FEBS Letters 586 (2012) 70–73

FEBS Letters

journal homepage: www.FEBSLetters.org

Length-dependent compaction of intrinsically disordered proteins

Vladimir N. Uversky [a,b,*], Carlo Santambrogio [c], Stefania Brocca [c], Rita Grandori [c,*]

[a] Department of Molecular Medicine, College of Medicine, University of South Florida, Tampa, FL 33612, USA
[b] Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia
[c] Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126 Milan, Italy

ARTICLE INFO

Article history:
Received 30 September 2011
Revised 8 November 2011
Accepted 20 November 2011
Available online 28 November 2011

Edited by Judit Ovádi

Keywords:
Compaction index
Hydrodynamic radius
Induced folding
Functional misfolding
Structural disorder
Sic1

ABSTRACT

This work investigates the effect of chain length on the degree of compaction of intrinsically disordered proteins (IDPs). The three main IDP types, native coil (NC), pre-molten globule (PMG) and molten globule (MG), are compared by means of a compaction index (CI) normalized for chain length. The results point out a strong variability of compactness as a function of chain length within each group, with larger proteins populating more compact states. While qualitative sequence features are responsible for the main differences among groups, chain length seems to have an unspecific effect modulating the extent of compaction within each group. The results are consistent with a cooperative character of the weak interactions responsible for chain collapse.

1. Introduction

Intrinsically disordered proteins (IDPs) are characterized by the lack of ordered 3D-structure, but can completely or partially fold upon binding to specific partners [1,2]. Molecular recognition by IDPs is interesting as an extreme example of conformational adaptation during binding, whose mechanism can help understand intermolecular interactions in general. Furthermore, these proteins perform key regulatory functions, such as cell-cycle and transcription regulation [1–5]. Structural disorder is crucial for functional peculiarities of IDPs, including binding promiscuity, binding plasticity (the ability to adjust to multiple, differently shaped interactors), fast association kinetics, and regulation by post-translational modifications [1–5]. In order to interpret molecular recognition by these proteins, it is essential to understand their conformational properties in the free state. Although recently developed NMR approaches provide rather accurate description of IDP conformational ensembles [6–8], low resolution techniques continue to play an important role in structural characterization of these highly dynamic systems.

Using combined information from size-exclusion chromatography and circular dichroism, the conformational states of IDPs have been classified as either native coil (NC), pre-molten globule (PMG) or molten globule (MG), according to increasing compaction [4,5,9]. The degree of IDP compaction has been shown to be affected, although to different extents, by several factors, such as net charge, hydrophobicity, proline content, histidine tag and secondary-structure propensity [10–12]. One of the most striking features of highly disordered proteins (NCs or PMGs) is abundance of charged residues and high net charge [13]. This feature seems to be crucial to maintain the polypeptide chain in an extended conformation [10], since sequences that are rich in uncharged, polar amino acids, although devoid of canonical hydrophobic residues, have been repeatedly shown to form heterogeneous ensembles of collapsed structures in aqueous solutions [10,14–17]. Based on the analysis of a set of highly charged polypeptides, it has been concluded that the net charge per residue can modulate the intrinsic preference of polypeptide backbones for collapsed structures [10]. Furthermore, extensive simulations by coarse-grained models, over a wide range of sequence hydrophobicity, net charge, and length, revealed that conformational properties of natively unfolded proteins can be described by a coil-to-globule transition in a charge/hydrophobicity sequence space [18]. In agreement with this hypothesis, it has been recently shown that nucleoporins (Nups) with low charge content possess more compact configurations, whereas highly charged Nups adopt more dynamic, extended, coil-like conformations [11]. This study intends to test

* Corresponding authors. Address: Department of Molecular Medicine, College of Medicine, University of South Florida, Tampa, FL 33612, USA. Fax: +1 813 9747357 (V.N. Uversky), Fax: +39 02 64483565 (R. Grandori).
E-mail addresses: vuversky@health.usf.edu (V.N. Uversky), rita.grandori@unimib.it (R. Grandori).

0014-5793/$36.00
doi:10.1016/j.febslet.2011.11.026

the hypothesis that the degree of an IDP compaction might also be affected by the protein chain length.

