Intrinsically disordered N-terminal domain of the *Helicoverpa armigera* Ultraspiracle stabilizes the dimeric form via a *scorpion-like* structure

Krzysztof Wycisk[a,*], Aneta Tarczewska[a], Magdalena Kaus-Drobek[b], Michał Dadlez[b], Rafał Hołubowicz[a], Zbigniew Pietras[b], Andrzej Dziembowski[b], Michał Taube[c], Maciej Kozak[c], Marek Orłowski[a], Andrzej Ożyhar[a,*]

[a] Department of Biochemistry, Faculty of Chemistry, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
[b] Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawińskiego 5A, 02-106 Warsaw, Poland
[c] Department of Macromolecular Physics, Faculty of Physics, Adam Mickiewicz University, Umultowska 85, 61-614 Poznań, Poland

ABSTRACT

Nuclear receptors (NRs) are a family of ligand-dependent transcription factors activated by lipophilic compounds. NRs share a common structure comprising three domains: a variable N-terminal domain (NTD), a highly conserved globular DNA-binding domain and a ligand-binding domain. There are numerous papers describing the molecular details of the latter two globular domains. However, very little is known about the structure-function relationship of the NTD, especially as an intrinsically disordered fragment of NRs that may influence the molecular properties and, in turn, the function of globular domains. Here, we investigated whether and how an intrinsically disordered NTD consisting of 58 amino acid residues affects the functions of the globular domains of the Ultraspiracle protein from *Helicoverpa armigera* (HaUsp). The role of the NTD was examined for two well-known and easily testable NR functions, i.e., interactions with specific DNA sequences and dimerization. Electrophoretic mobility shift assays showed that the intrinsically disordered NTD influences the interaction of HaUsp with specific DNA sequences, apparently by destabilization of HaUsp-DNA complexes. On the other hand, multi-angle light scattering and sedimentation velocity analytical ultracentrifugation revealed that the NTD acts as a structural element that stabilizes HaUsp homodimers. Molecular models based on small-angle X-ray scattering indicate that the intrinsically disordered NTD may exert its effects on the tested HaUsp functions by forming an unexpected *scorpion-like* structure, in which the NTD bends towards the ligand-binding domain in each subunit of the HaUsp homodimer. This structure may be crucial for specific NTD-dependent regulation of the functions of globular domains in NRs.

1. Introduction

Nuclear receptors (NRs) are a family of transcription factors that bind and respond to certain small lipophilic compounds, e.g., steroid hormones. Despite the many differences in the amino acid sequences of different NRs, the overall molecular structures of NRs are largely similar. Most of characterized NRs share a common structural organization consisting of at least four distinct functional domains [1]. There are two highly conserved well-folded domains of well-known structures as well as variable, often disordered domains of different lengths and amino acid compositions. The intrinsically disordered domains seem to

be dominant regulators of NR activation and activity in general. The first evolutionarily conserved domain is a DNA-binding domain (DBD) [2]. This domain contains two zinc-finger modules that allow the DBD to bind to specific response elements (REs) [3]. The second conserved domain is a ligand-binding domain (LBD), which is responsible not only for recognition of specific ligands but also for interactions with modulators, such as coactivators and corepressors, and this domain participates in homo- and heterodimerization of NRs [4]. The overall LBD fold comprises up to 12 α helices and often one β turn [5]. As has been shown for the RXR-α LBD, three of the central helices form a ligand-binding pocket (LBP), whereas the last helix is mobile. After the ligand

binds in the LBP, the movement of the last helix leads to the formation of activation function 2 (AF-2), which in turn provokes the dissociation of corepressors and association of coactivators [5]. An intrinsically disordered region (IDR) D is localized between the DBDs and LBDs. This region serves as a flexible hinge connecting these two folded domains, and in some NRs, this region contains a nuclear localization signal [4]. One of the most puzzling domains of NRs, in addition to the F domain, is the N-terminal domain (NTD). There are numerous papers describing the molecular properties of the DBD and LBD [2,5–8]. However, very little is known about the structure-function relationship of the NTD. The structures of NTDs are often predicted to be highly disordered [9,10], and in general, substantial differences in the sequences and lengths of the polypeptide chains of various NRs and various isoforms of the same NRs are characteristic features of NTDs [4,11]. The variability in the length and structural flexibility of the NTD are important for temporal interactions with either NR coregulators or multi-protein assemblies [12]. NTDs are known to interact with the C termini of NRs when the NTD possesses an FxxLF motif, which is similar to the LxxLL motif of coactivators interacting with AF-2 [13]. However, an in-depth study of the mineralocorticoid receptor (MR) revealed that the NTD of this receptor does not contain an FxxLF or LxxLL motif, and yet, N/C interaction is observed in the presence of a ligand [14]. As a result of alternative splicing, different isoforms of NRs are generated, which are often characterized by different spatial and temporal distributions within various cells. For example, alternative splicing gives rise to several isoforms of the glucocorticoid receptor (GR), which are characterized by different activities in a cell-dependent manner [15]. However, these studies were conducted on vertebrate NRs; when we consider insects, which are usually less complex organisms and are therefore studied as model organisms, the scope of knowledge regarding NTDs in these organisms is even smaller. This lack of knowledge might be due to the lengths of invertebrate NTDs, which are usually shorter and are therefore often difficult to study. Nevertheless, some observations have been made, for example, complex analyzes of ecdysone receptor (EcR) isoforms revealed that they have different NTDs, resulting in different spatial and temporal distributions of these isoforms within various tissues, which in turn leads to different developmental functions. More precisely, it has been suggested that the NTD of isoform A is a weaker transactivator than that of isoform B1 of the *Drosophila melanogaster* EcR [16]. A study of the isolated NTD of the Ultraspiracle (Usp) NR from *Aedes aegypti* clearly showed that this domain is intrinsically disordered, and this study also suggested an important role for the NTD in the dimerization process of this receptor [17].

Usp is a member of the NR family and is homologous to the mammalian retinoid X receptor (RXR) [18]. Usp is involved in the regulation of essential developmental processes, such as metamorphosis or molting, in arthropods [19]. Similar to RXR, which interacts with a variety of other proteins, the interaction of Usp with different proteins has also been demonstrated, e.g., with hormone receptor 38 (HR38) or the Seven-up protein [20,21]. However, the most prominent partner of Usp is EcR. In response to high concentration of 20-hydroxyecdysone (20E), these two NRs form a functional heterodimeric complex and activate the transcription of crucial genes involved in metamorphosis [22]. Various isoforms of Usp differ mainly in the NTD sequence, which often leads to specific differences in the functions of these isoforms, as was shown for the Usp from *Bombyx mori* or *A. aegypti* [23,24]. It seems to be essential to understand the role of NTD as a factor that defines in a specific way the structures of Usp isoforms, leading to differences in the functions of the isoforms. Theoretically, this role may be implemented in two ways. First, in what seems to be an obvious manner, the NTD might exhibit extrinsic functionality, for example, by directly interacting with regulatory proteins. Second, in a less obvious manner, intrinsically disordered NTDs may influence the molecular properties and in turn the functions of the remaining Usp domains that possess well-defined, stable tertiary structures. We decided to experimentally

investigate this second possibility by testing the potential role of the NTD in Usp dimer formation and in the interaction of Usp with specific REs. We chose to carry out our experiments using Usp isoform 1 from *Helicoverpa armigera* (HaUsp) (ACD74808.1), which along with *D. melanogaster, A. aegypti* and *B. mori* serves as a good model to study arthropod NRs. First, we obtained homogenous samples of full-length HaUsp and HaUsp lacking the NTD (HaUsp_ΔNTD), and subsequently, we investigated the dimerization capability of both forms. Numerous analyses with techniques such as analytical ultracentrifugation (AUC), size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) or small-angle X-ray scattering (SAXS) have shown undisputedly that an intrinsically disordered NTD is indispensable for effective dimerization of HaUsp. Using electrophoretic mobility shift assay (EMSA), we also showed that the protein-DNA interaction patterns of full-length HaUsp and HaUsp_ΔNTD differ significantly. Moreover, the specific complexes formed by HaUsp seem to be less stable than the complexes formed by HaUsp_ΔNTD under the conditions used for EMSA. Thus, the intrinsically disordered NTD, composed of 58 residues, significantly influences the DNA-binding properties of HaUsp. Molecular models reconstructed from SAXS data indicate that HaUsp dimers are most likely stabilized by *scorpion-like* structures of NTDs, which bend towards the respective LBDs. This proximity of NTDs and LBDs in *scorpion-like* structures may result in the interaction of these domains, possibly leading to rearrangement of the LBD structure and in turn to the ability of full-length HaUsp to form dimers with increased stability. Apparently, in the *scorpion-like* structure, some changes occur also in the DBD, which in turn results in effects on the interaction mode of the protein with specific DNA sequences. Thus, the intrinsically disordered NTD is an indispensable element that controls the molecular properties of HaUsp.

2. Materials and methods

2.1. Buffer compositions

All buffers were prepared at room temperature. The lysis buffer contained 20 mM Na_2HPO_4 and 150 mM NaCl (pH 7.0). Buffer A contained 50 mM Na_2HPO_4, 300 mM NaCl, 5 mM imidazole, 5% (v/v) glycerol, and 1 mM βmercaptoethanol (pH 7.0). Buffer B contained 50 mM Na_2HPO_4, 300 mM NaCl, 200 mM imidazole, 5% (v/v) glycerol, and 1 mM βmercaptoethanol (pH 7.0). Buffer C contained 50 mM Na_2HPO_4 and 150 mM NaCl (pH 7.0). Buffer D contained 50 mM Na_2HPO_4, 100 mM NaCl, 10% (v/v) glycerol, 0.5 mg/ml ovalbumin, and 1 mM β-mercaptoethanol (pH 7.8).

2.2. Construction of expression vectors

A cDNA clone encoding the full-length HaUsp (isoform 1) (EU526832.1) was obtained from GeneArt Life Technologies. The ordered sequence was previously optimized for *Escherichia coli* and flanked with restriction sites: *Bam*HI at the beginning and *Hind*III at the end. The obtained plasmid containing the DNA sequence was double-digested with *Bam*HI and *Hind*III and ligated into the corresponding sites of the pQE-80 l (Qiagen) vector. The same cDNA was used to obtain HaUsp lacking the 58 amino acids of the NTD (HaUsp_ΔNTD). First, by PCR, the correct insert was obtained. The reaction was prepared according to the manufacturer's instructions for Phusion polymerase (Thermo Fisher Scientific), and the primers are listed in Table 1. Then, the PCR product was double-digested with *Bam*HI and *Hind*III and ligated into the corresponding sites of the pQE-80 l vector. Sequences of the purified constructs were verified by DNA sequencing.

2.3. Expression and purification procedures

For efficient expression, 200 ml of LB medium containing antibiotics (35 µg/ml chloramphenicol and 50 µg/ml carbenicillin) was inoculated

Table 1
Sequences of primers used for PCR. Red-colored sequences correspond to restriction sites: *Bam*HI for the forward primer (HaUsp_ΔNTD_F) and *Hind*III for the reverse primer (HaUsp_ΔNTD_R). Lower-case sequences correspond to complementary sequences for both primers.

Primer	Sequence
HaUsp_ΔNTD_F	5' – GCC CGG GGA TCC cat ctg tgt agc att tgt gg – 3'
HaUsp_ΔNTD_R	5' – GCC CGG AAG CTT cat cat att ggt atc aat agg – 3'

with *E.coli* strain BL21(DE3)pLysS (Novagen) transformed with HaUsp or HaUsp_ΔNTD (containing a $6 \times$ His tag on the N-terminus). After an overnight incubation at 37 °C and 182 rpm, the starter culture was used to inoculate 4 l of TB medium with antibiotics (35 μg/ml chloramphenicol and 50 μg/ml carbenicillin). The final suspension was divided into 500 ml portions and incubated at 29 °C and 182 rpm. When the OD_{600} reached a value of 0.8-0.9, synthesis of the recombinant proteins was induced by the addition of 0.25 mM isopropyl-β-D-thiogalactopyranoside (IPTG). After 3 h of incubation and a subsequent 20 min centrifugation at 4500 × g at 4 °C, the bacterial cells were collected. Each obtained pellet was washed once with 24 ml of the lysis buffer per 1 l of original culture, resuspended in 15 ml of the lysis buffer and frozen at -80 °C. The frozen cells from 1 l of culture were lysed by thawing in a 25 °C water bath. As soon as the cells started to thaw, they were supplemented with an appropriate volume of phenylmethylsulfonyl fluoride (PMSF) and β-mercaptoethanol to final concentrations of 0.2 mg/ml and 10 mM, respectively. Then, DNase I and RNase A were added to a final concentration of 10 μg/ml of each enzyme, and the lysates were incubated at 4 °C until there was a loss of viscosity, i.e., nucleic acids were completely digested. The cell extract was clarified by 45 min of centrifugation at 20,000 × g at 4 °C. The soluble fraction was collected, supplemented with 0.2 mg/ml PMSF, and then purified using immobilized metal ion affinity chromatography. The cell lysate was incubated for 1 h at 4 °C at 4 rpm with 1 ml of Co^{2+}-TALON resin (Clontech), which had been previously equilibrated with buffer A. The resin was then loaded onto a disposable column and washed with 20 ml of buffer A. Then, the resin was transferred to a Tricorn 5/50 column (Amersham Biosciences), connected to an ÄKTAexplorer (Amersham Biosciences) system and operated at 0.5 ml/min at room temperature. The column was washed with at least 5 ml of buffer A, and then, the target protein was eluted with a linear gradient of buffer B for 30 min. One-milliliter fractions were collected, combined and concentrated to a total volume of 500 μl using an Amicon Ultra-4 centrifugal filter unit (Millipore). The sample was then injected into a Superdex 200 Increase 10/300 G l column (Amersham Biosciences) equilibrated with buffer C. Chromatography was performed on an ÄKTAexplorer (Amersham Biosciences) system operated at 0.5 ml/min at room temperature. Fractions containing pure recombinant protein were collected, combined, aliquoted and stored at -80 °C. The same expression and purification procedure was used for HaUsp_ΔNTD. Concentrations of the purified proteins were determined with a NanoDrop instrument (Thermo Scientific). The absorption coefficients, calculated according to the method proposed by Gil and von Hippel [25], for HaUsp and HaUsp_ΔNTD were 0.650 ml/(mg × cm) and 0.706 ml/(mg × cm), respectively. Protein content and purity were estimated after each stage of purification with SDS-PAGE and Coomassie Brilliant Blue R-250-stained gels. Electrophoretic mobility was determined for the final samples from the protein preparations (see Supplementary Information).

DmUsp used in EMSA was expressed according to the procedure and with expression vectors presented previously [26]. The purification procedure included desalting with 35% $(NH_4)_2SO_4$ before affinity chromatography with the Co^{2+}-TALON resin (Clontech). Proteins eluted with 200 mM imidazole were subsequently injected into a Superdex 200 Increase 10/300 G l column (Amersham Biosciences). Compared to the procedure described by Rymarczyk et al., this

procedure allowed us to obtain 1 mg of homogenous DmUsp from a 1-l culture.

