THE PROTEIN SOCIETY

Intrinsically disordered inhibitor of glutamine synthetase is a functional protein with random-coil-like pK_a values

Concetta Cozza,[1] José L. Neira,[2,3]* Francisco J. Florencio,[4] M. Isabel Muro-Pastor,[4] and Bruno Rizzuti[5]*

[1]Molecular Biophysics Laboratory, Department of Physics, University of Calabria, Rende, Italy
[2]Instituto de Biología Molecular y Celular, Universidad Miguel Hernández, Elche, Alicante, Spain
[3]Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Unidad Asociada IQFR-CSIC-BIFI, Universidad de Zaragoza, Zaragoza, Spain
[4]Instituto de Bioquímica Vegetal y Fotosíntesis, CSIC-Universidad de Sevilla, Seville, Spain
[5]CNR-NANOTEC, Licryl-UOS Cosenza and CEMIF.Cal, Department of Physics, University of Calabria, Rende, Italy

Received 16 January 2017; Accepted 10 March 2017
DOI: 10.1002/pro.3157
Published online 15 March 2017 proteinscience.org

Abstract: The sequential action of glutamine synthetase (GS) and glutamate synthase (GOGAT) in cyanobacteria allows the incorporation of ammonium into carbon skeletons. In the cyanobacterium *Synechocystis* sp. PCC 6803, the activity of GS is modulated by the interaction with proteins, which include a 65-residue-long intrinsically disordered protein (IDP), the inactivating factor IF7. This interaction is regulated by the presence of charged residues in both IF7 and GS. To understand how charged amino acids can affect the binding of an IDP with its target and to provide clues on electrostatic interactions in disordered states of proteins, we measured the pK_a values of all IF7 acidic groups (Glu32, Glu36, Glu38, Asp40, Asp58, and Ser65, the backbone C-terminus) at 100 m*M* NaCl concentration, by using NMR spectroscopy. We also obtained solution structures of IF7 through molecular dynamics simulation, validated them on the basis of previous experiments, and used them to obtain theoretical estimates of the pK_a values. Titration values for the two Asp and three Glu residues of IF7 were similar to those reported for random-coil models, suggesting the lack of electrostatic interactions around these residues. Furthermore, our results suggest the presence of helical structure at the N-terminus of the protein and of conformational changes at acidic pH values. The overall experimental and *in silico* findings suggest that local interactions and

Abbreviations: GOGAT, glutamate synthase; GS, glutamine synthetase; HSQC, heteronuclear single quantum correlation; IDP, intrinsically disordered protein; IF, inactivating factor; IF7, *Synechocystis* sp. PCC 6803 65-residue-long inactivating factor of glutamine synthetase; MD, molecular dynamics; NMR, nuclear magnetic resonance; R_G, radius of gyration; TSP, sodium trimethylsilyl [2,2,3,3-$^2H_4$] propionate.

Additional Supporting Information may be found in the online version of this article.

C.C. and J.L.N. contributed equally to this work.

Grant sponsor: Spanish Ministerio de Economia y Competitividad; Grant numbers: CTQ 2015-64445-R (to J.L.N.) and BFU2013-41712-P and BIO2016-75634P (to M.I.M.-P. and F.J.F.), with Fondo Social Europeo (ESF). Grant sponsor: Junta de Andalucía; Grant number: BIO-284 (to F.J.F). Grant sponsor: Generalitat Valenciana; Grant number: Prometeo 018/2013 (to J.L.N.).

Importance/Impact: Charge–charge interactions are fundamental to modulate protein function. They are believed to be especially important for intrinsically disordered proteins (IDPs) due to their usually high amount of titratable residues. We report that the 65-residue-long inactivating factor of glutamine synthetase (IF7) is a unique IDP that does not show differences in the pK_a values of its ionizable residues compared to those of amino acids in random-coil models.

*Correspondence to: José L. Neira, Instituto de Biología Molecular y Celular, Universidad Miguel Hernández, Avda. del Ferrocarril s/n, 03202 Elche, Alicante, Spain. E-mail: jlneira@umh.es or Bruno Rizzuti, CNR-NANOTEC, Licryl-UOS Cosenza and CEMIF.Cal, Department of Physics, University of Calabria, Ponte P. Bucci, 87036 Rende, Italy. E-mail: bruno.rizzuti@cnr.it

conformational equilibria do not play a role in determining the electrostatic features of the acidic residues of IF7.

Keywords: electrostatics; intrinsically disordered proteins; molecular dynamics; nuclear magnetic resonance; pK_a values; titration

Introduction

Ammonium assimilation takes place in most microorganisms by the sequential action of glutamine synthetase (GS) and glutamate synthase (GOGAT). GS is capable of catalyzing the ATP-dependent amidation of glutamic acid to yield glutamine. Subsequently, GOGAT transfers the amide group from glutamine to 2-oxoglutarate, yielding two molecules of glutamic acid. In this way, nitrogen is incorporated into carbon skeletons.[1,2] In cyanobacteria, the expression and activity of GS type I are modulated at the transcriptional and post-transcriptional levels.[3] In particular, there is a post-translational regulation of GS involving protein–protein interactions with a 65-residue-long (IF7) and a 149-residue-long (IF17) inactivating factors (IFs).[4–6] A maximal level of inactivation occurs *in vivo* when both IFs are present, although the presence of any of the two proteins is enough to produce some degree of GS inactivation.[4]

We have shown that both isolated IFs from *Synechocystis* sp. PCC 6803 are intrinsically disordered proteins (IDPs).[7,8] IDPs lack a well-defined globular three-dimensional structure, and they adopt an ensemble of rapidly interconverting conformations that in most cases, but not always, acquire an ordered structure upon binding to their targets. In general, IDPs have important roles in cell-cycle control, cellular signaling and transcription.[9] Among the advantages conferred by disorder are specificity, even at small binding strength, and promiscuity toward a large number of molecular partners.[9,10]

We have recently found that the binding between GS and IF7 is electrostatically driven,[11] although the binding reaction is not very fast; furthermore, mutational studies have shown that the substitution of Arg8, Arg21, and Arg28 residues of IF7 by glutamic acid abolishes the ability of the mutant protein to inactivate GS.[12] In this study, we investigated whether acidic residues (Glu and Asp) play a specific role in dictating the ensemble structure of IF7 in solution, and whether their ionization properties are different to those observed in random-coil models.[13,14] To understand how titratable groups can modulate protein binding, it is necessary to measure the pK_a values of ionizable residues and to identify the molecular determinants of their possible variations.[15,16] Deviations from the random-coil values may indicate the presence of electrostatic interactions, hydrogen bonds, hydrophobic effects, decreased solvent accessibility,[17–20] and even the propensity for large backbone fluctuations.[21–23]

