



HAL
open science

Structural disorder and induced folding within two cereal, ABA stress and ripening (ASR) proteins

Karama Hamdi, Edoardo Salladini, Darragh O'Brien, Sebastien Brier, Alexandre Chenal, Ines Yacoubi, Sonia Longhi

► **To cite this version:**

Karama Hamdi, Edoardo Salladini, Darragh O'Brien, Sebastien Brier, Alexandre Chenal, et al.. Structural disorder and induced folding within two cereal, ABA stress and ripening (ASR) proteins. Scientific Reports, 2017, 7, pp.15544. 10.1038/s41598-017-15299-4 . hal-01802824

HAL Id: hal-01802824

<https://hal.science/hal-01802824v1>

Submitted on 30 May 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

SCIENTIFIC REPORTS



OPEN

Structural disorder and induced folding within two cereal, ABA stress and ripening (ASR) proteins

Karama Hamdi¹, Edoardo Salladini², Darragh P. O'Brien³, Sébastien Brier⁴, Alexandre Chenal³, Ines Yacoubi¹ & Sonia Longhi²

Abstract Abscisic acid (ABA), stress and ripening (ASR) proteins are plant-specific proteins involved in plant response to multiple abiotic stresses. We previously isolated the ASR genes and cDNAs from durum wheat (TtASR1) and barley (HvASR1). Here, we show that HvASR1 and TtASR1 are consistently predicted to be disordered and further confirm this experimentally. Addition of glycerol, which mimics dehydration, triggers a gain of structure in both proteins. Limited proteolysis showed that they are highly sensitive to protease degradation. Addition of 2,2,2-trifluoroethanol (TFE) however, results in a decreased susceptibility to proteolysis that is paralleled by a gain of structure. Mass spectrometry analyses (MS) led to the identification of a protein fragment resistant to proteolysis. Addition of zinc also induces a gain of structure and Hydrogen/Deuterium eXchange-Mass Spectrometry (HDX-MS) allowed identification of the region involved in the disorder-to-order transition. This study is the first reported experimental characterization of HvASR1 and TtASR1 proteins, and paves the way for future studies aimed at unveiling the functional impact of the structural transitions that these proteins undergo in the presence of zinc and at achieving atomic-resolution conformational ensemble description of these two plant intrinsically disordered proteins (IDPs).

Abscisic acid (ABA), stress and ripening (ASR) proteins are a family of plant-specific proteins that have been reported in many species ranging from gymnosperms, (i.e. ginkgo)¹, to monocots (i.e. rice and maize)^{2,3}, and dicots (i.e. grape)⁴. Despite their broad occurrence in plants, ASR proteins lack orthologues in *Arabidopsis spp.* ASR proteins play a crucial role in plant response to multiple abiotic stresses⁵⁻⁷. While some ASR proteins are found exclusively in the nucleus, where they act as transcription factors regulating gene expression during stress response, other ASR proteins are localized in both the nucleus and the cytosol⁸. Beyond this, ASR proteins also act as chaperone-like proteins; for example, tomato ASR1 was found to protect enzymes against freezing or heat denaturation *in vitro*, and similar results were obtained with ASR proteins from plantain and lily⁹.

ASR proteins are small, hydrophilic and highly charged proteins¹⁰, for which structural data is scarce (for a recent review see⁹). Currently, the best structurally characterized ASR protein is tomato ASR1¹¹. Tomato ASR1 is a charged 13-kDa protein, enriched in Gly (7%), Ala, (13%), Glu (15%), His (15%), and Lys (17%)¹². It is over-expressed under water- and salt-stress conditions, and possesses a zinc-dependent DNA-binding activity⁶. Biophysical methods consistently show that tomato ASR1 is an intrinsically disordered protein (IDP) which folds under desiccation and upon addition of zinc ions¹¹. More recently, another ASR protein, namely the product of the MpASR gene from plantain, was also found to be intrinsically disordered¹⁰. In line with these findings, several studies have classified members of the ASR protein family within the superfamily of Late Embryogenesis Abundant (LEA) proteins. LEA proteins are proteins that accumulate in response to low water availability¹³⁻¹⁶, and most members of this superfamily are either entirely or partially disordered⁸.

In recent years, IDPs have attracted much attention. IDPs lack a well-defined structure in their native state and under physiological conditions¹⁷⁻²¹, and exist as highly dynamic and heterogeneous conformational ensembles

¹Laboratoire de Protection et d'Amélioration des Plantes, Centre de Biotechnologie de Sfax (CBS), Sfax, Tunisia.

²Aix-Marseille Univ, CNRS, Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR 7257, Marseille, France. ³Institut Pasteur, CNRS UMR 3528, Unité de Biochimie des Interactions Macromoléculaires, Département de Biologie Structurale et Chimie, Paris, France. ⁴Institut Pasteur, CNRS USR 2000, Unité de Spectrométrie de Masse Structurale et Protéomique, Paris, France. Ines Yacoubi and Sonia Longhi contributed equally to this work. Correspondence and requests for materials should be addressed to I.Y. (email: ines.bouchrityacoubi@cbs.rnrt.tn) or S.L. (email: Sonia.Longhi@afmb.univ-mrs.fr)

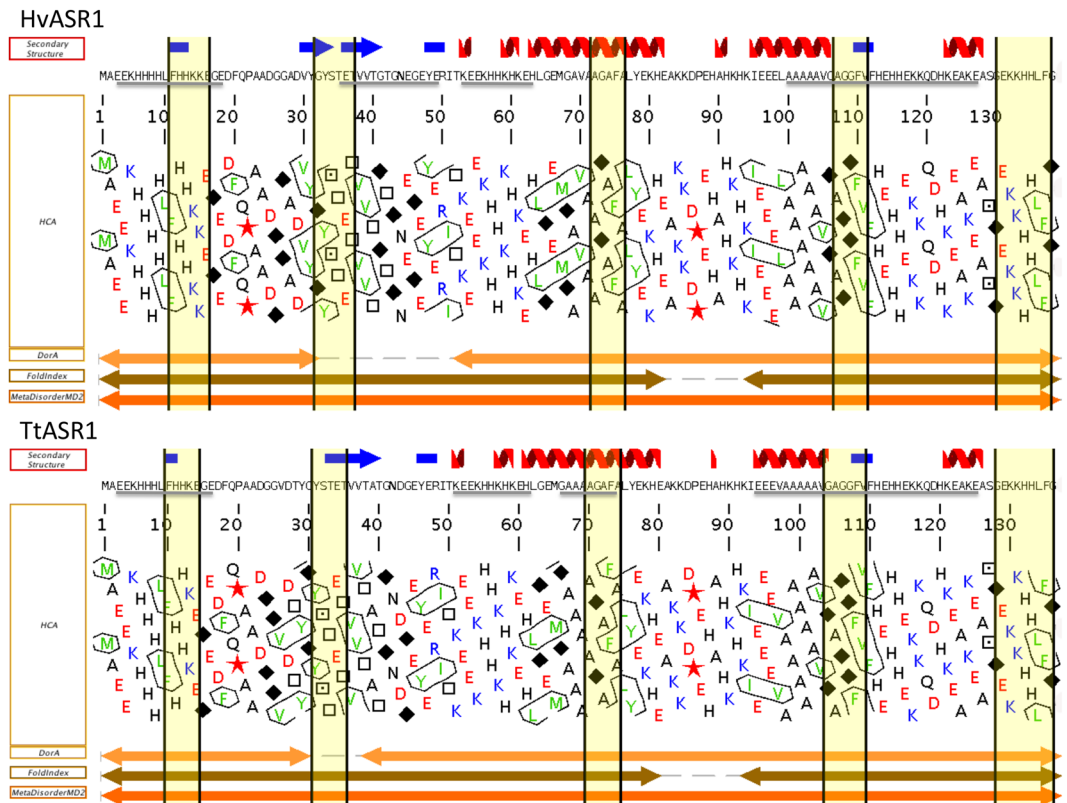


Figure 1. Disorder prediction of HvASR1 and TtASR1. MeDor outputs of HvASR1 (A) and TtASR1 (B) proteins. The sequences are represented as single, continuous horizontal lines below the predicted secondary structure elements. Low sequence complexity regions (HvASR1: aa 3–16, 35–49, 53–63, 100–127; TtASR1: aa 3–16, 51–61, 66–74 and 94–125) are shown by a grey bar below the sequence. Below the sequences are shown the HCA plots and the predicted regions of disorder that are represented by bi-directional arrows. MoREs, as predicted by MoRFpred, are highlighted in yellow.

in the absence of their partner(s) (for a recent review on IDPs, see^{22,23}). IDPs or proteins possessing intrinsically disordered regions (IDRs) are frequently involved in essential processes such as transcriptional and translational activation, chromatin remodeling, and signal transduction. Recent studies have investigated the abundance of intrinsic disorder (ID) in plant proteins^{8,24,25}. The inability of plants to move away from a danger has probably been a major driving factor in the development of their stress response. Plants exhibit a high ability to adapt to variable environmental conditions. Taking into consideration that ID serves as a determinant of interactivity, it has been proposed that it may contribute to such plasticity in plants²⁶.

We have previously isolated the ASR genes and cDNA from durum wheat (TtASR1) and barley (HvASR1) seedlings under abiotic stress conditions (accession numbers KX660743 and KX660744, respectively). Using computational analyses, we show that these proteins possess the peculiar sequence features of IDPs. We then report their bacterial expression, purification and characterization. Using various complementary biochemical and biophysical methods, we show that both TtASR1 and HvASR1 proteins are disordered. Their spectroscopic and hydrodynamic properties are nevertheless indicative of the presence of some residual secondary and tertiary structure. They were found to undergo a disorder-to-order transition in the presence of glycerol, zinc or 2,2,2-trifluoroethanol (TFE).

Results

Disorder predictions and sequence properties of HvASR1 and TtASR1. The amino acid sequences of HvASR1 and TtASR1 were analyzed using the metasever for predicting protein disorder (MeDor) and Genesilico MetaDisorder metasever (Fig. 1). From hydrophobic cluster analysis (HCA), both proteins appear depleted in hydrophobic clusters, indicative of protein disorder (Fig. 1). Additionally, the two proteins are predicted to be disordered over most of their sequence by DorA and FoldIndex (Supplementary Text 1). In further support of this, both proteins are predicted to be disordered over their entire sequence by MetaDisorderMD2 (Fig. 1), the most accurate disorder predictor according to CASP9²⁷. Although HvASR1 and TtASR1 are predominantly disordered, they do possess short regions locally enriched in hydrophobic clusters (Fig. 1). We have previously reported^{28–31}, that such regions often correspond to Molecular Recognition Elements (MoREs), i.e. short order-prone regions within IDRs with a propensity to undergo induced folding upon binding to a partner³². MoRFpred predicts five such regions (Fig. 1). By comparing the sequence composition of HvASR1 and TtASR1 to that of proteins within the SWISS-PROT database, both proteins were found to be enriched in the

Protein	Hydropathy	pI	N	f ₊	f ₋	NCPR	FCR	k	PDR
HvASR1	3.37	6.10	138	0.136	0.201	-0.065	0.34	0.101	2
TtASR1	3.40	5.99	136	0.131	0.204	-0.074	0.33	0.099	2

Table 1. Sequence properties of ASR1 proteins. Protein sequences examined here include the initial methionine and no vector-encoded, non-native residues. N: residue number. f₊: fraction of positively charged residues. f₋: fraction of negatively charged residues. FCR: fraction of charged residues (f₊ + f₋). NCPR: net charge per residue, value of the difference between the fraction of positively charged and negatively positively residues, (f₊ - f₋). The k-value was computed as described³⁸. PDR: phase diagram region as defined in³⁸.

most disorder-promoting amino acids (A, G, R, D, H, Q, K, S, E and P) and depleted in order-promoting residues (W, F, Y, I, M, L, V, C and T) (Supplementary Fig. S1)^{19,33}. Five amino acids (EAHKG) constitute ~68% of the total TtASR1 and HvASR1 residues (Supplementary Fig. S1). This biased sequence composition is also mirrored by the occurrence of four low sequence complexity regions, as identified by the SEG program¹⁹ (Fig. 1). According to the scale developed by³⁴, Glu (E) is the most disorder-promoting amino acid, and it is also the most abundant in both proteins, representing as much as 16.1% and 16.5% of all residues. Finally, HvASR1 and TtASR1 are also predicted to be disordered by their mean hydrophobicity/mean net charge ratio¹⁸, as judged from their location in the left-hand side of the RH-plot (Supplementary Fig. S2). Their CH-distance from the boundary hydrophathy is very similar (0.061 for HvASR1 and 0.062 for TtASR1), supporting a very similar degree of disorder.

Pappu and colleagues recently indicated that sequence polarity and linear distribution of opposite charges were the primary determinants of the chain dimensions and conformational classes of IDPs³⁵. Polar IDPs were found to favor collapsed ensembles in water, despite the absence of hydrophobic groups³⁶. To evaluate sequence polarity, we calculated the net charge per residue (NCPR, f₊ - f₋)^{36,37}, total fraction of charged residues (FCR, f₊ + f₋), and linear distribution of opposite charges (κ value)³⁸ (Table 1). For both proteins, the NCPR is negative and well below the threshold of 0.2 that discriminates collapsed from extended IDPs³⁶, suggesting that they would preferentially populate collapsed states. They belong to the category of weak polyelectrolytes/polyampholytes, where |f₊ - f₋| ≤ 0.2 and both f₊ and f₋ are small. In addition, they have an FCR value of ~0.3. FCR allows for the discrimination between compact globules and swollen coils, leading to a predictive diagram of states³⁵, where region 1 accommodates low-FCR sequences (either weak polyampholytes or polyelectrolytes) that adopt globule or tadpole-like conformations; region 3 houses high-FCR (strong polyampholytes) and low-NCPR sequences that adopt non-globular conformations swollen coil-like conformations (i.e. coil-like and hairpin-like); and region 2 is the boundary between regions 1 and 3, where proteins adopt conformations that likely represent a continuum of possibilities between 1 and 3. ASR1 proteins fall in phase diagram region (PDR) 2³⁵ (Supplementary Fig. S2). Furthermore, in region 3 and upper region 2, linear patterning of opposite charges plays a crucial role in defining the chain dimensions of swollen coils. This patterning is quantified by a single parameter, κ, that varies from 0 (when opposite charges are well mixed) to 1 (when opposite charges are segregated). The average chain dimension of IDPs (radius of gyration, R_g) shows strong anti-correlation with the κ value³⁸. The very low κ value of the ASR1 proteins suggests that these proteins adopt a rather extended and swollen, coil-like conformation.

In conclusion, all of the *in silico* analyses consistently converge to predict that HvASR1 and TtASR1 are members of the family of IDPs that adopt a swollen coil-like conformation.

Expression and purification of HvASR1 and TtASR1 proteins. To experimentally assess the disordered nature of HvASR1 and TtASR1, we cloned the cDNAs encoding full-length ASR1 proteins into the pGEX4-T1 expression vector that allows the inducible expression in *E. coli* of N-terminally glutathione S transferase (GST) tagged proteins. The tagged proteins were purified by affinity chromatography, followed by thrombin cleavage to remove the GST tag and size-exclusion chromatography (SEC) (Fig. 2). Both HvASR1 and TtASR1 exhibit an abnormally slow migration in SDS-PAGE, with an apparent molecular mass (MM) comprised between 20 and 25 kDa (expected MM ~16 kDa) (Fig. 2). MALDI-TOF and native electrospray ionization (ESI) mass spectrometry (MS) analyses yielded the exact molecular mass expected for both proteins (data not shown and Fig. 3). This aberrant migration during electrophoresis is a hallmark of IDPs and is often due to their typically high content of acidic and negatively charged residues, which results in a lower binding of Sodium dodecyl sulfate (SDS) than usual²¹. As a result, their apparent MM is often 1.2–1.8 times higher than that which is calculated from sequence data or measured by MS. Furthermore, we have previously reported that the degree of protein extension in solution is an additional parameter affecting the electrophoretic mobility of IDPs³⁹. The aberrant electrophoretic migration of HvASR1 and TtASR1 proteins constitutes the first experimental hint of their disordered nature.

