DOI: 10.1002/ange.201309127

Protein Dynamics

# Investigation of the Structure and Dynamics of the Capsid–Spacer Peptide 1–Nucleocapsid Fragment of the HIV-1 Gag Polyprotein by Solution NMR Spectroscopy**

*Lalit Deshmukh, Rodolfo Ghirlando, and G. Marius Clore**

***Abstract:*** *Structural studies of HIV-1 Gag, the primary structural polyprotein involved in retroviral assembly, have been challenging, owing to its flexibility and conformational heterogeneity. Using residual dipolar couplings, we show that the four structural units of the capsid (CA)–spacer peptide 1 (SP1)–nucleocapsid (NC) fragment of HIV-1 Gag (namely, the N- and C-terminal domains of capsid, and the N- and C-terminal Zn knuckles of nucleocapsid) have the same structures as their individually isolated counterparts, and tumble semi-independently of one another in the absence of nucleic acids. Nucleic acids bind exclusively to the nucleocapsid domain and fix the orientation of the two Zn knuckles relative to one another so that the nucleocapsid domain/nucleic acid complex behaves as a single structural unit. The low $^{15}N–\{^1H\}$ heteronuclear NOE values ($\leq 0.4$), the close to zero values for the residual dipolar couplings of the backbone amides, and minimal deviations from random-coil chemical shifts for the C-terminal tail of capsid and SP1, both in the absence and presence of nucleic acids, indicate that these regions are intrinsically disordered in the context of CA–SP1–NC.*

The assembly of human immunodeficiency virus 1 (HIV-1) is mediated by the polyprotein precursor Gag (group-specific antigen).[1] The organization of Gag is as follows: matrix (MA)–capsid (CA)–spacer peptide 1 (SP1)–nucleocapsid (NC)–spacer peptide 2 (SP2)–p6. There are five HIV-1 protease cleavage sites located at the junction of these six elements. Gag is expressed in the cytoplasm of the host cell and subsequently transported to the plasma membrane, where assembly into an immature spherical particle takes place. During the budding process from the cell surface, HIV-1 protease cleaves Gag in a sequential manner, leading to the formation of a mature virion with a conical viral capsid. The matrix regulates the binding of Gag to the cell membrane, the capsid protein forms the viral capsid, nucleocapsid binds viral nucleic acids, and the intrinsically disordered p6 interacts with cellular and viral proteins. The spacer peptides, SP1 and SP2, are also critical for viral assembly and are thought to control conformational changes in Gag that take place during viral maturation.

Although extensive structural information is available for the individual protein components of Gag,[1] structural studies of either full-length or extended fragments of Gag have been hindered by conformational heterogeneity, which prevents crystallization,[2] and extreme sensitivity to proteolysis.[3,4] We recently published an extensive NMR and X-ray solution scattering study of the full-length capsid protein and obtained a quantitative description of the conformational space sampled by its N- and C-terminal domains on the nano- to millisecond time scale.[5] NMR studies on fragments comprising matrix and the N-terminal domain of capsid,[6] and the C-terminal domain of capsid, SP1, and nucleocapsid,[4] showed extensive motion between the independently folded units. Small-angle neutron scattering on a monomeric mutant of the matrix–capsid–SP1–nucleocapsid fragment of Gag, in which Trp316 and Met317 (equivalent to residues 184 and 185 of capsid and located at the dimerization interface of the C-terminal domain of capsid) were mutated to alanine, suggested that the Gag polyprotein is folded over with the matrix domain that is located close to the nucleocapsid domain.[7] However, the W316A/M317A double mutation alters the positional distribution of the N- and C-terminal domains of capsid relative to the wild-type monomeric form,[5] which may significantly impact the interpretation of the scattering data. Further, interpretation of scattering data for a highly dynamic system is complex and cannot be reliably carried out in the absence of additional data (e.g. from NMR). Lastly, a model derived from cryo-electron tomography of immature HIV-1 virions proposed that a helical bundle of six SP1 spacer peptides stabilizes capsid hexamers,[8] but the low resolution of this study precludes any definitive verification of this model.

