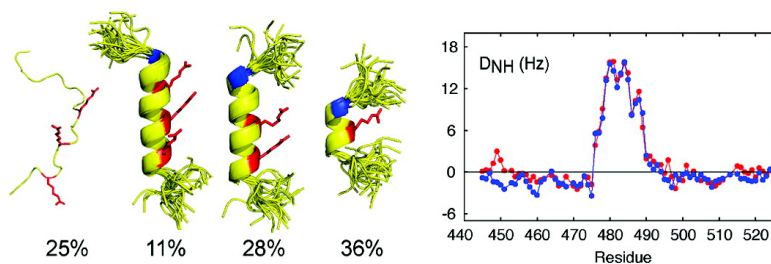Article

Quantitative Conformational Analysis of Partially Folded Proteins from Residual Dipolar Couplings: Application to the Molecular Recognition Element of Sendai Virus Nucleoprotein

Malene Ringkjøbing Jensen, Klaartje Houben, Ewen Lescop,
Laurence Blanchard, Rob W. H. Ruigrok, and Martin Blackledge

More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 1 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

Quantitative Conformational Analysis of Partially Folded Proteins from Residual Dipolar Couplings: Application to the Molecular Recognition Element of Sendai Virus Nucleoprotein

Malene Ringkjøbing Jensen,[†] Klaartje Houben,[†,§] Ewen Lescop,[†,‖]
Laurence Blanchard,[†,⊥] Rob W. H. Ruigrok,[‡] and Martin Blackledge*,[†]

Protein Dynamics and Flexibility, Institut de Biologie Structurale Jean-Pierre Ebel, CEA, CNRS, UJF UMR 5075, 41 Rue Jules Horowitz, Grenoble 38027, France, Unit for Virus Host Cell Interactions, UJF-EMBL-CNRS UMR 5233, 6 rue Jules Horowitz, BP 181 38042 Grenoble cedex 9, France, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands, CNRS, Institut de Chimie des Substances Naturelles, Laboratoire de Chimie et Biologie Structurales, 1 avenue de la terrasse, 91190 Gif sur Yvette, France, and CEA Cadarache, DSV/IBEB/SBVME/LEMiRE, UMR 6191 CNRS/CEA/Univ Aix-Marseille, Saint-Paul-lez-Durance F-13108, France

Received February 22, 2008; E-mail: martin.blackledge@ibs.fr

Abstract: A significant fraction of proteins coded in the human proteome do not fold into stable three-dimensional structures but are either partially or completely unfolded. A key feature of this family of proteins is their proposed capacity to undergo a disorder-to-order transition upon interaction with a physiological partner. The mechanisms governing protein folding upon interaction, in particular the extent to which recognition elements are preconfigured prior to formation of molecular complexes, can prove difficult to resolve in highly flexible systems. Here, we develop a conformational model of this type of protein, using an explicit description of the unfolded state, specifically modified to allow for the presence of transient secondary structure, and combining this with extensive measurement of residual dipolar couplings throughout the chain. This combination of techniques allows us to quantitatively analyze the level and nature of helical sampling present in the interaction site of the partially folded C-terminal domain of Sendai virus nucleoprotein (N$_{TAIL}$). Rather than fraying randomly, the molecular recognition element of N$_{TAIL}$ preferentially populates three specific overlapping helical conformers, each stabilized by an N-capping interaction. The unfolded strands adjacent to the helix are thereby projected in the direction of the partner protein, identifying a mechanism by which they could achieve nonspecific encounter interactions prior to binding. This study provides experimental evidence for the molecular basis of helix formation in partially folded peptide chains, carrying clear implications for understanding early steps of protein folding.

Introduction

A broad category of proteins (around 35% of the human proteome) do not fold into stable three-dimensional structures, but are either fully unfolded or contain unfolded regions of significant length.[1–3] These proteins play key roles in many physiological and pathological processes, including signaling, cell cycle control, molecular recognition, transcription and replication, as well as endocytosis and amyloidogenesis.[4,5] The conformational and thermodynamic properties of unfolded proteins also provide important information for understanding the molecular basis of protein folding and stability.[6–8] Despite the importance of characterizing this class of proteins, standard, single conformer-based approaches to structure determination necessarily fail to adequately describe such highly flexible systems.[9,10] It has therefore become essential to develop new tools for characterizing their rapidly fluctuating conformational behavior.

NMR is very well adapted to deal with this question, reporting on ensemble-averaged parameters that can be interpreted in terms of a continuum of interchanging conformations.[11–13] In

[†] Institut de Biologie Structurale Jean-Pierre Ebel.
[‡] UVHCI.
[§] Utrecht University.
[‖] CNRS, Institut de Chimie des Substances Naturelles.
[⊥] CEA Cadarache.
(1) Tompa, P. *Trends Biochem. Sci.* **2002**, *27*, 527–533.
(2) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z. *Biochemistry* **2002**, *41*, 6573–6582.
(3) Uversky, V. N. *Protein Sci.* **2002**, *11*, 739–756.
(4) Fink, A. L. *Curr. Opin. Struct. Biol.* **2005**, *15*, 35–41.
(5) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208.