Hydrodynamic properties of HvASR1 and TtASR1 proteins from SEC. To investigate the hydrodynamic properties of HvASR1 and TtASR1 we used analytical SEC. The two proteins were eluted from the gel filtration column as sharp peaks (data not shown), with a similar elution volume. The elution profiles were similar, irrespective of whether sodium phosphate or Tris/HCl buffer was used, and regardless of the NaCl concentration. The Stokes radius (R_s) was estimated to be 24.7 ± 2 Å for HvASR1, and 25.2 ± 2 Å for TtASR1 (see Table 2). These very high Stokes radius (R_s) values can't be ascribed to protein aggregation, since they are independent from protein concentration (data not shown). Different protein conformational classes have characteristic hydrodynamic dimensions and molecular mass correlations⁴⁰. IDPs have larger hydrodynamic dimensions compared to typical native globular proteins⁴⁰. By comparing the experimentally determined Stokes radius (R_s^{obs}) of HvASR1 and TtASR1 with the Stokes radii expected for various conformational states (native, molten globule, pre-molten globule, and denaturant-unfolded), the two proteins were found to have an R_s higher than the value expected for

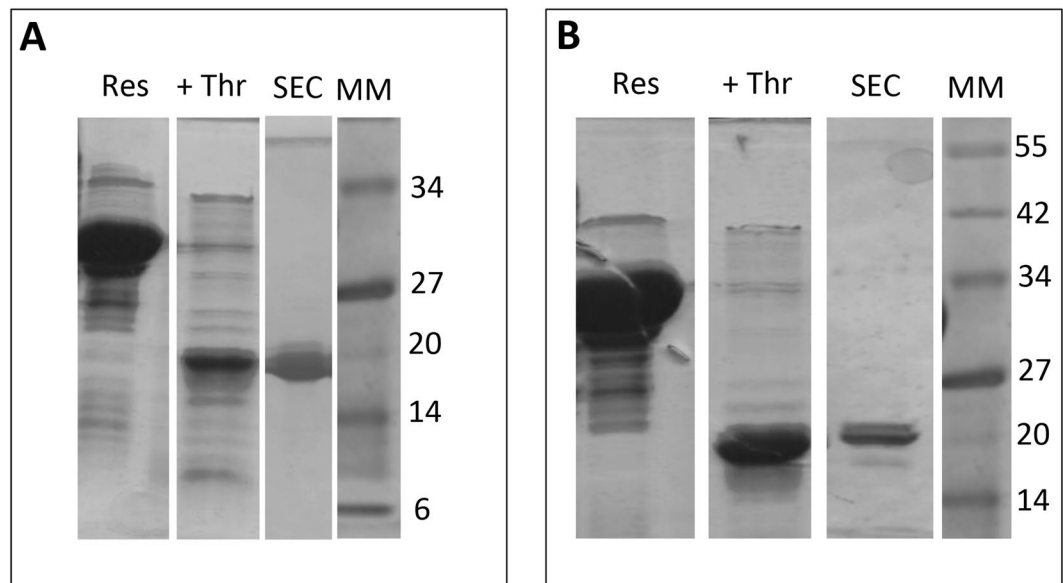


Figure 2. Purification of HvASR1 (A) and TtASR1 (B) proteins from *E. coli*. Coomassie blue staining of SDS-PAGE showing the affinity chromatography resin after extensive washing (Res), the released protein after thrombin cleavage (+Thrombin) and the final purified product as obtained after SEC (SEC). MM: molecular mass markers.

a monomeric, natively folded form (~ 20 Å) (Table 2). The experimentally observed values of the R_g are very close to those expected either for a premolten globule (PMG) form (i.e. an extended conformation possessing some residual structure⁴⁰) (~ 27.3 Å), or for a molten globule (MG) form (~ 22.3 Å), or for a folded dimer (~ 25.5 Å) (see ratios between R_g^{obs} and R_g^{PMG} or R_g^{MG} or R_g^{DimNF} in Table 2). These results suggest that HvASR1 and TtASR1 proteins are either folded dimers or IDPs adopting a MG or PMG conformation.

Native electrospray ionization mass spectrometry (ESI-MS) analysis of HvASR1 and TtASR1 proteins. To determine whether the ASR1 proteins are prevalently unfolded monomers or folded dimers, we carried out native electrospray ionization MS (ESI-MS) analysis of the HvASR1 and TtASR1 proteins. To ascertain the suitability of the approach to detect possibly occurring oligomeric species in a protein sample, we first carried out control experiments using alcohol dehydrogenase (ADH, a tetrameric protein)⁴¹. Results show that under the conditions herein used, ADH conserves its tetrameric organization (Supplementary Fig. S3). Results show that both TtASR1 and HvASR1 proteins are present as monomers in the gas phase (Fig. 3). No m/z values corresponding to dimeric forms of each were detected thus confirming that the large R_g observed in SEC studies reflects a predominantly unfolded monomeric species. The multiple and high charge states observed (up to +17) confirm the intrinsically disordered nature of the proteins.

Differential scanning fluorimetry of HvASR1 and TtASR1 proteins. The conformation of HvASR1 and TtASR1 was further explored by differential scanning fluorimetry (DSF). This method is used to monitor thermal transitions of proteins in the presence of a fluorescent dye that is highly fluorescent in non-polar environments, such as the hydrophobic pockets of (partly) unfolded proteins, and which is quenched in aqueous solutions and/or in the presence of native proteins (Supplementary Text 2)⁴². As shown in Supplementary Fig. S4, the experimentally observed profiles for HvASR1 and TtASR1 are consistent with lack of a stable 3D structure, as judged from their rather high basal fluorescence at 20 °C and from the flatness of their profile. These results thus confirm their disordered nature and advocate for a PMG rather than a MG conformation.

Conformational properties of HvASR1 and TtASR1 proteins from small angle X-ray scattering (SAXS) studies. Small-angle X-ray scattering (SAXS) is well suited to study flexible, low compactness or even extended macromolecules in solution^{43,44}. The SAXS curves and Guinier plots obtained at different protein concentrations are independent of protein concentration, indicating the absence of significant aggregation (data not shown). Each curve can be approximated by a straight line in the Guinier region ($qR_g < 1.0$). The slope gives the value of the radius of gyration, R_g , while the intercept of the straight line gives the $I(0)$, which is proportional to the molecular mass of the scatterer. Guinier analysis in the low q region gave an R_g of 34.6 ± 0.6 Å for HvASR1 and 35.5 ± 0.3 Å for TtASR1 at the highest protein concentration (Supplementary Fig. S5 and Fig. 4A and Table 3). Very similar values were obtained at lower concentrations (Table 4), and in good agreement with the values (35.7 ± 0.2 for HvASR1 and 35.8 ± 0.5 for TtASR1) determined from the pair distance distribution function $P(r)$ (Supplementary Fig. S5 and Fig. 4B and Table 3). The molecular mass determined from the extrapolated scattering intensity at zero angle $I(0)$ is 12.8 kDa for HvASR1 and 13.0 kDa for TtASR1. These values are close, although a bit smaller, to the values expected for a monomeric form of the two proteins (Table 3).

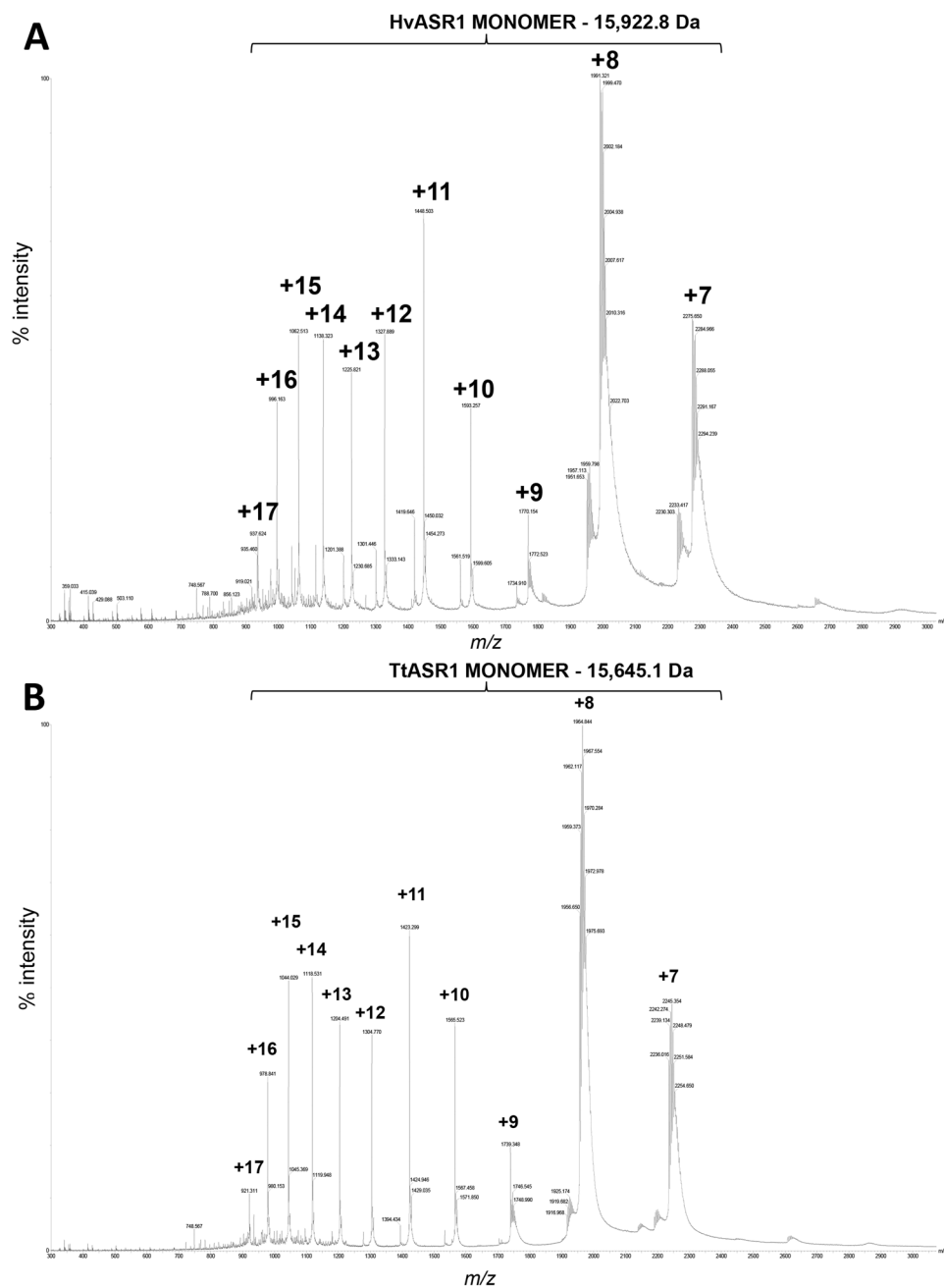


Figure 3. Native ESI-MS spectra of HvASR1 (A) and TtASR1 (B) in 250 mM ammonium acetate, pH 8 (Synapt G2-Si mass spectrometer, Waters).

Protein	Mass (Da)	R_s^{obs}	R_s^{NF}	R_s^{MG}	R_s^{PMG}	R_s^{U}	R_s^{IDP}	$R_s^{\text{Dim NF}}$	$R_s^{\text{obs}}/R_s^{\text{MG}}$	$R_s^{\text{obs}}/R_s^{\text{PMG}}$	$R_s^{\text{obs}}/R_s^{\text{U}}$	$R_s^{\text{obs}}/R_s^{\text{IDP}}$	$R_s^{\text{obs}}/R_s^{\text{Dim NF}}$	CI
HvASR1	15922	24.7	19.8	22.4	27.4	34.7	31.2	25.5	1.10	0.90	0.71	0.79	0.97	0.67
TtASR1	15645	25.2	19.7	22.3	27.2	34.4	31.0	25.5	1.13	0.92	0.73	0.81	0.99	0.62

Table 2. Stokes radii (R_s^{obs} , Å) as obtained by SEC and expected values for the various conformational states. R_s^{NF} : R_s expected for a natively folded form; R_s^{MG} : R_s expected for a molten globule (MG); R_s^{PMG} : R_s expected for a premolten globule (PMG); R_s^{U} : R_s expected for a fully unfolded form; R_s^{IDP} : R_s expected for an IDP based on the simple power-law model; $R_s^{\text{Dim NF}}$: R_s expected for dimeric, folded form. M: molecular mass calculated from the amino acid sequence of the recombinant protein. CI: compaction index.

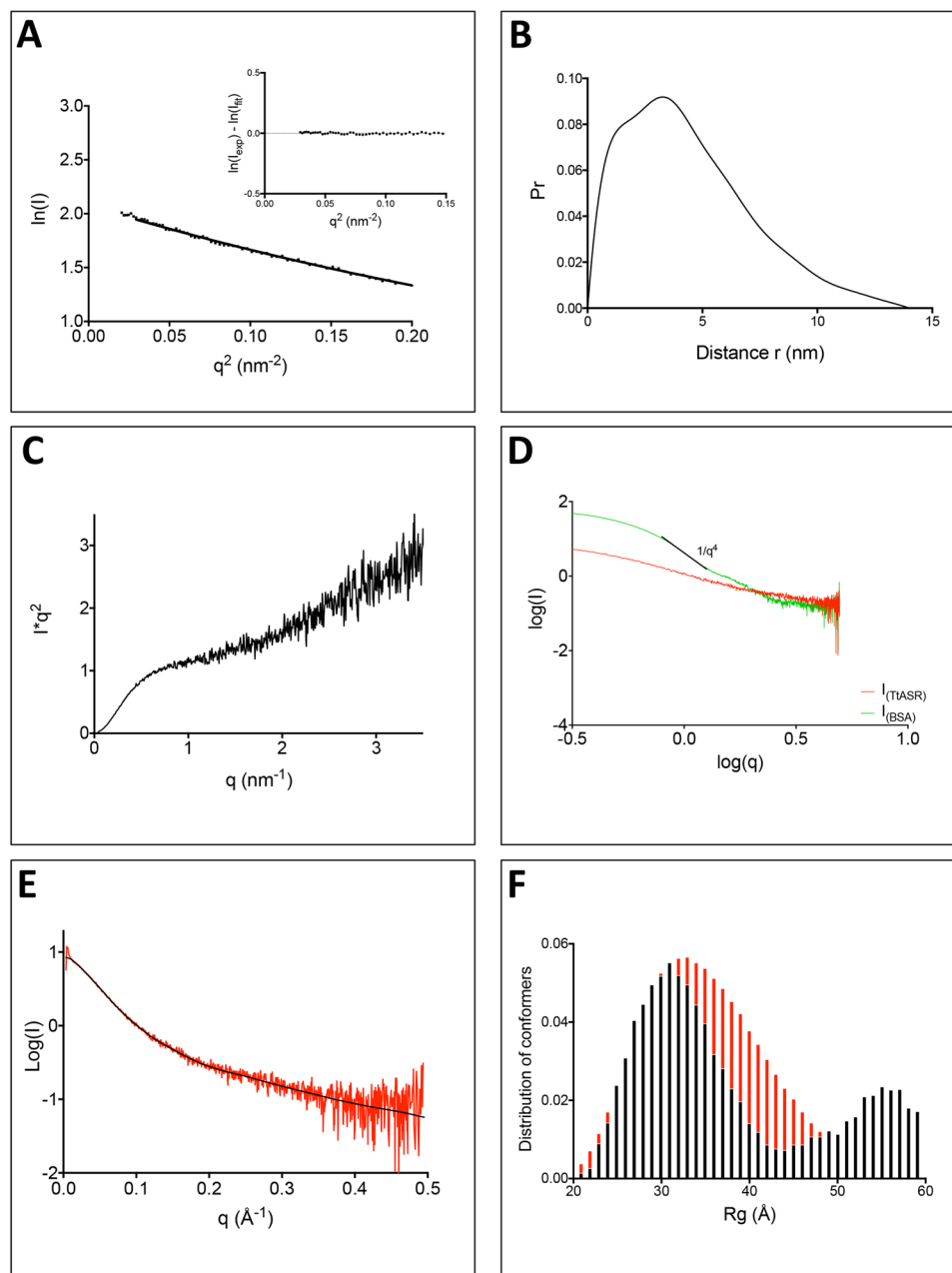


Figure 4. SAXS studies of TtASR1. (A) Representation of the Guinier plot for the protein at 3.0 g/L. Inset: residuals. (B) Pair distance distribution, $P(r)$, function of the data for the 3.0 g/L concentration. (C) Kratky plot of the SAXS data obtained at 1.5 g/L. (D) Log-log representation of TtASR1 and BSA SAXS patterns. (E) Experimental scattering curve of the protein at 3.0 g/L (red) and GAJOE fit (black). (F) Distribution of R_g of the ensemble of randomly generated conformers by Flexible-Meccano without constraints (red) and of the sub-ensemble of selected conformers using GAJOE (black).

