The N⁰-binding region of the vesicular stomatitis virus phosphoprotein is globally disordered but contains transient α -helices

Cédric Leyrat,¹ Malene Ringkjøbing Jensen,² Euripedes A. Ribeiro Jr.,¹ Francine C. A. Gérard,¹ Rob W. H. Ruigrok,¹ Martin Blackledge,² and Marc Jamin^{1*}

¹Unit of Virus Host Cell Interactions, UMI 3265 UJF-EMBL-CNRS, 38042 Grenoble Cedex 9, France

²Protein Dynamics and Flexibility, Institut de Biologie Structurale, UMR 5075 CEA-CNRS-UJF, Grenoble 38027, France

Received 1 October 2010; Revised 9 December 2010; Accepted 21 December 2010

DOI: 10.1002/pro.587

Published online 4 January 2011 proteinscience.org

Abstract: The phosphoprotein (P) of vesicular stomatitis virus (VSV) interacts with nascent nucleoprotein (N), forming the N⁰-P complex that is indispensable for the correct encapsidation of newly synthesized viral RNA genome. In this complex, the N-terminal region (P_{NTR}) of P prevents N from binding to cellular RNA and keeps it available for encapsidating viral RNA genomes. Here, using nuclear magnetic resonance (NMR) spectroscopy and small-angle X-ray scattering (SAXS), we show that an isolated peptide corresponding to the 60 first N-terminal residues of VSV P (P₆₀) and encompassing P_{NTR} has overall molecular dimensions and a dynamic behavior characteristic of a disordered protein but transiently populates conformers containing α -helices. The modeling of P₆₀ as a conformational ensemble by the ensemble optimization method using SAXS data correctly reproduces the α -helical content detected by NMR spectroscopy and suggests the coexistence of subensembles of different compactness. The populations and overall dimensions of these subensembles are affected by the addition of stabilizing (1M trimethylamine-*N*-oxide) or destabilizing (6M guanidinium chloride) cosolvents. Our results are interpreted in the context of a scenario whereby VSV P_{NTR} constitutes a molecular recognition element undergoing a disorder-to-order transition upon binding to its partner when forming the N⁰-P complex.

Keywords: rhabdovirus; vesicular stomatitis virus; phosphoprotein; SAXS; NMR; intrinsically disordered proteins; molecular recognition element

Introduction

In the last decade, our understanding of the relationships between protein structure and function was revolutionized by the discoveries that numerous proteins are intrinsically disordered (IDPs) or con-

tain intrinsically disordered regions (IDRs) and that these disordered proteins are implicated in various biological functions.^{1–5} These findings continuously raise new challenges in structural biology for predicting IDPs or IDRs from the amino acid sequence and for characterizing their structural properties. Nuclear magnetic resonance (NMR) spectroscopy^{6–8} and small-angle scattering^{9,10} revealed to be particularly suitable for providing detailed structural information on these proteins. Various algorithms were developed for identifying IDPs or for locating IDRs.^{11,12} Complete genome surveys revealed that not all kingdoms of life are equal when we consider the abundance of IDPs and IDRs encoded in the genomes. A large fraction (20–40%) of eukaryote

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

Grant sponsor: French ANR (ANRAGE); Grant number: ANR-07-001-01; Grant sponsor: FINOVI Foundation; Lyonbiopôle; French Government (MENRT Fellowship); EMBO Fellowship; Lundbeckfonden, Denmark.

*Correspondence to: Marc Jamin, Unit of Virus Host Cell Interactions (UVHCI), UMI 3265 UJF-EMBL-CNRS, 6, rue Jules Horowitz, B.P. 181, 38042 Grenoble Cedex 9, France. E-mail: jamin@embl.fr

genomes encodes IDP or proteins containing long IDRs, whereas they are much less abundant (<10%) in archaea and bacteria.^{13–15} Viruses, and in particular RNA viruses, occupy an intermediate position between eukaryotes and bacteria and exhibit a high rate of loosely packed proteins.¹⁶

The Rhabdoviridae is a family of nonsegmented negative-sense RNA viruses, which includes human (rabies virus—RV), animal (vesicular stomatitis virus—VSV), and plant pathogens. The machinery of RNA transcription and replication of these viruses is composed of the genomic RNA and of three viral proteins: the nucleoprotein (N), the phosphoprotein (P), and the large subunit of the RNA-dependent RNA polymerase (L). The RNA genome is encapsidated by N. Every N binds to nine nucleotides, forming a long helical N–RNA complex¹⁷ that serves as a template for both transcription and replication.¹⁸ The crystal structure of circular N–RNA complexes of both VSV and RV revealed that the N protein contains a N-terminal domain (N_{NTD}) and a C-terminal (N_{CTD}) domain separated by a positively charged groove, which binds the RNA.^{19,20} N_{NTD} and N_{CTD} act like jaws that close around and completely enwrap the RNA, forming multiple salt bridges with the phosphate backbone of the RNA. The phosphoprotein (P) is involved in different stages of the viral replication process.^{21,22} First, it associates with the L protein, forming active RNA polymerase complexes that catalyze both transcription and replication of the viral genome. In these complexes, L carries out the RNA synthesis and the 5' and 3' RNA processing required for the production of messenger RNAs, whereas P acts as an essential cofactor²¹ by attaching L to its N–RNA template.^{23,24} P possesses binding sites for both L and the N–RNA complex and thus ensures the correct positioning of the polymerase onto its template. Second, P is essential for encapsidating newly synthesized RNA genomes.^{25–27} P binds to nascent N molecules, forming a N⁰–P complex (the superscript ⁰ denotes the absence of RNA), which prevents N from binding to cellular RNAs and maintains N in a soluble form, available for the encapsidation of newly synthesized RNA genomes.^{22,25,27–30} In both VSV and RV P, the N-terminal region of the protein (P_{NTR}), residues 4–40 in RV and residues 11–30 in VSV, is sufficient for maintaining N in a soluble and monomeric form.^{29,30} Because P_{NTR} is rich in negatively charged residues, it was hypothesized that P_{NTR} binds in the RNA-binding cavity of N,^{29,31} thereby preventing the nonspecific attachment of N⁰ to cellular RNA.

Recently, we showed that both RV and VSV P form dimers in solution³² and are modular proteins made of structured domains and IDRs.¹² The consensus prediction of disordered regions in both RV and VSV P obtained by combining results from different disorder predictors suggested that the 39

N-terminal amino acids of VSV P and the 52 N-terminal amino acids of RV P are structured,^{12,33} whereas the following residues (aa 40–90 in VSV; aa 53–86 in RV) would be disordered. However, in a previous study, an isolated peptide corresponding to the N-terminal region of RV P (RV P₆₈; aa 1–68) appeared to be disordered in solution.¹² Static light scattering revealed that RV P₆₈ was monomeric, its Stokes' radius measured by size exclusion chromatography (SEC) corresponded to that expected for an unfolded protein of the same molecular mass, and its mean secondary structure content measured by circular dichroism (CD) was less than 5%.¹² These contradicting features, namely predictions of a folded domain in P₆₈ and experimental results showing that P₆₈ is disordered in solution, are reminiscent of some recently identified short protein motifs that fold only upon binding to their partners and were named molecular recognition elements (MoREs) or molecular recognition features.^{34–37}

To further investigate the structural properties of the N-terminal region of P, we studied a peptide corresponding to the 60 N-terminal residues of VSV P (P₆₀) by NMR spectroscopy and small-angle X-ray scattering (SAXS). We choose to work with a peptide of this length to ascertain that it extends beyond the N⁰-binding region (P_{NTR})³⁰ and the N-terminal region that was predicted to be structured.¹² Also, an equivalent peptide of 60 residues was shown to inhibit more efficiently the transcription and replication in RV than a peptide corresponding to the 40 N-terminal residues.³⁸ Although the NMR spectra recorded with VSV P₆₀ exhibit typical features of disordered proteins, transiently formed secondary structures were identified from NMR chemical shifts and the dynamics of the polypeptide chain was probed by ¹⁵N relaxation rates. In addition, the average compactness of P₆₀ was probed by SAXS, and a conformational ensemble that accounts for all experimental observations was built by modeling SAXS data with the ensemble optimization method (EOM). Also, the effects of denaturing or stabilizing cosolvents on the size distribution of the representative ensemble were measured by SAXS. Our results revealed that, in solution, P₆₀ is a highly dynamic protein transiently forming α -helical elements and populating compact conformers that could play a major role in the formation of the N⁰–P complex.

