VIP Real-Time NMR Spectroscopy | Very Important Paper

DOI: 10.1002/ange.201309940

Quantitative Analysis of Location- and Sequence-Dependent Deamination by APOBEC3G Using Real-Time NMR Spectroscopy**

*Ayako Furukawa, Kenji Sugase, Ryo Morishita, Takashi Nagata, Tsutomu Kodaki, Akifumi Takaori-Kondo, Akihide Ryo, and Masato Katahira**

Abstract: *The human antiretroviral factor APOBEC3G (A3G) deaminates the newly synthesized minus strand of the human immunodeficiency virus 1 (HIV-1), which results in the abolition of the infectivity of virus-infectivity-factor (Vif)-deficient HIV-1 strains.[1–6] A unique property of A3G is that it deaminates a CCC hot spot that is located close to the 5′ end more effectively than one that is less close to the 5′ end. However, the mechanism of this process is elusive as it includes nonspecific binding of A3G to DNA and sliding of A3G along the DNA strand. Therefore, this process cannot be analyzed by existing methods using the Michaelis–Menten theory. A new real-time NMR method has been developed to examine the nonspecific binding and the sliding processes explicitly, and it was applied to the analysis of the deamination by A3G. As a result, the location-dependent deamination can be explained by a difference in the catalytic rates that depend on the direction of the approach of A3G to the target cytidine. Real-time NMR experiments also showed that A3G deaminates CCCC tandem hotspots with little redundancy, which suggests that A3G efficiently mutates many CCC hotspots that are scattered throughout the HIV-1 genome.*

The enzyme APOBEC3G (A3G) possesses two consensus zinc-finger-type cytidine deaminase motifs (CD1 and CD2),[2] but only CD2 is catalytically active.[7,8] A3G preferably deaminates the third cytidine of a CCC sequence in single-stranded DNA (ssDNA).[9,10] It was reported that A3G nonspecifically binds to ssDNA and slides along ssDNA over 30 nm (69 nucleotides) without directional preference.[11,12] Interestingly, A3G deaminates CCC hot spots in a location-dependent fashion.[10] A 5′ to 3′ gradient of mutations in HIV RNA, which is transcribed from the minus strand DNA, was observed in vivo.[9,13] This would arise from the 3′ to 5′ deamination gradient of minus strand DNA by A3G. Deamination by A3G has been analyzed by various methods, including gel shift assays,[10,14,15] single-molecule fluorescence resonance energy transfer (FRET) spectroscopy,[11] and atomic force microscopy.[12,16] We previously demonstrated that real-time NMR spectroscopy can be utilized to monitor the deamination reaction and revealed that A3G CD2 deaminates the third cytidine of CC<u>C</u> much faster than the second one (C<u>C</u>C).[17] An advantage of real-time NMR spectroscopy over other methods is that it can directly detect a site-specific deamination reaction. Moreover, this method is sensitive to weak interactions because highly concentrated ssDNA (mM order) can be used for the NMR experiment. The real-time NMR method that we used for the analysis of the deamination process has become increasingly popular and has been used by other groups.[18,19]

Our real-time NMR method monitors the intensity change of the H5–H6 total correlation spectroscopy (TOCSY) peak of the third cytidine of CC<u>C</u> in real time. Using this method, we previously analyzed the deamination of a hot spot that solely exists in ssDNA. To gain insight into the mechanism that underlies the location-dependent deamination by A3G, we further developed this NMR method. First, to determine whether real-time NMR spectroscopy provides sufficiently high spectral and time resolution to distinguish the multiple deamination reactions that occur on ssDNA, we monitored the deamination reactions of ssDNA comprising two CCC hot spots (S_{2CCC} in Table 1; see also the Supporting Information). Obviously, the two hot spots were deaminated at different rates by full-length A3G (Figure 1a). Surprisingly, CD2 alone also deaminated the two hot spots in a location-dependent manner, exhibiting higher activity than full-length A3G (Figure 1b). The higher activity of A3G CD2 was previously observed by Chen and co-workers.[14] F-tests showed that the difference between the deamination rates for