(6) Daggett, V.; Fersht, A. R. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 497–502.
(7) Vendruscolo, M.; Paci, E.; Karplus, M.; Dobson, C. M. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 14817–14821.
(8) Dobson, C. M. *Nature* **2003**, *426*, 884–890.
(9) Mittag, T.; Forman-Kay, J. D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3–14.
(10) Vendruscolo, M. *Curr. Opin. Struct. Biol.* **2007**, *17*, 15–20.
(11) Smith, L. J.; Bolin, K. A.; Schwalbe, H.; MacArthur, M. W.; Thornton, J. M.; Dobson, C. M. *J. Mol. Biol.* **1996**, *255*, 494–506.
(12) Choy, W.-Y.; Mulder, F. A. A.; Crowhurst, K. A.; Muhandiram, D. R.; Millett, I. S.; Doniach, S.; Forman-Kay, J. D.; Kay, L. E. *J. Mol. Biol.* **2002**, *316*, 101–112.

particular, residual dipolar couplings (RDC)[14,15] have been shown to provide detailed information about structural propensities in the unfolded state at amino acid specific resolution.[16–20] Using a multiconformational ensemble approach based on the sampling of backbone dihedral angles from an amino acid specific structural database, inherent conformational properties and propensity for local structure have recently been identified in both intrinsically unfolded and chemically denatured proteins.[21–25]

A key feature of the complex relationship between structural dynamics and biological function in intrinsically unfolded proteins is the observed capacity of some members of this family to undergo a disorder-to-order transition on interaction with a physiological partner.[2,5,26,27] In these systems, representing a significant fraction of the proteome,[28] molecular recognition is often accompanied by local folding into a characteristic three-dimensional conformation. Understanding the physical basis of induced folding upon binding requires an accurate description of the conformational behavior of the prerecognition, free form of the protein. Recently, appreciable effort has been invested in the prediction of molecular recognition elements on the basis of primary sequence, using database mining based on predictors of disorder, secondary structure, and hydrophobic clusters and on the analysis of structure of the interaction sites in molecular complexes. [29–32]

The inherent flexibility of natively unfolded proteins has severely hindered the detailed experimental characterization of the prerecognition state. Nevertheless, the dynamics of polypeptide folding upon interaction have recently been studied using rotating frame relaxation, identifying the formation of initial encounter complexes via weak, nonspecific interactions that facilitate the formation of a partially folded state upon binding.[33] Such observations support the previously proposed "fly casting" mechanism that provides a theoretical framework for speeding up molecular recognition processes via the folding funnel.[34] Our goal is to develop a conformational model for partially folded proteins, using a recently developed explicit representation of the unfolded state,[22] specifically modified to allow for the presence of transiently formed secondary structure, and combining this with extensive measurement of RDCs to validate the ensemble description.

In this study, we have studied the conformational behavior of partially unfolded N_{TAIL}, the C-terminal domain of the nucleoprotein (N) of the Sendai virus (SeV), a paramyxovirus that causes bronchiolitis in mice and primates.[35] The RNA polymerase complex for replication and transcription of viral RNA in SeV is composed of protein L and the phosphoprotein P, a tetrameric modular protein comprising both folded and unfolded domains.[36,37] The matrix for this polymerase complex is a helical structure consisting of the viral RNA and multiple copies of the viral N protein.[38] Positioning of the polymerase requires an apparently dynamic interaction between the unstructured C-terminal tail of N (N_{TAIL}) and PX,[39] the C-terminal domain of P comprising three parallel helices and a long disordered strand, via which it is attached to the tetramerization domain.[22,40,41] Similar observations have been made concerning measles virus, another member of the *Paramyxoviridae* family.[42,43] The molecular recognition site of SeV N_{TAIL} appears as a nascently structured α-helix in free solution, further folding upon interaction with the C-terminal domain of P. The N_{TAIL} helix was proposed to orient such that the positively charged site of the N_{TAIL} helix interacts with a negative patch on the PX structure.[44] Our aim in this study was to quantitatively define the nature and extent of the α-helical sampling in the unbound form of N_{TAIL}.

Results

Local Structure in the Molecular Recognition Element of N_{TAIL}. The sequence of the C-terminal domain of Sendai virus nucleoprotein, N_{TAIL} (comprising amino acids 443−524), was submitted to the program AGADIR,[45] to estimate the level of helical propensity along the chain. This resulted in prediction

