Compaction Properties of an Intrinsically Disordered Protein: Sic1 and Its Kinase-Inhibitor Domain

Stefania Brocca,[†△] Lorenzo Testa,[†△] Frank Sobott,[‡] Maria Šamalikova,[†] Antonino Natalello,[†] Elena Papaleo,[†] Marina Lotti,[†] Luca De Gioia,[†] Silvia Maria Doglia,[†] Lilia Alberghina,[†] and Rita Grandori[†*]

[†]Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milan, Italy; and [‡]Department of Chemistry, University of Antwerp, Antwerp, Belgium

ABSTRACT IDPs in their unbound state can transiently acquire secondary and tertiary structure. Describing such intrinsic structure is important to understand the transition between free and bound state, leading to supramolecular complexes with physiological interactors. IDP structure is highly dynamic and, therefore, difficult to study by conventional techniques. This work focuses on conformational analysis of the KID fragment of the Sic1 protein, an IDP with a key regulatory role in the cell-cycle of *Saccharomyces cerevisiae*. FT-IR spectroscopy, ESI-MS, and IM measurements are used to capture dynamic and short-lived conformational states, probing both secondary and tertiary protein structure. The results indicate that the isolated Sic1 KID retains dynamic helical structure and populates collapsed states of different compactness. A metastable, highly compact species is detected. Comparison between the fragment and the full-length protein suggests that chain length is crucial to the stabilization of compact states of this IDP. The two proteins are compared by a length-independent compaction index.

INTRODUCTION

CKIs are key regulatory proteins that modulate kinase activities through the eukaryotic cell cycle. CKIs carry out their inhibitory function by the formation of ternary complexes with the target kinase and a cognate cyclin. The dynamic composition of kinase complexes throughout the cell cycle determines at which time and in which intracellular compartment regulatory phosphorylation events will take place (1).

Most CKIs belong to the class of IDPs (2,3). These proteins lack an ordered three-dimensional structure in their free state and undergo folding upon binding to specific interactors. Intrinsic structural disorder seems to favor multiplicity of interactions, independent regulation of affinity and specificity, regulation by post-translational modifications, and fast association kinetics (4). Nevertheless, no clear correlation emerges from large-scale surveys between degree of disorder and number of interactors (5). Structural characterization of IDPs in their free state is essential to the interpretation of their biochemical properties but it is technically challenging, due to their dynamic and heterogeneous conformation (6). The available evidence suggests that IDPs in the absence of interactors transiently populate partially ordered or compact states (7–13).

The protein Sic1 is a yeast CKI that binds to the target Cdk1 kinase in complex with the G1 cyclins Clb5,6 (14). The formation of these inhibitory complexes prevents phosphorylation of Cdk1 G1 substrates, which is required to enter S phase. The G1-to-S transition is executed only upon Sic1 degradation, which is, in turn, triggered by multiple Sic1 phosphorylations in its N-terminal region (10,15). Sic1 intracellular levels rise again in late M phase. Sic1 has also been implicated in other regulatory processes, such as mitosis exit (16), coordination between cell growth and cell cycle (17,18), and defense against osmotic stress (19). The Sic1 KID has been identified as the minimal protein fragment required for complementation of the ΔSic1 phenotype in vivo (20). Such a fragment corresponds to the C-terminal 70, out of 284, amino acids forming the native protein.

Previous work has shown that Sic1 is disordered in its whole length, with some intrinsic propensity to form ordered helical structure (7). The Sic1 molecule also displays remarkable compactness that makes it more similar to a collapsed chain than to a random coil (7,10). Such intrinsic tertiary structure can be denatured by acids and organic solvents (7). Fragments corresponding to the N- and C-terminal moieties have been produced, showing that the little content in secondary structure is distributed quite uniformly throughout the chain length, although the C-terminus is slightly more ordered than the N-terminus (21). More remarkable differences between the N- and C-terminal regions emerge from limited proteolysis experiments, which hint to the C-terminal ~1/3 of the protein as the most resistant fragment (7). Consistent with such evidence, conformational analysis by ESI-MS suggests that the Sic1 C-terminal domain is more structured than its complementary N-terminal domain (22). Thus, altogether, functional and structural features point to a modular organization of this protein, despite its disordered nature.

Submitted July 12, 2010, and accepted for publication February 22, 2011.

△Stefania Brocca and Lorenzo Testa contributed equally to this work.

*Correspondence: rita.grandori@unimib.it

Abbreviations used: CKI, cyclin-dependent kinase inhibitor; KID, kinase inhibitory domain; IDP, intrinsically disordered protein; IFSU, intrinsically folded structural unit; NMR, nuclear magnetic resonance; FT-IR, Fourier-transform infrared; ESI, electrospray ionization; MS, mass spectrometry; IM, ion mobility; CSD, charge-state distribution; FSD, Fourier-self-deconvoluted; R_h, hydrodynamic radius; *CI*, compaction index.

Editor: Bertrand Garcia-Moreno.

0006-3495/11/05/2243/10 $2.00

doi: 10.1016/j.bpj.2011.02.055

In vitro and in vivo degradation patterns suggest that the boundary of the more compact C-terminal region lies close to residue 186 (7,21).

The Sic1 C-terminal region is structurally and functionally related to the mammalian p21 and p27 KID that is located, instead, at the protein N-terminus (23). Detailed structural characterization of the p27 KID has been achieved by NMR and could be compared to the crystallographic structure of the ternary complex with cyclin A and Cdk2 (24,25). The residue-level description of secondary structure indicates that, in the absence of interactors, short chain segments transiently assume ordered backbone conformation. These secondary-structure elements resemble those observed in the bound state and correspond to docking elements of the interaction surfaces. Therefore, these IFSUs (4) represent on-pathway nuclei for complex formation. The mechanism of p27 binding and induced folding has been investigated by an array of different biophysical techniques (4). The ternary complex forms via a stepwise process where one end of the p27 KID first binds to cyclin A and the opposite end binds to Cdk2, covering the ATP-binding site. Thus, the inhibitor wraps on the surface of the cyclin A/Cdk2 complex in a highly extended conformation engaging its two ends in specific interactions with quite distant binding sites. As a result, it acts as a structural and thermodynamic staple in the ternary complex (26).

Despite the detailed characterization of p21 and p27 secondary structure, very little is known about KID tertiary structure for CKIs in the absence of interactors. The goal of this work is to describe the conformational properties of Sic1 KID in its free state, both in terms of secondary and tertiary structure. The highly dynamic nature of IDPs makes it difficult to capture structural features that might arise only transiently in the unbound state. Here, conformational transitions of Sic1 KID are analyzed by complementary biophysical methods. Secondary structure is monitored by FT-IR spectroscopy. Highly dynamic helical elements are detected. Protein tertiary structure is probed by nondenaturing gel filtration, ESI-MS, and ESI-IM-MS. The extent of ionization by electrospray reflects protein compactness in the original solution, at the moment of transfer to the gas phase. The most compact states result in the lowest charge states (27–29). ESI-IM-MS resolves distinct conformers in the gas phase, based on their different shape, and allows for direct detection of compact species (30–32). The molecular ensemble of the isolated KID fragment is found to interconvert between collapsed states of different compactness. A small fraction of the population is found in a highly compact state. Comparison to the full-length protein hints to a critical role of chain length in determining the overall compaction of this IDP.

MATERIALS AND METHODS

Sequence analysis

The values of net charge per residues were derived according to Pappu and co-workers (33). Predictions of the R_h were performed according to Forman-Kay and co-workers (34). Sequence hydrophobicity was calculated by the Kyte and Doolittle approximation (35), using a window size of 5 amino acids by the ProtScale Expasy tool (http://expasy.org/tools/). The hydrophobicity of each residue was normalized to a scale between 0 and 1.