Among various experimental techniques available, nuclear magnetic resonance (NMR) is perhaps the most general method to determine the individual pK_a values of residues, since signals assigned to specific atoms can be followed at most of the pH values.[24] Changes in chemical shifts of resonances of the titrating group (or nearby nuclei) will report on both the specific ionization sites (yielding the intrinsic pK_a values for a particular residue) and variations in the interactions with the other surrounding groups. Furthermore, molecular dynamics (MD) is one of the most common theoretical techniques that is routinely combined with NMR to model the folding properties of proteins[25] and can be used to investigate the conformational features and complex energy landscape of IDPs.[26,27]

Here, we carried out triple-resonance NMR experiments to measure the pK_a values for the side chains of the two aspartic (Asp40 and Asp58), three glutamic acids (Glu32, Glu36, and Glu38), and the main-chain C-terminal residue (Ser65) of IF7 at 100 m*M* NaCl. MD simulation was used to provide further details on protein conformations. We also compared the experimental pK_a values with those predicted from several on-line web-servers by using the structures obtained from our simulation. In this way, we could find out which interactions, if any, are relevant in the disordered ensemble of IF7 and in modulating its binding toward GS. Our results show that the acidic residues of IF7 have pK_a values similar to those measured in random-coil models, with insignificant deviations (within 0.3 units). Thus, both short- and long-range electrostatic interactions do not affect in a specific way the acidic residues of IF7 (even though they are clustered in a small protein region, i.e., the last 30 amino acids in the sequence), and they do not seem to play a driving role in binding to GS. Furthermore, our results also provide additional data for collections of pK_a values of residues in IDPs, which could help to clarify analogies and differences in their conformations compared to denatured protein states and random-coil regions in folded polypeptides.

Results

Experimental results

The pK_a values for Asp and Glu side chains and for the C-terminal Ser were obtained from triple resonance 3D C(CO)NH and HC(CO)NH spectra acquired at different pH values. We shall firstly

describe the titration of the two types of acidic residues and then of the C-terminal group.

Aspartic residues. In all monitored nuclei, the titration curves were monophasic (i.e., a single sigmoidal transition was observed). The triple-resonance NMR experiments showed that the intensities of the C_β correlations for the two Asp were larger than those for the C_α cross-peaks; that is, the strongest cross-peaks were those closest to the titrating carboxyl groups of the side chains. Similar results have been reported for titrations of the C_β and C_α protons of other IDPs.[28,29] We assume that the monitoring of the chemical shifts of the nearest nuclei to the ionizable moiety provides the most accurate measure of the real pK_a value of that moiety; that is, the nearby chemical environment has the largest effect on the NMR chemical shift of a particular nucleus.[30] It has been noted that these closest nuclei provide the most regular and reliable titrations, with the largest chemical shift changes.[30]

The pH dependence of the C_β chemical shifts ranged from ~39 to ~41 ppm [Fig. 1(A), Table I]; these values were similar to those observed in the protonated and deprotonated species of other disordered proteins or protein fragments.[28,29,31] In all cases, the Hill coefficients, n, for the titrations of the C_β resonances were ~1.0 (Table I). The pH-dependence of the C_α chemical shifts was smaller than that for the C_β, going from ~53 ppm (at acidic pH) to ~55 ppm (at physiological pH) [Fig. 1(B), Table I]; identical behavior has been found in other IDPs[28,29] and well-folded proteins.[30,32] The values of the pK_a for both residues, following the titration of the two nuclei, were identical (Table I). However, there were larger deviations from 1.0 for the n values of C_α than for those of the C_β resonances, probably due to the smaller variation of the chemical shifts of the C_α in the explored pH range; for instance, a Hill coefficient $n = 3 \pm 2$ was observed for the C_α of Asp58. This is due to the fact that the C_α is more distant than the C_β from the titrating side chain, and the closest nucleus to a particular titrating moiety is the one experiencing the largest effect on its chemical shifts.[30]

We also tried to follow the titration of the side chain of both Asp40 and Asp58 by following the chemical shifts of the H_β protons, although they can be affected both by nearby titrating groups and pH-dependent conformational changes[30,32] (see also Castañeda *et al.*[33] and references therein). In fact, the titration of Asp58 could not be monitored because of the very small difference in the δ_a and δ_b. The values of the chemical shifts of the H_β protons of Asp40 experienced a downfield shift (~0.07 ppm) at acidic pH values. A similar behavior has been observed for the H_β protons of both well-folded proteins[18,34] and IDPs,[29] although in these cases the

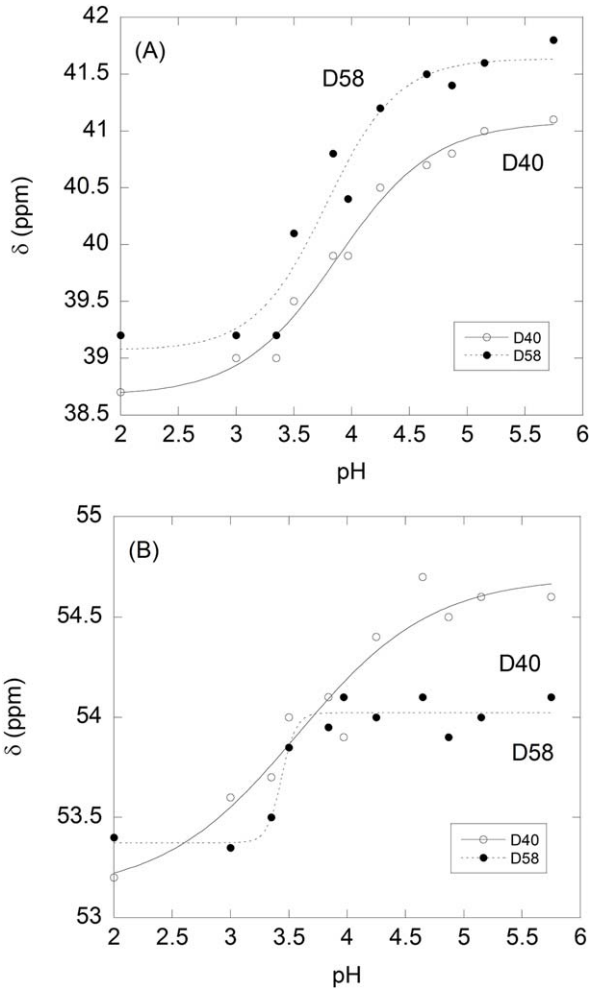

Figure 1. Titration curves for the Asp residues. (A) The pH profiles of the C_β resonances. (B) The pH profiles of the C_α resonances. In both panels, the lines are fittings to Eq. (1).

downfield variation is larger than in IF7 (~0.2 ppm). Given the sensitivity of protons to variations in the environment, these results suggest the presence of pH-induced conformational changes in IF7 under acidic conditions.[23,33]

The amide proton of Asp40 also titrated with a pK_a value similar, within the error, to that of the C_α and C_β resonances, although with $n > 1.0$. This is probably due to the small variation in the value of the NH chemical shift during the titration (~0.08 ppm, see Table I), as it was described above for the chemical shifts of the C_α of Asp58 at the different pH values. It has been shown that hydrogen bonds between amide protons and carboxylates can be identified on the basis of pH-dependent perturbations of the chemical shifts of the amide protons.[35] As the pH raises, the resonance of the hydrogen-bonded amide proton is downfield shifted (toward larger chemical shifts). In Asp40, such downfield shift was observed but its amplitude was small. These results suggest that the NH of Asp40 is hydrogen-bonded, possibly with its own side-chain.