(13) Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. *Science* **2002**, *295*, 1719–1722.
(14) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111–1114.
(15) Prestegard, J. H.; al-Hashimi, H. M.; Tolman, J. R. *Q. Rev. Biophys.* **2000**, *33*, 371–424.
(16) Alexandrescu, A. T.; Kammerer, R. A. *Protein Sci.* **2003**, *12*, 2132–2140.
(17) Louhivuori, M.; Pääkkönen, K.; Fredriksson, K.; Permi, P.; Lounila, J.; Annila, A. *J. Am. Chem. Soc.* **2003**, *125*, 15647–15650.
(18) Mohana-Borges, R.; Goto, N. K.; Kroon, G. J. A.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2004**, *340*, 1131–1142.
(19) Fieber, W.; Kristjansdottir, S.; Poulsen, F. M. *J. Mol. Biol.* **2004**, *339*, 1191–1199.
(20) Meier, S.; Güthe, S.; Kiefhaber, T.; Grzesiek, S. *J. Mol. Biol.* **2004**, *344*, 1051.
(21) Jha, A. K.; Colubri, A.; Freed, K.; Sosnick, T. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099–13104.
(22) Bernado, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W. H.; Blackledge, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17002–17007.
(23) Bernado, P.; Bertoncini, C.; Griesinger, C.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2005**, *127*, 17968–17969.
(24) Mukrasch, M. D.; Markwick, P. R. L.; Biernat, J.; von Bergen, M.; Bernado, P.; Griesinger, C.; Mandelkow, E.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 5235–5243.
(25) Meier, S.; Grzesiek, S.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 9799–9807.
(26) Demchenko, A. P. *J. Mol. Recognit.* **2001**, *14*, 42–61.
(27) Dyson, H. J.; Wright, P. E. *Curr. Opin. Struct. Biol.* **2002**, *12*, 54–60.
(28) Aloy, P.; Russell, R. B. *Nat. Biotechnol.* **2004**, *22*, 1317–1321.
(29) Fuxreiter, M.; Simon, I.; Friedrich, P.; Tompa, P. *J. Mol. Biol.* **2004**, *338*, 1015–1026.
(30) Vucetic, S.; Obradovic, Z.; Vacic, V.; Radivojac, P.; Peng, K.; Iakoucheva, L. M.; Cortese, M. S.; Lawson, J. D.; Brown, C. J.; Sikes, J. G.; Newton, C. D.; Dunker, A. K. *Bioinformatics* **2005**, *21*, 137–140.
(31) Vacic, V.; Oldfield, C. J.; Mohan, A.; Radivojac, P.; Cortese, M. S.; Uversky, V. N.; Dunker, A. K. *J. Proteome Res.* **2007**, *6*, 2351–2366.
(32) Tompa, P.; Fuxreiter, M. *Trends Biochem. Sci.* **2008**, *33*, 2–8.

(33) Sugase, K.; Dyson, H. J.; Wright, P. E. *Nature* **2007**, *447*, 1021–1025.
(34) Levy, Y.; Cho, S. S.; Onuchic, J. N.; Wolynes, P. G. *J. Mol. Biol.* **2005**, *346*, 1121–1145.
(35) Skiadopoulos, M. H.; Surman, S. R.; Riggs, J. M.; Elkins, W. R.; St. Claire, M.; Nishio, M.; Garcin, D.; Kolakofsky, D.; Collins, P. L.; Murphy, B. R. *Virology* **2002**, *297*, 153–160.
(36) Tarbouriech, N.; Curran, J.; Ebel, C.; Ruigrok, R. W.; Burmeister, W. P. *Virology* **2000**, *266*, 99–109.
(37) Tarbouriech, N.; Curran, J.; Ruigrok, R. W.; Burmeister, W. P. *Nat. Struct. Biol.* **2000**, *7*, 777–781.
(38) Arnheiter, H.; Davis, N. L.; Wertz, G.; Schubert, M.; Lazzarini, R. A. *Cell* **1985**, *41*, 259–267.
(39) Kolakofsky, D.; Le Mercier, P.; Iseni, F.; Garcin, D. *Virology* **2004**, *318*, 463–473.
(40) Blanchard, L.; Tabouriech, N.; Blackledge, M.; Timmins, P.; Burmeister, W. P.; Ruigrok, R. W. H.; Marion, D. *Virology* **2004**, *319*, 201–211.
(41) Houben, K.; Blanchard, L.; Blackledge, M.; Marion, D. *Biophys. J.* **2007**, *93*, 2830–2844.
(42) Kingston, R. L.; Hamel, D. J.; Gay, L. S.; Dahlquist, F. W.; Matthews, B. W. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8301–8306.
(43) Bourhis, J. M.; Canard, B.; Longhi, S. *Virology* **2006**, *344*, 94–110.
(44) Houben, K.; Marion, D.; Tarbouriech, N.; Ruigrok, R. W. H.; Blanchard, L. *J. Virol.* **2007**, *81*, 6807–6816.
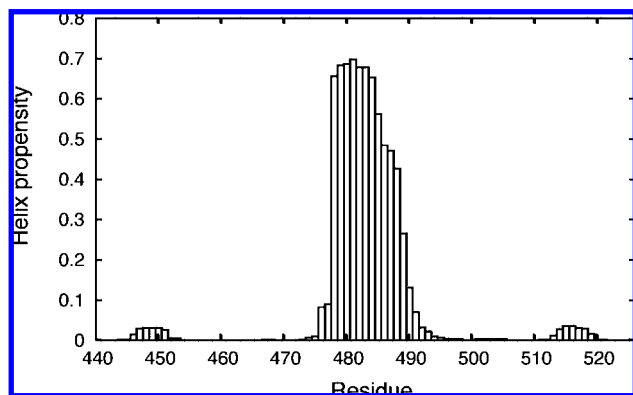(45) Lacroix, E.; Viguera, A. R.; Serrano, L. *J. Mol. Biol.* **1998**, *284*, 173–191.

