the FEBS Journal

Pliable natural biocide: Jaburetox is an intrinsically disordered insecticidal and fungicidal polypeptide derived from jack bean urease

Fernanda C. Lopes[1], Olena Dobrovolska[2], Rafael Real-Guerra[3], Valquiria Broll[1], Barbara Zambelli[2], Francesco Musiani[2], Vladimir N. Uversky[4,5,6], Célia R. Carlini[1,3,7] and Stefano Ciurli[2]

1 Graduate Program in Cellular and Molecular Biology – Center of Biotechnology, Federal University of Rio Grande do Sul, Porto Alegre, Brazil
2 Laboratory of Bioinorganic Chemistry, Department of Pharmacy and Biotechnology, University of Bologna, Italy
3 Department of Biophysics and Center of Biotechnology, Federal University of Rio Grande do Sul, Porto Alegre, Brazil
4 Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, USA
5 Institute for Biological Instrumentation, Russian Academy of Sciences, Pushchino, Moscow Region, Russia
6 Department of Biological Science, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia
7 Instituto do Cérebro, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brazil

Keywords
Canavalia ensiformis; circular dichroism; Jaburetox; light scattering; NMR spectroscopy; urease

Correspondence
S. Ciurli, Laboratory of Bioinorganic Chemistry, Department of Pharmacy and Biotechnology, University of Bologna, Via Giuseppe Fanin 40, 40127 Bologna, Italy
Fax: +39 051 209 6203
Tel: +39 051 2096204
E-mail: stefano.ciurli@unibo.it
V. N. Uversky, Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA
Fax: +1 813-974-7357
Tel: +1 813-974-5816
E-mail: vuversky@health.usf.edu
C. R. Carlini, Graduate Program in Cellular and Molecular Biology – Center of Biotechnology, Federal University of Rio Grande do Sul, Av. Bento Gonçalves 9500, Porto Alegre, RS, CEP 91501-970, Brazil
Fax: + 55 51 3308 7003
Tel: + 55 51 3308 7606
E-mail: ccarlini@ufrgs.br

(Received 9 October 2014, revised 13 January 2015, accepted 14 January 2015)

doi:10.1111/febs.13201

Jaburetox is a polypeptide derived from jack bean (*Canavalia ensiformis*) urease and toxic to a broad spectrum of insects, phytopathogenic filamentous fungi and yeasts of medical importance. The elucidation of the structural basis for the mode of action of Jaburetox is the focus of this multifaceted study. Jaburetox in solution is a monomer of 11.0 kDa featuring a large hydrodynamic radius, suggestive of a disordered polypeptide. The intrinsically disordered nature of Jaburetox was theoretically predicted by a comprehensive bioinformatics analysis and experimentally confirmed by light scattering as well as by circular dichroism and NMR spectroscopy. NMR signal assignment provided backbone secondary chemical shifts that indicated that Jaburetox has a low propensity to assume a stable secondary structure. ^{15}N relaxation studies revealed significant backbone mobility, especially in the N-terminal portion of the polypeptide. The solution structure of Jaburetox shows the presence of an α-helical motif close to the N terminus, together with two turn-like structures situated in the central portion of the protein and close to the C terminus. Similar regions were predicted as potential protein–protein interaction sites using computational tools. The knowledge of the structural properties of Jaburetox in solution is a key step to correlate its structural and biological activities.

Abbreviations
HSQC, heteronuclear single quantum coherence; IDP, intrinsically disordered protein; JBU, jack bean urease; LUV, large unilamellar vesicle; MALS, multiple-angle light scattering; MoRF, molecular recognition feature; QELS, quasi-elastic light scattering; SEC, size exclusion chromatography; SSP, secondary structure propensity; TCEP, tris(2-carboxyethyl)phosphine.

Introduction

Ureases (EC 3.5.1.5) are non-redox nickel dependent enzymes [1] that catalyze urea hydrolysis into ammonia and carbon dioxide [2–5]. Jack bean (*Canavalia ensiformis*) urease (JBU) (Fig. 1A) represents an important milestone in the history of biochemistry. It was the first enzyme ever crystallized [6], and also the first enzyme found to have nickel ions in its active site [7]. In addition to JBU, *C. ensiformis* displays two additional isoforms of urease, named JBURE-II [8,9] and canatoxin [10].

Canatoxin, a toxic protein isolated from *C. ensiformis* seeds [11] and identified 20 years later as an isoform of urease [10], displays insecticidal properties [12]. This toxicity involves an internal polypeptide of 10 kDa (pepcanatox), released from the protein upon hydrolysis by cathepsin-like digestive enzymes of insects such as *Callosobruchus maculatus*, *Rhodnius prolixus*, *Nezara viridula*, *Dysdercus peruvianus* and *Oncopeltus fasciatus* [13–16] (Fig. 1A,B). Based on the N-terminal sequence of pepcanatox and using as a template the cDNA of the urease isoform JBURE-II,

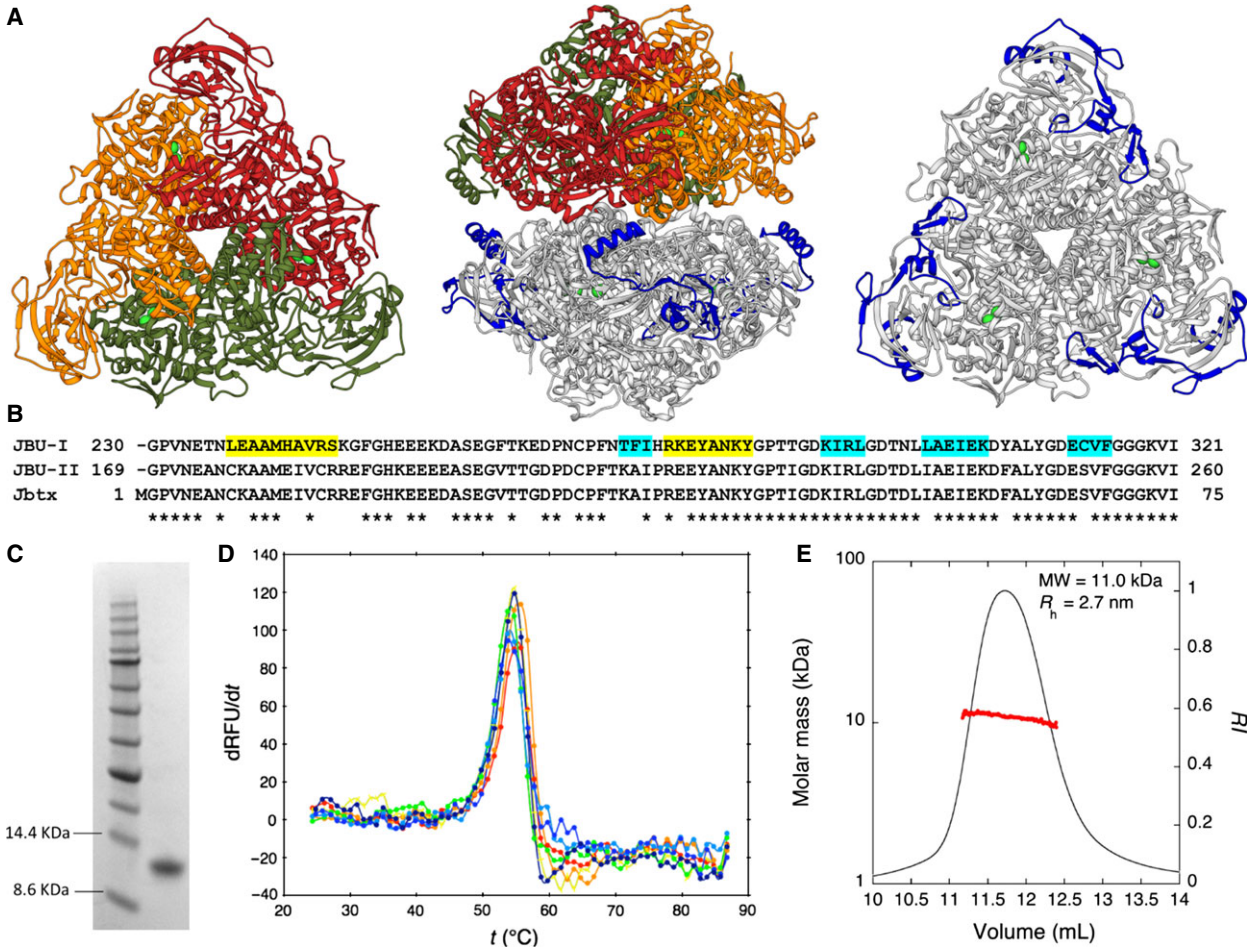


Fig. 1. Purification, stability and hydrodynamics of Jaburetox. (A) Ribbon scheme of the structure of jack bean urease: the α_6 hexamer is shown in the central panel, while the left and right panels display the two trimers that compose the overall protein tilted by 90° along the horizontal axis; the trimer on top is colored red, orange and green for the three α subunits, while the trimer on the bottom highlights in blue the protein portion that corresponds to the Jaburetox polypeptide; the nickel atoms in the active site are shown as green spheres. (B) Sequence alignment of Jaburetox (abbreviated as Jbtx) with the corresponding sequence of the two isoforms of urease in jack bean; α-helices and β-strands in the crystal structure of JBU-I are highlighted in yellow and cyan, respectively. (C) SDS/PAGE profile of the final purified Jaburetox (right lane) and the molecular mass reference markers (left lane). (D) Thermal denaturation assay of Jaburetox using Thermofluor. The plot represents the first derivative of the melting curve of the raw data obtained at pH 3.5 (red), pH 4.5 (orange), pH 5.5 (yellow), pH 6.5 (green), pH 7.5 (cyan), pH 8.5 (blue) and pH 9.5 (indigo). The T_m values correspond to the apex. (E) Molar mass distribution of Jaburetox eluted from a SEC column and evaluated using MALS-QELS. The solid line indicates the trace from the refractive index detector, whereas the dots are the weight-averaged molecular masses measured every second.

the recombinant polypeptide Jaburetox-2Ec (carrying a V5 epitope and a His-tag) was produced heterologously in *Escherichia coli* [17]. Jaburetox-2Ec shows a broader spectrum of insecticidal activity, which also includes insects not susceptible to the full-length canatoxin such as *Spodoptera frugiperda* [17] because the hydrolysis of the protein to release the polypeptide is no longer required. Jaburetox-2Ec causes inhibition of diuresis in *R. prolixus* by a mechanism that involves cGMP and disturbance of the transmembrane potential of Malpighian tubules [18]. Moreover, Jaburetox-2Ec permeabilizes large unilamellar vesicles (LUVs), displaying membrane-disruptive activity on acidic lipid bilayers [19]. A variant form of the recombinant polypeptide, corresponding to the His-tagged urease-derived sequence with no V5 epitope and simply named Jaburetox, also displays antifungal properties against filamentous fungi (*Mucor* sp. and *Penicillium herguei*) and yeasts (*Candida* (*Ca.*) *albicans*, *Ca. parapsilosis*, *Ca. tropicalis*, *Kluyveromyces marxianus*, *Pichia membranisfaciens* and *Saccharomyces cerevisiae*), in addition to its entomotoxic activity [20].

In order to understand the mode of action of Jaburetox-2Ec, molecular modeling studies were carried out that suggested the existence of a β-hairpin motif at the C-terminal portion [17,19]. This motif was indeed subsequently observed in the crystallographic structure of JBU [21], suggesting that it could be a factor for the membrane-disturbing activity of Jaburetox [19,22]. In order to confirm the importance of this β-hairpin for the biological activity of Jaburetox, three mutated versions of this polypeptide were analyzed: one lacking the β-hairpin motif, and two additional peptides corresponding to the N-terminal and the C-terminal portions of the recombinant Jaburetox, respectively [23]. The peptide lacking the β-hairpin motif showed all the properties of the wild-type Jaburetox, thus excluding this region as the biologically active portion of the molecule. On the other hand, injection assays into *O. fasciatus* and *R. prolixus* showed that only the N-terminal portion of Jaburetox carries its insecticidal activity. However, both parts caused almost the same mortality rate seen for the N-terminal peptide upon feeding to *R. prolixus*, indicating different modes of action for the two peptides depending on the different tissues of the insect. Although with different potencies, both the N- and C-terminal portions of Jaburetox caused the neuromuscular blockage of the cockroach *Phoetalia pallida* nerve-coxal muscle preparation, while both peptides were equipotent in inhibiting the fluid secretion in the Malpighian tubules of *R. prolixus* and in disrupting LUV membranes. These data suggested that the N-terminal portion carries the entomotoxic activity of Jaburetox, and that its C terminus probably contributes to the polypeptide activity by interacting with cell membranes [23]. Other studies have demonstrated that Jaburetox and its mutants can form well-resolved, highly cation-selective channels. The peptide corresponding to the N-terminal part of Jaburetox is more active in negative potentials, while the ion-channel activity of Jaburetox and the other mutants do not display voltage dependence [24].

The objective of the present study was the experimental characterization of the structural, dynamic and folding properties of Jaburetox. The hydrodynamic properties of the polypeptide were determined using size exclusion chromatography (SEC) coupled with light scattering experiments. The protein folding was examined using circular dichroism (CD), differential scanning fluorimetry and high resolution NMR spectroscopy. Based on NMR data, backbone mobility studies were carried out and the 3D structure of Jaburetox was determined. The properties that were established indicate the presence of a large ensemble of highly disordered conformers in solution with three more ordered segments.