Table I. *pK_a Values of Asp Residues in IF7 at 25°C*[a]

C_β resonances					C_α resonances			
Residue	pK_a	n	δ_a (ppm)	δ_b (ppm)	pK_a	n	δ_a (ppm)	δ_b (ppm)
Asp40[b]	3.8 ± 0.1	1.0 ± 0.1	38.7 ± 0.1	41.1 ± 0.1	3.6 ± 0.2	0.7 ± 0.3	53.1 ± 0.2	54.7 ± 0.2
	(3.5 ± 0.2)	(4 ± 2)	(8.09 ± 0.03)	(8.17 ± 0.07)				
Asp58[c]	3.8 ± 0.2	1.4 ± 0.4	39.0 ± 0.2	41.6 ± 0.2	3.6 ± 0.1	3 ± 2	53.3 ± 0.1	54.0 ± 0.1

[a] Reported errors are fitting errors to Eq. (1).
[b] The titration from the H_β yielded a pK_a value of 3.4 ± 0.1, with δ_a of 2.91 ± 0.01 and δ_b of 2.84 ± 0.01. The values within the parentheses are from the chemical shifts of the own amide proton.
[c] The titration of the H_β or NH could not be determined, because $|\delta_a - \delta_b| \leq 0.05$ ppm for each type of proton.

In contrast, we could not observe a titration for the NH of Asp58.

The pK_a values of both Asp were identical, within the error, to the value (p$K_a = 3.9$) measured for random-coil models.[13,14,30] Variations in pK_a values can be rationalized in terms of the free energy differences, according to: $\Delta G = (2.303)\ RT\ (\mathrm{p}K_a^{\mathrm{protein}} - \mathrm{p}K_a^{\mathrm{random-coil}})$. The ΔG values for both Asp are -0.1 ± 0.1 kcal mol^{-1}. These results suggest that in IF7 there were no energetically significant networks of electrostatically coupled interactions involving Asp.

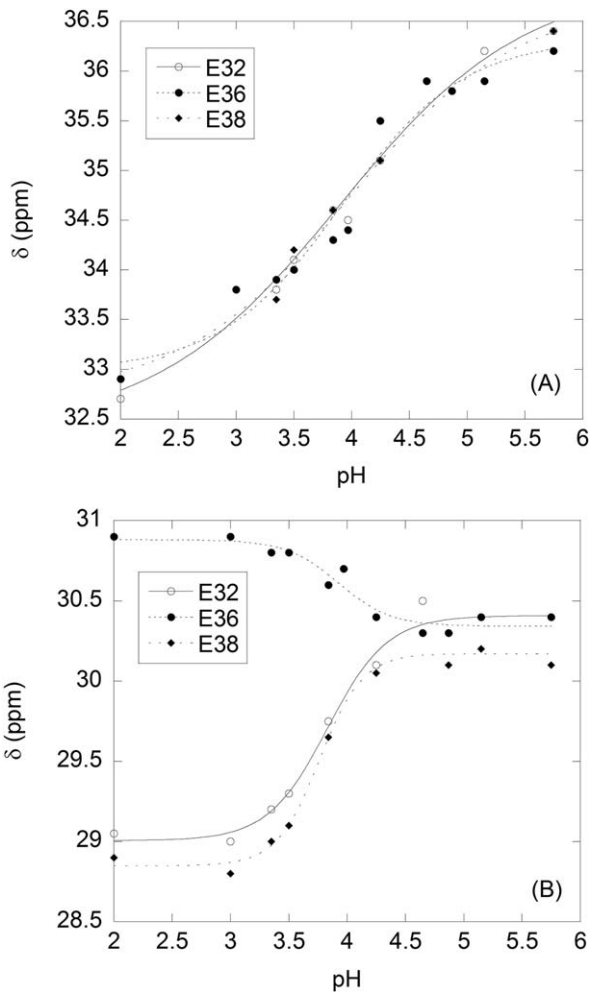

Figure 2. Titration curves for the Glu residues. (A) The pH profiles of the C_γ resonances. (B) The pH profiles of the C_β resonances. In both panels, the lines are fittings to Eq. (1).

Glutamic residues. In all monitored nuclei, the titration curves did have a single sigmoidal transition. The intensities of the C_γ cross-peaks for the three Glu residues were stronger than those of their C_β ones; furthermore, similarly as in Asp, the strongest cross-peaks were those closest to the titrating carboxyl groups of the side chains, as it has also been observed in well-folded proteins.[30,32] The pH-dependence of the C_γ resonances changed from ~33 (acidic pH) to ~37 ppm (physiological pH) [Fig. 2(A), Table II]. The pH dependence of the C_β resonances was from ~29 to ~30 ppm [Fig. 2(B)], and the variation of those of Glu36 was different. Interestingly enough, only the C_β resonances of Glu36 were shifted downfield at acidic pH values, suggesting that in this region there were pH-dependent conformational changes.[33] The pK_a values obtained for the three Glu residues at both C_β and C_γ were the same (Table II). The Hill coefficients for the titration curves of the C_γ resonances were close to 1.0, but those of the C_β cross-peaks were larger than unity. This is probably due to the small variation in the chemical shifts, as it happens with the C_α resonances of Asp (see above), and because the C_β is more distant from the titrating moiety.[30] We could not follow the titration of any of the H_β and H_γ of the Glu residues due either to overlapping or to their small amplitude change ($|\delta_a - \delta_b| \leq 0.05$ ppm).

The pK_a values of Glu32, Glu36, and Glu38 (Table II) were slightly smaller, but similar within the error, compared to that of random-coils (p$K_a = 4.3$).[13,14,30] The ΔG variations were around -0.4 ± 0.2 kcal mol^{-1}. As it happens with Asp, these results suggest that in IF7 there were no energetically significant networks of electrostatically coupled interactions involving the Glu residues.