Figure 1. Helical propensities present in N_{TAIL}. Helical propensity determined on the basis of the physical chemistry of the sequence using the program AGADIR.

of two regions with weak propensity at the N and C termini, and a central region (476−492) where propensities are predicted up to 70% (Figure 1). Analysis of secondary $^{13}C^\alpha$ chemical shifts revealed a similar pattern (Figure S1, Supporting Information). The central region (478−490) has previously been shown to interact with PX via electrostatic interactions involving a series of arginines situated on one side of the helical domain.[44] To investigate the conformational behavior of this partially formed molecular recognition element in more detail, $^1D_{HN}$, $^1D_{C\alpha H\alpha}$, $^1D_{C\alpha C'}$, and $^2D_{HNC'}$ RDCs were measured from N_{TAIL} aligned in liquid crystalline ethylene glycol/alcohol phase (Figure 2). $^1D_{HN}$ RDCs from the majority of the sequence follow the classically observed inverted bell-shaped distribution along the primary sequence, indicative of an unfolded peptide chain.[17] Additional structure is apparent, principally in the central region (476−493), but also in the two regions with low helical propensities as identified from AGADIR. In the central region, strongly positive $^1D_{HN}$ RDCs were measured, suggesting the presence of a helical motif. Positive $^1D_{HN}$ couplings can be understood in terms of the average orientation of the NH vectors within a helix being mainly aligned with the static magnetic field.[17,18,20] The origin of the periodicity of these $^1D_{HN}$ couplings, in addition to that present in the $^1D_{C\alpha H\alpha}$, and to some extent $^1D_{C\alpha C'}$ and $^2D_{HNC'}$ couplings, is less intuitively clear. Assuming that the helix is not deformed, such a periodicity can only occur if the effective orientation of the vectors on either side of the helix differs relative to the magnetic field, resulting in an effective tilt of the main axis of the helical element with respect to this axis.

To provide further insight into these observations, we used Flexible-Meccano,[22] an explicit ensemble molecular description, applicable to the interpretation of RDCs measured in dynamically fluctuating systems. Using this approach, we have recently shown that RDCs from partially folded and unfolded proteins can be predicted by randomly sampling distinct amino acid specific ϕ/ψ propensities to construct a large number of conformers.[23−25] RDCs are then predicted from the ensemble by averaging values predicted for each structure.

Systematic Analysis of RDCs Identifies Three Significantly Populated Helices. To characterize the conformational behavior of the central helical segment in N_{TAIL}, we have systematically investigated possible combinations and populations of continuous helical segments, from a minimum of 4 amino acids to a maximum of 20, from throughout the molecular recognition element segment (defined over the range 476−495, see Meth-
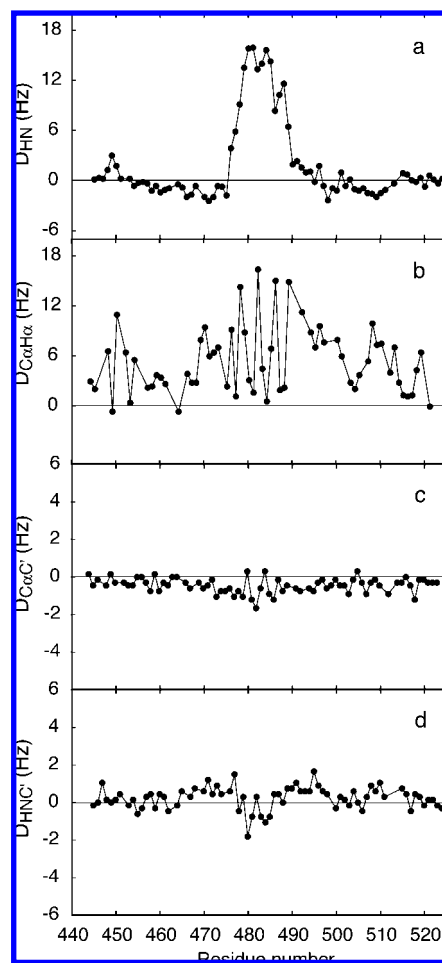

Figure 2. Residual dipolar couplings from N_{TAIL}. (a) $^1D_{HN}$, (b) $^1D_{C\alpha H\alpha}$, (c) $^1D_{C\alpha C'}$, and (d) $^2D_{HNC'}$ RDCs were measured from throughout the backbone of the protein aligned in liquid crystalline ethylene glycol/alcohol phase.

ods). Ensemble equilibria comprising combinations of increasing numbers of conformers ($n = 0, 1, 2, 3, 4$) of the 153 helical conformations were compared to the experimental data. In each case, the population of each member of the ensemble was optimized. Three parameters were thus optimized per helix (comprising the limits of each helix and its associated population) and one common scaling parameter per ensemble. Knowledge of these parameters allows a rigorous analysis of statistical significance of the improved data reproduction with increasing numbers of helical models present in the ensemble. Best fitting combinations are summarized in Table 1 and illustrated in Figure 3 for the $^1D_{HN}$ RDCs. A systematic and statistically significant decrease of the fitting function is found with increasing numbers of helical conformers until $n = 3$, beyond which no significant improvement is found. The best fitting combination found for $n = 2$ helices selects 476−488 and 479−484, representing a significant improvement as compared to the $n = 1$ model ($p < 0.0001$). This and the next best fitting $n = 2$ solution (478−492 and 479−484) together define the three conformers selected for the $n = 3$ solution. The $n = 3$ solution appears to be robust, again exhibiting a highly significant improvement over the $n = 2$ solution ($p < 0.0001$). The next-best fitting combinations for $n = 3$ are identical to the best fitting solution, except for the details of the termination of the least populated helix (see Table S1, Supporting Information, for details of less well fitting