Protein expression and purification

The *Escherichia coli* strain BL21-Rosetta 2(DE3) (Novagen, EMD Chemicals, Darmstadt, DE) was used for heterologous expression of $Sic1^{\Delta 214}$ fused to a His_6 tag at its C-terminus (21). A glycine residue is inserted between the His_6 tag and the natural C-terminus of the protein (21). Cultures in low-salt Luria-Bertani medium were grown at 37°C and 220 rpm to an optical density at 600 nm (OD_{600}) between 0.5 and 0.7. Cultures were then transferred to 30°C and induced with 200 μM isopropyl-β-D-thiogalactopyranoside. Cells were harvested after 2 h by centrifugation at 9400 $\times$ g at room temperature for 15 min and either immediately used for protein extraction or frozen at −20°C. Protein extraction and purification were performed as previously described (21). Protein concentration was assessed by the Bradford protein assay (Bio-Rad Laboratories, Hercules, CA).

FT-IR spectroscopy

The FT-IR absorption spectra of $Sic1^{\Delta 214}$ at a concentration of 800 μM in D_2O, 280 mM NaCl were collected in transmission using a Varian 670-IR spectrometer (Varian Australia Pty Ltd., Mulgrave VIC, AU), equipped with a nitrogen-cooled Mercury Cadmium Telluride detector, under accurate dry-air purging. A sample volume of 20 μL was placed in a temperature-controlled transmission cell with BaF_2 windows separated by a 150 μm teflon spacer. Spectra were collected under the following conditions: 2 cm^{-1} resolution, 25 kHz scan speed, 1000 scans coaddition, and triangular apodization. For temperature-variable experiments, the sample was heated from 10 to 100°C at a rate of 0.4°C/min.

The $Sic1^{\Delta 214}$ FT-IR spectra were obtained after subtraction of the solvent absorption, collected under the same conditions. Measured spectra were smoothed by a binomial function (11 points) and the second derivatives were obtained by the Savitsky–Golay method (third grade polynomial, 5 smoothing points). The FT-IR spectra in transmission of *Candida rugosa* Lipase 1 (CRL1) and α-synuclein in D_2O solutions were also measured at 2 cm^{-1} spectral resolutions and under conditions similar to the $Sic1^{\Delta 214}$ samples (13,36). Curve fitting of the Amide-I band as a linear combination of Gaussian components was performed as previously described (36,37). Briefly, the number of components and the initial values of their peak positions were taken from the second derivative and FSD spectra. The curve fitting was performed in two steps. First, peak wavenumbers were kept fixed and the other parameters were adjusted iteratively. In the second step, peak positions were allowed to change as well, to optimize the set of fitting functions. The final peak positions were found to be identical to those observed in the second derivative spectrum. The fractional area of each Gaussian function over the total area reflects the percentage of the corresponding secondary structure.

The above analysis was performed using the GRAMS/32 software (Galactic Industries Corporation, Salem, NH), with the exception of the FSD analysis, for which the Resolution-Pro software (Varian Australia Pty Ltd, Mulgrave VIC, AU) was employed.

Analytical gel filtration

Gel filtration was performed on an ÄKTA purifier liquid-chromatography system, using a prepacked, 30 cm $\times$ 1 cm SuperdexTM 75 HR column (GE Healthcare, Little Chalfont, UK). Chromatography was carried out at room temperature in 50 mM sodium phosphate, pH 8.0, 200 mM NaCl at a flow rate of 0.5 mL/min, and monitored by absorbance at 280 nm.

A calibration curve was constructed using the following standards (0.75 mg/mL): bovine aprotinin (6,500 Da), horse cytochrome *c* (12,400 Da), horse myoglobin (17,600 Da), chicken ovalbumin (45,000 Da), equine apoferritin (80,000 Da) (Sigma Aldrich, St. Louis, MO), and recombinant green fluorescent protein (29,800 Da). A second calibration curve was constructed to calculate the R_h values of $Sic1^{FL}$ and $Sic1^{\Delta 214}$, by plotting the R_h values of the standards (ovalbumin, myoglobin, cytochrome *c*, and aprotinin) against their relatve elution volume (38–40).

Mass spectrometry

ESI-MS experiments were performed on a hybrid quadrupole-time-of-flight mass spectrometer (QSTAR ELITE, Applied Biosystems, Foster City, CA) equipped with a nano-ESI sample source. The samples contained 10 μM $Sic1^{\Delta 214}$ in 50 mM ammonium acetate at the proper pH, which was adjusted by ammonium hydroxide or formic acid before protein addition. Metal-coated borosilicate capillaries (Proxeon, Odense, DK), with medium-length emitter tip of 1 μm internal diameter, were used for nanospray. The instrument was calibrated using renine inhibitor (1757.9 Da) (Applied Biosystems) and its fragment (109.07 Da) as standards. Spectra were acquired in the 500–2000 *m/z* range, with accumulation time 1 s, ion-spray voltage 1200–1400 V, declustering potential 60–80 V, and instrument interface at room temperature. Spectra were averaged over a time period of 2 min. Data analysis was performed by the program Analyst QS 2.0 (Applied Biosystems). Gaussian fitting of ESI-MS spectra was carried out on row data reporting ion relative intensity versus charge (22). These data were fitted by the minimal number of Gaussian functions leading to a stable fit. The relative peak areas were used to build the profile of the pH-dependent transition. Data from three independent acquisitions at each pH values were fitted by a Boltzmann sigmoidal function (final $R^2 > 0.98$). Fitting analyses were performed by the software OriginPro 7.5 (Originlab, Northampton, MA).

IM

IM-mass spectrometry was performed under nondenaturing conditions, by reconstituting samples into aqueous ammonium acetate buffer (pH 7.0) and removal of nonvolatile salts and buffers using micro bio-spin buffer exchange columns (Bio-Rad Laboratories). The protein sample (10 μM) was introduced into the mass spectrometer using nano-ESI with an Advion Triversa inlet system (Advion, Ithaca, NY). Mass-to-charge values and drift time of positive ions were determined using a Waters Synapt 1 T-wave IMS quadrupole-time-of-flight MS/MS system (Waters Corporation, Milford MA). Experimental parameters were chosen to optimize desolvation of sample ions, maintaining conditions that promote the transmission of intact protein ions, while possible disruption of native conformations is minimized (Sampling Cone 60 V, Extraction Cone 0 V, Trap Collision Energy 6.0 V, Transfer Collision Energy 4.0 V, Trap DC Bias 20.0 V, Backing pressure 1.47e0 mbar and Source Pressure 4.66e-4 mbar, IMS Wave Velocity 350 m/s, IMS Wave Height 8.0 V).

RESULTS AND DISCUSSION

Sequence analysis

Recent analyses of IDPs phase diagrams point to mean hydrophobicity and mean net charge as discriminating order parameters (33,34,41–43). The mean hydrophobicity value $<H>$ for $Sic1^{\Delta 214}$ is quite low, 0.33, and close to the value calculated for the full-length Sic1 ($Sic1^{FL}$) protein (0.38, Table 1). Such a low hydrophobicity places Sic1 and its fragment in the region of the natively unfolded proteins in the bidimensional diagram developed by Uversky and co-workers (42). This result is consistent with the already reported conclusion that Sic1 is disordered in its whole length (7). To evaluate sequence polarity, the net charge per residue was calculated as the difference between the fraction of positively charged residues (f_+) and the fraction of negatively charged residues (f_-), according to Pappu and co-workers (33). The $f_+ - f_-$ values for $Sic1^{\Delta 214}$ and $Sic1^{FL}$ are well below the recognized threshold of 0.2, discriminating globular and coil-like IDPs. These results suggest that both $Sic1^{\Delta 214}$ and $Sic1^{FL}$ would preferentially populate collapsed states. In particular, according to Pappu and co-workers (33), these proteins would belong to the category of the weak polyelectrolytes/polyampholytes, characterized by relatively small values of both parameters f_+ and f_-. Altogether, these sequence features would identify $Sic1^{\Delta 214}$ and $Sic1^{FL}$ as disordered globules in the three-dimensional phase diagram proposed by Pappu and co-workers (33), where they would lie quite close to the axis origin.

