A) as well as the observed growth rates after induction (B) of the ERD14 K3 mutant compared to the control and the wild type protein.

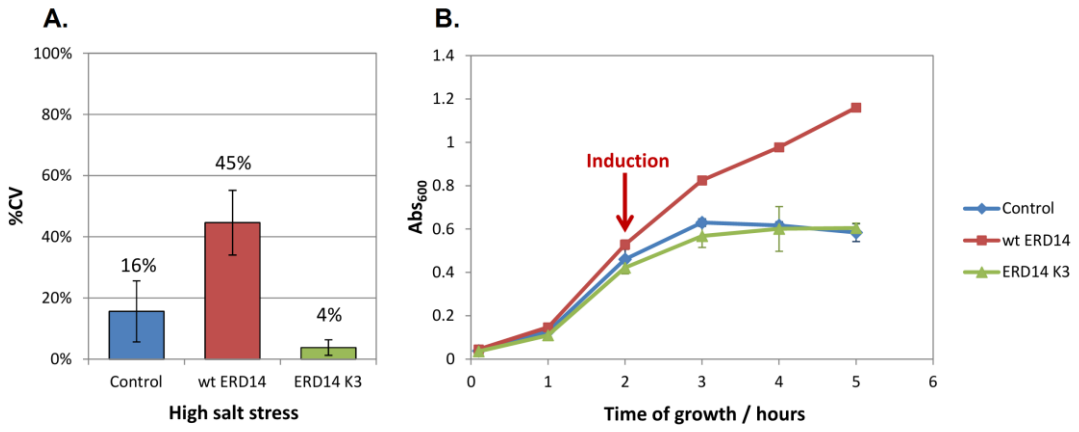

Figure 43. Relative cell viability (%CV) after high salt stress (A) and growth after induction (B) for control cells, and cells containing either WT ERD14 or the ERD14 K3 mutant.

In both cases (the chaperone assay as well as the growth test) the protective effect observed for WT ERD14 is completely lost through the deletion of the three K-segments, which unmistakably demonstrates the functional relevance of these regions of the protein.

It should be noted, however, that for a more thorough functional analysis resulting in a more detailed understanding of the chaperone function of ERD14, the investigations still need to be continued, involving e.g. the measurement of the *in vivo* chaperone effect of several more mutants of ERD14.

5. Conclusions

5.1. Structural and dynamic properties of ERD14 *in vitro* and *in vivo*

In vitro assessment of structure and dynamics of ERD14

The presented NMR results provide evidence of the almost completely disordered state of ERD14 under near-native conditions. This picture is in full agreement with previous data obtained by CD, limited proteolysis, heat resistance and 1D ^{1}H NMR (Kovács, Kalmár, et al., 2008).

By considering subtle details of SCS and relaxation data, small deviations from a random coil structure were detected, and five regions of restricted motion have been identified. These regions, which have a tendency for helix formation on the magnitude of about 5-15%, partly overlap with the conserved segments of ERD14 (see table 5 and figure 27). The relatively low values of helical preference suggest that the above regions are only partially structured and constantly fluctuate between a helically dense and an open or random coil state.

Furthermore, partial helicity might be an inherent, general property of the K-segments as observed chemical shifts are similar for the conserved amino acids of the different K-segments of ERD14 and also for K-segments of the other dehydrin assigned so far, namely dehydrin 1a from *V. riparia* (Findlater and Graether, 2009). A comparison of the $^{1}H^{N}$ and ^{15}N chemical shifts of the two fully conserved K-segments of dehydrin 1a and ERD14 is shown in table 6 (a full comparison of the chemical shifts of all K-segments within these proteins is listed in addendum 6.4).

The general structural analysis and comparison of the basic mutants of ERD14 shows similar properties for all tested deletion mutants, i.e. no major structural changes upon mutagenesis can be observed. From this it is concluded that no forms of tertiary structure, i.e. no molten-globule-like or long-range interactions, are present within ERD14, as these would be disrupted by deletions of long stretches of protein sequence and show up as a major change upon mutagenesis in one or several global properties (e.g. the hydrodynamic radius or the CD spectrum). Moreover, the apparent molecular weight observed by gel filtration for WT ERD14 as well as its mutants is about 6-7 times bigger than the respective calculated value,

indicating intensive hydration and an overall extended structure close to random coil (in contrast to molgen-globule or pre-molten-globule states). Furthermore, the observation that deletion mutants of ERD14 show the same overall disorder than the wild type protein indicates that these can be used to assess the importance of the deleted regions within the function of the protein because their absence does not alter general structural properties of the protein (compare this behavior to globular proteins where mutations induce structural *and* functional changes at the same time with effects that are hard to separate).

	dehydrin 1a		ERD14		difference	
K	8.265	122.3	8.289	122.337	-0.024	-0.037
I	8.171	122.897	8.162	122.268	0.009	0.629
K	8.402	126.032	8.363	125.317	0.039	0.715
E	8.36	122.707	8.319	121.609	0.041	1.098
K	8.3	122.668	8.259	122.031	0.041	0.637
L	8.283	125.258	8.267	124.934	0.016	0.324
P				136.354		
G	8.486	109.986	8.608	109.598	-0.122	0.388

Table 6. Comparison of the $^{1}H^{N}$ and ^{15}N chemical shifts for the two fully conserved K-segments of dehydrin 1a and ERD14. The good agreement between the chemical shift values indicates structural similarity of the K-segments.

Assessment of structure under live conditions: in-cell NMR

The same overall intrinsically disordered structure of ERD14 can be observed by NMR within the live cell as in dilute solution. In the first place it is emphasized that the present results further confirm the existence of IDPs in general and of dehydrins and ERD14 in particular, within the living cell in a state of disorder (as similarly observed for α-synuclein; Croke et al., 2008; McNulty et al., 2006), without undergoing binding and/or folding transitions (e.g. as observed for the binding of tau protein; Bodart et al., 2008).

However, disappearance of a few resonances of ERD14 indicates for some residues selective binding within the cell, which could be localized to regions within the conserved K-segments near the relatively restricted regions of partial helicity. These results underline the importance of the K-segments in binding – and therefore function – of ERD14 within the living cell, as has already long been suggested for dehydrins (compare section 1.2), but not confirmed within the live cell (except for the investigation of *E. coli* growth inhibition by dehydrins, claimed to be connected to either or both the ChP- and K-segments (Campos et al., 2006), as discussed in more detail further below).

5.2. Functional analysis of ERD14: *in vivo* chaperone function

In the present work an *in vivo* chaperone assay has been developed and tested. One of the three proposed stress procedures was optimized extensively, and feasibility was demonstrated on wild type ERD14 as well as one of its mutants.

Method development: the *in vivo* chaperone assay

A procedure for the measurement of the protective influence of overexpressed proteins on the short-term stress tolerance of *E. coli* cells, i.e., the *in vivo* chaperone function, has been developed, and three different stress effects have been tested for usability. Freezing of the *E. coli* cells as a form of cold stress turned out to depend in a large extent on cell density, and could not be reliably used for the measurement of the *in vivo* chaperone function (as different growth rates were observed for different constructs, which is hard to correct for, but the fact of different growth rates was further analyzed). Dehydration stress, as induced by the presence of high amounts of glycerol, has been preliminarily tested, though further optimization is needed in order to obtain a versatile tool for the analysis of chaperone proteins and their mutants. High salt stress, as induced by high amounts of NaCl, has been tested and successfully optimized. Also, feasibility has been demonstrated on ERD14 and one of its mutants. It is concluded that the *in vivo* chaperone assay presented herein using high salt stress can be used for both the confirmation (detection) of the *in vivo* chaperone activity of a protein, as well as for the mapping of an observed chaperone effect to the sequence through the comparison of different mutants of the protein. This has been demonstrated for ERD14.

It is emphasized that with the use of this assay for the analysis of further mutants of ERD14 (and of other ID chaperones), a more detailed mechanistic model of the chaperone function of ID chaperones can be deduced, leading to a general and deeper understanding of chaperone function.

An alternative method to test stress resistance

Furthermore, an alternative *in vivo* chaperone assay has also been presented and used for WT ERD14 in the form of a long-term salt tolerance measurement. However this test is only applicable to cells that continue to grow after induction of the target protein (e.g. the assay is not applicable to mutant ERD14 K3). This assay is similar to the ones used to assess the protective effect of LEA proteins within *E. coli* cells in two other studies (see section 1.3). However, in the present case upon long-term salt exposure a difference in the growth rate was

observed, while in the other two studies no change within the growth rate, but rather a shortening of the lag-phase of growth was observed. It is noted, though, that the performed experiments cannot be compared directly, as different experimental setups were applied (using concentrations between 0.5-1 M NaCl compared to less than 0.2 M NaCl, and the fact that different proteins and controls have been used).

Conclusions from the analysis of ERD14

ERD14 was shown herein to have a potent protective effect *in vivo* as observed under four different stress conditions, which means that ERD14 is not only a potent chaperone *in vitro*, but also *in vivo*. Combined with structural analysis, it can be stated that ERD14 is not only disordered inside the cell, but that this intrinsic disorder is also its functional state, as opposed to the general view that IDPs might acquire their functional state only through folding upon binding to their substrate (see also further below).

The analysis of the K3 deletion mutant of ERD14 proved that the *in vivo* chaperone function observed for ERD14 is connected to (at least one of) the conserved K-segments, as already proposed for dehydrins in general based on indirect measures and sequence characteristics. The results from the in-cell NMR analysis further corroborate this statement as they indicate the role of specific binding of these regions within the living cell. Therefore, both the structural and functional analysis results underline the importance of the K-segments in the function of ERD14 and indicate their functional significance within dehydrins in general.

However, the exact mode of function, could not be determined from these initial results. For the elucidation of a more elaborate model of function of ERD14, the analysis of further mutants (in progress, manuscript in preparation, Szalainé Ágoston et al., 2013) and perhaps the use of other techniques will be necessary. Nevertheless, the first results presented herein already give valuable information and confirm the importance of the K-segments within the *in vivo* function of the protein.

As a further result, it has been shown that ERD14 promotes the exponential growth of *E. coli* cells even after induction and during growth conditions with slightly elevated salt concentration, in contrast to the control (empty vector) or the K3-deficient mutant (where growth is retarded upon induction) and GST protein (where the elevated salt concentration slows growth). These results oppose those reported for the closest homologous dehydrin, ERD10, which was observed to inhibit *E. coli* cell growth (Campos et al., 2006). Based on the

effect of deletion mutants, the authors suggested that the cationic nature of the ChP-segment ("K-stutter") and the K-segment of the dehydrin might be responsible for this antibiotic effect. However, in their experiments, the full ERD10 sequence (or the sequences mutated on purpose) was not solely expressed, but also three shorter segments, and the presence of the additional peptides correlated with the antibiotic effect. In contrast, no expression of disturbing extra protein parts was observed within the present work. It is also noted that in the experiments by Campos et al., bacterial growth was observed in XLBlue cells, in contrast to the BL21 (DE3) Star cells used in the present work, which can induce differences in expression patterns as well as the active cellular background, including stress resistance and growth characteristics of the bacterial strain.

It has also been suggested that due to the growth inhibitory effect of dehydrins on *E. coli*, they might be of only limited use as a model organism in studying the protective effect of dehydrins (Tunnacliffe and Wise, 2007). However, the present results prove that an anti-microbial effect (i.e. growth inhibition) on *E. coli* cells is not a general property of dehydrins, nor their ChP- or K-segments. On the other hand, the *in vivo* chaperone assay developed and introduced in this work enables – being in this sense the first of its kind – testing for the protective effect of proteins that retard or inhibit *E. coli* cell growth through their induced overexpression, hereby providing a valuable method for testing of dehydrins in stress tolerance using *E. coli* as a model organism.

Furthermore, it is also noted that in the present investigation a three-fold protective effect of ERD14 has been shown *in vivo* in contrast to several other dehydrins, which were effective in only one or the other stress effect investigated (compare to section 1.2).

5.3. Conclusions on the structure-function relationship of ERD14 and intrinsically disordered chaperones in general

The intrinsic disorder of ERD14 was demonstrated *in vitro* (in detail), as well as *in vivo*. Next to the known *in vitro* chaperone function of ERD14, there is also evidence for its *in vivo* chaperone function. The importance of these results is twofold: it was shown by the structural investigations that ERD14 is an IDP, which retains its intrinsic disorder even in the cell, and it was also shown through the *in vivo* chaperone assay that intrinsic disorder is the active, functional state of ERD14. This contradicts the general conception in the IDP field that (most) IDPs have to fold upon binding to become active, i.e. "to slip into their active role", as was generally supposed so far to be the case for dehydrins as well (see section 1.1).

Furthermore, the presented results indicate that ERD14 is an IDP that does survive in the cell in a disordered state, which is – speaking of IDP proteins in general – still a subject of heated discussions. However, it should be added that the presented results were measured in a protease-deficient cell line of *E. coli* under continuous induction, and therefore these results cannot prove protease stability of ERD14 under native conditions. Nevertheless, the results indicate that the intrinsic disorder of ERD14 is its functional state within the living cell, which on its own implicates that this state has to survive within the cell at least for a short time.

Furthermore, it was shown that the conserved K-segments of ERD14 contain preformed helical elements for binding within the cell, and that these segments are essential for the function of the protein. The K-segment is the most conserved region of dehydrins, and the results presented herein confirm its importance in function.