Results

Isolated VSV P₆₀ is intrinsically disordered but exhibits transient secondary structures

The N-terminal domain of VSV P (P₆₀, aa 1–60) was produced in *Escherichia coli* as a recombinant protein containing a C-terminal His-tag (8 aa). The long-term stability of P₆₀ was significantly improved in the presence of a mixture of 50 mM L-Arg and

Figure 1. Molecular mass and Stokes' radius of isolated VSV P₆₀ measured by SEC/MALLS/RI. The line shows the SEC elution profile as monitored by refractometry (left axis). The crosses show the molecular mass calculated from light scattering and refractometry data across the elution peak (right axis). The Stokes's radius (R_S) was determined from the elution volume by calibrating the system with standard proteins of known Stokes' radius (see Materials and Methods).

50 mM L-Glu.³⁹ Under these conditions, an average molecular mass of 7.2 ± 0.7 kDa was determined by SEC combined with detection by multiangle laser light scattering and refractometry (SEC/MALLS/RI), indicating that P₆₀ was monomeric (MM calculated from the sequence = 8054 Da) and well behaved in solution up to concentrations of 2 mM (Fig. 1). Its Stokes' radius of 2.3 ± 0.1 nm indicated an extended conformation, closer to that expected for an unfolded protein of the same molecular mass ($R_S = 2.5$ nm) than to that expected for a globular protein ($R_S = 1.5$ nm).⁴⁰

SAXS profiles of P₆₀ were collected for scattering vector values ($Q = 4\pi(\sin\theta)/\lambda$) ranging from 0.1 to 3.0 nm^{-1} over a concentration range of 5–18 mg mL⁻¹. The profiles obtained at different protein concentrations had the same shape and were flat at low Q values, indicating the absence of significant aggregation [Fig. 2(a)]. The radius of gyration, R_g , values derived from the Guinier approximation at low Q value ($Q.R_g < 1.4$) [Fig. 2(b)] or from the distance distribution function ($P(r)$) [Fig. 2(c)] showed no significant dependence on protein concentration, confirming the absence of aggregation (Table I). The average R_g value of 2.4 ± 0.1 nm was close to the value of 2.5 nm calculated for a random coil of 68 amino acids (peptide + His-Tag) according to Flory's power-law dependence on chain length [Eq. (2)] and using parameters predicted for a random coil with excluded volume⁴¹ and to the value of 2.3 nm calculated from the same power-law dependence using parameters derived from experimental measurements of R_g for IDPs.¹⁰ The forward scattering intensity, $I(0)$, was proportional to protein concentration and yielded an estimated molecular mass of 11 ± 1 kDa, in agreement with SEC/MALLS/RI measurements,

confirming that P₆₀ was monomeric and well behaved in solution. The 2D ¹H-¹⁵N-HSQC NMR spectrum of P₆₀ exhibited narrow line widths and

Figure 2. Small-angle X-ray scattering experiments. (a) Scattering curve of P₆₀ recorded at different concentrations. In this panel and in panels b and c, the protein concentration was 5 mg mL⁻¹ (circles), 8 mg mL⁻¹ (squares), or 18 mg mL⁻¹ (diamonds). (b) Guinier plot. The fitted lines using $0.10 < Q < 0.28 \text{ nm}^{-1}$ yield the radii of gyration shown in Table I. The fit corresponds to a range of $0.5 < Q.R_g < 1.5$. The $I(0)$ value calculated from the intercept together with protein concentrations yielded a molecular mass estimate of 11 ± 1 kDa. (c) Distance distribution function. The distance distribution functions were calculated with GNOM by using D_{max} values shown in Table I. The surface areas under the curves were normalized to account for the differences in protein concentration.

Table I. Molecular Dimensions Calculated from SAXS Data

Cosolvent	Protein concentration (mg mL ⁻¹)	R_g (Guinier) (nm)	R_g ($P(r)$) (nm)	$D_{\max}$ (nm)
50 mM Arg/	5	2.4 ± 0.1	2.4 ± 0.1	7.4
50 mM Glu	8	2.4 ± 0.1	2.4 ± 0.1	7.4
	18	2.4 ± 0.1	2.4 ± 0.1	7.4
10 mM Arg/	5	2.6 ± 0.1	2.6 ± 0.1	7.9
10 mM Glu				
6M GdmCl	5	2.9 ± 0.1	2.8 ± 0.1	8.2
1M TMAO	5	2.3 ± 0.1	2.3 ± 0.1	7.0

R_g values were determined from the Guinier plot or from the distance distribution function ($P(r)$ vs. r). $D_{\max}$ values beyond which $P(r) = 0$ were adjusted in order that the R_g values obtained with the program GNOM agree with those obtained from the Guinier analysis.

poor dispersion of the amide proton chemical shifts (Fig. 3), in agreement with structural disorder in solution. These results indicate that VSV P₆₀ has overall molecular dimensions and NMR spectroscopic properties of a disordered protein, as shown previously for RV P₆₈.¹²

The far-UV CD spectrum exhibited a prominent shoulder near 222 nm and a minimum near 205 nm, suggesting the presence of α -helical structure (Fig. 4). Assuming that the molar ellipticity value at 222 nm reports only on the presence of α -helix, the average helical content of P₆₀ was estimated to be 7%.^{12,42} NMR spectroscopy is highly sensitive to local structure and therefore provides information about the location and population of fluctuating secondary structures within an otherwise flexible protein.^{8,43–47} After complete assignment of the backbone nuclei (C_α , C_β , C' , N, and HN), the chemical shifts and ¹⁵N nuclear relaxation rates were used to identify transient structural features within P₆₀. The secondary structure propensity (SSP) calculated by combining secondary chemical shifts of C_α and C_β showed continuous stretches of positive propensity (SSP > 0.1) for residues 2–12 (helix 1) and 25–38 (helix 2), with the highest population in the region 25–31, suggesting the presence of residual helical structures in these two regions of the protein [Fig. 5(a)].⁴⁵ It also showed a stretch of negative propensity for residues 53–57 characteristic of an extended β structure. Averaging these propensities over P₆₀ yielded α -helical and β -structure contents of 9 and 1%, respectively, in good agreement with the helical content estimated from the CD spectrum. The uniformly low values of the heteronuclear [¹H]-¹⁵N NOEs indicated high backbone flexibility on the picosecond to nanosecond time scale [Fig. 5(b)]. The presence of transient secondary structures is, however, suggested from the ¹⁵N nuclear relaxation rates, R_1 and R_2 [Fig. 5(c,d)]. In particular, the ¹⁵N R_2 rates are larger in the regions of the peptide that have helical propensity according to the SSP score.