TABLE 1 Disorder parameters for $Sic1^{\Delta 214}$ and $Sic1^{FL}$

	$Sic1^{\Delta 214}$		$Sic1^{FL}$	
	No tag	Histidine tag	No tag	Histidine tag
N	70	77	284	291
f_-	0.214	0.195	0.148	0.144
f_+	0.243	0.221	0.151	0.148
$f_+ - f_-$	0.029	0.026	0.004	0.004
$<H>$	0.33	0.32	0.38	0.37
Q	2	2	1	1
P_{pro}	0.043	0.039	0.095	0.093
Calculated R_h (Å)	21.64 (20.27)	22.72 (19.07)	44.15 (43.79)	44.70 (39.84)
R_h by gel filtration (Å)		20.57 ± 0.22		31.25 ± 3.53
CI by gel filtration		0.57 ± 0.03		0.77 ± 0.13

N, sequence length. $f_+ - f_-$, net charge per residue as defined by Pappu and co-workers (33), where f_+ and f_- are the fractions of negatively and positively charged residues, respectively. $<H>$, mean Kyte-Doolittle hydropathy score (35). Q, difference between the total number of negatively charged and positively charged residues. P_{pro}, fractional proline content. R_h was calculated by the simple power-law model $R_h = R_0 N^{\nu}$ (Eq. 2 of reference 34), with parameters for IDPs, $R_0 = 2.49$ Å and $\nu = 0.509$, or (values in parentheses) by the sequence-based model. $R_h = (AP_{pro} + B)(C|Q| + D)S_{his} \times R_0 N^{\nu}$ (Eq. 6 of reference 34), where $A = 1.24$, $B = 0.904$, $C = 0.00759$, $D = 0.963$, $S_{his} = 0.901$ (correction factor for histidine tag), $R_0 = 2.49$ and $\nu = 0.509$. The R_h values of folded and unfolded proteins used to calculate *CI* were derived from Eq. 2 of reference 34.

Secondary structure by FT-IR spectroscopy

The fragment produced in this work encompasses the sequence from residue 215 to the C-terminus and corresponds to the minimal Sic1 fragment retaining inhibitory activity (20,23). Protein expression and purification was carried out from a recombinant *E. coli* strain carrying

plasmid pET21a[SIC1$^{\Delta 214}$] as previously described (21). The purified fragment was employed for secondary structure analysis by FT-IR spectroscopy in the Amide-I region (1600 to 1700 cm^{-1}). Data were acquired in D_2O, because random coil and α-helix contributions largely overlap in H_2O (44). The Sic1$^{\Delta 214}$ absorption spectrum measured at 10°C is reported in Fig. 1 *A*. To resolve the Amide-I band constituents, FSD and second derivative were calculated from the raw spectrum. The results are reported in Fig. 1 *A*, together with the Amide-I curve fitting into Gaussian components. The second derivative spectrum is dominated by the ~1640 cm^{-1} band, which can be assigned to random-coil structure (44). In addition, other components were identified and their relative intensities quantified by Gaussian fitting from the ratio of their band area over the total Amide I area (36,44). They were found to be ~15% turns (1676 and 1670 cm^{-1}), ~30% α- and 3_{10}-helix (1656 cm^{-1}), ~40% random coil (1640 cm^{-1}), and ~15% β-strands (1627 cm^{-1}).

Comparison of Sic1$^{\Delta 214}$ second derivative spectrum with those of CRL1 (a natively folded α/β protein) and α-synuclein (a well-known IDP) highlights the disordered nature of Sic1$^{\Delta 214}$ (Supporting Material, Fig. S1). Indeed, its second derivative FT-IR spectrum is very similar to that of α-synuclein. Moreover, we should note that a protein containing ~30% of helical structure, such as CRL1, usually displays a strong and well-resolved peak in the second derivative spectrum (36). In the case of Sic1$^{\Delta 214}$, however, the contribution at 1656 cm^{-1} appears only as a shoulder of the main 1640 cm^{-1} peak, as expected for a component having a large band width (45). Indeed, in this case, the α-helical component is quite broad, with a band width comparable to that of the random-coil component. This result can be due to the presence of dynamic α-helical structures and possibly also to 3_{10} helices. It can be concluded that this protein fragment has a disordered nature, with a predominance of random coil and a certain intrinsic propensity to form ordered secondary structure elements.

FT-IR measurements were also performed at variable temperatures (Fig. 1, *B* and *C*). By heating the sample up to 100°C, the random-coil band around 1640 cm^{-1} in the second derivative spectra shifts to 1645 cm^{-1}, whereas the intensity of the 1656 cm^{-1} α-helical component seems to increase. As indicated by the profiles reported in Fig. 1 *C*, these effects are small but reproducible. Furthermore, the thermal transition is reversible. The results suggest that a minor structural rearrangement is induced by heat, in agreement with previous data on the full-length protein (7).

Hydrodynamic behavior by gel filtration

The hydrodynamic behavior of the Sic1$^{\Delta 214}$ was characterized by gel filtration (Fig. 2). When standard globular proteins are used for column calibration, the elution time of Sic1$^{\Delta 214}$ yields an apparent molecular mass of 18,381

