doi:10.1246/cl.2011.852
Published on the web July 16, 2011

Interstrand Crosslink for Discrimination of Methylated Cytosines

Chikara Dohno, Tomonori Shibata, and Kazuhiko Nakatani*
Department of Regulatory Bioorganic Chemistry, The Institute of Scientific and Industrial Research (ISIR), Osaka University, Ibaraki, Osaka 567-0047

(Received May 11, 2011; CL-110398; E-mail: nakatani@sanken.osaka-u.ac.jp)

We have developed a chemistry-based method to discriminate cytosine from methylated cytosine using a probe DNA bearing an aldehyde functionality. The probe DNA undergoes interstrand crosslinking with target DNA containing a non-methylated cytosine in a sequence-dependent manner, while the crosslink formation is significantly suppressed for both 5-methyl and N4-methyl cytosines.

Genomic information consists of a one-dimensional array of four chemical letters, adenine, guanine, thymine, and cytosine. Chemically modified nucleosides are often found in genomic DNA and can be considered as fifth and subsequent letters of DNA. Methylation of nucleobases is a commonly used chemical modification in many organisms, and three methylated nucleobases, 5-methylcytosine (5mC), N^4-methylcytosine (4mC), and N^6-methyladenine (6mA), are known as a natural component of genomic DNA.[1] In particular, 5mC has attracted considerable attention because of the biological and pathogenic importance in many eukaryotic cells. Patterns of C5-cytosine methylation in genomic DNA determine epigenetic gene regulation and cellular differentiation and are associated with many human cancers.[2] 6mA and 4mC are mostly found in bacterial genome and participate in genome defense, replication, mismatch repair, and control of gene expression.[3] These methylated nucleobases are introduced by DNA methyltransferases and provide additional information that is not encoded by four standard nucleobases.

Since 5mC plays significant roles in mammalian cells, many analytical methods for 5mC distribution have been developed over the past decade.[4] Bisulfite conversion, methylation-sensitive enzymatic digestion, and immunoprecipitation are conventional techniques for the methylome analysis.[4,5] Besides the most common bisulfite sequencing,[6] there have been several reports on methylation analysis based on selective chemical reactions.[7–10] We recently reported a chemistry-based method to detect N^6-methyladenine by the formation of a selective interstrand crosslink (ICL).[11] Here we demonstrated that the ICL formation could be also applicable for the detection of methylation status of cytosine in the predetermined target sequence. Electrophilic formyl group, when properly placed in the major groove of DNA, selectively reacts with exocyclic amino groups of cytosine but not with the methylated counterparts, 5mC and 4mC (Figure 1). This new chemical methodology will be applicable to detect all types of natural methylated bases of DNA.

While the chemical basis of strategies for discriminating 5mC from C is the susceptibility of C5–C6 double bonds to the nucleophilic addition[6,7] and oxidation reactions[8,9] depending on the presence of C5 methyl groups, we focused on nucleophilic reaction involving an exocyclic amino group at C4. Since the amino group is located in close proximity to the cytosine

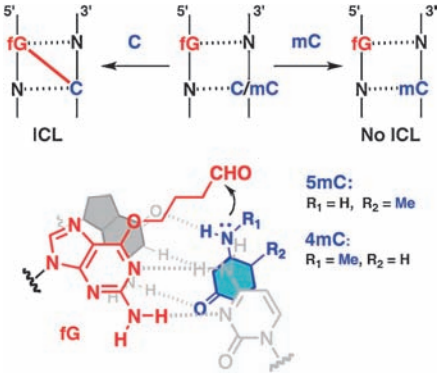

Figure 1. Schematic illustration of the discrimination of methylated cytosines, 5mC ($R_1 = H$, $R_2 = Me$) and 4mC ($R_1 = Me$, $R_2 = H$) by the interstrand crosslink (ICL) between ^{f}G and cytosine. The ICL formation is suppressed in the presence of the methyl group.

methylation sites, a nucleophilic attack of the amino group on the electrophilic group placed near the cytosine-N4 would be sensitive to base methylation. We previously demonstrated that N6-methylated and non-methylated adenine could be discriminated by ICL between the exocyclic amino group of adenine-N6 and ^{f}G, which has an electrophilic formyl group at the O6 position of guanine via three methylene linker (Figure 1) and can deliver the formyl group into close proximity to the target adenine base.[11] The presence of the methyl group at the N6 position of adenine suppresses the ICL formation probably due to steric hindrance. To know whether the ICL-based strategy is applicable to the methylated cytosine, ICL between ^{f}G and exocyclic amino group of cytosine-N4 was investigated in the presence and absence of methyl groups at C5 and N4.