Results

Jaburetox expression and purification

Jaburetox was expressed and purified as a His-tagged polypeptide in order to avoid additional steps in the purification protocol that might have led to decreased yields. Indeed, the presence of this tag does not interfere with the biological insecticidal [17] and fungicidal [20] activities of this urease-derived peptide. The expression protocol of Jaburetox was optimized in this work to typically yield 30 mg of Jaburetox per liter of culture. The previously described protocol had a protein yield of 10 $mg \cdot L^{-1}$ [20,23]. Screening of different culture conditions allowed a maximal yield of protein expression at 19 °C, 0.87 mM IPTG and 16 h of expression time to be obtained. Moreover, before protein induction, LB medium was exchanged with fresh M9 minimal medium and the biomass was four-fold concentrated, according to a previously reported method [25]. In addition to the advantage of increasing the cellular performance by removing secondary metabolites that might inhibit cellular growth, this method reduces isotope consumption for producing NMR samples: indeed, the majority of cell biomass is generated using unlabeled media, while labeled protein is produced in a reduced volume of isotopically labeled minimal medium. The purification protocol was improved by adding a SEC separation after the first

affinity chromatography purification step. The SDS/PAGE profile of the final purified protein is shown in Fig. 1C. Notably, although the predicted molecular mass determined from the amino acid sequence of Jaburetox with the six-histidine extension is 11 kDa, the protein exhibited a 2 kDa shift in the SDS/PAGE and migrated as a 13 kDa polypeptide. This aberrant migration during SDS/PAGE experiments is a typical feature of intrinsically disordered proteins (IDPs) [26,27], determined by their unusual amino acid compositions typically characterized by a high ratio of charged to hydrophobic residues. As a result, these disordered proteins interact less efficiently with SDS and are characterized by a decreased migration velocity, giving an apparent molecular mass that is higher than the real one [28].

Jaburetox stability by thermal differential scanning fluorimetry

Jaburetox tends to aggregate after long-term storage [19,23], a process that negatively affects the biological activity of this protein [29]. In order to find conditions to stabilize the protein solutions, the effect of pH was explored by performing thermal denaturation assays using differential scanning fluorimetry (Thermofluor [30]). In this assay, a fluorescent probe (SYPRO Orange) added to the protein solution binds to exposed hydrophobic residues, thereby increasing its fluorescence. Maximal fluorescence intensity is expected when the protein unfolds, which corresponds to a peak in the first derivative plot occurring at the melting temperature (T_m). Similar curves and melting temperatures were consistently observed in the pH range between 3.5 and 9.5 (Fig. 1D). These curves revealed the presence of an unfolding process occurring at an average melting temperature of 54.6 ± 0.9 °C. If Jaburetox were totally unfolded at room temperature, no melting transition would be observed [30], thus indicating the presence of some degree of folding that is largely independent of pH. Substitution of the reducing agent β-mercaptoethanol, previously used in the storage buffer [20,23], with tris (2-carboxyethyl)phosphine (TCEP) significantly decreased the tendency of Jaburetox to aggregate both at 25 °C and at low temperatures (−80 °C), and the same oligomeric state was obtained after freezing and thawing. No difference was observed in the melting temperatures in the absence and in the presence of TCEP, suggesting that the improved effect of TCEP versus β-mercaptoethanol is related to the more efficient reduction of disulfide bonds that could form over time by oxidation of cysteine thiol groups, and not to an overall stabilization of the protein fold.

Hydrodynamic properties and aggregation state

Light scattering measurements were carried out in order to study the aggregation state of Jaburetox in solution by combining in-line SEC, static multiple-angle light scattering (MALS) and dynamic quasi-elastic light scattering (QELS) (Fig. 1E). SEC experiments showed a single peak eluting from the column, indicating that, in solution, Jaburetox exists in a single and homogeneous oligomeric form. The latter, according to MALS measurement, presents a molar mass of 11.03 ± 0.01 kDa, fully consistent with the theoretical molar mass of the monomeric protein (10 951 Da) based on its amino acid sequence that includes the six histidines (http://web.expasy.org/protparam/). The hydrodynamic radius measured by QELS is 2.7 ± 0.1 nm. This value is larger than the expected value of 1.74 nm for a well-folded protein of the same molecular mass [31], and is very close to that predicted for intrinsically disordered pre-molten globular proteins (2.46 nm) or random coil proteins (2.77 nm) of the same molecular mass [31]. This observation supports the idea that Jaburetox exists in an extended conformation in solution, probably due to the lack of secondary or tertiary contacts. To experimentally evaluate the amount of secondary structure of the protein in solution, CD spectroscopy was applied.

Secondary structure of Jaburetox by circular dichroism

The CD spectrum of the protein under native conditions (Fig. 2A) presents features typical of a random coil conformation, with a minimum centered at 200 nm. No strong negative signals above 205 nm, characteristic of α-helix or β-sheet structures, were observed. Accordingly, the software CAPITO (http://capito.nmr.fli-leibniz.de/), a tool that distinguishes different folding states of polypeptides on the basis of their far-UV CD spectra [32], indicated that the CD spectrum of Jaburetox is compatible with a native ensemble of disordered conformations in a pre-molten globular state, featuring a small amount of secondary structure (Fig. 2B). The CD spectrum of the protein undergoes minor changes upon temperature increase (Fig. 2A), suggesting a modest influence of temperature on the structural distribution among the ensemble of protein conformers. In particular, at 90 °C the minimum at 200 nm shifts to 202 nm, and the ellipticity at 222 nm, typical of α-helical composition, slightly

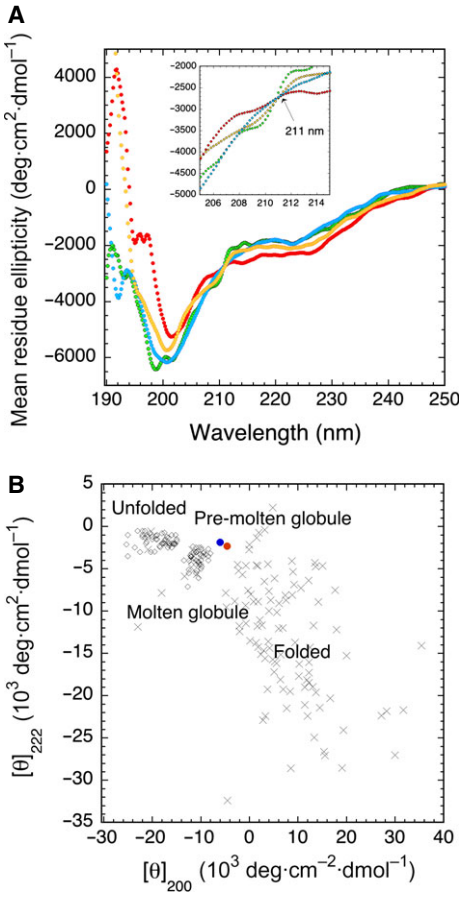


Fig. 2. Circular dichroic properties of Jaburetox. (A) CD spectrum of Jaburetox at 298 K (green), 323 K (yellow), 363 K (red) and 298 K after cooling temperature (cyan); the inset shows the region closer to the dichroic point at 211 nm. (B) Conformational classification of Jaburetox according to its CD spectrum performed using the web server CAPITO (http://capito.nmr.fli-leibniz.de/).

increases. In addition, a positive band appears at 192 nm indicative of the presence of some α-helices or β-strand elements. The spectrum of the protein at 50 °C is intermediate between the spectrum at 25 °C and at 90 °C. Overall, this reveals a small expansion of the secondary structure content at high temperature. This behavior was previously reported for IDPs [33,34] and can be explained by the known increased strength of hydrophobic interactions at higher temperatures. The effect is particularly significant in the case of strong electrostatic repulsions by a large number of electric charges that destabilize the hydrophobic attractions at lower temperatures [33]. The spectrum of the native Jaburetox at 25 °C was fully recovered after cooling the temperature, indicating that the changes induced by temperature onto the secondary structure composition are fully reversible (Fig. 2A). The superimposition of the spectra obtained at different temperatures reveals an isodichroic point at 211 nm, with a mean residue ellipticity of −2700 deg·cm^2·dmol^{-1} (Fig. 2A, inset). The presence of a wavelength in which the molar absorptivity is the same for two (or more) protein spectra is indicative of the presence of two prevalent conformational states in equilibrium. For proteins with prevalence of α-helices, isodichroic points are observed around 203 nm and are related to the helix–coil transition upon protein unfolding [35]. On the other hand, isodichroic points at higher wavelengths have been observed upon conformational transitions of IDPs, such as α-synuclein [36], and have been interpreted as indicative of conformational transitions within the random coil ensemble [37]. The presence of a conformational change between two different population states observed here is consistent with the thermal differential scanning fluorimetry experiments, which showed the presence of a transition induced by temperature.

Computational analysis of disorder propensity

In a previous study [23], the tertiary structure of Jaburetox was calculated by homology modeling, suggesting that Jaburetox contains very few secondary structure elements, with ~ 70% of the protein in the random coil conformation. A 13-residue α-helix was modeled in the N-terminal portion of the protein, while the C-terminal fragment was proposed to contain a β-hairpin. Molecular dynamics simulations suggested that the protein undergoes a further increase of the random coil conformation that, at the end of the simulation, comprised 84% of the protein structure, with the N-terminal portion losing the α-helical content and a short β-sheet appearing at the C terminus [23]. Previous molecular dynamics studies [17,19,22] also suggested that Jaburetox adopts a β-hairpin in its C-terminal region, as observed in the X-ray crystallographic structure of JBU [21].

Consistently with the intrinsically disordered nature of Jaburetox revealed by CD and hydrodynamic analyses, various computational tools indicated that this polypeptide is characterized by high intrinsic disorder propensity. First, we used the known fact that the amino acid compositions of ordered proteins and IDPs are very different, with disordered proteins being systematically enriched in disorder-promoting residues (A, R, G, Q, S, E, K and P) and depleted in order-promoting residues (W, Y, F, I, L, V, C and N) [38,39]. The results of this analysis are shown in Fig. 3A, which illustrates that Jaburetox is depleted in major order-promoting amino acids and enriched in

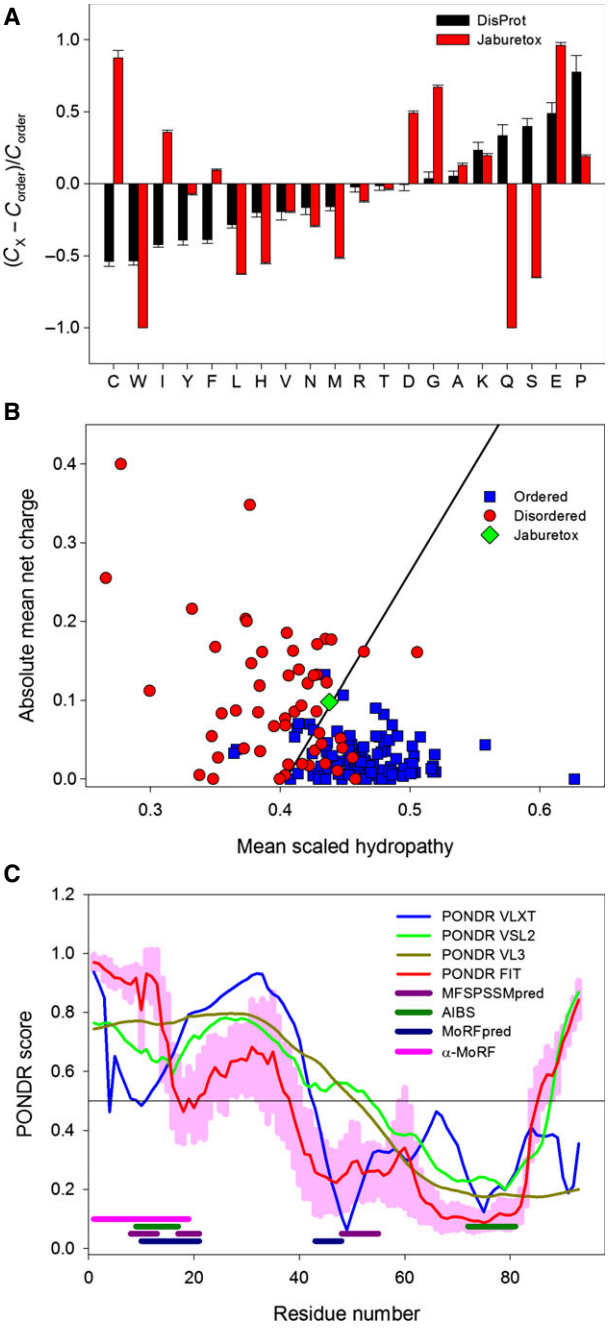


Fig. 3. Evaluation disorder propensity of Jaburetox. (A) Compositional analysis of Jaburetox in comparison with the composition of typical ordered proteins. The compositional profile of typical IDPs from the DisProt database is shown for comparison (black bars). Positive bars correspond to residues found more abundantly in Jaburetox than in ordered proteins, whereas negative bars show residues in which Jaburetox is depleted. (B) Charge–hydropathy plot for Jaburetox; data for ordered and disordered proteins are shown as blue squares and red circles, respectively, whereas the position of Jaburetox is shown as a green diamond. The black line represents a boundary separating compact and extended proteins. (C) Evaluating the per-residue intrinsic disorder propensity of Jaburetox using four members of the PONDR family: PONDR VL-XT (blue line), PONDR VSL2 (green line), PONDR VL3 (dark yellow line), PONDR-FIT (red line); sections with scores higher than 0.5 correspond to disordered regions. Light pink shadow around the PONDR-FIT line corresponds to standard errors of disorder prediction by PONDR-FIT. Location of potential binding sites predicted by MFSPSSMPRED, ANCHOR, general MORFPRED and α-MORF-PRED are shown as bold dark pink, dark green, dark blue and pink bars respectively.

some disorder-promoting residues, particularly D, G, A, K, E and P, relatively to typical ordered proteins. Also, the Uversky plot of mean net absolute charge versus mean hydrophobicity [40] (access to which is available at the PONDR® website http://www.pondr.com [41]) indicated that this protein is predicted as disordered even though it lies close to the boundary between proteins predicted to be extended and compact (Fig. 3B). Analysis of the per-residue disorder propensity of Jaburetox by a family of PONDR predictors revealed the presence of long disordered regions in this protein (PONDR scores above 0.5), especially in its N-terminal part (Fig. 3C).