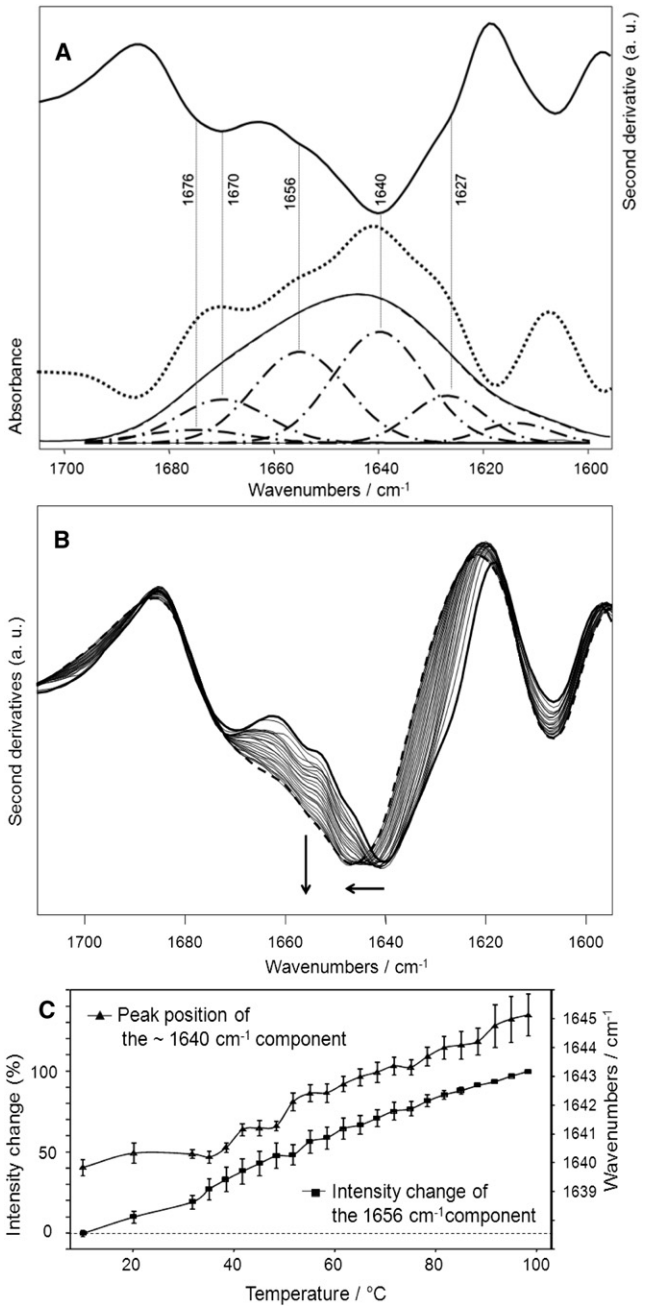

FIGURE 1 Sic1$^{\Delta 214}$ secondary structure by FT-IR spectroscopy. (*A*) Amide-I absorption spectrum (*continuous line*) and curve-fitted spectrum (*dashed line*) of Sic1$^{\Delta 214}$ in D_2O solution. The curve fitting of the spectrum as a linear combination of Gaussian components (*dot-dashed lines*) is superimposed to the measured spectrum. The Fourier self-deconvoluted (*dotted line*) and the second derivative (*continuous bold line*) spectra are also reported. (*B*) Amide-I second derivative spectra of Sic1$^{\Delta 214}$ in D_2O solution at different temperatures from 10°C (*bold continuous line*) to 100°C (*dashed line*). Arrows indicate the trend of spectral changes at increasing temperature. (*C*) Peak position of the ~1640 cm^{-1} component and intensity change of the 1656 cm^{-1} component as a function of temperature, both taken from the second derivative spectra.

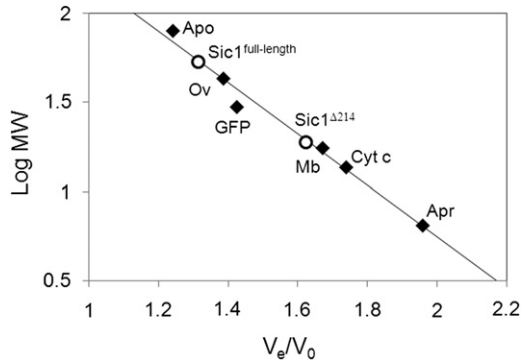

FIGURE 2 Hydrodynamic behavior of Sic1$^{\Delta 214}$ and Sic1^{FL} compared to standard globular proteins. Analytical gel filtration of ~100 μM protein in 50 mM sodium phosphate, pH 8.0, 200 mM NaCl. V_e, elution volume; V_o, void volume; Apo, apoferritin; Ov, ovalbumin; GFP, green fluorescent protein; Mb, myoglobin; Cyt c, cytochrome *c*; Apr, aprotinin. The reported values are the average from three independent experiments. Standard deviation was within 6.5%.

Da, well above the real value (9,293 Da, see below) (21). This result is consistent with a disordered nature of this protein. Based on the known R_h of the standards, the gel filtration data also allow estimating of the R_h value of Sic1$^{\Delta 214}$ (40). The results indicate a value of 20.57 ($\pm$ 0.22) Å. The same procedure has been applied, for comparison, to Sic1^{FL} yielding a value of 31.25 ($\pm$ 3.53) Å (Fig. 2, Table 1). That these values actually refer to the intact products and that no proteolytic degradation had occurred during the analysis has been checked by SDS-electrophoresis of the eluted material.

It would be interesting to compare the degree of compaction of Sic1$^{\Delta 214}$ and Sic1^{FL} with reference to globular and denatured proteins. Relevant to this point, the R_h values of folded, globular proteins and those of chemically denatured proteins display a well-characterized dependence on the number of residues, *N*, and can be fit by simple power-law equations (34). Thus, we define the compaction index (*CI*):

$$CI = \frac{R_h^D - R_h}{R_h^D - R_h^F},$$

where R_h is the experimental value for a given protein and R_h^D and R_h^F are the reference values calculated for a chemically denatured or a globular, folded protein of the same size. Different from the frequently used parameter R_h/R_h^D, *CI* does not depend on *N* and, therefore, it allows for comparison among proteins of different lengths. The *CI* value can vary between 0 and 1, with 0 indicating minimal compaction and 1 maximal compaction. This index could be applied to any structural feature linked to compactness (also the gyration radius, for instance), given that the reference curves for folded and unfolded proteins used in the calculation are built to fit the same parameter. By using the reference equations for R_h published by Forman-Kay and co-workers (34), we obtain *CI* values of 0.57 ($\pm$ 0.03) for Sic1$^{\Delta 214}$ and 0.77 ($\pm$ 0.13) for Sic1^{FL}. Slightly different values are obtained using the equations published by Uversky (40), namely, 0.63 ($\pm$ 0.02) for Sic1$^{\Delta 214}$ and 0.80 ($\pm$ 0.13) for Sic1^{FL}. These results indicate that the full-length protein displays higher overall compaction than the fragment. Because the isolated N-terminal moiety is even more disordered than the here considered C-terminal fragment (22), it can be conceived that either long-range coupling between the N- and C- termini or indirect effects of chain length play a role in the compaction of the full-length protein. Nevertheless, gel filtration offers relatively low resolution and low sensitivity, and captures only average features of the molecular ensemble.

Tertiary structure by ESI-MS and ESI-IM-MS

Finer characterization of tertiary structural features of Sic1$^{\Delta 214}$ was achieved by ESI-MS. Protein CSDs obtained by ESI-MS are strongly affected by the global compactness of the protein structure at the moment of transfer from the liquid to gas phase (28,46). It has been shown that the ionization behavior can effectively distinguish not only folded and unfolded protein, but also partially folded forms from both the above (27–29). Furthermore, this kind of analysis offers the typical advantage of MS methods, namely to describe heterogeneous samples by direct detection of the distinct components (29). Thanks to these features, ESI-MS is particularly well suited to the identification of labile compact forms that might arise transiently within a molecular population. The spectra of Sic1$^{\Delta 214}$ under nondenaturing conditions are characterized by sharply bimodal distributions, with a predominant form centered on the 10+ ion and a minor component centered on the 7+ ion (Fig. 3 *A*). Such a distribution is typical for coexistence of two distinct conformers of different compactness that do not interconvert on the time scale of the proton-transfer reactions that label protein molecules during the ESI process. The peak envelope at lower charge states (7+) would reflect the more compact form. Bimodal distributions have been observed on other Sic1-derived fragments (22). The relative contributions of the two components to the CSDs can be quantified by Gaussian fitting, upon transformation of the spectra to $x = z$ abscissa axis (22). The results of such a procedure are reported in the insets of Fig. 3. Although quantitation of different conformers by ESI-MS can only be taken indicatively (47), the apparent relative amount of the compact form in 50 mM ammonium acetate, pH 6.5 is quite small (~20%). Mass deconvolution yields a value of 9293.5 Da (+/− 0.1 Da), consistent with the sequence of the fragment with the His_6 tag and without initial methionine (calculated mass 9293.4 Da).