2.4. SDS-PAGE

Proteins samples were analyzed by SDS-PAGE using 4% polyacrylamide stacking gels and 12% polyacrylamide resolving gels developed in a Tris/glycine system [27]. An unstained protein molecular marker (Fermentas) was used. After electrophoresis, the proteins were stained with Coomassie Brilliant Blue R-250 [28].

2.5. Circular dichroism (CD)

Circular dichroism (CD) experiments were performed using a JASCO J-715 CD-spectropolarimeter. Measurements were performed at 20 °C (controlled with a Peltier-type temperature control system) with a scanning speed of 50 nm/min, data resolution of 1.0 nm and bandwidth of 1.0 nm A 1.0 nm-path-length cuvette was used for measurements, and HaUsp was used at a concentration of 10 μM. The spectra were collected in a spectral range of 190–260 nm in buffer C. The final spectrum was obtained after averaging three measurements. The measurements were corrected for the contribution of the buffer and converted to molar residual ellipticity units [29]. The secondary structure content was calculated using CDPro spectra deconvolution software developed by Sreerama and Woody with IBasis 4, which is a reference set of 43 proteins. The averages and standard deviations were determined for the results obtained from three algorithms: CONTILL, CDSSTR and SELCON3 [30].

2.6. Hydrogen-deuterium exchange monitored by mass spectrometry (HDX-MS)

Hydrogen-deuterium exchange (HDX) experiments were performed using HDX Manager with on-line digestion and separation using a nanoACQUITY UPLC System with *HDX technology coupled to a SYNAPT G2 HDMS instrument* (Waters).

Initially, a non-deuterated sample was prepared by mixing 10 μl of the HaUsp protein stock (45 μM) with 40 μl of buffer C, prepared in H_2O. The sample was then acidified with 10 μl of 2 M glycine-HCl and 4 M Gdm-HCl (pH 2.5) and digested on-line on an immobilized pepsin column (Poroszyme immobilized pepsin, ABI) with 0.07% formic acid in water as the mobile phase (flow rate, 200 μl/min). The resulting peptides were captured on a 2.1 × 5 mm C18 trap column (ACQUITY BEH C18 VanGuard pre-column, 1.7 μm resin; Waters) and then directly passed onto an ACQUITY UPLC-BEH C18 reversed-phase column (2.1 × 50 mm, 1.7 μm; Waters) and eluted using a gradient of 10%–35% acetonitrile in 0.1% (v/v) formic acid at a flow rate of 90 μl/min. The Manager system (Waters) was used to strictly maintain the temperatures of all fluids, valves, and columns at 0.5 °C, except for the pepsin column, which was maintained at 20 °C inside the temperature-controlled digestion compartment. For protein identification, mass spectra were acquired in MSE mode over an *m/z* range of 50-1950. The spectrometer parameters were as follows: ESI in positive mode; capillary voltage, 3 kV; sampling cone voltage, 35 V; extraction cone voltage, 3 V; source temperature, 80 °C; desolvation temperature, 175 °C; and desolvation gas flow, 800 l/h.

For deuterated samples, 5 μl of protein stock was diluted 10 fold with 45 μl of buffer C, prepared in D_2O. The exchange reaction was performed for 10 s, 1 min, 5 min, 30 min or 150 min and then immediately quenched in 2 M glycine-HCl and 4 M Gdm-HCl (pH 2.5) cooled on ice. The samples were analyzed the same way as the non-deuterated sample, but the C18 column outlet was coupled directly to the ion source of the spectrometer, which, additionally, was run in ion mobility mode. Each experiment was carried out in triplicate.

Two control HDX-MS experiments were performed to account for minimum and maximum exchange of peptides. To obtain minimum exchange, D_2O reaction buffer was added to the quenching buffer cooled on ice prior to the addition of protein stock. For maximum exchange, 5 μl of protein stock was mixed with 45 μl of D_2O reaction buffer, incubated overnight, and then mixed with quenching buffer and analyzed as described above.

2.7. HDX-MS data analysis

Peptides were identified using ProteinLynx Global SERVER software (PLGS, Waters). Peptide coverage map of the peptides generated from the on-line pepsin digestion is presented in the supplementary data (Fig. S4). The list of identified peptides was analyzed by the DynamX 3.0 program (Waters) with the following acceptance criteria: minimum intensity threshold of 3000 and minimum products per amino acid of 0.3. The isotopic envelopes of the peptides after exchange were analyzed using DynamX 3.0 with manual corrections wherever necessary. Final data containing molecular weights of all samples (M_{ex}) were exported to Excel (Microsoft Office) for calculations. Percent sample deuteration was calculated with a formula that takes into consideration the molecular weights from the minimum ($M_{ex}0$) and maximum exchange ($M_{ex}100$) values of a given peptide:

$$D(\%) = \frac{(M_{ex} - M_{ex}0)}{(M_{ex}100 - M_{ex}0)} \cdot 100\%$$

(1)

Error bars for percent deuteration D (%) were calculated as standard deviations of three independent experiments. Final graphs were plotted using OriginLab 8.5.

2.8. Electrophoretic mobility shift assay (EMSA)

EMSA [31] was performed using the following REs: *hsp27_EcRE* from the gene coding for the *D. melanogaster* heat shock protein [32] and *HR3_EcRE* from the gene coding for the *H. armigera* HR3 protein [33]. The sequences of the sense strands of the double-stranded oligonucleotides used in this experiment were as follows: *hsp27_EcRE*, 5′-AGCGACAAGGGTTCAATGCACTTGTCCAATGAA-3′; *HR3_EcRE*, 5′-GTCCCGGGGTCAATGAACTGCTGTGA-3′. Double-stranded oligonucleotides were labeled with [γ-32P]ATP using T4 polynucleotide kinase. The indicated amount of protein and approximately 40,000 cpm of [32P]-labeled probe were incubated for 30 min on ice in a final volume of 20 μl in buffer D containing 100 ng of poly(dI-dC). The receptor-DNA complexes were separated from the free DNA probe on a 5% polyacrylamide gel (20 × 16 × 0.15 cm) run in 0.25 × TBE pre-cooled to 4 °C overnight and pre-run at 160 V for 90 min. After applying the samples to the gel, electrophoresis was carried out at a temperature of 4 °C for 30 min at 270 V and then for 3 h at 200 V. The gels were dried under vacuum at 80 °C and exposed to imaging plates (Fujifilm, Tokyo, Japan). Fluorescence signals were scanned with a Fujifilm FLA-3000 fluorescent image analyzer (Raytest, Germany). Scans were read at a 50 μm resolution and with 16 bit quantitative image accuracy and then analyzed using AIDA Bio Package software (Raytest).

2.9. Analytical ultracentrifugation (AUC)

The sedimentation velocity AUC (SV-AUC) experiment was conducted in a Beckman Coulter ProteomeLab XL-I ultracentrifuge (software version 6.0, Beckman Coulter Inc., Brea, CA, USA). An An-60 Ti rotor and cells with two-channel charcoal-filled Epon® centerpieces were used. Four hundred microliters of HaUsp (0.15, 0.36 or 0.71 mg/ml) or HaUsp_ΔNTD (0.18, 0.37 or 0.81 mg/ml) in buffer C were centrifuged overnight at 50,000 rpm (approximately 200000 × g) at 20 °C. Sedimentation was monitored by absorbance at 280 nm at a 0.003-cm resolution in continuous mode.

Time-corrected data [32] were analyzed using SEDFIT [33]. The densities and dynamic viscosities of the buffer and partial specific volumes of HaUsp (0.734 ml/g) and HaUsp_ΔNTD (0.723 ml/g) at 20 °C were estimated using SEDNTERP [34]. Sedimentation coefficients (s), integrated water-corrected sedimentation coefficients (($s_{20,w}$)), frictional ratios (f/f_0) and molecular weights (MW_{app}) were calculated using the continuous $c(s)$ distribution model with at least 10 points per 1 S. Maximum entropy regularization with $p = 0.68$ was applied. The quality of the fits was assessed using RMSD values, residual distributions and residual histograms.

A sedimentation equilibrium AUC (SE-AUC) experiment was conducted for HaUsp (0.14, 0.35, 0.37, 0.72, 1.10 and 1.67 mg/ml) and HaUsp_ΔNTD (0.28, 0.94, 1.14, 1.44, 1.86 and 2.44 mg/ml) using the same 2channel cells as those used for SV; the sample volume was 170 μl. The experiment was performed at 5400, 10,700 and 18,000 rpm for HaUsp and at 6300, 10,700 and 18,000 rpm for HaUsp_ΔNTD at 8 °C. Every 3 h, intensity scans at 280 nm and 250 nm for the reference and sample sectors were collected in step mode at a 0.001-cm resolution. Ten replicates were accumulated for each scan. Equilibration of the samples was judged using SEDFIT. The data were processed using GUSSI [35] and fitted to a monomer-dimer equilibrium model using SEDPHAT [33].

2.10. Size-exclusion chromatography with multi-angle light scattering (SEC-MALS)

SEC-MALS analysis was performed using a high-performance liquid chromatography (HPLC) instrument (1260 Infinity LC, Agilent Technologies) equipped with a UV detector; Samples were monitored at wavelengths of 280, 254, and 215 nm. The HPLC instrument was connected to in-line detectors: a MALS detector (DAWN HELEOS II, Wyatt Technology) and a differential refractometer (Optilab T-rEX, Wyatt Technology). One hundred microliters of protein samples were loaded onto a Superdex 200 Increase 10/300 column (GE Healthcare) equilibrated with buffer C. Samples were run at room temperature at a flow rate of 0.5 ml/min. The results were analyzed using ASTRA v. 6 software (Wyatt Technology) in accordance with manufacturer's instructions.

2.11. Small-angle X-ray scattering (SAXS)

Full-length HaUsp and HaUsp_ΔNTD protein samples at concentrations of 1–8 mg/ml in buffer C were measured using a Xeuss 2.0 laboratory SAXS/WAXS system (Xenocs, Sassenage, France) equipped with MetalJet microfocus X-ray source (Excillum AB, Kista, Sweden) with a liquid gallium alloy target (X-ray λ = 0.134 nm) and a PILATUS3 R 1 M hybrid photon counting detector (Dectris AG, Baden-Daettwil, Switzerland). Protein samples were measured in a low-noise cell, and for each sample, 12 or 18 frames, 10 min each, were collected. Data measured from 0.012 Å⁻¹ to 0.58 Å⁻¹ were processed and checked for radiation damage using FOXTROT [34]. The radius of gyration (R_g) was calculated in Primus using the formula:

$$I(s) = I(0) \cdot exp\left(-\frac{1}{3}R_g^2 s^2\right)$$

(2)

and the data range was limited by a $sR_g < 1.3$ criterion [35]. The pair-distance distribution function ($P(r)$) and maximum diameter of the particle (D_{max}) were calculated in GNOM using the indirect Fourier-transform method [36]. Molecular weight (MW) was estimated by using

bovine serum albumin (BSA) as a reference protein of known molecular weight in solution (MW_{BSA}) and the formula:

$$MW_{sample} = MW_{BSA} \cdot \frac{I(0)_{sample}}{I(0)_{BSA}} \tag{3}$$

where $I(0)_{sample}$ and $I(0)_{BSA}$ are extrapolated intensities at a scattering angle of zero, obtained in Primus, for the measured sample and BSA, respectively.

2.12. Low-resolution modeling

Low-resolution modeling was conducted in DAMMIN [37], in which a simulated annealing Monte Carlo algorithm is used to create low-resolution models of proteins by starting from random assemblies of "dummy" atoms. Ten individual models obtained in DAMMIN were subsequently averaged and filtered in DAMAVER to obtain the final model [38]. In the modeling, two-fold symmetry was implemented, assuming dimeric states of the HaUsp and HaUsp_ΔNTD proteins.

Ensemble Optimization Method (EOM) modeling was conducted using models of the LBD (amino acids 169–424) and DBD (amino acids 69–141) of HaUsp created in the I-TASSER server [39], and the rest of the protein was represented as beads corresponding to individual residues [40,41]. For modeling of the dimers of full-length HaUsp and HaUsp_ΔNTD, the LBD domains were superposed on the crystal structure of the LBD domains of *D. melanogaster* Usp (PDB ID: 1HG4). Trimeric, tetrameric and pentameric assemblies of HaUsp variants were constructed using the SymmDock server [42]. An initial random pool of 10,000 models was generated in RanCh (version 2.1) [43]. For the dimer-oligomer equilibrium scenario, the starting pool was created by mixing the individual pools for the dimer and the different oligomeric states (trimer, tetramer or pentamer) of HaUsp. Half of the models were asymmetric, and the other half exhibited two-fold, three-fold, four-fold or five-fold symmetry. Final ensembles were selected from the starting pool using a genetic algorithm implemented in GAJOE (version 2.1) [43].

2.13. In silico analysis

The Uversky plot and VL-XT analysis were conducted using PONDR at http://pondr.com [44,45]. Protein backbone dynamics were predicted with DynaMine at http://dynamine.ibsquare.be [46,47]. Meta-Disorder at http://iimcb.genesilico.pl/metadisorder [48] was utilized for supporting disorder predictions. Disorder binding regions were predicted with ANCHOR at http://anchor.enzim.hu/ [49]. Domain ranges were determined using the Conserved Domain Database (CDD) at http://www.ncbi.nlm.nih.gov/cdd/ [50–53]. The 3D structure prediction of full-length HaUsp was performed with the I-TASSER server at http://zhanglab.ccmb.med.umich.edu/I-TASSER/ [39,41,54]. The resulting model was assessed according to a confidence score (C-score). The best model possessed C = -2.20. The top10 threading templates used by I-TASSER were as follows: 3dzyA, 4nqaA, 4nqaA, 4nqa, 4nqa, 3dzyA, 4nqa, 1g2nA, 3dzyA and 4iqrA. After performing the structure assembly simulation, I-TASSER used the TM-align structural alignment program to match the first I-TASSER model to all structures in the PDB library. Proteins structurally close to the target (HaUsp) or to its fragments in the PDB library (as identified by TM-align) were found as follows: 4nqaB, 2aclH, 1pq6B, 3a40X, 1z5xE, 1rkhA, 2nxxE, 3e00D, 1r1kD and 3dr1 A. The structure of the hRXR-alpha/hLXR-beta heterodimer on DNA (4nqaB) was considered reliable (TM-score = 0.772, RMSD = 1.49, IDEN = 0.285 and Cov = 0.792), and used for the data interpretation and showing the 3D model of the corresponding DBD and LBD of HaUsp.