The R_g value obtained for the ASR1 proteins ($\sim 35 \text{ \AA}$) is 2.25 times higher than that expected for globular, fully unfolded proteins and IDPs of the same size ($\sim 15.5 \text{ \AA}$) (see Eq. 10 in Methods)⁴⁵ On the other hand, the expected value for a fully unfolded form is $\sim 110 \text{ \AA}$ (see Eq. 12 in Methods). The R_g values expected for IDPs of the same length as the ASR1 proteins, as calculated using Flory's power law and parameters based on an IDP pool⁴⁶, are 33.9 \AA for HvASR1 and 33.6 \AA for TtASR1 (see Eq. 10 in Methods). The experimental R_g values are therefore in very good agreement with the expected R_g values using parameters inferred from IDPs, while they are three times smaller than those expected for fully unfolded forms. Denaturing conditions are indeed known to lead to excluded volume effects around the polypeptide chain, thereby resulting in higher values of R_g , compared to IDPs. The experimentally observed R_g values are ~ 1.7 times larger than those expected for globular proteins ($\sim 19 \text{ \AA}$), with an R_s equal to that experimentally observed for the ASR1 proteins ($\sim 25 \text{ \AA}$) (see Eq. 13 in Methods), thus suggesting an extended shape. They are both much higher than the expected value for a sphere ($\sim 12.8 \text{ \AA}$),

	HvASR1	TtASR1
Data-collection parameters		
Detector	Pilatus (1 M)	Pilatus (1 M)
Beam geometry	Bending magnet (BM29)	Bending magnet (BM29)
Wavelength (Å)	0.992	0.992
q Range (Å ⁻¹)	0.028–4.525	0.028–4.525
Exposure time (s)	1	1
Concentration range (g/L)	1–1.5	1–3
Temperature (°C)	20	20
Structural parameters		
I(0) (cm ⁻¹) (from Guinier)	12.7 ± 0.8	12.5 ± 0.5
R _g (nm) (from P(r))	3.57 ± 0.02	3.58 ± 0.05
R _g (nm) (from Guinier)	3.46 ± 0.06	3.55 ± 0.03
D _{max} (nm)	11.9	12.3
Molecular-mass determination (kDa)		
Molecular mass (MM) (from I(0))	12.8	13
Calculated MM from sequence	15.9	15.7
Software employed		
Primary data reduction	PRIMUS	PRIMUS
Data processing	GNOM	GNOM
Validation of structural models	CRY SOL/GAJOE	CRY SOL/GAJOE

Table 3. SAXS data collection and scattering-derived structural parameters (at the highest concentration) for HvASR1 and TtASR1.

Protein concentration (g/L)	R _g (nm) (Guinier)	D _{max} (nm)
HvASR1 1.0	3.51 ± 0.09	13.47
HvASR1 1.5	3.46 ± 0.06	11.90
TtASR1 1.0	3.31 ± 0.08	14.87
TtASR1 1.5	3.49 ± 0.06	11.75
TtASR1 2.0	3.58 ± 0.02	12.17
TtASR1 3.0	3.55 ± 0.03	12.31

Table 4. R_g (from Guinier) and D_{MAX} for HvASR1 and TtASR1 at the various protein concentrations.

with an R_g of ~25 Å, as determined from the volume of a sphere with either 141 or 143 residues (see Methods). The strong discrepancy between the experimentally observed R_g values and those expected for a globular/spherical form, together with the good agreement with those expected for IDPs, indicate that the ASR1 proteins are disordered. The distribution of internal distances, as inferred from the scattering curves obtained at the highest protein concentration, yielded a maximal internal dimension D_{max} of 119 Å for HvASR1 and of 123 Å for TtASR1 (Supplementary Fig. S5 and Fig. 4B). This large D_{max} also indicates that the proteins are extended.

The Kratky plots of the two ASR1 proteins (Supplementary Fig. S5 and Fig. 4C) have plateaus >1.0 nm⁻¹. The absence of a maximum clearly indicates that the proteins are not globular and do not possess a tightly packed core (Supplementary Text 3). In addition, from the log-log representation of the scattering curves, no region with a s⁻⁴ dependence above 0.1 Å⁻¹ was observed (Supplementary Fig. S5 and Fig. 4D), a feature indicative of the absence of a clear boundary between the solvent and the protein surface. By contrast, the curve of BSA exhibits a region with a s⁻⁴ dependence characteristic of the presence of a sharp interface between the solvent and a compact, well-folded domain, as stated in Porod's law⁴⁷.

We then assessed the distribution of conformations of the ASR1 proteins in solution. For each protein, a pool of 10,000 conformations was randomly generated using Flexible-Meccano. From this initial ensemble, a sub-ensemble of conformers that collectively reproduces the experimental SAXS data and represents the distribution of structures adopted by the protein in solution was selected. The average SAXS scattering curves back-calculated from the selected sub-ensembles reproduce correctly the experimental curves (χ² of 0.70 for HvASR1 and 0.78 for TtASR1) (Supplementary Fig. S5 and Fig. 4E). The R_g distribution of the initial ensembles is broad and symmetrical with values extending from 20 to 60 Å nm with a maximum frequency near 33 Å (Supplementary Fig. S5 and Fig. 4E, red bars). For both proteins, the R_g distribution of the selected sub-ensemble is a bit wider and bimodal, i.e. it displays two peaks, centered on 33 and 60 Å, suggesting the presence of two distinct sub-populations of conformers (Supplementary Fig. S5 and Fig. 4F, black bars). This implies that the scattering curves of the two ASR1 proteins do not reflect a randomly distributed ensemble of conformations. Note that successive and independent selections by GAJOE yielded similar ensembles of conformers, indicating that the distribution of the selected sub-ensemble was reproducible (data not shown).

Altogether, these analyses consistently confirm the disordered nature of the ASR1 proteins.

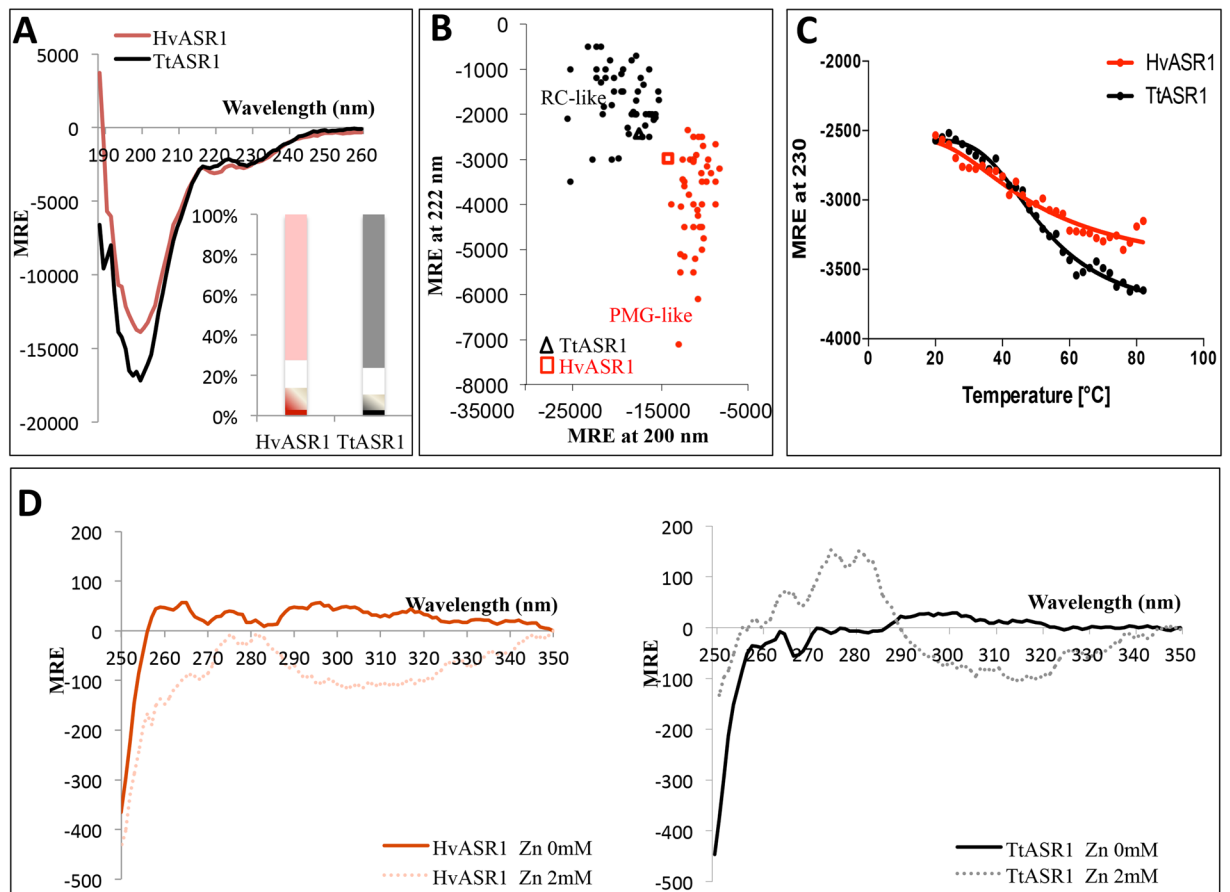


Figure 5. Analysis of HvASR1 and TtASR1 by circular dichroism. **(A)** Far-UV CD spectra of HvASR1 and TtASR1 at 0.1 mg/mL in 10 mM sodium phosphate pH 7 at 20 °C. Data are representative of one out of three independent measurements. The inset shows the secondary structure content of the two proteins, as derived using CDSSTR. Grey or light red: unordered; white: turns; black or red: helix; hatched: strand. **(B)** Plot of the molar residue ellipticity (MRE) at 222 nm and at 200 nm of a set of well-characterized unfolded, random coil-like (RC-like) or premolten globule-like (PMG-like) proteins (from⁴⁰). The position in the plot of HvASR1 and TtASR1 is highlighted by a square and a triangle, respectively. **(C)** Molar residue ellipticity (MRE) at 230 nm of HvASR1 and TtASR1 as a function of the temperature. The experimental points in the 20–80 °C range were fitted to a sigmoidal curve. Protein and buffer concentrations were the same as described in (A). **(D)** Near-UV CD spectra of HvASR1 and TtASR1 at 1 mg/mL in 10 mM sodium phosphate pH 7 in the absence or presence of 2 mM ZnSO₄ at 20 °C. Data are representative of one out of two independent acquisitions.

Circular dichroism (CD) studies of HvASR1 and TtASR1 proteins. To estimate the secondary structure content of HvASR1 and TtASR1, we recorded their circular dichroism (CD) spectra in the far ultraviolet (UV) region. Under native conditions and neutral pH, both proteins present spectra typical of proteins lacking any stable organized secondary structure, as judged from their large negative ellipticity at 200 nm, low amplitude in the 210–230 nm region, and low ellipticity at 190 nm (Fig. 5A). Spectral deconvolution revealed a high content (>60%) of unordered structure in both ASR1 proteins (see insets in Fig. 5A).

It has been previously noticed that IDPs can be subdivided in PMG-like and Random Coil-like (RC-like) forms as a function of their ellipticity values at 200 and 222 nm⁴⁰. As shown in Fig. 5B, HvASR1 and TtASR1 fall in the RC-like and PMG-like region, respectively, with both being located close to the boundary between the two classes. This indicates that TtASR1 is slightly more extended than HvASR1, a finding in agreement with SEC data that highlighted a slightly smaller compaction index (CI) for TtASR1 (see Table 2).