ICL between the ^{f}G-containing DNA and its complementary DNA containing a target cytosine base was analyzed by denaturing polyacrylamide gel electrophoresis (PAGE, Figure 2a). The ICL should be electrophoresed much slower than DNAs without any ICL under the denaturing conditions. After incubation of the ^{f}G-containing duplex in sodium cacodylate buffer (pH 6) for 8 h, slower mobility bands corresponding to the ICL were observed in sequence-dependent manner. When cytosine is located opposite ^{f}G (lane 5) or the 5′-adjacent base of ^{f}G (lanes 3 and 4), no ICL formation is observed. The ICLs were observed only for the sequence containing a cytosine opposite the 3′-adjacent base of ^{f}G, 5′-^{f}GN-3′/5′-CN-3′ (Figure 2a, lanes 1 and 2). The observed specificity is consistent with that of the previously reported ICL involving adenine-N6.[11,12b] The ICL reaction of the 5′-^{f}GG-3′/5′-CT-3′ sequence (Figure 2a, lane 1) reached a plateau phase after 4 h incubation with chemical yield of 11% as determined

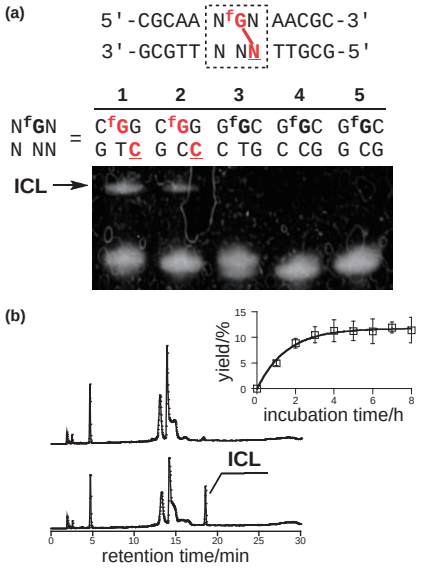

Figure 2. ICLs at a cytosine opposite the 3′-adjacent base of ^{f}G, 5′-^{f}GN-3′/5′-CN-3′. (a) A series of DNA duplexes containing a different core sequence were incubated at 30 °C for 8 h and then subjected to denaturing polyacrylamide gel electrophoresis. The crosslinked cytosine is underlined. (b) HPLC analysis of the ICL formation in the 5′-^{f}GG-3′/5′-CT-3′ sequence before and after incubation for 8 h. The sharp peak eluted at 18 min is attributed to the ICL, while the broad peak at around 15 min is attributed to non-ICL duplex. dT (5 min) was added as an internal standard. Inset: time course of the ICL formation.

by HPLC (Figure 2b). Matrix-assisted laser desorption/ionization (MALDI) time-of-flight mass spectrometry (TOF MS) measurements of the isolated ICL by HPLC showed recovery of the initial ^{f}G-containing DNA. This phenomenon is quite similar to that of ICL between adenine-N6 and ^{f}G.[11] The saturation of the ICL formation and the reproduction of the initial DNA during the MALDI-TOF MS measurements suggests the reversible nature of the ICL reaction between the aldehyde and the amino groups in aqueous media. The formation of carbinolamine and/or Schiff base is likely responsible for the ICL.[11–14]

Sequence-specific ICL formation between ^{f}G and cytosine-N4 allows us to use ^{f}G-containing DNA as the probe for the methylation status of cytosine. For discrimination of cytosine from methylated cytosine (5mC or 4mC),[15] we prepared a series of oligonucleotides containing a sequence, 5′-^{f}GG-3′/5′-(mC/C)N-3′, where the target cytosine is located at 5′ adjacent to the opposite base of ^{f}G. The ICL formation in the presence and absence of the methyl group was investigated by PAGE (Figure 3).[17] ICL band was observed for non-methylated cytosine with a different efficiency depending on the opposite base N (odd number lanes). Thymine base opposite ^{f}G (N = T) was found to be optimal for the ICL, most likely because the ^{f}G·T base pairing allows the formyl group to be placed in a proper arrangement for the ICL.[16] In the case of 4mC, in contrast, no ICL band was observed (Figure 3b). The presence of methyl group at N4 of cytosine effectively hinders ICL formation between ^{f}G and cytosine-N4.