The regions 3–14 and 26–34 of P₆₀ exhibited significant intrinsic helical propensity [Fig. 5(e)],^{48,49} and amino acids that can accept hydrogen bonds from free backbone amide groups and have high N-capping preferences⁵⁰ were found at the N-terminal extremity of each helical segment. The program AGADIR predicted intrinsic N-capping propensities near 5% for Asp2 and Asn3 (N_{cap} for helix 1) and of 13.5% for Asp25 (N_{cap} for helix 2).^{48,49} These results suggest that the fluctuating helical elements are locally encoded in the amino acid sequence and stabilized by short-range interactions.

To probe the effect of the remaining part of P on the conformation of the N-terminal region, we recorded an NMR spectrum of full-length VSV P. Because the full-length protein aggregates at low pH, the NMR spectrum was obtained at pH 7.5. VSV P is a dimer of 62 kDa, but only a limited number of narrow amide backbone resonances were observed with a restricted ¹H chemical shift dispersion (Supporting Information Fig. S1). Although these resonances were not assigned, the spectral features suggest that only residues located in flexible regions of VSV P were visible. A comparison with an NMR spectrum of P₆₀ recorded under identical conditions showed that most of the resonances visible at pH 7.5 match with resonances of full-length VSV P. These results clearly argued that the N-terminal part of VSV P is also globally disordered in the context of the full-length dimeric protein.

Modeling VSV P₆₀ as a conformational ensemble

Recently, different methods were developed for building low-resolution models based on SAXS data.^{51–54} Many of these *ab initio* methods are not suitable for modeling highly flexible molecules because they yield a unique structure that at best would provide an average model of the most frequent conformers present in solution.^{55,56} Because VSV P₆₀ appeared as a disordered protein with fluctuating helical structures, we used the EOM⁹ (Fig. 6). This method starts from a large ensemble of randomly generated conformers and selects a subensemble of conformers that collectively reproduces the experimental SAXS data and represents the distribution of structures adopted by the protein in solution.⁹ Based on the location of residual helical elements identified by NMR spectroscopy, an initial ensemble of 4500 conformers of P₆₀ was built using the Flexible Meccano algorithm⁵⁷ by pooling an ensemble of 500 conformers containing no helical structure with eight ensembles, of 500 conformers each, containing helices of different lengths (helix 1: aa 2–9 or 2–19, helix 2: aa 25–32 or 25–38) as described in Materials and Methods. The black curves in Figure 6 show the R_g and $D_{\max}$ distributions calculated for the initial ensemble of conformers. These distributions are broad and slightly

Figure 3. 2D [^1H - ^{15}N]-HSQC NMR spectrum of VSV P₆₀. The spectrum was recorded in 20 mM Tris, pH 6.0, 150 mM NaCl, 50 mM Arg, and 50 mM Glu at 10°C.

asymmetrical, with R_g values extending from 1.4 to 4.0 nm with a maximum frequency near 2.3 nm [Fig. 6(a)] and with D_{max} values extending from 4.0 to 12.0 nm with a maximum frequency near 7.3 nm, respectively [Fig. 6(b)]. A subensemble of 20 conformers was selected for which the average back-calculated SAXS curve reproduced correctly the experimental curve ($\chi = 0.630$) [Fig. 7(a)]. The red curves in Figure 6 show the R_g and D_{max} distributions of the EOM-selected subensemble. They have widths similar to those of the initial ensemble as expected for a disordered P₆₀ [red curve in Fig. 6(a,b)]. Figure

6(i) shows representative members of the selected ensemble exhibiting varying R_g values as indicated in Figure 6(a). However, the R_g distribution within the selected ensemble exhibited two maxima suggesting the presence of two distinct and almost equally important subpopulations of conformers with average R_g values centered on 2.0 and 2.9 nm [Fig. 6(a)]. The D_{max} distribution is not as clearly bimodal [Fig. 6(b)]. Variation in the number of conformers selected by the EOM procedure from 50 to 2 had no significant effect on the quality of the fit and yielded similar bimodal distributions of R_g values (data not

Figure 4. Circular dichroism spectroscopy of VSV P₆₀. CD spectra were recorded at 20°C in 20 mM Tris-HCl, pH 7.5, containing 50 mM Arg and 50 mM Glu (1), 1M TMAO (2), or 6M GdmCl (3).

shown). The presence of helical elements in the models introduced specific features in the scattering curve at Q values above 1 nm^{-1} [Fig. 7(c)], and therefore, the average helix content of the selected ensemble of about 7% was in good agreement with estimations obtained from CD and NMR. Models containing helical elements were found in both subpopulations of compact and extended conformers [Fig. 6(a,i)]. Successive and independent selections by EOM yielded similar results and similar ensembles of conformers showing that the bimodal distribution of the selected ensemble was reproducible. Also, a similar bimodal distribution was obtained independently by using another initial ensemble of 10,000 conformers devoid of any explicit secondary structure generated with the program RanCh⁹ (data not shown). A simulation indicated that, at least under some conditions, the program EOM is capable of revealing the existence of two different populations of different average size (Supporting Information Fig. S2). The EOM approach thus revealed features in the disordered ensemble that could not be detected from the average dimensions measured as R_s by SEC or as R_g by SAXS. In our case, the overall ensemble has average dimensions close to those calculated from random-coil statistics,^{10,41,58} whereas the EOM approach suggested the coexistence of two different subensembles with molecular dimensions lower and higher than those predicted for a random coil.

To test the effect of Arg and Glu, we also recorded SAXS data in the presence of 10 mM Arg/10 mM Glu. The average R_g value was slightly larger ($2.6 \pm 0.1 \text{ nm}$) than in 50 mM Arg/50 mM Glu, indicating that these cosolutes induce the compaction of the protein. The SAXS profile was also different [Fig. 7(a)], and the selected ensemble obtained with EOM showed a different distribution of population [Fig. 6(c,d)]. Although the R_g and D_{max} distributions of the selected ensemble conserved a

similar width than in 50 mM Arg and 50 mM Glu, the two subpopulations were not as clearly individualized, showing that a higher concentration of these cosolutes stabilizes the more compact subensemble.

The addition of stabilizing (trimethylamine-*N*-oxide, TMAO) or destabilizing (guanidinium chloride, GdmCl) cosolutes induced variations in the average helical content as shown by far-UV CD spectroscopy (Fig. 4). In 1M TMAO, a well-known stabilizing cosolute,⁵⁹ the ellipticity at 222 nm indicated a higher helical content (15%), whereas in 6M GdmCl, it indicated a lower helical content (<1%). These cosolutes also affected the average size of P₆₀, and analysis of the SAXS profiles yielded a higher average R_g values of $2.9 \pm 0.1 \text{ nm}$ in 6M GdmCl and a lower average R_g value of $2.3 \pm 0.1 \text{ nm}$ in 1M TMAO (Table I). The back-calculated SAXS curve from the ensemble of 20 conformers selected by the program EOM also adequately reproduced the experimental curve in the presence of these cosolutes [Fig. 7(b)]. The modeling of SAXS curves using the program EOM revealed that the variations of the average R_g value arose from changes in the subpopulations of compact and expanded conformers and from slight variations of the average R_g values of these two subpopulations. In the presence of 6M GdmCl, the subpopulation of compact conformers completely disappeared, and the width of the selected distributions was significantly reduced [Fig. 6(e,f)] when compared with those in 50 mM Arg/50 mM Glu [Fig. 6(a,b)]. A single maximum was observed in both distributions of R_g and D_{max} [Fig. 6(e,f)] with a maximum frequency at R_g value near 2.7 nm and D_{max} value near 8.2 nm. These values were slightly lower than the R_g and D_{max} values found for the extended subpopulation observed in the presence of 50 mM Arg/50 mM Glu, suggesting that the subensemble with the highest average R_g value contained extended conformers that could not be further unfolded by addition of 6M GdmCl. The R_g value of 2.9 nm found for this subensemble is larger than that of 2.5 nm predicted for a random coil of this size in a good solvent⁴¹ and also larger than that of 2.3 for an IDP of this size.¹⁰ Because P₆₀ has a high negative net charge at neutral pH, this discrepancy could result from charge repulsion leading to an expansion of the molecule.⁶⁰ GdmCl can act as both denaturant and salt and, therefore, could reduce repulsions between charged residues by screening charges, a behavior which would explain the little compaction of the structural ensemble observed in 6M GdmCl.