[*] Dr. A. Furukawa, Prof. T. Nagata, Prof. T. Kodaki, Prof. M. Katahira
Institute of Advanced Energy, Graduate School of Energy Science
Kyoto University
Gokasho, Uji, Kyoto 611-0011 (Japan)
E-mail: katahira@iae.kyoto-u.ac.jp
Dr. K. Sugase
Bioorganic Research Institute, Suntory Foundation for Life Sciences
1-1-1 Wakayamadai, Shimamoto, Mishima, Osaka 618-8503 (Japan)
Dr. R. Morishita
CellFree Sciences Co., Ltd., Ehime University Venture
Matsuyama, Ehime 790-8577 (Japan)
Prof. A. Takaori-Kondo
Department of Hematology and Oncology
Graduate School of Medicine, Kyoto University
54 Shogoin-Kawaracho, Sakyo-ku, Kyoto 606-8507 (Japan)
Dr. R. Morishita, Prof. A. Ryo
Department of Microbiology
Yokohama City University School of Medicine
3-9 Fukuura, Kanazawa-ku, Yokohama 236-0004 (Japan)

[**] We thank the Ministry of Education, Science, Sports and Culture of Japan for Grants-in Aid for Scientific Research (24121714, 25115507, and 25291013 to M.K.; 23570146 and 24113710 to T.N.), Japan Science and Technology (CREST; M.K.), the Sumitomo–Denko and Iwatani Foundations (M.K.), and the Japan Society for the Promotion of Science (A.F.). For funding for open access charges, we are grateful to the Ministry of Education, Science, Sports and Culture of Japan (24121714).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201309940.

Table 1: Oligonucleotides used in this study.

Name	Sequence
S_{2CCC}	ATTCCCAATTTTTTTTATACCCATTT
$S_{CCCssCCC}$	ATTCCCAATTTTTTTTTTTTTTTTTTTATACCCAT_{23}
$S_{CCCdsCCC}$	ATTCCCAATTCGCGAACGCGCAAGCGCATACCCAT_{23} GCGCTTGCGCGTTCGCG
$S_{5'CCC}$	TTACCCAT_{42}
S_{mCCC}	T_{21}ACCCAT_{22}
$S_{3'CCC}$	T_{41}ACCCATTT
S_{CCCC}	ATTCCCCATT
S_{CCCU}	ATTCCCUATT
S_{CCUC}	ATTCCUCATT
S_{CCCA}	ATTCCCAATT
S_{CCUA}	ATTCCUAATT
S_{CCUU}	ATTCCUUATT

The cytidines that are to be deaminated are underlined.

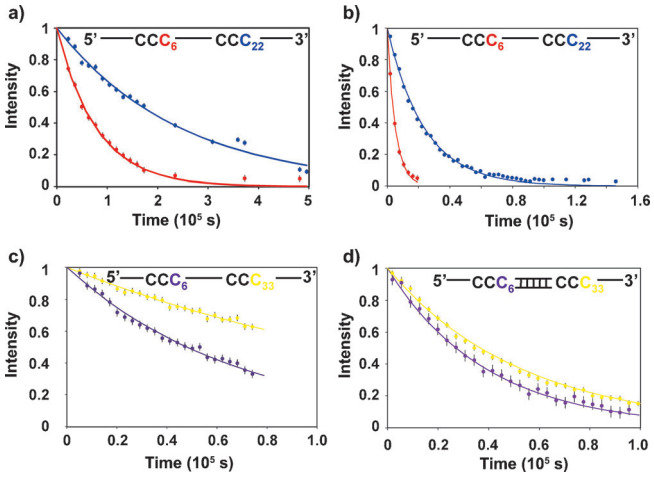

Figure 1. Real-time monitoring of deamination reactions at two CCC hot spots in ssDNA. a, b) Deamination reactions with full-length A3G (a) and CD2 (b) were monitored for C6 (●) and C22 (●) in S_{2CCC}. The deamination rates for full-length A3G are $1.3\times10^{-5}\pm4.0\times10^{-7}$ s^{-1} for C6 and $4.1\times10^{-6}\pm7.4\times10^{-8}$ s^{-1} for C22, and those for CD2 are $2.0\times10^{-4}\pm7.5\times10^{-6}$ s^{-1} for C6 and $4.2\times10^{-5}\pm5.0\times10^{-7}$ s^{-1} for C22. c, d) Deamination reactions of CD2 were monitored for C6 (●) and C33 (●) in $S_{CCCssCCC}$ (c) and $S_{CCCdsCCC}$ (d).