Table 1. Data Reproduction from Ensembles with Different Numbers of Helical Conformers

number of helical conformers	$\chi^{2a}/\Delta_f{}^b$	helical conformers[c]	% population[d]	significance[e]
1	427/96	476−488	52	
2	227/93	476−488, 479−484	37, 45	$p < 0.0001$
3	120/90	476−488, 479−484, 478−492	28, 36, 11	$p < 0.0001$
4	113/87	476−488, 479−484, 478−495, 479−492	31, 24, 6, 5	$p = 0.16$

[a] Target function measured over all 100 RDCs. [b] Number of degrees of freedom in the fit (number of couplings − number of optimized parameters). One helix implies the optimization of three parameters, starting amino acid, final amino acid, and the population. [c] Range of the invoked helices. [d] Population of the invoked helices. The remaining conformers are unfolded. [e] Significance of the improvement of this model as compared to the simpler model. Calculated using a standard F-test.
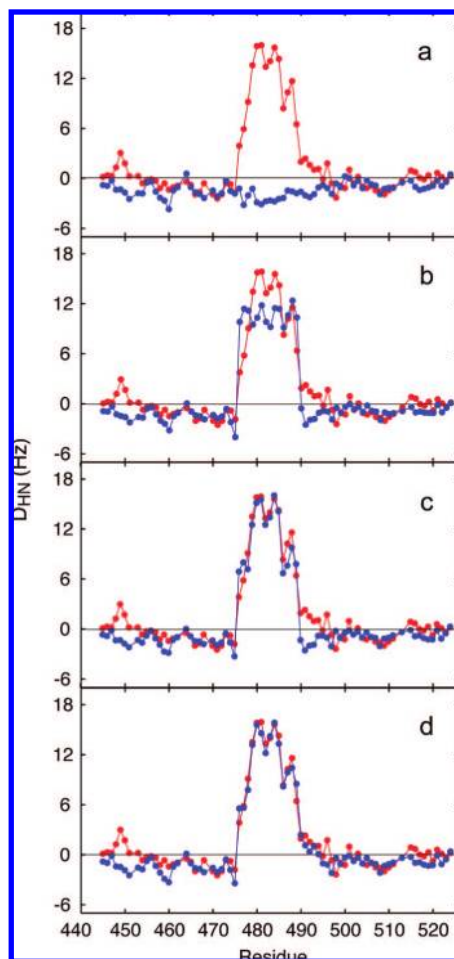

Figure 3. Reproduction of $^1D_{HN}$ RDCs using Flexible-Meccano statistical coil/explicit helix conformational ensembles. Simulations b–d (blue) were scaled by a common factor to best reproduce the experimental data (red), and given χ^2 values are summed over all four RDC types in the range 472−498. (a) No explicit helical elements. (b) One single helix (four fitting parameters – see Methods). The best fitting helix stretches from 476−488 and has an optimal population of 52%, with the remaining 48% sampling unfolded, statistical coil conformations. The χ^2 for this model is 427. (c) Two helical elements (seven fitting parameters – see Methods). The best fitting combinations of two helices comprise 476−488 and 479−484. These combinations have optimal populations of 37% and 45% and χ^2 of 227. (d) Three helical elements (10 fitting parameters – see Methods). The best fitting helices stretch from 476−488, 479−484, and 478−492 with optimal populations of 28 ± 1%, 36 ± 3%, and 11 ± 1%, with the remaining 25 ± 4% sampling unfolded, statistical coil conformations. The χ^2 for this model is 120.

combinations). Selection of 4 helical conformers provides no further improvement. The range and distribution of RDCs along the sequence, including the periodic nature of certain couplings, are reproduced throughout the helical region within acceptable
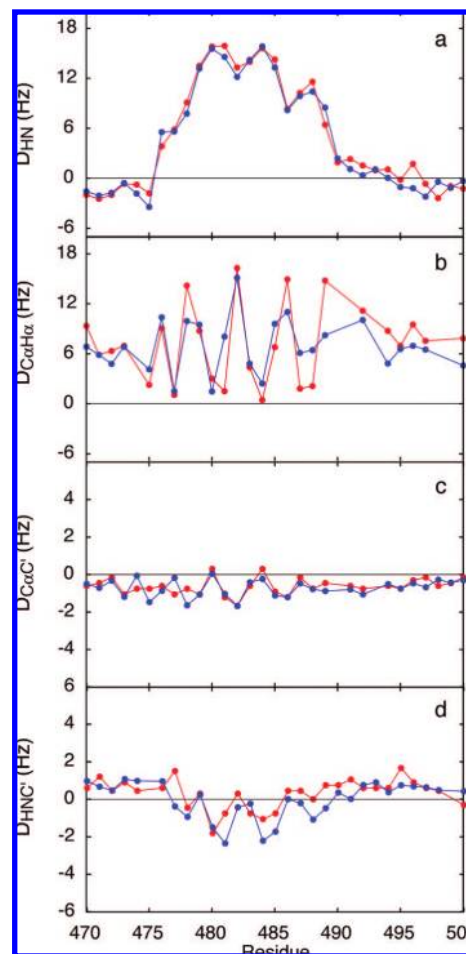

Figure 4. Reproduction of all four RDCs in the helical region of N_{TAIL}. Comparison of (a) $^1D_{HN}$, (b) $^1D_{C\alpha H\alpha}$, (c) $^1D_{C\alpha C'}$, and (d) $^2D_{HNC'}$ experimental RDCs (red) with associated simulated values (blue) using the optimal three-conformer model presented in Figure 3d. All simulations were scaled by the same factor to best reproduce the experimental data.

limits of experimental uncertainty (Figure 4). The three specific conformers sampled by N_{TAIL}, H1 (479−484), H2 (476−488), and H3 (478−492), are populated to levels estimated at 36%, 28%, and 11%, respectively, with the remaining 25% being unfolded (Figure 5).