3. Results

3.1. In silico analysis

The NTDs of NRs often do not possess stable tertiary structures and only residual secondary structures. Hence, these NTDs can be classified as IDRs [10,16,17,55]. In fact, disordered conformations provide many functional advantages; therefore, IDRs are often identified as fragments involved in a variety of signaling pathways. IDRs of proteins can be targets for posttranslational modifications such as phosphorylation [56]. The dynamic and flexible nature of disordered polypeptide chains allows diverse interactions with multiple partners, which may be important for molecular assembly processes, ensuring proper complex formation [56]. Recent findings also indicate that IDRs are important for the regulation of protein degradation [56,57]. Furthermore, during alternative splicing, different isoforms of a protein, often with variable IDR length, can be generated, which provides an additional level of functional control for a given protein [58]. Romero *et. al* [59] showed that regions of pre-mRNA lost in the alternative splicing are often intrinsically disordered and they presence in the protein products is associated with functional and regulatory diversity. Considering the multifunctionality of IDRs, it is important to investigate how an intrinsically disordered protein fragment can perform many functions and participate in many signaling pathways while constantly bearing a single amino acid sequence.

In the case of Usp NTD, the structure-function relationship in the context of the whole protein is not known. Thus, the main purpose of this work was to investigate the molecular properties of this domain and the potential influence of this domain on the molecular properties, and in turn the functions, of the remaining Usp domains, which possess well-defined, stable tertiary structure, such as the DBD [60,61] and LBD [7,8]. Prior to *in vitro* analyses, we decided to analyze the structural organization of the NTD *in silico* in the context of the full-length HaUsp. Great emphasis was placed on identifying intrinsically disordered regions in HaUsp. The CDD [52] indicated that the HaUsp DBD and LBD cover amino acid residues from positions 59 to 135 and 155 to 387, respectively, so the NTD of this protein is formed by the amino acid residues fragment from positions 1 to 58 (data not shown). Previous research has indicated that the N-terminal fragment of the Usp from *A. aegypti* is disordered [17]. To accurately delineate the IDRs in HaUsp, the PONDR VL-XT [62] algorithm was used. As seen in Fig. 1A, the highest propensity for ordered structure was observed for fragments with amino acid residues 60–100 and 190–380, defined by CDD as DBD and LBD, respectively. In contrast, a fragment including the amino acid residues at positions 1–60 and the hinge located between the predicted DBD and LBD showed high propensity for disorder. The results obtained using the PONDR VL-XT algorithm were also confirmed by the Dyna-Mine algorithm, which is designed for the prediction of protein backbone dynamics [46]. Obtained values close to 1 indicate stable conformations (ordered state), whereas values close to 0 correspond to highly dynamic conformations (disordered state) [47]. As shown in Fig. 1B, DynaMine identified the N-terminal fragment of HaUsp as being highly flexible and hence disordered. The calculated percentage of the disordered region (S^2 value lower than 0.69) was 24.8%. Supporting disorder predictions were conducted using MetaDisorder [48] and obtained results for MetaDisorderMD2, IUPred and DisProt (VSL2) (Fig. S1) were in good agreement with previous predictions. Finally, a charge-hydropathy plot was used to classify HaUsp as an ordered or disordered protein (Fig. 1C). This tool discriminates between order and disorder by plotting the net charge against the hydropathy [63]. The obtained results localized the full-length HaUsp in a group of ordered proteins; however, the relative proximity to the boundary can indicate that there also exist some disordered regions in this protein. Furthermore, HaUsp_ΔNTD exhibited an increased mean hydropathic value. Finally, when the sequence of only NTD was analyzed, the mean hydropathy value obtained was 0.3712, and on the plot, this domain is
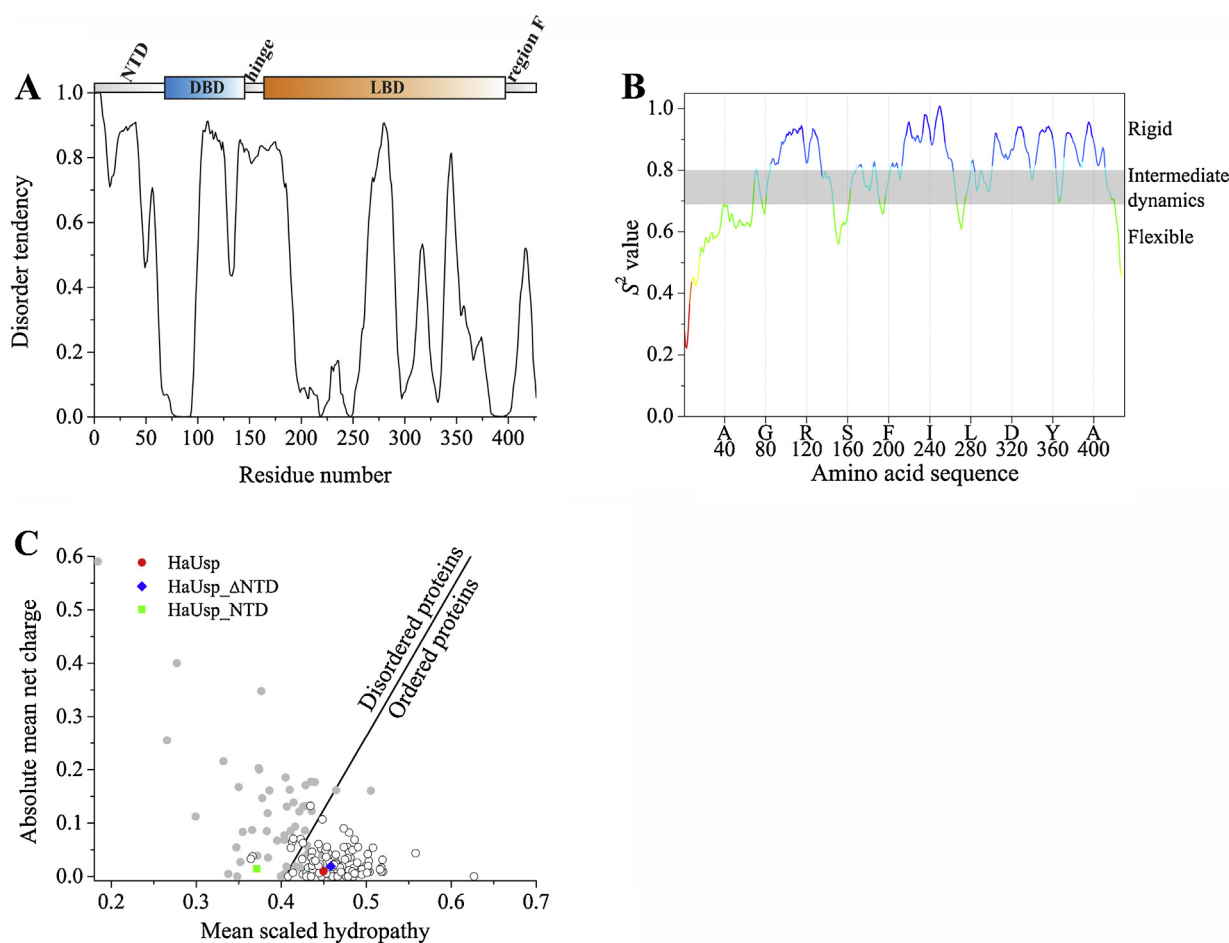
Fig. 1. *In silico* analysis of HaUsp. (A) Prediction of the degree of disorder in HaUsp calculated from the primary structure using the PONDR VL-XT [62] neural network predictor. A score of more than 0.5 indicates high probability of disorder. The top of the panel represents the domain structure of HaUsp based on the Conserved Domain Database. Blue represents the DNA-binding domain (DBD); orange represents the ligand-binding domain (LBD); and gray represents the N-terminal domain (NTD), hinge and region F. (B) Prediction of protein backbone dynamics with DynaMine. An S^2 value larger than 0.8 indicates high rigidity of the sequence, whereas an S^2 value lower than 0.69 indicates high flexibility, which is typical for disordered segments. Values between 0.69 and 0.8 are characteristic of context-dependent structural organization of polypeptide chains. (C) Charge-hydropathy plot, which divides proteins into globular (white dots) and disordered (gray dots) groups based on their absolute mean net charge versus mean scaled hydropathy. The black line indicates the boundary between these two groups. Full-length HaUsp is represented by a red circle; HaUsp_ΔNTD is represented with a blue diamond; and NTD alone is represented with a green square.

localized in the region occupied strictly by fully disordered proteins. Taken together, the results of the *in silico* analyses indicated that fragments predicted by CDD as DBD and LBD are characterized with ordered structure propensities. The NTD and the hinge region located between the DBD and LBD are predicted to be intrinsically disordered. Furthermore, when HaUsp amino acid sequence was analyzed with ANCHOR [49] it predicted one disordered binding region in the NTD (Fig. S2) which may be relevant for HaUsp functionality. The *in silico* analyses showed that the structural organization of HaUsp, which is characterized by mixed, partially disordered nature, is typical for NRs. The putative IDR-like characteristics of the NTD of HaUsp may be important for the activity of this protein.

3.2. CD analysis of HaUsp

CD spectroscopy is a rapid technique often used for determination of the secondary structures of proteins. Chromophores of the amides of polypeptide chain aligned in arrays unequally absorb left- and right-handed circularly polarized light. Different structural elements have been shown to exhibit characteristic maxima, negative and positive, in the far-UV range [64]; therefore, the shape of the CD spectrum is indicative of the type of secondary structure present in the analyzed protein. The obtained CD spectrum of HaUsp (Fig. 2) had deep negative
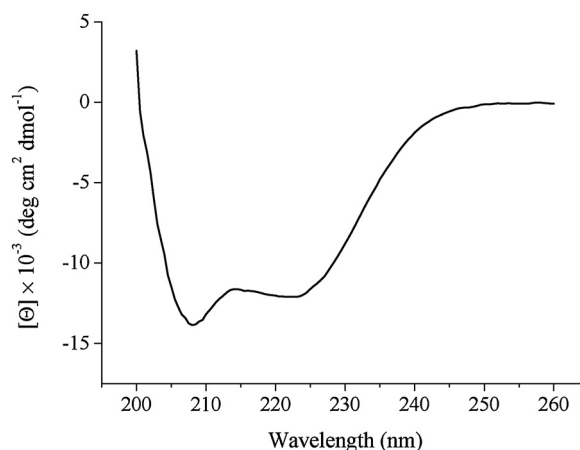

Fig. 2. CD spectrum of HaUsp. The far-UV spectra were recorded in 50 mM Na_2HPO_4 and 150 mM NaCl (pH 7.0) at room temperature. For measurement, 10 mM protein was used. The averaged spectrum of the native protein exhibits deep negative maxima at approximately 208 nm and 222 nm, which is characteristic of an ordered protein containing α-helices [29,65].

Table 2

Estimation of the HaUsp secondary structure content from the CD spectrum. The secondary structure content was calculated using the CDPro package with IBasis 4, which is a reference set of 43 proteins. The means ± standard deviations were calculated for results obtained from three algorithms: CONTINLL, CDSSTR and SELCON3 [30].

α helix (%)			β sheet (%)			Other (%)	
Regular	Distorted	Total	Regular	Distorted	Total	Turn	Unordered
24.4 ± 1.1	21.4 ± 0.9	45.8 ± 1.7	4.3 ± 1.2	5.5 ± 0.2	9.8 ± 1.3	21.0 ± 1.7	23.0 ± 1.6

maxima at approximately 208 and 222 nm, which are typical for ordered secondary structures [29,65]. This result was consistent with the results of the bioinformatic analyses presented above, which indicated that HaUsp is characterized by a high percentage of ordered structure. To estimate the types of secondary structures in HaUsp, the CDPro software package was used with the CONTINLL, CDSSTR and SELCON3 algorithms [30]. The obtained and averaged results are presented in Table 2. Deconvolution of the spectrum revealed that the most common type of secondary structure is α-helix, which constitutes 45.8 ± 1.7% of the secondary structure of the polypeptide chain. It has been demonstrated that the DBDs and LBDs of NRs are helical, especially LBDs, which can have up to 12 α helices [5]. Therefore, the obtained result may be an indication that these domains are well folded in the obtained HaUsp. β Sheets are relatively less abundant in the secondary structure, with values of 9.8 ± 1.3%. Most likely, β sheets are primarily present in the LBD, which is known to contain a common β turn [5]. Most importantly, deconvolution of the CD spectrum also revealed a significant percentage of disordered structure, with a value of nearly 25%. The obtained value is similar to the disordered content calculated by DynaMine (24.8%) analysis. Both of these results, together with the previous *in silico* predictions, suggest that the NTD may significantly impact these values, which strongly suggests a disordered nature for HaUsp NTD.

3.3. Investigation of intrinsically disordered regions by hydrogen-deuterium exchange mass spectrometry (HDX-MS)

In silico analyses along with CD spectroscopy indicated that some fragments (approximately 25%), most likely the NTD, of HaUsp are disordered. To directly map the IDRs in HaUsp, we utilized HDX-MS. In this technique, the rate of chemical exchange between deuterons in water and hydrogens on the amide groups of a protein is determined by the mass spectrometry. Amide groups deeply buried in the core of a protein or those participating in intramolecular bond formation are characterized by a very low rate of hydrogen exchange. In contrast, amides from the dynamic disordered regions of proteins exchange hydrogens very rapidly. Based on this difference, intrinsically disordered regions in a given protein can be identified [66]. The obtained results are presented in Fig. 3A, which is a graph presenting the percentage of deuteration in different regions of HaUsp after a 10-second incubation with D_2O. We also applied these results to a 3D model of the HaUsp molecule generated by I-TASSER [39] and visualized with PyMOL [67] (Fig. 3B). The results showed that the highest rate of exchange, approximately 80%, was observed for the peptides covering the NTD, which is consistent with the results of the bioinformatic analyses, indicating that this region is disordered. Notably, the algorithm of the I-TASSER tool also predicted this protein fragment to be extended and lacking secondary structure (Fig. 3B). In addition to the NTD, a high rate of isotopic exchange was also observed in the C-terminal fragment (F region) and in some short fragments of the LBD, which are likely highly exposed. The lowest exchange rates were observed for peptides originating from the DBD and LBD. The limited exchange rate of amide hydrogens in those two domains indicated that the hydrogens are engaged in the formation of hydrogen bonds stabilizing the secondary structure and are also deeply buried inside the folded structure. The

putative structure of these two domains is presented in the I-TASSER-derived structural model (Fig. 3B).

All the above results were consistent with the previous *in silico* analyses and the deconvolution analyses of the CD spectrum. However, the HDX-MS experiments allowed us to experimentally confirm that the NTD of HaUsp is intrinsically disordered.