It is known that protein–protein interactions are commonly mediated by disordered regions that undergo disorder to order transitions as a result of binding. Amino acid sequences of such disordered binding regions contain specific signals that can be identified by several specialized computational tools, such as A-MORF-PRED [42,43], general MORFPRED [44], MFSPSSMPRED [45] and ANCHOR [46]. Figure 3C shows that Jaburetox contains several potential disorder-based binding sites, such as α-MoRFs, general MoRFs, sites identified by MFSPSSMPRED and ANCHOR-indicated binding sites (AiBSs), which completely or partially overlap with each other. A larger agreement is observed among the various tools in the case of a potential binding site located in the N-terminal region (residues 1–21) compared to that found in the central portion (residues 43–55) or in the C-terminal region (residues 72–81). This is probably due to the fact that the potential binding site at the N terminus has a stronger α-helical signal, which is recognized by the computational tools used in this study.

The presence of these potential disorder-based binding sites is an indication that one of the functions of the disordered Jaburetox might be related to providing an interaction platform with various binding partners. Notably, earlier studies revealed that the majority of the biological activity of this protein resides preva-

lently in the N-terminal portion of Jaburetox [23], as opposed to the β-hairpin presumably located within the C-terminal portion [17,19,22]. In order to derive experimental structural information for Jaburetox, NMR spectroscopy was applied.

Assignment of the NMR spectra of Jaburetox

Figure 4 shows the ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectrum of Jaburetox. The spectrum is characterized by low signal dispersion in the proton dimension, indicative of a disordered state of the protein. Heteronuclear 2D and 3D triple resonance NMR spectra of Jaburetox were recorded and analyzed, and the assignment of the chemical shifts of backbone ^{1}H, ^{15}N and ^{13}C nuclei was obtained using the scalar connectivities derived by a computer-assisted resonance assignment software (CARA) [47], following a standard sequential assignment procedure. The identification of backbone ^{1}HN and ^{15}N amide peaks was obtained for 84 of the expected 86 residues of the protein, corresponding to 97.7% (not counting P3, P38, P41, P47 and P57, in addition to M1, usually featuring an $-NH_3^+$ group and not readily observable because of exchange with solvent, and G2, which is placed between the two unassigned residues M1 and P3). Residues E14, D61 and D77 were tentatively assigned based on their chemical shifts. The only unassigned residues were E49 and K62. Nearly complete assignments were achieved for the other backbone nuclei (100% for $^{13}C\alpha$, 100% for $^{13}C\beta$ and 85.2% for ^{13}CO) while side chain assignments are complete to 80.2% considering only aliphatic ^{1}H and ^{13}C nuclei. The resonance assignments were deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under accession number 19830.

Secondary structure propensity of Jaburetox from NMR chemical shifts

Disordered proteins tend to feature only few long-range tertiary contacts, and therefore local structural constraints on the backbone are important in order to describe the ensemble of possible conformers of the protein. In this perspective, the structural information encoded in the chemical shifts of backbone nuclei has emerged as a key indicator of IDP ensemble properties [48]. Therefore, the experimental chemical shifts of backbone Cα, Cβ and Hα nuclei of Jaburetox were used to predict the residue-specific secondary structure propensity (SSP) for this polypeptide in solution using the program SSP [49]. A positive SSP score indicates a propensity for α-structure, while a negative score indicates a propensity for β-structure or extended loops. Residues in fully formed α-helices and β-strands are given scores of +1 and −1, respectively. Figure 5A

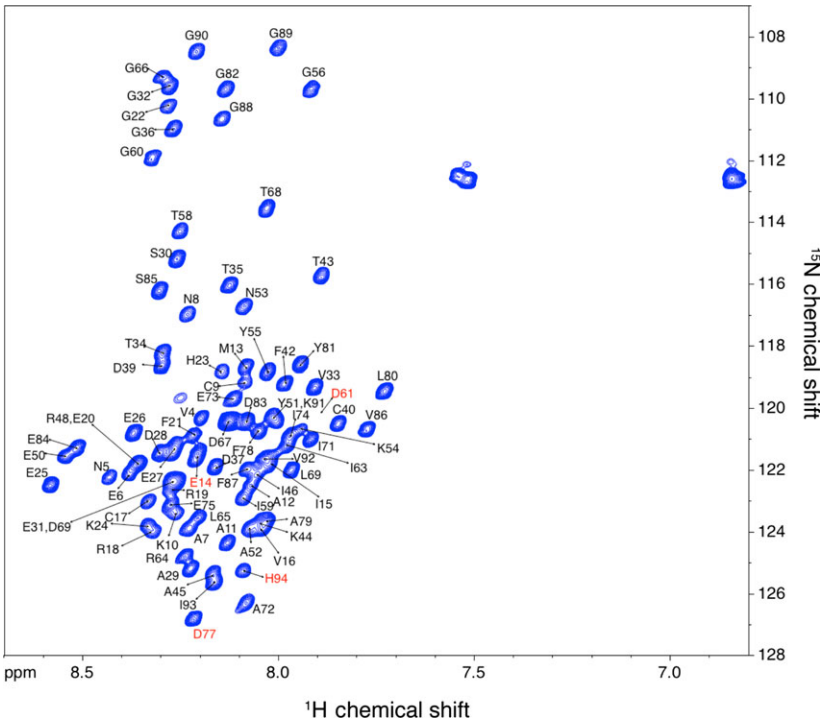


Fig. 4. *In vitro* NMR spectrum of Jaburetox. 800 MHz ^{1}H-^{15}N HSQC spectrum of Jaburetox in 90% H_2O, 10% D_2O, pH 6.5, $T = 298$ K. Assigned cross-peaks are labeled with one-letter amino acid type and sequence number. Tentatively assigned residues, as well as the initial histidine of the His tag, are indicated by red labels.

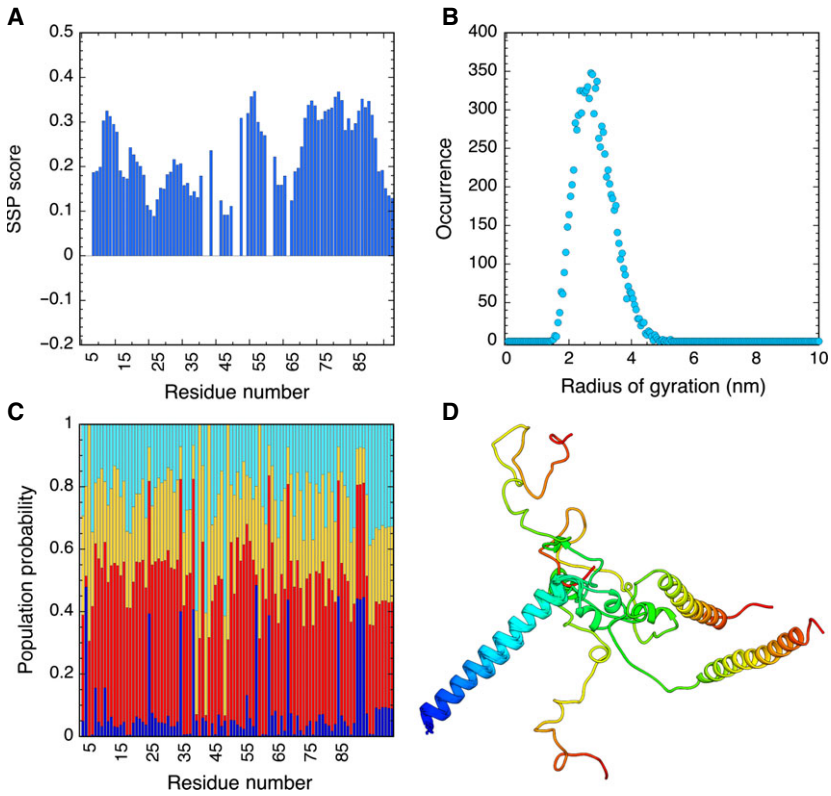


Fig. 5. Secondary structure of Jaburetox. (A) SSP scores calculated using Cα and Cβ chemical shifts for Jaburetox. Residues in fully formed α-helices and β-strands are expected to give scores of +1 and −1, respectively. (B) Radius of gyration distribution profile calculated for 10 000 conformers of Jaburetox using FLEXIBLE-MECCANO. (C) Population probability of secondary structure (cyan, β-sheet; yellow, polyproline II; red, α-helix; blue, random coil) calculated for 10 000 conformers of Jaburetox using FLEXIBLE-MECCANO. (D) Representative conformers of the structural clusters that are most populated by Jaburetox according to the FLEXIBLE-MECCANO/SSP analysis. Ribbons are colored from deep blue in the proximity of the N terminus to red at the C terminus. The structures are superimposed on residues 1–36 for clarity.

shows that Jaburetox is largely disordered, featuring only a small helical propensity, with slightly smaller SSP values in the N-terminal (larger predicted disorder) compared to the C-terminal portion (smaller predicted disorder). The average SSP score is 0.23 ± 0.08. Some correlation is observed between the SSP score based on chemical shifts and the score calculated using PONDR (Fig. 3). An analogous approach carried out using TALOS+ [50] did not yield any secondary structure with acceptable level of confidence, consistent with the picture obtained using SSP, suggesting a very small propensity to assume any organized secondary structure.

The results of the SSP analysis were used as input to calculate a structural ensemble representing Jaburetox in solution using FLEXIBLE-MECCANO [51]. This approach entails the use of amino acid specific statistical coil sampling to describe the unfolded state of the protein on the basis of the primary sequence; we further implemented the residue-by-residue conformational propensities, calculated by SSP as described above, into the calculation. The large range of gyration radii calculated using FLEXIBLE-MECCANO (Fig. 5B) suggests a largely diversified ensemble of conformers. The resulting average value of R_g (2.7 ± 0.7 nm), together with the hydrodynamic (Stokes, R_S) radius of 2.7 ± 0.1 nm determined experimentally using QELS (Fig. 1E), provides support for the nature of Jaburetox as an intrinsically disordered polypeptide that retains considerable residual structure in solution: indeed the R_g/R_S ratio is predicted to be ~ 0.775, ~ 0.9 or ~ 1.5 for proteins in a globular, pre-molten globular or fully unfolded state [52,53]. The probability of each residue existing in α-helix, β-sheet, polyproline II or random coil conformations was also calculated using FLEXIBLE-MECCANO (Fig. 5C), further supporting the preponderant presence of helices along the sequence, with slightly larger probability for helical propensity in the N-terminal portion of the polypeptide compared to the C-terminal region. Clustering analysis of the 10 000 structures of Jaburetox explicitly calculated using FLEXIBLE-MECCANO was performed using the g_cluster module of GROMACS 4.6 [54–57] and the GROMOS algorithm [58]. A 1.5 nm cutoff for the rmsd was used to include structures in the same cluster. A total of 71 clusters were identified, with the first five clusters accounting for 64% of the overall conformational ensemble. Figure 5D shows the superimposition of the representative structures of these five clusters selected by g_cluster. These show the consistent presence of an α-helix at the N terminus, of ~ 40 residues, followed by a coiled region of ~ 20 residues and by another ~ 40-residue long α-helix in the C-terminal region, observed in two

of the five most populated clusters covering ~ 42% of the overall ensemble. A parallel analysis performed with FLEXIBLE-MECCANO without the use of SSP scores yielded a structural ensemble characterized by the absence of conserved extended secondary structure elements. The consistency of the calculated structural ensemble, obtained using FLEXIBLE-MECCANO and SSP scores, with experimental data obtained using NMR-based distance constraints was tested by determining the structure of Jaburetox in solution.

