BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

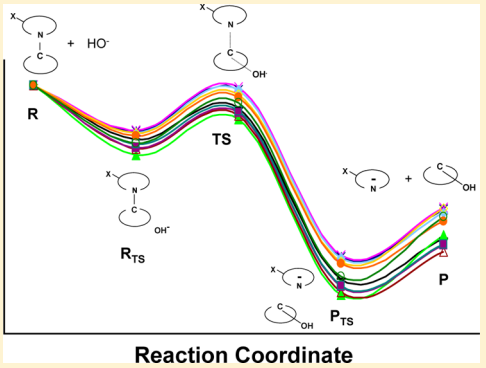
pubs.acs.org/biochemistry

A Density Functional Theory Study on the Kinetics and Thermodynamics of *N*-Glycosidic Bond Cleavage in 5-Substituted 2′-Deoxycytidines

Renee T. Williams and Yinsheng Wang*

Department of Chemistry-027, University of California, Riverside, California 92521-0403, United States

ABSTRACT: B3LYP/6-311+G(2d,p)//B3LYP/6-31+G(d) density functional theory calculations were employed to explore the kinetics and thermodynamics of gas-phase *N*-glycosidic bond cleavage induced by nucleophilic attack of C1′ with a hydroxide ion in 5-substituted 2′-deoxycytidines. The results showed that, among the 5-substituted 2′-deoxycytidine derivatives examined [XdC, where X = H (dC), CH_3 (medC), CH_2OH (hmdC), CHO (fmdC), COOH (cadC), F (FdC), or Br (BrdC)], fmdC and cadC exhibited the lowest energy barrier and largest exothermicity for *N*-glycosidic bond cleavage. These results paralleled previously reported nucleobase excision activities of human thymine DNA glycosylase (hTDG) toward duplex DNA substrates harboring a thymine and 5-substituted cytosine derivatives when paired with a guanine. Our study suggests that the inherent chemistry associated with the nucleophilic cleavage of *N*-glycosidic bond constitutes a major factor contributing to the selectivity of hTDG toward 5-substituted dC derivatives. These findings provided novel insights into the role of TDG in active cytosine demethylation.



Cytosine methylation at CpG sites is one of the best-characterized epigenetic modifications and is central to a variety of cellular processes, including retrotransposon silencing, genomic imprinting, X-chromosome inactivation, regulation of gene expression, and maintenance of epigenetic memory.[1] Aberrant cytosine methylation is linked to imprinting disorders such as Prader-Willi and Angelman syndromes and is implicated in diseases, including cancer.[2,3] In mammalian cells, DNA (cytosine-5)-methyltransferase 1 is known to convert nascent, replication-produced hemimethylated CpG sites to fully methylated ones,[4] and loss or inhibition of this process can give rise to passive DNA demethylation. However, the mechanism for active, DNA replication-independent demethylation in mammals remains poorly defined.[4] It was shown very recently that the ten-eleven translocation 1-3 (Tet 1-3) proteins can successively oxidize meC to hmC,[5,6] fmC,[7] and caC (Figure 1).[7,8] Additionally, hmC, fmC, and caC are found in mammalian DNA,[7−9] and human thymine DNA glycosylase (TDG) can readily excise fmC and caC, but not C, meC, or hmC, from duplex DNA.[10] These findings provide important evidence that supports a newly proposed mechanism of active cytosine demethylation, which involves iterative oxidation of meC to fmC and caC by Tet 1-3, TDG-mediated excision of fmC and caC, and the subsequent employment of base excision repair (BER) machinery to yield unmethylated cytosine at CpG sites, though the involvement of direct decarboxylation of caC in active cytosine demethylation cannot be excluded (Figure 1A).[11]