C-terminal Ser65. The titration of the backbone carboxyl of Ser65 was followed by the chemical shifts of the C_α and C_β resonances of Leu64 (the NH of Ser65 itself was downfield shifted at acidic pH values, and its titration yielded a pK_a value of 3.96 ± 0.04, with

Table II. *pK_a Values of Glu and C-Terminal Residues in IF7 at 25°C*[a]

	C_γ resonances				C_β resonances			
Residue	pK_a	n	δ_a (ppm)	δ_b (ppm)	pK_a	N	δ_a (ppm)	δ_b (ppm)
Glu32	3.9 ± 0.2	0.6 ± 0.2	32.4 ± 0.5	36.9 ± 0.5	3.9 ± 0.1	1.7 ± 0.3	29.00 ± 0.06	30.41 ± 0.05
Glu36	3.9 ± 0.3	0.8 ± 0.2	32.9 ± 0.3	36.3 ± 0.3	4.0 ± 0.2	1.9 ± 0.7	30.82 ± 0.05	30.34 ± 0.04
Glu38[b]	4.0 ± 0.2	0.6 ± 0.2	32.7 ± 0.4	36.7 ± 0.5	3.9 ± 0.2	2.3 ± 0.8	28.85 ± 0.08	30.17 ± 0.04
	(3.7 ± 0.1)	(1.6 ± 0.3)	(8.17 ± 0.02)	(8.25 ± 0.02)				
Ser65[c]	3.27 ± 0.04	1.6 ± 0.3	38.0 ± 0.2	42.4 ± 0.1	3.32 ± 0.03	2.0 ± 0.5	53.21 ± 0.08	55.30 ± 0.04

[a] The variations for the H_β and H_γ protons in all Glu residues were $|\delta_a - \delta_b| \leq 0.05$ ppm. Reported errors are fitting errors to Eq. (1).
[b] The titration from the NH is within the parentheses. The rest of the amide protons for the other residues had $|\delta_a - \delta_b| \leq 0.05$ ppm.
[c] Determined from the titrations of the C_β (left side) and C_α (right side of the Table) resonances of Leu64.

$|\delta_a - \delta_b| = 0.06$ ppm). Both Leu64 resonances gave the same pK_a value, within the error: 3.29 ± 0.06 (Table II). This value was smaller than the one normally found in random-coils (p$K_a = 3.7$),[13,14] but the variation could be rationalized as probably due to hydrogen-bonding with the OH of its own side-chain. The corresponding ΔG was -0.5 ± 0.1 kcal mol^{-1}.

Molecular simulation and theoretical predictions

Conformational propensities of the protein model. The MD technique was exploited to gain insight into the structure of IF7 in solution, by collapsing a model built in extended conformation through a long term (40 ns) simulation run in water. As shown in Figure 3, the radius of gyration (R_G) during the simulation levels off to a plateau value after about 30 ns. The curve could be fitted with a simple exponential decay (red line in Fig. 3), in analogy with the equilibration process of typical structural properties of proteins.[36] The asymptotic value was $R_G = 17 \pm 2$ Å, which is consistent with the Stokes radius we previously[7] determined by gel filtration experiments ($R_S = 15.6 \pm 0.2$ Å).

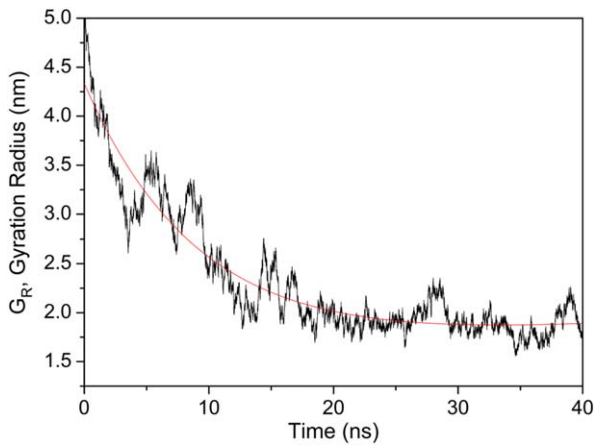

Figure 3. Protein radius of gyration as a function of simulation time. The curve (black) is fitted (adjusted $R^2 = 0.90$) with an exponential decay (red) function.

These observations rule out an over-compactness of the protein structure due to deficiencies of the force field, which is a common concern in the simulation of IDPs.[27,37] More importantly, as reported in Figure 4, a clear tendency toward the formation of transient helical structure in the region encompassing residues 3–28 was observed during the simulation. This result is in agreement with previous experimental data,[11] in which we have identified the presence of helical conformations in the protein N-terminal region. In particular, the simulation correctly reproduced the helical propensity in the Thr3-Arg13 region and around Ser27, while also predicting a tendency in the formation of some additional secondary structure for residues 17–20. Thus, the structures sampled in the time interval 30–40 ns can be considered representative states explored by IF7 in solution, although this does

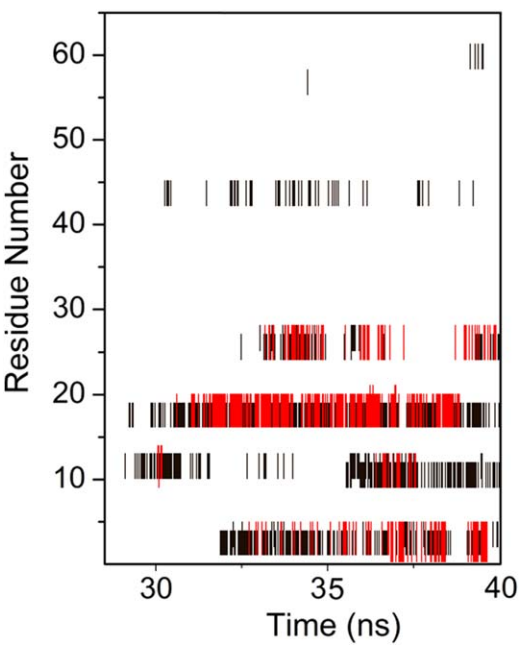

Figure 4. Protein residues with secondary structure as a function of simulation time. Residues showing (red) α-helical and (black) 3_{10}-helical structure at equilibrium (~30–40 ns) are shown.