Solution structure of Jaburetox by NMR

The 3D structure elucidation of folded proteins in solution relies on the availability of long-range distance information obtained from the nuclear Overhauser enhancement (NOE). Regions of the protein distant in the primary sequence but close in space in the folded structure give rise to NOEs that are utilized to determine the global fold. As expected, in the case of Jaburetox, no long-range interactions indicative of the presence of stable tertiary structure were observed. However, a few medium-range NOE constraints indicative of helical or turn-like structures were observed, corresponding to the regions V4–E20, R48–G56 and I63–I74. The solution structure of Jaburetox was calculated using CYANA based on geometrical constraints derived from these NOEs and dihedral angle constraints obtained from TALOS+ (Table 1). In total, 844 NOE-based upper distance limits and 12 φ and ψ torsion angles were used to derive the Jaburetox structure. The geometrical constraints and coordinate files of the Jaburetox structural ensemble were deposited in the Protein Data Bank under the accession code 2MM8. The NMR-derived solution structural ensemble of Jaburetox, featuring large values of rmsd, confirms its disordered fold, as predicted by earlier [23] and the present bioinformatics studies (Fig. 3). However, some elements of secondary structure are found in three different parts of this protein consistently for all structure conformers of the selected ensemble (Fig. 6, Table 1): a small α-helical motif at the N terminus (A12–V16), and two turn-like structures located in the middle of the protein (R48–G56) and at its C terminus (I63–E74). These regions with transient secondary structure coincide with the potential disorder-based binding sites identified by the bioinformatics analysis described above (Fig. 3C). The presence of these regions with more structured fold is consistent with the presence of a peak in the thermal shift assays that implies an unfolding process, highlighting the power of this methodology to predict the folding state of proteins. Figure 6 illustrates the Jaburetox structural ensemble with the lowest target function superimposed using the three different fragments with residual secondary structure. To allow a more detailed investigation of the structural organization of these regions, the structure of each fragment was calculated separately from the rest of the protein and further refined following the same protocols as for the full protein (Fig. 7, Table 1). Although bioinformatics analysis predicted a slightly longer N-terminal α-helix composed of 13 amino acids (V4–E20), in fact only five of them (A12–V16) were experimentally confirmed to be involved in the α-helix formation as indicated by the presence of the characteristic NOE distance constraints. The rest of the residues in this region (V4–A11) exhibited some α-helical propensity which, together with the backbone relaxation data, may suggest that the formation of a more extended α-helix in principle could take place in slightly different sample conditions.

Overall, the NMR-based structural ensemble of Jaburetox shows some features that are similar to the ensemble determined by FLEXIBLE-MECCANO using only sequence and secondary chemical shift information: the helical fragment found by NMR close to the N terminus is composed within the much longer helix predicted to exist using the FLEXIBLE-MECCANO/SSP approach but shows a reduced extension; concomitantly, the helical region suggested to appear in the C-terminal portion of Jaburetox is in fact absent in the structural family determined by distance and dihedral constraints by NMR, and only two short sequences with some consistent prevalence of a turn motif are found in the NMR structure. Overall, these observations can be interpreted as indicating an overestimation of the helical propensity in Jaburetox as derived from the combination of intrinsic sequence properties and chemical shift index. We speculate, however, that helical regions could indeed be formed under solution conditions so far unattained.

In order to explore the protein folding under more physiological conditions, the disordered nature of Jaburetox was investigated by in-cell NMR spectroscopy, a technique that allows the acquisition of structural data on biomolecules in the cytoplasm [59–62]. ^{1}H-^{15}N HSQC spectra were recorded on *E. coli* cells overexpressing Jaburetox upon induction with IPTG and growth in ^{15}N-labeled media. A typical spectrum, shown in Fig. 8, demonstrates that, although some modifications of the backbone are evident by chemical shift perturbations compared to the spectrum in solution (possibly derived from differences in the physical parameters of the milieu), Jaburetox disorder is preserved within the cell, and it is not an *in vitro* artifact

Table 1. Structural statistics and geometrical constraints derived from NMR for Jaburetox.

Restraints used in structure calculation	
NOE constraints for the full protein	844
NOE constraints for the V4–E20 fragment	190
NOE constraints for the R48–G56 fragment	63
NOE constraints for the I63–I74 fragment	168
Torsion angle constraints for the full protein	12
Structure statistics	
CYANA target function value for the full protein ($Å^2$)	281.52 ± 0.73
CYANA target function value for the V4–E20 fragment	44.7 ± 0.09
CYANA target function value for the R48–G56 fragment	11.8 ± 0.05
CYANA target function value for the I63–I74 fragment	65.8 ± 0.04
Mean global backbone rmsd for the full protein (Å)	9.36 ± 2.45
Mean global backbone rmsd for the V4–E20 fragment (Å)	0.31 ± 0.18
Mean global backbone rmsd for the R48–G56 fragment (Å)	1.07 ± 0.86
Mean global backbone rmsd for the I63–I74 fragment (Å)	0.70 ± 0.41
Mean global heavy atom rmsd for the full protein (Å)	10.43 ± 2.55
Mean global heavy atom rmsd for the V4–E20 fragment (Å)	1.11 ± 0.25
Mean global heavy atom rmsd for the R48–G56 fragment (Å)	2.41 ± 1.16
Mean global heavy atom rmsd for the I63–I74 fragment (Å)	1.35 ± 0.58
Amber energies (kcal·mol^{-1})	
Amber energy for the full-length protein	−1.98E+05
Amber energy for the V4–E20 fragment	−2.91E+04
Amber energy for the R48–G56 fragment	−2.37E+05
Amber energy for the I63–I74 fragment	−2.14E+04
PROCHECK *NMR Ramachandran statistics*	
Residues in favorable regions for the full-length protein (%)	32.9
Residues in favorable regions for the V4–E20 fragment (%)	22.9
Residues in favorable regions for the R48–G56 fragment (%)	66.7
Residues in favorable regions for the I63–E74 fragment (%)	12.5
Residues in allowed regions for the full-length protein (%)	55.3
Residues in allowed regions for the V4–E20 fragment (%)	43.6
Residues in allowed regions for the R48–G56 fragment (%)	16.7
Residues in allowed regions for the I63–E74 fragment (%)	31.2
Residues in generously allowed regions for the full-length protein (%)	9.4
Residues in generously allowed regions for the V4–E20 fragment (%)	19.3
Residues in generously allowed regions for the R48–G56 fragment (%)	0
Residues in generously allowed regions for the I63–E74 fragment (%)	42.5
Residues in disallowed regions for the full-length protein (%)	2.4
Residues in disallowed regions for the V4–E20 fragment (%)	14.3
Residues in disallowed regions for the R48–G56 fragment (%)	16.7
Residues in disallowed regions for the I63–E74 fragment (%)	13.8

caused by the purification process. Similarly, in-cell NMR experiments have been used to demonstrate that disorder is maintained in the case of α-synuclein overexpressed in *E. coli* cells [63] or for the tau-protein microinjected into *Xenopus* oocytes [64].

NMR studies of Jaburetox dynamics

The results of the computational and structural analyses of chemical shifts described above prompted us to investigate more directly the protein dynamics of Jaburetox using ^{15}N relaxation measurements. NMR relaxation parameters provide valuable insights into the internal molecular motions of unfolded or partially folded proteins. The ^{15}N relaxation rates R_1 and R_2 and the ^{1}H-^{15}N NOE values of all assigned backbone amide groups of Jaburetox were determined and are shown in Fig. 9A,B,C, respectively. The presence of local internal motions in the picosecond–nanosecond time scales contribute to the R_1, R_2 and NOE values, while conformational exchange processes occurring on the microsecond–millisecond time scale additionally contribute to increase the R_2 rates [65]. Therefore, the analysis of these parameters can provide information

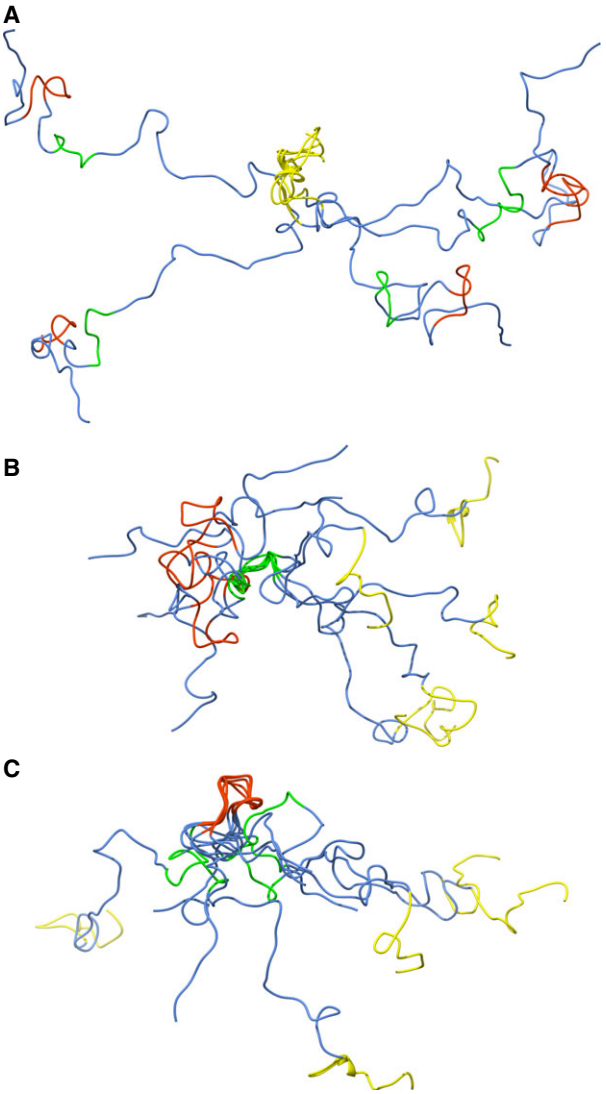


Fig. 6. Structure of Jaburetox by NMR. Ribbon scheme of the solution structure of Jaburetox calculated using CYANA, represented by the ensemble of five conformers with the lowest target function superimposed using three different and self-consistent protein fragments: (A) V4–E20, yellow; (B) R48–G56, green; (C) I63–I74, red. Figure made using CHIMERA [103].

on local backbone mobility of Jaburetox at different time scales. NOEs in particular are more sensitive to fast internal dynamics than R_1 and R_2 [66].

A qualitative analysis of the relaxation data for Jaburetox indicates that small NOE values are generally observed, indicative of a protein characterized by fast motion and high flexibility throughout the protein chain, consistently with the disordered nature of the protein revealed by the small chemical shift range of amide protons. The C-terminal region features larger NOE values compared to the N terminus, indicative of a relatively reduced mobility, in agreement with what is suggested by disorder predictions (see Fig. 3B). This difference somehow parallels the distinct biological activity of the two portions of Jaburetox, with the N terminus being involved in the toxicity of the protein against insects while the C terminus is not or less active [23]. Small or negative NOE values are found in the initial portion of the N terminus (4–10) and in the 30–40 region, indicating larger mobility in the nanosecond–picosecond time range. In the C-terminal region, several residues exhibit large values of R_2, which can be due either to a more structured conformation or to exchange processes occurring in the millisecond–microsecond time scale, or to both effects simultaneously.

A preliminary estimate of the rotational correlation time $\tau_m = 4.13 \pm 0.50$ ns was obtained according to $\tau_m = (1/2\omega_N)(6R_2/R_1 - 7)^{1/2}$ [67] using 14 experimental R_1 and R_2 relaxation rates selected by excluding residues characterized by significant internal mobility as shown by their small R_2 ($R_2 < (\bar{R}_2 - \sigma)$ and $(\bar{R}_2 - R_2)/R_2 > (\bar{R}_1 - R_1)/R_1$) and small NOE (NOE < 0.35). This value of τ_m was used to obtain a more accurate value by fitting the same 14 R_2/R_1 ratios to the general equation that correlates this ratio with the spectral densities assuming isotropic tumbling [66], an approach that yielded $\tau_m = 4.17 \pm 0.50$ ns. This value corresponds to a molecular mass of 7.0 ± 1.0 kDa estimated using the empirical relationship τ_m (ns) ~ 0.6 kDa determined for folded proteins [68]. This value supports the presence of the Jaburetox monomer in solution under the experimental conditions used, in agreement with the light scattering data reported above. The value estimated by NMR is slightly lower than expected, a difference possibly due to the large mobility of the protein that somehow simulates a smaller molecular mass.