Here, we present a solution NMR study of a 300-residue capsid–SP1–nucleocapsid (CA–SP1–NC) fragment of HIV-1 Gag extending from residues 133 to 432 (numbering for the HXB2 strain), based upon residual dipolar coupling (RDC) data, complemented by heteronuclear $^{15}N$ relaxation data, backbone chemical shifts, and analytical ultracentrifugation (see the Supporting Information for expression and purification). This construct is functionally relevant, as it forms both tubes and cones in the presence of RNA (Supporting Information, Figure S1), consistent with previous reports.[9]

The dimerization properties of CA–SP1–NC (Figure 1a), which are mediated by the C-terminal domain of capsid, are

[*] Dr. L. Deshmukh, Dr. R. Ghirlando, Dr. G. M. Clore
Laboratories of Chemical Physics and Molecular Biology
National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health
Bethesda, MD 20892-0520 (USA)
E-mail: mariusc@mail.nih.gov

[**] We thank Drs. Schwieters, Grishaev, Bayro, Jenkins, and Cai for useful discussions, Dr. Sundquist for the gift of CA–SP1–NC cDNA, and Drs. Baber, Garrett, and Gustchina for technical support. This work was supported by funds from the Intramural Program of the NIH, NIDDK, and from the Intramural AIDS Targeted Antiviral Program of the Office of the Director of the NIH (to G.M.C.).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201309127.

  

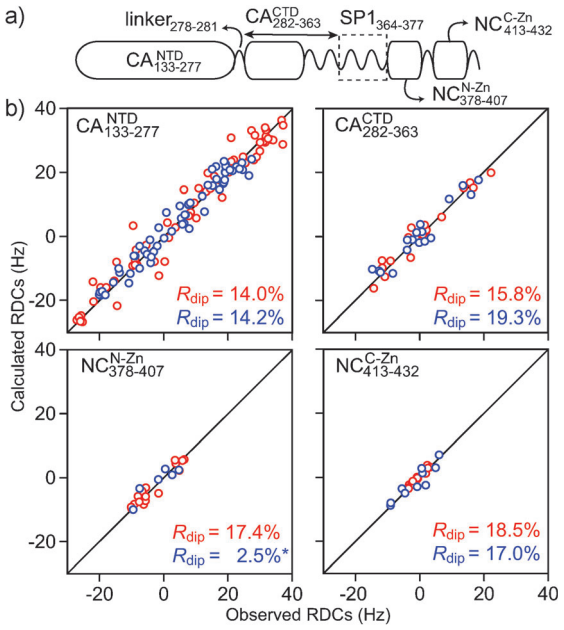


***Figure 1.*** RDC analysis of CA–SP1–NC with and without nucleic acids. a) Schematic representation of the structural domains of CA–SP1–NC. The N- and C-terminal domains of capsid are indicated as $CA^{NTD}_{133-277}$ and $CA^{CTD}_{282-363}$, respectively; the N- and C-terminal Zn knuckles of nucleocapsid as $NC^{N-Zinc}_{378-407}$ and $NC^{C-Zinc}_{413-432}$, respectively. b) Singular-value decomposition (SVD) analysis showing the agreement of the experimental backbone amide ($^1D_{NH}$) RDCs acquired in lipid bicelles (red: without DNA, blue: with ΔP(−) PBS DNA at a protein–DNA molar ratio of 1:2) with those calculated from the coordinates of the isolated capsid (PDB: 2M8L and 2M8N[5]) and nucleocapsid (PDB: 2M3Z[14]). The RDC *R*-factor, $R_{dip}$, is given by $\{\langle(D_{obs}\text{-}D_{calc})^2\rangle/(2\langle D_{obs}{}^2\rangle)\}^{1/2}$, where $D_{obs}$ and $D_{calc}$ are the observed and calculated RDC values, respectively. Note that severe resonance broadening is observed for the N-terminal Zn knuckle of nucleocapsid in the presence of DNA, resulting in far fewer RDCs (6 as compared to 17 without DNA) for SVD analysis, leading to an artificially low (2.5%) value of $R_{dip}$. PDB = protein data bank.