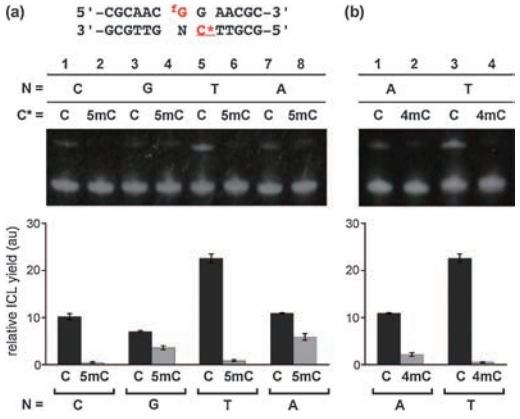

Figure 3. Discrimination between methylated and non-methylated cytosine by ICL with ^{f}G-containing DNA. (a) PAGE analysis for the discrimination between cytosine and 5mC. The ICL formation was investigated changing the opposite base of ^{f}G. Relative ICL yield was estimated from (band intensity of ICL)/(total band intensity). (b) Discrimination between cytosine and 4mC.

In the case of 5mC, the ICL formation was strongly suppressed, when the opposite base of ^{f}G is a pyrimidine base (N = C or T, Figure 3a, lanes 2 and 6). The methyl substitution at C5 position resulted in an approximately 30-fold decrease in the ICL formation (N = T). Although the C5 methyl group is not directly attached to the amino group responsible for the ICL, the presence of the methyl group hinders electrophilic attack of the formyl group of ^{f}G. When the opposite base of ^{f}G is purine (N = G or A), both efficiency of the crosslinking and discrimination were lower than that for pyrimidines (Figure 3a), indicating that 5′-(mC)G-3′ is not the best target sequence for ^{f}G. Since C5 methylation at non-CpG sites is limited to a particular type of cells,[5c] the sequence dependency on the opposite base needs to be improved for further applications.

In conclusion, we presented a chemistry-based methodology to discriminate cytosine methylation status using ^{f}G-containing DNA. The presence of a methyl group both at C5 and N4 of the cytosine strongly suppressed the ICL formation with the ^{f}G-containing DNA. Considering our previous results for 6 mA,[11] the strategy based on the nucleophilic reactions of nucleotide base amines is effective for detection of any of three methylated bases found in genomic DNA, C5, N4 of cytosine and N6 of adenine.

References and Notes

1 A. Jeltsch, *ChemBioChem* **2002**, *3*, 274.
2 a) K. D. Robertson, A. P. Wolffe, *Nat. Rev. Genet.* **2000**, *1*, 11. b) P. A. Jones, S. B. Baylin, *Nat. Rev. Genet.* **2002**, *3*, 415. c) K. D. Robertson, *Nat. Rev. Genet.* **2005**, *6*, 597.
3 a) D. Ratel, J.-L. Ravanat, F. Berger, D. Wion, *BioEssays* **2006**, *28*, 309. b) D. Wion, J. Casadesús, *Nat. Rev. Microbiol.* **2006**, *4*, 183. c) M. Ehrlich, G. Wilson, K. Kuo, C. Gehrke, *J. Bacteriol.* **1987**, *169*, 939.
4 a) T. Rein, M. L. DePamphilis, H. Zorbas, *Nucleic Acids Res.* **1998**, *26*, 2255. b) S. Beck, V. K. Rakyan, *Trends Genet.* **2008**, *24*, 231. c) M. M. Suzuki, A. Bird, *Nat. Rev. Genet.* **2008**, *9*, 465. d) P. W. Laird, *Nat. Rev. Genet.* **2010**,