The amphipathic character of the possible α-helix of the K-segments can, in principle, provide both membrane binding and binding of misfolded proteins at their temporarily exposed hydrophobic patches. In the first case, membrane binding can result in either membrane stabilization (however, ERD14 was shown not to alter membrane fluidity; Kovács, Kalmár, et al., 2008) or can represent a kind of anchoring of the chaperone to the membrane where the water retention potential of this overall highly charged protein might be needed most within the cell (however, using only temporarily folded helices, an eventual anchoring effect cannot be too strong). In the latter case of protein binding, ERD14 might help in the refolding of misfolded proteins either passively (by preventing aggregation through entropic exclusion) or actively (by providing some entropic energy for the rescue from a folding trap as suggested by the entropy transfer model, described in detail in section 1.5).

To test for these hypotheses and the possible roles of the extremely mobile hydrophobic patch within ERD14, of phosphorylation within the S-segment, and the eventual relevance of the length of the disordered chain in general (e.g. for the entropy transfer, or the entropic exclusion effect), various mutants will have to be tested using either the presented assay or similar investigations (in progress, manuscript in preparation; Szalainé Ágoston et al., 2013).

However, the protective effect of these proteins within the cell might be even more complex. Water retention, ion buffering and radical scavenging potentials of dehydrins may all operate in addition to the chaperone effect investigated in this work. This shows that although important results could be presented herein, still a lot of further investigations are needed to be able to fully understand the mode of action of dehydrins, and/or of intrinsically disordered chaperones in general.

6. Addenda

6.1. Assigned chemical shifts of ERD14

All assigned resonances of wild type ERD14 at 288 K and pH 6.5, in 10 mM MES buffer are listed below. Missing values indicate missing resonances, while "-" indicates nuclei that are not present in the given amino acid. Chemical shifts that are grayed out and marked (*) had to be excluded from further analysis due to changes in the sample (observed within the HSQC and H(CA)CON spectra), as discussed in section 4.1.1.

Amino acid number	name	H^N	N	C^O	C^α	C^β	H^α
1	MET						
2	ALA						
3	GLU	8.745	120.478				4.310
4	GLU	8.686	123.468	176.210	56.284	30.237	4.275
5	ILE	8.441	124.002	176.160	60.904	38.278	4.118
6	LYS	8.530	126.489	176.004	55.863	33.178	4.354
7	ASN	8.564	120.771	174.695	53.108	38.922	4.692
8	VAL	8.224	122.260	174.376	59.790	32.533	4.412
9	PRO	-	139.429	177.017	63.046	32.070	4.397
10	GLU	8.614	121.277	176.626	56.801	30.071	4.197
11	GLN	8.439	121.010	175.770	55.630	29.658	4.310
12	GLU	8.510	123.017	176.243	56.188	30.306	4.280
13	VAL	8.338	123.956	174.405	59.908	32.416	4.386
14	PRO	-	139.720	176.754	62.965	32.150	4.389
15	LYS	8.510	122.686	176.704	56.098	33.087	
16	VAL	8.250	122.160	175.787	61.842	33.030	4.103
17	ALA	8.559	128.722	177.841	52.405	19.227	4.393
18	THR	8.238	114.133	174.701	61.735	69.766	4.309
19	GLU	8.558	123.266	176.572	56.508	30.188	
20	GLU	8.551	122.410	176.691	56.636	30.249	4.295
21	SER	8.475	117.289	174.772	58.306	63.743	4.476
22	SER	8.461	118.309	174.429	58.306	63.743	4.461
23	ALA	8.358	126.026	177.775	52.709	19.138	4.312
24	GLU	8.358	120.087	176.697	56.549	30.115	4.254
25	VAL	8.269	121.758	176.634	62.487	32.474	4.147
26	THR	8.301	118.040	174.370	62.070	69.766	4.332
27	ASP	8.390	122.999	176.457	54.339	40.974	4.608
28	ARG	8.412	122.116	177.157	56.625	30.235	4.244

Amino acid number	name	H^N	N	C^O	C^α	C^β	H^α
29	GLY	8.546	109.242	174.536	45.436	-	
30	LEU	7.980	121.097	177.323	55.511	42.146	4.233
31	PHE	8.159	119.297	175.728	57.636	39.568	4.569
32	ASP	8.199	121.477	176.319	54.213	40.865	4.533
33	PHE	8.112	120.163	176.236	58.384	39.098	4.513
34	LEU	8.158	122.302	177.931	55.335	41.912	4.244
35	GLY	8.045	108.905	174.051	45.340	-	3.865
	GLY	-	-	-	-	-	3.865
36	LYS	8.101	120.967	176.697	55.980	33.002	4.290
37	LYS	8.473	123.404	176.754	56.192	32.935	4.257
38	LYS	8.459	123.680	176.267	56.347	32.960	4.233
39	ASP	8.463	121.922	176.286	54.163	40.981	4.596
40	GLU	8.505	122.299	176.640	56.506	30.247	4.342
41	THR	8.365	115.964	174.341	62.154	69.682	4.275
42	LYS	8.408	125.657	174.463	54.046	32.474	4.631
43	PRO	-	137.479	177.034	63.046	31.989	4.397
44	GLU	8.677	121.176	176.543	56.508	30.130	4.229
45	GLU	8.476	122.469	176.292	56.488	30.229	4.346
46	THR	8.378	119.118	172.703	59.895	69.682	4.554
47	PRO	-	139.399	176.844	62.885	32.070	4.442
48	ILE	8.356	121.992	176.189	60.982	38.629	4.103
49	ALA	8.484	128.867	177.816	52.463	19.168	4.325
50	SER	8.395	115.741	174.943	58.473	63.743	4.385
51	GLU	8.568	122.750	176.399	57.035	29.778	4.188
52	PHE	8.158	119.934	175.787	57.887	39.484	4.589
53	GLU	8.173	122.151	176.136	56.454	30.364	4.205
54	GLN	8.390	121.801	175.906	55.712	29.274	
55	LYS	8.392	123.174	176.432	55.903	32.984	
56	VAL	8.175	121.477	175.770	62.135	32.950	4.043
57	HIS	8.646	123.938	174.496	55.453	29.661	4.695
58	ILE	8.344	124.391	175.946	60.728	38.747	4.174
59	SER	8.500	120.835	173.951	58.091	63.718	4.448
60	GLU	8.485	124.492	174.323	54.222	29.719	4.591
61	PRO	-	137.236	176.679	62.725	32.070	4.404
62	GLU	8.585	122.787	174.694	54.339	29.485	4.520
63	PRO	-	137.293	176.795	62.965	32.070	4.382
64	GLU	8.569	121.728	176.502	56.508	30.306	
65	VAL	8.292	122.851	175.882	62.164	32.614	4.035
66	LYS	8.474	125.867	176.317	55.980	32.978	4.284
67	HIS	8.635	121.443	174.949	55.778	29.712	4.645
68	GLU	8.679	123.017	176.548	56.684	30.364	4.288

Amino acid number	name	H^N	N	C^O	C^α	C^β	H^α
69	SER	8.627	117.951	174.925	58.389	63.743	4.444
70	LEU	8.441	124.268	177.932	56.006	41.853	4.275
71	LEU	8.105	121.046	178.009	55.912	41.968	4.214
72	GLU	8.144	120.817	177.098	57.153	30.012	4.169
73	LYS	8.198	121.504	177.062	56.949	32.734	4.184
74	LEU	8.140	121.902	177.482	55.453	42.146	4.244
75	HIS	8.371	119.254	175.002	55.712	29.278	4.664
76	ARG	8.424	122.741	176.589	56.391	30.833	4.339
77	SER	8.549	117.208	174.453	58.473	63.659	4.457
78	ASP	8.454	122.548	176.548	54.280	41.033	4.676
79	SER	8.378	116.389	174.990	58.557	63.659	4.483
80	SER	8.492	118.174	174.866	58.706	63.672	4.494
81	SER	8.412	117.825	174.760	58.473	63.743	4.529
82	SER	8.439	118.013	174.750	58.389	63.743	4.529
83	SER	8.465	118.085	174.772	58.445	63.639	4.529
84	SER	8.469	118.088	174.779	58.407	63.690	4.529
85	SER	8.467	118.111	174.808	58.521	63.665	4.476
86	GLU	8.476	122.612	176.671	56.653	30.071	4.302
87	GLU	8.402	121.379	176.701	56.430	30.236	
88	GLU	8.499	122.382	177.215	56.732	30.229	4.279
89	GLY	8.548	110.238	174.577	45.172	-	3.988
	GLY	-	-	-	-	-	3.988
90	SER	8.385	115.954	174.831	58.557	63.827	4.465
91	ASP	8.608	122.161	177.163	54.749	40.915	4.606
92	GLY	8.411	108.973	174.925	45.758	-	
93	GLU	8.263	120.950	177.269	56.977	29.895	4.214
94	LYS	8.324	121.562	177.280	56.986	32.571	4.225
95	ARG	8.260	121.056	176.685	56.904	30.599	
96	LYS	8.258	122.311	176.850	56.589	32.885	4.244
97	LYS	8.331	122.682	176.773	56.408	32.950	4.260
98	LYS	8.387	123.303	176.671	56.455	32.992	4.254
99	LYS	8.436	123.349	176.550	56.207	32.948	
100	GLU	8.489	123.174	176.292	56.162	30.460	4.283
101	LYS	8.498	123.792	176.439	56.196	32.989	4.288
102	LYS	8.469	123.809	176.391	55.981	33.069	4.286
103	LYS	8.528	125.136	174.636	54.261	32.326	4.577
104	PRO	-	137.293	177.134	62.990	32.206	4.513
105	THR	8.466	115.236	174.801	61.903	69.682	4.381
106	THR	8.200	116.156	174.423	61.735	69.766	4.361
107	GLU	8.516	123.781	176.457	56.418	30.248	
108	VAL	8.219	121.395	176.089	62.365	32.704	4.061

Amino acid number	name	H^N	N	C^O	C^α	C^β	H^α
109	GLU	8.481	125.317	176.416	56.508	30.364	4.283
110	VAL	8.374	123.533	176.218	62.487	32.650	4.034
111	LYS	8.524	126.012	176.935	56.156	32.896	4.320
112	GLU	8.578	123.082	177.025	57.121	30.071	4.173
113	GLU	8.651	121.489	176.933	57.013	29.988	
114	GLU	8.384	121.907	176.874	56.757	30.089	4.211
115	LYS	8.279	122.467	176.972	56.590	32.709	4.249
116	LYS	8.360	122.343	177.488	56.918	32.845	4.254
117	GLY	8.502	109.808	174.358	45.340	-	3.924
	GLY	-	-	-	-	-	3.924
118	PHE	8.128	120.503	176.260	58.618	39.333	4.508
119	MET	8.284	121.397	176.478	55.687	32.474	
120	GLU	8.289	121.888	176.939	57.131	30.012	
121	LYS	8.249	121.853	176.998	56.747	32.546	4.225
122	LEU	8.138	122.752	177.724	55.511	42.088	4.248
123	LYS	8.207	121.676	176.720	56.566	32.767	4.223
124	GLU	8.243	121.490	176.213	56.482	30.306	4.193
125	LYS	8.314	122.402	176.283	55.987	32.885	4.283
126	LEU	8.344	125.071	175.312	52.874	41.384	4.606
127	PRO	-	136.240	177.653	63.325	31.955	4.412
128	GLY	8.558	109.324	174.028	45.196	-	3.914
	GLY	-	-	-	-	-	3.914
129	HIS	8.305	118.821	174.843	55.712	29.864	4.640
130	LYS	8.436	123.385	176.218	56.066	33.176	4.303
131	LYS	8.587	125.071	174.776	54.296	32.240	4.576
132	PRO	-	137.606	177.256	63.206	31.989	4.389*
133	GLU	8.764	121.204	176.502	56.566	29.895	4.259*
134	ASP	8.367	121.582	177.057	54.276	41.091	4.600*
135	GLY	8.486	110.093	174.606	45.423	-	4.022*
	GLY	-	-	-	-	-	4.022*
136	SER	8.279	115.984	174.447	58.640	63.743	4.399*
137	ALA	8.352	125.853	177.755	52.304	19.138	4.355*
138	VAL	8.082	119.518	175.941	62.070	32.708	4.054
139	ALA	8.380	128.290	177.198	52.170	19.168	4.286
140	ALA	8.292	123.975	177.058	51.925	19.312	4.275
141	ALA	8.301	125.209	175.501	50.294	17.937	4.567
142	PRO	-	135.842	176.747	62.645	31.989	
143	VAL	8.338	121.650	176.136	62.339	32.642	4.045*
144	VAL	8.385	126.214	175.831	62.055	32.748	4.097*
145	VAL	8.470	128.010	173.927	59.615	32.580	4.401*
146	PRO	-	141.167	174.078*	61.363*		4.662*

Amino acid number	name	H^N	N	C^O	C^α	C^β	H^α
147	PRO	-	136.116*	174.722*	61.192*		4.708*
148	PRO	-	135.273	177.009	62.645	31.989	4.442*
149	VAL	8.294	120.863	176.260	62.413	32.767	
150	GLU	8.565	125.115	176.201	56.215	30.381	4.290
151	GLU	8.490	123.238	175.864	56.125	30.477	4.212
152	ALA	8.405	125.585	177.289	52.405	19.168	4.214
153	HIS	8.524	118.774	172.553	53.203	28.944	4.957
154	PRO	-	137.508	176.976	63.046	32.070	4.459
155	VAL	8.452	121.259	176.295	62.350	32.711	4.075
156	GLU	8.589	125.462	176.305	56.280	30.364	4.279
157	LYS	8.515	123.816	176.446	56.207	32.906	
158	LYS	8.487	123.106	177.149	56.443	33.027	4.294
159	GLY	8.570	110.327	174.258	45.235	-	
160	ILE	8.056	120.257	176.590	61.555	38.512	4.112
161	LEU	8.363	125.268	177.523	55.349	41.985	4.323
162	GLU	8.321	121.934	176.543	56.742	30.188	
163	LYS	8.289	122.337	176.876	56.590	32.709	4.244
164	ILE	8.162	122.268	176.614	61.490	38.336	4.075
165	LYS	8.363	125.317	176.663	56.742	32.767	4.211
166	GLU	8.319	121.609	176.213	56.587	30.380	4.061
167	LYS	8.259	122.031	176.289	55.864	32.860	4.261
168	LEU	8.267	124.934	175.270	52.932	41.443	4.610
169	PRO	-	136.354	177.659	63.492	31.872	4.418
170	GLY	8.608	109.598	173.585	44.921	-	
171	TYR	7.974	120.724	175.102	57.973	38.864	4.464
172	HIS	8.244	124.377	171.733	52.815	29.426	4.793
173	PRO	-	136.724	176.778	62.725	32.070	
174	LYS	8.628	122.060	176.986	56.332	32.949	4.357
175	THR	8.339	115.984	174.618	61.568	69.766	4.422
176	THR	8.313	117.181	174.536	61.735	69.850	4.386
177	VAL	8.294	122.766	176.301	62.430	32.709	
178	GLU	8.535	124.760	176.646	56.625	30.130	
179	GLU	8.464	122.382	176.580	56.611	30.309	
180	GLU	8.481	122.458	176.493	56.356	30.200	
181	LYS	8.413	123.238	176.587	56.247	32.841	4.287
182	LYS	8.414	123.183	176.416	56.194	32.968	4.317
183	ASP	8.435	121.471	175.970	54.456	41.033	4.578
184	LYS	8.239	121.198	175.687	56.156	33.061	4.316
185	GLU	8.094	127.333	181.268	58.032	30.892	

Data deposition

The chemical shift assignments of ERD14 presented herein have also been deposited in the Biological Magnetic Resonance Data Bank (BMRB; http://www.bmrb.wisc.edu) under the BMRB accession number 16876.