similar to those of the isolated capsid protein,[5] as judged by analytical ultracentrifugation (Supporting Information, Table S1). As in the case of capsid,[5] the time scale of monomer–dimer exchange is intermediate on the chemical shift scale. As a result, some $^1H_N/^{15}N$ cross-peaks for residues at or close to the dimerization interface of the capsid C-terminal domain, as well as within the linker connecting the N- and C-terminal capsid domains, are broadened out in the $^1H$–$^{15}N$ TROSY correlation spectrum (Supporting Information, Figure S2). Some regions of SP1 (residues 368–372) and the N-terminal end (residues 381–388) of the nucleocapsid domain are also broadened out, presumably as a result of rapid proton exchange between the backbone amides and water (Supporting Information, Figure S3). However, the similarity between the $^1H_N/^{15}N$ backbone chemical shifts of the structured domains of CA–SP1–NC with those of the isolated capsid (residues 133–363) and nucleocapsid (residues 379–432) constructs indicates that these domains retain their fold within the context of CA–SP1–NC and likely tumble semi-independently of one another.

Addition of excess DNA to CA–SP1–NC, either in the form of a linear single-stranded 30 mer $(TG)_{15}$ or a 14 mer hairpin structure known as the DNA (−) primer binding site[10] (5′d-GTCCCTGTTCGGGC, hereafter referred to as ΔP(−)PBS DNA) results in the formation of a 1:1 CA–SP1–NC/DNA complex (Supporting Information, Figure S4 and S5), accompanied by significant $^1H_N/^{15}N$ chemical shift perturbations within the nucleocapsid domain but not in other regions of CA–SP1–NC (Supporting Information, Figure S2B,C). We therefore conclude that interaction with nucleic acids is limited exclusively to the nucleocapsid domain. In addition, cross-peaks for the N-terminal region of nucleocapsid become visible upon addition of DNA, as expected since this region in the isolated nucleocapsid construct is known to adopt an ordered conformation upon DNA binding[11] (Supporting Information, Figure S3).

RDC measurements in a liquid crystalline medium of neutral bicelles provide information on bond vector orientations, molecular shape, and interdomain dynamics.[5,12,13] There is excellent agreement between the observed backbone amide ($^1D_{NH}$) RDCs obtained both in the absence and presence of DNA and those calculated from the coordinates of the individual domains of capsid[5] and nucleocapsid,[14] indicating that the structure of each domain remains essentially unaltered in the context of CA–SP1–NC (Figure 1). Of note, however, is that the magnitude of the principal component $(D_a^{NH})$ of the alignment tensor, as well as the rhombicity $(\eta)$ of the tensor is very different for the four structural domains (i.e., the N- and C-terminal domains of capsid, and the N- and C-terminal Zn knuckles of nucleocapsid; Table 1), indicative of large, rigid body interdomain motions. Thus, the value of $D_a^{NH}$ for the N-terminal domain of capsid is twice that for the C-terminal domain of capsid and the N-terminal Zn knuckle of nucleocapsid, and close to one order of magnitude larger than that for the C-terminal Zn knuckle of nucleocapsid. Moreover, whereas the align-

***Table 1:*** Magnitude of RDC alignment tensors for the structural domains of CA–SP1–NC.

| Domain[a] | without DNA | | with DNA | |
|---|---|---|---|---|
| | $D_a^{NH}$ [Hz] | $\eta$[b] | $D_a^{NH}$ [Hz] | $\eta$[b] |
| $CA^{NTD}_{133-277}$ (60/59) | 18.3 | 0.33 | 12.6 | 0.34 |
| $CA^{CTD}_{282-363}$ (19/17)[c] | −10.5 | 0.65 | −9.2 | 0.64 |
| $NC^{N-Zn}_{378-407}$ (17/6) | 8.1 | 0.19 | −5.7 | 0.59 |
| $NC^{C-Zn}_{413-432}$ (13/10) | 2.1 | 0.49 | −6.2 | 0.51 |