the two hot spots is statistically significant for both full-length A3G and CD2. The same result was obtained with ssDNA in which two units of S_{2CCC}, ATTCCCAAT and ATACCCATT, had swapped positions. These results confirmed that CD2 alone causes the location-dependent deamination. Although catalytically inactive CD1 also supposedly contributes to the location-dependent deamination process, we assumed that the intrinsic characteristics of CD2 are important for the location-dependent deamination by full-length A3G. We focused on CD2 in the following experiments. Subsequently, we investigated whether our real-time NMR method can be used to examine the sliding of A3G along ssDNA. As A3G binds to ssDNA, but not to dsDNA,[10] intervening dsDNA should block the sliding of A3G. As shown in Figure 1c and d, the deamination rates of two hot spots in a substrate with short dsDNA between them were almost the same, whereas the rates were location-dependent without dsDNA ($S_{CCCdsCCC}$ and $S_{CCCssCCC}$ in Table 1). This finding clearly indicates that real-time NMR spectroscopy can sense the sliding of A3G CD2 along ssDNA.

We have thus confirmed that real-time NMR spectroscopy is sufficiently sensitive to location-dependent deamination reactions and sliding processes of A3G CD2. Based on these results, we designed experiments and constructed a kinetic model to quantitatively analyze the deamination reaction. As A3G binds to ssDNA nonspecifically, the binding rate should depend on the length of the ssDNA and the concentrations of A3G and ssDNA. On the other hand, the duration of the sliding to reach a hot spot should depend on the length of the ssDNA and the position of a hot spot. Therefore, the kinetics of the deamination reaction can be determined by real-time NMR studies with different concentrations of ssDNA and A3G CD2, and with ssDNAs of different lengths and with different positions of hot spots. According to this idea, we developed a kinetic model for the analysis of real-time NMR data (see the Supporting Information). The deamination reaction that is monitored by NMR spectroscopy as a change of intensity, $I(t)$, is expressed as:

$$I(t) = I_0\exp(-k_{\mathrm{deami}}t) \tag{1}$$

I_0 is the initial intensity, which is proportional to the concentration of ssDNA, and k_{deami} represents the apparent deamination rate, which is calculated as detailed below:

$$k_{\mathrm{deami}} = \frac{[^NS][^NE]}{S_0K_{\mathrm{d}}}\left\{k_{\mathrm{cat}(3'\rightarrow5')}(1-\alpha^n)+k_{\mathrm{cat}(5'\rightarrow3')}\left(1-\alpha^{N-n+1}\right)\right\} \tag{2}$$

$$[^NS] = \frac{1}{2}\left(-{}^N\beta+S_0-E_0+\sqrt{({}^N\beta-S_0+E_0)^2+4E_0{}^N\beta}\right) \tag{3}$$

$$[^NE] = \frac{1}{2}\left(-{}^N\beta-S_0+E_0+\sqrt{({}^N\beta+S_0-E_0)^2+4E_0{}^N\beta}\right) \tag{4}$$

$$\alpha = k_s/(k_s+k_{off}), \quad {}^N\beta = \frac{K_{\mathrm{d}}}{2\{N+1-(1-\alpha^{N+1})/(1-\alpha)\}} \tag{5}$$

N and n represent the total number of nucleotides in ssDNA and the position of the reacting cytidine from the 3' end, respectively. According to this kinetic model, A3G CD2 binds to any nucleotide of ssDNA with an association rate constant k_{on} and dissociates from ssDNA with a dissociation rate constant k_{off}. A3G CD2 slides along ssDNA in both directions at a sliding rate k_s (Figure 2a). As the ssDNAs used are relatively short (49–57 nucleotides), we assumed that the sliding of A3G CD2 does not change direction during a single sliding event. The structure of the A3G–ssDNA complex has not been determined yet; therefore, it is still unknown how A3G recognizes a hot spot. On the other hand, the structure of free A3G CD2 has already been determined, and a protruding bump that hangs over one side of the catalytic pocket, which includes the catalytic residue Glu259 and the zinc ion, was revealed (Supporting Information, Figure S1a).[14,17,20] We assumed that this bump interferes with the accommodation of the target cytidine into the catalytic pocket; this process depends on the direction of the approach of A3G CD2 (Figure S1b,c). Then, we incorporated two cata-