3.4. Influence of the NTD on the interaction with DNA investigated by EMSA

NRs regulate gene expression by binding to specific DNA sequences called REs [68]. At present, not much is known about whether and how intrinsically disordered fragments of invertebrate NRs, including NTDs, influence NR-DNA interactions. Given that the NTD of HaUsp is most likely fully disordered, we decided to investigate whether the presence of this N-terminal IDR is important for the interaction of HaUsp with specific REs. For this purpose, the interaction of full-length HaUsp and HaUsp_ΔNTD with DNA probes was investigated using EMSA [31]. Two different [^{32}P]-labeled probes were used, each of which contained an ecdysone-response element (EcRE). One of these probes carried the sequence of *hsp27_EcRE* [32] and the second carried the sequence of *HR3_EcRE* [33]. For each EcRE, three different concentrations of HaUsp and HaUsp_ΔNTD were used, and the full-length *D. melanogaster* Usp (DmUsp), the interaction of which with *hsp27_EcRE* has been described previously [26], was used as a positive control.

As shown in Fig. 4, all three recombinant proteins were functional as these proteins interacted with specific EcREs. Furthermore, at higher protein concentrations, a stronger signal corresponding to the protein-DNA complex was observed along with depletion of the free probe signal, most likely caused by the interactions with the proteins. This result indicated the presence of concentration-dependent interactions. For the positive control, i.e., DmUsp interacting with *hsp27_EcRE,* two defined bands, corresponding to monomeric and dimeric protein-DNA complexes, could be observed (Fig. 4, lane 15), as described previously [26]. A similar result was obtained for HaUsp_ΔNTD. In this case, formation of two different protein-DNA complexes was also observed, for both *hsp27_EcRE* (Fig. 4, lanes 5, 6, 7) and *HR3_EcRE* (Fig. 4, lanes 12, 13, 14), although the proportion of the complex with greater mobility was larger (compare lanes 6 and 13 with lane 15). The presence of NTD in HaUsp resulted in a change in the interaction pattern with specific DNA sequences. In particular, for *HR3_EcRE,* significant reduction in the intensity of the main band, which was characterized by high mobility, was observed together with the absence of the sharp band, characterized by low mobility (Fig. 4, lanes 9, 10, 11). On the other hand, for *hsp27_EcRE*, no band or very weak bands were observed, corresponding to a stable protein-DNA complexes under the experimental conditions (Fig. 4, lanes 2, 3, 4). The simplest explanation for these differences is the low affinity of full-length HaUsp to specific DNA sequences. However, comparison of free probe signals corresponding to HaUsp and HaUsp_ΔNTD suggests that both of these signals were titrated to a greater extent in HaUsp than in HaUsp_ΔNTD. This observation was particularly evident for the *HR3_EcRE* probe (compare lanes 10 and 11 with 13 and 14, respectively). This result may suggest that complexes formed by full-length HaUsp are less stable in the conditions used for EMSA. Nonetheless, it is clear that the intrinsically disordered NTD,
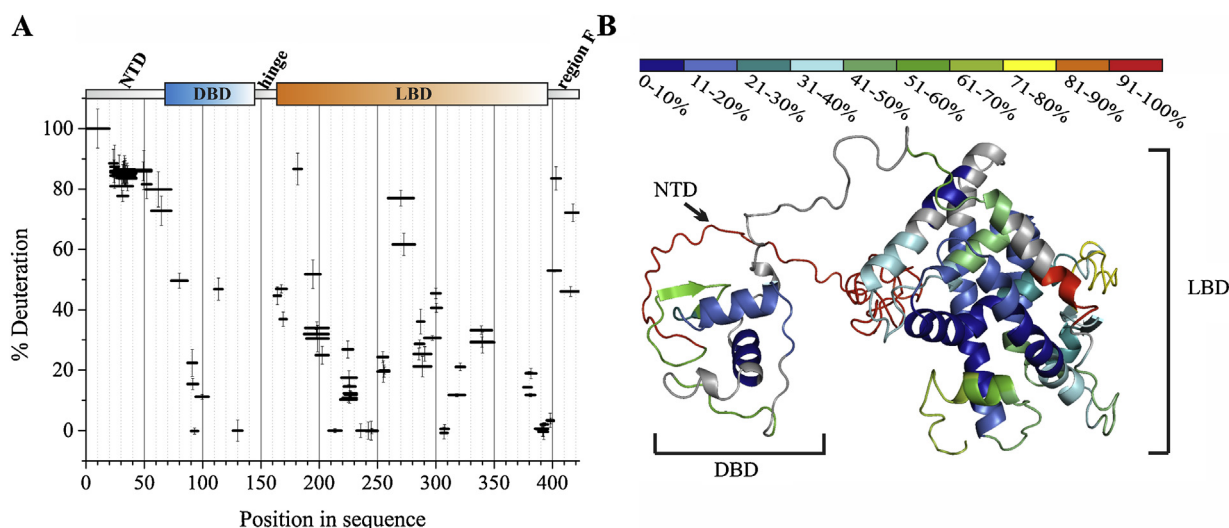
A

B

Fig. 3. HDX of HaUsp. (A) Deuteration percentage of peptide fragments from HaUsp with an exchange time of 10 s. The positioning of peptides in the sequence is shown along the x-axis, represented with a horizontal bar with a length corresponding to a particular peptide. The position of the bar along the y-axis indicates the fraction exchanged after 10 s. The vertical error bars are standard deviations calculated from three independent experiments. The top of the panel represents the domain structure of the protein. (B) 3D model of HaUsp. The structure was generated using the I-TASSER web-based tool [39] and the sequence of recombinant HaUsp (including a 6×His tag). Subsequently, the results were visualized with PyMOL [67]. Segments of the protein are colored with the degree of deuteration in a time of 10 s according to the upper scale. The brackets and arrows indicate specific domains.
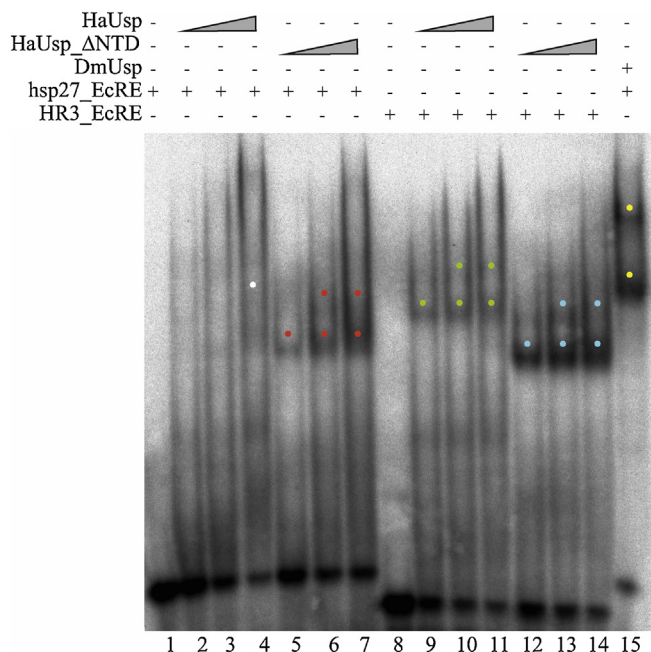

Fig. 4. EMSA result of HaUsp and HaUsp_ΔNTD. EMSA was performed with full-length HaUsp (lanes 2–4 and 9–11) or HaUsp_ΔNTD (lanes 5–7 and 12–14). DmUsp was used as a positive control (lane 15). Lanes 1–7 correspond to *hsp27_EcRE*. Lane 1 is a control without any protein. Lanes 2–4 represent increasing concentrations of HaUsp: 250 nM, 625 nM and 1250 nM. Lanes 5–7 represent increasing concentrations of HaUsp_ΔNTD: 250 nM, 625 nM and 1250 nM. Lanes 8–14 correspond to *HR3_EcRE*. Lane 8 is a control without any protein. Lanes 9–11 represent increasing concentrations of HaUsp: 250 nM, 625 nM and 1250 nM. Lanes 12–14 represent increasing concentrations of HaUsp_ΔNTD: 250 nM, 625 nM and 1250 nM. Lane 15 is a positive control of DmUsp (625 nM) over *hsp27_EcRE*. Signals corresponding to free probes are visible on the bottom. Dots represent the formed protein-DNA complexes discussed in the text.

Table 3
Hydrodynamic properties of HaUsp and HaUsp_ΔNTD obtained by SV-AUC analysis.

Protein	c [mg/ml]	s [S]	$\overline{s_{20,w}}$ [S]	f/f_0	R_h [nm]	MW_{app} [kDa]	Fraction of signal [%]
HaUsp	0.15	3.08	3.28	1.375	3.21	43.9	96.4
		5.18	5.50		4.16	95.6	3.7
	0.36	3.24	3.44	1.395	3.36	48.3	97.5
		5.30	5.64		4.30	101.5	2.7
	0.71	2.97	3.21	1.357	3.11	41.8	39.2
		3.63	3.92		3.44	56.4	61.7
		5.21	5.63		4.12	96.9	3.3
HaUsp_ΔNTD	0.18	3.07	3.26	1.350	3.13	43.1	100
	0.37	2.81	2.99	1.328	2.93	36.9	32.1
		3.44	3.66		3.24	49.8	92.1
	0.81	2.85	3.03	1.354	3.03	38.7	26.2
		3.62	3.85		3.42	55.4	74.6

composed of 58 residues, significantly influences the DNA-binding properties of HaUsp.

3.5. Hydrodynamic properties of HaUsp and HaUsp_ΔNTD

Homo- or heterodimerization of NRs is necessary for the function of these proteins as transcription factors. Our EMSA showed that, in its active form, HaUsp forms various protein-DNA complexes. Here, we utilized AUC to determine the hydrodynamic properties of full-length HaUsp and HaUsp lacking NTD and to test whether the presence of the intrinsically disordered NTD may influence the oligomerization propensity of the protein. SV experiments were carried out to determine the shape-related parameters of HaUsp and HaUsp_ΔNTD. The results obtained showed an increase in the s values in a concentration-dependent manner for both forms of the protein (Table 3, Fig. 5). For full-length HaUsp, at each concentration, more than one peak was observed. At the lowest analyzed concentration (0.15 mg/ml), in the *c(s)*
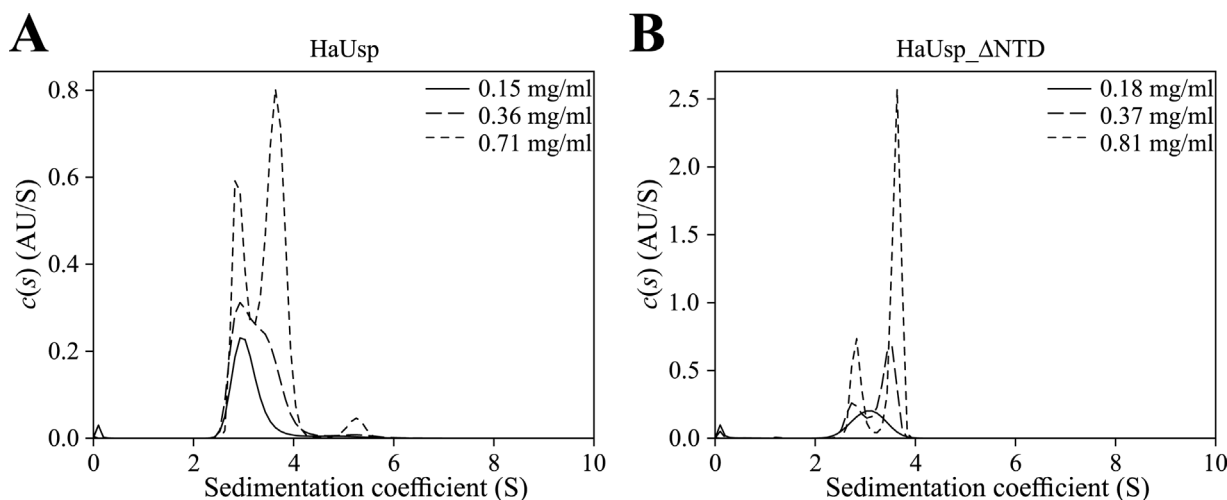
Fig. 5. Oligomerization of HaUsp and HaUsp_ΔNTD assessed by SV-AUC. (A) *c(s)* distribution of HaUsp (B) *c(s)* distribution of HaUsp_ΔNTD. The calculations were performed in SEDFIT using a continuous *c(s)* distribution model.

distribution, the main peak (96.4%) was characterized with an *s* value of 3.08 S (Fig. 5A, Table 3). At the intermediate concentration (0.36 mg/ml), the peak was asymmetric and was apparently composed of at least two different peaks. At the highest concentration (0.71 mg/ml), this one peak was resolved into additional peaks - two taller peaks with *s* values of 2.97 S and 3.63 S and one smaller peak with an *s* value of 5.21 S. These changes indicated the presence of weak, reversible self-association of HaUsp [69,70]. At the highest concentration, the apparent molecular weight (MW_{app} = 41.8 kDa) of the peak with the lowest *s* value (2.97 S) (Table 3) indicated the presence of a monomer, the molecular weight of which, calculated from the protein sequence (MW_{seq}), is 48,342 Da. The intermediate species at this concentration (3.63 S), with MW_{app} = 56.4 kDa, cannot be attributed to a stable oligomer. Instead, that population consisted of a mixture of monomers and oligomers, likely dimers that undergo fast and reversible association [71,72]. Finally, the MW_{app} of the last peak at this concentration (5.21 S), calculated to be 96.9 kDa, suggested that homodimers of HaUsp are present.

For HaUsp_ΔNTD, an irregular *c(s)* distribution was also observed. At 0.18 mg/ml, a single peak with an *s* value of 3.07 S was observed. At higher concentrations, this peak resolved into two peaks, characterized by *s* values of 2.81 S and 3.44 S at 0.37 mg/ml and 2.85 S and 3.62 S at 0.81 mg/ml. The increased *s* value of the main peak (Table 3, indicated by the light gray background) indicated that weak, reversible dimerization also occurs in the case of HaUsp_ΔNTD, (Fig. 5B). However, no peak corresponding strictly to the dimeric form was observed, which indicated that deletion of NTD decreases the tendency for dimerization and/or stability of homodimers.

An SE experiment was performed to determine the stability of the oligomers of HaUsp and HaUsp_ΔNTD (Fig. 6). The best fits for both forms of the protein were obtained for a monomer-dimer equilibrium model. The estimated monomer *MW* values of 47.0 kDa for HaUsp and 41.8 kDa for HaUsp_ΔNTD are consistent with the MW_{seq} values (48,342 and 42,347 Da, respectively). The dissociation constant (K_d) of HaUsp was determined to be 31.7 µM. The K_d of HaUsp_ΔNTD was higher than that of HaUsp, with a value of 46.4 µM (Table 4). To eliminate possible degradation of protein samples were analyzed with SDS-PAGE and results are presented in the supplementary (Fig. S5).

The obtained K_d values indicated that the interaction between HaUsp monomers is relatively weak, which explains the modest dimerization of the protein observed in the SV experiments. Deletion of the NTD further decreased the strength of the dimeric complex, as indicated by the increase in K_d. This observation is also consistent with the results obtained in the SV experiment. It can be concluded that the

intrinsically disordered NTD influences the dimerization of HaUsp, stabilizing the oligomeric form.