6.2. NMR relaxation data of ERD14

All measured NMR relaxation data of wild type ERD14 at 288 K and pH 6.5, in 10 mM MES buffer are listed below with the respective standard deviations of the measurement. The table caption "hetNOE" refers to heteronuclear ^{15}N-{^{1}H} NOE ratios. SD refers to the respective standard deviations.

Amino acid number	name	R_1 / sec^{-1}	SD (R_1)	R_2 / sec^{-1}	SD (R_2)	$R_{1\rho}$ / sec^{-1}	SD ($R_{1\rho}$)	hetNOE	SD (hetNOE)
1	Met								
2	Ala								
3	Glu	0.831	0.079	2.036	0.090	1.339	0.129	-1.031	0.052
4	Glu	1.255	0.040	1.674	0.038	1.731	0.009	-0.749	0.001
5	Ile	1.425	0.014	1.957	0.017	2.016	0.023	-0.419	0.003
6	Lys	1.476	0.036	2.208	0.038	2.247	0.013	-0.342	0.010
7	Asn	1.444	0.083	2.435	0.079	2.514	0.017	-0.126	0.012
8	Val	1.554	0.026	3.305	0.039	3.634	0.020	0.064	0.001
9	Pro								
10	Glu	1.559	0.016	2.966	0.019	2.990	0.021	-0.032	0.002
11	Gln	1.553	0.033	3.118	0.041	3.115	0.010	0.002	0.000
12	Glu	1.541	0.040	3.138	0.043	3.244	0.011	0.009	0.005
13	Val	1.509	0.003	3.205	0.012	3.215	0.020	-0.046	0.002
14	Pro								
15	Lys	1.528	0.017	3.117	0.042	3.215	0.015	-0.043	0.001
16	Val	1.536	0.021	3.621	0.045	3.672	0.013	0.050	0.002
17	Ala	1.478	0.030	3.113	0.042	3.033	0.011	-0.074	0.004
18	Thr	1.404	0.045	2.912	0.061	2.934	0.023	-0.098	0.002
19	Glu	1.510	0.038	3.096	0.068	3.141	0.015	-0.044	0.005
20	Glu	1.522	0.049	3.017	0.057	3.079	0.012	-0.046	0.001
21	Ser	1.449	0.081	2.888	0.061	2.925	0.018	-0.080	0.001
22	Ser	1.405	0.066	2.939	0.099	2.927	0.018	-0.058	0.003
23	Ala	1.493	0.052	3.085	0.068	3.169	0.008	-0.016	0.005
24	Glu	1.514	0.032	3.150	0.042	3.151	0.021	0.010	0.003
25	Val	1.564	0.018	3.757	0.035	3.772	0.022	0.077	0.002
26	Thr	1.569	0.054	3.628	0.065	3.691	0.012	0.071	0.001
27	Asp	1.622	0.057	3.737	0.078	3.709	0.014	0.123	0.005
28	Arg	1.654	0.054	4.114	0.055	4.148	0.011	0.219	0.005
29	Gly	1.619	0.053	4.044	0.084	3.923	0.029	0.235	0.003
30	Leu	1.685	0.029	4.655	0.050	4.713	0.011	0.283	0.008
31	Phe	1.662	0.034	4.554	0.052	4.604	0.012	0.247	0.009
32	Asp	1.638	0.035	4.815	0.056	4.794	0.015	0.202	0.008
33	Phe	1.647	0.043	4.697	0.061	4.735	0.014	0.266	0.008

Amino acid number	name	R_1 / sec^{-1}	SD (R_1)	R_2 / sec^{-1}	SD (R_2)	$R_{1\rho}$ / sec^{-1}	SD ($R_{1\rho}$)	hetNOE	SD (hetNOE)
34	Leu	1.646	0.028	4.600	0.040	4.695	0.057	0.235	0.019
35	Gly	1.629	0.048	4.198	0.074	4.039	0.012	0.243	0.009
36	Lys	1.637	0.032	4.789	0.054	4.878	0.013	0.233	0.014
37	Lys	1.633	0.038	4.026	0.070	4.092	0.009	0.184	0.007
38	Lys	1.609	0.043	4.292	0.061	4.294	0.018	0.177	0.012
39	Asp	1.606	0.048	4.026	0.089	4.067	0.015	0.160	0.002
40	Glu	1.591	0.042	3.617	0.051	3.675	0.010	0.120	0.002
41	Thr	1.544	0.068	3.672	0.090	3.698	0.011	0.135	0.002
42	Lys	1.512	0.036	3.782	0.053	3.923	0.013	0.015	0.004
43	Pro								
44	Glu	1.522	0.017	3.971	0.027	3.978	0.011	0.011	0.005
45	Glu	1.603	0.027	3.398	0.037	3.428	0.017	0.069	0.002
46	Thr	1.448	0.041	3.987	0.052	4.102	0.030	0.037	0.003
47	Pro								
48	Ile	1.502	0.006	3.647	0.018	3.815	0.016	-0.003	0.001
49	Ala	1.519	0.024	3.750	0.036	3.690	0.016	0.003	0.002
50	Ser	1.420	0.054	3.691	0.069	3.727	0.016	0.041	0.006
51	Glu	1.519	0.025	4.294	0.049	4.340	0.012	0.054	0.003
52	Phe	1.591	0.038	4.042	0.048	4.108	0.009	0.107	0.004
53	Glu	1.603	0.030	4.137	0.048	4.244	0.010	0.112	0.005
54	Gln	1.559	0.043	4.207	0.065	4.286	0.015	0.094	0.003
55	Lys	1.584	0.055	3.596	0.063	3.655	0.018	0.097	0.003
56	Val	1.545	0.029	4.529	0.039	4.490	0.016	0.081	0.005
57	His	1.481	0.077	4.243	0.117	4.396	0.024	0.032	0.006
58	Ile	1.453	0.050	4.179	0.123	4.241	0.012	-0.032	0.002
59	Ser	1.417	0.046	4.704	0.086	4.662	0.014	-0.079	0.003
60	Glu	1.389	0.030	4.615	0.089	4.697	0.022	-0.051	0.003
61	Pro								
62	Glu	1.508	0.031	4.405	0.055	4.558	0.024	0.032	0.005
63	Pro								
64	Glu	1.483	0.014	4.742	0.040	4.721	0.009	0.035	0.006
65	Val	1.486	0.014	4.098	0.056	4.308	0.007	0.057	0.003
66	Lys	1.528	0.028	4.833	0.071	4.836	0.014	0.057	0.005
67	His	1.495	0.051	5.402	0.066	5.308	0.026	0.134	0.005
68	Glu	1.486	0.068	4.771	0.111	4.845	0.031	0.147	0.003
69	Ser	1.519	0.062	5.184	0.103	5.200	0.022	0.192	0.011
70	Leu	1.608	0.056	4.776	0.100	4.926	0.008	0.223	0.016
71	Leu	1.629	0.034	4.651	0.039	4.638	0.021	0.244	0.019
72	Glu	1.616	0.024	5.552	0.038	5.482	0.035	0.243	0.012
73	Lys	1.638	0.039	4.817	0.052	4.805	0.027	0.216	0.010
74	Leu	1.668	0.031	5.015	0.055	5.033	0.009	0.250	0.033

Amino acid number	name	R_1 / sec^{-1}	SD (R_1)	R_2 / sec^{-1}	SD (R_2)	$R_{1\rho}$ / sec^{-1}	SD ($R_{1\rho}$)	hetNOE	SD (hetNOE)
75	His	1.465	0.043	4.119	0.068	4.120	0.027	0.099	0.011
76	Arg	1.587	0.084	4.492	0.115	4.478	0.026	0.218	0.001
77	Ser	1.506	0.097	3.798	0.126	3.867	0.038	0.186	0.003
78	Asp	1.531	0.087	3.530	0.084	3.544	0.031	0.123	0.002
79	Ser	1.447	0.097	3.161	0.094	3.145	0.028	0.044	0.008
80	Ser	1.447	0.096	2.964	0.089	3.056	0.048	0.018	0.001
81	Ser	1.413	0.072	2.762	0.110	2.775	0.034	-0.018	0.001
82	Ser	1.380	0.080	2.656	0.087	2.577	0.026	-0.076	0.008
83	Ser	1.422	0.076	2.755	0.107	2.729	0.023	-0.035	0.000
84	Ser	1.443	0.097	2.844	0.091	2.870	0.030	-0.017	0.001
85	Ser	1.435	0.087	2.874	0.107	2.894	0.028	-0.019	0.002
86	Glu	1.587	0.071	3.203	0.092	3.255	0.016	0.087	0.001
87	Glu	1.633	0.035	3.166	0.045	3.173	0.010	0.084	0.003
88	Glu	1.612	0.037	3.441	0.047	3.481	0.011	0.104	0.001
89	Gly	1.590	0.050	3.301	0.079	3.159	0.017	0.111	0.007
90	Ser	1.544	0.075	3.592	0.079	3.674	0.018	0.164	0.005
91	Asp	1.595	0.079	3.997	0.102	4.087	0.020	0.201	0.001
92	Gly	1.614	0.060	4.158	0.096	4.023	0.014	0.235	0.007
93	Glu	1.671	0.039	4.993	0.080	5.092	0.015	0.276	0.015
94	Lys	1.663	0.038	4.941	0.068	5.008	0.015	0.224	0.004
95	Arg	1.675	0.060	4.850	0.068	4.831	0.011	0.289	0.001
96	Lys	1.536	0.031	3.501	0.038	3.664	0.014	0.022	0.004
97	Lys	1.649	0.039	4.713	0.075	4.792	0.017	0.260	0.007
98	Lys	1.609	0.049	4.472	0.054	4.401	0.029	0.178	0.006
99	Lys	1.601	0.051	3.146	0.069	3.303	0.012	0.109	0.002
100	Glu	1.521	0.036	4.369	0.055	4.390	0.015	0.073	0.002
101	Lys	1.530	0.046	4.348	0.055	4.409	0.013	0.109	0.003
102	Lys	1.582	0.053	4.352	0.116	4.460	0.021	0.143	0.004
103	Lys	1.532	0.034	4.621	0.084	4.706	0.021	0.087	0.003
104	Pro								
105	Thr	1.480	0.058	4.502	0.085	4.643	0.010	0.080	0.003
106	Thr	1.470	0.062	4.572	0.116	4.733	0.014	0.092	0.002
107	Glu	1.534	0.038	4.388	0.091	4.496	0.019	0.082	0.001
108	Val	1.502	0.019	5.003	0.040	5.013	0.011	0.119	0.002
109	Glu	1.513	0.012	4.998	0.075	5.097	0.020	0.029	0.012
110	Val	1.497	0.010	4.780	0.043	4.869	0.027	0.063	0.006
111	Lys	1.538	0.010	5.841	0.073	5.903	0.016	0.073	0.006
112	Glu	1.549	0.022	5.444	0.060	5.507	0.019	0.175	0.004
113	Glu	1.547	0.028	5.705	0.065	5.631	0.024	0.163	0.005
114	Glu	1.581	0.031	4.492	0.075	4.484	0.036	0.149	0.010
115	Lys	1.606	0.032	5.214	0.066	5.280	0.020	0.189	0.004