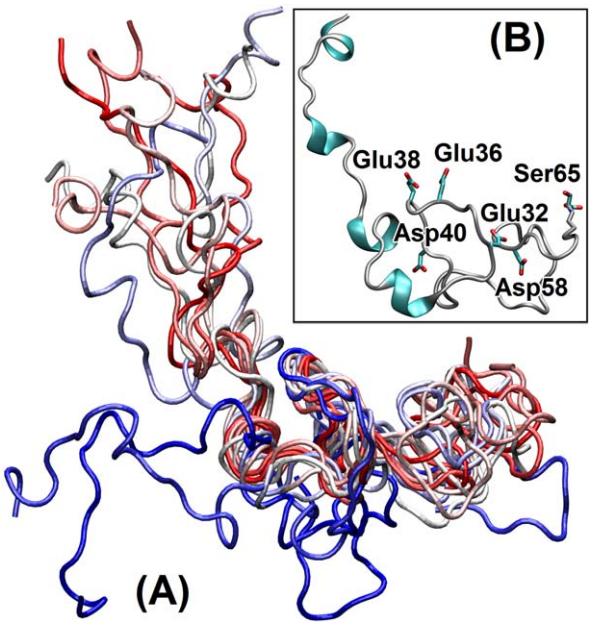

Figure 5. Simulated protein conformations. (A) Ensemble of structures sampled at 1 ns in the time interval 30-40 ns (color scale is from blue to red); MultiProt[38] was used for molecule superposition. (B) Protein snapshot representing the average conformation, with helical structure and acidic residues explicitly shown; the N-terminus is on the upper left and the C-terminus on the right.

not prove that they form a complete statistical ensemble of its whole conformational space.[25]

Visual inspection of the aligned protein structures[38] revealed further details of the conformational propensities of IF7. The snapshots reported in Figure 5 show that the protein forms a tadpole-like ensemble of structures, which is one of the most typical conformations for weak polyampholytes in solution.[39] The protein N-terminus had a clear tendency to be very dynamic and approximately elongated, favoring the nucleation of helical structures. In contrast, the C-terminal region of the protein sequence [where the acidic residues are located, see Fig. 5(B)] was more compact and globule-like and showed smaller fluctuations of the backbone. The side-chains of the ionizable residues were solvent-exposed and prevented a hydrophobic collapse of the protein structure; occasionally, they formed transient hydrogen-bonds, mostly with the amide group of the same residue.

Structure-based estimation of pK_a values. The protein structures obtained in simulation were used to investigate the titration values observed in our experiments from a theoretical point of view, with the help of three popular algorithms (DelPhiPKa,[40] H++,[41] and PROPKA[42]) used for the prediction of pK_a values. Protein snapshots extracted at intervals of 1 ns showed a variability in the calculated titration values of 0.07–0.22 units, which is comparable to the experimental uncertainties obtained in our NMR experiments (0.1–0.2 units of pK_a). The H++ tool outperformed the other two in predicting the titration midpoints for all Glu and Asp residues: pK_a values were closer to the experimental values in 85% of the cases. One of the advantages in the use of H++ was the flexibility in the implementation of the continuum solvent methodology used,[41] which allowed us to adjust both the dielectric constant of the protein and the ionic strength of the solution. In particular, values of the dielectric constant in the range 20–40 were found to be more appropriate to model the highly solvent-exposed residues of IF7, compared to values of 8–10 commonly used for well-folded proteins.[40,41]

As described for NUPR1,[29] another IDP, the theoretical pK_a values obtained by using H++ (Table III) were not accurate enough to be considered useful predictions. In fact, discrepancies with the experimental pK_a values for IF7 were >0.4 units on average (0.75 in the worst case). On the other hand, a systematic overestimation was noticed for all the pK_a values predictions. Furthermore, compared to random-coil values,[13,14,30] all the calculated pK_a values for IF7 were systematically higher for acidic residues and lower for basic ones. Although an effect of the empiric force field used in simulation could not be ruled out, we believe that this result was largely due to deficiencies in parameterization of the prediction algorithms. In fact, the same effect was observed also for DelPhiPKa, but not for PROPKA. The latter performed poorly compared to H++, except for an excellent estimate of the pK_a value of the C-terminus (the discrepancy with the experimental pK_a was 0.01 units).

To correct the systematic effects observed, we plotted the pK_a values predicted by H++ versus the experimental ones (Supporting Information Fig. S1) and fitted them to a line constrained to pass through the point of coordinates (7;7). This resulted in a slope of 1.1, suggesting that the values can be improved by using the relationship $pK_a^{(corr)} = ((pK_a^{(theor)} - 7) \cdot 1.1) + 7$, where $pK_a^{(corr)}$ is the corrected value (Table III). Adjustments to each single estimated pK_a value were up to 0.30 and

Table III. *Comparison of Experimental pK_a Values with those Obtained by H++,[41] Empirically Corrected, and Expected for Random-Coil Models.*[13,14]

Residue	Exp.	H++	H++ (corr)[a]	Random-coil models
Glu32	3.9	4.50	4.22	4.3
Glu36	4.0	4.38	4.12	4.3
Glu38	4.0	4.40	4.14	4.3
Asp40	3.8	3.99	3.69	3.9
Asp58	3.8	3.95	3.65	3.9
Ser65 (C-term.)	3.3	4.05	3.75	3.7

[a] The empirical correction is applied by using the equation $pK_a^{(corr)} = ((pK_a^{(theor)} - 7) \cdot 1.1) + 7$.

improved the agreement with the experimentally determined acidic values. The average discrepancy of $pK_a^{(corr)}$ with the experimental one was 0.2, mostly due to the poor prediction of the titration value of the C-terminal residue.

Discussion

Electrostatic interactions in proteins are mainly caused by the charged states of titratable residues. These interactions are known to play key roles in determining the structural and dynamical features of proteins, ultimately driving their functional properties. This conclusion is expected to be especially true in the case of IDPs, whose sequence is dominated by the presence of charged residues. Even relatively small shifts ($\sim$1 unit) in the pK_a values of a residue in IDPs can have a significant effect ($>$1 kcal/mol) in determining the ΔG between the unfolded state and a folded (or bound) state.[22] Nevertheless, our current understanding of how charge locations affect the function of natively unfolded proteins is still incomplete and requires further additional studies.

We contributed to this effort by using NMR to investigate the titration properties of the 65-residue-long IF7. Our study reveals that this IDP had pK_a values that agreed (within their uncertainty) with those found in random-coil models.[13,14] We also tested the reliability of three well-known algorithms (DelPhiPKa,[40] H++,[41] and PROPKA[42]) for predicting pK_a values on an experimentally validated ensemble of structures of IF7 obtained in MD simulation. Some caution should be exercised in the use of such structures to determine general properties of IF7, because they could not be an accurate representation of the whole conformational space accessible to the protein in solution. However, this ensemble of conformations (i) predicted the helical structure in the N-terminal region of the protein, as independently observed by NMR;[11] and, (ii) had the same radius of gyration as the one determined by size exclusion chromatography[7], suggesting that it sampled realistic IF7 states in solution, and it could be used to predict the titration properties of IF7 acidic residues. Although it is generally accepted that theoretical methods cannot predict pK_a values better than 0.5 units for folded proteins,[13,14,16] our results show that a small empirical adjustment was sufficient to reproduce with a reasonable accuracy the titration values for the acidic residues of IF7, and suggest that predictions may work well for IDPs in general.