N_{TAIL} Preferentially Samples Specific N-Capped Helical Conformers in Solution. This analysis reveals that the molecular recognition element of N_{TAIL} cannot be represented by a single helical element in equilibrium with an unfolded form, but that specific helical elements that overlap in the sequence are required to describe the ensemble, of which two are populated to over 25%.

Examination of the primary sequence of the helices shows that one of the arginines involved in molecular recognition
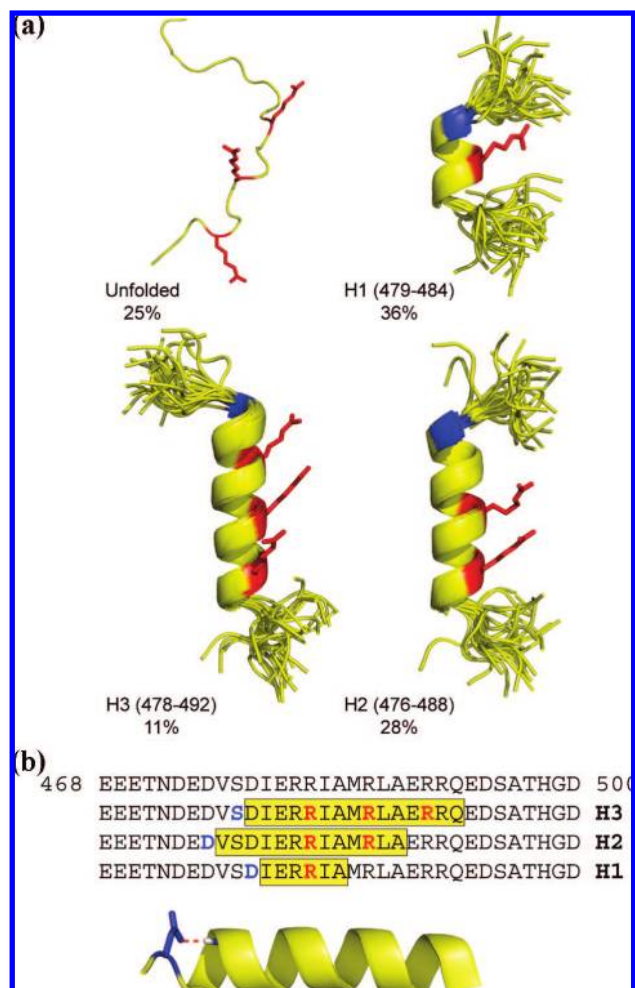
Figure 5. Molecular representation of the proposed conformational equilibrium of N_{TAIL} in solution. (a) The four conformations are presented as: a single structure, representing the $25 \pm 4\%$ unfolded conformers, the shortest helical element, comprising 6 amino acids 479–484, populated at a level of $36 \pm 3\%$, 476–488 populated at $28 \pm 1\%$, and a longer stretch 478–492 populated to a level of $11 \pm 1\%$. The molecular recognition site arginines are shown in red. Twenty randomly selected conformers are shown for each of the helical segments to illustrate the directionality of the adjacent chains projected from the helix caps. (b) Primary sequence of the central region of N_{TAIL} showing the positions of the arginines responsible for the interaction with the phosphoprotein (red) and the N-capping amino acids (blue). The cartoon figure illustrates an N-capping Asp side chain–backbone interaction.

(R482) is at the center of the shortest element (H1); similarly H2 comprises two of the arginines involved in the interaction at its center on two consecutive helical turns (R482, R486), with H3 additionally containing four arginines, three of which are positioned on the same side of the helix (R482, R486, and R490). The presence of these arginines at the center of the helical motifs may optimize accessibility of the side chains, thereby optimizing conditions for electrostatic interaction with the negatively charged cleft on the surface of the phosphoprotein PX domain, as well as maximizing structural stability of the side chains within the flexible chain due to the restricted rotameric propensities of arginine side chains within helical elements.[46]

It is particularly notable that all three of the helical segments identified from the RDC analysis are preceded by an N-capping

(46) Dunbrack, R. L.; Karplus, M. *Nat. Struct. Biol.* **1994**, *1*, 334–340.

amino acid (Figure 5b). In the case of the two most populated helices, H1 and H2, the N-capping amino acids are aspartic acids (D478 and D475), while in the case of H3 this is a serine (S477). Aspartic acid and serine are the most prevalently found helix capping residues in high-resolution folded structural databases[47] (and Figure S2, Supporting Information), with the capping amino acid side chain forming a hydrogen bond with backbone amides at positions 2 or 3 in the helix (Figure 5b).[48–50] Site-directed mutagenesis has recognized the importance of N-capping residues in nucleating and stabilizing helical structures,[51,52] and the observation that the populated helices in N_{TAIL} are all nucleated by stabilizing N-capping motifs suggests a natural mechanism designed to promote helical populations. This mechanism alleviates the problem of establishing and maintaining helical structure in a highly flexible protein and reduces the energetic and entropic cost of folding from a completely unfolded chain. Preformation of the helix will also create an electrostatic interaction surface that will reduce the effective conformational space to be searched for interaction partners on the partner protein PX.