Amino acid number	name	R_1 / sec^{-1}	SD (R_1)	R_2 / sec^{-1}	SD (R_2)	$R_{1\rho}$ / sec^{-1}	SD ($R_{1\rho}$)	hetNOE	SD (hetNOE)
116	Lys	1.558	0.038	5.453	0.072	5.464	0.015	0.197	0.008
117	Gly	1.545	0.056	5.280	0.129	5.203	0.022	0.214	0.002
118	Phe	1.564	0.047	5.663	0.079	5.718	0.017	0.269	0.005
119	Met	1.572	0.039	5.688	0.092	5.643	0.020	0.205	0.008
120	Glu	1.573	0.029	4.933	0.079	4.955	0.031	0.159	0.007
121	Lys	1.593	0.024	4.885	0.079	4.990	0.021	0.166	0.008
122	Leu	1.621	0.020	5.580	0.058	5.577	0.011	0.209	0.008
123	Lys	1.597	0.029	4.985	0.062	5.008	0.031	0.215	0.009
124	Glu	1.579	0.029	5.058	0.056	4.990	0.007	0.175	0.004
125	Lys	1.611	0.024	4.669	0.048	4.792	0.019	0.157	0.014
126	Leu	1.573	0.016	4.235	0.063	4.263	0.017	0.110	0.011
127	Pro								
128	Gly	1.563	0.056	4.072	0.085	4.156	0.043	0.174	0.004
129	His	1.458	0.084	3.755	0.134	3.717	0.061	0.171	0.007
130	Lys	1.565	0.055	3.935	0.112	4.042	0.024	0.140	0.007
131	Lys	1.574	0.035	3.666	0.062	3.716	0.009	0.050	0.010
132	Pro								
133	Glu	1.605	0.038	3.510	0.059	3.509	0.010	0.076	0.008
134	Asp	1.501	0.053	3.415	0.049	3.598	0.017	0.023	0.005
135	Gly	1.470	0.071	2.889	0.080	2.898	0.013	-0.034	0.006
136	Ser	1.479	0.072	3.208	0.096	3.205	0.014	0.055	0.002
137	Ala	1.432	0.072	2.591	0.091	2.731	0.020	-0.101	0.004
138	Val	1.317	0.024	2.618	0.031	2.558	0.028	-0.290	0.000
139	Ala	1.345	0.033	2.820	0.047	2.804	0.013	-0.329	0.001
140	Ala	1.255	0.030	2.957	0.043	2.996	0.018	-0.317	0.000
141	Ala	1.219	0.018	3.782	0.049	3.814	0.018	-0.308	-0.001
142	Pro								
143	Val	1.414	0.015	4.016	0.016	4.143	0.013	-0.153	-0.001
144	Val	1.399	0.003	3.795	0.040	3.788	0.015	-0.146	0.001
145	Val	1.403	0.006	4.888	0.063	4.615	0.036	-0.100	0.001
146	Pro								
147	Pro								
148	Pro								
149	Val	1.423	0.002	4.382	0.019	4.409	0.017	-0.009	0.006
150	Glu	1.469	0.014	4.149	0.034	4.129	0.011	-0.026	0.007
151	Glu	1.490	0.030	4.374	0.060	4.407	0.015	0.040	0.001
152	Ala	1.493	0.030	3.720	0.059	3.685	0.011	0.004	0.003
153	His	1.358	0.060	3.954	0.098	4.036	0.020	-0.084	-0.002
154	Pro								
155	Val	1.472	0.016	3.870	0.030	3.814	0.031	-0.026	0.008
156	Glu	1.545	0.026	3.954	0.058	4.049	0.015	0.014	0.003

Amino acid number	name	R_1 / sec^{-1}	SD (R_1)	R_2 / sec^{-1}	SD (R_2)	$R_{1\rho}$ / sec^{-1}	SD ($R_{1\rho}$)	hetNOE	SD (hetNOE)
157	Lys	1.538	0.047	4.323	0.096	4.423	0.012	0.093	0.001
158	Lys	1.542	0.046	4.003	0.065	3.833	0.011	0.088	0.003
159	Gly	1.521	0.055	3.740	0.107	3.561	0.017	0.079	0.001
160	Ile	1.532	0.025	4.268	0.036	4.241	0.014	0.152	0.005
161	Leu	1.612	0.020	4.888	0.059	4.900	0.012	0.128	0.013
162	Glu	1.575	0.026	4.619	0.043	4.666	0.009	0.134	0.012
163	Lys	1.605	0.032	4.907	0.065	4.836	0.012	0.162	0.010
164	Ile	1.644	0.025	4.359	0.038	4.535	0.022	0.206	0.011
165	Lys	1.613	0.021	4.798	0.059	4.900	0.012	0.130	0.013
166	Glu	1.463	0.017	4.448	0.042	4.572	0.024	0.050	0.007
167	Lys	1.598	0.030	4.421	0.079	4.638	0.026	0.174	0.003
168	Leu	1.585	0.016	4.274	0.025	4.297	0.009	0.160	0.021
169	Pro								
170	Gly	1.492	0.075	3.925	0.101	3.823	0.015	0.142	0.016
171	Tyr	1.560	0.056	4.314	0.087	4.294	0.012	0.270	0.023
172	His	1.554	0.068	3.997	0.113	4.068	0.031	0.166	0.009
173	Pro								
174	Lys	1.577	0.091	4.085	0.103	4.141	0.021	0.174	0.011
175	Thr	1.530	0.073	3.678	0.104	3.663	0.015	0.124	0.010
176	Thr	1.540	0.091	3.524	0.089	3.592	0.016	0.110	0.008
177	Val	1.577	0.029	3.514	0.034	3.587	0.014	0.070	0.006
178	Glu	1.637	0.022	3.401	0.047	3.461	0.012	0.067	0.005
179	Glu	1.623	0.030	3.256	0.035	3.237	0.015	0.085	0.002
180	Glu	1.603	0.027	3.398	0.040	3.388	0.007	0.074	0.001
181	Lys	1.611	0.049	2.833	0.048	2.920	0.010	0.029	0.003
182	Lys	1.601	0.042	2.894	0.044	2.999	0.017	0.028	0.004
183	Asp	1.539	0.048	2.279	0.047	2.417	0.007	-0.130	0.006
184	Lys	1.422	0.035	2.073	0.032	2.050	0.013	-0.320	0.001
185	Glu	1.224	0.008	1.468	0.010	1.510	0.009	-0.762	0.010

Data deposition

The relaxation parameters of ERD14 presented herein have also been deposited in the Biological Magnetic Resonance Data Bank (BMRB; http://www.bmrb.wisc.edu) under the BMRB accession number 16876.

6.3. List of disappearing peaks (in-cell NMR)

Results of the in-cell NMR spectrum in dependence of the sequence of ERD14 are presented below. Raw results (i.e. observed changes compared to the reference) from the in-cell NMR spectrum and from the dextran control spectrum are listed as well as the final, corrected results in words and numbers (disappearance score, for details see section 4.1.4). For comparison, conserved segments and motionally restricted regions are indicated as well.

Amino acid number	name	in-cell NMR: raw data	Control experiment: dextrane	in-cell NMR: final result	Final dis-apperance score:	Conserved segments	Regions of Restricted Motion
1	**Met**			unknown	0		
2	**Ala**			unknown	-0.25		
3	**Glu**			unknown	-0.4		
4	**Glu**	steady		steady	-0.4		
5	**Ile**	steady		steady	-0.6		
6	**Lys**	shifted		unknown	-0.6		
7	**Asn**	steady		steady	-0.4		
8	**Val**			unknown	-0.4		
9	**Pro**			unknown	-0.4		
10	**Glu**	steady		steady	-0.2		
11	**Gln**			unknown	-0.4		
12	**Glu**			unknown	-0.4		
13	**Val**	steady		steady	-0.2		
14	**Pro**			unknown	-0.2		
15	**Lys**			unknown	-0.4		
16	**Val**			unknown	-0.2		
17	**Ala**	steady		steady	-0.2		
18	**Thr**	shifted		unknown	-0.2		
19	**Glu**			unknown	-0.4		
20	**Glu**			unknown	-0.2		
21	**Ser**	steady		steady	-0.2		
22	**Ser**	shifted		unknown	-0.2		
23	**Ala**			unknown	-0.2		
24	**Glu**	intensity loss		unknown	0	K	
25	**Val**			unknown	0	K	
26	**Thr**	shifted		unknown	0	K	
27	**Asp**			unknown	0	K	
28	**Arg**			unknown	0	K	region 1
29	**Gly**	disappearing	disappearing	unknown	0.2	K	region 1
30	**Leu**	disappearing	disappearing	unknown	0.2	K	region 1
31	**Phe**	disappearing		disappearing	0.4	K	region 1
32	**Asp**			unknown	0.4	K	region 1

Amino acid number	name	in-cell NMR: raw data	Control experiment: dextrane	in-cell NMR: final result	Final dis-apperance score:	Conserved segments	Regions of Restricted Motion
33	**Phe**	disappearing		disappearing	0.6	K	region 1
34	**Leu**			unknown	0.6	K	region 1
35	**Gly**	disappearing		disappearing	0.6	K	region 1
36	**Lys**	disappearing		disappearing	0.4	K	region 1
37	**Lys**			unknown	0.4	K	region 1
38	**Lys**			unknown	0.2	K	
39	**Asp**			unknown	-0.2	K	
40	**Glu**			unknown	-0.2	K	
41	**Thr**	steady		steady	-0.2	K	
42	**Lys**			unknown	-0.4	K	
43	**Pro**			unknown	-0.4	K	
44	**Glu**	steady		steady	-0.2		
45	**Glu**			unknown	-0.2		
46	**Thr**	shifted		unknown	-0.2		
47	**Pro**			unknown	0		
48	**Ile**			unknown	-0.2		
49	**Ala**	shifted		unknown	-0.2		
50	**Ser**	steady		steady	-0.2		
51	**Glu**			unknown	-0.2		
52	**Phe**	shifted		unknown	-0.2		
53	**Glu**			unknown	0		
54	**Gln**			unknown	0		
55	**Lys**			unknown	0.2		
56	**Val**			unknown	0.2		
57	**His**	disappearing		disappearing	0.2		
58	**Ile**	disappearing	disappearing	unknown	0		
59	**Ser**	intensity loss		unknown	0		
60	**Glu**	steady		steady	-0.2		
61	**Pro**			unknown	-0.2		
62	**Glu**			unknown	-0.4		
63	**Pro**			unknown	-0.2		
64	**Glu**	steady		steady	-0.2		
65	**Val**	intensity loss		unknown	-0.2		
66	**Lys**	disappearing	disappearing	unknown	-0.2		
67	**His**			unknown	0		
68	**Glu**			unknown	0		region 2
69	**Ser**	disappearing	disappearing	unknown	0.2		region 2
70	**Leu**	intensity loss		unknown	0.2		region 2
71	**Leu**	disappearing		disappearing	0.2		region 2
72	**Glu**	intensity loss		unknown	0.2		region 2
73	**Lys**			unknown	0.2		region 2
74	**Leu**			unknown	0	S	region 2

Amino acid number	name	in-cell NMR: raw data	Control experiment: dextrane	in-cell NMR: final result	Final dis-apperance score:	Conserved segments	Regions of Restricted Motion
75	**His**	intensity loss		unknown	0	S	region 2
76	**Arg**			unknown	0	S	
77	**Ser**	shifted		unknown	0	S	
78	**Asp**			unknown	0	S	
79	**Ser**	shifted		unknown	0	S	
80	**Ser**			unknown	0	S	
81	**Ser**			unknown	0	S	
82	**Ser**			unknown	0	S	
83	**Ser**			unknown	0	S	
84	**Ser**			unknown	0	S	
85	**Ser**			unknown	0	S	
86	**Glu**			unknown	0		
87	**Glu**			unknown	-0.2	ChP	
88	**Glu**			unknown	-0.4	ChP	
89	**Gly**	steady		steady	-0.6	ChP	
90	**Ser**	steady		steady	-0.8	ChP	
91	**Asp**	steady		steady	-0.8	ChP	
92	**Gly**	steady		steady	-0.6	ChP	region 3
93	**Glu**			unknown	-0.4	ChP	region 3
94	**Lys**			unknown	-0.2	ChP	region 3
95	**Arg**			unknown	0	ChP	region 3
96	**Lys**			unknown	0.2	ChP	region 3
97	**Lys**			unknown	0.2	ChP	
98	**Lys**	disappearing		disappearing	0.2	ChP	
99	**Lys**			unknown	0.2	ChP	
100	**Glu**			unknown	0.2	ChP	
101	**Lys**			unknown	0	ChP	
102	**Lys**			unknown	0	ChP	
103	**Lys**	intensity loss		unknown	0	ChP	
104	**Pro**			unknown	-0.2		
105	**Thr**	intensity loss		unknown	-0.2		
106	**Thr**	steady		steady	-0.2		
107	**Glu**			unknown	-0.2		
108	**Val**			unknown	-0.2		
109	**Glu**	disappearing	disappearing	unknown	0		region 4
110	**Val**	intensity loss		unknown	0		region 4
111	**Lys**	disappearing	disappearing	unknown	0		region 4
112	**Glu**			unknown	0		region 4
113	**Glu**	intensity loss	disappearing	unknown	0	K	region 4
114	**Glu**			unknown	0	K	region 4
115	**Lys**			unknown	0	K	region 4
116	**Lys**			unknown	0.2	K	region 4