To the best of our knowledge, pK_a values in IDPs have been determined only for α-synuclein[28] and NUPR1.[29] They both have values close to the ones of random-coils for most of, but not all, their acidic residues. We have previously argued[29] that the presence of Pro residues could help to shift the pK_a values of nearby acidic residues with respect to those of random-coils, suggesting yet an additional role in IDPs for such a multivalent amino acid.[43] In fact, Pro could act as a disorder-promoting residue that enhances the global propensity for protein backbone fluctuations, and yet may reduce the local conformational variety and solvent accessibility in specific regions of the sequence. In IF7, we note that Pro residues are underrepresented (one single occurrence is found, whereas Pro accounts for 6.3% of all amino acids in general,[44] and is often more abundant in IDPs), which could help to explain the random-coil-like behavior observed in the pK_a values.

Dissimilarities both in the amount and distribution of charged residues along the sequence could be another key issue that influences the titration values of ionizable residues in IDPs. By using molecular simulations, Pappu and coworkers[39,45] have shown how the charge content in highly charged polypeptides can modulate their conformational propensity, which has a direct influence on the deviations of pK_a values from random-coil ones.[21–23] Due to the small number of IDPs studied so far, it is not yet clear how this impacts the titration values in functional proteins, whose charge distribution is under evolutionary constraints—and thus is not random. IF7 is a weak polyampholyte (22% charged residues) compared to α-synuclein and NUPR1 (28% and 30%, respectively), whereas their net charge per residue is similar (6%). Some of the properties of IF7 may be explained by its asymmetric charge distribution, with an extended polyelectrolytic N-terminus containing only basic residues, and a globular polyampholytic C-terminal region. However, NUPR1[29] also has a polyelectrolytic (acidic only) N-terminal region, and α-synuclein[28] has three domains with different electrostatic features (an amphipathic N-terminus, a hydrophobic central region, and an acidic and Pro-rich C-terminus). Thus, further studies will be required to collect additional experimental data useful to understand the titration properties of IDPs.

Deviations of pK_a values from random-coil ones have been observed in a variety of polypeptide chains that include completely disordered proteins[15,19,20,46,47] and unfolded fragments.[31,48] In this context, IF7 appears to constitute a notable exception and suggests that the distribution of charges (and their associated network of long- and short-range electrostatic interactions) in unfolded protein sequences does not necessarily result in altered titration values of residues.

Conclusion

The properties of unfolded proteins are difficult to determine both in experiment and simulation, because they depend upon their labile conformational ensemble. The assessment of the ionization

properties of IDPs and the determinants of their pK_a values provides detailed information on the general nature of such disordered systems. This work is one of the still few reports on the titration properties of an IDP. The use of NMR spectroscopy allowed us to measure the pK_a values of acidic residues. The results were also compared with titration values predicted through theoretical models, based on protein structures obtained in an independent, unbiased MD simulation. The overall findings concur to suggest that the pK_a values of IF7 were indistinguishable, within their experimental uncertainties, from those obtained in model compounds. This behavior is uncommon among unfolded proteins and constitutes a remarkable novelty in IDPs. To the best of our knowledge, this constitutes the first reported case of a functional protein showing random-coil-like pK_a values of its ionizable residues.

Experimental Procedures

Materials

Deuterium oxide was obtained from Apollo Scientific (UK). Sodium trimethylsilyl [2,2,3,3-$^2H_4$] propionate (TSP), deuterated acetic acid, its sodium salt, and the commercial buffers (pH 4.0, 7.0, and 10.0) for pH-meter calibration were from Sigma (Spain). Dialysis tubing, with a molecular weight cut-off of 3500 Da, was from Spectrapor (Spectrum Laboratories, Japan). Standard suppliers were used for all other chemicals. Water was deionized and purified on a Millipore system.

Protein expression and purification

The recombinant IF7 protein was expressed and purified as described by using His-tagged vectors.[7,8] Protein stocks were run in SDS-PAGE gels and found to be >97% pure. Protein concentration was determined from the absorbance of individual amino acids.[49]

NMR sample preparation

Samples used for NMR experiments contained ~1.5 mM of ^{13}C, ^{15}N-labeled IF7 dissolved in a 90% H_2O and 10% D_2O mixture (v/v) with further addition of 100 m*M* NaCl. Spectra were acquired between pH 2.0 and 6.0. Attempts to follow titration curves at higher pH were unsuccessful due to solvent-exchange of the amide resonances.[11] The pH of the NMR sample was adjusted by adding aliquots of concentrated (11 *M*) solutions of either DCl (deuterium chloride) or NaOD (sodium deuteroxide). NMR samples were contained in 5 mm tubes (Wilmad) with an initial volume of 700 μL, larger than the recommended value (500 μL), to compensate for sample losses during the pH measurements and the pipetting.

The pH measurements were carried out in a Radiometer Analytical Ion-check 10 (Denmark) using an ultrathin glass Crison electrode (VWR Spain). Three-point electrode calibrations (4.0, 7.0, and 10.0) were done before and after each pH titration. Solution pH was measured before and after each experiment, with typical differences of 0.05–0.1 pH units; the average of such measurements was taken as the value for each particular pH. To ensure the stability of the pH measurements we followed the same protocol described elsewhere.[29]

NMR spectroscopy

The NMR experiments were acquired on a Bruker Avance DRX-500 spectrometer equipped with a room-temperature triple resonance probe and *z*-pulse field gradients. Experiments were acquired at 25°C; probe temperature was calibrated with a methanol NMR standard.[50] TSP was used as an internal reference; the ^{15}N and ^{13}C chemical shifts were indirectly referenced as described.[51] The intrinsic pH-dependence of the TSP was taken into account and corrected.[18,52] The peaks in the 2D- and 3D-NMR spectra at the different pH values were identified by using the assignments at pH 4.5 and 25°C (BMRB number 25921).[11]

The pH-titration of ^{13}C, ^{15}N-labeled IF7 was followed with 3D C(CO)NH, HC(CO)NH[53,54] and 2D-HSQC spectra.[55] The spectral acquisition parameter and processing sets [performed by using Topsin 2.1 (Bruker)] have been described elsewhere.[29]

Titration data fitting

The simplest model (the Henderson-Hasselbalch equation) for determining the pK_a values of ionizable groups can be extended to account for remote interactions with other chemical groups by introducing the Hill coefficient, n, as an additional parameter:[30]

$$\delta(pH) = \frac{\delta_b + \delta_a 10^{n(\mathrm{p}K_a - \mathrm{pH})}}{1 + 10^{n(\mathrm{p}K_a - \mathrm{pH})}} \quad (1)$$

where δ(pH) is the measured chemical shift at any pH; δ_b is the chemical shift associated with the unprotonated (basic) residue; δ_a is the chemical shift associated with the protonated (acidic) species. The values determined for the Hill coefficient are lower or higher than 1 in the case of, respectively, negative or positive co-operativity in the ionization due to the surrounding chemical environment. Fitting of data to Eq. (1) was carried out by using Kaleidagraph (Synergy software). We also carried out fitting of the titration curves, according to the same equation, without the use of the Hill coefficient. However, inclusion of the Hill coefficient improved the regression coefficient and χ^2 in all the curves for any of the nuclei monitored (as judged by a *F*-test of both parameters in both fittings in each titration, with a 95% probability).