[a] The first and second numbers refer to the number of RDCs without and with DNA, respectively. Only RDCs in secondary structure elements are included in the SVD analysis. [b] The maximum value of the rhombicity $\eta$ is 2/3. When $\eta = 0$, the tensor is axially symmetric. [c] For the C-terminal domain of capsid, the values of $D_a^{NH}$ and $\eta$ reported in the table are apparent values obtained by fitting against the coordinates of a single subunit. The C-terminal domain is an exchanging mixture of monomer and dimer. When the RDC data at three different concentrations of CA–SP1–NC (0.05, 0.1 and 0.2 mM in subunits) are fitted to the appropriate population-weighted mixtures of a monomer and dimer (as determined from analytical ultracentrifugation, Supporting Information, Table S1), the values of $D_a^{NH}$ and $\eta$ are 10.1 Hz and 0.26 for the dimer, and 23.1 Hz and 0.51 for the monomer in the absence of DNA. In the presence of DNA, the corresponding values are 11.5 Hz and 0.31 for the dimer, and 16.0 Hz and 0.58 for the monomer.

ment tensor is close to axially symmetric for the N-terminal Zn knuckle of nucleocapsid, it is close to fully rhombic for the C-terminal domain of capsid and the C-terminal Zn knuckle of nucleocapsid. Upon binding to DNA, the two Zn knuckles of nucleocapsid have the same alignment tensor, indicating that their orientation relative to each other is fixed on the DNA (Table 1 and Figure 2 bottom right panel). Thus, in the presence of DNA, there are three structural units in CA–SP1–NC that tumble semi-independently of one another, namely the N- and C-terminal domains of capsid and the nucleocapsid/DNA complex.

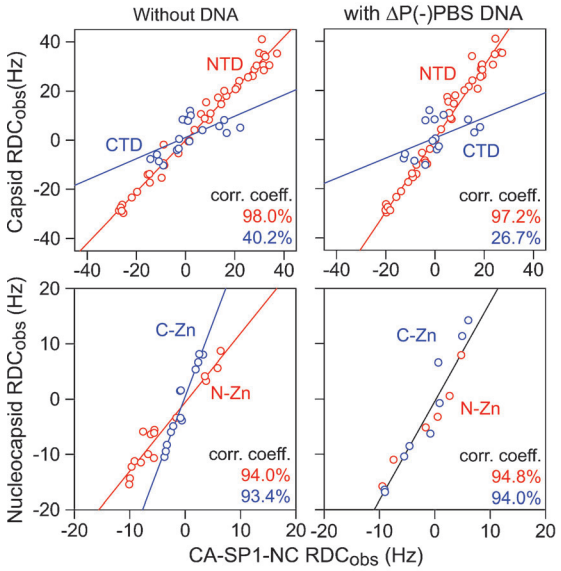


***Figure 2.*** Correlation of backbone amide RDCs for CA–SP1–NC with those for the isolated capsid (residues 133–363, top) and nucleocapsid (residues 379–432, bottom) constructs in the absence (left panels) and presence (right panels) of ΔP(−)PBS DNA. RDCs for the N-terminal domain (NTD) of capsid and the N-terminal Zn knuckle (N-Zn) of nucleocapsid are shown in red; the RDCs for the C-terminal domain (CTD) of capsid and the C-terminal Zn knuckle (C-Zn) of nucleocapsid are shown in blue. RDC data for capsid were taken from Ref. [5].