By contrast, in the presence of 1M TMAO, both subensembles remained equally populated [Fig. 6(g)]. They became slightly more compact than in 50 mM Arg/50 mM Glu with maximal frequencies centered on R_g values of 1.9 and 2.7 nm in agreement with the smaller experimental average R_g and D_{max}

Figure 5. NMR dynamics of isolated VSV P₆₀. (a) Secondary structure propensity (SSP). The SSP parameter was calculated from C_α and C_β secondary chemical shifts. (b) ¹H-¹⁵N heteronuclear NOEs. (c, d) Relaxation rates of longitudinal, R_1 , and transverse, R_2 , magnetization of backbone ¹⁵N nuclei. The shaded area highlights the regions predicted to be helical from the SSP parameter (aa 2–12 and 24–39). NMR data were recorded at 600 MHz and 10°C using a peptide sample in 20 mM Tris, pH 6.0, 150 mM NaCl, 50 mM Arg, and 50 mM Glu. (e) Intrinsic helical propensity. Bar graph showing the intrinsic helical propensity per residue calculated with the program AGADIR.⁴⁸

Figure 6. Conformational ensemble selection and effects of stabilizing and destabilizing cosolutes. Ensembles of 20 conformers that collectively reproduce the experimental curves were selected from the initial ensemble. In each figure, the black curve shows the R_g and D_{max} distributions calculated for the initial ensemble of conformers, whereas the red curve shows the R_g and D_{max} distributions of the selected ensemble that fit the experimental SAXS data. (a) R_g and (b) D_{max} distribution from the EOM analysis in 50 mM Glu and 50 mM Arg. (c) R_g and (d) D_{max} distribution from the EOM analysis in 10 mM Glu and 10 mM Arg. (e) R_g and (f) D_{max} distributions from the EOM analysis in 6M GdmCl. (g) R_g and (h) D_{max} distributions from the EOM analysis in 1M TMAO. (i) Members of the conformational ensemble. The cartoon models show some of the selected models at varying R_g values, highlighting the presence of residual helical structures. The chains are colored from the N-terminal (blue) to the C-terminal (red). The numbers refer to their position in the distribution shown in Figure 6(a).

Figure 7. Reproducing SAXS curves. (a) Effects of Arg and Glu on the SAXS curve of P₆₀. The SAXS curve in 50 mM Arg and 50 mM Glu was obtained at three different protein concentrations (Fig. 2), and the EOM analysis was performed on the curve at 5 mg mL⁻¹ to allow direct comparison with SAXS curves obtained in 10 mM Arg/10 mM Glu, 1M TMAO, and 6M GdmCl that were also obtained at 5 mg mL⁻¹. The fitted curve calculated using the EOM procedure for P₆₀ in 10 mM Arg/10 mM Glu is shown in blue and that in 50 mM Arg/50 mM Glu is shown in red. Data were recorded at 20°C in 20 mM Tris-HCl, pH 7.5, buffer containing the corresponding cosolutes. (b) Effects of 1M TMAO and 6M GdmCl on the SAXS curve of P₆₀. The fitted curve calculated using the EOM procedure for P₆₀ in 50 mM Arg/50 mM Glu is shown in red, that in 1M TMAO in green, and that in 6M GdmCl in magenta. Data were recorded at 20°C in 20 mM Tris-HCl, pH 7.5, buffer containing the corresponding cosolutes. (c) Simulated SAXS curves of ensembles containing different amounts of helical elements. The curves are average SAXS profiles calculated for the 500 models that contain no helical elements (no helix) and for the 500 models that contain helices in the regions 2–19 and 25–38 (32% helical content).

values measured directly from the SAXS curves (Table I). The distribution of $D_{\max}$ values in the selected ensemble also appeared slightly bimodal with a maximum at a $D_{\max}$ value of 8.2 nm and a shoulder near 6.2 nm [Fig. 6(h)]. Also, in the presence of 1M TMAO, the models selected by the EOM procedure contained more residual α -helical structure in average and longer helices (the average helical content in the selected ensemble was 30%) (data not shown) in agreement with CD experiments.

Discussion

The conformational ensemble representing P₆₀

The narrow line widths in the NMR spectrum and the limited chemical shift dispersion reflect the rapid interconversion between heterogeneous conformations of VSV P₆₀. ¹⁵N relaxation experiments show that the protein is highly flexible with the exception of two regions, aa 2–12 and aa 25–38, that exhibit an increased level of order. This observation is supported by secondary chemical shifts that indicate that both regions transiently populate α -helical conformations. The presence of α -helix is confirmed by CD spectrum, and the SSP parameter derived from the C_{α} and C_{β} chemical shifts provides an estimate of the time and ensemble average population of these helical elements near 9%. Short elements of secondary structure such as α -helices, β -strands, or turns are locally encoded and often stable in water.^{61–63} α -Helices form very quickly in the 10- to 100-ns time range in agreement with the timescale probed by ¹⁵N relaxation measurements.⁶⁴ NMR experiments thus provide a dynamic picture of P₆₀ where the polypeptide chain is highly flexible but populates fluctuating α -helices in two regions.

The hydrodynamic radius of P₆₀ measured by SEC and its radius of gyration measured by SAXS are close to those expected for denatured proteins⁵⁸ or for IDPs¹⁰ of the same size. The dimensions of both classes of proteins follow power-law dependences on chain length characteristic of random-coil behavior.^{65,66} However, the random-coil model is insensitive to the presence of stiff elements in an otherwise flexible chain, and the presence of transient formation of α -helical elements in a disordered polypeptide chain is coherent with random-coil statistics,^{41,67,68} as already pointed out by Tanford⁶⁶ and demonstrated by simulations.^{41,69} The size distribution (R_g) of our selected ensemble has a width similar to that of the initial ensemble, showing that P₆₀ samples a global conformational space predicted for a random coil although it contains on average 9% of fluctuating α -helices.

Recently, using NMR and SAXS experiments, unfolded or IDPs have been modeled as structural ensembles to convey the highly dynamic character of these proteins.^{9,36,44,57,70} Here, we combined

information from NMR experiments showing the existence of two fluctuating α -helical elements with SAXS data to build an ensemble of 20 different conformations that together represent the dynamic structural organization of VSV P₆₀. The ensemble of 20 conformations selected by fitting the SAXS curve exhibits an average helical content in agreement with estimations obtained by CD and NMR and correctly captures the dynamical behavior of the polypeptide chain. The structure of the selected models must, however, be considered with care. The models selected by the EOM procedure only represent plausible conformers that together reproduce the SAXS data. They should by no means be taken as actual structures. No correlation has been found between the presence of helical conformation and the size of the conformers, and residual helical structures are found in conformers of different compactness. This argues that α -helices in P₆₀ are stabilized by local interactions as suggested by the predicted and observed propensity of the amino acid sequence to form helices in solution and, therefore, their presence is independent of the compactness of the peptide chain. However, it is also probable that the information content of SAXS data is insufficient to correlate the presence of helical structure with the degree of compaction of the chain.