We also monitored the ellipticity at 230 nm as a function of temperature (Fig. 5C). In both cases, increasing the temperature globally results in an increase in the ellipticity (Fig. 5C). While the profile obtained for TtASR1 could be satisfactorily fitted to a sigmoidal curve (with a calculated inflection point of folding of 51 ± 8 °C), the quality of the fit for HvASR1 was poor, leading to a much less reliable estimation of the folding temperature (61 °C) (Fig. 5C). Regardless of these subtle differences, the profiles indicated that both ASR1 proteins undergo temperature-induced folding, as already observed for other IDPs^{48,49}. This heat-induced folding is reversible, with the spectra reverting to their initial appearance after re-cooling at room temperature (data not shown). To rule out any possible temperature-induced aggregation, we monitored the HT voltage associated with the transmission of

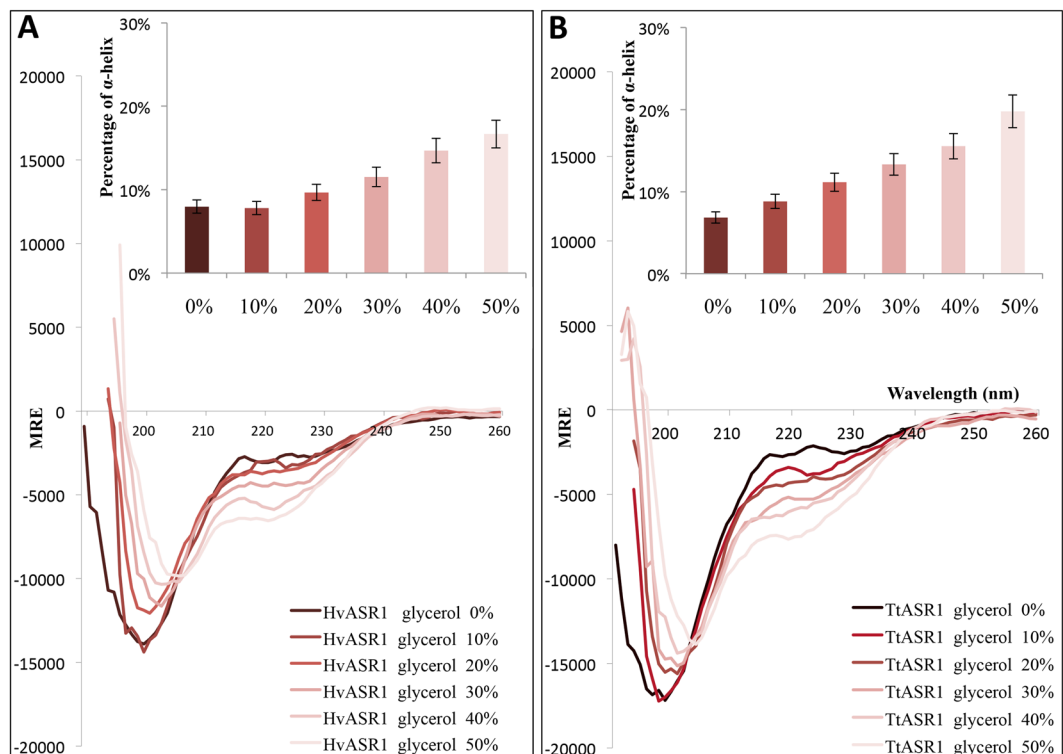


Figure 6. Far-UV CD spectra of HvASR1 (A) and TtASR1 (B) in the presence of increasing concentrations of glycerol at 20 °C. Proteins were at 0.1 mg/mL in 10 mM sodium phosphate pH 7. Data are shown to the point up to which the dyna voltage was in the permissible range. Data are representative of one out of two independent acquisitions. The insets show the α -helical content as derived as described in Methods (see Eq. 14). MRE: molar residue ellipticity.

light through the proteins in the CD spectropolarimeter during heating of the proteins. We observed no significant changes in the HT voltage associated with heating, indicating a lack of aggregation (data not shown).

We also recorded the CD spectra of HvASR1 and TtASR1 in the near ultraviolet region (250–350 nm). Near-UV CD spectra reflect the environment of the aromatic amino acid side chains and provide information about the tertiary structure of proteins⁵⁰. The near-UV CD spectra of both proteins are characterized by a weak intensity, and by the presence of a broad negative band (Fig. 5D), indicating that unlike typical globular proteins, these two proteins lack a hydrophobic core containing oriented aromatic residues, i.e. they lack a stable tertiary structure⁵¹.

Folding induced by glycerol. As desiccation has been previously reported to induce folding of tomato ASR1¹¹, as well as of a number of LEA proteins^{52,53}, we sought to assess whether HvASR1 and TtASR1 also undergo some degree of folding under such conditions. In this regard, we recorded their far-UV CD spectra in the presence of increasing concentrations of glycerol, a condition that mimics dehydration. As shown in Fig. 6, addition of increasing glycerol concentrations triggers a gradual increase in the α -helical content. Note that for both proteins, the spectra display an isodichroic point around 205 nm, a behavior indicative of a two-state transition.

To ascertain whether the gain of structure observed in the presence of glycerol arises from dehydration or from crowding effects (i.e. increased excluded volume effects), we also recorded the far-UV CD spectra of both ASR1 proteins in the presence of 30% of sucrose. The addition of sucrose does not induce a pronounced folding in the two ASR1 proteins and triggers only a modest increase in the α -helical content of HvASR1 (Supplementary Fig. S6), that is not mirrored in TtASR1 (Supplementary Fig. S6). In the latter however, the overall content in disorder slightly decreases, as judged from the decrease in the amplitude of the negative peak at 200 nm upon addition of sucrose. In neither case, the structural transition is comparable to that observed with glycerol, advocating for a mechanism where the protein probably folds not because of excluded volume effects, but rather because hydrogen bonds with water molecules are replaced by intramolecular ones.

Folding induced by TFE. Many IDPs undergo some degree of folding upon binding to their partners/ligands²³. To further study the potential of HvASR1 and TtASR1 to adopt a stable regular secondary structure, their CD spectra were recorded in the presence of increasing concentrations of TFE (Fig. 7). TFE is a secondary structure stabilizer that mimics the hydrophobic conditions that proteins experience when binding hydrophobic patches in their targets, and is commonly used to probe hidden structural propensities of IDPs (Supplementary Text 4)^{54–56}. In the presence of TFE, both HvASR1 and TtASR1 show a notable increase in ordered structure. Their α -helicity increases with increasing TFE concentrations (see inset in Fig. 7). In the presence of 50% of TFE, the

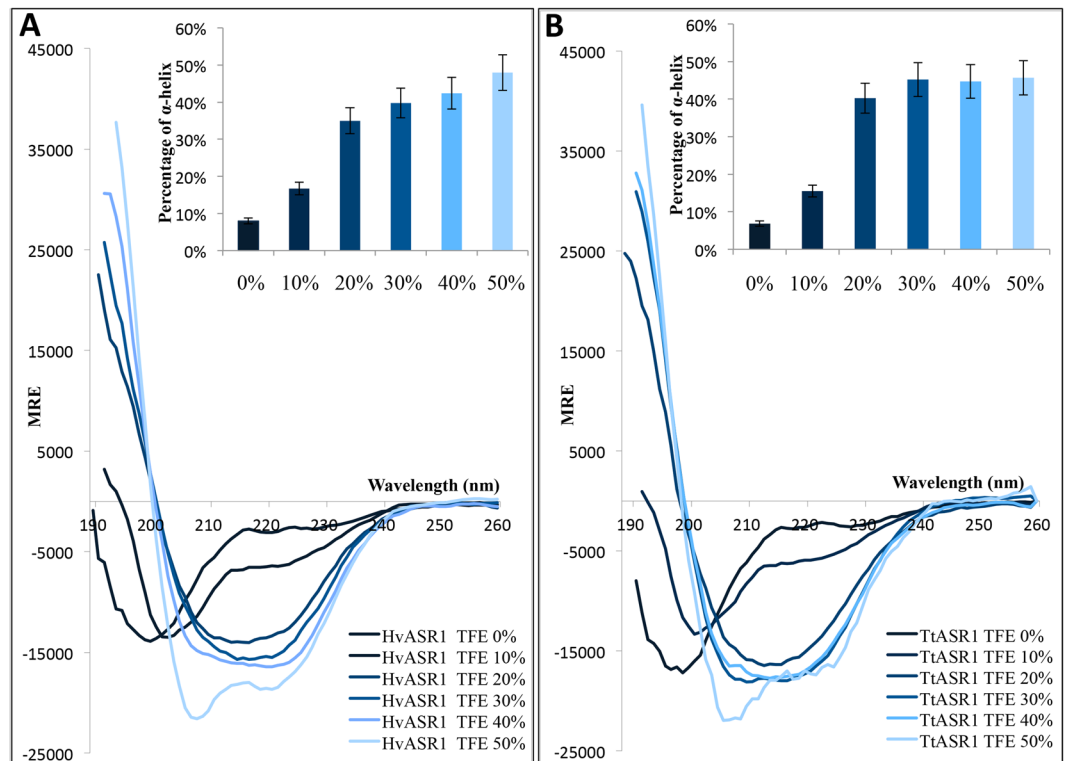


Figure 7. Far-UV CD spectra of HvASR1 (A) and TtASR1 (B) in the absence or presence of increasing TFE concentrations at 20°C. Proteins were at 0.1 mg/mL in 10 mM sodium phosphate pH 7. Data are shown to the point up to which the dyna voltage was in the permissible range. Data are representative of one out of two independent acquisitions. The insets show the α -helical content as derived as described in Methods (see Eq. 14). MRE: molar residue ellipticity.

spectra exhibit a clear α -helical character, as judged from the appearance of the characteristic double minima at 208 and 222 nm (Fig. 7). Note that for both proteins, no isodichroic point is discernable, arguing for a more than two-state TFE-induced structural transition.

Protease sensitivity in the absence and presence of TFE and mass spectrometry analyses. We also investigated HvASR1 and TtASR1 by limited proteolysis either in the presence or absence of TFE. Limited proteolysis is widely used to identify folded domains within (modular) proteins⁵⁷. Proteolytic cleavage sites within unfolded regions are cut first, with more structured regions remaining undigested⁵⁷. A corollary of this is that IDPs are hypersensitive to proteolysis^{21,58}. For digestion, we used thermolysin, a TFE-resistant enzyme that has broad substrate specificity. The latter property allows the identification of cleavage sites solely on the basis of their location in flexible/exposed regions. HvASR1 and TtASR1 were submitted to limited thermolysin digestion either in the presence or absence of 15% TFE, i.e. the minimum concentration to obtain a fragment resistant to proteolysis (Fig. 8). In the absence of TFE, the two proteins are readily degraded after an incubation as short as one hour, and entirely digested after 6 hours (Fig. 8), a behavior that is consistent with an overall high solvent accessibility and disordered nature. Conversely, they show greater resistance to digestion in the presence of 15% TFE (Fig. 8). For both proteins, a few resistant fragments could be detected, of which one was clearly more resistant to digestion than the others, even after an incubation period as long as 24 hours (Fig. 8, arrow 2). Mass spectrometry (MS) analysis allowed the resistant fragment of HvASR1 and TtASR1 to be mapped to the regions encompassing residues 99–122 and 98–118 of both proteins, respectively. Notably, this fragment covers a predicted α -helix and MoRE in each (see Fig. 1). It is therefore reasonable to assume that under these conditions, this fragment is folded as an α -helix, preventing its cleavage and digestion by thermolysin. This region may therefore represent a secondary structure element that is involved in the disorder-to-order transition that the proteins may undergo upon binding their physiological partner(s).

Effect of zinc. As tomato ASR1 was previously shown to bind zinc ions and undergo Zn-induced folding^{11,59}, we checked whether HvASR1 and TtASR1 retain this ability. Binding experiments to a sepharose resin, previously equilibrated with ZnSO_4 , unveiled that the two proteins are retained on the resin after extensive washing (Supplementary Fig. S7). The possibility that the proteins could stick to the resin itself was checked and ruled out (Supplementary Fig. S7). Concomitantly, these experiments also unveiled that the two proteins are equally able to bind to a resin previously equilibrated with NiSO_4 (Supplementary Fig. S7). Incidentally, these findings also have a practical interest, as they point to a behavior that could be exploited for purification purposes.

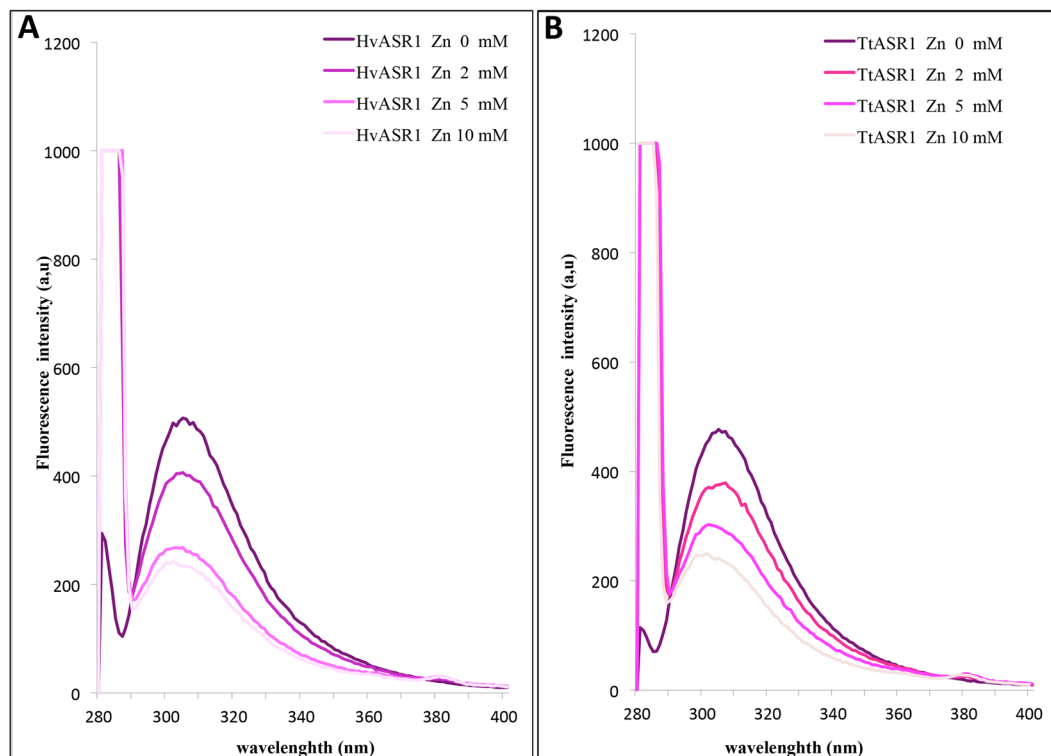


Figure 8. SDS–PAGE analysis of limited digestion of HvASR1 and TtASR1 by thermolysin both in the presence and absence of 15% TFE at 25 °C. The extent of thermolysin digestion of the purified proteins at different time intervals (0, 1, 3, 6 and 24 h) in the presence of 15% TFE is shown. The extent of digestion of the two proteins, as observed after 1 and 6 h in the absence of TFE is also shown. The black arrow shows undigested ASR1 proteins (1) and the red arrow a fragment resistant to proteolysis (2). M: molecular mass markers. The amino acid sequence of the proteins is shown and the region corresponding to the resistant fragment is shown in red. The underlined region corresponds to the region, shared by both ASR1 proteins, that becomes more structured in the presence of zinc (see also Fig. 11).

We next assessed the impact of zinc ions on ASR1 structure. We first recorded the fluorescence emission spectra of the proteins either in the presence or absence of increasing zinc concentrations (Fig. 9). As expected for Tyr-containing proteins, both ASR1 proteins exhibit a maximum close to 303 nm. Interestingly, the intensity of the emission peak is gradually quenched with increasing zinc concentration, reflecting a decrease in the solvent exposure of Tyr residues. This behavior is indicative of a zinc-induced gain of tertiary structure.

Near-UV CD studies further support this conclusion (Supplementary Text 5). For both proteins, the addition of 2 mM ZnSO_4 results in a spectral modification, consistent with the gain of some tertiary structure (Fig. 5D). In particular, a peak between 275 and 282 nm, which corresponds to Tyr residues⁶⁰, becomes discernible, reflecting a conformational change in the environment of Tyr residues (Fig. 5D). Surprisingly however, the near-UV CD spectra obtained in the presence of zinc differ for the two proteins, a finding that cannot be easily rationalized, as the two proteins have the same number of Tyr (and Phe) residues whose location in the sequence is also conserved.

We then recorded the far-UV CD spectra in the presence of zinc (Fig. 10). The addition of 2.5 mM ZnSO_4 does not trigger a gain in α -helical structure, but rather, leads to a decrease in the amplitude of the negative peak at 200 nm, consistent with a decrease in disorder content. By combining 10% TFE and 2 mM ZnSO_4 , a much more pronounced gain of structure is observed (Fig. 10), resulting in an increase in α -helical content, as judged from the appearance of the characteristic double minima at 208 and 222 nm (Fig. 10). While the addition of Zn^{2+} and 10% TFE does not promote an additional gain in α -helicity in HvASR1 (Fig. 10A), the reverse is observed for TtASR1, where the effects appear cumulative (Fig. 10B). These differences in folding propensities might reflect subtle functional differences between the two proteins. In agreement with a zinc-induced gain of structure, zinc binding also results in a decreased susceptibility of both ASR1 proteins to trypsin digestion (data not shown), as already reported for tomato ASR1^{11,59}.