2. Methodological approach

One way to evaluate the degree of protein compaction is to measure the hydrodynamic radius, R_h, of a given protein in solution and to compare the measured value with the reference values of normally folded and completely unfolded proteins [4,5]. A frequently used parameter is R_h/R_h^c, where R_h^c is the R_h value calculated for a random coil polypeptide of the same length as the considered protein. However, for a more general comparison, one has to take into account the dependence of R_h on chain length. In fact, the equilibrium conformations of globular and denatured proteins are characterized by very different dependency of hydrodynamic properties on chain length [4,5,19,20]. The R_h values of folded and completely unfolded proteins vary with the number of residues, N, according to different power-laws [12]. As a result, these curves diverge significantly as N increases. Due to these features of the reference curves, R_h/R_h^c can only be used to compare the extent of compaction among proteins of similar size. For instance, natively folded, globular proteins, which all display similar compaction degree, will give rise to R_h/R_h^c values that vary from almost 1 to ~0.4 as N increases [12]. In order to allow for comparison among proteins of different sizes, it is necessary to normalize for the maximal possible variation of R_h at each value of N. Therefore, we use the previously defined compaction index, CI [21,22]:

$$\mathrm{CI} = (R_h^c - R_h)/(R_h^c - R_h^f) \quad (1)$$

where R_h^c and R_h^f are the reference values for fully unfolded (random coil) and folded, globular proteins of the same size as the examined protein. Its value ideally ranges from 0 for minimal compaction and 1 for maximal compaction. This normalization makes the definition of CI independent of chain length, enabling comparison of structural compaction among proteins of different sizes. This means that the calculation of the parameter is not biased by chain length, while obviously not implying any constraint on its variability with length. Whether or not its value will be constant for a given set of proteins, or how it will vary with N, depends on the function that will describe $R_h = f(N)$ in the specific case. Thus, CI is an adequate tool to explore possible effects of chain length on chain compactness. Its variations with N would reveal actual differences in chain compactness, free from the systematic bias that affects, instead, R_h/R_h^c.

The R_h values of IDPs can be derived from the known equations of R_h versus protein size for NCs, PMGs and MGs [4,5,9]. These equations were derived from the available viscometry, gel-filtration chromatography and dynamic light scattering data on the hydrodynamic dimensions of globular proteins in different conformational states and on the hydrodynamic behavior of 60 IDPs under conditions of neutral pH and physiological salt concentrations (100–150 mM). It is important to remember that charge interactions are known to modulate dimensions of IDPs [23]. Therefore, the hydrodynamic properties of IDPs are extremely sensitive to the pH and ionic strength of the solution. Furthermore, the hydrodynamic measurements were performed in a broad range of protein concentrations, to ensure that the derived hydrodynamic dimensions are not concentration dependent. Available data provide us with a large list of experimentally determined R_h values for proteins of different lengths, classified according to the conformational class and IDP type. The size of the considered proteins varies between 25 and 1000 residues. In this range falls the vast majority of the predicted IDPs, with mean length values of 457 for the *Mus musculus* proteome and 319 for *Escherichia coli* [24]. Since the original equations were expressed as R_h functions of protein mass, and since protein molecular mass, in turn, correlates with chain length, we reanalyzed the original data to obtain the chain length dependence of the R_h for NCs, PMGs and MGs, as well as for folded and chemically unfolded globular proteins:

$$\log R_h^{\mathrm{NC}} = (0.454 \pm 0.017) + (0.493 \pm 0.008) \times \log N \quad (2)$$

$$\log R_h^{\mathrm{PMG}} = (0.587 \pm 0.029) + (0.402 \pm 0.012) \times \log N \quad (3)$$

$$\log R_h^{\mathrm{MG}} = (0.629 \pm 0.051) + (0.334 \pm 0.021) \times \log N \quad (4)$$

$$\log R_h^{f} = (0.525 \pm 0.012) + (0.358 \pm 0.005) \times \log N \quad (5)$$

$$\log R_h^{c} = (0.385 \pm 0.017) + (0.543 \pm 0.007) \times \log N \quad (6)$$

These equations are simple fits of empiric data and are not based on a theoretical physical model. Furthermore, their validity is limited to the range spanned by the experimental points and there is no theoretical argument allowing any extrapolation outside such a range. The same is true for the reference curves themselves. For instance, those curves intersect at $N = 5.7$, leading to a meaningless behavior of the model for very small values of N. It should be noted to this regard that also simulations yield an ill-defined coil-to-globule transition for the case of a penta-peptide [25]. However, the present analysis focuses exclusively on the size range where experimental points are available. The empiric Eqs. (2)–(6) have been used here to compute CI. The values of R_h^c and R_h^f have been used as the references for fully unfolded and folded, globular proteins [4,5].