Molecular modeling and dynamics

A model of IF7 was built by using the Molefacture plugin of VMD[56] and simulated with the GROMACS software.[57] The protein backbone was assembled in extended conformation (dihedral angles $\varphi = -180°$ and $\psi = 180°$), with bends in correspondence of Pro residues. All amino acids were parameterized using the Amber 99SB-ILDN force field.[58] To mimic the experimental conditions, the protonation of titratable residues was adjusted to acidic pH, and counterions were added. The protein structure was first equilibrated in an isochoric–isothermal run performed *in vacuo*, with removal of both translation and rotation around the center of mass. This first simulation slightly reduced the R_G, from 6 to 5 nm. Afterwards, the protein was placed in a rhombic dodecahedron box of 4000 nm^2 and solvated with ~133,000 water molecules, for which the TIP3P model was used.[59] A production run in the isobaric-isothermal ensemble at 300 K and 10^5 Pa was performed for 40 ns, controlling the electrostatic interactions through the PME method.[60]

Prediction of pK_a values

Expected pK_a values were calculated on the basis of the protein structures obtained in simulation, by using three web-accessible servers: DelPhiPKa, H++ and PROPKA,[40–42] all generating predictions based on continuum solvation models. Their accuracy in determining the correct values for an IDP was initially tested as described.[29] The dielectric constant for water was set to 80 in all cases, whereas all the other adjustable parameters were set to fit the experimental values. In particular, the best results were obtained by using internal dielectric constant values in the range 20–40, which are appropriate for highly-solvated protein regions.[41]

Acknowledgments

The authors thank two anonymous reviewers for their helpful suggestions. They deeply thank May García, María del Carmen Fuster, and Javier Casanova for technical assistance.

References

1. Leigh JA, Dodsworth JA (2007) Nitrogen regulation in bacteria and archaea. Annu Rev Microbiol 61:349–377.
2. Luque I, Forchammer KA, Nitrogen assimilation and C/N balance sensing. In: Herrero A, Flores E, Eds. (2008) The cyanobacteria: molecular biology, genetics and evolution. Norwich, UK: Caister Academic Press, pp 335–382.
3. Muro-Pastor MI, Reyes JC, Florencio FJ (2005) Ammonium assimilation in cyanobacteria. Photosynt Res 83: 135–150.
4. García-Domínguez M, Reyes JC, Florencio FJ (1999) Glutamine synthetase inactivation by protein–protein interaction. Proc Natl Acad Sci USA 96:7161–7166.
5. Reyes JC, Florencio FJ (1995) A novel mechanism of glutamine synthetase inactivation by ammonium in the *cyanobacterium Synechocystis* sp. PCC 6803. Involvement of an inactivating protein. FEBS Lett 367:45–48.
6. Galmozzi CV, Saelices L, Florencio FJ, Muro-Pastor MI (2010) Posttranscriptional regulation of glutamine synthetase in the filamentous cyanobacterium *Anabaena* sp. PCC 7120: differential expression between vegetative cells and heterocysts. J Bacteriol 192:4701–4711.
7. Muro-Pastor MI, Barrera FN, Reyes JC, Florencio FJ, Neira JL (2003) The inactivating factor of glutamine synthetase, IF7, is a "natively unfolded" protein. Protein Sci 12:1443–1454.
8. Saelices L, Galmozzi CV, Florencio FJ, Muro-Pastor MI, Neira JL (2011) The inactivating factor of glutamine synthetase IF17 is an intrinsically disordered protein, which folds upon binding to its target. Biochemistry 50:9767–9778.
9. Wright PE, Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol 16:18–29.
10. Zhou HX (2011) Intrinsic disorder: signaling via highly specific but short-lived association. Trends Biochem Sci 37:43–48.
11. Pantoja-Uceda D, Neira JL, Saelices L, Robles-Rengel R, Florencio FJ, Muro-Pastor MI, Santoro J (2016) Dissecting the binding between glutamine synthetase and its two natively unfolded protein inhibitors. Biochemistry 55:3370–3382.
12. Saelices L, Galmozzi CV, Florencio FJ, Muro-Pastor MI (2011) Mutational analysis of the inactivating factors, IF7 and IF17 from *Synechocystis* sp. PCC 6803: critical role of arginine amino acid residues for glutamine synthetase inactivation. Mol Microbiol 82:964–975.
13. Grimsley GR, Scholtz JM, Pace CN (2009) A summary of the measured pK values of the ionizable groups in folded proteins. Protein Sci 18:247–251.
14. Pace CN, Grimsley GR, Scholtz JM (2009) Protein ionizable groups: pK values and their contribution to protein stability and solubility. J Biol Chem 284:13285–13289.
15. Swint-Kruse L, Robertson AD (1995) Hydrogen bonds and the pH dependence of ovomucoid third domain stability. Biochemistry 34:4724–4732.
16. Alexov E, Mehler EL, Baker N, Baptista AM, Huang Y, Milletti F, Nielsen JE, Farrell D, Carstensen T, Olsson MHM, Shen JK, Warwicker J, Williams S, Word JM (2011) Progress in the prediction of pK_a values in proteins. Proteins 79:3260–3275.
17. Song J, Laskowski M, Jr, Qasim MA, Markley JL (2003) NMR determination of pK_a values for Asp, Glu, His, and Lys mutants at each variable contiguous enzyme-inhibitor contact position of the turkey ovomucoid third domain. Biochemistry 42:2847–2856.
18. Kuhlman B, Luisi DL, Young P, Raleigh DP (1999) pK_a values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions. Biochemistry 38:4896–4903.
19. Oliveberg M, Arcus VL, Fersht AR (1995) pK_A values of carboxyl groups in the native and denatured states of barnase: the pK_A values of the denatured state are on average 0.4 units lower than those of model compounds. Biochemistry 34:9424–9433.
20. Tan YJ, Oliveberg M, Davis B, Fersht AR (1995) Perturbed pK_A-values in the denatured states of proteins. J Mol Biol 254:980–992.