Amino acid number	name	in-cell NMR: raw data	Control experiment: dextrane	in-cell NMR: final result	Final dis-apperance score:	Conserved segments	Regions of Restricted Motion
117	**Gly**	disappearing	disappearing	unknown	0.2	K	region 4
118	**Phe**	disappearing		disappearing	0.2	K	region 4
119	**Met**			unknown	0.2	K	region 4
120	**Glu**			unknown	0.2	K	region 4
121	**Lys**			unknown	0	K	region 4
122	**Leu**	disappearing	disappearing	unknown	0	K	region 4
123	**Lys**			unknown	0	K	region 4
124	**Glu**			unknown	0.2	K	region 4
125	**Lys**			unknown	0.2	K	
126	**Leu**	disappearing		disappearing	0.4	K	
127	**Pro**			unknown	0.4	K	
128	**Gly**	disappearing		disappearing	0.4	K	
129	**His**	disappearing	disappearing	unknown	0.2	K	
130	**Lys**			unknown	0.2	K	
131	**Lys**	intensity loss		unknown	-0.2	K	
132	**Pro**			unknown	-0.2	K	
133	**Glu**	steady		steady	-0.4	K	
134	**Asp**			unknown	-0.6	K	
135	**Gly**	steady		steady	-0.6		
136	**Ser**	steady		steady	-0.4		
137	**Ala**			unknown	-0.4		
138	**Val**	shifted		unknown	-0.2		
139	**Ala**	shifted		unknown	0.2		
140	**Ala**	intensity loss		unknown	0.2		
141	**Ala**	disappearing		disappearing	0.2		
142	**Pro**			unknown	0.2		
143	**Val**			unknown	0		
144	**Val**	intensity loss		unknown	-0.2		
145	**Val**	steady	disappearing	steady	-0.2		
146	**Pro**			unknown	-0.2		
147	**Pro**			unknown	-0.2		
148	**Pro**			unknown	0		
149	**Val**			unknown	0		
150	**Glu**	intensity loss		unknown	0		
151	**Glu**			unknown	0.2		
152	**Ala**			unknown	0.2		
153	**His**	disappearing		disappearing	0.2		
154	**Pro**			unknown	0.2		
155	**Val**			unknown	0.2		
156	**Glu**	intensity loss		unknown	0	K	
157	**Lys**			unknown	-0.2	K	
158	**Lys**			unknown	0	K	

Amino acid number	name	in-cell NMR: raw data	Control experiment: dextrane	in-cell NMR: final result	Final dis-apperance score:	Conserved segments	Regions of Restricted Motion
159	**Gly**	steady		steady	0	K	
160	**Ile**	disappearing		disappearing	0	K	
161	**Leu**			unknown	0	K	
162	**Glu**			unknown	0.2	K	
163	**Lys**			unknown	0	K	region 5
164	**Ile**			unknown	0	K	region 5
165	**Lys**			unknown	0	K	region 5
166	**Glu**			unknown	0.2	K	
167	**Lys**			unknown	0.2	K	
168	**Leu**	disappearing		disappearing	0.4	K	
169	**Pro**			unknown	0.6	K	
170	**Gly**	disappearing		disappearing	0.6	K	
171	**Tyr**	disappearing		disappearing	0.4		
172	**His**	disappearing	disappearing	unknown	0.4		
173	**Pro**			unknown	0.2		
174	**Lys**	shifted		unknown	0		
175	**Thr**			unknown	0		
176	**Thr**			unknown	0		
177	**Val**	intensity loss		unknown	0		
178	**Glu**	intensity loss		unknown	0		
179	**Glu**			unknown	0		
180	**Glu**			unknown	0		
181	**Lys**			unknown	0		
182	**Lys**			unknown	0		
183	**Asp**			unknown	0		
184	**Lys**			unknown	0		
185	**Glu**	shifted		unknown	0		

6.4. Comparison of the chemical shifts of K-segments

The table below compares the H and N chemical shifts of the conserved K-segments within the two dehydrins assigned so far (Findlater and Graether, 2009; Szalainé Ágoston et al., 2011). Non-conserved residues are grayed out. It should be taken into account that non-conserved residues also influence the chemical shift of their conserved neighbors. Segment Ka of ERD14 is not listed as its sequence similarity to the classical K-segment is too low to be comparable in this way.

DHN K1	H	N	**DHN K2**	H	N	**ERD14 Kb**	H	N	**ERD14 Kc**	H	N
Q	8.478	121.429	R	8.323	121.460	E	8.589	125.462	E	8.384	121.907
Q	8.422	121.758	K	8.295	121.431	K	8.515	123.816	K	8.279	122.467
K	8.387	122.791	K	8.297	122.182	K	8.487	123.106	K	8.36	122.343
G	8.489	109.911	G	8.539	109.537	G	8.57	110.327	G	8.502	109.808
M	8.183	119.713	M	8.456	120.863	I	8.056	120.257	F	8.128	120.503
M	8.396	120.956	K	8.503	123.057	L	8.363	125.268	M	8.284	121.397
E	8.344	121.961	E	8.345	121.335	E	8.321	121.934	E	8.289	121.888
K	8.265	122.300	K	8.267	121.951	K	8.289	122.337	K	8.249	121.853
I	8.171	122.897	I	8.088	122.061	I	8.162	122.268	L	8.138	122.752
K	8.402	126.032	K	8.311	125.124	K	8.363	125.317	K	8.207	121.676
E	8.360	122.707	E	8.402	121.205	E	8.319	121.609	E	8.243	121.490
K	8.300	122.668	R	8.391	122.802	K	8.259	122.031	K	8.314	122.402
L	8.283	125.258	I	8.209	124.029	L	8.267	124.934	L	8.344	125.071
P			P			P		136.354	P		136.240
G	8.486	109.986	G	8.509	109.603	G	8.608	109.598	G	8.558	109.324

7. List of References

Alipanahi, B., Gao, X., Karakoc, E., Li, S.C., Balbach, F., Feng, G., Donaldson, L., and Li, M. (2011). Error tolerant NMR backbone resonance assignment and automated structure generation. J Bioinform Comput Biol **9**: 15–41.

Alsheikh, M.K., Heyen, B.J., and Randall, S.K. (2003). Ion binding properties of the dehydrin ERD14 are dependent upon phosphorylation. J. Biol. Chem. **278**: 40882–40889.

Alsheikh, M.K., Svensson, J.T., and Randall, S.K. (2005). Phosphorylation regulated ion-binding is a property shared by the acidic subclass dehydrins. Plant Cell and Environment **28**: 1114–1122.

Asghar, R., Fenton, R., DeMason, D., and Close, T. (1994). Nuclear and cytoplasmic localization of maize embryo and aleurone dehydrin. Protoplasma **177**: 87–94.

Aswad, D.W., Paranandi, M.V., and Schurter, B.T. (2000). Isoaspartate in peptides and proteins: formation, significance, and analysis. J. Pharm. Biomed. Anal. **21**: 1129–1136.

Bayer, M.E. (1967). Response of Cell Walls of Escherichia coli to a Sudden Reduction of the Environmental Osmotic Pressure. J. Bacteriol. **93**: 1104–1112.

Bekei, B., Rose, H.M., Herzig, M., Dose, A., Schwarzer, D., and Selenko, P. (2012). In-cell NMR in mammalian cells: part 1. Methods Mol. Biol. **895**: 43–54.

Bekei, B., Rose, H.M., Herzig, M., and Selenko, P. (2012). In-cell NMR in mammalian cells: part 2. Methods Mol. Biol. **895**: 55–66.

Belle, A., Tanay, A., Bitincka, L., Shamir, R., and O'Shea, E.K. (2006). Quantification of protein half-lives in the budding yeast proteome. PNAS **103**: 13004–13009.

Benschop, J.J., Mohammed, S., O'Flaherty, M., Heck, A.J.R., Slijper, M., and Menke, F.L.H. (2007). Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis. Mol. Cell Proteomics **6**: 1198–1214.

Berjanskii, M.V. and Wishart, D.S. (2007). The RCI server: rapid and accurate calculation of protein flexibility using chemical shifts. Nucleic Acids Res **35**: W531–W537.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucl. Acids Res. **28**: 235–242.

Bermel, W., Bertini, I., Felli, I.C., Gonnelli, L., Koźmiński, W., Piai, A., Pierattelli, R., and Stanek, J. (2012). Speeding up sequence specific assignment of IDPs. J. Biomol. NMR **53**: 293–301.

Bermel, W., Bertini, I., Felli, I.C., Lee, Y.-M., Luchinat, C., and Pierattelli, R. (2006). Protonless NMR Experiments for Sequence-Specific Assignment of Backbone Nuclei in Unfolded Proteins. J. Am. Chem. Soc. **128**: 3918–3919.

Bertoncini, C.W., Rasia, R.M., Lamberto, G.R., Binolfi, A., Zweckstetter, M., Griesinger, C., and Fernandez, C.O. (2007). Structural characterization of the intrinsically unfolded protein beta-synuclein, a natural negative regulator of alpha-synuclein aggregation. J. Mol. Biol. **372**: 708–722.

Bhattacharyya, J. and Das, K.P. (1999). Molecular chaperone-like properties of an unfolded protein, alpha(s)-casein. J. Biol. Chem. **274**: 15505–15509.

Bodart, J.-F., Wieruszeski, J.-M., Amniai, L., Leroy, A., Landrieu, I., Rousseau-Lescuyer, A., Vilain, J.-P., and Lippens, G. (2008). NMR observation of Tau in Xenopus oocytes. J. Magn. Reson. **192**: 252–257.

Bokor, M., Csizmók, V., Kovács, D., Banki, P., Friedrich, P., Tompa, P., and Tompa, K. (2005). NMR relaxation studies on the hydrate layer of intrinsically unstructured proteins. Biophys J **88**: 2030–7.

Braun, D., Wider, G., and Wüthrich, K. (1994). Sequence-corrected 15N " random coil" chemical shifts. J. Am. Chem. Soc. **116**: 8466–8469.

Bravo, L.A., Gallardo, J., Navarrete, A., Olave, N., Martínez, J., Alberdi, M., Close, T.J., and Corcuera, L.J. (2003). Cryoprotective activity of a cold-induced dehydrin purified from barley. Physiologia Plantarum **118**: 262–269.

Bray, E.A. (1994). Alterations in gene expression in response to water deficit. In Stress-Induced Gene Expression in Plants, A.S. Basra, ed (Harwood Academic Publishers: Newark, NJ), pp. 1–23.

Bray, E.A. (1993). Molecular Responses to Water Deficit. Plant Physiol **103**: 1035–1040.

Brini, F., Hanin, M., Lumbreras, V., Amara, I., Khoudi, H., Hassairi, A., Pagès, M., and Masmoudi, K. (2007). Overexpression of wheat dehydrin DHN-5 enhances tolerance to salt and osmotic stress in Arabidopsis thaliana. Plant Cell Rep **26**: 2017–2026.

Brini, F., Saibi, W., Amara, I., Gargouri, A., Masmoudi, K., and Hanin, M. (2010). Wheat dehydrin DHN-5 exerts a heat-protective effect on beta-glucosidase and glucose oxidase activities. Biosci. Biotechnol. Biochem. **74**: 1050–1054.

Bryant, J.E., Lecomte, J.T.J., Lee, A.L., Young, G.B., and Pielak, G.J. (2006). Cytosol has a small effect on protein backbone dynamics. Biochemistry **45**: 10085–10091.

Bryant, J.E., Lecomte, J.T.J., Lee, A.L., Young, G.B., and Pielak, G.J. (2005). Protein dynamics in living cells. Biochemistry **44**: 9275–9279.

Bundi, A. and Wüthrich, K. (1979). Use of amide 1H-nmr titration shifts for studies of polypeptide conformation. Biopolymers **18**: 299–311.

Burz, D.S., Dutta, K., Cowburn, D., and Shekhtman, A. (2006a). In-cell NMR for protein-protein interactions (STINT-NMR). Nat Protoc **1**: 146–152.

Burz, D.S., Dutta, K., Cowburn, D., and Shekhtman, A. (2006b). Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR). Nature Methods **3**: 91–93.

Camilloni, C., De Simone, A., Vranken, W.F., and Vendruscolo, M. (2012). Determination of Secondary Structure Populations in Disordered States of Proteins Using Nuclear Magnetic Resonance Chemical Shifts. Biochemistry **51**: 2224–2231.

Campbell, S.A. and Close, T.J. (1997). Dehydrins: Genes, Proteins, and Associations with Phenotypic Traits. New Phytologist **137**: 61–74.

Campos, F., Zamudio, F., and Covarrubias, A.A. (2006). Two different late embryogenesis abundant proteins from Arabidopsis thaliana contain specific domains that inhibit Escherichia coli growth. Biochem. Biophys. Res. Commun. **342**: 406–413.

Chakrabortee, S., Boschetti, C., Walton, L.J., Sarkar, S., Rubinsztein, D.C., and Tunnacliffe, A. (2007). Hydrophilic protein associated with desiccation tolerance exhibits broad protein stabilization function. Proc. Natl. Acad. Sci. U. S. A. **104**: 18073–18078.

Chatterjee, A., Kumar, A., Chugh, J., Srivastava, S., Bhavesh, N., and Hosur, R. (2005). NMR of unfolded proteins. J. Chem. Sci. **117**: 3–21.

Cheng, Y., LeGall, T., Oldfield, C.J., Mueller, J.P., Van, Y.-Y.J., Romero, P., Cortese, M.S., Uversky, V.N., and Dunker, A.K. (2006). Rational drug design via intrinsically disordered protein. Trends Biotechnol. **24**: 435–442.

Cheng, Z., Targolli, J., Huang, X., and Wu, R. (2002). Wheat LEA genes, PMA80 and PMA1959, enhance dehydration tolerance of transgenic rice (Oryza sativa L.). Molecular Breeding **10**: 71–82.

Cino, E.A., Karttunen, M., and Choy, W.-Y. (2012). Effects of molecular crowding on the dynamics of intrinsically disordered proteins. PLoS ONE **7**: e49876.

Clerici, M., Mourão, A., Gutsche, I., Gehring, N.H., Hentze, M.W., Kulozik, A., Kadlec, J., Sattler, M., and Cusack, S. (2009). Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. EMBO J. **28**: 2293–2306.

Close, T.J. (1996). Dehydrins: emergence of a biochemical role of a family of plant dehydration proteins. Physiol. Plant. **97**: 795–803.