TDG is an enzyme that removes thymine (T)[12] or uracil[13] when paired with guanine as well as a variety of cytosine derivatives[14] located at the 5′-CpG-3′ site in duplex DNA. Crystal structure analysis revealed that the sequence-specific mismatch recognition of TDG may be attributed to electrostatic interactions of the N1H and $N2H_2$ groups of the opposing G, relative to the damaged base, with the backbone amides of A274 and P280 within the enzyme.[15,16] These initial structures were determined for the complexes formed between the catalytic domain of human TDG and double-stranded DNA (dsDNA) containing an abasic site; thus, they did not provide insight into the mechanism of nucleobase recognition. Only very recently were X-ray crystal structures determined for the catalytic domain of human TDG in complex with dsDNA harboring a nonhydrolyzable cadC analogue or the corresponding catalytically inactive mutant with cadC-containing DNA.[17] It was found that the active site of human TDG recognizes the 5-carboxylate group of cadC through a network of hydrogen bonds in a small pocket formed by the side chains of A145 and N157, and the backbone atoms of H150, H151, and Y152.[17]

To the best of our knowledge, no computational analysis of the kinetics and thermodynamics associated with the cleavage of the *N*-glycosidic bond in the context of TDG-mediated selective cleavage of caC and fmC has been reported. Thus, we used density functional theory (DFT) calculations to computationally assess the activation energy (E_a) and thermochemistry (E_{rxn}) of excision of a nucleobase from thymidine (dT) and 5-X-2′-deoxycytidines (XdC), where X is H (dC), CH_3 (medC),

Received: June 15, 2012
Revised: July 17, 2012
Published: July 18, 2012

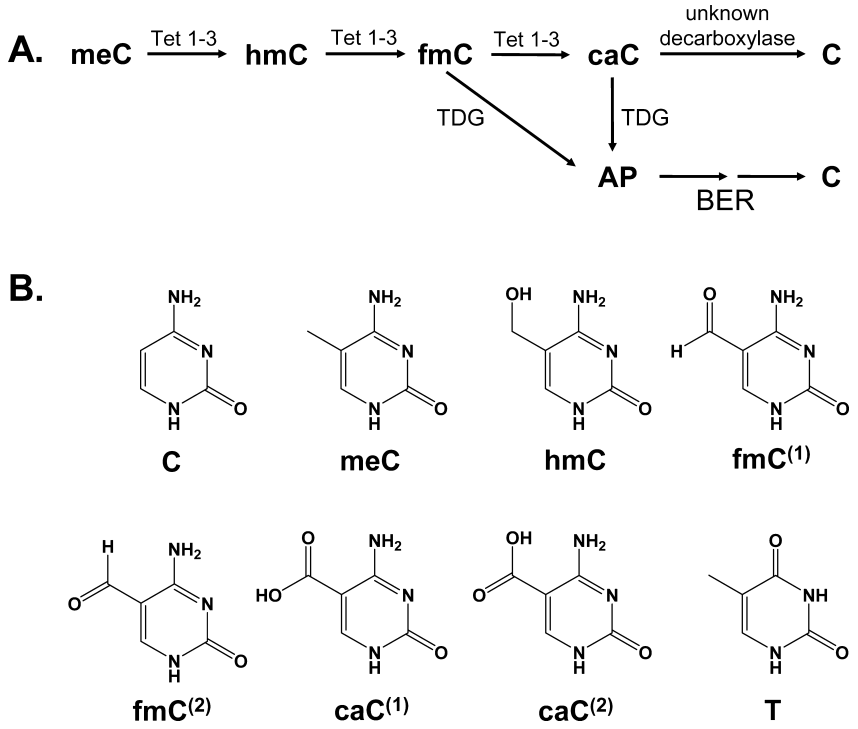


Figure 1. (A) Proposed mechanism for active DNA demethylation of meC to C. (B) Chemical structures of nucleobases examined in this work.