The extent of motion can be assessed by comparing the observed RDCs measured on CA–SP1–NC with those measured for the isolated capsid and nucleocapsid constructs (Figure 2). The RDCs for the N-terminal domain of capsid, and the two Zn knuckles of nucleocapsid measured for CA–SP1–NC are highly correlated to those measured for the isolated constructs, both in the absence and presence of DNA (Figure 2, left and right panels, respectively). This suggests that the motional characteristic of the N-terminal domain of capsid and the two Zn knuckles of nucleocapsid are similar in the context of the CA–SP1–NC, capsid, and nucleocapsid constructs. The corresponding correlations for the C-terminal domain of capsid, however, are poor (Figure 2, top panels). This is due to two factors: First, the dimerization equilibrium is a little different in isolated capsid[5] and the CA–SP1–NC construct (Supporting Information, Table S1) so that the proportion of monomer and dimer is altered. Second, the C-terminal domain of capsid serves as an anchor for the non-associating N-terminal domain of capsid and nucleocapsid domain, and is therefore influenced by the overall shape of the molecule, which is necessarily different for the capsid construct versus CA–SP1–NC.

What of the structural and dynamic characteristics of the C-terminal tail of capsid (residues 353–363) and the SP1 linker (residues 364–377) in the context of CA–SP1–NC? On the basis of peptide studies in which this region becomes helical upon reduction of solvent polarity (e.g. by addition of trifluoroethanol), it has been proposed that SP1 undergoes a coil-to-helix transition during viral assembly triggered by Gag binding to nucleic acids.[7,18] The values of the RDCs for both the capsid C-terminal tail and SP1 are close to zero, and the $^{15}$N–{$^{1}$H} heteronuclear NOE values are less than 0.4, indicative of high local mobility on the nanosecond time scale (Figure 3 and 4, and Supporting Information, Table S2).

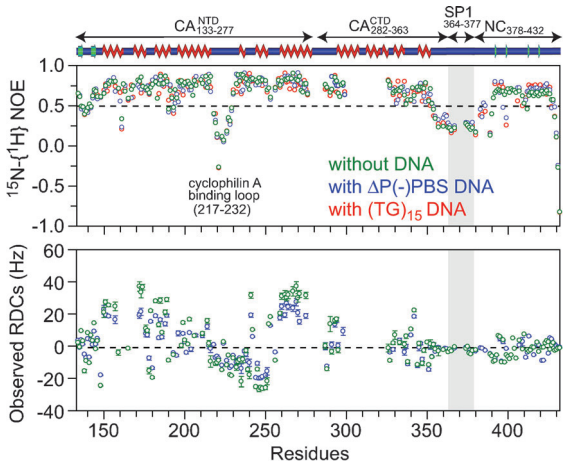


***Figure 3.*** $^{15}$N–{$^{1}$H} heteronuclear NOE (top) and backbone amide RDC (bottom) profiles for CA–SP1–NC in the absence and presence of DNA. The domain delineation and secondary structure elements of CA–SP1–NC are indicated above the panels. The semitransparent grey rectangle marks the location of the SP1 spacer (residues 364–377). The experimental data were recorded at a spectrometer $^{1}$H frequency of 800 MHz.

These values are unaffected by binding of nucleic acids to CA–SP1–NC. The absence of any structural ordering is confirmed by analysis of backbone chemical shifts using both δ2D[15] (Figure 4) and Talos+[16] (Supporting Information, Figure S3), thus indicating that residues 353–377 are intrinsically disordered and unaffected by the addition of DNA. The downfield-shifted Cα and upfield-shifted Cβ chemical shifts for Met367, however, are suggestive of a possible kink in the polypeptide chain at that location. Finally, nothing definitive can be said about the conformational characteristics of residues 368–372 of SP1 as their $^{1}$H$_N$/$^{15}$N cross-peaks were not observed. Nevertheless, the fact that there is extensive interdomain motion between the C-terminal domain of capsid and the N-terminal Zn knuckle of nucleocapsid suggests that residues 368–372 of SP1 are also highly mobile.