Close, T.J., Kortt, A.A., and Chandler, P.M. (1989). A cDNA-based comparison of dehydration-induced proteins (dehydrins) in barley and corn. Plant Mol. Biol. **13**: 95–108.

Cokol, M., Nair, R., and Rost, B. (2000). Finding nuclear localization signals. EMBO Rep. **1**: 411–415.

Croke, R.L., Sallum, C.O., Watson, E., Watt, E.D., and Alexandrescu, A.T. (2008). Hydrogen exchange of monomeric α-synuclein shows unfolded structure persists at physiological temperature and is independent of molecular crowding in Escherichia coli. Protein Sci **17**: 1434–1445.

Cuming, A.C. (1999). LEA proteins. In Seed proteins, P.R. Shewry and R. Casey, eds (Kluwer Academic Pub), pp. 753–780.

Csizmók, V., Szőllősi, E., Friedrich, P., and Tompa, P. (2006). A Novel Two-dimensional Electrophoresis Technique for the Identification of Intrinsically Unstructured Proteins. Molecular & Cellular Proteomics **5**: 265 –273.

Dalgarno, D.C., Levine, B.A., and Williams, R.J. (1983). Structural information from NMR secondary chemical shifts of peptide alpha C-H protons in proteins. Biosci. Rep. **3**: 443–452.

Danyluk, J., Perron, A., Houde, M., Limin, A., Fowler, B., Benhamou, N., and Sarhan, F. (1998). Accumulation of an acidic dehydrin in the vicinity of the plasma membrane during cold acclimation of wheat. Plant Cell **10**: 623–638.

Dosztányi, Z., Csizmók, V., Tompa, P., and Simon, I. (2005a). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics **21**: 3433–3434.

Dosztányi, Z., Csizmók, V., Tompa, P., and Simon, I. (2005b). The Pairwise Energy Content Estimated from Amino Acid Composition Discriminates between Folded and Intrinsically Unstructured Proteins. J. Mol. Biol. **347**: 827–839.

Dumont, F., Marechal, P.-A., and Gervais, P. (2004). Cell Size and Water Permeability as Determining Factors for Cell Viability after Freezing at Different Cooling Rates. Appl. Environ. Microbiol. **70**: 268–272.

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C., and Obradovic, Z. (2001). Intrinsically disordered protein. J. Mol. Graph. Model. **19**: 26–59.

Dunker, A.K. and Obradovic, Z. (2001). The protein trinity—linking function and disorder. Nature Biotechnology **19**: 805–806.

Dure, L. (1993). A repeating 11-mer amino acid motif and plant desiccation. Plant J **3**: 363–369.

Dure, L., Greenway, S.C., and Galau, G.A. (1981). Developmental biochemistry of cottonseed embryogenesis and germination: changing messenger ribonucleic acid populations as shown by in vitro and in vivo protein synthesis. Biochemistry **20**: 4162–4168.

Dure, L.I., Crouch, M., Harada, J., Ho, T.-H.D., Mundy, J., Quatrano, R., Thomas, T., and Sung, Z.R. (1989). Common amino acid sequence domains among the LEA proteins of higher plants. Plant Mol Biol **12**: 475–486.

Dyson, H.J. and Wright, P.E. (2004). Unfolded proteins and protein folding studied by NMR. Chem. Rev. **104**: 3607–3622.

Egerton-Warburton, L., Balsamo, R., and Close, T. (1997). Temporal accumulation and ultrastructural localization of dehydrins in Zea mays L. Physiol. Plant. **101**: 545–555.

Findlater, E.E. and Graether, S.P. (2009). NMR assignments of the intrinsically disordered K2 and YSK2 dehydrins. Biomol. NMR Assignments **3**: 273–275.

Fiorito, F., Hiller, S., Wider, G., and Wüthrich, K. (2006). Automated Resonance Assignment of Proteins: 6 DAPSY-NMR. J Biomol NMR **35**: 27–37.

Fuxreiter, M., Simon, I., Friedrich, P., and Tompa, P. (2004). Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J. Mol. Biol **338**: 1015–1026.

Garay-Arroyo, A., Colmenero-Flores, J.M., Garciarrubio, A., and Covarrubias, A.A. (2000). Highly Hydrophilic Proteins in Prokaryotes and Eukaryotes Are Common during Conditions of Water Deficit. Journal of Biological Chemistry **275**: 5668–5674.

Goddard, T.D. and Kneller, D.G. SPARKY 3.

Gorovits, B.M. and Horowitz, P.M. (1995). The molecular chaperonin cpn60 displays local flexibility that is reduced after binding with an unfolded protein. J. Biol. Chem. **270**: 13057–13062.

Gossert, A.D., Hiller, S., and Fernández, C. (2011). Automated NMR Resonance Assignment of Large Proteins for Protein−Ligand Interaction Studies. J. Am. Chem. Soc. **133**: 210–213.

Goyal, K., Walton, L.J., and Tunnacliffe, A. (2005). LEA proteins prevent protein aggregation due to water stress. Biochem. J. **388**: 151–157.

Hanin, M., Brini, F., Ebel, C., Toda, Y., Takeda, S., and Masmoudi, K. (2011). Plant dehydrins and stress tolerance. Plant Signal Behav **6**: 1503–1509.

Hara, M., Fujinaga, M., and Kuboi, T. (2005). Metal binding by citrus dehydrin with histidine-rich domains. J. Exp. Bot. **56**: 2695–2703.

Hara, M., Fujinaga, M., and Kuboi, T. (2004). Radical scavenging activity and oxidative modification of citrus dehydrin. Plant Physiol. Biochem. **42**: 657–662.

Hara, M., Shinoda, Y., Tanaka, Y., and Kuboi, T. (2009). DNA binding of citrus dehydrin promoted by zinc ion. Plant, Cell Environ. **32**: 532–541.

Hara, M., Terashima, S., Fukaya, T., and Kuboi, T. (2003). Enhancement of cold tolerance and inhibition of lipid peroxidation by citrus dehydrin in transgenic tobacco. Planta **217**: 290–298.

Hara, M., Terashima, S., and Kuboi, T. (2001). Characterization and cryoprotective activity of cold-responsive dehydrin from Citrus unshiu. Journal of Plant Physiology **158**: 1333–1339.

Honjoh, K.I., Oda, Y., Takata, R., Miyamoto, T., and Hatano, S. (1999). Introduction of the hiC6 Gene, which Encodes a Homologue of a Late Embryogenesis Abundant (LEA) Protein, Enhances Freezing Tolerance of Yeast. Journal of Plant Physiology **155**: 509–512.

Huang, Y. and Liu, Z. (2009). Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "fly-casting" mechanism. J. Mol. Biol. **393**: 1143–1159.

Hundertmark, M. and Hincha, D.K. (2008). LEA (Late Embryogenesis Abundant) proteins and their encoding genes in Arabidopsis thaliana. BMC Genomics **9**: 118.

Ikeya, T., Sasaki, A., Sakakibara, D., Shigemitsu, Y., Hamatsu, J., Hanashima, T., Mishima, M., Yoshimasu, M., Hayashi, N., Mikawa, T., Nietlispach, D., Wälchli, M., Smith, B.O., Shirakawa, M., Güntert, P., and Ito, Y. (2010). NMR protein structure determination in living E. coli cells using nonlinear sampling. Nat Protoc **5**: 1051–1060.

Imai, R., Chang, L., Ohta, A., Bray, E.A., and Takagi, M. (1996). A lea-class gene of tomato confers salt and freezing tolerance when expressed in Saccharomyces cerevisiae. Gene **170**: 243–248.

Ingram, J. and Bartels, D. (1996). The Molecular Basis of Dehydration Tolerance in Plants. Annu Rev Plant Physiol Plant Mol Biol **47**: 377–403.

Inomata, K., Ohno, A., Tochio, H., Isogai, S., Tenno, T., Nakase, I., Takeuchi, T., Futaki, S., Ito, Y., Hiroaki, H., and Shirakawa, M. (2009). High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. Nature **458**: 106–109.

Ismail, A.M., Hall, A.E., and Close, T.J. (1999). Allelic variation of a dehydrin gene cosegregates with chilling tolerance during seedling emergence. Proc Natl Acad Sci U S A **96**: 13566–13570.

Ito, J., Batth, T.S., Petzold, C.J., Redding-Johanson, A.M., Mukhopadhyay, A., Verboom, R., Meyer, E.H., Millar, A.H., and Heazlewood, J.L. (2011). Analysis of

the Arabidopsis cytosolic proteome highlights subcellular partitioning of central plant metabolism. J. Proteome Res. **10**: 1571–1582.

Ito, Y. and Selenko, P. (2010). Cellular structural biology. Curr. Opin. Struct. Biol **20**: 640–648.

Iturriaga, G., Schneider, K., Salamini, F., and Bartels, D. (1992). Expression of desiccation-related proteins from the resurrection plant Craterostigma plantagineum in transgenic tobacco. Plant Mol Biol **20**: 555–558.

Jensen, A.B., Goday, A., Figueras, M., Jessop, A.C., and Pagès, M. (1998). Phosphorylation mediates the nuclear targeting of the maize Rab17 protein. Plant J. **13**: 691–697.

Keller, R.L.J. (2004). The Computer Aided Resonance Assignment Tutorial (CANTINA Verlag: Goldau, Switzerland).

Kim, T.D., Paik, S.R., and Yang, C.H. (2002). Structural and functional implications of C-terminal regions of alpha-synuclein. Biochemistry **41**: 13782–90.

Kjaergaard, M., Brander, S., and Poulsen, F.M. (2011). Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. J Biomol NMR **49**: 139–149.

Kjaergaard, M. and Poulsen, F.M. (2011). Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. J. Biomol. NMR **50**: 157–165.

Kleffmann, T., Russenberger, D., Von Zychlinski, A., Christopher, W., Sjölander, K., Gruissem, W., and Baginsky, S. (2004). The Arabidopsis thaliana chloroplast proteome reveals pathway abundance and novel protein functions. Curr. Biol. **14**: 354–362.

Kline, K.G., Barrett-Wilt, G.A., and Sussman, M.R. (2010). In planta changes in protein phosphorylation induced by the plant hormone abscisic acid. PNAS **107**: 15986–15991.

Koag, M.C., Fenton, R.D., Wilkens, S., and Close, T.J. (2003). The binding of maize DHN1 to lipid vesicles. Gain of structure and lipid specificity. Plant Physiol. **131**: 309–316.

Kovács, D., Ágoston, B., and Tompa, P. (2008). Disordered plant LEA proteins as molecular chaperones. Plant Signaling Behav. **3**: 710–713.

Kovács, D., Kalmár, E., Torok, Z., and Tompa, P. (2008). Chaperone activity of ERD10 and ERD14, two disordered stress-related plant proteins. Plant Physiol **147**: 381–90.

Kovács, D. and Tompa, P. (2012). Diverse functional manifestations of intrinsic structural disorder in molecular chaperones. Biochem. Soc. Trans. **40**: 963–968.

Kruger, C., Berkowitz, O., Stephan, U.W., and Hell, R. (2002). A metal-binding member of the late embryogenesis abundant protein family transports iron in the phloem of Ricinus communis L. J. Biol. Chem. **277**: 25062–25069.

Lan, Y., Cai, D., and Zheng, Y.-Z. (2005). Expression in Escherichia coli of Three Different Soybean Late Embryogenesis Abundant (LEA) Genes to Investigate Enhanced Stress Tolerance. Journal of Integrative Plant Biology **47**: 613–621.

Li, C., Charlton, L.M., Lakkavaram, A., Seagle, C., Wang, G., Young, G.B., Macdonald, J.M., and Pielak, G.J. (2008). Differential dynamical effects of macromolecular crowding on an intrinsically disordered protein and a globular protein: implications for in-cell NMR spectroscopy. J. Am. Chem. Soc. **130**: 6310–6311.

Lindner, R.A., Kapur, A., Mariani, M., Titmuss, S.J., and Carver, J.A. (1998). Structural alterations of alpha-crystallin during its chaperone action. Eur. J. Biochem. **258**: 170–183.

Lippens, G., Amniai, L., Wieruszeski, J.-M., Sillen, A., Leroy, A., and Landrieu, I. (2012). Towards understanding the phosphorylation code of tau. Biochem. Soc. Trans. **40**: 698–703.

Lisse, T., Bartels, D., Kalbitzer, H.R., and Jaenicke, R. (1996). The recombinant dehydrin-like desiccation stress protein from the resurrection plant Craterostigma plantagineum displays no defined three-dimensional structure in its native state. Biol. Chem. **377**: 555–561.

Liu, Y. and Zheng, Y. (2005). PM2, a group 3 LEA protein from soybean, and its 22-mer repeating region confer salt tolerance in Escherichia coli. Biochem. Biophys. Res. Commun. **331**: 325–332.

Löhr, F., Pfeiffer, S., Lin, Y.-J., Hartleib, J., Klimmek, O., and Rüterjans, H. (2000). HNCAN pulse sequences for sequential backbone resonance assignment across proline residues in perdeuterated proteins. J. Biomol. NMR **18**: 337–346.

Löhr, F. and Rüterjans, H. (1998). Detection of Nitrogen-Nitrogen J-Couplings in Proteins. J. Magn. Reson. **132**: 130–137.

Lukin, J.A., Gove, A.P., Talukdar, S.N., and Ho, C. (1997). Automated probabilistic method for assigning backbone resonances of (13C,15N)-labeled proteins. Journal of Biomolecular NMR **9**: 151–166.

Lüttge, U., Beck, E., and Bartels, D. eds. (2011). Plant Desiccation Tolerance (Springer).