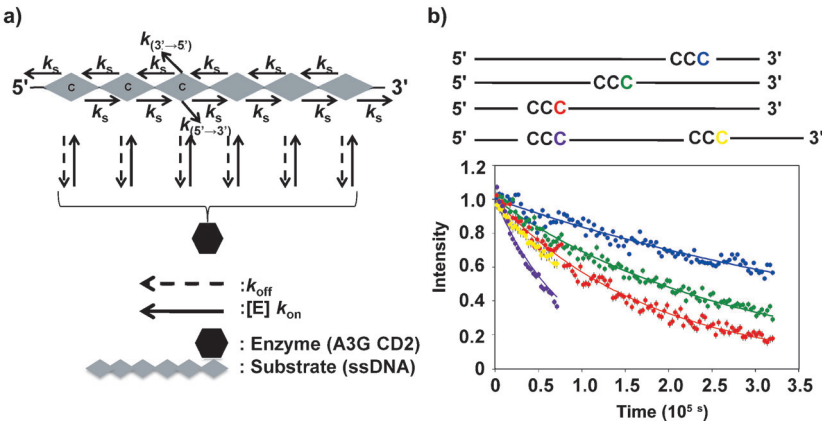

Figure 2. Quantitative analysis of the location-dependent deamination reaction using real-time NMR spectroscopy. a) Kinetic model of the deamination reaction by A3G. b) Real-time NMR data for C6 of $S_{5'CCC}$ (●), C26 of S_{mCCC} (●), C45 of $S_{3'CCC}$ (●), and both C6 (●) and C33 (●) of $S_{CCCssCCC}$ were fitted to Eq. (1).

lytic rates, $k_{cat(3'\rightarrow5')}$ and $k_{cat(5'\rightarrow3')}$, into the model, which are the catalytic rates for the cases where A3G CD2 approaches the target cytidine from downstream and upstream, respectively. As A3G slides along ssDNA without directional preference,[12] the location-dependent deamination is difficult to explain without the two different k_{cat} values. These k_{cat} values represent the rates of the entry of ssDNA into the catalytic pocket and of its deamination. A global fit of multiple real-time NMR data collected under different conditions (see above) provided the parameters α, K_d ($=k_{off}/k_{on}$), $k_{cat(3'\rightarrow5')}$, and $k_{cat(5'\rightarrow3')}$. However, it is difficult to separate α into k_s and k_{off} because neither the sliding nor the dissociation process can be perturbed experimentally without affecting the other process. The NMR data were fitted to Eq. (1) using the program GLOVE.[21]

For this analysis, we obtained real-time NMR data for $S_{5'CCC}$, S_{mCCC}, and $S_{3'CCC}$ (Table 1) aside from the data for the aforementioned two hot spots of $S_{CCCssCCC}$. These data are in good agreement with Eq. (1), yielding values for $k_{cat(3'\rightarrow5')}$ and $k_{cat(5'\rightarrow3')}$ of 68 s^{-1} and 14 s^{-1}, respectively (Figure 2b). The closer a CCC hot spot is located to the 5′ end, the more chance it has of being detected by A3G approaching from downstream rather than from upstream; therefore, a CCC hot spot that is located close to the 5′ end is deaminated more rapidly than one that is less close to the 5′ end.

For further applications of our real-time NMR method, we analyzed the deamination of CCCC, in which two CCC sequences overlap (S_{CCCC} in Table 1). The clusters of three to six consecutive cytidines are scattered throughout the HIV genome, and CCCC is the second most abundant cluster after CCC. Unfortunately, the NMR peaks that need to be monitored during the deamination reactions overlapped, for example, the TOCSY peaks of the fourth cytidines of CCCC and CCUC. Therefore, the time courses of each deamination reaction could not be monitored separately (Figure 3a and b). However, the results of the deamination reactions were characterized using the NMR spectrum of the same ssDNA whose deamination reactions had reached completion. As a result, the third and fourth cytidines of CCCC were found to be deaminated by A3G CD2 with equal efficiency because the intensities of peaks 2 (CCUC) and 3 (CCCU + CCUU) were nearly identical. This finding is different from that of a previous study with a gel shift assay; in the earlier case, the third cytidine was deaminated more efficiently than the fourth cytidine.[15] This discrepancy could be due to the difference in the duration of the monitoring between our NMR method (6 h) and the gel shift assay (3 min). Real-time NMR spectroscopy can monitor multiple reactions directly and simultaneously using a single sample, whereas the gel shift assay cannot. Therefore, our results should provide more accurate information.