Relaxation data were analyzed using the reduced spectral density mapping approach [69–72]. This represents a more suitable method than the model-free approach [73–76] because it does not make assumptions about the nature of the correlation function that describes the overall rotational diffusion, nor about the value of ^{15}N chemical shift anisotropy. In addition, it does not require knowledge of the protein structure, making it a more reliable approach in the case of IDPs. The spectral density function $J(\omega)$ describes the relative populations of HN bond vector fluctuations at different frequencies, namely at the frequency corresponding to the overall protein tumbling, as well as at ω_N and $0.87\omega_H$. $J(0)$ is influenced by low-frequency motions (nanoseconds) along with some fluctuations occurring on the millisecond to microsecond time scale arising from chemical exchange, $J(0.87\omega_H)$ is influ-

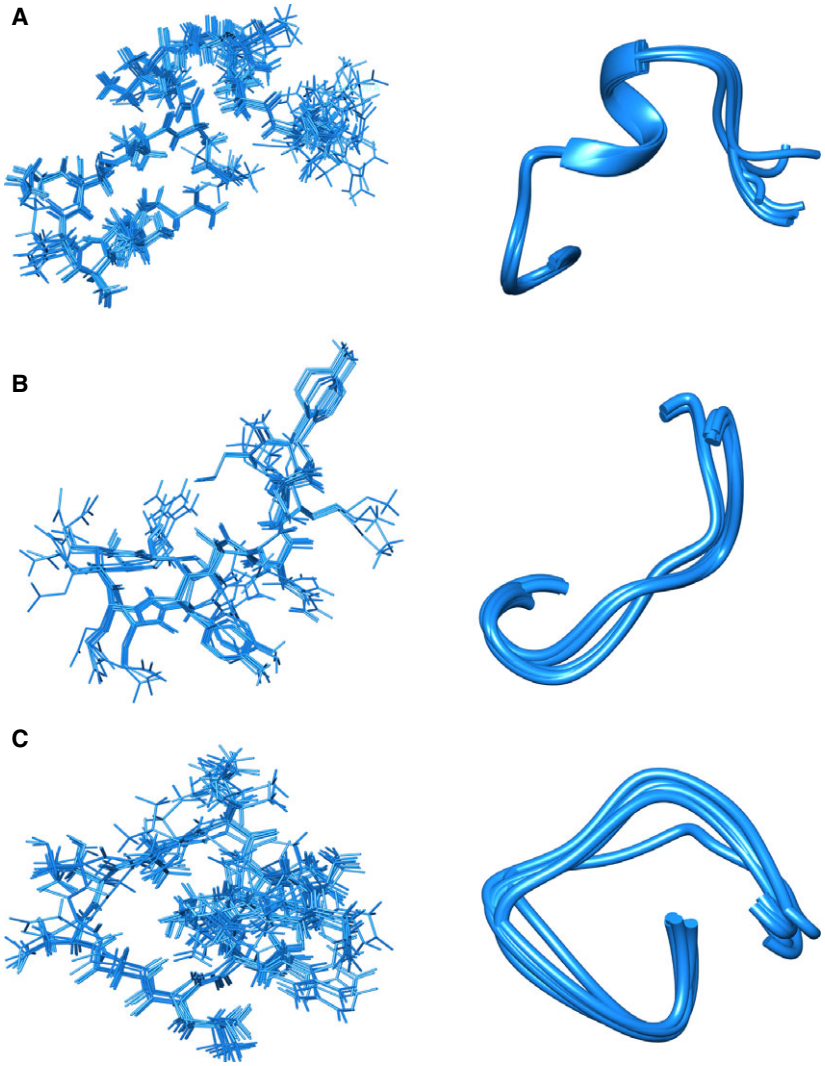


Fig. 7. Structure of Jaburetox ordered portions by NMR. Superimposition of the 10 conformers with the lowest target function represented as sticks (left panel) and ribbon (right panel) for the Jaburetox fragments V4–E20 (A), R48–G56 (B), I63–I74 (C). Figure made using CHIMERA [103].

enced by high-frequency motions (picoseconds) and $J(\omega_N)$ is the most sensitive to motions in the intermediate time scale. The results of this analysis are shown in Fig. 9D,E,F. Smaller than average values of $J(0)$ are observed in the N-terminal region of Jaburetox compared to the C-terminal region, indicating faster motions. $J(\omega_N)$ is particularly small at the beginning of the protein sequence, highlighting even faster dynamics. In the C-terminal region, a few residues display much higher values of $J(0)$, suggesting a contribution from conformational exchange phenomena. Consistently, $J(0.87\omega_H)$ decreases progressively from the N to the C terminus, paralleling a decrease of internal mobility. The possibility exists that exchange with the solvent, in addition to conformational processes, could influence the accurate determination of the relaxation times. Therefore, this treatment only provides a qualitative picture of the protein dynamics of Jaburetox and a comparative view of mobility along the protein sequence.

Discussion

The determination of the solution structure of plant proteins and peptides with insecticidal and/or antifungal activities is poorly explored in the literature. Some examples of antifungal polypeptides studied by NMR are Psd1, a 46-residue recombinant defensin from *Pisum sativum* [77,78], Rs-AFP1, a 51-residue defensin isolated from seeds of *Raphanus sativus* L. [79], Ib-AMP1, a 20-residue peptide derived from seeds of *Impatiens balsamina* [80], and EAFP2, a 41-residue polypeptide from *Eucommia ulmoides* [81]. Other plant-derived insecticides structurally characterized by

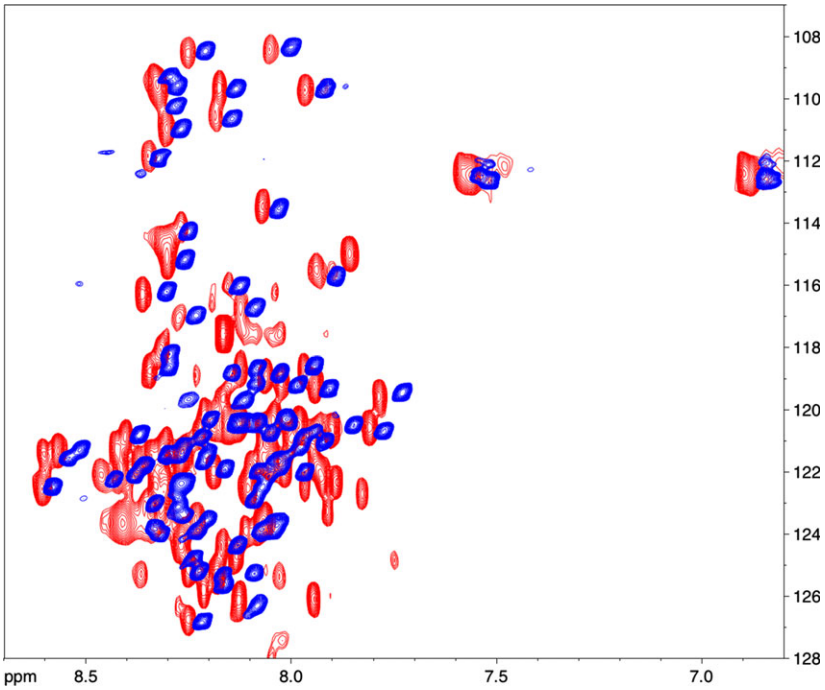


Fig. 8. In-cell NMR spectrum of Jaburetox. 950 MHz ^{1}H-^{15}N HSQC spectrum of Jaburetox-expressing *E. coli* cells induced with IPTG (shown in red) overlapped with the spectrum of isolated Jaburetox in solution (in blue). The experimental conditions are described in the text.

NMR are *Vigna radiata* VrD1, a plant defensin [82], Kalata B2, a cyclic peptide extracted from *Oldenlandia affinis* [83], PA1b, a 37-residue cysteine-rich plant defense peptide isolated from *P. sativum*, and NaD1, a 47-residue polypeptide isolated from flowers of *Nicotiana alata* [84]. In all these cases, a stable secondary and tertiary structure was determined.

In this work, we determined the solution properties of Jaburetox, characterizing this polypeptide as an IDP. We further used solution NMR spectroscopy, a powerful tool for studying IDPs [85–88], to establish the properties of the structural ensemble of Jaburetox conformations.

Considering the membrane-disturbing properties of Jaburetox [19,23,24], we would expect that the polypeptide would change its conformation to a folded state when in contact with membranes. The full assignment of Jaburetox reported here will allow us to perform NMR studies of this toxic polypeptide in contact with lipid micelles mimicking insect and fungus membranes [78], determining changes in its conformation properties based on modifications of the ^{1}H-^{15}N HSQC spectrum. Many IDPs are known to bind efficiently to artificial and natural membranes and this interaction is accompanied by a dramatic increase in their α-helical content [89]. It may be also that these conformational changes will not happen, since there are some IDPs that do not require protein folding to be active. Indeed, function can arise from any of these conformations and transitions between them [33,90]. With the currently established NMR assignment of Jaburetox, the study of its interaction with different targets is now possible, aiming to find a receptor to Jaburetox in insects and fungi.

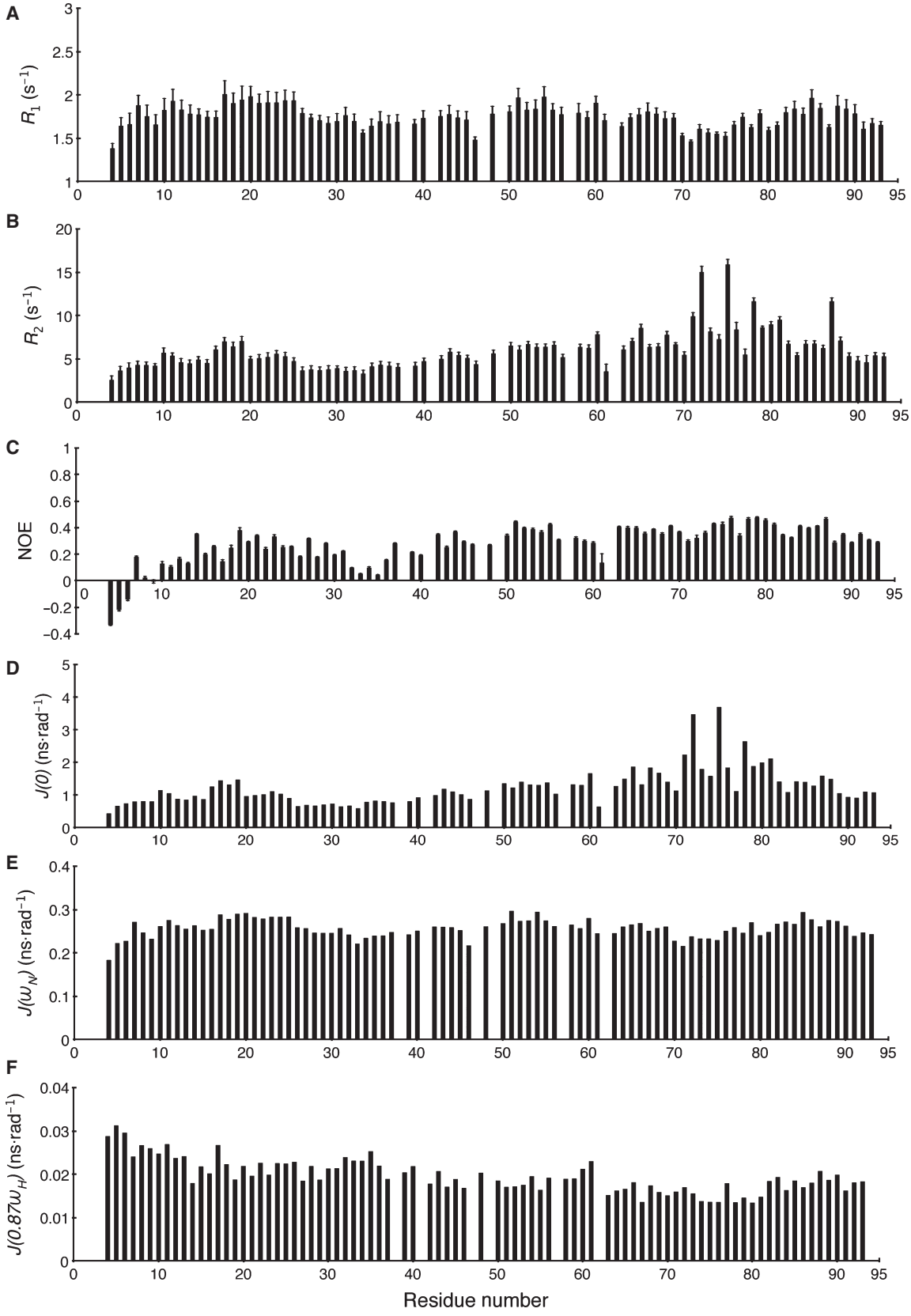
Current efforts in our laboratories are aimed at further exploring the folding state of Jaburetox in *E. coli* cells, as well as *in vitro* in the presence of different crowding agents, membranes and non-interacting proteins, in order to understand whether some part of the protein undergoes significant conformational change in the cellular environment. In addition, we are setting up a method to determine the protein structure in cells of transgenic soybean plants that express the peptide using NMR spectroscopy.

The structural characterization of Jaburetox as an IDP is an important step towards understanding its toxic properties. Besides the structural peculiarities, the use of NMR, a powerful tool, could help us go further to unravel the pathways involved in the toxicity of Jaburetox, paving the way to exploring the potential applications for this naturally occurring polypeptide.