Mäntylahti, S., Hellman, M., and Permi, P. (2011). Extension of the HA-detection based approach: (HCA)CON(CA)H and (HCA)NCO(CA)H experiments for the main-chain assignment of intrinsically disordered proteins. J. Biomol. NMR **49**: 99–109.

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

Marsh, J.A., Singh, V.K., Jia, Z., and Forman-Kay, J.D. (2006). Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. Protein Sci **15**: 2795–2804.

McCubbin, W.D., Kay, C.M., and Lane, B.G. (1985). Hydrodynamic and optical properties of the wheat germ Em protein. Can. J. Biochem. Cell Biol. **63**: 803–811.

McNulty, B.C., Young, G.B., and Pielak, G.J. (2006). Macromolecular crowding in the Escherichia coli periplasm maintains alpha-synuclein disorder. J. Mol. Biol. **355**: 893–897.

Mehta, P.A., Rebala, K.C., Venkataraman, G., and Parida, A. (2009). A diurnally regulated dehydrin from Avicennia marina that shows nucleo-cytoplasmic localization and is phosphorylated by Casein kinase II in vitro. Plant Physiol. Biochem. **47**: 701–709.

Merutka, G., Jane Dyson, H., and Wright, P.E. (1995). "Random coil" 1H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. Journal of Biomolecular NMR **5**: 14–24.

Mille, Y., Beney, L., and Gervais, P. (2002). Viability of Escherichia coli after combined osmotic and thermal treatment: a plasma membrane implication. Biochimica et Biophysica Acta (BBA) - Biomembranes **1567**: 41–48.

Mohan, A., Oldfield, C.J., Radivojac, P., Vacic, V., Cortese, M.S., Dunker, A.K., and Uversky, V.N. (2006). Analysis of Molecular Recognition Features (MoRFs). J. Mol. Biol. **362**: 1043–1059.

Morris, G.J., Winters, L., Coulson, G.E., and Clarke, K.J. (1986). Effect of osmotic stress on the ultrastructure and viability of the yeast Saccharomyces cerevisiae. J. Gen. Microbiol. **132**: 2023–2034.

Mouillon, J.-M., Eriksson, S.K., and Harryson, P. (2008). Mimicking the plant cell interior under water stress by macromolecular crowding: disordered dehydrin proteins are highly resistant to structural collapse. Plant Physiol. **148**: 1925–1937.

Mouillon, J.M., Gustafsson, P., and Harryson, P. (2006). Structural investigation of disordered stress proteins. Comparison of full-length dehydrins with isolated peptides of their conserved segments. Plant Physiol **141**: 638–50.

Mulder, F.A.A., Lundqvist, M., and Scheek, R.M. (2010). Nuclear Magnetic Resonance Spectroscopy Applied to (Intrinsically) Disordered Proteins. In Instrumental Analysis of Intrinsically Disordered Proteins, V.N. Uversky and S. Longhi, eds (John Wiley & Sons, Inc.), pp. 59–87.

Nair, R. and Rost, B. (2005). Mimicking cellular sorting improves prediction of subcellular localization. J. Mol. Biol. **348**: 85–100.

Nikolovski, N., Rubtsov, D., Segura, M.P., Miles, G.P., Stevens, T.J., Dunkley, T.P.J., Munro, S., Lilley, K.S., and Dupree, P. (2012). Putative glycosyltransferases and other plant golgi apparatus proteins are revealed by LOPIT proteomics. Plant Physiol. **160**: 1037–1051.

Nylander, M., Svensson, J., Palva, E.T., and Welin, B.V. (2001). Stress-induced accumulation and tissue-specific localization of dehydrins in Arabidopsis thaliana. Plant Mol. Biol **45**: 263–279.

Ochoa-Alfaro, A.E., Rodríguez-Kessler, M., Pérez-Morales, M.B., Delgado-Sánchez, P., Cuevas-Velazquez, C.L., Gómez-Anduro, G., and Jiménez-Bremont, J.F. (2012). Functional characterization of an acidic SK(3) dehydrin isolated from an Opuntia streptacantha cDNA library. Planta **235**: 565–578.

Ogino, S., Kubo, S., Umemoto, R., Huang, S., Nishida, N., and Shimada, I. (2009). Observation of NMR signals from proteins introduced into living mammalian cells by reversible membrane permeabilization using a pore-forming toxin, streptolysin O. J. Am. Chem. Soc. **131**: 10834–10835.

Oldfield, C.J., Meng, J., Yang, J.Y., Yang, M.Q., Uversky, V.N., and Dunker, A.K. (2008). Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. BMC Genomics **9**: S1.

Onitsuka, M., Kamikubo, H., Yamazaki, Y., and Kataoka, M. (2008). Mechanism of induced folding: Both folding before binding and binding before folding can be realized in staphylococcal nuclease mutants. Proteins **72**: 837–847.

Palmer, A.G., 3rd (2004). NMR characterization of the dynamics of biomacromolecules. Chem. Rev. **104**: 3623–3640.

Park, S.M., Jung, H.Y., Kim, T.D., Park, J.H., Yang, C.-H., and Kim, J. (2002). Distinct roles of the N-terminal-binding domain and the C-terminal-solubilizing domain of alpha-synuclein, a molecular chaperone. J. Biol. Chem. **277**: 28512–28520.

Pelah, D. and Cohen, E. (2005). Cellular response of Chlorella zofingiensis to exogenous selenium. Plant Growth Regulation **45**: 225–232.

Peng, Y., Reyes, J.L., Wei, H., Yang, Y., Karlson, D., Covarrubias, A.A., Krebs, S.L., Fessehaie, A., and Arora, R. (2008). RcDhn5, a cold acclimation-responsive dehydrin from Rhododendron catawbiense rescues enzyme activity from dehydration effects in vitro and enhances freezing tolerance in RcDhn5-overexpressing Arabidopsis plants. Physiologia Plantarum **134**: 583–597.

Pielak, G.J. (2007). Retraction. Biochemistry **46**: 8206.

Plaxco, K.W., Morton, C.J., Grimshaw, S.B., Jones, J.A., Pitkeathly, M., Campbell, I.D., and Dobson, C.M. (1997). The effects of guanidine hydrochloride on the 'random coil' conformations and NMR chemical shifts of the peptide series GGXGG. Journal of Biomolecular NMR **10**: 221–230.

Puhakainen, T., Hess, M.W., Mäkelä, P., Svensson, J., Heino, P., and Palva, E.T. (2004). Overexpression of multiple dehydrin genes enhances tolerance to freezing stress in Arabidopsis. Plant Mol. Biol. **54**: 743–753.

Radivojac, P., Iakoucheva, L.M., Oldfield, C.J., Obradovi, Z., Uversky, V.N., and Dunker, A.K. (2007). Intrinsic Disorder and Functional Proteomics. Biophysical Journal **92**: 1439–1456.

Realini, C., Rogers, S.W., and Rechsteiner, M. (1994). KEKE motifs. Proposed roles in protein-protein association and presentation of peptides by MHC class I receptors. FEBS Lett. **348**: 109–113.

Richarz, R. and Wüthrich, K. (1978). Carbon-13 NMR chemical shifts of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly- X-L- Ala-OH. Biopolymers **17**: 2133–2141.

Riera, M., Figueras, M., López, C., Goday, A., and Pagès, M. (2004). Protein kinase CK2 modulates developmental functions of the abscisic acid responsive protein Rab17 from maize. PNAS **101**: 9879–9884.

Rinne, Kaikuranta, Van der Plas LH, and Van der Schoot C (1999). Dehydrins in cold-acclimated apices of birch (Betula pubescens ehrh.): production, localization and potential role in rescuing enzyme function during dehydration. Planta **209**: 377–388.

Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., and Dunker, A.K. (2001). Sequence Complexity of Disordered Protein. Proteins: Struct. Funct. Gen. **42**: 38–48.

Ruibal, C., Salamó, I.P., Carballo, V., Castro, A., Bentancor, M., Borsani, O., Szabados, L., and Vidal, S. (2012). Differential contribution of individual dehydrin genes from Physcomitrella patens to salt and osmotic stress tolerance. Plant Sci. **190**: 89–102.

Saavedra, L., Svensson, J., Carballo, V., Izmendi, D., Welin, B., and Vidal, S. (2006). A dehydrin gene in Physcomitrella patens is required for salt and osmotic stress tolerance. Plant J. **45**: 237–249.

Sakai, T., Tochio, H., Tenno, T., Ito, Y., Kokubo, T., Hiroaki, H., and Shirakawa, M. (2006). In-cell NMR spectroscopy of proteins inside Xenopus laevis oocytes. J. Biomol. NMR **36**: 179–188.

Sakakibara, D., Sasaki, A., Ikeya, T., Hamatsu, J., Hanashima, T., Mishima, M., Yoshimasu, M., Hayashi, N., Mikawa, T., Wälchli, M., Smith, B.O., Shirakawa, M., Güntert, P., and Ito, Y. (2009). Protein structure determination in living cells by in-cell NMR spectroscopy. Nature **458**: 102–105.

Šamalíková, M., Santambrogio, C., and Grandori, R. (2010). Mass Spectrometry Tools for the Investigation of Structural Disorder and Conformational Transitions in Proteins. In Instrumental Analysis of Intrinsically Disordered Proteins, V.N. Uversky and S. Longhi, eds (John Wiley & Sons, Inc.), pp. 627–652.

Schallus, T., Fehér, K., Ulrich, A.S., Stier, G., and Muhle-Goll, C. (2009). Structure and dynamics of the human muscle LIM protein. FEBS Lett. **583**: 1017–1022.

Schreiber, G. and Keating, A.E. (2011). Protein binding specificity versus promiscuity. Current Opinion in Structural Biology **21**: 50–61.

Schwarzinger, S., Kroon, G.J., Foss, T.R., Chung, J., Wright, P.E., and Dyson, H.J. (2001). Sequence-dependent correction of random coil NMR chemical shifts. J. Am. Chem. Soc. **123**: 2970–2978.

Schwarzinger, S., Kroon, G.J., Foss, T.R., Wright, P.E., and Dyson, H.J. (2000). Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. J. Biomol. NMR **18**: 43–48.

Selenko, P., Serber, Z., Gadea, B., Ruderman, J., and Wagner, G. (2006). Quantitative NMR analysis of the protein G B1 domain in Xenopus laevis egg extracts and intact oocytes. Proc. Natl. Acad. Sci. U.S.A. **103**: 11904–11909.

Serber, Z., Keatinge-Clay, A.T., Ledwidge, R., Kelly, A.E., Miller, S.M., and Dötsch, V. (2001). High-resolution macromolecular NMR spectroscopy inside living cells. J. Am. Chem. Soc. **123**: 2446–2447.

Sickmeier, M., Hamilton, J.A., LeGall, T., Vacic, V., Cortese, M.S., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V.N., Obradovic, Z., and Dunker, A.K. (2007). DisProt: the Database of Disordered Proteins. Nucleic Acids Res. **35**: D786–793.

Siminovitch, D., Wilson, C.M., and Briggs, D.R. (1953). Studies on the Chemistry of the Living Bark of the Black Locust in Relation to Its Frost Hardiness. V. Seasonal Transformations and Variations in the Carbohydrates: Starch-Sucrose Interconversions 1. Plant Physiol **28**: 383–400.

De Simone, A., Cavalli, A., Hsu, S.-T.D., Vranken, W., and Vendruscolo, M. (2009). Accurate Random Coil Chemical Shifts from an Analysis of Loop Regions in Native States of Proteins. J. Am. Chem. Soc. **131**: 16332–16333.

Slonczewski, J.L., Rosen, B.P., Alger, J.R., and Macnab, R.M. (1981). pH homeostasis in Escherichia coli: measurement by 31P nuclear magnetic resonance of methylphosphonate and phosphate. Proc Natl Acad Sci U S A **78**: 6271–6275.

Soulages, J.L., Kim, K., Arrese, E.L., Walters, C., and Cushman, J.C. (2003). Conformation of a group 2 late embryogenesis abundant protein from soybean. Evidence of poly (L-proline)-type II structure. Plant Physiol **131**: 963–975.

Steponkus, P.L., Uemura, M., Joseph, R.A., Gilmour, S.J., and Thomashow, M.F. (1998). Mode of action of the COR15a gene on the freezing tolerance of Arabidopsis thaliana. Proc. Natl. Acad. Sci. U.S.A. **95**: 14570–14575.

Svensson, J., Palva, E.T., and Welin, B. (2000). Purification of recombinant Arabidopsis thaliana dehydrins by metal ion affinity chromatography. Protein Expr Purif **20**: 169–178.

Swire-Clark, G.A. and Marcotte, W.R., Jr (1999). The wheat LEA protein Em functions as an osmoprotective molecule in Saccharomyces cerevisiae. Plant Mol. Biol. **39**: 117–128.

Szalainé Ágoston, B., Kovács, D., Tompa, P., and Perczel, A. (2011). Full backbone assignment and dynamics of the intrinsically disordered dehydrin ERD14. Biomol NMR Assign **5**: 189–193.

Szalainé Ágoston, B., Magos, Z., Kalmár, L., Felméry, A., Perczel, A., and Tompa, P. (2013). Functional disorder of plant dehydrin ERD14 in vivo.

Szász, C., Alexa, A., Tóth, K., Rakács, M., Langowski, J., and Tompa, P. (2011). Protein Disorder Prevails under Crowded Conditions. Biochemistry **50**: 5834–5844.

Tamiola, K., Acar, B., and Mulder, F.A.A. (2010a). NMR chemical shifts to characterize protein order and disorder.

Tamiola, K., Acar, B., and Mulder, F.A.A. (2010b). Sequence-specific random coil chemical shifts of intrinsically disordered proteins. J. Am. Chem. Soc. **132**: 18000–18003.