In addition, a detailed analysis of the population distribution in the EOM-selected ensemble and the effects of denaturing and stabilizing cosolutes suggest the coexistence of two distinct subpopulations with different degrees of compactness. The EOM proved capable of revealing the coexistence of two subpopulations exhibiting a large difference in R_g value.⁷¹ Here, we used a simulation to demonstrate that EOM is also capable of distinguishing between unimodal and bimodal distributions under certain conditions (Supporting Information Fig. S2). In our approach, the addition of cosolutes was not intended to mimic physiological conditions although the intracellular medium is complex and contains large concentrations of various cosolutes. The addition of 50 mM Arg/50 mM Glu was used to improve solubility and prevent aggregation,³⁹ and the addition of TMAO and GdmCl was used to probe the conformational space accessible to P₆₀. The mechanism by which Arg and Glu achieve their effects has not been clearly demonstrated.^{39,72} Because the effects are observed at low concentration of cosolutes, they may result from binding to protein surface, and it has been suggested that the charges of the cosolutes interact with opposite charges on the protein surface, while the hydrophobic parts of the amino acids mask exposed hydrophobic patches.³⁹ In another study, it was suggested that Arg and Glu induce the ordering of residues at the protein surface and act as crowding agents.⁷² In our case, the addition of 50 mM Arg and 50 mM Glu increases the population of

compact conformers and therefore reduces the average size of the molecule.

The addition of TMAO or GdmCl was used to probe the conformational space that is accessible to the polypeptide chain and revealed how the peptide adapts to variations of the environment. Addition of GdmCl and TMAO changes the average size of the molecule and the population distribution selected by the EOM program. In particular, in 6M GdmCl, the width of the size distribution is substantially reduced and the distribution becomes unimodal [Fig. 6(e)], whereas in 1M TMAO, the bimodal distribution becomes more clearly visible [Fig. 6(g)]. These results suggest the presence of one subpopulation that is more compact than that predicted by random-coil statistics and one that is more expanded. Some IDPs are more compact than predicted from random-coil statistics because of the presence of long-range interactions involving hydrophobic clusters,^{73,74} hydrogen bonding,⁷⁵ or charge–charge attractions.^{60,76} Others appear more expanded because of charge–charge repulsions.^{77–79} Charge–charge interactions thus seem to play a critical role in modulating the dimensions of the conformational ensemble of IDPs^{60,76}; repulsions between like charges can lead to the expansion of the peptide chain, whereas attractions between opposite charges can lead to collapse. VSV P₆₀ is rich in charged residues, containing 15 acidic residues (D and E), seven basic residues (K and R), and one histidine residue (net charge at neutral pH of -7 or -8 if H is charged or neutral) and therefore different interactions between charged residues could explain the existence of the two different conformational subensembles. The salt properties of GdmCl can explain the disappearance in 6M GdmCl of the compact subensemble if its formation is driven by opposite charge interactions. Conversely, TMAO favors compaction of random coils,⁸⁰ and we observed slight reductions of the average dimensions of both selected conformational subensembles. Also, TMAO induces helix formation,⁸¹ and we observed a larger α -helical content as judged by CD spectroscopy and a higher number of models containing helical elements in the EOM-selected ensemble.

In conclusion, our data show that P₆₀ is highly flexible and has overall molecular dimensions of a random-coil chain, but it contains helical structures and can adopt compact and extended conformers. The presence of fluctuating helical elements and of a distinct population of compact conformers suggests that P₆₀ may exist in conformers that prefigure its structure in the N⁰–P complex and could be involved in the recognition of N⁰. However, the distribution of population observed here *in vitro* depends on environmental conditions and need not reflect the distribution under physiological conditions. Also, in its natural context, the peptide is part of a large

dimeric protein, and although our data indicate that P₆₀ is also flexible and disordered in full-length VSV P, it is possible that the remaining part of P influences the distribution of population of the two ensembles.

Role for a MoRE in the chaperone function of P

The discovery here that, in solution, P₆₀ populates preorganized conformers containing helical elements suggests a mechanism by which P_{NTR} could fold upon binding to its N⁰ partner.³ P_{NTR} could thus be a MoRE, that is, a short motif within a disordered protein that promotes binding to a partner.^{36,37,82–84} Structural ensembles of disordered proteins that fold upon binding to a partner usually contain conformations similar to those adopted in the final complex.^{36,70,85} The fluctuating α -helices might prefigure structural elements involved in the formation of the N⁰–P complex. The first helix (aa 8–12) is amphipathic exposing three positively charged residues and two negatively charged residues on one face of the helix and four hydrophobic residues on the other face. The second helix (aa 25–38) is hydrophilic exposing charged residues around the helix and containing only two hydrophobic residues (I27 and Y37). Both helices could thus interact with positively charged residues lining the RNA-binding groove of N⁰. One might speculate that the highly flexible character of isolated P₆₀ helps the protein to get into the RNA-binding groove and allows the correct positioning of the helical elements.

In the viral replication cycle, the N⁰–P complex is not an end product. It forms only transiently as an intermediate during the synthesis of new nucleocapsids. The N⁰-bound P must be outcompeted by the newly synthesized genomic RNA, and, therefore, the affinity of P for N⁰ cannot be too high. The mechanism of folding upon binding provides a means of specific recognition without the corollary of high affinity.^{3,86} Indeed, the folding or the adoption of a rigid structure by the ligand when it binds to its receptor will generally lead to the formation of multiple specific intermolecular interactions that confer a great specificity and, consequently, a high affinity. If the ligand is disordered in its unbound form, this strong binding energy is opposed by the high entropy of the disordered protein. We might thus speculate that, by being disordered in its free form, P_{NTR} recognizes N⁰ with high specificity but moderate affinity, thus allowing its displacement by the newly synthesized viral RNA.

Materials and Methods

Cloning, expression, and purification of VSV P₆₀

The cDNA-encoding VSV P₆₀ (aa 1–60) was amplified and cloned into a pET28a(+) plasmid between the NcoI and XhoI restriction sites using standard

PCR techniques. A His6-tag, including a linker of two amino acids (Glu-Leu), was introduced at the C-terminal extremity. The construct was verified by standard dideoxy sequencing. The plasmid was transfected into *E. coli* strain BL21(DE3). ¹⁵N and ¹³C protein samples for NMR spectroscopy were produced in M9 minimal medium supplemented with 1.0 g L⁻¹ of ¹⁵NH₄Cl, 2.0 g L⁻¹ of ¹³C glucose, and MEM vitamins (Gibco). Cells were grown in the medium containing 100 μ g mL⁻¹ of kanamycin at 37°C until OD₆₀₀ reached 0.6–0.8 A.U. At this point, IPTG at a final concentration of 0.5 mM was added. After an incubation of 3 h, cells were harvested by centrifugation and resuspended in 20 mM Tris-HCl, 150 mM NaCl, 50 mM Glu, and 50 mM Arg, pH 7.5 (Buffer A) supplemented with antiproteases (CompleteTM Protease Inhibitor Cocktail Tablets, Roche), and cells were disrupted by sonication. The extract was centrifuged at 20,000g during 30 min at 4°C. The supernatant was filtered (0.45 μ m), loaded onto a Ni²⁺ resin column, and pre-equilibrated with Buffer A. The resin was washed with three bed volumes of Buffer A, then with three bed volumes of 20 mM Tris-HCl, 1.5M NaCl, and 10 mM imidazole at pH 7.5, and the protein was eluted using Buffer A supplemented with 400 mM imidazole. The recombinant protein was loaded onto a SEC column (HiLoad 16/60 Superdex 75 prep grade, GE Healthcare) pre-equilibrated with Buffer A. Separations were performed at a flow rate of 0.5 mL min⁻¹. The purified protein was concentrated (Vivaspin—5000 MWCO) and stored at 4°C. The purity of the protein samples was assessed by SDS-PAGE. Protein concentration was measured by RI.