Materials and methods

Jaburetox production and purification

E. coli BL21(DE3) (Novagen, Madison, WI, USA) cells were transformed by heat shock with the construct pET23a-*Jaburetox* that expressed a protein containing a six-histidine tag at the C terminus [20]. Jaburetox analyzed



in this study is the polypeptide corresponding to the residues 169–260 of the urease JBURE-II from *C. ensiformis* (UniProt ID Q8H6V8) containing the additional N-terminal methionine residue and the E195D substitution. Jaburetox expression was optimized in 25 mL cultures grown for 16 h, testing different values of temperature (16, 19, 26, 33 and 37 °C) and IPTG concentration (0.1, 0.23, 0.55, 0.87 and 1.0 mM) using a methodology that involved a total of 12 experiments to map a calculated response surface [91]. In all cases, the cells obtained after centrifugation (4500 ***g***) were incubated for 30 min at room temperature in 300 μL of 50 mM Tris/HCl buffer at pH 7.8, containing 60 μg of lysozyme, 6 μg of DNase and 0.3 mM $MgCl_2$. The resulting suspension of cell lysate and membranes was centrifuged to remove cell debris, and the supernatant was added to 50 μL of a slurry of Ni^{2+}-loaded Sepharose FF (GE Healthcare) resin. The yield of soluble Jaburetox produced under the different growth conditions was then quantified by evaluating the amount of protein bound to the resin using SDS/PAGE in 12% polyacrylamide gels and the software IMAGEJ 1.47 (National Institutes of Health, Bethesda, USA). The regression and graphical analysis of the data were performed using the STATISTICA 6.0 software (Statsoft, USA). After the best conditions for protein expression were selected, a recombinant cell pre-inoculum of 50 mL was cultured overnight, at 37 °C and 150 r.p.m., in 20 mL LB broth, supplemented with 1% glucose and 50 $\mu g \cdot mL^{-1}$ of carbenicillin. Subsequently, cells were inoculated in 1 L of LB broth containing 50 $\mu g \cdot mL^{-1}$ of carbenicillin and grown at 37 °C, 150 r.p.m., until D_{600} was 0.6–0.7. Cells were collected by centrifuging the culture at 4500 ***g*** for 30 min, and then the cells were resuspended in 250 mL of an M9 salt solution (6 $g \cdot L^{-1}$ Na_2HPO_4, 3 $g \cdot L^{-1}$ KH_2PO_4, 0.5 $g \cdot L^{-1}$ NaCl, 0.246 $g \cdot L^{-1}$ $MgSO_4$) and centrifuged again. The cell pellet was finally resuspended in 250 mL of M9 medium supplemented with 1.25 $g \cdot L^{-1}$ of $(NH_4)_2SO_4$ and 4 $g \cdot L^{-1}$ of glucose. Protein expression was induced after 1 h by addition of IPTG to a final concentration of 0.87 mM in the presence of 50 $\mu g \cdot mL^{-1}$ carbenicillin [25]. Cells expressing the protein were cultured for an additional 16 h at 20 °C. For isotope-labeled samples used in NMR experiments, 99% ^{13}C-labeled glucose and 98% ^{15}N-labeled ammonium sulfate (Sigma Aldrich) were used.

Cells were harvested by centrifugation at 11 300 ***g*** for 10 min at 4 °C and resuspended in 30 mL of buffer A (50 mM Tris/HCl buffer, pH 7.5, 500 mM NaCl) containing 5 mM imidazole and 200 $\mu g \cdot mL^{-1}$ lysozyme. The cell suspension was incubated at room temperature for 20 min, and then DNase I (20 $\mu g \cdot mL^{-1}$) was added followed by an additional incubation at 37 °C for 20 min. Subsequently, the cells were passed three times through a French pressure cell (SLM, Aminco) at 20 000 $lb \cdot in^{-2}$. The supernatant was separated by centrifugation at 27 200 ***g*** for 40 min at 4 °C and loaded onto a Ni^{2+}-loaded 5 mL His-Trap HP column (GE Healthcare), previously equilibrated with buffer A. Unbound proteins were washed from the column using 10 volumes of buffer A at a flow rate of 1 $mL \cdot min^{-1}$. Elution of Jaburetox was performed using 100 mL of a gradient of buffer A containing 5–500 mM imidazole. The fractions containing Jaburetox were collected, concentrated and loaded onto a Superdex 75 16/60 column (GE Healthcare) pre-equilibrated with buffer B (50 mM phosphate pH 7.5, 1 mM EDTA and 1 mM TCEP) and eluted at a flow rate of 1 $mL \cdot min^{-1}$. Sample purity was assessed by SDS/PAGE using NuPAGE Novex 12% Bis-Tris gels (Life Technologies) stained with ProBlue Safe Stain (Giotto Biotech). Fractions containing the protein were pooled and protein concentration was measured by Bradford assay [92].

Computational analysis of the intrinsic disorder propensity of Jaburetox

To better understand and visualize the relationships between sequence peculiarities and propensity for intrinsic disorder, the amino acid composition of the Jaburetox was analyzed using a tool developed for visualizing amino acid composition biases in proteins [39]. To this end, the fractional difference in amino acid compositions between Jaburetox and a set of representative ordered proteins with low mutual sequence identity [93] selected from the PDB [94], as well as the fractional difference in amino acid compositions between the latter group and a set of typical experimentally validated IDPs from the DisProt database [95], were calculated for each amino acid residue. The fractional difference is calculated as $(C_X - C_{order})/C_{order}$ values, where C_X is the content of a given residue in a protein/protein set of interest (Jaburetox or typical disordered proteins) and C_{order} is the corresponding value for the sample set of ordered proteins from the PDB [39]. Positive and negative values indicate residues in a given set that have more and less order, respectively. Confidence intervals were estimated using per-protein bootstrapping with 10 000 iterations.

Four different disorder predictors of the PONDR family evaluated the intrinsic disorder propensity of Jaburetox: (a) PONDR VL-XT [41], which applies various compositional probabilities and hydrophobic measures of amino acid as the input features of artificial neural networks for the prediction; although it is no longer the most accurate predictor, it is very sensitive to the local compositional biases and is thus capable of identifying potential molecular interaction motifs; (b) PONDR VSL2B [96], which is suitable for accurate evaluation of short and long disordered regions;

Fig. 9. Dynamics of Jaburetox by NMR. Backbone amide ^{15}N relaxation properties recorded at 298 K and 18.8 T on ^{15}N-labeled Jaburetox. (A) Longitudinal R_1 relaxation rate; (B) transverse R_2 relaxation rates; (C) steady-state heteronuclear ^1H-^{15}N NOE; values of $J(0)$ (D), $J(\omega_N)$ (E) and $J(0.87\omega_H)$ (F) resulting from a reduced spectral density analysis of the relaxation data for Jaburetox.

(c) PONDR VL3 [97], which is suitable for finding long disordered regions; (d) PONDR-FIT [98], a meta-predictor that combines the six individual PONDR predictors VL-XT, VSL2, VL3, FoldIndex, IUPred and TopIDP; PONDR-FIT is moderately more accurate than each of the component predictors. Access to these predictors is provided by the DisProt database (http://www.disprot.org/metapredictor.php).

Molecular recognition features (MoRFs), defined as short order-prone motifs within a long disordered region and able to undergo disorder to order transitions during the binding to a specific partner, usually have much higher content of aliphatic and aromatic amino acids than disordered regions in general. Due to these peculiarities, MoRF regions are frequently observed as sharp dips in the corresponding plots representing per-residue distribution of PONDR VL-XT disorder scores. Hence, based on the PONDR VL-XT prediction and a number of other attributes (such as helical propensity), the α-MoRF regions in Jaburetox could be identified [42,43]. Intrinsically disordered regions can fold, upon interaction with specific binding partners, not only to α-helical but also to β-strand, irregular or complex structures; therefore another computational tool, MORFPRED, was used to find all MoRF types in Jaburetox (α, β, coil and complex) [44]. The MFSPSSMPRED tool was additionally used, which is a masked, filtered and smoothed position-specific scoring matrix-based predictor for finding potential MoRF regions in disordered proteins based on contextual local evolutionary conservation [45]. Finally, additional potential binding sites in Jaburetox were identified by the ANCHOR algorithm [46]. This approach relies on the pairwise energy estimation approach developed for the general disorder prediction method IUPRED, being based on the hypothesis that long regions of disorder contain localized potential binding sites that cannot form enough favorable intra-chain interactions to fold on their own, but are likely to gain stabilizing energy by interacting with a globular protein partner. Here we are using the term ANCHOR-indicated binding site (AIBS) to identify a region of a protein suggested by the ANCHOR algorithm to have significant potential to be a binding site for an appropriate but typically unidentified partner protein.

Hydrodynamic properties of Jaburetox

The molecular mass and hydrodynamic radius of Jaburetox in solution were determined using a combination of SEC, MALS and QELS. Jaburetox (140 μM, 400 μL) in buffer B was loaded onto a Superdex 75 10/300 GL column (GE Healthcare), pre-equilibrated with the same buffer, and eluted at room temperature at a flow rate of 0.6 $mL \cdot min^{-1}$. The column was connected downstream to a multi-angle laser light (690.0 nm) scattering (MALS) DAWN EOS photometer and to a 90° angle quasi-elastic (dynamic) light scattering (QELS) device (Wyatt Technology). The concentration of the eluted protein was determined using a refractive index detector (Optilab DSP, Wyatt). Values of 0.185 $mL \cdot g^{-1}$ for the refractive index increment (dn/dc) and of 1.330 for the solvent refractive index were used. Molecular masses were determined from a Zimm plot, using the Zimm equation [99] with a fitting degree of 1. Data were analyzed using the ASTRA 4.90.07 software (Wyatt Technology), following the manufacturer's indications.

Circular dichroism spectroscopy

The CD spectra of Jaburetox (24 μM in 50 mM phosphate, pH 7.5) were measured using a Jasco 810 spectropolarimeter flushed with N_2 and a cuvette with 0.1 cm path length. The spectra were registered from 190 to 240 nm with 0.2 nm intervals, at 25 °C and 90 °C. Ten spectra for each condition were accumulated and averaged to achieve an appropriate signal-to-noise ratio. The spectrum of the buffer was subtracted. The web server CAPITO (CD Analysis and Plotting Tool, http://capito.nmr.fli-leibniz.de/index.php) [32] was used to evaluate the fold state (unfolded, pre-molten globule, molten globule and globular) of Jaburetox according to its CD spectrum.

Differential scanning fluorimetry

Differential scanning fluorimetry experiments were performed using the Slice pH (HR2-070) kit from Hampton Research (USA). The experiments were conducted in an Mx3000P qPCR system (Agilent). Fluorescent signals were acquired with excitation and emission wavelengths of 495 and 520 nm, respectively. Temperature scans were performed from 25 °C to 90 °C in 1 $°C \cdot min^{-1}$ increments. A solution of 500 × SYPRO Orange (Life Technologies) in 100% DMSO was prepared from the 5000 × stock solution and diluted 100-fold in MilliQ water to prepare the working solution (5×). Experiments were performed in a 96-well plate, each well containing 20 μL of final solution. The solutions were prepared with 2 μL kit buffer (final concentration 100 mM), 2 μL Jaburetox (final concentration 6.4 μM) and 4 μL SYPRO Orange 5 × working solution [100]. The fluorescence data were acquired and the curves were analyzed using the ORIGIN PRO8 software (1991–2007 Origin Lab Corporation). The melting temperatures (T_m) for every condition were calculated from the first derivative of the melting peak.

NMR spectroscopy data collection and analysis for backbone assignment

NMR spectra were acquired at 298 K on a Bruker Avance 800 spectrometer, operating at the proton nominal frequency of 800.13 MHz (18.8 T) on NMR samples containing 1 mM of ^{15}N and ^{13}C uniformly labeled Jaburetox in 50 mM phosphate buffer, pH 6.5, 1 mM EDTA, 1 mM

TCEP, in 90% H_2O/10% D_2O. The spectrometer was equipped with a TXI 5-mm triple resonance cryo-probe with pulsed field gradients along the *z*-axis, using the acquisition parameters provided in Table 2. 2D ^{15}N-^{1}H HSQC, 2D ^{1}H-^{13}C HSQC, 3D HNCA, HNcoCA, HNCACB, HNcoCACB, HNCO, HNcaCO, HCCH-TOCSY and ^{15}N-^{1}H/^{13}C-^{1}H NOESY experiments were used to obtain the sequential backbone and side chain resonance assignment. In these pulse schemes, water suppression is achieved using selective pulse and transverse signal cancellation with pulsed field gradients associated with a flip-back pulse. The NMR data were processed using TOPSPIN 3.2 (Bruker BioSpin). Spectral analysis for resonance assignment was performed using CARA 1.8.4.2 [47].

NMR spectroscopy data collection and analysis for backbone mobility

The experiments for the determination of ^{15}N longitudinal ($R_1 = 1/T_1$) and transverse ($R_2 = 1/T_2$) relaxation rates, and of the ^{1}H-^{15}N cross-relaxation rate measured via steady-state heteronuclear ^{1}H-^{15}N NOE, were acquired on a Bruker Avance 800 spectrometer, operating at the proton nominal frequency of 800.13 MHz (18.8 T) using NMR samples containing 1 mM of ^{15}N-labeled Jaburetox in 50 mM phosphate buffer, pH 6.5, 1 mM EDTA, 1 mM TCEP, in 90% H_2O/10% D_2O at 298 K, using phase-sensitive gradient-enhanced sequences (Table 2) [66,70]. In these pulse schemes, water suppression is achieved using selective pulse and transverse signal cancellation with pulsed field gradients with a flip-back pulse in order to avoid saturation of water magnetization that could affect the NOE values of solvent-exposed backbone amide protons. The NMR experiments used to measure relaxation rates consist of a series of ^{1}H-^{15}N HSQC experiments in which spectra are acquired at different time intervals after the radiofrequency pulse. T_1 measurements were based on inversion-recovery type experiments recorded using 10 different delays: 10, 150, 300, 450, 600, 750, 900, 1150, 1300 and 1500 ms. T_2 measurements were carried out using Carr–Purcell–Meiboom–Gill (CPMG) spin-echo pulse sequences recorded with 10 different delays prior to the 180° refocusing pulse: 16, 50, 100, 130, 160, 200, 250, 300, 350 and 400 ms. Recycle delays of 4 s were used in both experiments. ^{1}H-^{15}N NOE values were obtained recording two sets of spectra in the presence and absence of a 6 s proton saturation period. The two spectra were recorded in an interleaved manner to ensure identical conditions in the two experiments. Spectra were processed using TOPSPIN 3.2 (Bruker BioSpin) and peak intensities were analyzed using DYNAMICS CENTER 2.1.5 (Bruker BioSpin). The values of R_1 and R_2 were calculated by fitting the peak intensities in ^{1}H-^{15}N HSQC spectra acquired at different relaxation delays to a two-parameter exponential decay function, using a nonlinear Marquard algorithm. The errors on the parameters derived from the fit were estimated from the inverse of the weighted curvature matrix, using a confidence level of 95%. The heteronuclear NOE values were estimated as the ratio of the intensities in the saturated versus non-saturated spectra. In all cases, the uncertainty on the intensities was derived from the standard deviation of the noise in each spectrum. The DYNAMICS CENTER 2.1.5 (Bruker BioSpin) software was used to carry out the quantitative analysis of the relaxation data.