Specific N-Capping Projects the Unfolded Chain toward the Binding Partner. It is of significant interest that the two most populated conformers (H1 and H2) differ by one helical turn at both ends. This has the consequence of projecting the unfolded chain at the ends of the helix in a specific and conserved direction in the majority of the helical conformers. This projection does not point away from the partner in the bound complex, as defined by the position of the arginines on one side of the helix, but rather in the same overall direction as the arginine side chains (Figure 5). This observation is only valid for the first few amino acids preceding and following the helix, beyond which the statistical coil sampling of the chain will result in very diffuse directionality, but may be important in terms of nonspecific interactions away from the binding site, which facilitate formation of the complex.[34]

Discussion

The molecular mechanisms governing the folding of intrinsically disordered proteins upon interaction remain largely unknown. In this respect, the extent to which regions of the protein that contribute to binding and function are preconfigured prior to interaction remains an important question that can prove difficult to resolve in highly flexible systems. Here, we have studied the structural properties of the partially ordered recognition element of N_{TAIL} from Sendai virus, whose interaction site exhibits helical propensity, but cannot be described by a single conformational state. We show that despite broad conformational heterogeneity, analysis of extensive backbone RDCs in terms of an ensemble of interconverting structures allows remarkable insight into the structural details of the conformational ensemble. In particular, we demonstrate that rather than fraying in a random fashion, the molecular recognition sequence of N_{TAIL} preferentially populates three specific helical conformers. The most populated conformers differ in length by approximately one

(47) Lovell, S. C.; Davis, I. W.; Arendall, W. B., III; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. *Proteins* **2003**, *50*, 437–450.

(48) Richardson, J. S.; Richardson, D. C. *Science* **1988**, *240*, 1648–1652.

(49) Doig, A. J.; Macarthur, M. W.; Stapley, B. J.; Thornton, J. M. *Protein Sci.* **1997**, *6*, 147–155.

(50) Aurora, R.; Rose, G. D. *Protein Sci.* **1998**, *7*, 21–38.

(51) Serrano, L.; Fersht, A. R. *Nature* **1989**, *342*, 296–299.

(52) Serrano, L.; Sancho, J.; Hirshberg, M.; Fersht, A. R. *J. Mol. Biol.* **1992**, *227*, 544–559.

helical turn at both termini and enclose the recognition site amino acids. The limits of the populated helical segments exhibit commonly observed N-capping motifs, suggesting that the preferred conformations are stabilized by these motifs and that these selected conformers are encoded in the primary sequence of the protein.

The directionality of the unfolded strand adjacent to the helix is strongly correlated to the capping position, indicating that the termination of the helices selectively controls for specific directionality in the unfolded strands immediately adjacent to the helical elements. The tendency to project in the direction of the partner protein can be rationalized on the basis of the possible existence of short-lived and nonspecific encounter complexes mediated by amino acids outside the molecular recognition element. Our understanding of the mechanisms governing protein folding upon interaction has been greatly enhanced by theoretical prediction,[34] and more recent observation,[33] of weakly binding, nonspecific encounter complexes formed between partially folded proteins and their partners. Although we currently have no further experimental evidence for the existence of such interactions in this particular system, we appear to have identified a mechanism by which the partially folded form of the protein could project the adjacent region of the unfolded strand in the most appropriate direction to achieve such "fly casting" interactions.

This work adds to our understanding of the interaction of the SeV PL–polymerase complex with the viral N–RNA complex. On the side of the tetrameric phosphoprotein, the surface interaction with N_{TAIL} is formed by the negatively charged surface of a three-helical bundle, PX. The K_D of the interaction of N_{TAIL} and PX is about 60 μM.[44] This relatively weak local affinity would suggest that at any given time the polymerase complex will have to ensure two or three PX–N_{TAIL} interactions to remain associated with the N–RNA matrix. For the polymerase to move forward during transcription or replication, PX–N_{TAIL} interactions need to be made and broken very rapidly. The combined requirements are therefore for a strong and dynamic affinity between PX and N_{TAIL} that would be favored by the proposed "fly casting" process, promoting fast interaction of the projected positively charged amino acids of N_{TAIL} with the negative surface of PX.

More generally, this study provides further experimental evidence for the molecular basis of nascent helix formation in partially folded chains and identifies mechanisms by which the primary sequence can be exploited to control these events. These observations have clear implications for our understanding of the early steps of protein folding.