CH_2OH (hmdC), CHO (fmdC), COOH (cadC), F (FdC), or Br (BrdC), via hydroxide ion-mediated nucleophilic attack on C1′ of 2′-deoxyribose in the gas phase. Results from DFT calculations at the B3LYP/6-311+G(2d,p)//B3LYP/6-31+G(d) level of theory predicted that, among all the dC derivatives examined, cleavage of the *N*-glycosidic bond in fmdC and cadC exhibits the lowest energy barriers and highest exothermicity. These results are consistent with the largest k_{max} (maximal rate) for nucleobase excision observed for duplex DNA harboring these two modified nucleosides measured under single-turnover conditions.[10]

COMPUTATIONAL METHODS

All calculations were conducted using DFT methods included in the Gaussian 09W package.[18] Geometry optimizations (including transition states) and frequency calculations (without scaling) were performed at the B3LYP/6-31+G(d) level of theory in the gas phase. The transition state structures were confirmed by the presence of only one imaginary frequency. The structures of the initial reactant and product complexes (R_{TS} and P_{TS}, respectively) were determined on the basis of manual displacements of the optimized transition state (TS) structure. Those initial structures were then optimized [B3LYP/6-31+G(d)] to ensure a minimum, where no imaginary frequency was identified. Basis set superposition error (BSSE) corrections were calculated for R_{TS}, TS, and P_{TS} at the B3LYP/6-311+G(2d,p) level of theory on B3LYP/6-31+G(d)-optimized geometries [termed B3LYP/6-311+G(2d,p)//B3LYP/6-31+G(d) within the text] using counterpoise (fragments = 2) to account for molecular interactions between basis functions that are within proximity. Single-point energies (SPEs) were also determined from B3LYP/6-311+G(2d,p)//B3LYP/6-31+G(d) calculations.

Our choice of basis sets stemmed from earlier studies showing that 6-31G(d) geometries yield final relative energies in excellent agreement with geometries obtained using the larger basis set, 6-31+G(d,p).[19,20] These studies also showed that the C3′ and C5′ hydroxyl groups of the 2′-deoxyribose interact with the nucleophile and/or nucleobase, which cannot occur in duplex DNA. As such, all C3′ and C5′ hydroxyl groups in the nucleosides and deoxyribose structures modeled herein were replaced with methoxyl groups for more accurate calculations. Although modified, these structures are still termed "2′-deoxy" within the text. On the grounds that an Asn140-activated H_2O molecule serves as a nucleophile attacking C1′ to induce the *N*-glycosidic bond cleavage,[17] we employed hydroxide ion as the nucleophile for modeling the TDG-mediated cleavage of the pyrimidine nucleosides.

The total energies (in kilojoules per mole) for R_{TS}, TS, and P_{TS} were calculated as follows: $E = (E_{SPE} + E_{BSSE})_{B3LYP/6\text{-}311+G(2d,p)} + (E_{ZPEC})_{B3LYP/6\text{-}31+G(d)}$, where E_{ZPEC} is the zero-point energy correction (ZPEC) extracted from the frequency calculations. Additionally, the B3LYP/6-31+G(d)-optimized TS complex was used to generate electrostatic potential maps with an isosurface of 0.02 electron/Å^3. The activation energy (E_a) and reaction energy (E_{rxn}) were determined by taking the difference in total energy between TS and R_{TS} and between P_{TS} and R_{TS}, respectively. We also calculated the acidity (kilojoules per mole) in terms of the change in enthalpy (ΔH) at the N1 position of thymine and 5-X-cytosine-substituted derivatives at the B3LYP/6-31+G(d) level of theory as follows: $\Delta H = E_{H^+} + E_{XC^-} - E_{XC}$, where H^+, XC^-, and XC represent a proton, a deprotonated (N1 position) nucleobase, and a neutral nucleobase, respectively.