To confirm that the two peak envelopes correspond to distinct conformational states, the effects of acids and organic solvents have been investigated (Fig. 3 *B* and *C*). By the addition of 1% formic acid, the more compact

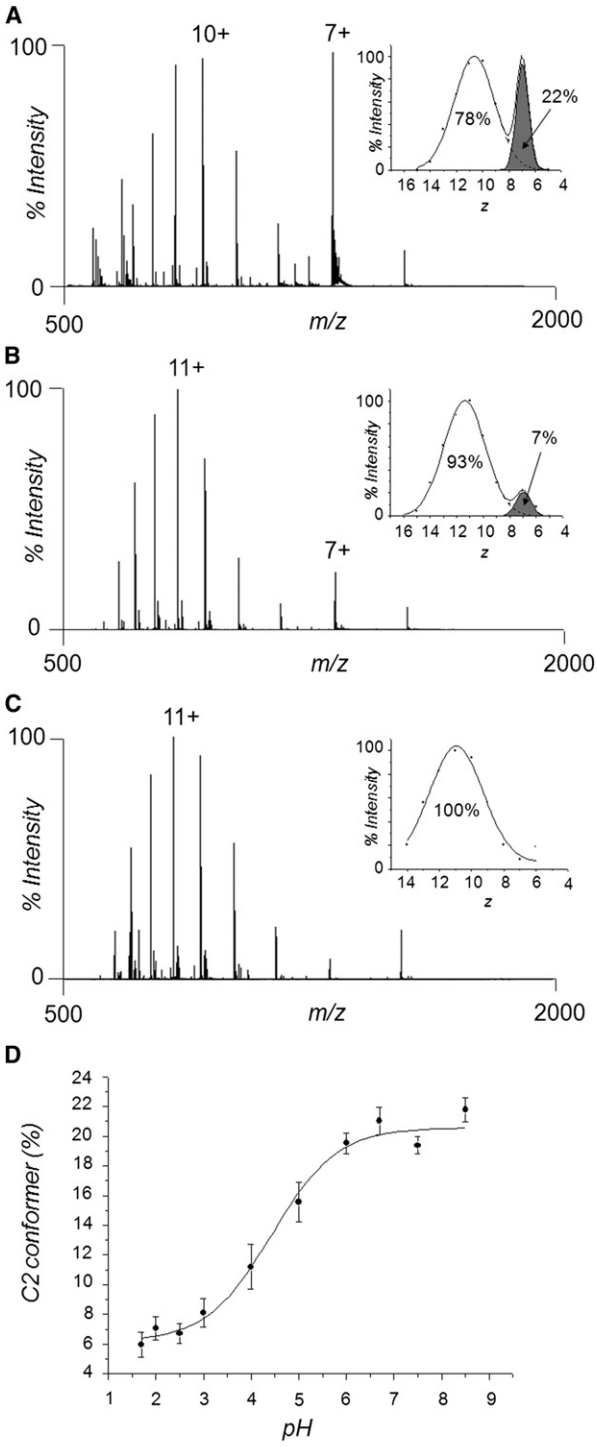

FIGURE 3 Sic1$^{\Delta 214}$ tertiary structure and unfolding. Nano-ESI-MS spectra of 10 μM protein in (*A*) 50 mM ammonium acetate, pH 6.5; (*B*) 50 mM ammonium acetate, 1% formic acid (pH ~3); and (*C*) 50 mM ammonium acetate, 1% formic acid, 50% acetonitrile. The main peak of each component is labeled by the corresponding charge state. The results of Gaussian fitting of the CSDs are shown in the insets. (*D*) Profile of the pH transition, showing the relative amount of the collapsed species C2 over the total population as a function of pH. The line represents data fitting by a sigmoidal curve. Error bars indicate the standard deviation from three independent acquisitions. (*E*) Amino acid sequence of Sic1$^{\Delta 214}$.

form decreases and the highly charged component accumulates. By the further addition of 50% acetonitrile, the compact form disappears completely. These results suggest that the component centered on the 7+ ion has considerable tertiary structure that is destabilized by acids and organic solvents. The pH dependence reveals a quite cooperative transition (Fig. 3 *D*). This response to pH might reflect the transition from disordered globule to swollen coil at increasing $f_+ - f_-$ in the phase diagram proposed by Pappu and co-workers (33). The pH response of this protein fragment is consistent with its amino acid sequence, counting 15 acidic residues (Fig. 3 *E*). Although the His_6 tag does not change significantly the pI (9.13 with tag and 9.11 without tag) its presence might affect the shape of the transition. These effects of denaturing agents on the CSD provide important control evidence that the low-charge component in Sic1$^{\Delta 214}$ spectra is not artifactual. The high-charge component, instead, responds to denaturing agents only by a minor shift in the main charge state from 10+ to 11+. Therefore, the 11+ form can be identified as the most unfolded species, whereas the 10+ component is characterized by modest intrinsic compactness. We name the 10+ and 7+ components, respectively, C1 and C2, referring to collapsed species of different compactness.

Altogether, this evidence strongly suggests that the component centered on the 7+ ion reflects a metastable compact state of this molecule in solution. No significant changes were observed by varying the buffer concentration between 10 and 50 mM. It should be underscored that a net charge 7+ reflects a surprisingly high compactness for this protein fragment. Indeed, it closely matches the average charge (7.5) calculated for a globular, folded protein of this size, according to the empiric charge-to-mass relation in protein ESI-MS (48). The effect of pH has been investigated also by FT-IR spectroscopy. The second derivative spectrum at pH 1.5 displays a small reduction of the shoulder around 1627 cm^{-1}, which characterizes the spectrum at pH 6.5 (Fig. S2). This difference suggests that also minor rearrangements in protein secondary structure take place at acidic pH. The region influenced by these changes is that of β-strands.

The existence of compact conformations of the here considered proteins was further tested by ESI-IM-MS measurements (Figs. 4 and 5). This method adds a second dimension to conventional ion sorting by ESI-MS, based on their mobility through a buffer gas under the influence of an electric field (30–32). Under these conditions, IM is directly related to the shape and overall topology of the ion, leading to separation of the ions on the basis of their structural compactness. Although protein charging during electrospray is affected by the structural properties of the