To gain further insights into the precise character of this structural transition, and to precisely identify the protein region where this transition takes place, we performed Hydrogen/Deuterium eXchange–Mass Spectrometry (HDX–MS) analysis.

The unlabeled ASR1 proteins were first digested with both *A. saitoi* protease Type XIII and pepsin to obtain peptides for monitoring local-level effects. Sequence coverage maps of all peptides of TtASR1 and HvASR1 selected for HDX–MS analysis are displayed in Supplementary Fig. S8. Excellent sequence coverage was achieved in both cases. Digestion of TtASR1 gave rise to 33 unique peptides identified from their accurate masses and

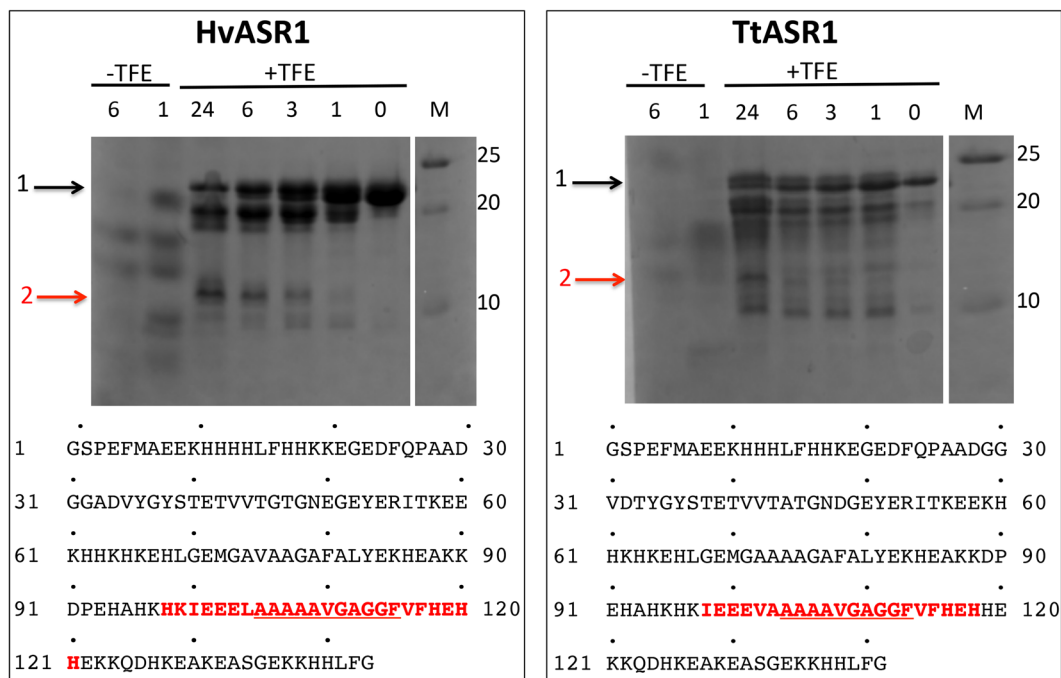


Figure 9. Fluorescence emission spectra of HvASR1 (A) and TtASR1 (B) at 20 °C. Proteins were at 1 μM in 10 mM sodium phosphate buffer pH7 supplemented with various concentrations of ZnSO₄.

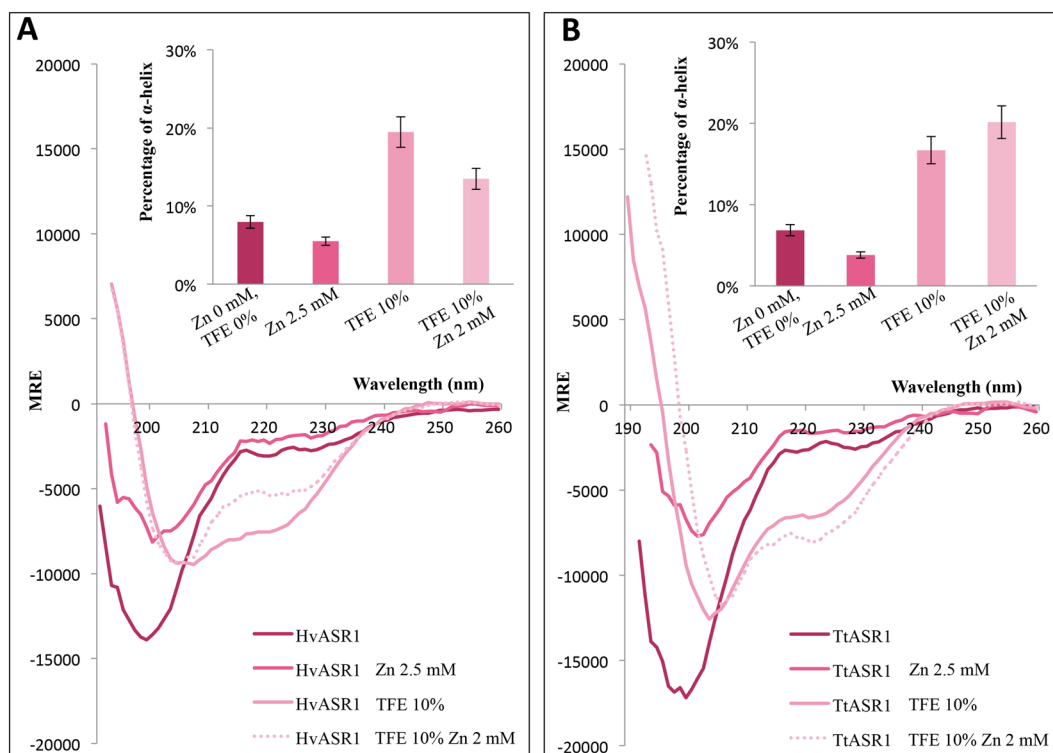


Figure 10. Far-UV CD spectra of HvASR1 (A) and TtASR1 (B) in the presence or absence of either 2.5 mM ZnSO₄ or 10% TFE or 10% TFE + 2 mM ZnSO₄ at 20 °C. Proteins were at 0.1 mg/mL in 10 mM sodium phosphate pH 7. Data is shown to the point at which the dyna voltage was in the permissible range. Data is representative of one out of two independent acquisitions. The insets show the α-helical content as derived as described in Methods (see Eq. 14). MRE: molar residue ellipticity.

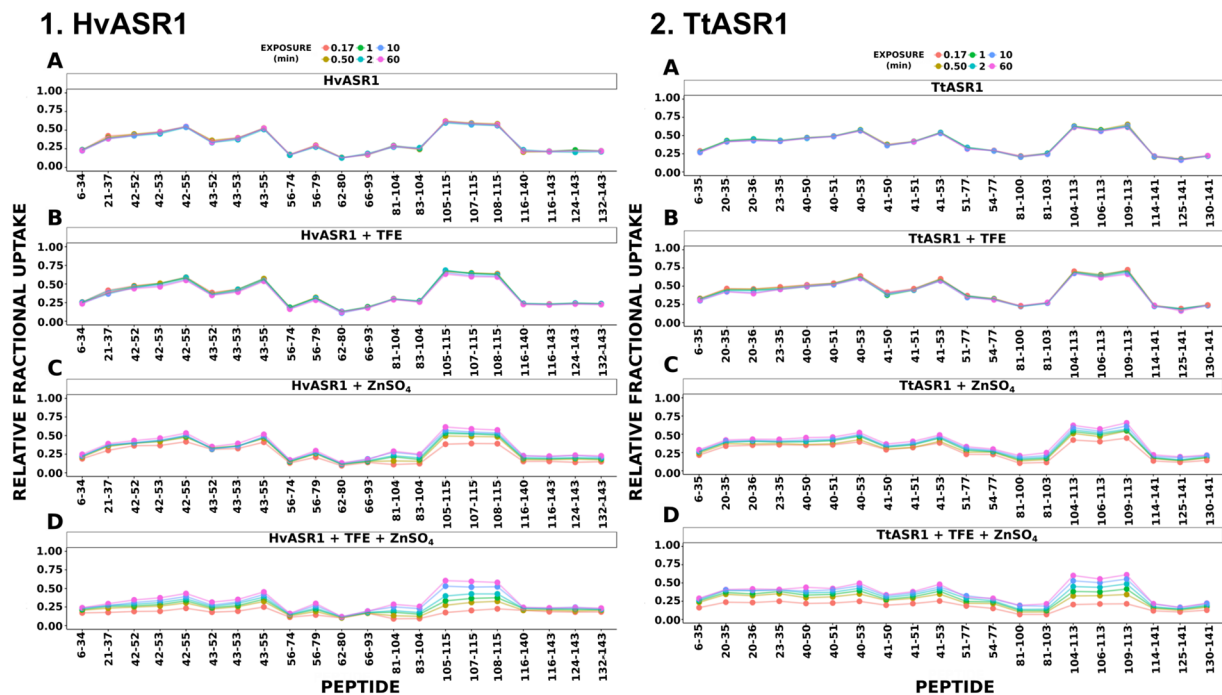


Figure 11. HDX-MS analysis of HvASR1 and TtASR1. The relative fractional uptake was plotted as a function of peptide position for the full-length ASR1 protein, where each color and dot represents a specific time point. The HDX-MS activity for the native proteins is displayed in panel A, in the presence of 10% TFE in Panel B, in the presence of 2 mM $ZnSO_4$ in Panel C, and in the presence of both 10% TFE and 2 mM $ZnSO_4$ in Panel D. Both ASR1 proteins are intrinsically disordered in their native state. Structure formation occurs throughout the proteins in the presence of zinc and TFE, the most dynamic and accessible region of which being located between residues 81 to 115 for HvASR1 and 104 to 113 for TtASR1.

product ion spectra. A total of 20 peptides were brought forward for HDX data analysis, corresponding to a sequence coverage of 92.2% (Supplementary Fig. S8). Similarly, for HvASR1, from a total of 29 identified peptides, 21 were selected for HDX-MS analysis, covering 93.7% of the total protein sequence (Supplementary Fig. S8). Amide hydrogen exchange rates within a protein are determined by both the solvent accessibility and structure of the protein. Those amide hydrogens which are fully solvent exposed and not participating in structural elements exchange rapidly, whilst those maintaining secondary structural elements exchange more slowly, due to hydrogen bonding⁶¹. Accordingly, amide hydrogens can be used as a structural probe to discriminate IDRs from those regions with inherent structural features⁶².

The HDX-MS behavior of HvASR1 was comparable to that of TtASR1, with each protein giving equivalent levels of deuterium uptake in each condition sampled (Fig. 11). In the first instance, HvASR1 and TtASR1 were found to be devoid of dynamic HDX-MS activity in the absence of either TFE or $ZnSO_4$ in the timescale of the experiment (Panels A, Fig. 11). Both proteins reached their maximum exchange after 10 sec of labeling, indicating that HvASR1 and TtASR1 are predominantly unfolded in their native state. For both proteins, the presence of 10% TFE resulted in no observable dynamic HDX behavior at the equilibrium (Panels B, Fig. 11). The addition of 2 mM $ZnSO_4$ induced a weak dynamic HDX-MS behavior, indicative of structure formation, throughout the entire polypeptide chain of both proteins (Panels C, Fig. 11). Notably, the addition of both 10% TFE and 2 mM $ZnSO_4$ in a combined fashion resulted in an more pronounced change in magnitude of HDX dynamics (Panels D, Fig. 11). Moreover, the region spanning residues 105–115 in HvASR1, and that spanning residues 104–113 in TtASR1, are the most dynamic elements, (Panels D, Fig. 11). Furthermore, this region is also the most accessible in terms of solvent exposure.

To summarize, both TtASR1 and HvASR1 proteins behave similarly by HDX-MS. In the absence of additives, both proteins are essentially unstructured. Each protein contains a segment that is more dynamic and accessible upon zinc and TFE binding, compared to the majority of flanking sequences.

Discussion

In the course of evolution, plants have evolved various mechanisms allowing them to adapt to abiotic stresses such as drought, low temperature, and high salinity. ID likely plays a crucial role in this ability to adapt to a varying environment. Indeed, the lack of a rigid 3D protein structure confers both binding promiscuity (ability to interact with multiple partners) and binding plasticity (ability to undergo binding-induced folding to accommodate diverse binding sites for different partners). These properties play an important role in cellular processes by conferring functional advantages in stress response, signaling and regulation²⁵. In line with this, analysis of the proteome of *A. thaliana* indicated that approximately 23% of its proteins are mostly disordered²⁴, with IDPs

being over-represented in functional categories such as signaling, development, cell cycle regulation and stress response⁸. Furthermore, plant proteins exhibit significantly higher degrees of ID compared to human proteins⁸. In spite of this compelling computational evidence, experimental knowledge of plant IDPs is lacking, with only few being well characterized²⁴. Dehydrins (e.g. ERD10 and ERD14) and proteins of the GRAS family are among the few plant IDPs that have been experimentally described^{63,64}. Dehydrins are proteins involved in global cell protection during the highly compact dry state characteristic of plant seeds, while proteins of the GRAS family play an important role in plant development and signal transduction cascades⁶⁴. Among the most intensively studied plant IDPs are LEA proteins⁸. The induction of expression under water deficit and/or osmotic stress are features shared by LEA and ASR proteins¹⁰. In addition, HvASR1 and TtASR1 can be considered as hydrophilins, as they have a high content in Gly (10%) and a high hydrophilicity (54%). LEA proteins comprise the largest group of hydrophilins. Accordingly, various studies have proposed to classify ASR proteins within this superfamily^{13,15,16}, but this remains debatable⁶⁵.

In an effort to provide additional insights into the ASR family, we have focused on the characterization of two ASR1 proteins encoded by genes that we have recently isolated. In the current study, we not only show that HvASR1 and TtASR1 are consistently predicted to be disordered by several predictors, but we also provide experimental evidence for their disordered state. In particular, we show that (i) they have an aberrant electrophoretic migration, which is a hallmark of protein disorder, (ii) they have Stokes radii, inferred from SEC that are close to the values expected for PMG conformations, (iii) they lack highly populated secondary and tertiary structure as inferred from both far- and near-CD, (iv) they lack hydrophobic cavities, as judged from DSF, (v) they are hypersensitive to proteolysis and (vi) they possess hydrodynamic parameters and SAXS scattering profiles typical of IDPs, and finally, (vii) they have no HDX-MS dynamic behavior.

Our conformational studies show that HvASR1 and TtASR1 have very similar hydrodynamic radii and radii of gyration. In agreement with this, and also with the high similarity in their charges, they have the same electrophoretic behavior in native PAGE (data not shown). The spectroscopic and hydrodynamic parameters of HvASR1 and TtASR1 indicate that, in spite of their overall extended nature, these proteins are not fully unfolded, but rather, conserve some residual compactness due to the presence of transiently populated secondary and/or tertiary structure. In addition, the R_g distribution of the selected sub-ensembles of HvASR1 and TtASR1 suggests the existence of two different populations characterized by a different average size. The two ASR1 proteins thus sample in solution two different subpopulations differing in their extent of compactness.