3. Results and discussion

The plot in Fig. 1 reports the CI profiles obtained for the three IDP classes. The results show the expected differences in compactness among the three groups. The average CI value calculated from the equations, for N varying from 25 to 1000, are (0.21 ± 0.08) for NCs, (0.50 ± 0.21) for PMGs, and (0.86 ± 0.15) for MGs. Most importantly, there is a clear effect of chain length on the extent

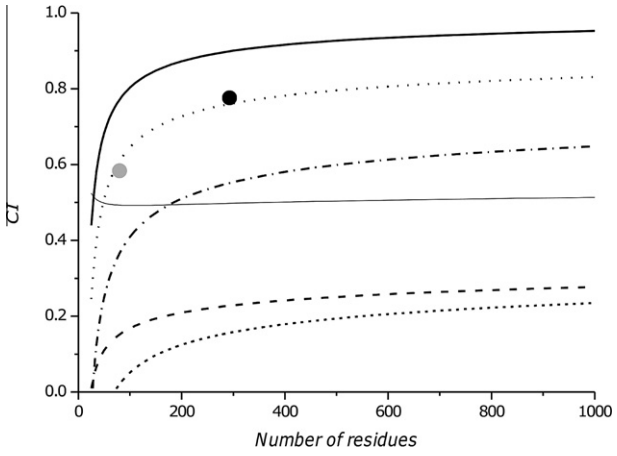

Fig. 1. CI dependence on chain length for NCs (long-dashed line), PMGs (dot-dashed line), and MGs (solid line) IDPs. The circles indicate the values for full-length Sic1 (black) and its C-terminal fragment (gray) [21]. The dotted line represents data fitting by a linear combination of the PMG and MG equations ($0.6 \times \mathrm{MG} + 0.4 \times \mathrm{PMG}$). The thin line shows the CI calculated for the generic equation $y = aN^b$ obtained as the best fit of the curve of isocompactness $y = 1/2R_h^c + 1/2R_h^f$. The short-dashed line shows the CI calculated from the R_h equation for a self-avoiding random coil obtained by computational simulations in an athermal solvent [26]. Data fitting and other calculations were done by the program Origin 7.0 (Originlab, Northampton, MA, USA). The R_h values of the proteins used in this work are listed in references [5,9].

of compaction within each group. The profiles derived from experimental data are compared in Fig. 1 with the profile based on computational simulations for unfolded proteins in an athermal solvent ([26] and personal communication). In such a model system, the protein interacts with itself just as well as with the solvent. This assumption is introduced in order to simulate random-coil polymers in solution, accounting for the excluded-volume effect [26]. The CI profile calculated from the simulation data lies between zero and the experimental profile for the NC class. The model system, too, displays some length-dependence compaction. Thus, simulations of random-flight polymer chains with excluded volume give results similar to what is seen with unfolded proteins, somewhat in between NCs under non-denaturing conditions and unfolded proteins in the presence of denaturants (CI = 0).

It is important to underscore that the IDP behavior reported in Fig. 1 is not merely a consequence of the equations and definitions employed in the present analysis. A first important observation to this regard is that constant CI values would be possible within the here considered model. The curves of isocompactness are defined by