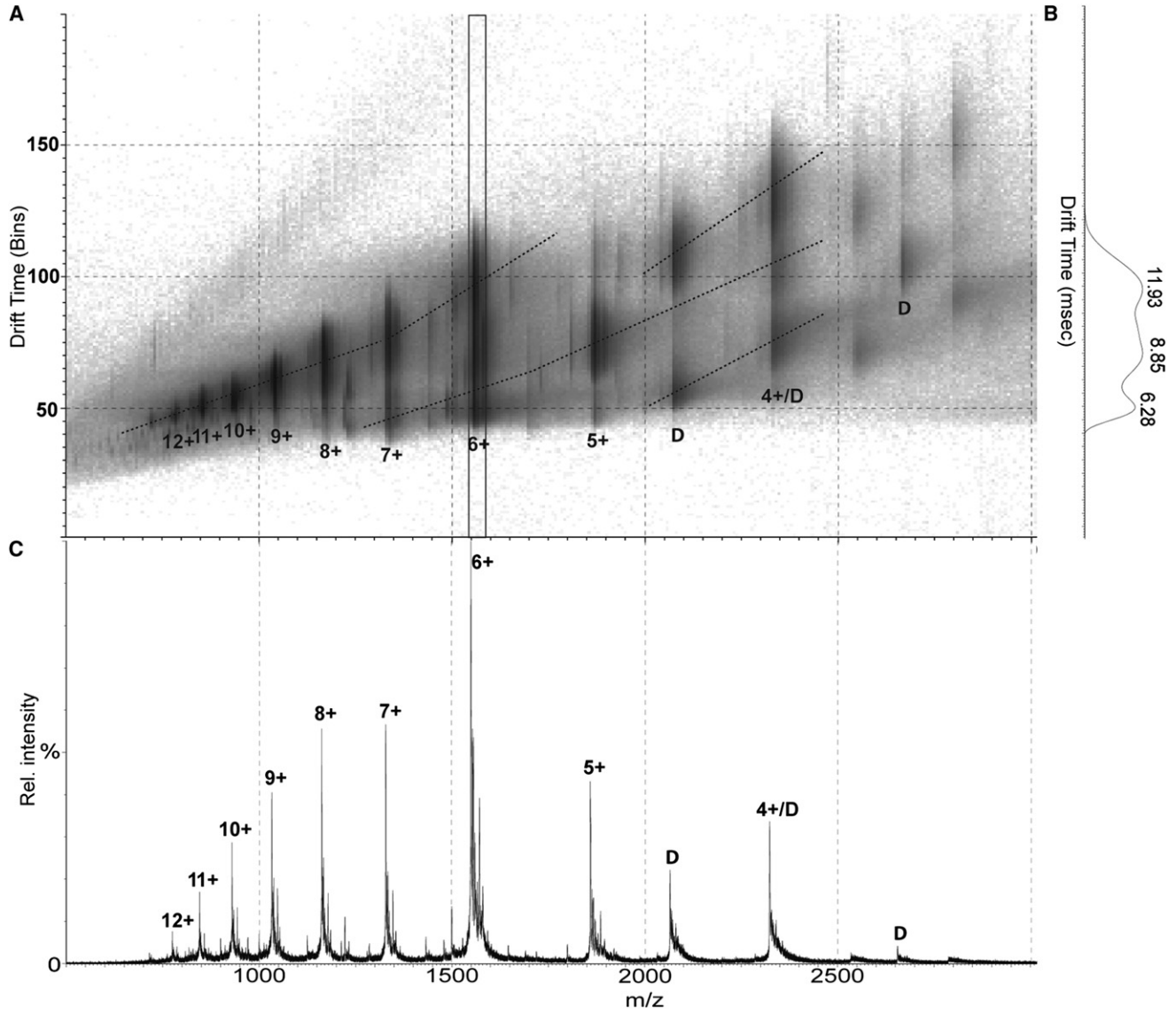

FIGURE 4 Analysis of Sic1$^{\Delta 214}$ distinct conformers in the gas phase with nondenaturing nano-ESI-IMS-MS. (*A*) IM-plot reporting the drift time measurements for each charge state. The dotted lines connect corresponding conformations that differ in IM due to their different charges. (*B*) Arrival time distribution for the 6+ charge state. (*C*) Corresponding nano-ESI-MS spectrum. The infused sample was 5 μM protein in 50 mM ammonium acetate pH 6.5.

protein in solution, IM is, obviously, affected by the actual conformation held by the protein in the gas phase. Folded or partially folded conformations can survive upon desolvation for low charge states. Highly charged ions, instead, tend to open into more extended conformations due to electrostatic repulsions, especially when they become activated due to collisions with background gas during transit of the mass spectrometer. The drift time plots obtained with Sic1$^{\Delta 214}$ (Fig. 4), under conditions that minimize thermal (collisional) activation of the ions, reveal two well-resolved populations. Because IM depends on the charge of the ions as well, particles with the same collisional cross sections come to lie on oblique lines (indicated in the figure). The species with the lowest mobility is enriched at high charge states and represents the least compact conformer. A second species with higher mobility is detectable and is, therefore, characterized by higher compactness. This species is maximally enriched at the charge states 5+ and 6+. The arrival time distribution for charge state 6+ is reported in Fig. 4 *B*. These results provide direct evidence of the existence of compact conformations of Sic1$^{\Delta 214}$, clearly distinguishable from the extended form of the protein. Therefore, IM analysis supports the hypothesis that the observed bimodal CSD is due to coexistence of at least two distinct conformers.

The spectrum reported in Fig. 4 *C* also reveals a small fraction of dimer-specific peaks at high *m/z* values, suggesting that Sic1$^{\Delta 214}$ might have an intrinsic propensity to dimerize. Dimers can be identified in the high-*m/z* range

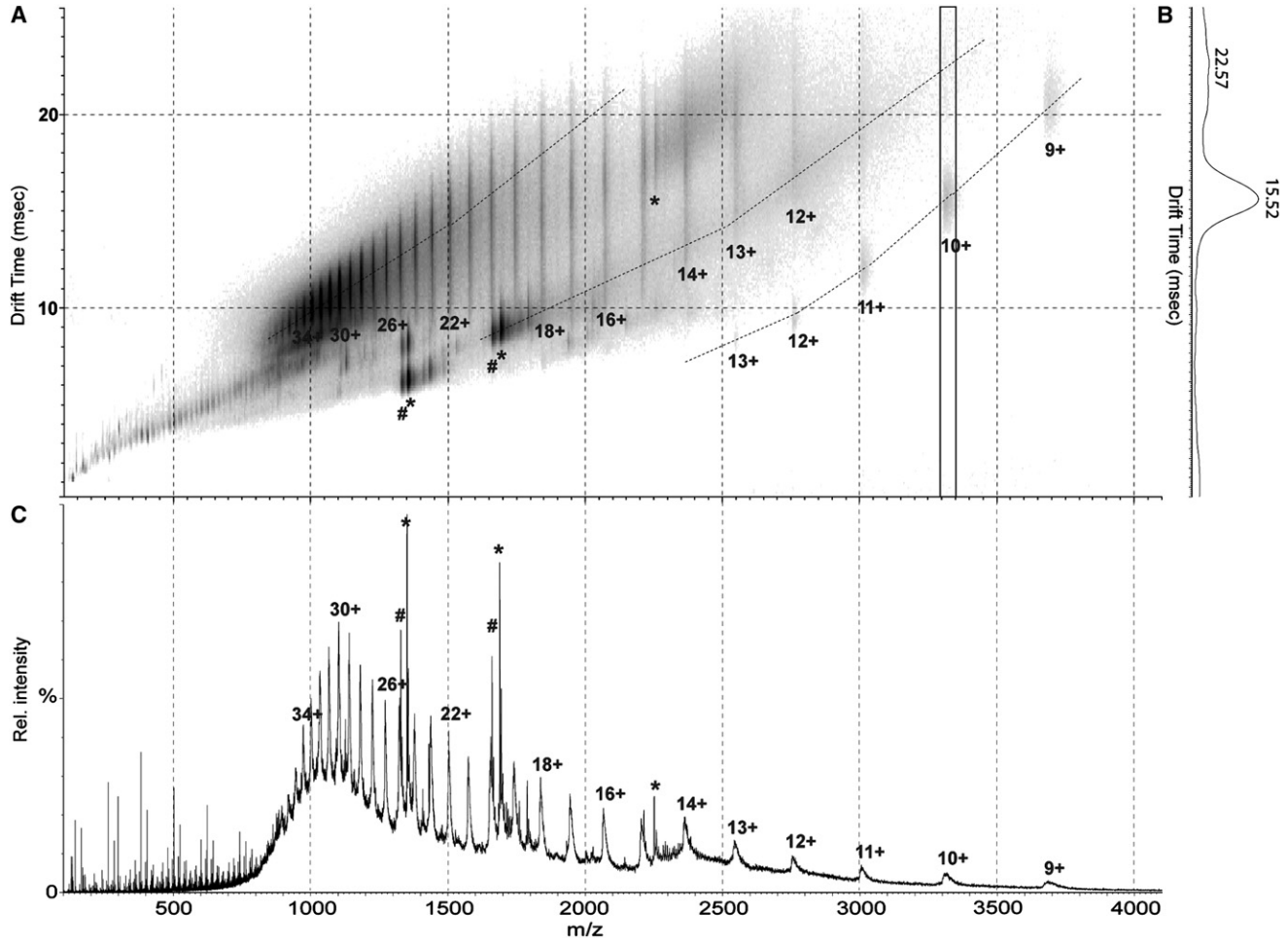

FIGURE 5 Analysis of Sic1^{FL} distinct conformers in the gas phase with nondenaturing nano-ESI-IMS-MS. (*A*) IM-plot reporting the drift time measurements for each charge state. The dotted lines connect corresponding conformations that differ in IM due to their different charges. (*B*) Arrival time distribution for the 10+ charge state. (*C*) Corresponding nano-ESI-MS spectrum. The infused sample was 5 μM protein in 50 mM ammonium acetate pH 6.5. The signals of proteolytic fragments are labeled with * and # in panels A and C.

of the IM-plot (Fig. 4 *A*). However, no dimers are detected for the same sample under the other instrumental setup employed for this work (Fig. 3). Therefore, Sic1 dimers must be highly unstable and prone to dissociate under electrospray conditions. More detailed studies will be needed to evaluate the structural and functional relevance of this dimeric species. It cannot be ruled out that they simply derive from low-level nonspecific association. The dimeric assemblies also reflect the existence of two conformational states of the protein subunits.