Previous studies have shown that desiccation triggers a conformational transition from an unfolded to an α -helical conformation in tomato ASR1 and MpASR proteins^{10,11}. Here, using CD studies in the presence of glycerol, we show that this property is also conserved in HvASR1 and TtASR1. This latter feature constitutes an additional point of commonality with LEA proteins^{52,53}. Addition of Zn^{2+} ions triggers a conformational transition from an unfolded to an α -helical conformation in tomato ASR1 and MpASR proteins^{10,11}. By contrast, in the case of soybean ASR protein (GmASR), Zn^{2+} was shown to induce (reversible) aggregation. The authors speculated that the expression of GmASR might be up regulated to buffer the concentration of Zn^{2+} , thus alleviating metal toxicity under stressed conditions⁶⁶. In the case of tomato ASR1, addition of Zn^{2+} ions not only triggers a gain of structure, but also induces homodimerization^{11,59}. Note that we could not assess possible Zn-induced homodimerization of ASR1 proteins, as the addition of $ZnSO_4$ in SEC systematically led to the precipitation of the protein onto the column.

While it is plausible that the Zn-induced structural transition observed in HvASR1 and TtASR1 plays a role in DNA-binding, as observed for tomato ASR1⁵⁹, it is also tempting to speculate that it can also serve to protect membranes against drought or freezing, by analogy with other LEA proteins that were shown to fold upon dehydration and to bind to membranes in their folded state⁶⁷. HDX-MS studies allowed us to precisely identify a region common to both proteins which is more dynamic and solvent exposed after addition of Zn^{2+} than the rest of the protein (see underlined fragment in Fig. 8). In the case of HvASR1, this region extends to include residues 81–104. Notably, this region contains an amino acid stretch (PEHAHKHK) that is also conserved in tomato ASR1, and that was previously shown to bind to zinc⁵⁹.

According to CD analyses, TtASR1 and HvASR1 undergo α -helical folding in the presence of TFE. This likely reflects their inherent propensity to undergo conformational change as part of their physiological function, as for instance, during ligand binding. Even if TFE is known to stabilize α -helices more than β -strands, it is worth emphasizing that gain of α -helicity in the presence of TFE is not a general rule and hence truly reflects the inherent structural propensities of the protein under study. For instance, (i) the acidic activator domain of GCN4 forms little or no α -helix in TFE concentrations as high as 30%, and folds mostly as β -sheets in 50% TFE⁶⁸, and (ii) the intrinsically disordered dehydrin Rab18 has an α -helical content as low as 2% in the presence of 90% TFE⁶⁹. To further support the argument that TFE does not promote non-native folding, the intrinsically disordered N_{TAIL} domains undergo α -helical folding in the presence of 20% TFE exclusively within a region known to undergo partner-induced α -helical folding^{70,71}. All of these considerations argue that TFE has the ability to increase the structural propensity of IDPs.

Many IDPs undergo (partial) folding upon binding to their partner(s)/ligand(s), with short, transiently structured regions, referred to as MoREs, being primarily involved in these events³². We show that HvASR1 and TtASR1 possess short regions locally enriched in hydrophobic clusters that correspond to predicted MoREs. The limited proteolysis experiments carried out in the presence of TFE allowed the identification of a resistant fragment in both ASR1 proteins. Interestingly, this fragment contains a predicted α -helix and a predicted MoRE, and also overlaps with the region that becomes more structured upon addition of Zn^{2+} . It therefore corresponds to a *bona fide* MoRE undergoing folding coupled to zinc binding.

Given the increasing interest that is being paid to IDPs, and the paucity of studies pertaining the structural characterization of plant IDPs in general, and of ASR proteins in particular, this work represents an important contribution that sheds light on fundamental aspects of plant IDPs. Unraveling the complex biology of

multifunctional ASR proteins is challenging. The present study, by providing the first experimental characterization of two ASR1 proteins and by unveiling their disordered nature, paves the way towards future studies that will reveal how ASR functional regulations and involvement in stress response is achieved *via* their flexibility. It also sets the basis for achieving atomic-resolution conformational ensemble description of these two plant IDPs. Finally, taking into account previous reports that established a relationship between structural disorder and protein interactivity, the present results suggest that HvASR1 and TtASR1 proteins are likely to be involved in manifold protein-protein and/or protein-DNA interactions.

Methods

Disorder predictions and in silico analysis of amino acid composition. Sequence accession numbers are KX660743 for HvASR1 and KX660744 for TtASR1. Disordered regions within HvASR1 and TtASR1 were identified using the MeDor (<http://www.vazymolo.org/MeDor/>)⁷², and the Genesilico MetaDisorder (<http://iimcb.genesilico.pl/metadisorder/metadisorder.html>)⁷³ metaservers for the prediction of disorder. These metaservers collect disorder and secondary structure predictions from servers available on the web. Beyond canonical predictors, MeDor also incorporates HCA⁷⁴. DorA (Disorder Analyser) is a predictor that identifies regions of disorder based on the combined use of a disorder scoring matrix and HCA. It works by first detecting regions that exhibit an amino acid composition bias towards disorder-promoting residues and then by filtering them to eliminate segments rich in hydrophobic clusters.

Low complexity regions were identified using SEG (<http://mendel.imp.ac.at/METHODS/seg.server.html>)⁷⁵, using trigger window length $[W] = 12$, trigger complexity $[K(1)] = 2.2$ and extension complexity $[K(2)] = 2.5$.

Disordered binding regions, i.e. regions with a propensity to undergo folding upon binding to a partner, were identified using ANCHOR (<http://anchor.enzim.hu/>)⁷⁶ and MoRFpred (<http://biomine-ws.ece.ualberta.ca/MoRFpred/index.html>)⁷⁷.

HvASR1 and TtASR1 were also submitted to charge/hydropathy analysis, a binary predictor allowing globular proteins to be distinguished from unstructured ones based on the ratio of their net charge (R) *versus* their hydropathy (H)¹⁸. A protein is predicted as disordered if $H < [(R + 1.151)/2.785]$. The RH-plot was generated by choosing this option on the main page of the PONDR server (<http://www.pondr.com/cgi-bin/PONDR/pondr.cgi>). Analysis of sequence attributes, as defined in³⁸, were carried out using the CIDER server (<http://pappulab.wustl.edu/CIDER/>). The letter was also used to generate the phase diagram plots.

Deviations in amino acid composition of HvASR1 and TtASR1 were analyzed and computed as already described⁷⁸ using the average amino acid frequencies of the SWISS-PROT database (as obtained from <http://us.expasy.org/spot>) as the reference value. The average amino acid frequencies of the SWISS-PROT database roughly corresponds to the mean composition of proteins in nature. If the average composition of an amino acid X in SWISS-PROT proteins is CSP_X , and CP_X is the composition of X within a protein P, deviation from the composition of X of SWISS-PROT proteins was defined for P as $(CP_X - CSP_X)/CSP_X$.

Proteins expression and purification. The full-length TtASR1 and HvASR1 cDNAs, encoding a protein of 136 and 138 residues respectively, were cloned in the pGEX4T-1 expression vector into the EcoRI site using standard restriction and ligation techniques. This vector allows the bacterial expression of the protein of interest as a fusion protein appended to a cleavable glutathione S transferase (GST) tag. Removal of the GST tag by thrombin results in a protein with an N-terminal, vector-encoded GSPEF amino acid extension. The sequence of the coding region was checked by sequencing performed in center of biotechnology of sfax (CBS) analysis service using the automated sequencer (ABI PRISM 3100 PRE) and found to conform to expectations.

The *E. coli* strain BL21 was used for the expression of the recombinant proteins. Cultures were grown overnight to saturation in LB medium containing $100 \mu\text{g mL}^{-1}$ ampicillin. An aliquot of the overnight culture was diluted 1/50 into 1 L of LB medium and grown at 37°C . When the optical density at 600 nm (OD_{600}) reached 0.5–0.8, isopropyl β -D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.5 mM, and the cells were grown at 37°C for 4 additional hours. The induced cells were harvested, washed and collected by centrifugation (4000 rpm, 20 min). The resulting pellets were frozen at -20°C .

Each bacterial pellet was resuspended in 25 mL of buffer A (50 mM sodium phosphate pH 7, 300 mM NaCl) supplemented with 0.1 mg mL^{-1} lysozyme, $10 \mu\text{g mL}^{-1}$ DNase I, MgSO_4 20 mM and half a tablet of EDTA-free protease inhibitor cocktail (Sigma). After 20 min of incubation with gentle agitation, the cells were disrupted by sonication (using a 750 W sonicator and 3 cycles of 30 s each at 45% power output). The lysates were clarified by centrifugation at 14000 rpm for 30 min at 4°C . The clarified supernatant was incubated with 1 mL of glutathione Sepharose 4B resin (GE, Healthcare), previously equilibrated with buffer A, for 1 h with gentle shaking at 4°C . Then, the resin was washed three times with buffer A. To remove the GST-tag, 1 mL of Phosphate Buffered Saline (PBS) pH 7.3 containing 5 U of bovine thrombin (Calbiochem) was added to the resin. After a 16 h incubation at 22°C , the non-retained fraction was recovered by pelleting the resin. To ensure optimal recovery of the protein of interest from the resin, the latter was washed with 2 mL of PBS, thus yielding a final volume of 3 mL. The presence of the protein was checked by SDS-PAGE. Each protein was further purified by SEC. To this end, the sample (typically 6 mL at 2.5 mg/mL) was loaded onto a Superdex 75 16/60 column (GE, Healthcare) using a fast protein liquid chromatography (FPLC) Äkta system (GE, Healthcare) and eluted in PBS pH 7.3.

Protein concentrations were calculated using the theoretical absorption coefficients at 280 nm as obtained using the program ProtParam at the EXPASY server. For both proteins, typical purification yields were of 0.5 mg of purified protein *per* liter of bacterial culture.

Size Exclusion Chromatography and calculation of hydrodynamic radii. The hydrodynamic radii (Stokes radii, R_s) of the HvASR1 and TtASR1 proteins were estimated by analytical SEC. Typically 2.6 mg mL⁻¹ of purified protein was injected. The SEC buffer was PBS pH 7.3.

The Stokes radii of proteins eluted from the SEC column were deduced from a calibration curve obtained using globular proteins of known molecular mass (MM, in Daltons) and whose R_s (in Å) was calculated according to⁷⁹:

$$\log(R_s^{\text{Obs}}) = 0.369 * (\log \text{MM}) - 0.254 \quad (1)$$

The R_s (in Å) of a natively folded (R_s^{NF}), fully unfolded state in urea (R_s^{U}) and natively unfolded premolten globule (PMG) (R_s^{PMG}) protein with a molecular mass (MM) (in Daltons) were calculated according to²⁰:

$$\log(R_s^{\text{NF}}) = 0.357 * (\log \text{MM}) - 0.204 \quad (2)$$

$$\log(R_s^{\text{U}}) = 0.521 * \log(\text{MM}) - 0.649 \quad (3)$$

$$\log(R_s^{\text{PMG}}) = 0.392 * (\log \text{MM}) - 0.210 \quad (4)$$

$$\log(R_s^{\text{MG}}) = 0.334 * (\log \text{MM}) - 0.053 \quad (5)$$

The R_s (in Å) of a natively folded dimeric form (R_s^{DimNF}), was calculated as:

$$\log(R_s^{\text{DimNF}}) = 0.357 * (\log \text{MM} * 2) - 0.204 \quad (6)$$

The R_s of an IDP with N residues was also calculated according to⁸⁰ using the simple power-law model:

$$R_s^{\text{IDP}} = R_0 N^\nu \quad (7)$$

where $R_0 = 2.49$ and $\nu = 0.509$. The compaction index (CI) is expressed as according to⁸¹:

$$\text{CI} = (R_s^{\text{U}} - R_s^{\text{obs}}) / (R_s^{\text{U}} - R_s^{\text{NF}}) \quad (8)$$

This parameter, which allows comparison between proteins of different lengths, can vary between 0 and 1, with 0 indicating minimal compaction and 1 maximal compaction.

Native ESI-MS analysis. We carried out native electrospray ionization mass spectrometry (ESI-MS) studies to decipher the macromolecular composition of TtARS1 and HvASR1, *i.e.* to assess whether they are monomers or dimers. As a control, we first measured ADH to confirm that the MS parameters used were suitable to maintain non-covalent complexes. Prior to analysis, all proteins were buffer-exchanged into 250 mM ammonium acetate, pH 8.0 using Micro Bio-SpinTM 6 Columns (Bio-Rad). All proteins were infused at a concentration of 15 μM. Native ESI-MS was performed on a Synapt G2-Si mass spectrometer (Waters). The capillary voltage was set to 1.8–2.1 kV, the source temperature was 30 °C, the sampling cone was set to 150 V, the source offset to 150 V, and the trap gas flow to 4 mL/min. The latter was adjusted to 5 mL/min for ADH.

Differential scanning fluorimetry (DSF). DSF monitors thermal unfolding of proteins in the presence of a fluorescent dye⁸². A solution of SYPRO Orange (5000× stock solution) was diluted in water to yield a 7× working solution. This experiment was conducted using a PCR instrument (Biorad) and 96-well plates containing 25 μL of mixture *per* well. Each well contained 21.5 μL of protein solution (HvASR1 and TtASR1 at 1 mg mL⁻¹) and 3.5 μL of SYPRO Orange working solution. Fluorescent signals were acquired with excitation and emission wavelengths at 485 nm and 625 nm, respectively. Temperature scans were performed from 20 °C to 90 °C.

SAXS measurements and calculation of the radius of gyration. All SAXS measurements were carried out at the European Synchrotron Radiation Facility (ESRF) on beamline BM29 at a working energy of 12.5 KeV. The sample-to-detector distance of the X-rays was 2.847 m, leading to scattering vectors q ranging from 0.028 to 4.525 nm⁻¹. The scattering vector is defined as $q = 4\pi/\lambda \sin\theta$, where 2θ is the scattering angle. The exposure time was optimized to reduce radiation damage.

SAXS data were collected at 20 °C using purified protein samples (30 μL each). Protein concentrations were as follows: 1.0 and 1.5 g/L for HvASR1 and 1.0 and 1.5, 2.0 and 3.0 g/L for TtASR1. Both proteins were in PBS pH 7.3 buffer containing 5 mM DTT.

Samples were loaded in a fully automated sample charger. Ten exposures of 10 s each were made for each protein concentration and data were combined to give the average scattering curve for each measurement. Any data points affected by aggregation, possibly induced by radiation damages were excluded. The profiles obtained in the range 1.0–1.5 g/L for HvASR1, and 1.0–3.0 g/L for TtASR1 had the same shape and were flat at low q values indicating the absence of significant aggregation. Then, we used the higher concentration (1.5 g/L for HvASR1 and 3.0 g/L for TtASR1) to obtain maximal information at high resolution.

The data were analyzed using the ATSAS program package⁸³. Data reductions were performed using the established procedure available at BM29, and buffer background runs were subtracted from sample runs. The R_g

and forward intensity at zero angle $I(0)$ were determined with the program PRIMUS⁸⁴ according to the Guinier approximation at low q values, in a $q \cdot R_g$ range up to 1.3:

$$\ln[I(Q)] = \ln[I_0] - \frac{Q^2 R_g^2}{3} \quad (9)$$

The forward scattering intensities were calibrated using bovine serum albumin as reference. The R_g and pair distance distribution function, $P(r)$, were calculated with the program GNOM⁸⁵. The maximum dimension (D_{\max}) value was adjusted such that the R_g value obtained from GNOM agreed with that obtained from the Guinier analysis.