Calculation of the structure of Jaburetox using NMR spectroscopy

NOE cross-peaks were identified, assigned and integrated in ^{15}N-edited and ^{13}C-edited 3D NOESY-HSQC spectra

Table 2. NMR acquisition parameters of Jaburetox.

Experiment	Label	Sequence	TD	SI	SW (ppm)
^{1}H-^{15}N HSQC	^{15}N	hsqcfpf3gpphwg	1024 × 256	1024 × 1024	11.2 × 40
^{1}H-^{13}C HSQC	^{13}C	hsqcctetgpsisp	1536 × 256	1024 × 512	11.2 × 76
HNCA	^{15}N/^{13}C	hncagp3d	1024 × 64 × 110	512 × 64 × 128	11.2 × 22 × 44
HNCO	^{15}N/^{13}C	hncogp3d	1024 × 32 × 96	1024 × 32 × 128	11.2 × 36 × 9.2
CBCAcoNH	^{15}N/^{13}C	cbcaconhgp3d	1024 × 64 × 128	512 × 64 × 128	11.2 × 22 × 62
HNCACB	^{15}N/^{13}C	hncacbgp3d	1024 × 64 × 128	512 × 64 × 128	11.2 × 22 × 62
HNcoCA	^{15}N/^{13}C	hncocagp3d	1024 × 64 × 110	512 × 64 × 128	11.2 × 22 × 44
HNCACO	^{15}N/^{13}C	hncacogp3d	1024 × 32 × 96	512 × 32 × 64	11.2 × 36 × 9.2
HCCH-TOCSY	^{15}N/^{13}C	hcchdigp3d2	1536 × 64 × 128	1024 × 64 × 128	11.2 × 76 × 76
^{13}C-NOESY	^{15}N/^{13}C	noesyhsqcetgp3d	1536 × 56 × 164	1024 × 64 × 128	11.2 × 76 × 11.2
^{15}N-NOESY	^{15}N/^{13}C	noesyhsqcf3gp193d	1536 × 48 × 236	512 × 64 × 256	11.2 × 22 × 11.2
^{1}H-^{15}N HSQC (for R_1)	^{15}N	hsqct1etf3gpsi	1536 × 256	2048 × 1024	11.2 × 22
^{1}H-^{15}N HSQC (for R_2)	^{15}N	hsqct2etf3gpsi	1536 × 256	1024 × 1024	11.2 × 22
^{1}H-^{15}N HSQC (for NOEs)	^{15}N	hsqcnoef3gpsi	1536 × 512	2048 × 1024	11.2 × 22
^{1}H-^{15}N HSQC (for in-cell NMR)	^{15}N	sfhmqcf3gpph	1024 × 128	2048 × 128	16 × 50

TD is the total number of real+imaginary points collected. SI is the total number of real+imaginary points before the Fourier transform. SW is the spectral width.

using CARA 1.8.4.2 [47]. The CALIBA subroutine in CYANA 2.1 (www.cyana.org) was used to convert cross-peak intensities to distance constraints. Dihedral angle constraints were derived using TALOS+ [50]. The structure was calculated using the torsion angle dynamics program CYANA 2.1. A total of 100 random conformers were annealed in 8000 steps using NOE and dihedral angle constraints. Five conformers were selected on the basis of their lowest and consistent target function values. Their structures were placed in truncated octahedral water boxes using a 10-Å thick buffer zone of solvent around the protein. The Amber ff99SB force field [101] for the protein and the TIP3P water model [102] were used. Ions were added to reproduce the experimental ion concentration used for NMR structures. Each system was geometry optimized using GROMACS 4.6 [54–57]. The quality of the obtained structures was analyzed using the PROCHECK PDB validation server (http://deposit.pdb.org/validate/).

NMR spectroscopy data collection in cell

BL21(DE3) *E. coli* cells expressing Jaburetox were grown in 70 mL of LB medium containing 100 $\mu g \cdot mL^{-1}$ ampicillin, at 28 °C for 16 h. Subsequently, cells were centrifuged at 2000 ***g*** for 15 min at 25 °C, and gently resuspended into 50 mL of M9 medium containing $(^{15}NH_4)_2SO_4$. After 15 min, protein overexpression was induced with 0.8 mM IPTG and continued for 4 h. Then, 25 mL of the cellular culture was centrifuged at 2000 ***g*** for 15 min and gently resuspended with 300 μL of M9 medium containing 10% v/v D_2O. The obtained NMR sample reached a final volume of 600 μL and contained a 50% v/v cellular suspension. NMR spectra were acquired at 310 K on a Bruker Avance 950 spectrometer operating at the proton nominal frequency of 950.20 MHz (22.3 T). The spectrometer was equipped with a TCI 5-mm cryo-probe with pulsed field gradients along the *z*-axis, using a SOFAST-HMQC pulse sequence and the acquisition parameters provided in Table 2.

Acknowledgements

Massimo Lucci and Enrico Luchinat are thanked for acquiring respectively *in vitro* and in-cell NMR spectra at the Center for Magnetic Resonance (Sesto Fiorentino, Italy). Dr Cristian Follmer is acknowledged for critically reading the first version of the manuscript.

Funding

This project was financed by the Brazilian agency Conselho de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Projeto Pesquisador Visitante PVE 054/2012, Ciências sem Fronteiras (Science without Borders). F.C.L received from CAPES a split PhD fellowship at the University of Bologna for this project. O.D. was supported by a fellowship financed by Specialty Fertilizer Products (Leawood, KS, USA). The research is partially funded by CIRMMP (Consorzio Interuniversitario di Risonanze Magnetiche di Metallo-Proteine) [104]. The WeNMR project (European FP7 e-Infrastructure grant, contract no. 261572, www.wenmr.eu) supported by the European Grid Initiative (http://www.egi.eu/) through the national GRID Initiatives of Belgium, France, Italy, Germany, the Netherlands, Poland, Portugal, Spain, UK, South Africa, Malaysia, Taiwan, the Latin America GRID infrastructure via the Gisela project (http://www.gisela-grid.eu/), the International Desktop Grid Federation (http://desktopgridfederation.org/) with its volunteers and the US Open Science Grid (http://www.opensciencegrid.org/) are acknowledged for the use of web portals, computing and storage facilities.

Author contributions

FCL produced and purified Jaburetox, and carried out differential scanning fluorimetry experiments; OD analyzed the NMR data, performed the assignment, and calculated the structural and dynamic properties of Jaburetox; RRG produced labeled Jaburetox for initial NMR studies; VB complemented experiments on differential scanning fluorimetry; VNU carried out the bioinformatics analysis; BZ carried out light scattering and CD experiments, and in-cell NMR; FM carried out computational analysis of the experimental and theoretical structures; SC processed and analyzed the NMR data; CRC and SC designed the research; all authors contributed to data analysis and wrote the paper.

References

1 Maroney MJ & Ciurli S (2013) Nonredox nickel enzymes. *Chem Rev* **114**, 4206–4228.
2 Mobley HLT & Hausinger RP (1989) Microbial urease: significance, regulation and molecular characterization. *Microbiol Rev* **53**, 85–108.
3 Karplus PA, Pearson MA & Hausinger RP (1997) 70 years of crystalline urease: what have we learned? *Acc Chem Res* **30**, 330–337.
4 Krajewska B (2009) Ureases I. Functional, catalytic and kinetic properties: a review. *J Mol Cat B: Enzymatic* **59**, 9–21.
5 Zambelli B, Musiani F, Benini S & Ciurli S (2011) Chemistry of Ni^{2+} in urease: sensing, trafficking, and catalysis. *Acc Chem Res* **44**, 520–530.

6 Sumner JB (1926) The isolation and crystallization of the enzyme urease. *J Biol Chem* **69**, 435–441.
7 Dixon NE, Gazzola TC, Blakeley RL & Zermer B (1975) Jack bean urease (EC 3.5.1.5). A metalloenzyme. A simple biological role for nickel? *J Am Chem Soc* **97**, 4131–4133.
8 Pires-Alves M, Grossi-de-Sá MF, Barcellos GB, Carlini CR & Moraes MG (2003) Characterization and expression of a novel member (JBURE-II) of the urease gene family from jackbean [*Canavalia ensiformis* (L.) DC]. *Plant Cell Physiol* **44**, 139–145.
9 Mulinari F, Becker-Ritt AB, Demartini DR, Ligabue-Braun R, Stanisçuaski F, Verli H, Fragoso RR, Schroeder EK, Carlini CR & Grossi-de-Sá MF (2011) Characterization of JBURE-IIb isoform of *Canavalia ensiformis* (L.) DC urease. *BBA-Proteins Proteom* **1814**, 1758–1768.
10 Follmer C, Barcellos GB, Zingali RB, Machado OL, Alves EW, Barja-Fidalgo C, Guimarães JA & Carlini CR (2001) Canatoxin, a toxic protein from jack beans (*Canavalia ensiformis*), is a variant form of urease (EC 3.5.1.5): biological effects of urease independent of its ureolytic activity. *Biochem J* **360**, 217–224.
11 Carlini CR & Guimarães JA (1981) Isolation and characterization of a toxic protein from *Canavalia ensiformis* (jack bean) seeds, distinct from concanavalin A. *Toxicon* **19**, 667–675.
12 Carlini CR, Oliveira AE, Azambuja P, Xavier-Filho J & Wells MA (1997) Biological effects of canatoxin in different insect models: evidence for a proteolytic activation of the toxin by insect cathepsin-like enzymes. *J Econ Entomol* **90**, 340–348.
13 Ferreira Da Silva CT, Gombarovits MEC, Masuda H, Oliveira CM & Carlini CR (2000) Proteolytic activation of canatoxin, a plant toxic protein, by insect cathepsin-like enzymes. *Arch Insect Biochem* **44**, 162–171.
14 Carlini CR & Grossi-de-Sá MF (2002) Plant toxic proteins with insecticidal properties. A review on their potentialities as bioinsecticides. *Toxicon* **40**, 1515–1539.
15 Staniscuaski F, Ferreira-DaSilva C, Mulinari F, Pires-Alves M & Carlini C (2005) Insecticidal effects of canatoxin on the cotton stainer bug *Dysdercus peruvianus* (Hemiptera: Pyrrhocoridae). *Toxicon* **45**, 753–760.
16 Defferrari MS, Demartini DR, Marcelino TB, Pinto PM & Carlini CR (2011) Insecticidal effect of *Canavalia ensiformis* major urease on nymphs of the milkweed bug *Oncopeltus fasciatus* and characterization of digestive peptidases. *Insect Biochem Molec* **41**, 388–399.
17 Mulinari F, Staniscuaski F, Bertholdo-Vargas L, Postal M, Oliveira-Neto O, Rigden D, Grossi-de-Sa M & Carlini C (2007) Jaburetox-2Ec: an insecticidal peptide derived from an isoform of urease from the plant *Canavalia ensiformis*. *Peptides* **28**, 2042–2050.
18 Stanisçuaski F, Te Brugge V, Carlini CR & Orchard I (2009) *In vitro* effect of *Canavalia ensiformis* urease and the derived peptide Jaburetox-2Ec on *Rhodnius prolixus* Malpighian tubules. *J Insect Physiol* **55**, 255–263.
19 Barros PR, Stassen H, Freitas MS, Carlini CR, Nascimento MA & Follmer C (2009) Membrane-disruptive properties of the bioinsecticide Jaburetox-2Ec: implications to the mechanism of the action of insecticidal peptides derived from ureases. *Biochim Biophys Acta* **1794**, 1848–1854.
20 Postal M, Martinelli AH, Becker-Ritt AB, Ligabue-Braun R, Demartini DR, Ribeiro SF, Pasquali G, Gomes VM & Carlini CR (2012) Antifungal properties of *Canavalia ensiformis* urease and derived peptides. *Peptides* **38**, 22–32.
21 Balasubramanian A & Ponnuraj K (2010) Crystal structure of the first plant urease from jack bean: 83 years of journey from its first crystal to molecular structure. *J Mol Biol* **400**, 274–283.
22 Balasubramanian A, Balaji N, Gautham N & Ponnuraj K (2013) Molecular dynamics simulation and molecular modelling studies on the insecticidal domain from jack bean urease. *Mol Simulat* **39**, 357–366.
23 Martinelli AH, Kappaun K, Ligabue-Braun R, Defferrari MS, Piovesan AR, Stanisçuaski F, Demartini DR, Dal Belo CA, Almeida CG, Follmer C *et al.* (2014) Structure-function studies on jaburetox, a recombinant insecticidal peptide derived from jack bean (*Canavalia ensiformis*) urease. *Biochim Biophys Acta* **1840**, 935–944.
24 Piovesan AR, Martinelli AH, Ligabue-Braun R, Schwartz J-L & Carlini CR (2014) *Canavalia ensiformis* urease, Jaburetox and derived peptides form ion channels in planar lipid bilayers. *Arch Biochem Biophys* **547**, 6–17.
25 Marley J, Lu M & Bracken C (2001) A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* **20**, 71–75.
26 Iakoucheva LM, Kimzey AL, Masselon CD, Smith RD, Dunker AK & Ackerman EJ (2001) Aberrant mobility phenomena of the DNA repair protein XPA. *Protein Sci: Pub Protein Soc* **10**, 1353–1362.
27 Receveur-Brechot V, Bourhis JM, Uversky VN, Canard B & Longhi S (2006) Assessing protein disorder and induced folding. *Proteins* **62**, 24–45.
28 Kozlowska M, Tarczewska A, Jakob M, Szpotkowski K, Wojtas M, Rymarczyk G & Ozyhar A (2014) Calponin-like chd64 is partly disordered. *PLoS One* **9**, e96809.
29 Follmer C, Pereira FV, Da Silveira NP & Carlini CR (2004) Jack bean urease (EC 3.5.1.5) aggregation