The same analysis has been carried out on the intact protein, Sic1^{FL} (Fig. 5). In this case, the ESI-IM-MS 2D plots reveal three distinct populations (Fig. 5 *A*). The predominant one displays the lowest mobility and corresponds, therefore, to the most extended conformation. As expected, this species is enriched at the highest charge states (34+ to 20+). A second population with intermediate mobility is detectable for the charge states between 20+ and 12+. Finally, a well-resolved series of peaks corresponding to the highest mobility is resolved in the low-charge region of the plot. This species is detectable for the charge states between 13+ and 9+ and is highly enriched for the 10+ ions. The arrival time distribution for this charge state is reported in Fig. 5 *B*. These data indicate that, under nondenaturing conditions, collapsed states can be identified also for the intact protein. The analysis of Sic1^{FL} CSD (Fig. 5 *C*) indicates that the molecular ensemble is more homogeneous than observed for Sic1$^{\Delta 214}$. Indeed the CSD is unimodal, rather than bimodal, with only some tailing to the right side. Thus, the large majority of Sic1^{FL} molecules are found in a compact state. That the component centered around the 30+ ion corresponds to a partially folded species in solution is also supported by previous denaturation studies (7). Although the presence of histidine tags has been shown to slightly affect compactness of IDPs (34), the differences emerging from this comparison should be ascribed to the different primary structure of the two proteins, because both products contain a His_6 tag.

Fig. S3 summarizes the results reported here, comparing calculated and experimental R_h values for both $Sic1^{\Delta 214}$ and $Sic1^{FL}$. The curve describing the dependence of R_h on chain length for IDPs in general (34) predicts the highest values. Slightly smaller values are calculated by the more accurate model that uses sequence-specific information (34). The average R_h measured by gel filtration for $Sic1^{\Delta 214}$ is very close to the calculated value. We attribute this R_h value to the highly predominant C1 component. Therefore, the C1 conformer of $Sic1^{\Delta 214}$ approximates the average compactness of IDPs for this chain length. The C2 conformer is not included in this plot because it could not be resolved by gel filtration, but would lie well below C1. The average experimental value for $Sic1^{FL}$ is much below both theoretical values calculated for the intact protein. This result is consistent with MS data, indicating that most of the $Sic1^{FL}$ protein is found in a highly collapsed state. Altogether, these results indicate that both $Sic1^{FL}$ and $Sic1^{\Delta 214}$ have a propensity to populate conformational states with higher compactness than predicted by the average behavior of IDPs.

CONCLUSION

Rapidly growing evidence indicates that IDPs in their unbound state (7–12,33,34) not only are far from resembling a true random coil, but are also quite different from chemically denatured proteins. However, the degree of compaction for IDPs is more variable than observed for folded, globular proteins (34). Sic1 is an example of IDP populating collapsed states of particularly high compactness. Nevertheless, compact states are destabilized in the isolated C-terminal fragment, relative to the native protein. Indeed, only a minor fraction of the molecular population is found in a compact state for $Sic1^{\Delta 214}$, consistently with the quite low *CI* value measured by gel filtration. The results reported here suggest that chain length could play an important role in stabilizing compact states of poorly foldable sequences.

The picture that emerges for the isolated Sic1 KID fragment is that of a highly dynamic structure, in which transient secondary and tertiary structure can develop. The detailed NMR studies performed on the related proteins p21 and p27 have shown that IFSUs are on-pathway for complex formation with the binding partners cyclin and kinase (4). However, highly compact tertiary structure might not be productive for complex formation, because the inhibitor seems to bind to its interactors in an extended conformation (23,24). More studies will be needed to understand whether the compact states detected here are on-pathway intermediates in the formation of Sic1 physiological complexes. Furthermore, it still remains to be clarified which interactions are responsible for chain compaction and which residues they involve.

The highly dynamic and heterogeneous molecular ensemble representing unfolded proteins poses a major challenge to structural characterization by biophysical methods. Short-lived structured states can, indeed, easily escape detection. Nevertheless, compact states of IDPs have been detected by several different biophysical techniques, such as small-angle x-ray scattering (10,49), atomic-force microscopy (50), Förster resonance energy transfer (51), ESI-MS (13), ESI-IM-MS (52), NMR studies on translational diffusion coefficients (10), and paramagnetic relaxation enhancement (10,12). The ion-sorting process inherent to MS measurements offers the possibility to directly assess species distributions in complex mixtures, allowing detection of metastable conformational states. Combining this information with structural data derived from IM measurements can offer an in-depth description of highly dynamic molecular systems.

SUPPORTING MATERIAL

Three figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(11)00308-0.

This work was supported by the grants FIRB/Italbionet to L.A., FAR ("Fondo Ateneo per la Ricerca") of the University Milano-Bicocca to R.G., M.L., S.B., and S.M.D., and by postdoctoral fellowships of the University of Milano-Bicocca to A.N. and M.Š.

REFERENCES

1. Alfieri, R., M. Barberis, …, L. Alberghina. 2009. Towards a systems biology approach to mammalian cell cycle: modeling the entrance into S phase of quiescent fibroblasts after serum stimulation. *BMC Bioinformatics.* 10 (*Suppl 12*):S16.
2. Fink, A. L. 2005. Natively unfolded proteins. *Curr. Opin. Struct. Biol.* 15:35–41.
3. Turoverov, K. K., I. M. Kuznetsova, and V. N. Uversky. 2010. The protein kingdom extended: ordered and intrinsically disordered proteins, their folding, supramolecular complex formation, and aggregation. *Prog. Biophys. Mol. Biol.* 102:73–84.
4. Galea, C. A., Y. Wang, …, R. W. Kriwacki. 2008. Regulation of cell division by intrinsically unstructured proteins: intrinsic flexibility, modularity, and signaling conduits. *Biochemistry.* 47:7598–7609.
5. Schnell, S., S. Fortunato, and S. Roy. 2007. Is the intrinsic disorder of proteins the cause of the scale-free architecture of protein-protein interaction networks? *Proteomics.* 7:961–964.
6. Eliezer, D. 2009. Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 19:23–30.
7. Brocca, S., M. Samalíková, …, R. Grandori. 2009. Order propensity of an intrinsically disordered protein, the cyclin-dependent-kinase inhibitor Sic1. *Proteins.* 76:731–746.
8. Galea, C. A., A. Nourse, …, R. W. Kriwacki. 2008. Role of intrinsic flexibility in signal transduction mediated by the cell cycle regulator, p27 Kip1. *J. Mol. Biol.* 376:827–838.
9. Tran, H. T., A. Mao, and R. V. Pappu. 2008. Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *J. Am. Chem. Soc.* 130:7380–7392.
10. Mittag, T., J. Marsh, …, J. D. Forman-Kay. 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.
11. Ganguly, D., and J. Chen. 2009. Atomistic details of the disordered states of KID and pKID. Implications in coupled binding and folding. *J. Am. Chem. Soc.* 131:5214–5223.