3.6. Full-length HaUsp forms higher oligomeric structures in solution

Subsequently, we decided to analyze the influence of the intrinsically disordered NTD on the oligomerization of HaUsp via an independent technique. For this purpose, HaUsp and HaUsp_ΔNTD were analyzed by SEC-MALS. In this technique, proteins are first separated by SEC based on their hydrodynamic properties, and subsequently, the MWs of specific populations of molecules can be determined with a MALS detector. This technique allows shape-independent determination of the *MWs* of analyzed macromolecules [73]. HaUsp and HaUsp_ΔNTD were analyzed at different concentrations to validate the occurrence of concentration-dependent dimerization (Table 5). For HaUsp_ΔNTD, only one peak was observed, whereas the chromatograms for HaUsp contained some additional minor peaks at smaller elution volumes (Fig. 7.), indicating the presence of higher oligomers. Moreover, when the sample concentration was increased, the main peak was observed to shift to lower elution volumes. The *MW* calculated for 0.1 mg/ml HaUsp_ΔNTD was close to the theoretical value obtained for the monomer (MW_{seq} = 42,347 Da), and this value increased to 63.4 kDa for the highest measured concentration (5.0 mg/ml). This value was almost 1.5 times higher than the MW_{seq}, which may be because the observed peak corresponds to a mixture of monomers and dimers in the sample. The AUC experiment indicated that the self-association/dissociation of HaUsp_ΔNTD is fast; here we see that this process is too fast to obtain separate fractions of monomers and dimers. For full-length HaUsp, at 0.1 mg/ml, one main peak with a *MW* of 47.93 kDa was observed, which is close to the theoretical value obtained for the monomers (MW_{seq} = 48,342 Da). Similar to HaUsp_ΔNTD, at 5.0 mg/ml, the determined *MW* of the fraction from the main peak was 1.5 times higher than the theoretical value obtained for the monomer. However, in contrast to HaUsp_ΔNTD, for full-length HaUsp, some peaks of higher *MW* values were also detected, and starting from 1.0 mg/ml, the value of these peaks were close to the *MW* of the theoretical dimer (MW_{seq} = 96,684 Da) (Fig. 7, Table 5).

The above results indicated that both proteins have propensity for rapid reversible self-association, which is reflected in the observed shift of the main peak with increasing concentration and in the absence of two defined peaks corresponding to monomer and dimer [70]. However, for full-length HaUsp, additional peaks have been identified, and the *MW* values of these peaks have been determined. Starting from 1.0 mg/ml, the *MW* values calculated for these additional peaks
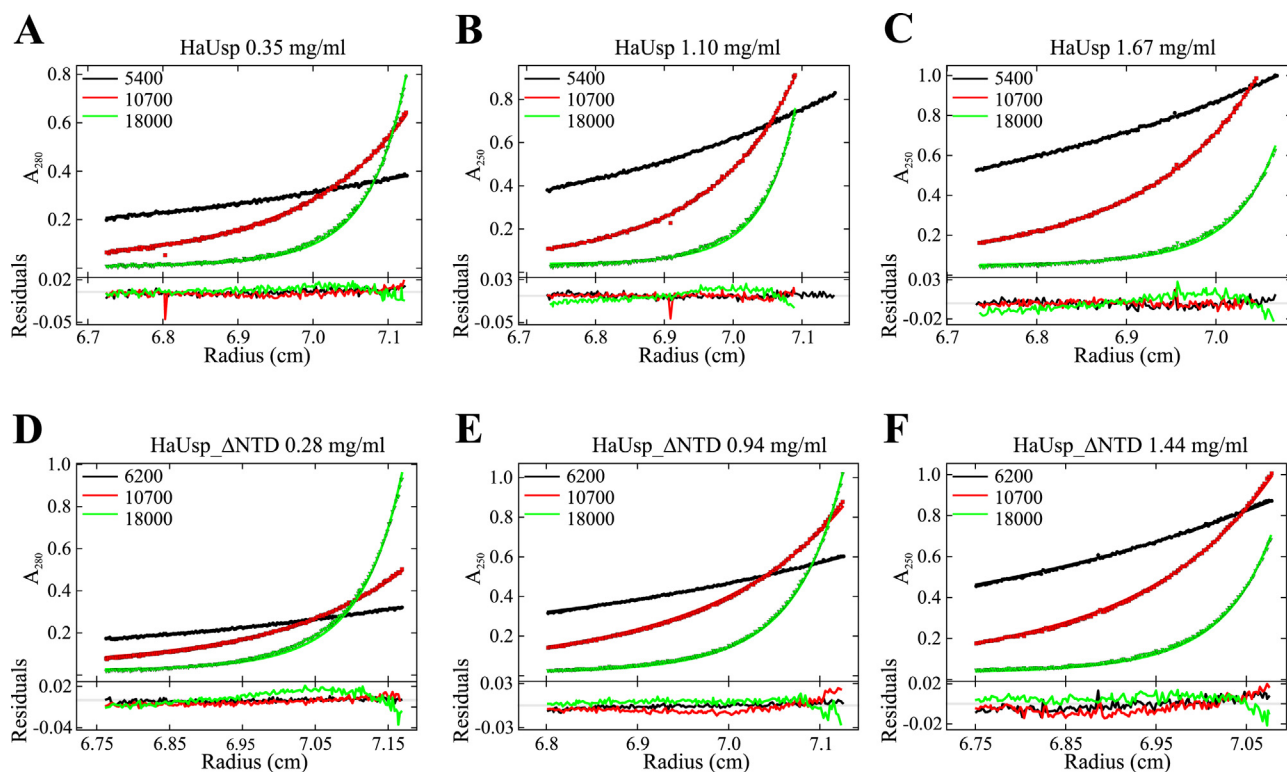
Fig. 6. SE-AUC analysis of HaUsp and HaUsp_ΔNTD. Representative results of SE-AUC experiments performed at 8 °C for (A) 0.35, (B) 1.10 and (C) 1.67 mg/ml HaUsp and (D) 0.28, (E) 0.94 and (F) 1.44 mg/ml HaUsp_ΔNTD. The samples were centrifuged at the speeds indicated in the figures (in rpm). Sets of SE-AUC experiments for each protein were globally fitted to a monomer-dimer equilibrium model in SEDPHAT software.

Table 4

Results of SE-AUC experiments. The calculations were conducted using a monomer-dimer equilibrium model in SEDPHAT software.

Protein	MW_{fit} [kDa]	MW_{seq} [Da]	Log(K_a)	$K_d = 10^{-\log(K_a)}$ [μM]	χ^2
HaUsp	47.0	48,342	4.499	31.7	1.458
HaUsp_ΔNTD	41.8	42,347	4.333	46.4	1.086

Table 5

Calculated molecular weights characterizing the peaks resolved by SEC-MALS. The main peaks are indicated in bold.

Protein	c [mg/ml]	MW [kDa]
HaUsp	0.1	**47.93**
	0.3	**51.08**/67.63
	0.5	**51.73**/54.56/84.08
	1.0	**56.52**/97.11
	2.0	**61.76**/102.30
	5.0	**69.54**/98.00
HaUsp_ΔNTD	0.1	40.93
	0.3	44.18
	0.5	46.98
	1.0	51.88
	2.0	57.43
	5.0	63.40

corresponded to the theoretical *MW* of the dimer, which indicated that the intrinsically disordered NTD is indispensable for effective dimerization of HaUsp. The NTD likely facilitates this process or stabilizes the formed dimers of HaUsp.

3.7. Scorpion-like dimeric structure can stabilize the HaUsp dimer

The above results showed that the intrinsically disordered NTD is an important determinant of the molecular properties of HaUsp and that deletion of the NTD changes the manner of interaction of HaUsp with specific DNA sequences and the ability of HaUsp to dimerize. To gain further insight into the molecular basis of these observations, we decided to carry out SAXS analyses of the full-length HaUsp and HaUsp_ΔNTD and then used the obtained data for structure modeling. Representative scattering curves are presented in Fig. 8A as a Kratky plot. Compared to the typical bell-shaped curves exhibited by globular proteins, for HaUsp and HaUsp_ΔNTD, we observed plots that are characteristic for partially disordered proteins [74], which confirms the results of the *in silico* analysis and the CD and HDX-MS data presented above. As the first step of the SAXS analysis, the basic structural parameters - radius of gyration (R_g) and maximum intramolecular distance (D_{max}) - were calculated (Table 6). The calculated R_g values for full-length HaUsp and for HaUsp_ΔNTD increased with increasing concentrations of analyzed proteins (Fig. 8B and C). Likewise, when D_{max} was calculated, higher values were observed for higher protein concentrations (Table 6). The pair-distribution function ($P(r)$), which reflects the frequencies of the distances (r) within a protein molecule, was calculated for both forms of the protein by indirect Fourier transformation of the experimental solution scattering data. The $P(r)$ functions for every analyzed concentration of HaUsp and HaUsp_ΔNTD are presented in Fig. 8D and E, respectively. The obtained asymmetrical profiles of the $P(r)$ functions indicated that both proteins have elongated shapes [75]. For each concentration, the *MW* was calculated by comparing the scattering intensity extrapolated to the zero angle ($I(0)$) with the $I(0)$ of BSA [76]. The results showed that the calculated *MW* value of HaUsp_ΔNTD at the lowest concentration (1.0 mg/ml) was
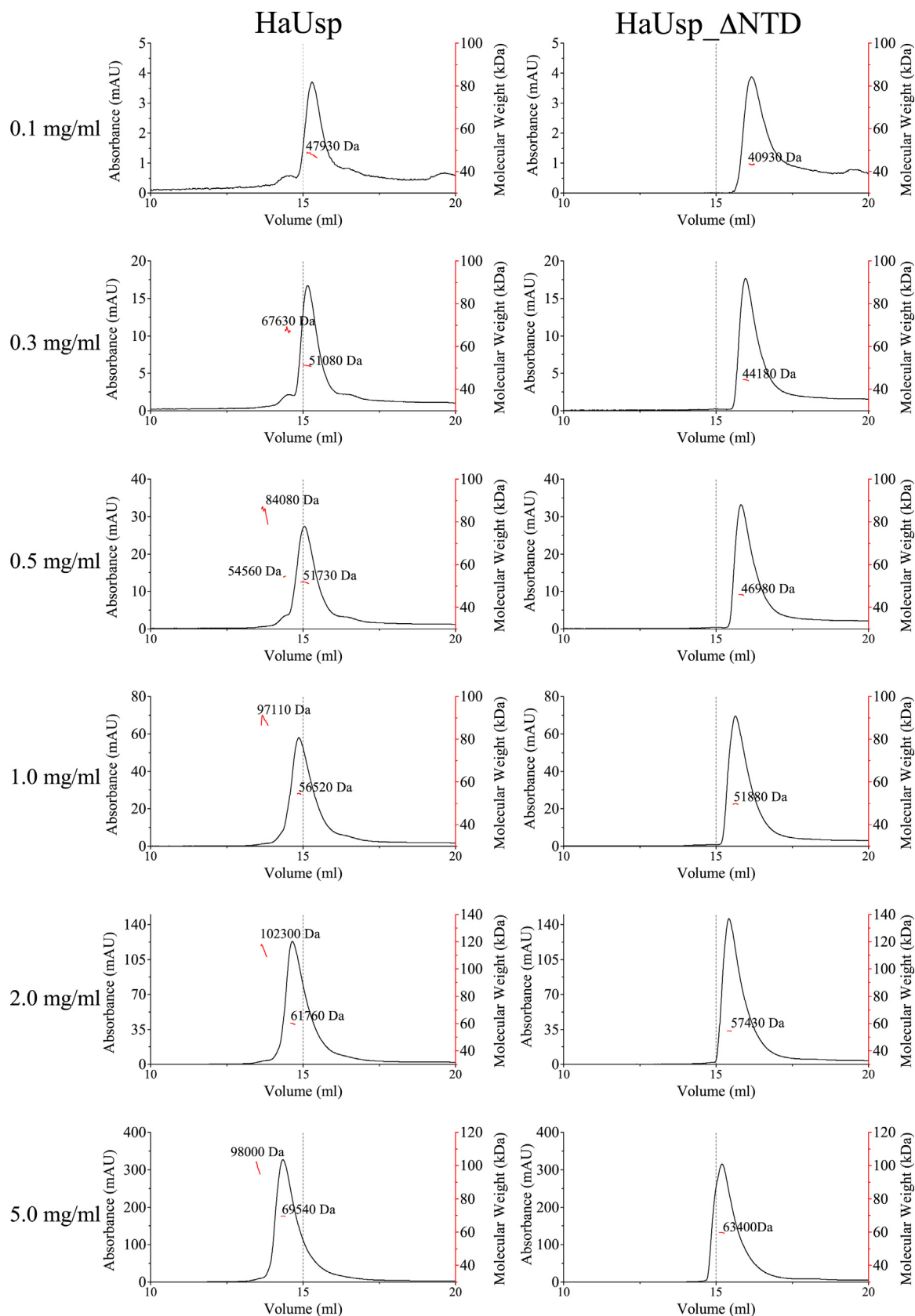
Fig. 7. Molecular weight determination by SEC-MALS. Analysis of the oligomeric states of HaUsp and HaUsp_ΔNTD by SEC-MALS. Average *MW* values as measured by MALS are given by the selected peaks. The left column corresponds to the full-length HaUsp, and the right column corresponds to HaUsp_ΔNTD. In each chromatogram, the black line corresponds to the absorbance measured at 280 nm, and the red lines correspond to the average *MW* values (right axis).
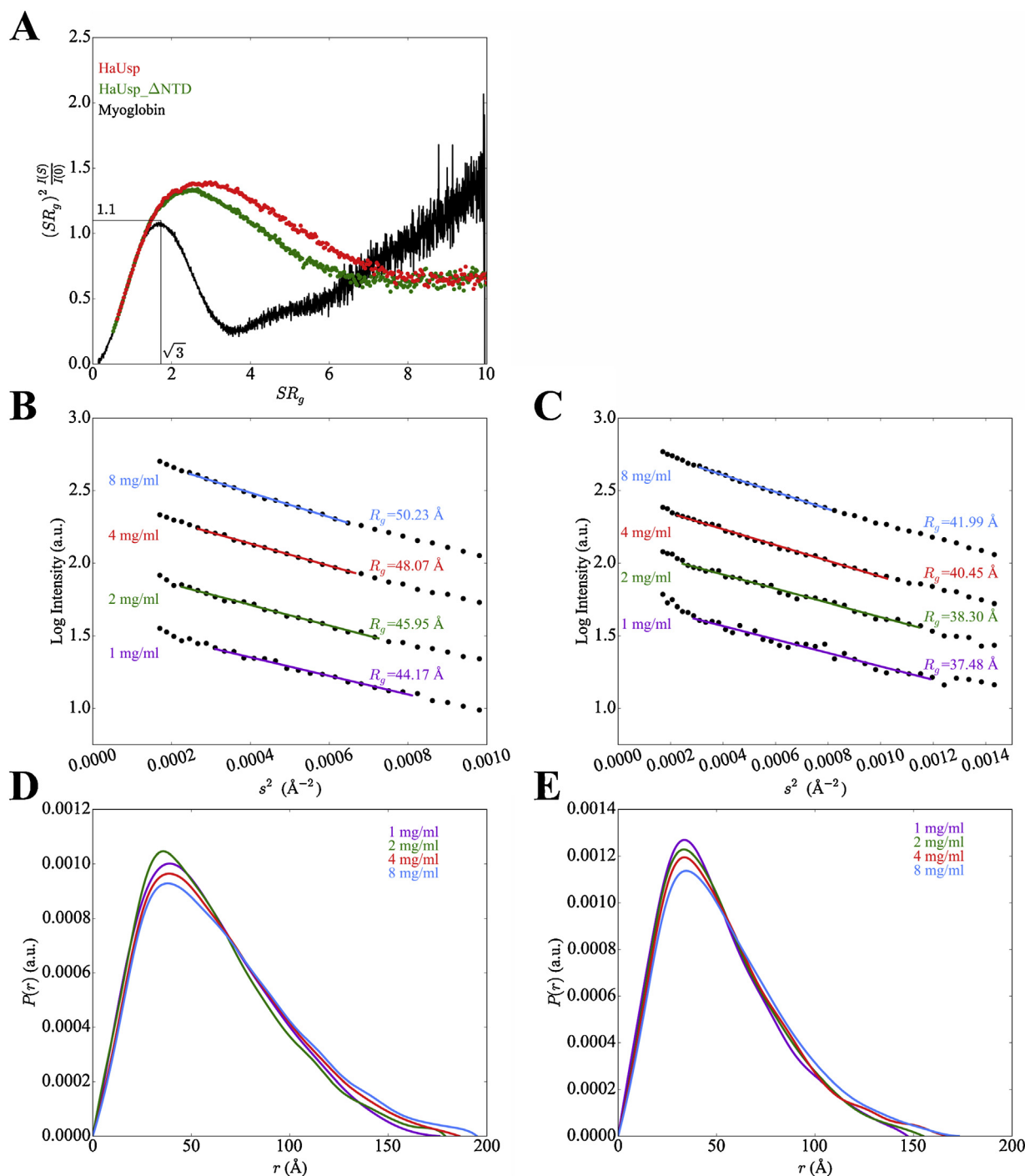
Fig. 8. SAXS analysis of HaUsp and HaUsp_ΔNTD in solution. (A) Kratky plot of the solution scattering data. The scattering SAXS curves of HaUsp (red color) and HaUsp_ΔNTD (dark green color) show a broad, asymmetric peak and a plateau, which is typical of partially disordered proteins. Representative scattering curve of globular myoglobin (id: SASDAH2) is shown in black. (B) Guinier plots of different concentrations of HaUsp. The estimated R_g value is given next to each curve in the corresponding color. (C) Guinier plots of different concentrations of HaUsp_ΔNTD. The estimated R_g value is given next to each curve in the corresponding color. (D) Pair-distance distribution function characterizing the HaUsp protein. The shape of the $P(r)$ function is characteristic of elongated proteins. (E) Pair-distance distribution function characterizing the HaUsp_ΔNTD protein. The shape of the $P(r)$ function is characteristic of elongated proteins.