The deamination of CCCU into CCUU is redundant and may be unnecessary for abolition of the HIV infectivity. However, the above experiment could not provide information on the deamination efficiency for CCCU because the NMR peaks of CCUU and CCUC overlapped. Thus, we monitored the deamination reactions for CCCA (non-deaminated site) and CCCU (monodeaminated site) to determine whether or not A3G has a preference for either of the two sequences, S_{CCCA} and S_{CCCU} (Table 1). In this case, the deamination reactions were monitored by one-dimensional ^{1}H NMR spectroscopy to increase spectral resolution (Figure 3c and d). The results revealed that A3G preferentially deaminates a non-deaminated site rather than a monodeami-

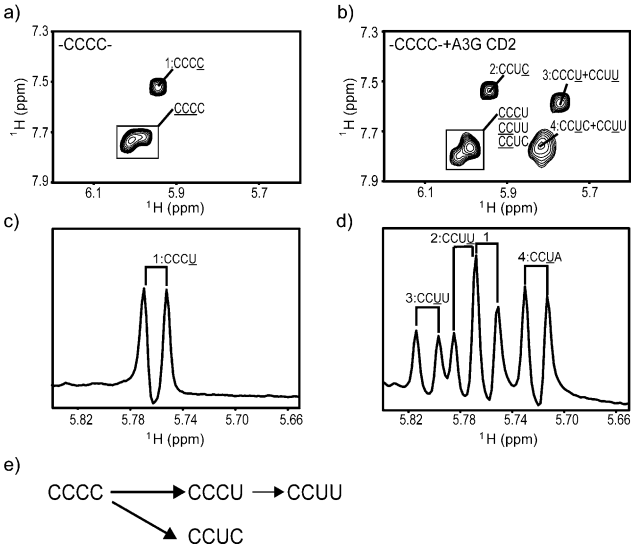

Figure 3. Deamination reactions of CCCC by A3G CD2. a, b) TOCSY spectrum of S_{CCCC} (a) and TOCSY spectrum recorded six hours after the addition of A3G CD2 (b). The peaks are labeled with the sequences of the corresponding substrates or deaminated products in which the assigned residues are underlined (see the Supporting Information). c, d) ^{1}H NMR spectrum of an equimolar mixture of S_{CCCA} and S_{CCCU} (c), and ^{1}H NMR spectrum recorded 40 minutes after the addition of A3G CD2 (d). Peaks 1 to 4 were assigned by recording the NMR spectra of S_{CCUU} and S_{CCUA}. e) Two possible deamination reaction pathways for CCCC.

nated site because the intensity of the peak for the deaminated product CCUA (peak 4) was greater than that of the peak corresponding to CCUU (peak 3). A3G would exploit this sequential preference to efficiently deaminate multiple hot spots that are scattered throughout the HIV genome with a small amount of redundant deamination of CCCC.

In conclusion, we analyzed real-time NMR data for the location-dependent deamination by A3G CD2 using a newly derived kinetic model. As a result, the location-dependent deamination can be explained by the difference in the rate constants k_{cat} of CD2, which depend on the direction of the approach to the target cytidine. We also characterized deamination reactions of CCCC by real-time NMR spectroscopy, even though multiple reactions occurred simultaneously. Previously, it was proposed that the location-dependent deamination is caused by two binding orientations on ssDNA and the existence of a region where either less or no deamination occurs (ca. 30 nucleotides) at the 3′ end of the ssDNA.[10,16] However, this model cannot explain the location-dependent deamination that is observed for long ssDNA (70 nucleotides)[15] because the contribution of the dead region to the location dependency becomes smaller. On the other hand, our model can explain the location-dependent deamination reaction even for such a long ssDNA fragment.

Recently, it was proposed that the proteins of the APOBEC family are involved in epigenesis.[22] They may contribute to the removal of an epigenetic marker, a methyl group of a 5-methyl cytosine, through deamination of 5-hydroxymethyl cytosine. Our real-time NMR method could also be applied to examine such DNA modifications and to provide insight into epigenesis. Furthermore, this method can also be used for the analysis of post-translation modifications of proteins.