RESULTS AND DISCUSSION

Activation Energy and Reaction Energy. Results from DFT calculations at the B3LYP/6-311+G(2d,p)//B3LYP/6-31+G(d) level of theory predicted that, among all the dC derivatives examined, the cleavages of the *N*-glycosidic bonds in

dx.doi.org/10.1021/bi300797q | *Biochemistry* 2012, 51, 6458−6462

Table 1. Properties for the Thymidine and C5-Substituted 2′-Deoxycytidine Derivatives

nucleoside	E_a[a] (kJ/mol)	E_{rxn}[a] (kJ/mol)	acidity[a] (kJ/mol)	k_{max} (37 °C)[b] (min^{-1})	k_{max} (22 °C)[c] (min^{-1})
thymidine	43.1	−194.3	1390	1.83 ± 0.04	0.32 ± 0.01
2′-deoxycytidine (dC)	58.3	−174.9	1436		(1.2 ± 0.05) × 10^{-5}
5-methyl-dC	59.6	−173.8	1440		
5-hydroxymethyl-dC	53.7	−178.9	1434		<1.4 × 10^{-5}
5-formyl-dC (1)	49.6	−193.6	1370	2.64 ± 0.09	0.61 ± 0.04
5-formyl-dC (2)	46.8	−197.9	1372		
5-carboxyl-dC (1)	48.5	−193.4	1383	0.47 ± 0.01	0.14 ± 0.01
5-carboxyl-dC (2)	46.1	−196.4	1384		
5-fluoro-dC	54.7	−188.2	1416		0.035 ± 0.004
5-bromo-dC	56.7	−185.5	1402		0.008 ± 0.001

[a]From this work. [b]From ref 10. [c]From ref 14.

fmdC and cadC exhibit the lowest energy barriers and highest exothermicities (Table 1 and Figure 2). These results are consistent with the largest k_{max} (maximal rate) for nucleobase excision observed for these two modified nucleosides measured under single-turnover conditions, whereas TDG exhibited no activities toward dC, medC, or hmdC under physiological conditions.[14]

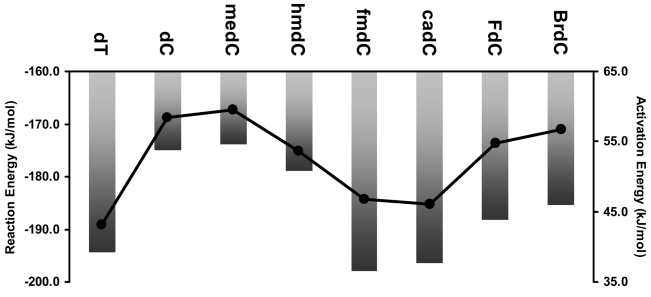


Figure 2. Reaction energies (E_{rxn}, bars) and activation energies (E_a, points) for cleavage of the *N*-glycosidic bond in dT and various 5-X-dC derivatives.

Using E_a as a benchmark, dT (E_a = 43.1 kJ/mol), fmdC (E_a = 46.8 kJ/mol, for the conformer with the lower energy barrier), and cadC (E_a = 46.1 kJ/mol, for the conformer with the lower energy barrier) display significantly lower energy barriers for *N*-glycosidic bond cleavage than all the other nucleosides assessed. This result correlates well with the largest k_{max} values obtained for these three nucleosides under single-turnover conditions (Table 1). In this vein, the electrostatic potential maps of the transition states obtained for all nucleosides assessed herein showed slightly less electron density on the N1 position of thymine, fmC, and caC than the other nucleosides (Figure 3). This result suggests that the transition state is more stabilized for these three nucleosides via better delocalization of the negative charge formed on N1 arising from partial glycosidic bond cleavage. Such stabilization of the transition state could be attributed, in part, to the electron-withdrawing nature of the C5 substituents (see more discussion about this below).[10,14]

The DFT calculation results also predict that the hydroxide-mediated cleavages of the *N*-glycosidic bond are thermodynamically more favorable for dT (E_{rxn} = −194.3 kJ/mol), and the lower-energy conformers of fmdC (E_{rxn} = −197.9 kJ/mol) and cadC (E_{rxn} = −196.4 kJ/mol), than for the other nucleosides examined (Table 1 and Figure 2).