A pool of random-coil conformers was generated using Flexible-Meccano⁸⁶. An optimized sub-ensemble of conformations that agrees with the experimental scattering curve was selected from the large conformational ensemble using GAJOE⁸⁷. The maximum size of the final optimized ensemble was set to 50.

The theoretical value of R_g (in Å) expected for an IDP was calculated using Flory's equation according to⁴⁶:

$$R_g = R_0 N^\nu \quad (10)$$

where N is the number of amino acid residues, R_0 is 2.54 ± 0.01 and ν is 0.522 ± 0.01 .

The theoretical value of R_g (in Å) expected for a globular protein was calculated according to⁴⁵:

$$R_g^{\text{NF}} = \sqrt{(3/5)4.75 N^{0.29}} \quad (11)$$

The theoretical value of R_g (in Å) expected for a fully unfolded form was calculated according to⁴⁵:

$$\text{Log}(R_g^{\text{U}}) = 0.58 \text{Log}(N) + 0.80 \quad (12)$$

The theoretical radius of gyration (R_g , in Å) expected for a globular protein with a hydrodynamic radius R_s was calculated according to⁴⁵:

$$R_g^{\text{NF}} = (3/5)^{1/2} R_s \quad (13)$$

HvASR1 and TtASR1 consist of 143 and 141 residues, respectively, including vector-encoded residues. Using an average volume of 134 \AA^3 per residue for proteins, the radius of a sphere with volume $V = 4/3 \pi R_s^3$ would be 16.6 \AA in the case of HvASR1, and 16.5 \AA in the case of TtASR1. According to Eq. 13, the corresponding R_g would be 12.8 \AA for both ASR1 proteins.

Circular dichroism (CD) measurements. CD spectra of HvASR1 and TtASR1 were measured using a Jasco 810 dichrograph, flushed with N_2 and equipped with a Peltier thermoregulation system. One-mm or 1-cm thick quartz cuvettes were used for far- and near-UV CD measurements, respectively. Proteins concentrations were 0.1 mg mL^{-1} and 1 mg mL^{-1} for far- and near-UV CD studies, respectively. Far-UV CD spectra were measured between 190 and 260 nm, while near-UV CD spectra were recorded between 250 and 350 nm. Unless differently specified, CD spectra were recorded in 10 mM sodium phosphate pH 7 at 20 °C. The scanning speed was 20 nm/min, with data pitch of 0.2 nm. Each spectrum is the average of three acquisitions. The spectrum of buffer was subtracted from the protein spectrum. Spectra were smoothed using the "means-movement" smoothing procedure implemented in the Spectra Manager package.

Far-UV CD spectra were also recorded in the presence of increasing concentrations (from 20 to 50%) of TFE. Mean molar ellipticity values per residue (MRE) were calculated as $(\Theta) = 3300M\Delta A/(lcn)$, where l is the path length in cm, n is the number of residues, M is the molecular mass in Daltons and c is the concentration of the protein in mg mL^{-1} . Numbers of amino acid residues are 141 for TtASR1 and 143 for HvASR1. Molecular masses are 15,645 Da for TtASR1 and 15,922 Da for HvASR1.

In order to study protein unfolding, measurements at a fixed wavelength of 230 nm were performed in the temperature range of 20 °C–80 °C, with data pitch 1 °C and a temperature slope of 1 °C/min and protein concentrations of 0.1 mg mL^{-1} .

The DICHROWEB website (<http://dichroweb.cryst.bbk.ac.uk/html/home.shtml>), which was supported by grants to the BBSRC Centre for Protein and Membrane Structure and Dynamics (CPMSD)⁸⁸, was used to analyze the experimental data in the 190–260 nm range. The content in the various types of secondary structure was estimated using the CDSSTR deconvolution method with the reference protein set⁷⁸⁹.

The α -helical content was estimated as follows:

For the estimation of the percentage of residues adopting an α -helical conformation in the presence of various additives, such as glycerol, TFE, sucrose and $ZnSO_4$, we analyzed the value of molar ellipticity at 220 nm observed under each condition and divided it by the value expected for a protein whose all residues adopt an α -helical conformation (100% α -helix). The latter was calculated according to the following empirical relationship⁹⁰:

$$(\Theta)_{220} = -39\,500 \times [1 - (2.57/n)] \quad (14)$$

where n is the number of residues.

Limited proteolysis by thermolysin. Limited proteolysis was used to identify possible folded fragments within the HvASR1 and TtASR1 proteins. The proteins (at 1 mg mL^{-1}) were incubated with thermolysin in 20 mM Tris/HCl pH 7.8 at 25 °C supplemented with 15% TFE. A thermolysin stock solution at $800 \mu\text{g mL}^{-1}$ was used in this experiment. Protease to protein substrate ratios were 1:100 (w/w). The extent of proteolysis was evaluated

by SDS–PAGE analysis of 20 μL aliquots removed from the reaction mixture over a time course (0, 1, 3, 6 and 24 hours), added to 5 μL of 5x loading sample buffer and boiled for 5 min to inactivate the protease. Proteins incubated with only thermolysin without TFE were used as controls.

Mass spectrometry (MALDI-TOF). Mass analysis of the purified ASR1 proteins was performed using a MALDI-TOF-TOF Bruker Ultraflex III spectrometer (Bruker Daltonics, Wissembourg, France) controlled by the Flexcontrol 3.0 package (Build 51). This instrument was used at a maximum accelerating potential of 25 kV and was operated in the linear mode with m/z range from 600 to 3500. Samples (1 μL containing 15 pmol) were mixed with an equal volume of α -Cyano-4-hydroxycinnamic acid matrix solution, spotted on the target, then dried at room temperature for 10 min.

Mass spectral analysis of the protein fragments, as obtained upon thermolysin limited digestion of the purified HvASR1 and TtASR1 proteins, was performed as follows. After SDS-PAGE separation, the bands of interest were excised from the gel. The bands were then digested with trypsin (0.25 μg trypsin per μg of protein substrate). For each protein band, mass analyses were performed on a MALDI-TOF-TOF Bruker Ultraflex III spectrometer as described above, except that the instrument was operated in reflector mode. The mass standards were either autolytic tryptic peptides used as internal standards or peptide standards (Bruker Daltonics). Following MS analysis, MS/MS analyses were performed on the most intense peaks to identify the amino acid sequence of the protein band.

Binding of HvASR1 and TtASR1 proteins to Zn and Ni. The ability of both HvASR1 and TtASR1 proteins to bind to zinc and nickel ions was assessed by incubating the purified proteins (25 μL at 1 mg/mL in PBS) with 60 μL of a sepharose fast-flow resin previously loaded with either ZnSO_4 or NiSO_4 (100 mM each) and equilibrated in PBS. After a 3 hour incubation with gentle agitation at 4 $^\circ\text{C}$, the resin was washed (4 times with 1 mL of PBS) and analyzed by SDS-PAGE by loading 12 μL .

Fluorescence spectroscopy. Fluorescence spectra of the Tyr residues in both ASR1 proteins were recorded by using a Cary Eclipse (Varian) equipped with a front face fluorescence accessory at 20 $^\circ\text{C}$, with 5 nm excitation and 5 nm emission bandwidths. The excitation wavelength was 280 nm, and the emission spectra were recorded between 280 and 400 nm. Spectra were recorded using a 1-ml quartz fluorescence cuvette containing 1 mg/mL of protein in PBS. ZnSO_4 was directly added to the protein solution to reach the desired concentration.

HDX-MS analyses. For each of the two ASR1 proteins, we monitored various conditions: A) ASR1 protein alone, B) ASR1 in the presence of 10% TFE, C) ASR1 in the presence of 2 mM ZnSO_4 and D) ASR1 in the presence of both 10% TFE and 2 mM ZnSO_4 . Prior to addition of the deuterated buffer, all solutions were equilibrated for 1 h at room temperature. Continuous labeling was performed at 20 $^\circ\text{C}$ for $t = 0.16, 0.5, 1, 2, 10,$ and 60 min. The labeling reaction for each experimental condition was performed using a deuterium solution supplemented with either TFE or ZnSO_4 , as above. The final deuterium concentration in each experiment was 72%. Aliquots of 10 pmol of protein were withdrawn at each experimental time point and quenched upon mixing of the deuterated sample with ice-cold 0.5% formic acid solution, achieving a final pH of 2.5. Quenched samples were immediately snap-frozen in liquid nitrogen and stored at -80 $^\circ\text{C}$ for approximately 24 h. Undeuterated controls were prepared using an identical procedure. Triplicate analyses were performed for each time point and condition for all HDX-MS analyses.

Prior to MS analysis, samples were rapidly thawed on ice. To minimize back exchange, the LC solvent line, injection valve, and sample loop were maintained at 0 $^\circ\text{C}$ with the aid of a cooled HDX Manager (Waters Corporation, Milford, MA). For peptide analysis, samples were initially digested with protease Type XIII from *Aspergillus saitoi* for 5 min on ice, followed by sequential digestion using an in-house prepared cartridge of immobilized pepsin beads (Thermo Scientific, Rockford, IL), for 2 min at 100 $\mu\text{L}/\text{min}$ and 20 $^\circ\text{C}$. Peptic peptides were rapidly desalted and concentrated using a Vanguard C4 pre-column (1.7 μm , 2.1 \times 5 mm; Waters), and separated using an ACQUITY UPLC™ BEH C18 column (1.7 μm , 1 \times 100 mm). ASR1 peptides were separated over a 10 min gradient of 5–40% ACN at 40 $\mu\text{L}/\text{min}$ and at 0 $^\circ\text{C}$. The LC flow was directed to a Synapt™ G2-Si HDMS™ mass spectrometer (Waters) that was equipped with ESI and lock-mass correction using Glu-Fibrinogen peptide. Mass spectra were acquired in positive-ion mode over the m/z range of 50–1800 using a data-independent acquisition scheme (MS^E) whereby exact mass information is collected at both low and high collisional energies for collisional induced dissociation. To enable local-level analysis, unlabeled TtASR1 and HvASR1 samples were digested in duplicate to build a peptide coverage map.

Peptide identification was *via* the Protein Lynx Global Server (PLGS) 3.0 (Waters). Oxidation of methionines and carbamylation of N-terminal and lysine residues were set as variable modifications. The sequence coverage map was plotted using DynamX 3.0 HDX software (Waters). D_2O uptake at the peptide level was extracted and visualized in uptake charts, difference plots, and heat maps, also performed in DynamX. HDX-MS results were further analyzed using MEMHDX⁹¹.

Data availability statement. All data herein presented are available upon request.

References

- Shen, G. *et al.* Molecular cloning, characterization and expression of a novel jasmonate-dependent defensin gene from *Ginkgo biloba*. *J Plant Physiol* **162**, 1160–1168, <https://doi.org/10.1016/j.jplph.2005.01.019> (2005).
- Riccardi, F., Gazeau, P., de Vienne, D. & Zivy, M. Protein changes in response to progressive water deficit in maize. Quantitative variation and polypeptide identification. *Plant Physiol* **117**, 1253–1263 (1998).
- Vaidyanathan, R., Kuruville, S. & Thomas, G. Characterization and expression pattern of an abscisic acid and osmotic stress responsive gene from rice. *Plant Science (Netherlands)* (1999).

4. Cakir, B. *et al.* A grape ASR protein involved in sugar and abscisic acid signaling. *Plant Cell* **15**, 2165–2180 (2003).
5. Jeanneau, M. *et al.* Improvement of drought tolerance in maize: towards the functional validation of the Zm-Asr1 gene and increase of water use efficiency by over-expressing C4-PEPC. *Biochimie* **84**, 1127–1135 (2002).
6. Kalifa, Y. *et al.* The water- and salt-stress-regulated Asr1 (abscisic acid stress ripening) gene encodes a zinc-dependent DNA-binding protein. *Biochem J* **381**, 373–378, <https://doi.org/10.1042/BJ20031800> (2004).
7. Liu, H. Y. *et al.* Characterization of a novel plantain Asr gene, MpAsr, that is regulated in response to infection of *Fusarium oxysporum* f. sp. cubense and abiotic stresses. *J Integr Plant Biol* **52**, 315–323, <https://doi.org/10.1111/j.1744-7909.2010.00912.x> (2010).
8. Marin, M. & Ott, T. Intrinsic disorder in plant proteins and phytopathogenic bacterial effectors. *Chem Rev* **114**, 6912–6932, <https://doi.org/10.1021/cr400488d> (2014).
9. Gonzalez, R. M. & Iusem, N. D. Twenty years of research on Asr (ABA-stress-ripening) genes and proteins. *Planta* **239**, 941–949, <https://doi.org/10.1007/s00425-014-2039-9> (2014).
10. Dai, J. R. *et al.* MpAsr encodes an intrinsically unstructured protein and enhances osmotic tolerance in transgenic *Arabidopsis*. *Plant Cell Rep* **30**, 1219–1230, <https://doi.org/10.1007/s00299-011-1030-1> (2011).
11. Goldgur, Y. *et al.* Desiccation and zinc binding induce transition of tomato abscisic acid stress ripening 1, a water stress- and salt stress-regulated plant-specific protein, from unfolded to folded state. *Plant Physiol* **143**, 617–628, <https://doi.org/10.1104/pp.106.092965> (2007).
12. Iusem, N. D., Bartholomew, D. M., Hitz, W. D. & Scolnik, P. A. Tomato (*Lycopersicon esculentum*) transcript induced by water deficit and ripening. *Plant Physiol* **102**, 1353–1354 (1993).
13. Battaglia, M., Olvera-Carrillo, Y., Garcarrubio, A., Campos, F. & Covarrubias, A. A. The enigmatic LEA proteins and other hydrophilins. *Plant Physiol* **148**, 6–24, <https://doi.org/10.1104/pp.108.120725> (2008).
14. Tunnacliffe, A. & Wise, M. J. The continuing conundrum of the LEA proteins. *Naturwissenschaften* **94**, 791–812 (2007).
15. Caramelo, J. J. & Iusem, N. D. When cells lose water: Lessons from biophysics and molecular biology. *Prog Biophys Mol Biol* **99**, 1–6, <https://doi.org/10.1016/j.pbiomolbio.2008.10.001> (2009).
16. Hunault, G. & Jaspard, E. LEAPdb: a database for the late embryogenesis abundant proteins. *BMC Genomics* **11**, 221, <https://doi.org/10.1186/1471-2164-11-221> (2010).
17. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* **293**, 321–331, <https://doi.org/10.1006/jmbi.1999.3110> (1999).
18. Uversky, V. N., Gillespie, J. R. & Fink, A. L. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* **41**, 415–427 (2000).
19. Romero, P. *et al.* Sequence complexity of disordered protein. *Proteins* **42**, 38–48 (2001).
20. Uversky, V. N. What does it mean to be natively unfolded? *Eur J Biochem* **269**, 2–12 (2002).
21. Tompa, P. Intrinsically unstructured proteins. *Trends Biochem Sci* **27**, 527–533 (2002).
22. Uversky, V. N. & Dunker, A. K. Multiparametric analysis of intrinsically disordered proteins: looking at intrinsic disorder through compound eyes. *Analytical chemistry* **84**, 2096–2104 (2012).
23. Habchi, J., Tompa, P., Longhi, S. & Uversky, V. N. Introducing protein intrinsic disorder. *Chem Rev* **114**, 6561–6588, <https://doi.org/10.1021/cr400514h> (2014).
24. Kragelund, B. B., Jensen, M. K. & Skriver, K. Order by disorder in plant signaling. *Trends Plant Sci* **17**, 625–632, <https://doi.org/10.1016/j.tplants.2012.06.010> (2012).
25. Sun, X., Rikkerink, E. H., Jones, W. T. & Uversky, V. N. Multifarious roles of intrinsic disorder in proteins illustrate its broad impact on plant biology. *Plant Cell* **25**, 38–55, <https://doi.org/10.1105/tpc.112.106062> (2013).
26. Pietrosemoli, N., García-Martín, J. A., Solano, R. & Pazos, F. Genome-wide analysis of protein disorder in *Arabidopsis thaliana*: implications for plant environmental adaptation. *PLoS One* **8**, e55524 (2013).
27. Monastyrskyy, B., Fidelis, K., Moullet, J., Tramontano, A. & Kryshchuk, A. Evaluation of disorder predictions in CASP9. *Proteins* **79**(Suppl 10), 107–118, <https://doi.org/10.1002/prot.23161> (2011).
28. Ferron, F., Longhi, S., Canard, B. & Karlin, D. A practical overview of protein disorder prediction methods. *Proteins-Structure Function and Bioinformatics* **65**, 1–14, <https://doi.org/10.1002/prot.21075> (2006).
29. Longhi, S., Lieutaud, P. & Canard, B. Conformational disorder. *Methods in molecular biology* **609**, 307–325, https://doi.org/10.1007/978-1-60327-241-4_18 (2010).
30. Lieutaud, P., Ferron, F., Habchi, J., Canard, B. & Longhi, S. Predicting protein disorder and induced folding: a practical approach (2013).
31. Lieutaud, P., Ferron, F. & Longhi, S. Predicting Conformational Disorder. *Methods in molecular biology* **1415**, 265–299, https://doi.org/10.1007/978-1-4939-3572-7_14 (2016).
32. Mohan, A. *et al.* Analysis of molecular recognition features (MoRFs). *J Mol Biol* **362**, 1043–1059, <https://doi.org/10.1016/j.jmb.2006.07.087> (2006).
33. Dunker, A. K. *et al.* Intrinsically disordered protein. *J Mol Graph Model* **19**, 26–59 (2001).
34. Campen, A. *et al.* TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept Lett* **15**, 956–963 (2008).
35. Das, R. K., Ruff, K. M. & Pappu, R. V. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr Opin Struct Biol* **32**, 102–112, <https://doi.org/10.1016/j.sbi.2015.03.008> (2015).
36. Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L. & Pappu, R. V. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc Natl Acad Sci USA* **107**, 8183–8188, <https://doi.org/10.1073/pnas.0911107107> (2010).
37. Müller-Späh, S. *et al.* Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences* **107**, 14609–14614 (2010).
38. Das, R. K. & Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences* **110**, 13392–13397 (2013).
39. Blocquel, D., Habchi, J., Gruet, A., Blangy, S. & Longhi, S. Compaction and binding properties of the intrinsically disordered C-terminal domain of Henipavirus nucleoprotein as unveiled by deletion studies. *Mol Biosyst* **8**, 392–410, <https://doi.org/10.1039/c1mb05401e> (2012).
40. Uversky, V. N. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* **11**, 739–756, <https://doi.org/10.1110/ps.4210102> (2002).
41. Raj, S. B., Ramaswamy, S. & Plapp, B. V. Yeast alcohol dehydrogenase structure and catalysis. *Biochemistry* **53**, 5791–5803, <https://doi.org/10.1021/bi5006442> (2014).
42. Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N. & Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical biochemistry* **357**, 289–298, <https://doi.org/10.1016/j.ab.2006.07.027> (2006).
43. Receveur-Brechot, V. & Durand, D. How random are intrinsically disordered proteins? A small angle scattering perspective. *Current protein & peptide science* **13**, 55–75 (2012).
44. Bernado, P. & Svergun, D. I. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol Biosyst* **8**, 151–167, <https://doi.org/10.1039/c1mb05275f> (2012).
45. Wilkins, D. K. *et al.* Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **38**, 16424–16431 (1999).