Terakawa, T. and Takada, S. (2011). Multiscale ensemble modeling of intrinsically disordered proteins: p53 N-terminal domain. Biophys. J. **101**: 1450–1458.

Theillet, F.-X., Binolfi, A., Liokatis, S., Verzini, S., and Selenko, P. (2011). Paramagnetic relaxation enhancement to improve sensitivity of fast NMR methods: application to intrinsically disordered proteins. J. Biomol. NMR **51**: 487–495.

Thomashow, M.F. (1999). PLANT COLD ACCLIMATION: Freezing Tolerance Genes and Regulatory Mechanisms. Annual Review of Plant Physiology and Plant Molecular Biology **50**: 571–599.

Thongwichian, R. and Selenko, P. (2012). In-cell NMR in Xenopus laevis oocytes. Methods Mol. Biol. **895**: 33–41.

Tompa, P. (2012). Intrinsically disordered proteins: a 10-year recap. Trends Biochem. Sci. **37**: 509–516.

Tompa, P. (2002). Intrinsically unstructured proteins. Trends Biochem. Sci. **27**: 527–533.

Tompa, P. (2009). Structure and Function of Intrinsically Disordered Proteins (CRC Press).

Tompa, P. (2005). The interplay between structure and function in intrinsically unstructured proteins. FEBS Lett. **579**: 3346–3354.

Tompa, P., Bánki, P., Bokor, M., Kamasa, P., Kovács, D., Lasanda, G., and Tompa, K. (2006). Protein-water and protein-buffer interactions in the aqueous solution of an intrinsically unstructured plant dehydrin: NMR intensity and DSC aspects. Biophys J **91**: 2243–9.

Tompa, P. and Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperones. FASEB J **18**: 1169–1175.

Tompa, P. and Fuxreiter, M. (2008). Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends Biochem. Sci. **33**: 2–8.

Tompa, P., Szász, C., and Buday, L. (2005). Structural disorder throws new light on moonlighting. Trends Biochem. Sci. **30**: 484–489.

Tunnacliffe, A. and Wise, M.J. (2007). The continuing conundrum of the LEA proteins. Naturwissenschaften **94**: 791–812.

Uversky, V. and Longhi, S. (2010). Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure and Conformation (John Wiley & Sons).

Uversky, V.N. (2011a). Flexible nets of malleable guardians: intrinsically disordered chaperones in neurodegenerative diseases. Chem. Rev. **111**: 1134–1166.

Uversky, V.N. (2011b). Intrinsically disordered proteins from A to Z. Int. J. Biochem. Cell Biol. **43**: 1090–1103.

Uversky, V.N. (2002). Natively unfolded proteins: a point where biology waits for physics. Protein Sci. **11**: 739–756.

Uversky, V.N. (2010). Seven lessons from one IDP structural analysis. Structure **18**: 1069–1071.

Uversky, V.N., Gillespie, J.R., and Fink, A.L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins: Structure, Function, and Bioinformatics **41**: 415–427.

Vassilev, L.T., Vu, B.T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C., Fotouhi, N., and Liu, E.A. (2004). In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. Science **303**: 844–848.

Wang, Y., Benton, L.A., Singh, V., and Pielak, G.J. (2012). Disordered Protein Diffusion under Crowded Conditions. J Phys Chem Lett **3**: 2703–2706.

Wang, Y. and Jardetzky, O. (2002). Probability-based protein secondary structure identification using combined NMR chemical-shift data. Protein Sci **11**: 852–861.

Wells, M., Tidow, H., Rutherford, T.J., Markwick, P., Ringkjobing Jensen, M., Mylonas, E., Svergun, D.I., Blackledge, M., and Fersht, A.R. (2008). Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. Proc. Natl. Acad. Sci. U. S. A. **105**: 5762–5767.

Wise, M.J. (2003). LEAping to conclusions: A computational reanalysis of late embryogenesis abundant proteins and their possible roles. BMC Bioinformatics **4**: 52.

Wise, M.J. and Tunnacliffe, A. (2004). POPP the question: what do LEA proteins do? Trends Plant Sci **9**: 13–17.

Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S., and Sykes, B.D. (1995). 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. J. Biomol. NMR **5**: 67–81.

Wishart, D.S. and Nip, A.M. (1998). Protein chemical shift analysis: a practical guide. Biochem. Cell Biol. **76**: 153–163.

Wishart, D.S. and Sykes, B.D. (1994). The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J. Biomol. NMR **4**: 171–180.

Wishart, D.S., Sykes, B.D., and Richards, F.M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry **31**: 1647–1651.

Wisniewski, M., Webb, R., Balsamo, R., Close, T.J., Yu, X.-M., and Griffith, M. (1999). Purification, immunolocalization, cryoprotective, and antifreeze activity of PCA60: A dehydrin from peach (Prunus persica). Physiologia Plantarum **105**: 600–608.

Wright, P.E. and Dyson, H.J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol **293**: 321–31.

Xu, J., Zhang, Y., Guan, Z., Wei, W., Han, L., and Chai, T. (2008). Expression and function of two dehydrins under environmental stresses in Brassica juncea L. Molecular Breeding **21**: 431–438.

Xu, J., Zhang, Y.X., Wei, W., Han, L., Guan, Z.Q., Wang, Z., and Chai, T.Y. (2008). BjDHNs Confer Heavy-metal Tolerance in Plants. Mol Biotechnol **38**: 91–98.

Yin, Z., Rorat, T., Szabala, B.M., Ziółkowska, A., and Malepszy, S. (2006). Expression of a Solanum sogarandinum SK3-type dehydrin enhances cold tolerance in transgenic cucumber seedlings. Plant Science **170**: 1164–1172.

Zhang, H., Neal, S., and Wishart, D.S. (2003). RefDB: A database of uniformly referenced protein chemical shifts. Journal of Biomolecular NMR **25**: 173–195.

Zhang, L., Ohta, A., Takagi, M., and Imai, R. (2000). Expression of plant group 2 and group 3 lea genes in Saccharomyces cerevisiae revealed functional divergence among LEA proteins. J. Biochem. **127**: 611–616.

Zhang, X., Ervin, E.H., and LaBranche, A.J. (2006). Metabolic Defense Responses of Seeded Bermudagrass during Acclimation to Freezing Stress. Crop Science **46**: 2598.

Zhang, Y., Li, J., Yu, F., Cong, L., Wang, L., Burkard, G., and Chai, T. (2006). Cloning and expression analysis of SKn-type dehydrin gene from bean in response to heavy metals. Mol Biotechnol **32**: 205–217.

Zilberstein, D., Agmon, V., Schuldiner, S., and Padan, E. (1984). Escherichia coli intracellular pH, membrane potential, and cell growth. J Bacteriol **158**: 246–252.

Zimmerman, S.B. and Minton, A.P. (1993). Macromolecular crowding: biochemical, biophysical, and physiological consequences. Annu Rev Biophys Biomol Struct **22**: 27–65.

Zybailov, B., Rutschow, H., Friso, G., Rudella, A., Emanuelsson, O., Sun, Q., and Van Wijk, K.J. (2008). Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. PLoS ONE **3**: e1994.

8. Summary

Dehydrins are a class of plant stress proteins that belong to the family of late embryogenesis abundant (LEA) proteins. They are known to be involved in the late stages of plant embryogenesis, i.e. seed development, as well as in the plant's response to particular abiotic stress effects, such as cold, dehydration and high salinity. However, their exact role within stress response is unknown, although several investigations have aimed at the determination of the molecular function of dehydrins. Possible functions that have been suggested range from hydration buffers to chaperones, from membrane stabilizers to radical scavengers, and may be different (i.e., not uniform) among different dehydrins. Mostly their conserved regions, the so called segments, are believed to be involved in these functions. The most important segment seems to be the K-segment, which is present in each and every dehydrin at least once, though other segments such as the Y-, S- and ChP-segments are, if present within a given dehydrin, highly conserved as well. From a structural point of view, dehydrins have been shown to be intrinsically disordered proteins, i.e., proteins that do not adopt a well-defined folded structure, but stay flexible even in their native environment.

The aim of the present work was to investigate the dehydrin ERD14 (Early Response to Dehydration 14) from *Arabidopsis thaliana* from a structural as well as functional point of view, under *in vitro* as well as *in vivo* conditions.

For the structural characterization of ERD14, an in depth analysis by NMR was performed under *in vitro* conditions. At the same time, a comparison and validation of the currently used methods for the interpretation of secondary chemical shifts was carried out with special focus on their use on intrinsically disordered proteins. Next, a special NMR technique, called in-cell NMR, was used to analyze structural features of ERD14 within the environment of a living *E. coli* cell. Hereby, it could be shown that ERD14 is intrinsically disordered in dilute solution as well as within the living cell, except for five regions of partial helicity, three of which seem to undergo binding within the cell. The five regions of partial helicity overlap with the conserved regions of ERD14 (K-, S- and ChP-segments), and show helical propensities of about 5-15%. The three binding regions fall within the three K-segments of the protein, indicating the possible functional importance of these regions.

The goal of the functional analysis of ERD14 was to show whether the known *in vitro* chaperone function of the protein is also relevant under *in vivo* conditions. Therefore, an *in vivo* chaperone assay using *E. coli* cells and three different stress treatments, as well as an extra alternative approach, have been developed and implemented. It was shown that ERD14 has potent *in vivo* chaperone function under all four tested conditions and that this function is connected to the presence of the K-segments.

In conclusion, both the structural as well as the functional analysis underline the importance of the K-segments of ERD14 under *in vivo* conditions and confirm the intrinsic disorder of ERD14 in its functional state within the living cell.

Összefoglaló

A dehidrinek növényi stressz fehérjék, melyek a LEA (Late Embryogenesis Abundant) fehérjecsaládba tartoznak. Fontos szerepet töltenek be egyrészt a növényi mag fejlődésében, másrészt a növény abiotikus stressz elleni védekezésében, vagyis hideg, szárazság és magas sótartalom tűrésében. Annak ellenére, hogy már sok vizsgálatban megcélozták a dehidrinek molekuláris funkciójának feltárását, a pontos szerepük a stresszválasz során még mindig homályos. Sokféle lehetséges funkciót vetettek már fel illetve vizsgáltak, a vízpuffertől a chaperon (dajka) hatásig, a membrán stabilizátortól a gyökfogóig, akár oly módon is, hogy az egyes dehidrinek esetleg különböző – vagyis nem egységes – funkciót tölthetnek be. Főként a konzervált régiókat, az úgynevezett szegmenseket, hozzák összefüggésbe az egyes vizsgált funkciókkal. Ezek közül a K-szegmens a legfontosabb, mely minden dehidrinben legalább egyszer előfordul, míg az ugyancsak konzervált Y-, S- és ChP-szegmensek jelenléte nem minden esetben adott. Szerkezeti oldalról nézve a dehidrinek úgynevezett funkcionálisan rendezetlen fehérjék (intrinsically disorderd proteins, IDPs), vagyis olyan fehérjék, melyek természetes körülmények között sem vesznek fel egyetlen határozott térszerkezetet, hanem állandó mozgásban, konformációs átrendeződésben maradnak.

A jelen munka célkitűzése az ERD14 (Early Response to Dehydration 14) nevű dehidrin részletes szerkezeti elemzése és funkcionális vizsgálata, *in vitro* és *in vivo* körülmények között. Az ERD14 szerkezeti elemzéséhez először is egy részletes *in vitro* NMR vizsgálatot végeztem. Ehhez kapcsolódóan összevetettem a másodlagos kémiai eltolódások kiszámítására és kiértékelésére jelenleg használatban lévő módszereket, és megvizsgáltam használhatóságukat funkcionálisan rendezetlen fehérjék analízisében. Ezután alkalmaztam egy speciális NMR technikát, az úgynevezett sejten belüli (in-cell) NMR spektroszkópiát, hogy az ERD14 szerkezeti jellemzőit élő *E. coli* sejten belül is megvizsgálhassam. Ezek során igazoltam, hogy az ERD14 sejten belül ugyanúgy, mint híg oldatban, funkcionálisan rendezetlen fehérje. Szekvenciáján belül kimutattam öt olyan régiót, mely részleges, 5-15%-os hajlamot mutat α-helikális másodlagos szerkezet felvételére. Ezen régiók átfedésben vannak a fehérje konzervált szegmenseivel (K-, S- és ChP-szegmensek). Továbbá megállapítottam, hogy három a K-szegmensekkel egybeeső szekvenciarész az élő sejten belül kötődésben vesz részt, mely e régiók lehetséges funkcionális szerepére utal.

A funkcionális vizsgálat az ERD14 ismert *in vitro* chaperon hatásának *in vivo* körülmények közötti visszaigazolására irányult. Ennek érdekében egy új vizsgálati módszert dolgoztam ki, mely háromféle abiotikus stresszhelyzet *E. coli* sejtekre gyakorolt hatását méri, és melyet egy további, alternatív megközelítést alkalmazó módszerrel egészítettem ki. Ezek használhatóságát az eredeti ERD14 és az egyik ERD14 mutáns fehérjéken mutattam be. A kísérletek azt igazolták, hogy az ERD14 mind a négy vizsgált körülmény esetén nagymértékben képes megvédeni az élő sejteket a pusztulástól, valamint hogy ez a funkciója összefügg a K-szegmensek jelenlétével.

Összefoglalóan elmondható, hogy mind a szerkezeti, mind a funkcionális elemzés alátámasztja a K-szegmensek *in vivo* jelentőségét, és az élő sejten belül is bizonyítja az ERD14 funkcionálisan rendezetlen állapotát.