11, 191.
5 a) X. Zhang, J. Yazaki, A. Sundaresan, S. Cokus, S. W.-L. Chan, H. Chen, I. R. Henderson, P. Shinn, M. Pellegrini, S. E. Jacobsen, J. R. Ecker, *Cell* **2006**, *126*, 1189. b) S. J. Cokus, S. Feng, X. Zhang, Z. Chen, B. Merriman, C. D. Haudenschild, S. Pradhan, S. F. Nelson, M. Pellegrini, S. E. Jacobsen, *Nature* **2008**, *452*, 215. c) R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q.-M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, J. R. Ecker, *Nature* **2009**, *462*, 315.
6 a) H. Hayatsu, Y. Wataya, K. Kai, *J. Am. Chem. Soc.* **1970**, *92*, 724. b) R. Shapiro, R. E. Servis, M. Welcher, *J. Am. Chem. Soc.* **1970**, *92*, 422. c) M. Frommer, L. E. McDonald, D. S. Millar, C. M. Collis, F. Watt, G. W. Grigg, P. L. Molloy, C. L. Paul, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 1827.
7 H. Ohmori, J.-i. Tomizawa, A. M. Maxam, *Nucleic Acids Res.* **1978**, *5*, 1479.
8 a) A. Okamoto, K. Tainaka, T. Kamei, *Org. Biomol. Chem.* **2006**, *4*, 1638. b) K. Tanaka, K. Tainaka, T. Kamei, A. Okamoto, *J. Am. Chem. Soc.* **2007**, *129*, 5612. c) A. Okamoto, *Org. Biomol. Chem.* **2009**, *7*, 21.
9 a) T. Rein, D. A. Natale, U. Gärtner, M. Niggemann, M. L. DePamphilis, H. Zorbas, *J. Biol. Chem.* **1997**, *272*, 10021. b) S. Bareyt, T. Carell, *Angew. Chem., Int. Ed.* **2008**, *47*, 181.
10 a) K. Tanabe, H. Yamada, S.-i. Nishimoto, *J. Am. Chem. Soc.* **2007**, *129*, 8034. b) D. Kato, N. Sekioka, A. Ueda, R. Kurita, S. Hirono, K. Suzuki, O. Niwa, *J. Am. Chem. Soc.* **2008**, *130*, 3716. c) M. Ogino, Y. Taya, K. Fujimoto, *Chem. Commun.* **2008**, 5996.
11 C. Dohno, T. Shibata, K. Nakatani, *Chem. Commun.* **2010**, *46*, 5530.
12 a) S. Dutta, G. Chowdhury, K. S. Gates, *J. Am. Chem. Soc.* **2007**, *129*, 1852. b) J. T. Sczepanski, A. C. Jacobs, M. M. Greenberg, *J. Am. Chem. Soc.* **2008**, *130*, 9646.
13 a) H.-Y. H. Kim, M. Voehler, T. M. Harris, M. P. Stone, *J. Am. Chem. Soc.* **2002**, *124*, 9324. b) I. D. Kozekov, L. V. Nechev, M. S. Moseley, C. M. Harris, C. J. Rizzo, M. P. Stone, T. M. Harris, *J. Am. Chem. Soc.* **2003**, *125*, 50. c) Y.-J. Cho, H.-Y. Kim, H. Huang, A. Slutsky, I. G. Minko, H. Wang, L. V. Nechev, I. D. Kozekov, A. Kozekova, P. Tamura, J. Jacob, M. Voehler, T. M. Harris, R. S. Lloyd, C. J. Rizzo, M. P. Stone, *J. Am. Chem. Soc.* **2005**, *127*, 17686.
14 a) M. Manoharan, L. K. Andrade, P. D. Cook, *Org. Lett.* **1999**, *1*, 311. b) C. Dohno, A. Okamoto, I. Saito, *J. Am. Chem. Soc.* **2005**, *127*, 16681. c) T. Angelov, A. Guainazzi, O. D. Schärer, *Org. Lett.* **2009**, *11*, 661.
15 a) G. Vilkaitis, S. Klimasauskas, *Anal. Biochem.* **1999**, *271*, 116. b) A. Bart, M. W. van Passel, K. van Amsterdam, A. van der Ende, *Nucleic Acids Res.* **2005**, *33*, e124. c) B. A. Flusberg, D. R. Webster, J. H. Lee, K. J. Travers, E. C. Olivares, T. A. Clark, J. Korlach, S. W. Turner, *Nat. Methods* **2010**, *7*, 461.
16 a) M. W. Kalnik, B. F. L. Li, P. F. Swann, D. J. Patel, *Biochemistry* **1989**, *28*, 6182. b) G. A. Leonard, J. Thomson, W. P. Watson, T. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9573. c) S. L. Ginell, J. Vojtechovsky, B. Gaffney, R. Jones, H. M. Berman, *Biochemistry* **1994**, *33*, 3487.
17 Supporting Information is available electronically on the CSJ-Journal Web site, http://www.csj.jp/journals/chem-lett/index.html.