between the theoretical values obtained for a monomer (MW_{seq} = 42,347 Da) and a dimer (MW_{seq} = 84,695 Da) (Table 6), and the value increased to the theoretical value obtained for a dimer at 8.0 mg/ml. However, for HaUsp, the MW calculated at 1.0 mg/ml was close to the theoretical value obtained for a dimer (MW_{seq} = 96,684 Da), and with increasing concentration, this value increased further. These results suggested that when the lowest concentration of HaUsp_ΔNTD was measured, there was probably a mixture of monomers and dimers in the

sample, while in the sample of HaUsp, dimers were more abundant than monomers.

Next, *ab initio* low-resolution models of both proteins were created using the DAMMIN program [37], assuming dimeric states (Fig. 9). Data from the 8.0 mg/ml samples were used, and the models obtained and visualized with the Chimera program [77] exhibited good fits with the experimental data (not shown), and in both cases, the models were characterized by elongated shapes, suggesting the formation of dimers

Table 6
Radius of gyration (R_g), maximum intramolecular distance (D_{max}) and molecular weight (MW) calculated for different protein concentrations.

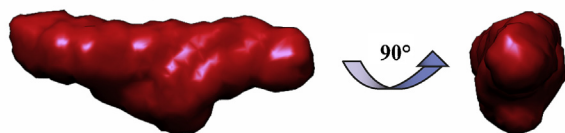| Protein	c [mg/ml]	R_g [Å]	D_{max} [Å]	MW [kDa]	Monomer MW_{seq} [Da]
HaUsp	8.0	50.23 ± 0.21	195.0	110.00	48,342
	4.0	48.07 ± 0.33	186.2	101.65	
	2.0	45.95 ± 0.39	179.0	87.52	
	1.0	44.17 ± 0.87	176.0	98.71	
HaUsp_ΔNTD	8.0	41.99 ± 0.21	173.0	83.87	42,347
	4.0	40.45 ± 0.22	168.5	76.55	
	2.0	38.30 ± 0.80	155.0	70.22	
	1.0	37.48 ± 1.89	147.0	59.93	

HaUsp

HaUsp_ΔNTD

Fig. 9. *Ab initio* low-resolution models of HaUsp and HaUsp_ΔNTD. Low-resolution structures of HaUsp and HaUsp_ΔNTD modeled with the DAMMIN program. Models were created assuming p2 symmetry, and the subsequently obtained results were visualized with the Chimera program [77]. The obtained structures indicated elongated shapes of both proteins with probable dimers of LBDs in the center.

with LBDs in the center. This observation was consistent with the *P(r)* functions, which also indicated elongated shapes of both forms of HaUsp. To gain deeper insight into the structures of the formed oligomers, more reliable models of HaUsp and HaUsp_ΔNTD were developed. The dimer of the LBD was generated from residues 169–424 of the I-TASSER model, and subsequently, the two LBDs were superposed over the structure of the *D. melanogaster* LBD (PDB ID: 1HG4). The DBD was prepared from residues 69–141 of the same I-TASSER model. Next, these atomic models were used for modeling HaUsp and HaUsp_ΔNTD structures using EOM. The components of the proteins chains that were not represented by the selected atomic models were represented as dummy-atom residues [40,43]. A population of ten thousand independent models was generated using the RanCh program [43] in the ATSAS package [35]. Half the population of models was generated without defined local symmetry (p1), and the other half possessed local p2 symmetry. Subsequently, these models were used as a starting pool in the GAJOE program (version 2.1) [43] in the ATSAS package; this program selects a pool of models that best fit the experimental data. As experimental data, the results from the 8.0 mg/ml samples were used for HaUsp_ΔNTD, and those from the 4.0 mg/ml sample were used for HaUsp, because at these concentrations, we obtained good fits to the experimental data - χ^2 = 0.965 for HaUsp and χ^2 = 1.213 for HaUsp_ΔNTD. Interestingly, for 8.0 mg/ml HaUsp, a good fit could be obtained only when the pool containing a mixture of dimers and higher conformers was used. The best results were obtained with pools of dimers and trimers, dimers and tetramers or dimers and pentamers (Fig. S6, Fig. S7 and Fig. S8, respectively). Therefore, the oligomeric state formed at the highest concentration cannot be clearly identified. Nonetheless, these results clearly indicate that in contrast to HaUsp_ΔNTD, which appears to form dimeric complexes at most, full-

length HaUsp has a tendency to form higher-order oligomers as well. The fits of the best profiles determined by EOM modeling of the SAXS data are presented in Fig. 10A (HaUsp) and 10B (HaUsp_ΔNTD). The histograms of the R_g values, characterizing the initial pools of models and selected frequencies of models after minimization, are presented in Fig. 10C (HaUsp) and 10D (HaUsp_ΔNTD). As a result of EOM modeling, we obtained depictions of the types of conformations potentially adopted by the HaUsp and HaUsp_ΔNTD dimers, which are shown in Fig. 11A and 11B. The obtained results showed that the NTD, which is intrinsically disordered and therefore pliable and without stable structure, is spaced from the core of the HaUsp dimer and adopts different conformations that are equally acceptable in terms of their proportion in the total population. However, surprisingly, the model that had the highest contribution to the final pool (44.44%) (supplementary data) was a model in which intrinsically disordered NTDs bend towards the respective LBDs, forming a structure that resembles the bended tail of a scorpion (*scorpion-like* structure). In the case of HaUsp_ΔNTD, the model that had the highest contribution to the final pool (42.57%) (supplementary data) was similar to the most frequent model of HaUsp; however, this model lacked the bended NTD. This proximity of the NTD and LBD in the *scorpion-like* structure may result in the interaction of these domains, which may lead to the rearrangement of the LBD, manifesting itself in the ability of the full-length HaUsp to form dimers with increased stability. Presumably, in the *scorpion-like* structure, some changes also occur in the DBD, which in turn affects the interaction of the protein with specific DNA sequences.

4. Discussion

The NR family consists of transcription factors that play crucial regulatory roles in the transcription process and signaling pathways. NRs are important for proper development, growth, procreation, cell differentiation, proliferation, apoptosis, homeostasis and many other cellular processes [78]. The molecular structures of NRs are highly preserved and include two globular domains: DBD and LBD. In addition to these domains, NRs also contain fragments that do not possess well-preserved structures, such as NTD and the hinge located between DBD and LBD [9]. The NTD and hinge are the most variable fragments of NRs in terms of length and composition of the polypeptide chain as well as three-dimensional structure. Often, flexible disordered NTDs determine the functional identity of different isoforms of the same protein [16] and can modulate the functions of the highly conserved DBDs and LBDs [13]. In fact, disordered regions are more frequently present on proteins termini than in the center of polypeptide chains and it has been postulated that disordered proteins termini evolved to preform significant and specific functions [79]. However, knowledge of the molecular properties of invertebrate NTDs and how these NTDs are associated with NR function remains limited and comes from only a few studies [16,17,55].

In this study, we asked whether intrinsically disordered NTDs may influence the molecular properties of the remaining domains of a representative member of invertebrate NRs. As a model invertebrate NR, we consider HaUsp, and the role of the NTD was examined for two well-known and easily testable NR functions: interactions with specific DNA sequences and dimerization. To determine whether HaUsp NTD is intrinsically disordered and to eventually determine the extent of this disorder, we performed a number of *in silico* analyses. The results suggested that the NTD was characterized by a high tendency for disorder and high flexibility of the backbone. Furthermore, when the NTD alone was applied to the charge-hydropathy plot, it was clearly localized in the disordered protein compartment, while HaUsp and HaUsp_ΔNTD were localized in the ordered protein compartment. The results of the *in silico* analyses were subsequently confirmed with CD spectroscopy, and even though the obtained CD spectrum had a shape that was typical for ordered proteins, deconvolution of this spectrum indicated that the protein contained almost 25% disordered regions.
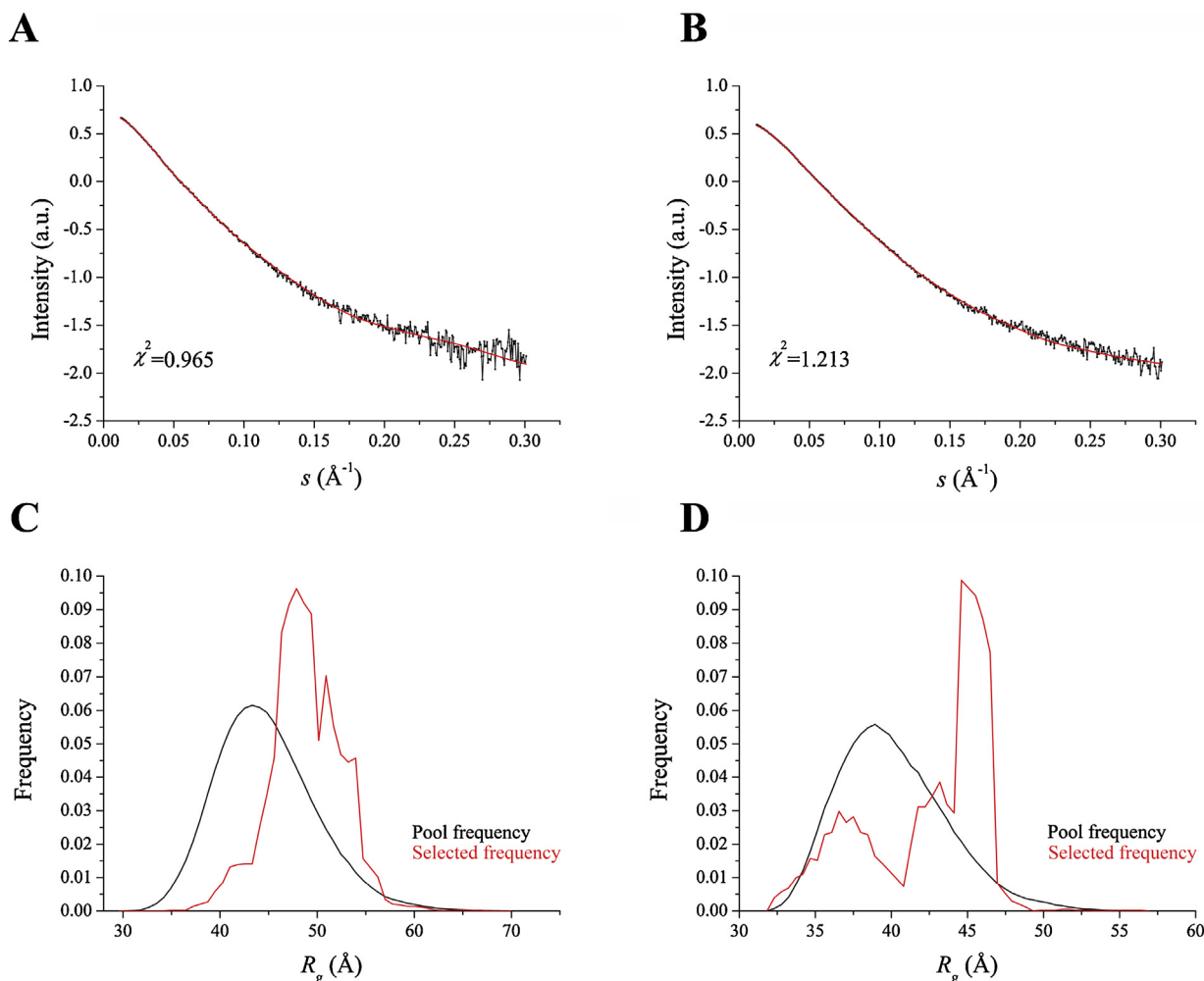
A

$\chi^2 = 0.965$

B

$\chi^2 = 1.213$

C

Pool frequency
Selected frequency

D

Pool frequency
Selected frequency

Fig. 10. Modeling of HaUsp and HaUsp_ΔNTD. (A) The fit (red) of the best profiles determined by EOM modeling of the SAXS data (black) of HaUsp. (B) The fit (red) of the best profiles determined by EOM modeling of the SAXS data (black) of HaUsp_ΔNTD. (C) Distribution of the R_g profile of the starting population of models (black) and profiles of the final optimized ensemble (red) of HaUsp. (D) Distribution of the R_g profile of the starting population of models (black) and profiles of the final optimized ensemble (red) of HaUsp_ΔNTD.