To probe the biological implications of an intrinsically disordered SP1 in the context of CA–SP1–NC, we looked at

  

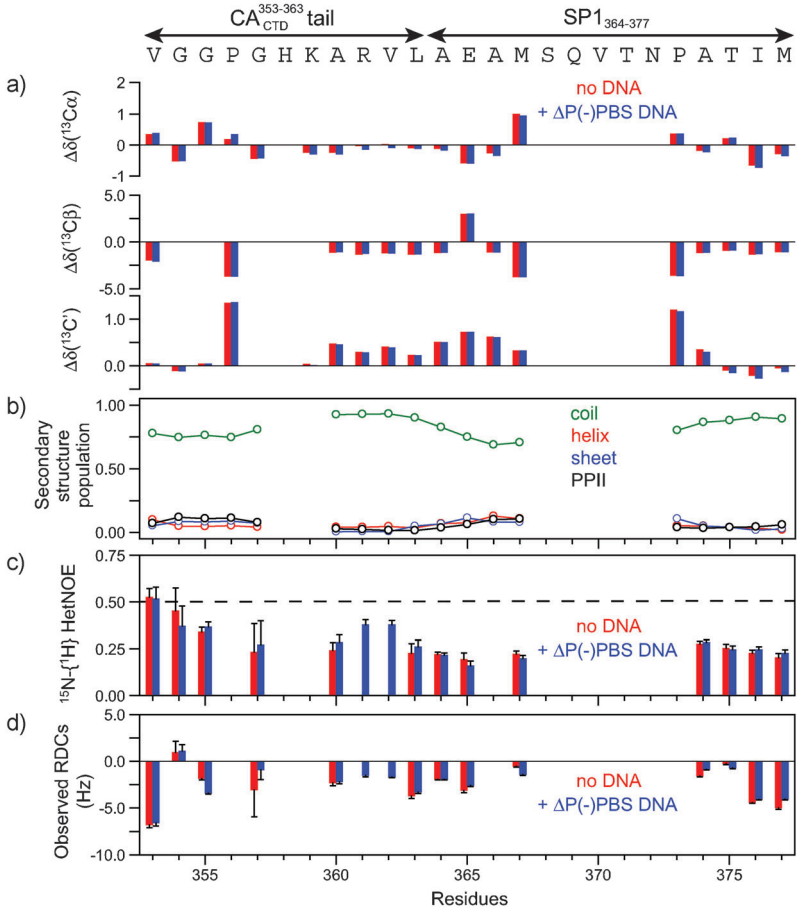


***Figure 4.*** Conformational analysis of the C-terminal tail of capsid (residues 353–363) and the SP1 linker (residues 364–377) in the context of CA–SP1–NC. a) Deviation of $^{13}C$ backbone chemical shifts from random coil,[17] b) secondary structure population derived from the backbone shifts using δ2D,[15] c) $^{15}N$–{$^{1}H$} NOEs, and d) backbone amide RDCs. In panels a, c, and d, the red and blue bars represent the data in the absence and presence of DNA, respectively. In panel b, the circles and lines represent the calculated secondary structure populations in the absence and presence of DNA, respectively.

cleavage rates by HIV-1 protease (Supporting Information, Figure S6). In the absence of nucleic acids (at pH 6.5), cleavage is very slow and approximately equal at the CA/SP1 and SP1/NC sites. But upon addition of nucleic acids, cleavage is greatly accelerated and occurs much more rapidly at the SP1/NC site, in agreement with in vitro and in vivo observations.[19] Based on the NMR data, SP1 remains intrinsically disordered in the presence of nucleic acids. We speculate that when the two Zn knuckles of nucleocapsid move independently of one another, access of HIV-1 protease to the CA/SP1 and SP1/NC cleavage sites is hindered, either through transient interactions or steric obstruction. However, once the two Zn knuckles behave as a single unit upon nucleic acid binding, access of HIV-1 protease to the SP1/NC site is considerably enhanced.