$$(R_h^c - y)/(R_h^c - R_h^f) = k \tag{7}$$

implying

$$y = (1-k)R_h^c + kR_h^f, \tag{8}$$

which can be satisfied for any value of k between 0 and 1. Nonetheless, Eq. (8) also shows that the condition for CI = const. = k is that the function y, describing dependence of R_h on N, is a linear combination of power-law functions. A simple power-law function, like those generally used to fit R_h data versus N, would never yield CI = const. Therefore, it is important to examine how much CI could vary for a generic power-law function and, in particular, what is the minimal variation implied by such a model. To retrieve this information, the parameters of simple power-law equations $y = aN^b$ were optimized to approach at the best the condition CI = const. = k (Eq. (7)) for different values of k. The best fits provide examples of simple power-law functions that closely approximate curves of isocompactness. Fig. 1 reports the CI profile obtained for the power-law function with CI $\sim$ 0.5. The same results for other values of k are summarized in Fig. S1 of the Supplementary material. All the power-law functions identified by this procedure, indeed, give rise to very flat CI profiles. In general, infinite equations could be found that generate an almost constant CI, even within a simple power-law model describing R_h dependence on N. The variations in CI calculated for IDPs and reported in Fig. 1 are much more pronounced, particularly for N below 200. It is important to note that 36% of the proteins in UniProt (http://www.uniprot.org/) are shorter than 200 residues.

To further describe the landscape of possible CI profiles, a set of equations $y = aN^b$ was generated by random assignment of the parameters. These functions and the relative CI profiles are reported in Fig. S2 of the Supplementary material. As can be seen, CI profiles can be very diverse. For particular values of the parameters, CI can even markedly decrease as N increases, in the typical range of protein chain length (80–400 residues). Therefore, we can conclude that the data reported in Fig. 1 reveal a trend that would not be necessarily expected a priori on the basis of the here employed mathematical model. These data strongly support the hypothesis that protein size is another factor that can affect IDP compaction.

One possible interpretation of the observed effect is that weak, intramolecular interactions can be unspecifically amplified by an increase in protein size. CI increases rapidly with chain length for low molecular weight, while the increase is very slow and almost constant at high molecular weight. This effect is consistent with some degree of cooperativity of weak interactions. Thus, it seems that qualitative features of the amino acid sequence are responsible for the major differences among NCs, PMGs and MGs, whereas a unspecifically effect of chain length modulates the degree of compaction within each group.

The relative role of weak forces (hydrophobic interactions, hydrogen bonds, electrostatic interactions, and van der Waals contacts) mediating such length effect remains to be investigated. The modest effect on NCs, compared to PMG-like and MG-like IDPs, would be compatible with the involvement of hydrophobic residues. Indeed, highly extended IDPs are characterized by low overall hydrophobicity and high net charge [13]. In particular, NCs contain, on average, fewer hydrophobic residues than PMGs and MGs [4]. It is conceivable that, for such hydrophilic sequences, length increase might not be enough to promote noticeable compaction. However, further experimental and theoretical studies will be needed to elucidate the specific contribution of each type of interaction.

It should also be noted that many IDPs, while mostly disordered, have transient elements of preformed secondary structure, which are highly interaction-prone and are used by those IDPs for binding to specific partners [27]. The existence of such preformed binding sites enables faster and more effective interactions of IDPs with their targets [27]. In the unbound state, these elements can be involved in a set of intramolecular interactions, which are off-pathway for complex formation [28]. Based on these observations, a functional misfolding concept was formulated [29], according to which potential binding elements transiently formed in highly disordered IDPs interact with each other intramolecularly and are sequestered inside structural cages, preventing them from unnecessary and unwanted interactions with non-native binding partners. Obviously, such functional misfolding is accompanied by a partial compaction of the polypeptide chain. Since the probability of forming and docking such elements increases with protein length [27], the probability of functional misfolding increases as well. Such a mechanism could prevent that interaction propensity towards non-natural partners increases with IDP chain length.

Previous analysis of the yeast IDP Sic1 (a cyclin-dependent kinase inhibitor) and its fragments had first suggested a length effect on protein compaction [21]. This hypothesis was based on the observation that the isolated kinase inhibitory domain (KID) displays a smaller CI value than the full-length protein, although partial proteolysis and other order-propensity parameters suggest that this is the most ordered/compact region of the protein [30,31]. The CI values for Sic1 and its C-terminal fragment are reported as circles in Fig. 1. These points can be fit by a curve representing a linear combination of the MG and PMG equations ($0.6 \times$ MG + $0.4 \times$ PMG). In agreement with previous evidence [21,30,31], these results suggest that Sic1 is a relatively compact IDP. That the two experimental points lie on the same curve can be rationalized by the relative homogeneous amino acidic composition [30] and secondary structure propensity [32] of this protein. Since Sic1 binds to the cyclin-kinase complex in an extended conformation, compact forms are likely unproductive for complex formation. Hence, as previously suggested [21], loss of tertiary structure could be required for association, implying a binding-induced unfolding event in addition to the well known binding-induced folding. The engagement of part of the chain in interactions with partners will reduce the effective length sensed by the chain, reducing compaction. Such a mechanism could mediate binding-induced unfolding.