21. Müller-Späth S, Soranno A, Hirschfeld V, Hofmann H, Rüegger S, Reymond L, Nettels D, Schuler B (2010) Charge interactions can dominate the dimensions of intrinsically disordered proteins. Proc Natl Acad Sci USA 107:14609–14614.
22. Cho JH, Meng W, Sato S, Kim EY, Schindelin H, Raleigh DP (2014) Energetically significant networks of coupled interactions within an unfolded protein. Proc Natl Acad Sci USA 111:12079–12084.
23. Richman DE, Majumdar A, García-Moreno BE (2014) pH dependence of conformational fluctuations of the protein backbone. Proteins 82:3132–3143.
24. Forsyth WR, Antosiewicz JM, Robertson AD (2002) Empirical relationships between protein structure and carboxyl pK_a values in proteins. Proteins 48:388–403.
25. Rizzuti B, Daggett V (2013) Using simulations to provide the framework for experimental protein folding studies. Arch Biochem Biophys 531:128–135.
26. Espinoza-Fonseca LM, Ilizaliturri-Flores I, Correa-Basurto J (2012) Backbone conformational preferences of an intrinsically disordered protein in solution. Mol Biosyst 8:1798–1805.
27. Knott M, Best RB (2012) A preformed binding interface in the unbound ensemble of an intrinsically disordered protein: evidence from molecular simulations. PLoS Comput Biol 8:e1002605-1–10.
28. Croke RL, Patil SM, Quevreaux J, Kendall DA, Alexandrescu AT (2011) NMR determination of pK_a values in α-synuclein. Protein Sci 20:256–269.
29. Neira JL, Rizzuti B, Iovanna JL (2016) Determinants of the pK_a values of ionizable residues in an intrinsically disordered protein. Arch Biochem Biophys 598: 18–27.
30. Webb H, Tynan-Connolly BM, Lee GM, Farrell D, O'Meara F, Søndergaard CR, Teilum K, Hewage C, McIntosh LP, Nielsen JE (2010) Remeasuring HEWL pK_a values by NMR spectroscopy: methods, analysis, accuracy and implications for theoretical pK_a calculations. Proteins 79:685–702.
31. Pujato M, Navarro A, Versace R, Mancusso R, Ghose R, Tasayco ML (2006) The pH-dependence of amide chemical shift of Asp/Glu reflects its pK_a in intrinsically disordered proteins with only local interactions. Biochim Biophys Acta 1764:1227–1233.
32. Hass MAS, Mulder FAA (2015) Contemporary NMR studies of protein electrostatics. Annu Rev Biophys 44: 53–75.
33. Castañeda CA, Fitch CA, Majumdar A, Khangulov V, Schlessman JL, García-Moreno BE (2009) Molecular determinants of the pK_a values of Asp and Glu residues in staphylococcal nuclease. Proteins 77:570–588.
34. Schaller W, Robertson AD (1995) pH, ionic strength, and temperature dependences of ionization equilibria for the carboxyl groups in turkey ovomucoid third domain. Biochemistry 34:4714–4723.
35. Bundi A, Wüthrich K (1979) Use of amide ^{1}H-NMR titration shifts for studies of polypeptide conformation. Biopolymers 18:299–311.
36. Walton EB, Van Vliet KJ (2006) Equilibration of experimentally determined protein structures for molecular dynamics simulation. Phys Rev E 74:061901–1-8.
37. Piana S, Donchev AG, Robustelli P, Shaw DE (2015) Water dispersion interactions strongly influence simulated structural properties of disordered protein states. J Phys Chem B 119:5113–5123.
38. Shatsky M, Nussinov R, Wolfson HJ (2004) A method for simultaneous alignment of multiple protein structures. Proteins 56:143–156.
39. Das RK, Pappu RV (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proc Natl Acad Sci USA 13:13392–13397.
40. Wang L, Zhang M, Alexov E (2015) DelPhiPKa web server: predicting pK_a of proteins, RNAs and DNAs. Bioinformatics 32:614–615.
41. Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A (2005) H++: a server for estimating pK_a and adding missing hydrogens to macromolecules. Nucl Acids Res 33:W368–W371.
42. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: Consistent treatment of internal and surface residues in empirical pK_a predictions. J Chem Theory Comput 7:525–537.
43. Theillet F-X, Kalmar L, Tompa T, Han K-H, Selenko P, Dunker AK, Daughdrill GD, Uversky VN (2013) The alphabet of intrinsic disorder. Intrinsically Disord Proteins 1:e24360-1–13.
44. Morgan AA, Rubenstein E (2013) Proline: the distribution, frequency, positioning, and common functional roles of proline and polyproline sequences in the human proteome. PLoS One 8:e53785-1–9.
45. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. Proc Natl Acad Sci USA 107:8183–8188.
46. Tollinger M, Forman-Kay JD, Kay LE (2002) Measurement of side-chain carboxyl pK_a values of glutamate and aspartate residues in an unfolded protein by multinuclear NMR spectroscopy. J Am Chem Soc 124:5714–5717.
47. Meng W, Raleigh DP (2011) Analysis of electrostatic interactions in the denatured state ensemble of the N-terminal domain of L9 under native conditions. Proteins 79:3500–3510.
48. Pujato M, Bracken C, Mancuso R, Cataldi M, Tasayco ML (2005) pH dependence of amide chemical shifts in natively disordered polypeptides detects medium-range interactions with ionizable residues. Biophys J 89: 3293–3302.
49. Gill SC, von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem 182:319–326.
50. Cavanagh JF, Wayne J, Palmer III AG, Skelton NJ (1996) Protein NMR spectroscopy: principles and practice. San Diego, CA: Academic Press.
51. Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfiled E, Markley JL, Sykes BD (1995) ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. J Biomol NMR 6:135–140.
52. De Marco A (1977) pH dependence of internal references. J Magn Reson 26:527–528.
53. Logan TM, Oljniczak ET, Xu RX, Fesik SW (1993) A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments. J Biomol NMR 3: 225–231.
54. Montelione GT, Kyons BA, Emerson SD, Tashiro M (1992) An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. J Am Chem Soc 114: 10974–10975.
55. Bodenhausen G, Ruben D (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. Chem Phys Lett 69:185–189.
56. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph Model 14:33–38.

57. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29: 845–854.
58. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78:1950–1958.
59. 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.
60. Essmann U (1995) A smooth particle mesh Ewald method. J Chem Phys 103:8577–8593.