Received: November 15, 2013
Published online: January 29, 2014

Keywords: deamination · DNA · enzyme kinetics · quantitative analysis · NMR spectroscopy

[1] A. M. Sheehy, N. C. Gaddis, J. D. Choi, M. H. Malim, *Nature* **2002**, *418*, 646 – 650.
[2] R. S. Harris, M. T. Liddament, *Nat. Rev. Immunol.* **2004**, *4*, 868 – 877.
[3] D. Lecossier, F. Bouchonnet, F. Clavel, A. J. Hance, *Science* **2003**, *300*, 1112.
[4] B. Mangeat, P. Turelli, G. Caron, M. Friedli, L. Perrin, D. Trono, *Nature* **2003**, *424*, 99 – 103.
[5] H. Zhang, B. Yang, R. J. Pomerantz, C. Zhang, S. C. Arunachalam, L. Gao, *Nature* **2003**, *424*, 94 – 98.
[6] R. S. Harris, K. N. Bishop, A. M. Sheehy, H. M. Craig, S. K. Petersen-Mahrt, I. N. Watt, M. S. Neuberger, M. H. Malim, *Cell* **2003**, *113*, 803 – 809.
[7] G. Hache, M. T. Liddament, R. S. Harris, *J. Biol. Chem.* **2005**, *280*, 10920 – 10924.
[8] F. Navarro, B. Bollman, H. Chen, R. Konig, Q. Yu, K. Chiles, N. R. Landau, *Virology* **2005**, *333*, 374 – 386.
[9] Q. Yu, R. Konig, S. Pillai, K. Chiles, M. Kearney, S. Palmer, D. Richman, J. M. Coffin, N. R. Landau, *Nat. Struct. Mol. Biol.* **2004**, *11*, 435 – 442.
[10] L. Chelico, P. Pham, P. Calabrese, M. F. Goodman, *Nat. Struct. Mol. Biol.* **2006**, *13*, 392 – 399.
[11] L. S. Shlyakhtenko, A. Y. Lushnikov, A. Miyagi, M. Li, R. S. Harris, Y. L. Lyubchenko, *Biochemistry* **2012**, *51*, 6432 – 6440.
[12] G. Senavirathne, M. Jaszczur, P. A. Auerbach, T. G. Upton, L. Chelico, M. F. Goodman, D. Rueda, *J. Biol. Chem.* **2012**, *287*, 15826 – 15835.
[13] R. Suspène, C. Rusniok, J. P. Vartanian, S. Wain-Hobson, *Nucleic Acids Res.* **2006**, *34*, 4677 – 4684.
[14] L. G. Holden, C. Prochnow, Y. P. Chang, R. Bransteitter, L. Chelico, U. Sen, R. C. Stevens, M. F. Goodman, X. S. Chen, *Nature* **2008**, *456*, 121 – 124.
[15] R. Nowarski, E. Britan-Rosich, T. Shiloach, M. Kotler, *Nat. Struct. Mol. Biol.* **2008**, *15*, 1059 – 1066.
[16] L. Chelico, C. Prochnow, D. A. Erie, X. S. Chen, M. F. Goodman, *J. Biol. Chem.* **2010**, *285*, 16195 – 16205.
[17] A. Furukawa, T. Nagata, A. Matsugami, Y. Habu, R. Sugiyama, F. Hayashi, N. Kobayashi, S. Yokoyama, H. Takaku, M. Katahira, *EMBO J.* **2009**, *28*, 440 – 451.
[18] S. Harjes, W. C. Solomon, M. Li, K. M. Chen, E. Harjes, R. S. Harris, H. Matsuo, *J. Virol.* **2013**, *87*, 7008 – 7014.
[19] I. J. Byeon, J. Ahn, M. Mitra, C. H. Byeon, K. Hercik, J. Hritz, L. M. Charlton, J. G. Levin, A. M. Gronenborn, *Nat. Commun.* **2013**, *4*, 1890.
[20] K. M. Chen, E. Harjes, P. J. Gross, A. Fahmy, Y. Lu, K. Shindo, R. S. Harris, H. Matsuo, *Nature* **2008**, *452*, 116 – 119.
[21] K. Sugase, T. Konuma, J. C. Lansing, P. E. Wright, *J. Biomol. NMR* **2013**, *56*, 275 – 283.
[22] N. Bhutani, D. M. Burns, H. M. Blau, *Cell* **2011**, *146*, 866 – 872.