NMR spectroscopy

The resonance assignment of VSV P₆₀ was carried out using a double-labeled (¹³C, ¹⁵N) sample of the peptide in 20 mM Tris-HCl, 150 mM NaCl, 50 mM Glu, and 50 mM Arg with 10% D₂O adjusted to pH 6.0 to avoid loss of resonances due to fast amide proton exchange with the solvent protons. The protein concentration was 0.9 mM. All NMR experiments were carried out at 10°C and a ¹H resonance frequency of 599.67 MHz. The assignment was obtained from a series of BEST-type triple resonance experiments^{87,88}: HNCO, intraresidue HN(CA)CO, HN(CO)CA, intraresidue HNCA, HN(COCA)CB, and intraresidue HN(CA)CB. The spectra were acquired with a sweep width of 8 kHz and 512 complex points in the ¹H dimension, and a sweep width of 1.2 kHz and 32 complex points in the ¹⁵N dimension. For the ¹³C dimension, the spectra were acquired with a sweep width of 1.2 kHz and 60 complex points (HNCO and intraresidue HN(CA)CO), 3 kHz and 110 complex points (HN(CO)CA and intraresidue HNCA), and 10 kHz and 100 complex points (HN(COCA)CB and intraresidue HN(CA)CB). All

spectra were processed with NMRPipe⁸⁹ and analyzed using SPARKY.⁹⁰ Automatic assignment of the resonances on the basis of the SPARKY peak lists was done using the program MARS.⁹¹

¹⁵N R_1 , R_2 (CPMG) relaxation and $\{^1\text{H}\}$ -¹⁵N steady-state nuclear Overhauser effect (nOe) experiments were acquired using standard pulse sequences⁹² on an ¹⁵N-labeled sample of P₆₀ with a protein concentration of 1.4 mM. The buffer conditions were as described above for the double-labeled sample. The spectra were acquired with a sweep width of 7.0 kHz and 512 complex points in the ¹H dimension, and a sweep width of 1.2 kHz and 250 complex points in the ¹⁵N dimension.

The magnetization decay was sampled at (0, 100, 200, 400, 600, 800, 1100, 1500, and 1900) ms for longitudinal and at (10, 30, 50, 70, 90, 130, 170, 210, and 250) ms for transverse relaxation, and the peak heights were used to extract the relaxation rates. To obtain estimates of the errors on the relaxation rates, a repeat measurement of one of the relaxation delays was performed in each case (600 ms for R_1 and 70 ms for R_2).

For the heteronuclear nOes, the amide protons were saturated using a 3-s WALTZ16 decoupling scheme that in the reference experiment was replaced by a 3-s delay. The recycle delay in both experiments was set to 2 s. Heteronuclear NOE values were obtained from the ratio between signal intensities in the saturated and the reference experiments, where the standard deviation in the noise was taken as a measure for the error in the signal intensity. The experiment was repeated twice, and the NOE values were averaged.

SEC-MALLS-RI

SEC was performed with an S75 Superdex column (GE Healthcare) equilibrated with 20 mM Tris/HCl at pH 7.5 containing 150 mM NaCl. Separations were performed at 20°C with a flow rate of 0.5 mL min⁻¹. Fifty microliters of P₆₀ solution at a concentration of 470 μM (3.8 mg mL⁻¹) was injected. The peptide samples were prepared in the equilibration buffer supplemented with 50 mM Arg and 50 mM Glu. Online MALLS detection was performed with a DAWN-EOS detector (Wyatt Technology, Santa Barbara, CA) using a laser emitting at 690 nm. Data were analyzed, and weight-averaged molecular masses (M_w) were calculated using the ASTRA software (Wyatt Technology, Santa Barbara, CA) as described previously.³² The column was calibrated with proteins of known Stokes' radii (R_S)³⁵: bovine serum albumin ($R_S = 3.4$ nm), RNase A ($R_S = 1.9$ nm), ovalbumin ($R_S = 3.0$ nm), β-lactoglobulin ($R_S = 2.7$ nm), and chymotrypsinogen ($R_S = 2.3$ nm).⁴⁰

Circular dichroism spectroscopy

Far-UV CD spectra were recorded at 20°C on a JASCO model J-810 CD spectropolarimeter equipped

with a Peltier temperature controller. VSV P₆₀ was diluted to a final concentration of about 950 μM, in 20 mM Tris-HCl, pH 7.5, containing 150 mM NaCl, 50 mM Arg, and 50 mM Glu. Spectra were measured in a cuvette with a path length of 0.1 mm. After subtracting the blank signal, the CD signal (in millidegrees) was converted to mean molar residue ellipticity (in deg cm² dmol⁻¹), and helix content was calculated by using formalism derived from isolated peptide⁴² as described previously.¹²

Small-angle X-ray scattering experiments

The monodispersity of the samples used in SAXS experiments was checked by SEC combined with detection by MALLS and RI.^{32,93} SAXS data were collected at the European Synchrotron Radiation Facility (Grenoble, France) on beamline ID14-3. The sample-to-detector distance was 1 m, and the wavelength of the X-rays was 0.0995 nm. Samples were contained in a 1.9-mm-wide quartz capillary. The time of exposure was optimized for reducing radiation damage. Data acquisition was performed at 20°C. Protein concentrations ranged from 5 to 18 mg mL⁻¹. Data reduction was performed using the established procedure available at ID14-3, and buffer background runs were subtracted from sample runs.

The radius of gyration and forward intensity at zero angle ($I(0)$) were determined with the programs PRIMUS⁹⁴ by using the Guinier approximation at low scattering vector values, Q ($Q = \frac{4\pi \sin \theta}{\lambda}$), in a Q R_g range up to 1.3:

$$\ln I(Q) = \ln I(0) - \frac{R_g^2 Q^2}{3} \quad (1)$$

The forward scattering intensity was calibrated using bovine serum albumin and lysozyme as references. The radius of gyration and pairwise distance distribution function, $P(r)$, were calculated with the program GNOM.⁹⁵ The maximum dimension ($D_{\max}$) value was adjusted so that the R_g value obtained from GNOM agreed with that obtained from the Guinier analysis.

The radius of gyration, R_g , was calculated using the power-law dependence established for random-coil polymers⁶⁵:

$$R_g = R_0 N^\nu \quad (2)$$

where R_0 is the persistence length, N is the number of residues in the peptide, and ν is the exponential scaling factor. Values for R_0 and ν were taken either from the random-coil polymer theory⁴¹ or from experiments with IDPs.¹⁰

A conformational ensemble of 4500 conformers was generated with the program Flexible-Meccano⁵⁷ by pooling nine subensembles of 500 conformers in

which helical element of different lengths were imposed in the putative helical regions of P₆₀ identified by NMR spectroscopy as follows: Ensemble 1: no helix; Ensemble 2: aa 2–9; Ensemble 3: aa 2–19; Ensemble 4: aa 25–32; Ensemble 5: aa 25–38; Ensemble 6: aa 2–9 and aa 25–32; Ensemble 7: aa 2–9 and aa 25–38; Ensemble 8: aa 2–19 and aa 25–32; and Ensemble 9: aa 2–19 and aa 25–38. All helices were invoked with a population of 100%. Side chains for these conformations were added with the program SCCOMP.⁹⁶ An optimized ensemble of conformations that agrees with the experimental SAXS data was selected from the large conformational ensemble using the EOM software.⁹ The size of the optimized ensemble was varied from 50 to 2.