12. Salmon, L., G. Nodet, …, M. Blackledge. 2010. NMR characterization of long-range order in intrinsically disordered proteins. *J. Am. Chem. Soc.* 132:8407–8418.

13. Natalello, A., F. Benetti, …, R. Grandori. 2011. Compact conformations of α-synuclein induced by alcohols and copper. *Proteins.* 79:611–621.

14. Deshaies, R. J. 1997. Phosphorylation and proteolysis: partners in the regulation of cell division in budding yeast. *Curr. Opin. Genet. Dev.* 7:7–16.

15. Deshaies, R. J., and J. E. Ferrell, Jr. 2001. Multisite phosphorylation and the countdown to S phase. *Cell.* 107:819–822.

16. López-Avilés, S., O. Kapuy, …, F. Uhlmann. 2009. Irreversibility of mitotic exit is the consequence of systems-level feedback. *Nature.* 459:592–595.

17. Coccetti, P., R. L. Rossi, …, L. Alberghina. 2004. Mutations of the CK2 phosphorylation site of Sic1 affect cell size and S-Cdk kinase activity in Saccharomyces cerevisiae. *Mol. Microbiol.* 51:447–460.

18. Barberis, M., E. Klipp, …, L. Alberghina. 2007. Cell size at S phase initiation: an emergent property of the G1/S network. *PLOS Comput. Biol.* 3:e64.

19. Yaakov, G., A. Duch, …, F. Posas. 2009. The stress-activated protein kinase Hog1 mediates S phase delay in response to osmostress. *Mol. Biol. Cell.* 20:3572–3582.

20. Hodge, A., and M. Mendenhall. 1999. The cyclin-dependent kinase inhibitory domain of the yeast Sic1 protein is contained within the C-terminal 70 amino acids. *Mol. Gen. Genet.* 262:55–64.

21. Brocca, S., L. Testa, …, M. Lotti. 2011. Defining structural domains of an intrinsically disordered protein: Sic1, the cyclin-dependent kinase inhibitor of *Saccharomyces cerevisiae*. *Mol. Biotechnol.* 47:34–42.

22. Testa, L., S. Brocca, …, R. Grandori. 2011. Electrospray ionization-mass spectrometry conformational analysis of isolated domains of an intrinsically disordered protein. *Biotechnol. J.* 6:96–100.

23. Barberis, M., L. De Gioia, …, L. Alberghina. 2005. The yeast cyclin-dependent kinase inhibitor Sic1 and mammalian p27Kip1 are functional homologues with a structurally conserved inhibitory domain. *Biochem. J.* 387:639–647.

24. Russo, A. A., P. D. Jeffrey, …, N. P. Pavletich. 1996. Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. *Nature.* 382:325–331.

25. Sivakolundu, S. G., D. Bashford, and R. W. Kriwacki. 2005. Disordered p27Kip1 exhibits intrinsic structure resembling the Cdk2/cyclin A-bound conformation. *J. Mol. Biol.* 353:1118–1128.

26. Bowman, P., C. A. Galea, …, R. W. Kriwacki. 2006. Thermodynamic characterization of interactions between p27(Kip1) and activated and non-activated Cdk2: intrinsically unstructured proteins as thermodynamic tethers. *Biochim. Biophys. Acta.* 1764:182–189.

27. Šamalikova, M., C. Santambrogio, and R. Grandori. 2010. Mass-spectometry tools for the investigation of structural disorder and conformational transitions in proteins. *In* Instrumental Analysis of Intrinsically Disordered Proteins. V. N. Uversky and S. Longhi, editors. John Wiley & Sons, Hoboken, NJ. 629–652.

28. Kaltashov, I. A., and R. R. Abzalimov. 2008. Do ionic charges in ESI MS provide useful information on macromolecular structure? *J. Am. Soc. Mass Spectrom.* 19:1239–1246.

29. Borysik, A. J., S. E. Radford, and A. E. Ashcroft. 2004. Co-populated conformational ensembles of beta2-microglobulin uncovered quantitatively by electrospray ionization mass spectrometry. *J. Biol. Chem.* 279:27069–27077.

30. Smith, D. P., K. Giles, …, A. E. Ashcroft. 2007. Monitoring copopulated conformational states during protein folding events using electrospray ionization-ion mobility spectrometry-mass spectrometry. *J. Am. Soc. Mass Spectrom.* 18:2180–2190.

31. Smith, D. P., T. W. Knapman, …, A. E. Ashcroft. 2009. Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies. *Eur. J. Mass Spectrom. (Chichester, Eng.).* 15:113–130.

32. Uetrecht, C., R. J. Rose, …, A. J. Heck. 2010. Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* 39:1633–1655.

33. Mao, A. H., S. L. Crick, …, R. V. Pappu. 2010. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. USA.* 107:8183–8188.

34. Marsh, J. A., and J. D. Forman-Kay. 2010. Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.* 98:2383–2390.

35. Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105–132.

36. Natalello, A., D. Ami, …, S. M. Doglia. 2005. Secondary structure, conformational stability and glycosylation of a recombinant Candida rugosa lipase studied by Fourier-transform infrared spectroscopy. *Biochem. J.* 385:511–517.

37. Vila, R., I. Ponte, …, P. Suau. 2001. Induction of secondary structure in a COOH-terminal peptide of histone H1 by interaction with the DNA: an infrared spectroscopy study. *J. Biol. Chem.* 276:30898–30903.

38. Irvine, G. B. 2001. Determination of molecular size by size-exclusion chromatography (gel filtration). *Curr. Protoc. Cell. Biol.*, 5.5.

39. Mohr, D., S. Frey, …, D. Görlich. 2009. Characterisation of the passive permeability barrier of nuclear pore complexes. *EMBO J.* 28:2541–2553.

40. Uversky, V. N. 1993. Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry.* 32:13288–13298.

41. Uversky, V. N. 2002. What does it mean to be natively unfolded? *Eur. J. Biochem.* 269:2–12.

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

43. Müller-Späth, S., A. Soranno, …, B. Schuler. 2010. From the Cover: Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. USA.* 107:14609–14614.

44. Arrondo, J. L. R., and F. M. Goñi. 1999. Structure and dynamics of membrane proteins as studied by infrared spectroscopy. *Prog. Biophys. Mol. Biol.* 72:367–405.

45. Susi, H., and D. M. Byler. 1986. Resolution-enhanced Fourier transform infrared spectroscopy of enzymes. *Methods Enzymol.* 130: 290–311.

46. Ashcroft, A. E. 2010. Mass spectrometry and the amyloid problem—how far can we go in the gas phase? *J. Am. Soc. Mass Spectrom.* 21:1087–1096.

47. Kuprowski, M. C., and L. Konermann. 2007. Signal response of coexisting protein conformers in electrospray mass spectrometry. *Anal. Chem.* 79:2499–2506.

48. Heck, A. J., and R. H. Van Den Heuvel. 2004. Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom. Rev.* 23:368–389.

49. Bernadó, P., E. Mylonas, …, D. I. Svergun. 2007. Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* 129:5656–5664.

50. Brucale, M., M. Sandal, …, B. Samorì. 2009. Pathogenic mutations shift the equilibria of alpha-synuclein single molecules towards structured conformers. *ChemBioChem.* 10:176–183.

51. Nettels, D., S. Müller-Späth, …, B. Schuler. 2009. Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proc. Natl. Acad. Sci. USA.* 106:20740–20745.

52. Bernstein, S. L., D. Liu, …, J. R. Winkler. 2004. Alpha-synuclein: stable compact and extended monomeric structures and pH dependence of dimer formation. *J. Am. Soc. Mass Spectrom.* 15:1435–1443.