In conclusion, the results of this study suggest that chain length represents an important factor determining the compaction degree of IDPs, in addition to sequence features. Further experimental evidence is needed to clarify the relative role of weak interactions promoting IDP compaction. Intrinsic structural compactness could

be important to protect IDPs from non-physiological interactions and from protein degradation. Furthermore, it may also provide an additional level of regulation by post-translational modifications, such as phosphorylation.

Acknowledgments

We thank David Goldenberg for insightful discussions. This study was supported by the Grant FAR (Fondo Ateneo per la Ricerca) to R.G. and S.B. and by the National Science Foundation Grant 324 EF 0849803 and the Program of the Russian Academy of Sciences for the "Molecular and cellular biology" to V.N.U.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2011.11.026.

References

[1] Uversky, V.N. (2011) Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. Chem. Soc. Rev. 40, 1623–1634.
[2] Mittag, T., Orlicky, S., Choy, W.Y., Tang, X., Lin, H., Sicheri, F., Kay, L.E., Tyers, M. and Forman-Kay, J.D. (2008) Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. Proc. Natl. Acad. Sci. USA 105, 17772–17777.
[3] Dunker, A.K., Cortese, M.S., Romero, P., Iakoucheva, L.M. and Uversky, V.N. (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J. 272, 5129–5148.
[4] Uversky, V.N. (2002) Natively unfolded proteins: a point where biology waits for physics. Protein Sci. 11, 739–756.
[5] Uversky, V.N. (2002) What does it mean to be natively unfolded? Eur. J. Biochem. 269, 2–12.
[6] Jensen, M.R., Markwick, P.R., Meier, S., Griesinger, C., Zweckstetter, M., Grzesiek, S., Bernadó, P. and Blackledge, M. (2009) Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. Structure 17, 1169–1185.
[7] Jensen, M.R., Communie, G., Ribeiro, E.A.J., Martinez, N., Desfosses, A., Salmon, L., Mollica, L., Gabel, F., Jamin, M., Longhi, S., Ruigrok, R.W. and Blackledge, M. (2011) Intrinsic disorder in measles virus nucleocapsids. Proc. Natl. Acad. Sci. USA 108, 9839–9844.
[8] Schneider, R., Huang, J.R., Yao, M., Communie, G., Ozenne, V., Mollica, L., Salmon, L., Jensen, M.R. and Blackledge, M. (2011). Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy. Mol. Biosyst. doi: 10.1039/C1MB05291H.
[9] Tcherkasskaya, O., Davidson, E.A. and Uversky, V.N. (2003) Biophysical constraints for protein structure prediction. J. Proteome Res. 2, 37–42.
[10] Mao, A.H., Crick, S.L., Vitalis, A., Chicoine, C.L. and Pappu, R.V. (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. Proc. Natl. Acad. Sci. USA 107, 8183–8188.
[11] Yamada, J., Phillips, J.L., Patel, S., Goldfien, G., Calestagne-Morelli, A., Huang, H., Reza, R., Acheson, J., Krishnan, V.V., Newsam, S., Gopinathan, A., Lau, E.Y., Colvin, M.E., Uversky, V.N. and Rexach, M.F. (2010) A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. Mol. Cell. Proteomics 9, 2205–2224.
[12] Marsh, J.A. and Forman-Kay, J.D. (2010) Sequence determinants of compaction in intrinsically disordered proteins. Biophys. J. 98, 2383–2390.
[13] Uversky, V.N., Gillespie, J.R. and Fink, A.L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 41, 415–427.
[14] Dougan, L., Li, J., Badilla, C.L., Berne, B.J. and Fernandez, J.M. (2009) Single homopolypeptide chains collapse into mechanically rigid conformations. Proc. Natl. Acad. Sci. USA 106, 12605–12610.