HDX-MS analysis showed that the NTD had the highest tendency for rapid hydrogen exchange, and the lowest rate of deuteration was observed in the peptides corresponding to the DBD and LBD. These data indicate that these two domains are globular and have poorly accessible exchangeable amide hydrogens, while the NTD is disordered and has accessible hydrogens. These results demonstrated that HaUsp has a typical NR molecular structure, with a relatively long intrinsically disordered NTD. Therefore, HaUsp is a good model protein for further molecular analyses of the putative effect of NTDs on NR conformation and function.

The EMSA performed revealed that the intrinsically disordered NTD, consisting of 58 residues, significantly influences the DNA-binding properties of HaUsp. Our results suggested that the presence of the NTD decreases the stability of oligomers and significantly affects the pattern of protein-DNA interaction, as evidenced by the lower intensity of the bands obtained for full-length HaUsp than those obtained for HaUsp_ΔNTD using the same DNA probe. Surprisingly, DmUsp, which contains an NTD, is characterized by a similar interaction pattern as that observed for HaUsp lacking the NTD. This observation clearly suggests that the NTD of DmUsp plays a different structural role than the NTD of the homologous HaUsp. This result is consistent with the idea that intrinsically disordered NTDs are structural elements that determine the specificity of cognate NRs. We investigated two different DNA sequences; the first sequence was *hsp27_EcRE* from *D.*

melanogaster, which has already been shown to bind to Usp proteins from other organisms [26]. The second EcRE was *HR3_EcRE*, which has been shown to bind to *H. armigera* Usp [33]. For both sequences, at higher protein concentrations, bands, characterized with less mobility, corresponding to potential dimers, could be observed. This result is similar to the results obtained for *D. melanogaster* Usp, for which it has been shown that increasing concentrations of the protein favor dimer formation on *hsp27_EcRE* [26]. However, in this study, bands corresponding to the dimer were weak, and in the case of HaUsp, these bands were often not observed. This result could be due to differences in the hinge region, since this region is known to influence DNA-binding [80,81], and BLAST analysis [82] of the HaUsp (19 residues) and DmUsp hinge regions (58 residues) revealed no matches. However, even though HaUsp and HaUsp_ΔNTD have the same hinge, some significant differences in their interactions with DNA were observed. We suggest that the NTD of HaUsp can either suppress protein-DNA interaction in the absence of additional cofactors or can induce the formation of many different, unstable complexes by full-length HaUsp; these complexes are often higher oligomers, which yield barely visible bands.

Two independent techniques, AUC and SEC-MALS, were used to determine whether the intrinsically disordered NTD is important for the dimerization process, which is known to be dependent on globular LBD. Both methods provided similar results, indicating that HaUsp can undergo reversible self-association, either as the full-length protein or
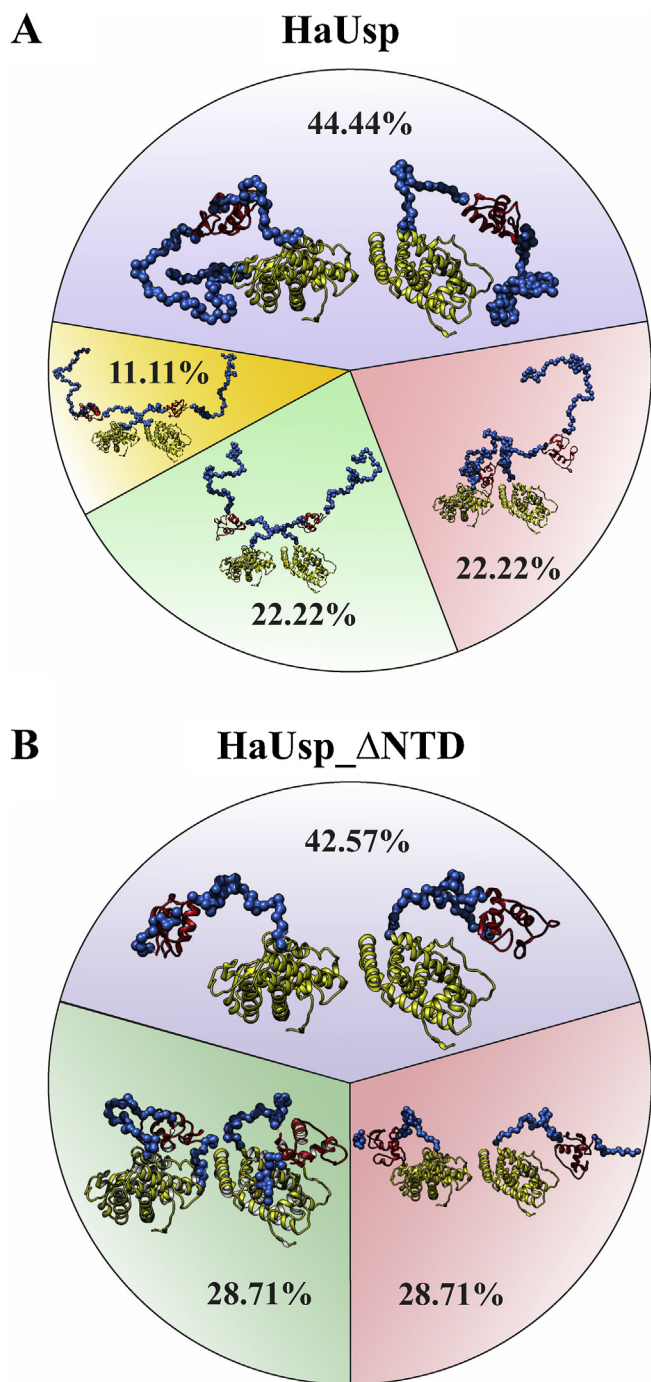
A HaUsp

B HaUsp_ΔNTD

Fig. 11. Representation of models of HaUsp and HaUsp_ΔNTD obtained with EOM. (A) The population of EOM models potentially adopted by the HaUsp is presented as a distribution of frequencies on a pie chart. Specific models were characterized with the following R_g values (starting from the most frequent and moving clockwise): 48.05 Å, 51.30 Å, 50.31 Å, 52.54 Å. (B) The population of EOM models potentially adopted by the HaUsp_ΔNTD is presented as a distribution of frequencies on a pie chart. Specific models were characterized with the following R_g values (starting from the most frequent and moving clockwise): 45.64 Å, 45.67 Å, 36.60 Å. DBDs are indicated in red, and LBDs are indicated in yellow.

without the NTD. SEC-MALS provided chromatograms that exhibited significant elution peak shifts to smaller elution volumes with increasing concentrations, which is a hallmark of reversible self-association [70]. SV-AUC-SV, used as an analogous method, also indicated evidence of self-association [69]. However, both techniques also

provided results that suggest a significant influence of the intrinsically disordered NTD on dimer stabilization, as evidenced by the presence of additional peaks corresponding to such oligomers. Furthermore, when we examined both forms with SE-AUC, we were able to determine the K_d value of the dimerization process, which was 31.7 μM for full-length HaUsp and 46.4 μM for HaUsp_ΔNTD. The obtained values clearly indicated that the NTD is crucial for stabilization of the full-length HaUsp dimer since deletion of this domain resulted in decreased affinity.

SAXS, which is a technique that allows the determination of biophysical parameters and overall structures of molecules in solution, confirmed the previous observations. SAXS analysis confirmed the role of the intrinsically disordered NTD in dimer formation. Furthermore, the obtained data allowed the generation of models that provide a molecular explanation for previous observations. Consistent with these models, formation of a *scorpion-like* structure in the case of full-length HaUsp may explain the stabilization of the dimers. The obtained Kratky plot was consistent with previous CD spectroscopy results, indicating that HaUsp is a partially disordered protein. Subsequently, the determined R_g and calculated *MW* matched SV-AUC and SEC-MALS observations, suggesting that the dimerization of full-length HaUsp and HaUsp_ΔNTD is concentration dependent. However, we obtained *MW* values corresponding to potential dimers at lower concentrations of HaUsp than HaUsp_ΔNTD, indicating that the presence of the NTD somehow facilitated dimer formation. The *P(r)* function indicated that both proteins are characterized by elongated shapes, which was confirmed with a low-resolution model prepared with the DAMMIN program. Based on these models and previous studies [83], we assumed that the centers of these dimers contain dimers of LBDs. Based on this assumption, we created reliable models with the EOM program and obtained structural parameters of the proteins and a possible explanation for dimer stabilization. First, the radii of gyration of the most frequent models for each protein were close to the average R_g obtained with the Guinier method - 48.07 Å (Guinier method) and 48.05 Å (model) for HaUsp and 41.99 Å (Guinier method) and 45.64 Å (model) for HaUsp_ΔNTD. Therefore, these models can be considered to be good representatives of the whole population of molecules in the solution. Second, as presented in Fig. 11, the most frequent models of each protein were similar, differing only in the presence of the NTD in full-length HaUsp, which creates a loop that bends towards the LBD, forming a *scorpion-like* structure of each subunit. Based on these results, we suggest that NTD stabilizes the dimers by interacting with specific fragments of the LBD, leading to conformational changes in this domain, which may in turn increase the exposed surface area for interactions between the two LBDs. An additional observation that indicates that the NTD clearly interacts with the LBD is that activation function 1 (AF-1) is located in the NTD and AF-2 is located in the LBD, and AF-1 and AF-2 have already been suggested to interact with each other [84,85]. Considering the observed differences in the interactions of HaUsp and HaUsp_ΔNTD with HREs, we hypothesize that the formation of more stable dimers in solution by HaUsp can apparently affect the protein-DNA interaction pattern. It is possible that a coactivator or another compound is necessary to weaken the interaction between the NTD and LBD to allow the formation of defined protein-DNA complexes. In the case of HaUsp_ΔNTD, there is no NTD present, so other compounds are not necessary for the formation of such complexes.

In summary, in this paper, we present complex studies of HaUsp, with an emphasis on the NTD, examining the potential influence of the disordered nature of this domain on the molecular properties, and hence the functions, of the remaining domains of the NR. As revealed by *in silico* analysis, CD spectroscopy and HDX, HaUsp may be a good model of NRs for studies of structure-function relationships. Our results clearly demonstrate that the intrinsically disordered NTD is an important structural element that strongly affects interactions of HaUsp with specific HREs and affects the dimerization of HaUsp. We were able to study the structure of HaUsp in solution and construct reliable structural models of HaUsp dimers based on SAXS data. All the obtained

results allowed us to conclude that full-length HaUsp has a tendency to form more stable dimers in solution, which can be a result of the formation of *scorpion-like* structures. We hypothesize that the function of disordered NTDs is to stabilize HaUsp dimers via the *scorpion-like* structure, and this can also influence the interaction of the protein with DNA. This function may be crucial for more rigorous regulation of the main function of NRs, which is the regulation of transcriptional processes. Further studies should focus on determination of the association and dissociation rate constants since we suggested that there is rapid reversible self-association. Additionally, it is necessary to investigate the characteristics of the interaction between the NTD and LBD and map the amino acid residues that may interact with each other.

Acknowledgments

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jsbmb.2018.06.011.

References

[1] V. Laudet, H. Gronemeyer, The Nuclear Receptor : Factsbook, Academic Press, San Diego, 2002.

[2] J.W. Schwabe, L. Chapman, J.T. Finch, D. Rhodes, The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements, Cell 75 (1993) 567–578.

[3] M. Vogtli, C. Elke, M.O. Imhof, M. Lezzi, High level transactivation by the ecdysone receptor complex at the core recognition motif, Nucleic Acids Res. 26 (1998) 2407–2414.

[4] Y. Nakagawa, V.C. Henrich, Arthropod nuclear receptors and their role in molting, FEBS J. 276 (2009) 6128–6157.

[5] W. Bourguet, M. Ruff, P. Chambon, H. Gronemeyer, D. Moras, Crystal structure of the ligand-binding domain of the human nuclear receptor RXR-alpha, Nature 375 (1995) 377–382.

[6] M.S. Lee, S.A. Kliewer, J. Provencal, P.E. Wright, R.M. Evans, Structure of the retinoid X receptor alpha DNA binding domain: a helix required for homodimeric DNA binding, Science 260 (1993) 1117–1121.

[7] G.M. Clayton, S.Y. Peak-Chew, R.M. Evans, J.W. Schwabe, The structure of the ultraspiracle ligand-binding domain reveals a nuclear receptor locked in an inactive conformation, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 1549–1554.

[8] I.M. Billas, L. Moulinier, N. Rochel, D. Moras, Crystal structure of the ligand-binding domain of the ultraspiracle protein USP, the ortholog of retinoid X receptors in insects, J. Biol. Chem. 276 (2001) 7465–7474.

[9] D.N. Lavery, I.J. McEwan, Structural characterization of the native NH2-terminal transactivation domain of the human androgen receptor: a collapsed disordered conformation underlies structural plasticity and protein-induced folding, Biochemistry 47 (2008) 3360–3369.

[10] M.D. Krasowski, E.J. Reschly, S. Ekins, Intrinsic disorder in nuclear hormone receptors, J. Proteome Res. 7 (2008) 4359–4372.

[11] I.J. McEwan, D. Lavery, K. Fischer, K. Watt, Natural disordered sequences in the amino terminal domain of nuclear receptors: lessons from the androgen and glucocorticoid receptors, Nuclear Recept Signal 5 (2007) e001.

[12] B. He, S. Bai, A.T. Hnat, R.I. Kalman, J.T. Minges, C. Patterson, E.M. Wilson, An androgen receptor NH2-terminal conserved motif interacts with the COOH terminus of the Hsp70-interacting protein (CHIP), J. Biol. Chem. 279 (2004) 30643–30653.

[13] B. He, J.T. Minges, L.W. Lee, E.M. Wilson, The FXXLF motif mediates androgen receptor-specific interactions with coregulators, J. Biol. Chem. 277 (2002) 10226–10235.

[14] J.B. Pippal, Y. Yao, F.M. Rogerson, P.J. Fuller, Structural and functional characterization of the interdomain interaction in the mineralocorticoid receptor, Mol. Endocrinol. 23 (2009) 1360–1370.

[15] R. Kumar, E.B. Thompson, Influence of flanking sequences on signaling between the activation function AF1 and DNA-binding domain of the glucocorticoid receptor, Arch. Biochem. Biophys. 496 (2010) 140–145.