Methods

Sample Preparation. SeV N_{TAIL} (comprising the sequence 443−524) was expressed and purified as previously described.[44] The NMR samples were prepared at a concentration of 0.3 mM in 50 mM potassium phosphate buffer at pH 6.0 with 50 mM NaCl, 10% D_2O, and a protease inhibitor cocktail (Complete; Bohringer Mannheim). Alignment of the protein was achieved with poly ethylene glycol (PEG C12E5, Sigma) and 1-hexanol (residual ^{2}H splitting of 21 Hz).[53]

NMR Spectroscopy. Chemical shift assignment of a shorter construct of SeV N_{TAIL} (N_{TAIL}(443−501))[54] was extended to assign the longer construct (N_{TAIL}(443−524)) studied here using standard

triple resonance experiments. $^1D_{HN}$, $^1D_{C\alpha H\alpha}$, $^1D_{C\alpha C'}$, and $^2D_{HNC'}$ residual dipolar couplings were measured at 298 K using BEST-type[55,56] 3D HNCO- and HN(CO)CA experiments modified to allow for spin-coupling measurements in the ^{13}C dimension.[57,58] NMR spectra were processed in NMRPipe,[59] and peak splittings were extracted using the program Sparky.[60] Uncertainties in experimental dipolar couplings were assessed using repeat measurements and were estimated as 1 Hz for $^1D_{HN}$, 1.5 Hz for $^1D_{C\alpha H\alpha}$ (3 Hz for some overlapped peaks), 0.5 Hz for $^1D_{C\alpha C'}$, and 0.75 Hz for $^2D_{HNC'}$.

Flexible-Meccano Calculations. Flexible-Meccano (FM) uses a Monte-Carlo sampling technique based on amino-acid propensity and side chain volume. The details of the algorithm have been presented elsewhere.[22] Alignment tensors and RDCs for each conformer were calculated using PALES,[61] an atomic resolution approach to alignment tensor prediction. Specific N-capping database conformational potentials were extracted for Asp and Ser residues and incorporated into the Flexible-Meccano protocol in a standard way.

FM calculations were also performed using specified ϕ/ψ distributions in addition to the standard residue-dependent distributions. Gaussian distributions with standard deviation of 3° centered on α-helical conformations for ϕ and ψ {−65°,−40°} were randomly introduced in a cooperative manner. For each helix length, an entire FM simulation of 10 000 conformers was carried out.

Helices of different lengths (from 4−20 amino acids) were explicitly incorporated into the statistical coil approach, thereby creating 153 different ensembles, each populated at 100% of the specified helix. RDCs predicted from these ensembles were then used to identify the most populated helical conformers by comparing experimental data to simulated RDCs from combinations of ensembles from unfolded and partially helical chains. All possible helical segments were systematically combined, and their populations were optimized to best reproduce the experimental data using the following relationships:

$$D_{ij,\text{eff}} = \sum_{k=1,n} p_k D_{ij}^k + \left(1 - \sum_{k=1,n} p_k\right) D_{ij}^U \qquad (1)$$

where p_k are the populations of the n helical conformers, for which D_{ij}^k are the individual predicted couplings between nuclei i and j, and D_{ij}^U are the couplings from the unfolded state. These effective couplings are compared to experimental data using the expression:

$$\chi^2 = \sum (D_{ij,\text{eff}} - D_{ij,\text{exp}})^2 / \sigma_{ij}^2 \qquad (2)$$

where σ represents the uncertainty on the experimental coupling values. The absolute level of alignment is not known, requiring an additional scaling of all averaged couplings. The total number of optimized parameters comprises three per helix (beginning, end, and population) and one common scaling parameter. Standard statistical F-tests were applied to test the significance in the improvement in data fitting. Uncertainties in populations were estimated from experimental noise based Monte Carlo simulations. This kind of fitting procedure bears some resemblance to approaches developed by Forman-Kay and co-workers to develop ensemble descriptions that are in agreement with experimental data from

(53) Ruckert, M.; Otting, G. *J. Am. Chem. Soc.* **2000**, *122*, 7793−7797.
(54) Marion, D.; Tarbouriech, N.; Ruigrok, R. W. H.; Burmeister, W. P.; Blanchard, L. *J. Biomol. NMR* **2001**, *21*, 75−76.

(55) Schanda, P.; Van Melckebeke, H.; Brutscher, B. *J. Am. Chem. Soc.* **2006**, *128*, 9042−9043.
(56) Lescop, E.; Schanda, P.; Brutscher, B. *J. Magn. Reson.* **2007**, *187*, 163−169.
(57) Brutscher, B. *J. Magn. Reson.* **2001**, *151*, 332−338.
(58) Champier, L.; Sibille, N.; Bersch, B.; Brutscher, B.; Blackledge, M.; Coves, J. *Biochemistry* **2002**, *41*, 3770−3780.
(59) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277−293.
(60) Goddard, T. D.; Kneller, D. G. University of California, 2003.
(61) Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791−3792.

partially folded or unfolded proteins where population weights are assigned to pregenerated conformers to calculate ensembles of structures.[62]

(62) Marsh, J. A.; Neale, C.; Jack, F. E.; Choy, W.-Y.; Lee, A. Y.; Crowhurst, K. A.; Forman-Kay, J. D. *J. Mol. Biol.* **2007**, *367*, 1494–1510.

Supporting Information Available: Two figures showing the $^{13}C_\alpha$ NMR secondary chemical shifts and the propensity of different amino acids to be in the N-capping position in folded structures derived from a database of high-resolution crystallographic structures. One table showing data reproduction from ensembles with different numbers of helical conformers. This material is available free of charge via the Internet at http://pubs.acs.org.

JA801332D