[15] Walters, R.H. and Murphy, R.M. (2009) Examining polyglutamine peptide length: a connection between collapsed conformations and increased aggregation. J. Mol. Biol. 393, 978–992.
[16] Wang, X., Vitalis, A., Wyczalkowski, M.A. and Pappu, R.V. (2006) Characterizing the conformational ensemble of monomeric polyglutamine. Proteins 63, 297–311.
[17] Tran, H.T., Mao, A. and Pappu, R.V. (2008) Role of backbone–solvent interactions in determining conformational equilibria of intrinsically disordered proteins. J. Am. Chem. Soc. 130, 7380–7392.
[18] Ashbaugh, H.S. and Hatch, H.W. (2008) Natively unfolded protein stability as a coil-to-globule transition in charge/hydropathy space. J. Am. Chem. Soc. 130, 9536–9542.
[19] Tcherkasskaya, O. and Uversky, V.N. (2003) Polymeric aspects of protein folding: a brief overview. Protein Pept. Lett. 10, 239–245.
[20] Uversky, V.N. (2003) Protein folding revisited. A polypeptide chain at the folding-misfolding-non-folding cross-roads: which way to go? Cell. Mol. Life Sci. 60, 1852–1871.
[21] Brocca, S., Testa, L., Sobott, F., Šamalikova, M., Natalello, A., Papaleo, E., Lotti, M., De Gioia, L., Doglia, S.M., Alberghina, L. and Grandori, R. (2011) Compaction properties of an intrinsically disordered protein: sic1 and its kinase-inhibitor domain. Biophys. J. 100, 2243–2252.
[22] Wilkins, D.K., Grimshaw, S.B., Receveur, V., Dobson, C.M., Jones, J.A. and Smith, L.J. (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques 38, 16424–16431.
[23] Müller-Späth, S., Soranno, A., Hirschfeld, V., Hofmann, H., Rüegger, S., Reymond, L., Nettels, D. and Schuler, B. (2010) Charge interactions can dominate the dimensions of intrinsically disordered proteins. Proc. Natl. Acad. Sci. USA 107, 14609–14614.
[24] Oldfield, C.J., Cortese, M.S., Brown, C.J., Uversky, V.N. and Dunker, A.K. (2005) Comparing and combining predictors of mostly disordered proteins. Biochemistry 44, 1989–2000.
[25] Vitalis, A., Wang, X. and Pappu, R.V. (2008) Atomistic simulations of the effects of polyglutamine chain length and solvent quality on conformational equilibria and spontaneous homodimerization. J. Mol. Biol. 384, 279–297.
[26] Goldenberg, D.P. (2003) Computational simulation of the statistical properties of unfolded proteins. J. Mol. Biol. 326, 1615–1633.
[27] Oldfield, C.J., Cheng, Y., Cortese, M.S., Romero, P., Uversky, V.N. and Dunker, A.K. (2005) Coupled folding and binding with alpha-helix-forming molecular recognition elements. Biochemistry 44, 12454–12470.
[28] Marsh, J.A., Dancheck, B., Ragusa, M.J., Allaire, M., Forman-Kay, J.D. and Peti, W. (2010) Structural diversity in free and bound states of intrinsically disordered protein phosphatase 1 regulators. Structure 18, 1094–1103.
[29] Uversky, V.N. (2011) Intrinsically disordered proteins may escape unwanted interactions via functional misfolding. Biochim. Biophys. Acta 1814, 693–712.
[30] Brocca, S., Šamalikova, M., Uversky, V.N., Lotti, M., Vanoni, M., Alberghina, L. and Grandori, R. (2009) Order propensity of an intrinsically disordered protein, the cyclin-dependent-kinase inhibitor Sic1. Proteins 76, 731–746.
[31] Testa, L., Brocca, S., Šamalikova, M., Santambrogio, C., Alberghina, L. and Grandori, R. (2011) Electrospray ionization–mass spectrometry conformational analysis of isolated domains of an intrinsically disordered protein. Biotechnol. J. 6, 96–100.
[32] Brocca, S., Testa, L., Šamalikova, M., Grandori, R. and Lotti, M. (2011) Defining structural domains of an intrinsically disordered protein: Sic1, the cyclin-dependent kinase inhibitor of *Saccharomyces cerevisiae*. Mol. Biotechnol. 47, 34–42.