Acidity. We also calculated the enthalpy of deprotonation, that is, the acidity of N1 for each nucleobase of interest, and we found that T, fmC, and caC had enthalpies that were lower than those of other 5-substituted cytosine derivatives examined (Table 1). Given that a decrease in deprotonation enthalpy reflects an increase in acidity, and thus, a greater stability of the monoanion that forms upon cleavage of the *N*-glycosidic

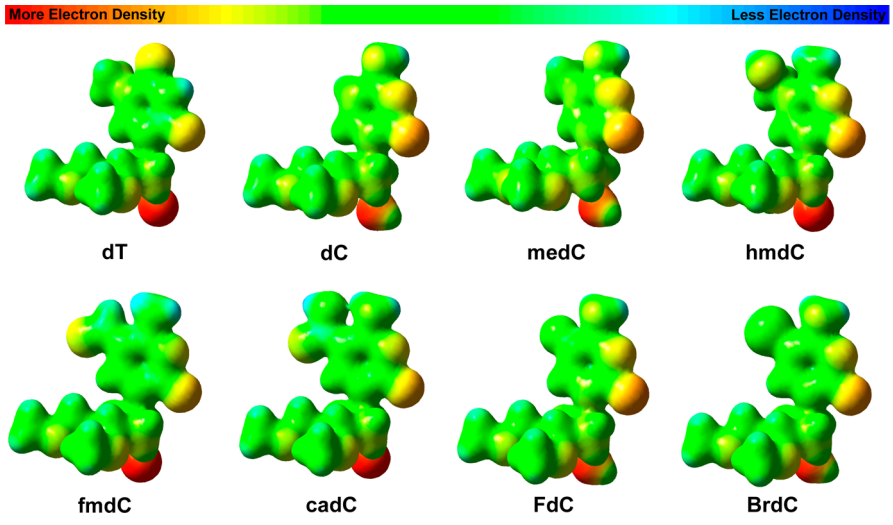


Figure 3. Elestrostatic potential maps of transition state structures for the hydroxide-mediated nucleophilic cleavage of dT and 5-substituted 2′-deoxycytidines examined in this study.

dx.doi.org/10.1021/bi300797q | *Biochemistry* 2012, 51, 6458−6462

Table 2. Bond Lengths and Dihedral Angles Determined for the Transition State and the Reactant and Product Complexes

	reactant complex			transition state structure			product complex		
	bond length (Å)		dihedral (deg)	bond length (Å)		dihedral (deg)	bond length (Å)		dihedral (deg)
nucleoside	N1−C1′	C1′−OH	∠(O1′C1′N1C2)	N1−C1′	C1′−OH	∠(O1′C1′N1C2)	N1−C1′	C1′−OH	∠(O1′C1′N1C2)
thymidine	1.519	3.091	197.5	1.988	2.364	175.8	4.413	1.428	27.1
2′-deoxycytidine (dC)	1.511	3.000	193.3	2.015	2.301	180.3	4.413	1.431	35.8
5-methyl-dC	1.511	2.993	192.7	2.017	2.302	179.0	4.473	1.431	28.8
5-hydroxymethyl-dC	1.513	3.000	191.5	2.005	2.329	178.6	4.497	1.430	27.2
5-formyl-dC (1)	1.477	2.767	168.2	1.972	2.360	180.7	4.533	1.428	27.3
5-formyl-dC (2)	1.473	2.743	176.0	1.965	2.372	185.3	4.549	1.428	29.9
5-carboxyl-dC (1)	1.477	2.773	171.6	1.975	2.309	182.4	4.537	1.415	31.6
5-carboxyl-dC (2)	1.476	2.765	174.4	1.971	2.368	184.1	4.538	1.485	30.8
5-fluoro-dC	1.513	3.000	192.4	1.999	2.311	180.1	4.493	1.430	30.5
5-bromo-dC	1.516	3.000	194.6	1.998	2.310	186.3	4.441	1.429	31.0

bond,[20] we report a strong correlation among higher N1 acidity, a lower barrier (E_a), greater exothermicity (E_{rxn}), and larger observed k_{max} values.