Acknowledgments

The authors thank the Partnership for Structural Biology for the excellent structural biology environment.

References

- Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 293:321–331.
- Dunker AK, Brown CJ, Obradovic Z (2002) Identification and functions of usefully disordered proteins. *Adv Protein Chem* 62:25–49.
- Wright PE, Dyson HJ (2009) Linking folding and binding. *Curr Opin Struct Biol* 19:31–38.
- Tomba P (2002) Intrinsically unstructured proteins. *Trends Biochem Sci* 27:527–533.
- Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 41:415–427.
- Eliez D (2009) Biophysical characterization of intrinsically disordered proteins. *Curr Opin Struct Biol* 19:23–30.
- Dyson HJ, Wright PE (2004) Unfolded proteins and protein folding studied by NMR. *Chem Rev* 104:3607–3622.
- Jensen MR, Markwick PR, Meier S, Griesinger C, Zweckstetter M, Grzesiek S, Bernado P, Blackledge M (2009) Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. *Structure* 17:1169–1185.
- Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* 129:5656–5664.
- Bernado P, Blackledge M (2009) A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. *Biophys J* 97:2839–2845.
- He B, Wang K, Liu Y, Xue B, Uversky VN, Dunker AK (2009) Predicting intrinsic disorder in proteins: an overview. *Cell Res* 19:929–949.
- Gérard FCA, Ribeiro EA, Leyrat C, Ivanov I, Blondel D, Longhi S, Ruigrok RWH, Jamin M (2009) Modular organization of rabies virus phosphoprotein. *J Mol Biol* 388:978–996.
- Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 337:635–645.
- Oldfield CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, Dunker AK (2005) Comparing and combining predictors of mostly disordered proteins. *Biochemistry* 44:1989–2000.
- Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ (2000) Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform* 11:161–171.
- Tokuriki N, Oldfield CJ, Uversky VN, Berezovsky IN, Tawfik DS (2009) Do viral proteins possess unique biophysical features? *Trends Biochem Sci* 34:53–59.
- Fields BN, Knipe DM, Howley PM, Chanock RM, Melnik TP, Roizman B, Straus SE (1996) *Fields' virology*, 3rd ed. New York: Lippincott-Raven Publishers.
- Arntner H, Davis NL, Wertz G, Schubert M, Lazzarini RA (1985) Role of the nucleocapsid protein in regulating vesicular stomatitis virus RNA synthesis. *Cell* 41:259–267.
- Albertini AA, Wernimont AK, Muziol T, Ravelli RB, Clapier CR, Schoehn G, Weissenhorn W, Ruigrok RW (2006) Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* 313:360–363.
- Green TJ, Zhang X, Wertz GW, Luo M (2006) Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* 313:357–360.
- Emerson SU, Yu Y (1975) Both NS and L proteins are required for in vitro RNA synthesis by vesicular stomatitis virus. *J Virol* 15:1348–1356.
- Masters PS, Banerjee AK (1988) Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA. *J Virol* 62:2658–2664.
- Emerson SU, Schubert M (1987) Location of the binding domains for the RNA polymerase L and the ribonucleocapsid template within different halves of the NS phosphoprotein of vesicular stomatitis virus. *Proc Natl Acad Sci USA* 84:5655–5659.
- Mellon MG, Emerson SU (1978) Rebinding of transcriptase components (L and NS proteins) to the nucleocapsid template of vesicular stomatitis virus. *J Virol* 27:560–567.
- Peluso RW, Moyer SA (1988) Viral proteins required for the in vitro replication of vesicular stomatitis virus defective interfering particle genome RNA. *Virology* 162:369–376.
- Peluso RW (1988) Kinetic, quantitative, and functional analysis of multiple forms of the vesicular stomatitis virus nucleocapsid protein in infected cells. *J Virol* 62:2799–2807.
- Howard M, Wertz G (1989) Vesicular stomatitis virus RNA replication: a role for the NS protein. *J Gen Virol* 70:2683–2694.
- Mavrakis M, Iseni F, Mazza C, Schoehn G, Ebel C, Gentzel M, Franz T, Ruigrok RW (2003) Isolation and characterisation of the rabies virus N degrees-P complex produced in insect cells. *Virology* 305:406–414.
- Mavrakis M, Mehoulas S, Real E, Iseni F, Blondel D, Tordo N, Ruigrok RW (2006) Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA. *Virology* 349:422–429.
- Chen M, Ogino T, Banerjee AK (2007) Interaction of vesicular stomatitis virus P and N proteins: identification of two overlapping domains at the N-terminus of P that are involved in N0-P complex formation and encapsidation of viral genome RNA. *J Virol* 81:13478–13485.
- Hudson LD, Condra C, Lazzarini RA (1986) Cloning and expression of a viral phosphoprotein: structure suggests

- vesicular stomatitis virus NS may function by mimicking an RNA template. *J Gen Virol* 67:1571–1579.
32. Gérard F, Ribeiro E, Albertini A, Zaccari G, Ebel C, Ruigrok R, Jamin M (2007) Unphosphorylated Rhabdoviridae phosphoproteins form elongated dimers in solution. *Biochemistry* 46:10328–10338.
 33. Ribeiro EA, Jr, Favier A, Gérard FC, Leyrat C, Brutscher B, Blondel D, Ruigrok RW, Blackledge M, Jamin M (2008) Solution structure of the C-terminal nucleoprotein-RNA binding domain of the vesicular stomatitis virus phosphoprotein. *J Mol Biol* 382:525–538.
 34. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6:197–208.
 35. Sugase K, Dyson HJ, Wright PE (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447:920–921.
 36. Jensen MR, Houben K, Lescop E, Blanchard L, Ruigrok RW, Blackledge M (2008) Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. *J Am Chem Soc* 130:8055–8061.
 37. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN (2006) Analysis of molecular recognition features (MoRFs). *J Mol Biol* 362:1043–1059.
 38. Castel G, Chteoui M, Caignard G, Prehaud C, Mehoulas S, Real E, Jallet C, Jacob Y, Ruigrok RW, Tordo N (2009) Peptides that mimic the amino-terminal end of the rabies virus phosphoprotein have antiviral activity. *J Virol* 83:10808–10820.
 39. Golovanov AP, Hautbergue GM, Wilson SA, Lian LY (2004) A simple method for improving protein solubility and long-term stability. *J Am Chem Soc* 126:8933–8939.
 40. Uversky VN (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. Fitzkee NC, Rose GD (2004) Reassessing random-coil statistics in unfolded proteins. *Proc Natl Acad Sci USA* 101:12497–12502.
 42. Luo P, Baldwin RL (1997) Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry* 36:8413–8421.
 43. Wishart DS, Sykes BD, Richards FM (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol* 222:311–333.
 44. Jensen MR, Salmon L, Nodet G, Blackledge M (2010) Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts. *J Am Chem Soc* 132:1270–1272.
 45. Marsh JA, Singh VK, Jia Z, Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci* 15:2795–2804.
 46. Eliezer D, Chung J, Dyson HJ, Wright PE (2000) Native and non-native secondary structure and dynamics in the pH 4 intermediate of apomyoglobin. *Biochemistry* 39:2894–2901.
 47. Spera S, Bax A (1991) Empirical correlation between protein backbone conformation and C α and C β ^{13}C nuclear magnetic resonance chemical shifts. *J Am Chem Soc* 113:5490–5492.
 48. Munoz V, Serrano L (1994) Elucidating the folding problem of helical peptides using empirical parameters. *Nat Struct Biol* 1:399–409.
 49. Lacroix E, Viguera AR, Serrano L (1998) Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J Mol Biol* 284:173–191.
 50. Doig AJ, Baldwin RL (1995) N- and C-capping preferences for all 20 amino acids in alpha-helical peptides. *Protein Sci* 4:1325–1336.
 51. Franke D, Svergun DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Cryst* 42:342–346.
 52. Svergun DI, Petoukhov MV, Koch MH (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80:2946–2953.
 53. Konarev P, Petoukhov M, Volchkov V, Svergun DI (2006) ATSAS 2.1, a program package for small-angle scattering data analysis. *J Appl Cryst* 39:277–286.
 54. Putnam CD, Hammel M, Hura GL, Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* 40:191–285.
 55. Bernado P (2010) Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. *Eur J Biophys* 39:769–780.
 56. Heller WT (2010) Influence of multiple well defined conformations on small-angle scattering of proteins in solution. *Acta Cryst D* 61:33–44.
 57. Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RW, Blackledge M (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc Natl Acad Sci USA* 102:17002–17007.
 58. Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, Dothager RS, Seifert S, Thiyagarajan P, Sosnick TR, Hasan MZ, Pande VS, Ruczinski I, Doniach S, Plaxco KW (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc Natl Acad Sci USA* 101:12491–12496.
 59. Bolen DW, Baskakov IV (2001) The osmophobic effect: natural selection of a thermodynamic force in protein folding. *J Mol Biol* 310:955–963.
 60. Muller-Spáth S, Soranno A, Hirschfeld V, Hofmann H, Ruegger 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.
 61. Baldwin RL, Rose GD (1999) Is protein folding hierarchical? I. Local structure and peptide folding. *Trends Biochem Sci* 24:26–33.
 62. Fitzkee NC, Fleming PJ, Gong H, Panasik N, Jr, Street TO, Rose GD (2005) Are proteins made from a limited parts list? *Trends Biochem Sci* 30:73–80.
 63. Dyson HJ, Wright PE (2002) Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. *Adv Protein Chem* 62:311–340.
 64. Eaton WA, Munoz V, Hagen SJ, Jas GS, Lapidus LJ, Henry ER, Hofrichter J (2000) Fast kinetics and mechanisms in protein folding. *Annu Rev Biophys Biomol Struct* 29:327–359.
 65. Flory PJ (1953) Principles of polymer chemistry. Ithaca, NY: Cornell University Press.
 66. Tanford C (1968) Protein denaturation. *Adv Protein Chem* 23:121–282.
 67. Millett IS, Doniach S, Plaxco KW (2002) Toward a taxonomy of the denatured state: small angle scattering