[16] M. Nocula-Lugowska, G. Rymarczyk, M. Lisowski, A. Ozyhar, Isoform-specific variation in the intrinsic disorder of the ecdysteroid receptor N-terminal domain,

[17] J. Pieprzyk, A. Zbela, M. Jakób, A. Ożyhar, M. Orłowski, Homodimerization propensity of the intrinsically disordered N-terminal domain of Ultraspiracle from Aedes aegypti, Biochim. Biophys. Acta 1844 (2014) 1153–1166.

[18] L.M. Riddiford, Juvenile hormone action: a 2007 perspective, J. Insect. Physiol. 54 (2008) 895–901.

[19] M.X. Huang, J. Du, B.J. Su, G.D. Zhao, W.D. Shen, Z.G. Wei, The expression profile and promoter analysis of ultraspiracle gene in the silkworm Bombyx mori, Mol. Biol. Rep. 41 (2014) 7955–7965.

[20] J. Zhu, K. Miura, L. Chen, A.S. Raikhel, Cyclicity of mosquito vitellogenic ecdysteroid-mediated signaling is modulated by alternative dimerization of the RXR homologue Ultraspiracle, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 544–549.

[21] J.D. Sutherland, T. Kozlova, G. Tzertzinis, F.C. Kafatos, Drosophila hormone receptor 38: a second partner for Drosophila USP suggests an unexpected role for nuclear receptors of the nerve growth factor-induced protein B type, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 7966–7970.

[22] T.P. Yao, W.A. Segraves, A.E. Oro, M. McKeown, R.M. Evans, Drosophila ultraspiracle modulates ecdysone receptor function via heterodimer formation, Cell 71 (1992) 63–72.

[23] G. Tzertzinis, A. Malecki, F.C. Kafatos, BmCF1, a Bombyx mori RXR-type receptor related to the Drosophila ultraspiracle, J. Mol. Biol. 238 (1994) 479–486.

[24] S.F. Wang, C. Li, J. Zhu, K. Miura, R.J. Miksicek, A.S. Raikhel, Differential expression and regulation by 20-hydroxyecdysone of mosquito ultraspiracle isoforms, Dev. Biol. 218 (2000) 99–113.

[25] S.C. Gill, P.H. von Hippel, Calculation of protein extinction coefficients from amino acid sequence data, Anal. Biochem. 182 (1989) 319–326.

[26] G. Rymarczyk, I. Grad, A. Rusek, K. Oswiecimska-Rusin, A. Niedziela-Majka, M. Kochman, A. Ozyhar, Purification of Drosophila melanogaster ultraspiracle protein and analysis of its A/B region-dependent dimerization behavior in vitro, Biol. Chem. 384 (2003) 59–69.

[27] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 227 (1970) 680–685.

[28] G. Fairbanks, T.L. Steck, D.F. Wallach, Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane, Biochemistry 10 (1971) 2606–2617.

[29] S.M. Kelly, T.J. Jess, N.C. Price, How to study proteins by circular dichroism, Biochim. Biophys. Acta 1751 (2005) 119–139.

[30] N. Sreerama, R.W. Woody, Computation and analysis of protein circular dichroism spectra, Methods Enzymol. 383 (2004) 318–351.

[31] M. Fried, D.M. Crothers, Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis, Nucleic Acids Res. 9 (1981) 6505–6525.

[32] A. Ożyhar, M. Strangmann-Diekmann, H.H. Kiltz, O. Pongs, Characterization of a specific ecdysteroid receptor-DNA complex reveals common properties for invertebrate and vertebrate hormone-receptor/DNA interactions, Eur. J. Biochem. 200 (1991) 329–335.

[33] W. Liu, M.J. Cai, C.C. Zheng, J.X. Wang, X.F. Zhao, Phospholipase Cgamma1 connects the cell membrane pathway to the nuclear receptor pathway in insect steroid hormone signaling, J. Biol. Chem. 289 (2014) 13026–13041.

[34] R.V.G. Girardot, J. Perez, M. Ounsy, FOXTROT: a JAVA-Based Application to Reduce and Analyse SAXS and WAXS Piles of 2D Data at Synchrotron Soleil, CanSAS-VIII Abstract, J-PARC, Tokai, Japan, (2015).

[35] D. Franke, M.V. Petoukhov, P.V. Konarev, A. Panjkovich, A. Tuukkanen, H.D.T. Mertens, A.G. Kikhney, N.R. Hajizadeh, J.M. Franklin, C.M. Jeffries, D.I. Svergun, ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions, J. Appl. Crystallogr. 50 (2017) 1212–1225.

[36] D.I. Svergun, Determination of the regularization parameter in indirect-transform methods using perceptual criteria, J. Appl. Crystallogr. 25 (1992) 495–503.

[37] D.I. Svergun, Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing, Biophys. J. 76 (1999) 2879–2886.

[38] V.V.S. Volkov, D.I. Svergun, Uniqueness of ab-initio shape determination in small-angle scattering, J. Appl. Cryst. 36 (2003) 860–864.

[39] Y. Zhang, I-TASSER server for protein 3D structure prediction, BMC Bioinform. 9 (2008) 40.

[40] P. Bernado, E. Mylonas, M.V. Petoukhov, M. Blackledge, D.I. Svergun, Structural characterization of flexible proteins using small-angle X-ray scattering, J. Am. Chem. Soc. 129 (2007) 5656–5664.

[41] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, The I-TASSER suite: protein structure and function prediction, Nat. Methods 12 (2015) 7–8.

[42] D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, H.J. Wolfson, PatchDock and SymmDock: servers for rigid and symmetric docking, Nucleic Acids Res. 33 (2005) W363–367.

[43] G. Tria, H.D. Mertens, M. Kachala, D.I. Svergun, Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering, IUCrJ 2 (2015) 207–217.

[44] X. Li, P. Romero, M. Rani, A.K. Dunker, Z. Obradovic, Predicting protein disorder for N-, C-, and internal regions, Genome Inf. Ser. Workshop Genome Inf. 10 (1999) 30–40.

[45] Romero, Obradovic, K. Dunker, Sequence data analysis for Long disordered regions prediction in the calcineurin family, Genome Inf. Ser. Workshop Genome Inf. 8 (1997) 110–124.

[46] E. Cilia, R. Pancsa, P. Tompa, T. Lenaerts, W.F. Vranken, From protein sequence to dynamics and disorder with DynaMine, Nat. Commun. 4 (2013) 2741.

[47] E. Cilia, R. Pancsa, P. Tompa, T. Lenaerts, W.F. Vranken, The DynaMine webserver: predicting protein dynamics from sequence, Nucleic Acids Res. 42 (2014) W264–270.

[48] L.P. Kozlowski, J.M. Bujnicki, MetaDisorder: a meta-server for the prediction of

intrinsic disorder in proteins, BMC Bioinform. 13 (2012) 111.

[49] Z. Dosztanyi, B. Meszaros, I. Simon, ANCHOR: web server for predicting protein binding regions in disordered proteins, Bioinformatics 25 (2009) 2745–2746.

[50] A. Marchler-Bauer, S.H. Bryant, CD-search: protein domain annotations on the fly, Nucleic Acids Res. 32 (2004) W327–331.

[51] A. Marchler-Bauer, S. Lu, J.B. Anderson, F. Chitsaz, M.K. Derbyshire, C. DeWeese-Scott, J.H. Fong, L.Y. Geer, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, J.D. Jackson, Z. Ke, C.J. Lanczycki, F. Lu, G.H. Marchler, M. Mullokandov, M.V. Omelchenko, C.L. Robertson, J.S. Song, N. Thanki, R.A. Yamashita, D. Zhang, N. Zhang, C. Zheng, S.H. Bryant, CDD: a conserved domain database for the functional annotation of proteins, Nucleic Acids Res. 39 (2011) D225–229.

[52] A. Marchler-Bauer, M.K. Derbyshire, N.R. Gonzales, S. Lu, F. Chitsaz, L.Y. Geer, R.C. Geer, J. He, M. Gwadz, D.I. Hurwitz, C.J. Lanczycki, F. Lu, G.H. Marchler, J.S. Song, N. Thanki, Z. Wang, R.A. Yamashita, D. Zhang, C. Zheng, S.H. Bryant, CDD: NCBI's conserved domain database, Nucleic Acids Res. 43 (2015) D222–226.

[53] A. Marchler-Bauer, Y. Bo, L. Han, J. He, C.J. Lanczycki, S. Lu, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, F. Lu, G.H. Marchler, J.S. Song, N. Thanki, Z. Wang, R.A. Yamashita, D. Zhang, C. Zheng, L.Y. Geer, S.H. Bryant, CDD/SPARCLE: functional classification of proteins via subfamily domain architectures, Nucleic Acids Res. 45 (2017) D200–D203.

[54] A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction, Nat. Protoc. 5 (2010) 725–738.

[55] A. Dziedzic-Letka, G. Rymarczyk, T.M. Kaplon, A. Gorecki, A. Szamborska-Gbur, M. Wojtas, P. Dobryszycki, A. Ozyhar, Intrinsic disorder of Drosophila melanogaster hormone receptor 38 N-terminal domain, Proteins 79 (2011) 376–392.

[56] A.K. Dunker, C.J. Brown, J.D. Lawson, L.M. Iakoucheva, Z. Obradovic, Intrinsic disorder and protein function, Biochemistry 41 (2002) 6573–6582.

[57] P. Tompa, Intrinsically unstructured proteins, Trends Biochem. Sci. 27 (2002) 527–533.

[58] R.J. Weatheritt, T.J. Gibson, Linear motifs: lost in (pre)translation, Trends Biochem. Sci. 37 (2012) 333–341.

[59] P.R. Romero, S. Zaidi, Y.Y. Fang, V.N. Uversky, P. Radivojac, C.J. Oldfield, M.S. Cortese, M. Sickmeier, T. LeGall, Z. Obradovic, A.K. Dunker, Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 8390–8395.

[60] S. Devarakonda, J.M. Harp, Y. Kim, A. Ozyhar, F. Rastinejad, Structure of the heterodimeric ecdysone receptor DNA-binding complex, EMBO J. 22 (2003) 5827–5840.

[61] M. Jakob, R. Kolodziejczyk, M. Orlowski, S. Krzywda, A. Kowalska, J. Dutko-Gwozdz, T. Gwozdz, M. Kochman, M. Jaskolski, A. Ozyhar, Novel DNA-binding element within the C-terminal extension of the nuclear receptor DNA-binding domain, Nucleic Acids Res. 35 (2007) 2705–2718.

[62] P. Romero, Z. Obradovic, X. Li, E.C. Garner, C.J. Brown, A.K. Dunker, Sequence complexity of disordered protein, Proteins 42 (2001) 38–48.

[63] V.N. Uversky, J.R. Gillespie, A.L. Fink, Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 41 (2000) 415–427.

[64] S.M. Kelly, N.C. Price, The application of circular dichroism to studies of protein folding and unfolding, Biochim. Biophys. Acta 1338 (1997) 161–185.

[65] W.C. Johnson Jr, Secondary structure of proteins through circular dichroism

[66] spectroscopy, Ann. Rev. Biophys. Biophys. Chem. 17 (1988) 145–166.

[66] M.J. Chalmers, S.A. Busby, B.D. Pascal, M.R. Southern, P.R. Griffin, A two-stage differential hydrogen deuterium exchange method for the rapid characterization of protein/ligand interactions, J. Biomol. Tech. 18 (2007) 194–204.

[67] W.L. DeLano, The PyMOl User's Manual, DeLanon Scientific, San Carlo, 2002.

[68] M. Schubiger, J.W. Truman, The RXR ortholog USP suppresses early metamorphic processes in Drosophila in the absence of ecdysteroids, Development 127 (2000) 1151–1159.

[69] P.H. Brown, A. Balbo, P. Schuck, Characterizing protein-protein interactions by sedimentation velocity analytical ultracentrifugation, Curr. Protoc. Immunol. (2008) 15 Chapter 18 (2008) Unit 18.

[70] T.F. Lerch, P. Sharpe, S.J. Mayclin, T.E. Edwards, E. Lee, H.D. Conlon, S. Polleck, J.C. Rouse, Y. Luo, Q. Zou, Infliximab crystal structures reveal insights into self-association, MAbs 9 (2017) 874–883.

[71] S.A. Ali, N. Iwabuchi, T. Matsui, K. Hirota, S. Kidokoro, M. Arai, K. Kuwajima, P. Schuck, F. Arisaka, Reversible and fast association equilibria of a molecular chaperone, gp57A, of bacteriophage T4, Biophys. J. 85 (2003) 2606–2618.

[72] H. Zhao, A. Balbo, P.H. Brown, P. Schuck, The boundary structure in the analysis of reversibly interacting systems by sedimentation velocity, Methods 54 (2011) 16–30.

[73] R.B. Gillis, A.J. Rowe, G.G. Adams, S.E. Harding, A review of modern approaches to the hydrodynamic characterisation of polydisperse macromolecular systems in biotechnology, Biotechnol. Genet. Eng. Rev. 30 (2014) 142–157.

[74] V. Receveur-Brechot, D. Durand, How random are intrinsically disordered proteins? A small angle scattering perspective, Curr. Protein Pept. Sci. 13 (2012) 55–75.

[75] P. Bernado, D.I. Svergun, Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering, Mol. Biosyst. 8 (2012) 151–167.

[76] E.S. Mylonas, D.I. Svergun, Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering, J. Appl. Crystallogr. (2007) 40.

[77] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera–a visualization system for exploratory research and analysis, J Comput. Chem. 25 (2004) 1605–1612.

[78] J.P. Renaud, D. Moras, Structural studies on nuclear receptors, Cell. Mol. Life Sci. 57 (2000) 1748–1769.

[79] V.N. Uversky, The most important thing is the tail: multitudinous functionalities of intrinsically disordered protein termini, FEBS Lett. 587 (2013) 1891–1901.

[80] A. Haelens, T. Tanner, S. Denayer, L. Callewaert, F. Claessens, The hinge region regulates DNA binding, nuclear translocation, and transactivation of the androgen receptor, Cancer Res. 67 (2007) 4514–4523.

[81] S. Khorasanizadeh, F. Rastinejad, Nuclear-receptor interactions on DNA-response elements, Trends Biochem. Sci. 26 (2001) 384–390.

[82] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.

[83] A. Tamrazi, K.E. Carlson, J.R. Daniels, K.M. Hurth, J.A. Katzenellenbogen, Estrogen receptor dimerization: ligand binding regulates dimer affinity and dimer dissociation rate, Mol. Endocrinol. 16 (2002) 2706–2719.

[84] B. He, E.M. Wilson, The NH(2)-terminal and carboxyl-terminal interaction in the human androgen receptor, Mol. Genet Metab. 75 (2002) 293–298.

[85] F.M. Rogerson, P.J. Fuller, Interdomain interactions in the mineralocorticoid receptor, Mol. Cell Endocrinol. 200 (2003) 45–55.