Molecular Geometry. The glycosidic bond length (N1−C1′), the distance between C1′ of the 2′-deoxyribose and the nucleophile (C1′−OH), and the dihedral angle about the glycosidic bond (∠O1′C1′N1C2) are listed in Table 2. As expected, there is a decrease in the C1′−OH distance, which is accompanied by an increase in the N1−C1′ bond length from the reactant complex to the transition state; however, dT, fmdC, and cadC bear noticeably shorter *N*-glycosidic bond lengths in the transition state structure than in the other dC derivatives. In the product complexes, both the C1′−OH and N1−C1′ bond lengths are comparable among all nucleosides.

It has been shown that, for cleavage of the *N*-glycosidic bond to occur, the nucleophile positioned underneath the 2′-deoxyribose ring in the reactant complex must first move closer to C1′ of the *N*-glycosidic bond.[20] Once this occurs, a transition state forms in which the partial cation generated on the 2′-deoxyribose is stabilized by charges on both the nucleophile and the departing nucleobase. Thus, the net charge is smaller when the *N*-glycosidic bond length in the transition state structure is shorter. This results in a greater stabilizing effect of the monoanion, which may lead to elevated TDG activity.[20]

CONCLUSIONS

In this work, we employ density functional theory to investigate the kinetics (E_a) and thermodynamics (E_{rxn}) for the hydroxide-mediated nucleophilic cleavage of the *N*-glycosidic bond in dT and 5-substituted dC derivatives. We provided computational evidence in support of a newly proposed mechanism for active cytosine demethylation where meC is iteratively oxidized to hmC, fmC, and then caC, followed by TDG-mediated nucleobase excision, and subsequent employment of the BER machinery to generate the unmethylated cytosine.[5−11] Specifically, we demonstrated that selective TDG activity toward fmC and caC compared to that toward C, meC, and hmC might arise partly from the inherent chemistry associated with nucleophilic cleavage of *N*-glycosidic bonds in these nucleosides. These cleavages exhibit lower energy barriers and greater exothermicities than the corresponding cleavages of other dC derivatives examined. We further showed that this dependence of energy barrier might be linked to the length of the *N*-glycosidic bond in the transition state structure, such that cleavage of the departing anionic nucleobase and the approach of the nucleophile stabilize the partial cation formed on the 2′-deoxyribose, thus rendering the reaction more energetically favorable. This is in contrast to previous reports suggesting that the k_{max} for forming the monoanion depends on the electronic substituent constant (σ_m) of X at the C5 position of the pyrimidine ring.[10,14] However, this theory has its shortcomings upon comparison of nucleobases such as fmC (k_{max} = 0.61 min^{-1}) and FC (k_{max} = 0.035 min^{-1}), which have similar σ_m values, 0.35 and 0.34, respectively, but markedly different activities toward TDG. Instead, we reason that assessment of the kinetics, thermodynamics, and molecular geometry involved in the cleavage of the *N*-glycosidic bond provides a more accurate prediction of varied hTDG activity toward a host of 5-substituted dC derivatives.

AUTHOR INFORMATION

Corresponding Author

*E-mail: yinsheng.wang@ucr.edu. Telephone: (951) 827-2700. Fax: (951) 827-4713.

Funding

We thank the National Institutes of Health for supporting this research (Grants R01 CA101864 and R01 DK082779-03S2).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Benjamin Moore, Dr. Thomas Morton, and Dr. Fernando Clemente for their assistance with select calculations.

ABBREVIATIONS

TDG, thymine DNA glycosylase; BER, base excision repair; DFT, density functional theory; TS, transition state.

REFERENCES

(1) Bird, A. (2002) DNA Methylation Patterns and Epigenetic Memory. *Genes Dev. 16*, 6−21.