Received: October 18, 2013
Published online: December 11, 2013

**Keywords:** nucleic acids · molecular dynamics · proteins · residual dipolar couplings · viruses

[1] a) W. I. Sundquist, H. G. Krausslich, *Cold Spring Harbor Perspect. Med.* **2012**, *2*, a006924; b) A. A. Waheed, E. O. Freed, *AIDS Res. Hum. Retroviruses* **2012**, *28*, 54; c) N. M. Bell, A. M. Lever, *Trends Microbiol.* **2013**, *21*, 136.
[2] T. A. Bharat, N. E. Davey, P. Ulbrich, J. D. Riches, A. de Marco, M. Rumlova, C. Sachse, T. Ruml, J. A. Briggs, *Nature* **2012**, *487*, 385.
[3] S. Campbell, A. Rein, *J. Virol.* **1999**, *73*, 2270.
[4] J. L. Newman, E. W. Butcher, D. T. Patel, Y. Mikhaylenko, M. F. Summers, *Protein Sci.* **2004**, *13*, 2101.
[5] L. Deshmukh, C. D. Schwieters, A. Grishaev, R. Ghirlando, J. L. Baber, G. M. Clore, *J. Am. Chem. Soc.* **2013**, *133*, 16133.
[6] C. Tang, Y. Ndassa, M. F. Summers, *Nat. Struct. Biol.* **2002**, *9*, 537.
[7] S. A. Datta, L. G. Temeselew, R. M. Crist, F. Soheilian, A. Kamata, J. Mirro, D. Harvin, K. Nagashima, R. E. Cachau, A. Rein, *J. Virol.* **2011**, *85*, 4111.
[8] E. R. Wright, J. B. Schooler, H. J. Ding, C. Kieffer, C. Fillmore, W. I. Sundquist, G. J. Jensen, *EMBO J.* **2007**, *26*, 2218.
[9] B. K. Ganser, S. Li, V. Y. Klishko, J. T. Finch, W. I. Sundquist, *Science* **1999**, *283*, 80.
[10] S. Bourbigot, N. Ramalanjaona, C. Boudier, G. F. Salgado, B. P. Roques, Y. Mely, S. Bouaziz, N. Morellet, *J. Mol. Biol.* **2008**, *383*, 1112.
[11] a) R. N. De Guzman, Z. R. Wu, C. C. Stalling, L. Pappalardo, P. N. Borer, M. F. Summers, *Science* **1998**, *279*, 384; b) G. K. Amarasinghe, R. N. De Guzman, R. B. Turner, K. J. Chancellor, Z. R. Wu, M. F. Summers, *J. Mol. Biol.* **2000**, *301*, 491.
[12] M. Zweckstetter, A. Bax, *J. Am. Chem. Soc.* **2000**, *122*, 3791 – 3792.
[13] D. T. Braddock, M. Cai, J. L. Baber, Y. Huang, G. M. Clore, *J. Am. Chem. Soc.* **2001**, *123*, 8634.
[14] N. Goudreau, O. Hucke, A. M. Faucher, C. Grand-Maitre, O. Lepage, P. R. Bonneau, S. W. Mason, S. Titolo, *J. Mol. Biol.* **2013**, *425*, 1982.
[15] C. Camilloni, A. De Simone, W. F. Vranken, M. Vendruscolo, *Biochemistry* **2012**, *51*, 2224.
[16] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, *J. Biomol. NMR* **2009**, *44*, 213.
[17] M. Kjaergaard, S. Brander, F. M. Poulsen, *J. Biomol. NMR* **2011**, *49*, 139.
[18] N. Morellet, S. Druillennec, C. Lenoir, S. Bouaziz, B. P. Roques, *Protein Sci.* **2005**, *14*, 375 – 386.
[19] a) R. J. Tritch, Y. E. Cheng, F. H. Yin, S. Erickson-Viitanen, *J. Virol.* **1991**, *65*, 922; b) S. C. Pettit, M. D. Moody, R. S. Wehbie, A. H. Kaplan, P. V. Nantermet, C. A. Klein, R. Swanstrom, *J. Virol.* **1994**, *68*, 8017; c) K. Wiegers, G. Rutter, H. Kottler, U. Tessmer, H. Hohenberg, H. G. Krausslich, *J. Virol.* **1998**, *72*, 2846; d) G. Mirambeau, S. Lyonnais, R. J. Gorelick, *RNA Biol.* **2010**, *7*, 724.