- studies of unfolded proteins. *Adv Protein Chem* 62: 241–262.
68. Jha AK, Colubri A, Freed KF, Sosnick TR (2005) Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc Natl Acad Sci USA* 102: 13099–13104.
 69. Miller WG, Goebel CV (1968) Dimensions of protein random coils. *Biochemistry* 7:3925–3935.
 70. Lange OF, Lakomek NA, Fares C, Schroder GF, Walter KF, Becker S, Meiler J, Grubmuller H, Griesinger C, de Groot BL (2008) Recognition dynamics up to micro-seconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 320:1471–1475.
 71. Bernado P, Perez Y, Svergun DI, Pons M (2008) Structural characterization of the active and inactive states of Src kinase in solution by small-angle X-ray scattering. *J Mol Biol* 376:492–505.
 72. Blobel J, Schmidl S, Vidal D, Nisius L, Barnado P, Millet O, Brunner E, Pons M (2007) Protein tyrosine phosphatase oligomerization studied by a combination of ^{15}N NMR relaxation and ^{129}Xe NMR. Effect of buffer containing arginine and glutamic acid. *J Am Chem Soc* 129:5946–5953.
 73. Felitsky DJ, Lietzow MA, Dyson HJ, Wright PE (2008) Modeling transient collapsed states of an unfolded protein to provide insights into early folding events. *Proc Natl Acad Sci USA* 105:6278–6283.
 74. Mittag T, Orlicky S, Choy WY, Tang X, Lin H, Sicheri F, Kay LE, Tyers M, Forman-Kay JD (2008) Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc Natl Acad Sci USA* 105: 17772–17777.
 75. Moglich A, Joder K, Kiefhaber T (2006) End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proc Natl Acad Sci USA* 103: 12394–12399.
 76. 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.
 77. Uversky VN, Gillespie JR, Millett IS, Khodyakova AV, Vasiliev AM, Chernovskaya TV, Vasilenko RN, Kozlovskaya GD, Dolgikh DA, Fink AL, Doniach S, Abramov VM (1999) Natively unfolded human prothymosin α adopts partially folded collapsed conformation at acidic pH. *Biochemistry* 38:15009–15016.
 78. Permyakov SE, Millett IS, Doniach S, Permyakov EA, Uversky VN (2003) Natively unfolded C-terminal domain of caldesmon remains substantially unstructured after the effective binding to calmodulin. *Proteins* 53: 855–862.
 79. Uversky VN, Li J, Souillac P, Millett IS, Doniach S, Jakes R, Goedert M, Fink AL (2002) Biophysical properties of the synucleins and their propensities to fibrillate: inhibition of α -synuclein assembly by β - and γ -synucleins. *J Biol Chem* 277: 11970–11978.
 80. Qu Y, Bolen CL, Bolen DW (1998) Osmolyte-driven contraction of a random coil protein. *Proc Natl Acad Sci USA* 95:9268–9273.
 81. Celinski SA, Scholtz JM (2002) Osmolyte effects on helix formation in peptides and the stability of coiled-coils. *Protein Sci* 11:2048–2051.
 82. Bourhis JM, Johansson K, Receveur-Brechot V, Oldfield CJ, Dunker KA, Canard B, Longhi S (2004) The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon binding to their physiological partner. *Virus Res* 99:157–167.
 83. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK (2005) Coupled folding and binding with α -helix-forming molecular recognition elements. *Biochemistry* 44:12454–12470.
 84. Fuxreiter M, Simon I, Friedrich P, Tompa P (2004) Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J Mol Biol* 338:1015–1026.
 85. Gsponer J, Christodoulou J, Cavalli A, Bui JM, Richter B, Dobson CM, Vendruscolo M (2008) A coupled equilibrium shift mechanism in calmodulin-mediated signal transduction. *Structure* 16:736–746.
 86. Dyson HJ, Wright PE (2002) Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* 12:54–60.
 87. Schanda P, Van Melckebeke H, Brutscher B (2006) Speeding up three-dimensional protein NMR experiments to a few minutes. *J Am Chem Soc* 128: 9042–9043.
 88. Lescop E, Schanda P, Brutscher B (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J Magn Reson* 187: 163–169.
 89. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277–293.
 90. Goddard TD, Kneller DG (2003) SPARKY 3. San Francisco: University of California.
 91. Jung YS, Zweckstetter M (2004) Mars—robust automatic backbone assignment of proteins. *J Biomol NMR* 30:11–23.
 92. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* 33: 5984–6003.
 93. Wyatt PJ (1998) Submicrometer particle sizing by multiangle light scattering following fractionation. *J Colloid Interface Sci* 197:9–20.
 94. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) PRIMUS: a windows PC-based system for small-angle scattering data analysis. *J Appl Cryst* 36:1277–1282.
 95. Semenyuk AV, Svergun D (1991) GNOM—a program package for small-angle scattering data processing. *J Appl Cryst* 24:537–540.
 96. Eyal E, Najmanovich R, McConkey BJ, Edelman M, Sobolev V (2004) Importance of solvent accessibility and contact surfaces in modeling side-chain conformations in proteins. *J Comput Chem* 25: 712–724.