(2) Lindahl, T. (1993) Instability and Decay of the Primary Structure of DNA. *Nature 362*, 709−715.

(3) Robertson, K. D. (2005) DNA Methylation and Human Disease. *Nat. Rev. Genet. 6*, 597−610.

(4) Chen, Z.-x., and Riggs, A. D. (2011) DNA Methylation and Demethylation in Mammals. *J. Biol. Chem. 286*, 18347−18353.

(5) Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., and Zhang, Y. (2010) Role of Tet Proteins in 5mC to 5hmC Conversion, ES-Cell Self-Renewal and Inner Cell Mass Specification. *Nature 466*, 1129−1133.

(6) Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., and Rao, A. (2009) Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science 324*, 930−935.
(7) Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., and Zhang, Y. (2011) Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science 333*, 1300−1303.
(8) He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.-X., Zhang, K., He, C., and Xu, G.-L. (2011) Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science 333*, 1303−1307.
(9) Pfaffeneder, T., Hackner, B., Truß, M., Münzel, M., Müller, M., Deiml, C. A., Hagemeier, C., and Carell, T. (2011) The Discovery of 5-Formylcytosine in Embryonic Stem Cell DNA. *Angew. Chem., Int. Ed. 50*, 7008−7012.
(10) Maiti, A., and Drohat, A. C. (2011) Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine. *J. Biol. Chem. 286*, 35334−35338.
(11) Wu, H., and Zhang, Y. (2011) Mechanisms and Functions of Tet Protein-Mediated 5-Methylcytosine Oxidation. *Genes Dev. 25*, 2436−2452.
(12) Wiebauer, K., and Jiricny, J. (1989) *In vitro* Correction of G o T Mispairs to G o C Pairs in Nuclear Extracts from Human Cells. *Nature 339*, 234−236.
(13) Neddermann, P., and Jiricny, J. (1994) Efficient Removal of Uracil from G.U Mispairs by the Mismatch-specific Thymine DNA Glycosylase from HeLa Cells. *Proc. Natl. Acad. Sci. U.S.A. 91*, 1642−1646.
(14) Bennett, M. T., Rodgers, M. T., Hebert, A. S., Ruslander, L. E., Eisele, L., and Drohat, A. C. (2006) Specificity of Human Thymine DNA Glycosylase Depends on N-Glycosidic Bond Stability. *J. Am. Chem. Soc. 128*, 12510−12519.
(15) Maiti, A., Morgan, M. T., and Drohat, A. C. (2009) Role of Two Strictly Conserved Residues in Nucleotide Flipping and N-Glycosylic Bond Cleavage by Human Thymine DNA Glycosylase. *J. Biol. Chem. 284*, 36680−36688.
(16) Maiti, A., Morgan, M. T., Pozharski, E., and Drohat, A. C. (2008) Crystal Structure of Human Thymine DNA Glycosylase Bound to DNA Elucidates Sequence-Specific Mismatch Recognition. *Proc. Natl. Acad. Sci. U.S.A. 105*, 8890−8895.
(17) Zhang, L., Lu, X., Lu, J., Liang, H., Dai, Q., Xu, G.-L., Luo, C., Jiang, H., and He, C. (2012) Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat. Chem. Biol. 8*, 328−330.
(18) Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J. A., Jr., Peralta, J. E., Ogliaro, F., Bearparke, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Keith, T., Kobayashi, R., Normande, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, O., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. (2010) *Gaussian 09*, revision B.01, Gaussian, Inc., Wallingford, CT.
(19) Millen, A. L., Archibald, L. A. B., Hunter, K. C., and Wetmore, S. D. (2007) A Kinetic and Thermodynamic Study of the Glycosidic Bond Cleavage in Deoxyuridine. *J. Phys. Chem. B 111*, 3800−3812.
(20) Millen, A. L., and Wetmore, S. D. (2009) Glycosidic Bond Cleavage in Deoxynucleotides: A Density Functional Study. *Can. J. Chem. 87*, 850−863.