46. Bernado, P. & Blackledge, M. A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. *Biophysical journal* **97**, 2839–2845, <https://doi.org/10.1016/j.bpj.2009.08.044> (2009).
47. Porod, G. *Small-angle X-ray scattering* (London Academic Press, 1982).
48. Brocca, S. *et al.* Order propensity of an intrinsically disordered protein, the cyclin-dependent-kinase inhibitor Sic1. *Proteins* **76**, 731–746, <https://doi.org/10.1002/prot.22385> (2009).
49. Kumar, N. *et al.* Intrinsically disordered protein from a pathogenic mesophile *Mycobacterium tuberculosis* adopts structured conformation at high temperature. *Proteins* **71**, 1123–1133, <https://doi.org/10.1002/prot.21798> (2008).
50. Kelly, S. M. & Price, N. C. The use of circular dichroism in the investigation of protein structure and function. *Current protein and peptide science* **1**, 349–384 (2000).
51. Woody, R. W. Circular dichroism of intrinsically disordered proteins. *Instrumental analysis of intrinsically disordered proteins: Assessing structure and conformation* 303–321 (2010).
52. Hinch, D. K. & Thalhammer, A. LEA proteins: IDPs with versatile functions in cellular dehydration tolerance. *Biochem Soc Trans* **40**, 1000–1003, <https://doi.org/10.1042/BST20120109> (2012).
53. Navarro-Retamal, C. *et al.* Molecular dynamics simulations and CD spectroscopy reveal hydration-induced unfolding of the intrinsically disordered LEA proteins COR15A and COR15B from *Arabidopsis thaliana*. *Phys Chem Chem Phys* **18**, 25806–25816, <https://doi.org/10.1039/c6cp02272c> (2016).
54. Tell, G. *et al.* Structural and functional properties of the N transcriptional activation domain of thyroid transcription factor-1: similarities with the acidic activation domains. *Biochem J* **329**(Pt 2), 395–403 (1998).
55. Hua, Q. X., Jia, W. H., Bullock, B. P., Habener, J. F. & Weiss, M. A. Transcriptional activator-coactivator recognition: nascent folding of a kinase-inducible transactivation domain predicts its structure on coactivator binding. *Biochemistry* **37**, 5858–5866, <https://doi.org/10.1021/bi9800808> (1998).
56. Dahlman-Wright, K. & McEwan, I. J. Structural studies of mutant glucocorticoid receptor transactivation domains establish a link between transactivation activity *in vivo* and alpha-helix-forming potential *in vitro*. *Biochemistry* **35**, 1323–1327, <https://doi.org/10.1021/bi952409k> (1996).
57. Fontana, A. *et al.* Probing protein structure by limited proteolysis. *Acta Biochim Pol* **51**, 299–321, 035001299 (2004).
58. Receveur-Brechot, V., Bourhis, J. M., Uversky, V. N., Canard, B. & Longhi, S. Assessing protein disorder and induced folding. *Proteins: Structure, Function and Bioinformatics* **62**, 24–45 (2006).
59. Rom, S. *et al.* Mapping the DNA- and zinc-binding domains of ASR1 (abscisic acid stress ripening), an abiotic-stress regulated plant specific protein. *Biochimie* **88**, 621–628, <https://doi.org/10.1016/j.biochi.2005.11.008> (2006).
60. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim Biophys Acta* **1751**, 119–139, <https://doi.org/10.1016/j.bbapap.2005.06.005> (2005).
61. Wales, T. E. & Engen, J. R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom Rev* **25**, 158–170, <https://doi.org/10.1002/mas.20064> (2006).
62. O'Brien, D. P. *et al.* Structural models of intrinsically disordered and calcium-bound folded states of a protein adapted for secretion. *Sci Rep* **5**, 14223, <https://doi.org/10.1038/srep14223> (2015).
63. Kovacs, D., Kalmar, E., Torok, Z. & Tompa, P. Chaperone activity of ERD10 and ERD14, two disordered stress-related plant proteins. *Plant physiology* **147**, 381–390 (2008).
64. Pazos, F., Pietrosemoli, N., Garcia-Martin, J. A. & Solano, R. Protein intrinsic disorder in plants. *Front Plant Sci* **4**, 363, <https://doi.org/10.3389/fpls.2013.00363> (2013).
65. Jaspard, E., Macherel, D. & Hunault, G. Computational and statistical analyses of amino acid usage and physico-chemical properties of the twelve late embryogenesis abundant protein classes. *PLoS One* **7**, e36968, <https://doi.org/10.1371/journal.pone.0036968> (2012).
66. Li, R.-H., Liu, G.-B., Wang, H. & Zheng, Y.-Z. Effects of Fe³⁺ and Zn²⁺ on the structural and thermodynamic properties of a soybean ASR protein. *Bioscience, biotechnology, and biochemistry* **77**, 475–481 (2013).
67. Thalhammer, A., Hundertmark, M., Popova, A. V., Seckler, R. & Hinch, D. K. Interaction of two intrinsically disordered plant stress proteins (COR15A and COR15B) with lipid membranes in the dry state. *Biochim Biophys Acta* **1798**, 1812–1820, <https://doi.org/10.1016/j.bbamem.2010.05.015> (2010).
68. Van Hoy, M., Leather, K. K., Kodadek, T. & Johnston, S. A. The acidic activation domains of the GCN4 and GAL4 proteins are not alpha helical but form beta sheets. *Cell* **72**, 587–594 (1993).
69. Mouillon, J.-M., Gustafsson, P. & Harryson, P. Structural investigation of disordered stress proteins. Comparison of full-length dehydrins with isolated peptides of their conserved segments. *Plant Physiology* **141**, 638–650 (2006).
70. Belle, V. *et al.* Mapping α -helical induced folding within the intrinsically disordered C-terminal domain of the measles virus nucleoprotein by site-directed spin-labeling EPR spectroscopy. *Proteins: Structure, Function, and Bioinformatics* **73**, 973–988 (2008).
71. Martinho, M. *et al.* Assessing induced folding within the intrinsically disordered C-terminal domain of the Henipavirus nucleoproteins by site-directed spin labeling EPR spectroscopy. *Journal of Biomolecular Structure and Dynamics* **31**, 453–471 (2013).
72. Lieutaud, P., Canard, B. & Longhi, S. MeDor: a metasever for predicting protein disorder. *BMC Genomics* **9**(Suppl 2), S25, <https://doi.org/10.1186/1471-2164-9-S2-S25> (2008).
73. Kozłowski, L. P. & Bujnicki, J. M. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics* **13**, 111, <https://doi.org/10.1186/1471-2105-13-111> (2012).
74. Callebaut, I. *et al.* Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cell Mol Life Sci* **53**, 621–645 (1997).
75. Wootton, J. C. Non-globular domains in protein sequences: automated segmentation using complexity measures. *Comput Chem* **18**, 269–285 (1994).
76. Dosztanyi, Z., Meszaros, B. & Simon, I. ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* **25**, 2745–2746, <https://doi.org/10.1093/bioinformatics/btp518> (2009).
77. Disfani, F. M. *et al.* MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics* **28**, i75–83, <https://doi.org/10.1093/bioinformatics/bts209> (2012).
78. Habchi, J., Mamelli, L., Darbon, H. & Longhi, S. Structural disorder within Henipavirus nucleoprotein and phosphoprotein: from predictions to experimental assessment. *PLoS One* **5**, e11684, <https://doi.org/10.1371/journal.pone.0011684> (2010).
79. Uversky, V. N. Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry* **32**, 13288–13298 (1993).
80. Marsh, J. A. & Forman-Kay, J. D. Sequence determinants of compaction in intrinsically disordered proteins. *Biophysical journal* **98**, 2383–2390, <https://doi.org/10.1016/j.bpj.2010.02.006> (2010).
81. Brocca, S. *et al.* Compaction properties of an intrinsically disordered protein: Sic1 and its kinase-inhibitor domain. *Biophysical journal* **100**, 2243–2252, <https://doi.org/10.1016/j.bpj.2011.02.055> (2011).
82. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2**, 2212–2221, <https://doi.org/10.1038/nprot.2007.321> (2007).
83. Petoukhov, M. V. *et al.* New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Cryst* **45**, 342–350 (2012).
84. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Cryst* **36**, 1277–1282 (2003).

85. Svergun, D. Determination of the regularization parameters in indirect-trasform methods using perceptual criteria. *J. Appl. Cryst.* **25**, 495–503 (1992).
86. Ozenne, V. *et al.* Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* **28**, 1463–1470, <https://doi.org/10.1093/bioinformatics/bts172> (2012).
87. Tria, G., Mertens, H. D. T., Kachala, M. & Svergun, D. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCr* **2**, 202–217 (2015).
88. Whitmore, L. & Wallace, B. A. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* **32**, W668–673, <https://doi.org/10.1093/nar/gkh371> (2004).
89. Sreerama, N. & Woody, R. W. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Analytical biochemistry* **287**, 252–260, <https://doi.org/10.1006/abio.2000.4880> (2000).
90. Chemes, L. B., Alonso, L. G., Noval, M. G. & de Prat-Gay, G. Circular dichroism techniques for the analysis of intrinsically disordered proteins and domains. *Methods in molecular biology* **895**, 387–404, https://doi.org/10.1007/978-1-61779-927-3_22 (2012).
91. Hourdel, V. *et al.* MEMHDX: an interactive tool to expedite the statistical validation and visualization of large HDX-MS datasets. *Bioinformatics* **32**, 3413–3419, <https://doi.org/10.1093/bioinformatics/btw420> (2016).

Acknowledgements

This work was supported jointly by French and Tunisian funding/research agencies. On the French side, this work was carried out with the financial support of the CNRS to S.L. E.S. is supported by a joint doctoral fellowship from the Direction Générale de l'Armement (DGA) and Aix-Marseille University. D.O.B., S.B. and A.C. are supported by the CNRS, CACSICE (Equipex ANR-11-EQPX-0008), Fondation Recherche Médicale (FRM DBS20140930771) and Institut Pasteur (PasteurInnov2015-197 and PTR451). On the Tunisian side, this work is supported by L'Oreal-Unesco for women in Science program-(Pan Arab Fellowship 2013), the Tunisian Ministry of Higher Education and Scientific Research and AUF (Collège en Biotechnologies végétales et agroalimentaire). We thank Patrick Fourquet, from the mass spectrometry platform of the Centre de Recherche en Cancérologie de Marseille (CRCM) for mass spectrometry analyses. We thank Julia Chamot-Rooke for providing access to the HDX-MS platform at Institut Pasteur, Paris. We also thank Julien Perard (ESRF) for its help in SAXS data collection, and the ESRF synchrotron for beamtime allocation. We are also grateful to Gerlind Sulzenbacher (AFMB lab) for efficiently managing the AFMB BAG.

Author Contributions

S.L. and I.Y. conceived and planned the experiments. K.H. and E.S. performed all experiments, except for those of HDX-MS and ESI-MS. The latter experiments were conceived by D.O.B., S.B. and A.C., D.O.B. and S.B. performed the HDX-MS and ESI-MS data acquisition, processing and interpretation. ES purified the samples for SAXS analysis and performed all the SAXS measurements. All the authors analyzed the data. K.H. and S.L. wrote the paper, but all the authors contributed to the writing. E.S. is the author of Figure 4 and S5.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-017-15299-4>.

Competing Interests: The authors declare that they have no competing interests.

Accession numbers: The nucleotide sequences of the HvASR1 and TtASR1 genes have been deposited within the Gene database under accession numbers KX660743 and KX660744, respectively.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017