monitored by dynamic and static light scattering. *Biophys Chem* **111**, 79–87.
30 Boivin S, Kozak S & Meijers R (2013) Optimization of protein purification and characterization using Thermofluor screens. *Protein Expr Purif* **91**, 192–206.
31 Uversky VN (2002) What does it mean to be natively unfolded? *Eur J Biochem* **269**, 2–12.
32 Wiedemann C, Bellstedt P & Görlach M (2013) CAPITO–a web server-based analysis and plotting tool for circular dichroism data. *Bioinformatics* **29**, 1750–1757.
33 Uversky VN (2009) Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *Protein J* **28**, 305–325.
34 Zambelli B, Cremades N, Neyroz P, Turano P, Uversky VN & Ciurli S (2012) Insights in the (un) structural organization of *Bacillus pasteurii* UreG, an intrinsically disordered GTPase enzyme. *Mol BioSyst* **8**, 220–228.
35 Corrêa DHA & Ramos CHI (2009) The use of circular dichroism spectroscopy to study protein folding, form and function. *African J Biochem Res* **3**, 164–173.
36 Anderson VL, Ramlall TF, Rospigliosi CC, Webb WW & Eliezer D (2010) Identification of a helical intermediate in trifluoroethanol-induced alpha-synuclein aggregation. *Proc Natl Acad Sci USA* **107**, 18850–18855.
37 Park SHS, Shalongo WW & Stellwagen EE (1997) The role of PII conformations in the calculation of peptide fractional helix content. *Protein Sci* **6**, 1694–1700.
38 Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN & Dunker AK (2007) Intrinsic disorder and functional proteomics. *Biophys J* **92**, 1439–1456.
39 Vacic V, Uversky VN, Dunker AK & Lonardi S (2007) Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics* **8**, 211–217.
40 Uversky VN, Gillespie JR & Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **41**, 415–427.
41 Romero P, Obradovic Z, Li X, Garner EC, Brown CJ & Dunker AK (2001) Sequence complexity of disordered protein. *Proteins* **42**, 38–48.
42 Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN & Dunker AK (2005) Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* **44**, 12454–12470.
43 Cheng Y, Oldfield CJ, Meng J, Romero P, Uversky VN & Dunker AK (2007) Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. *Biochemistry* **46**, 13468–13477.
44 Disfani FM, Hsu WL, Mizianty MJ, Oldfield CJ, Xue B, Dunker AK, Uversky VN & Kurgan L (2012) MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics* **28**, i75–i83.
45 Fang C, Noguchi T, Tominaga D & Yamana H (2013) MFSPSSMpred: identifying short disorder-to-order binding regions in disordered proteins based on contextual local evolutionary conservation. *BMC Bioinformatics* **14**, 300.
46 Dosztanyi Z, Meszaros B & Simon I (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* **25**, 2745–2746.
47 Keller R & Wuthrich K (2004) Computer-Aided Resonance Assignment (CARA). Verlag Goldau, Cantina, Switzerland.
48 Jensen MR, Salmon L, Nodet G & Blackledge M (2010) Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. *J Am Chem Soc* **132**, 1270–1272.
49 Marsh JA, Singh VK, Jia Z & Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between α-and γ-synuclein: implications for fibrillation. *Protein Sci: Pub Protein Soc* **15**, 2795–2804.
50 Shen Y, Delaglio F, Cornilescu G & Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* **44**, 213–223.
51 Ozenne V, Bauer F, Salmon L, Huang JR, Jensen MR, Segard S, Bernado P, Charavay C & Blackledge M (2012) Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* **28**, 1463–1470.
52 Gast K, Damaschun H, Misselwitz R, Muller-Frohne M, Zirwer D & Damaschun G (1994) Compactness of protein molten globules: temperature-induced structural changes of the apomyoglobin folding intermediate. *EBJ* **23**, 297–305.
53 Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones JA & Smith LJ (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **38**, 16424–16431.
54 Berendsen HJC, Vanderspoel D & Vandrunen R (1995) Gromacs – a Message-Passing Parallel Molecular-Dynamics Implementation. *Comput Phys Commun* **91**, 43–56.
55 Lindahl E, Hess B & Van Der Spoel D (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model* **7**, 306–317.

56 Van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE & Berendsen HJC (2005) Gromacs: Fast, flexible, and free. *J Comput Chem* **26**, 1701–1718.

57 Hess B, Kutzner C, van der Spoel D & Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* **4**, 435–447.

58 Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF & Mark AE (1999) Peptide folding: when simulation meets experiment. *Angew Chem Int Ed* **38**, 236–240.

59 Reckel S, Hänsel R, Löhr F & Dötsch V (2007) In-cell NMR spectroscopy. *Prog Nucl Mag Res Sp* **51**, 91–101.

60 Selenko P & Wagner G (2007) Looking into live cells with in-cell NMR spectroscopy. *J Struct Biol* **158**, 244–253.

61 Pielak GJ, Li C, Miklos AC, Schlesinger AP, Slade KM, Wang G-F & Zigoneanu IG (2009) Protein nuclear magnetic resonance under physiological conditions. *Biochemistry* **48**, 226–234.

62 Sakakibara D, Sasaki A, Ikeya T, Hamatsu J, Hanashima T, Mishima M, Yoshimasu M, Hayashi N, Mikawa T & Wälchli M (2009) Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature* **458**, 102–105.

63 McNulty BC, Young GB & Pielak GJ (2006) Macromolecular crowding in the *Escherichia coli* periplasm maintains α-synuclein disorder. *J Mol Biol* **355**, 893–897.

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

65 Palmer AG III (1993) Dynamic properties of proteins from NMR spectroscopy. *Curr Opin Biotechnol* **4**, 385–391.

66 Kay LE, Torchia DA & Bax A (1989) Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**, 8972–8979.

67 Fushman D, Weisemann R, Thüring H & Rüterjans H (1994) Backbone dynamics of ribonuclease T_1 and its complex with 2′ GMP studied by two-dimensional heteronuclear NMR spectroscopy. *J Biomol NMR* **4**, 61–78.

68 Rossi P, Swapna G, Huang YJ, Aramini JM, Anklin C, Conover K, Hamilton K, Xiao R, Acton TB & Ertekin A (2010) A microscale protein NMR sample screening pipeline. *J Biomol NMR* **46**, 11–22.

69 Peng JW & Wagner G (1992) Mapping of spectral density functions using heteronuclear NMR relaxation measurements. *J Magn Reson* **98**, 308–332.

70 Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD & Kay LE (1994) Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* **33**, 5984–6003.

71 Ishima R & Nagayama K (1995) Protein backbone dynamics revealed by quasi spectral density function analysis of amide N-15 nuclei. *Biochemistry* **34**, 3162–3171.

72 Farrow NA, Zhang O, Szabo A, Torchia DA & Kay LE (1995) Spectral density function mapping using ^{15}N relaxation data exclusively. *J Biomol NMR* **6**, 153–162.

73 Lipari G & Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc* **104**, 4546–4559.

74 Lipari G & Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J Am Chem Soc* **104**, 4559–4570.

75 Clore GM, Szabo A, Bax A, Kay LE, Driscoll PC & Gronenborn AM (1990) Deviations from the simple two-parameter model-free approach to the interpretation of ^{15}N nuclear magnetic relaxation of proteins. *J Am Chem Soc* **112**, 4989–4991.

76 D'Auvergne EJ & Gooley PR (2008) Optimisation of NMR dynamic models II. A new methodology for the dual optimisation of the model-free parameters and the Brownian rotational diffusion tensor. *J Biomol NMR* **40**, 121–133.

77 Almeida MS, Cabral K, Kurtenbach E, Almeida FC & Valente AP (2002) Solution structure of *Pisum sativum* defensin 1 by high resolution NMR: plant defensins, identical backbone with different mechanisms of action. *J Mol Biol* **315**, 749–757.

78 de Medeiros LN, Angeli R, Sarzedas CG, Barreto-Bergter E, Valente AP, Kurtenbach E & Almeida FCL (2010) Backbone dynamics of the antifungal Psd1 pea defensin and its correlation with membrane interaction by NMR spectroscopy. *Biochim Biophys Acta* **1798**, 105–113.

79 Fant F, Vranken W, Broekaert W & Borremans F (1998) Determination of the three-dimensional solution structure of *Raphanus sativus* antifungal protein 1 by ^{1}H NMR. *J Mol Biol* **279**, 257–270.

80 Patel SU, Osborn R, Rees S & Thornton JM (1998) Structural studies of Impatiens balsamina antimicrobial protein (Ib-AMP1). *Biochemistry* **37**, 983–990.

81 Huang R-H, Xiang Y, Tu G-Z, Zhang Y & Wang D-C (2004) Solution structure of *Eucommia* antifungal peptide: a novel structural model distinct with a five-disulfide motif. *Biochemistry* **43**, 6005–6012.

82 Liu YJ, Cheng CS, Lai SM, Hsu MP, Chen CS & Lyu PC (2006) Solution structure of the plant defensin VrD1 from mung bean and its possible role in insecticidal activity against bruchids. *Proteins* **63**, 777–786.
83 Jennings CV, Rosengren KJ, Daly NL, Plan M, Stevens J, Scanlon MJ, Waine C, Norman DG, Anderson MA & Craik DJ (2005) Isolation, solution structure, and insecticidal activity of kalata B2, a circular protein with a twist: do Möbius strips exist in nature? *Biochemistry* **44**, 851–860.
84 Lay FT, Schirra HJ, Scanlon MJ, Anderson MA & Craik DJ (2003) The three-dimensional solution structure of NaD1, a new floral defensin from *Nicotiana alata* and its application to a homology model of the crop defense protein alfAFP. *J Mol Biol* **325**, 175–188.
85 Mittag T & Forman-Kay JD (2007) Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* **17**, 3–14.
86 Jensen MR, Ruigrok RW & Blackledge M (2013) Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr Opin Struct Biol* **23**, 426–435.
87 Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, Mylonas E, Svergun DI, Blackledge M & Fersht AR (2008) Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc Natl Acad Sci USA* **105**, 5762–5767.
88 Mittag T, Marsh J, Grishaev A, Orlicky S, Lin H, Sicheri F, Tyers M & Forman-Kay JD (2010) Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. *Structure* **18**, 494–506.
89 Tompa P (2012) Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci* **37**, 509–516.
90 Uversky VN (2013) Unusual biophysics of intrinsically disordered proteins. *BBA-Proteins Proteom* **1834**, 932–951.
91 Haaland PD (1989) Experimental Design in Biotechnology. CRC press, New York, NY.
92 Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248–254.
93 Griep S & Hobohm U (2009) PDBselect 1992-2009 and PDBfilter-select. *Nucleic Acids Res* **38**, D318–D319.
94 Rose PW, Beran B, Bi C, Bluhm WF, Dimitropoulos D, Goodsell DS, Prlic A, Quesada M, Quinn GB, Westbrook JD *et al.* (2011) The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res* **39**, D392–D401.
95 Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN *et al.* (2007) DisProt: the database of disordered proteins. *Nucleic Acids Res* **35**, D786–D793.
96 Peng K, Radivojac P, Vucetic S, Dunker AK & Obradovic Z (2006) Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* **7**, 208.
97 Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK & Obradovic Z (2005) Optimizing long intrinsic disorder predictors with protein evolutionary information. *J Bioinformat Computat Biol* **3**, 35–60.
98 Xue B, Dunbrack RL, Williams RW, Dunker AK & Uversky VN (2010) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta* **1804**, 996–1010.
99 Zimm BH (1948) The dependence of the scattering of light on angle and concentration in linear polymer solutions. *J Phys Colloid Chem* **52**, 260–267.
100 Sheth PR, Ramanathan L, Ranchod A, Basso AD, Barrett D, Zhao J, Gray K, Liu YH, Zhang R & Le HV (2010) Expression, purification, stability optimization and characterization of human Aurora B kinase domain from *E. coli*. *Arch Biochem Biophys* **503**, 191–201.
101 Hornak V, Abel R, Okur A, Strockbine B, Roitberg A & Simmerling C (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct Funct Bioinform* **65**, 712–725.
102 Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW & Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926–935.
103 Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC & Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612.
104 Wassenaar TA, van Dijk M, Loureiro-Ferreira N, van der Schot G, de Vries SJ, Schmitz C, van der Zwan J, Boelens R, Giachetti A, Ferella L, *et al.* (2012) WeNMR: Structural Biology on the Grid. *J Grid Comp* **10**, 743–767.