JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Communication
pubs.acs.org/JACS

Replacement of 2′-Deoxycytidine by 2′-Deoxycytidine Analogues in the *E. coli* Genome

Angad P. Mehta, Han Li, Sean A. Reed, Lubica Supekova, Tsotne Javahishvili, and Peter G. Schultz*

The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States

Supporting Information

ABSTRACT: Several modified bases have been observed in the genomic DNA of bacteriophages, prokaryotes, and eukaryotes that play a role in restriction systems and/or epigenetic regulation. In our efforts to understand the consequences of replacing a large fraction of a canonical nucleoside with a modified nucleoside, we previously replaced around 75% of thymidine (T) with 5′-hydroxymethyl-2′-deoxyuridine (5hmU) in the *Escherichia coli* genome. In this study, we engineered the pyrimidine nucleotide biosynthetic pathway using T4 bacteriophage genes to achieve approximately 63% replacement of 2′-deoxycytidine (dC) with 5-hydroxymethyl-2′-deoxycytidine (5hmC) in the *E. coli* genome and approximately 71% replacement in plasmids. We further engineered the glucose metabolic pathway to transform the 5hmC into glucosyl-5-hydroxymethyl-2′-deoxycytidine (5-gmC) and achieved 20% 5-gmC in the genome and 45% 5-gmC in plasmid DNA.

Several modified bases such as N6-methyladenine (m^6A), 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), 5-carboxycytosine, and 2′-deoxy-5-(hydroxymethyl)uridine (5hmU) (Figure 1A) have been observed in the genomes of bacteriophages, prokaryotes, and eukaryotes.[1,2] m^6A and 5mC play an important role in restriction modification systems in bacteria.[3] In eukaryotes, including mammals, 5mC and 5hmC are stable epigenetic modifications, where 5mC plays a key role in regulating transcription. Cytosine modification−5hmC, which is most abundant in brain, has been shown to be enriched in the gene bodies of highly expressed genes; binding of MeCP2 to these gene bodies has been proposed to play a role in the organization of chromatin structure and in gene expression.[4] On the other hand, in the T4 bacteriophage 168 kb genome, a complete replacement of 2′-deoxycytidine (dC) with 5hmC is observed. Moreover, most of these 5hmC bases are *O*-glucosylated.[5−7] Inspired by the extent of these DNA modifications in bacteriophages, we asked whether we could achieve similar degrees of substitution in the *Escherichia coli* genome and better understand the consequences of the replacement of a large fraction of a nucleoside with a modified nucleoside.

As a first step, we had previously engineered the thymidine biosynthetic pathway to incorporate 5′-hydroxymethyl-2′-deoxyuridine (5hmU) in the *E. coli* genome and achieved approximately 75% replacement of thymidine (T) with 5hmU.[8]

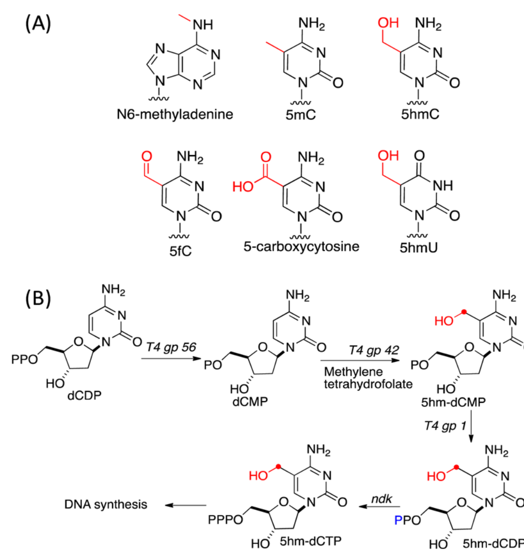

Figure 1. (A) Modified bases such as N6-methyladenine, 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), 5-carboxycytosine, and 2′-deoxy-5-(hydroxymethyl)-uridine (5hmU) have been observed in the genomes of either bacteriophages, prokaryotes, or eukaryotes. (B) Engineered metabolic network to biosynthesize 5hmC-incorporated DNA.

Similarly, thymidine replacement by feeding of 5-bromouracil or 5-chlorouracil has also been previously reported by Hanawalt and Mutzel, respectively.[9,10] Although dC analogues have been studied for their mutagenicity and transcriptional regulation[11−14] and in vitro incorporation of 5hmC and 5-gmC has been achieved in DNA templates,[15] to the best of our knowledge, replacement of a large fraction of dC with dC analogues in genomic DNA has not been studied. Herein we investigated the possibility of replacement of a large fraction of dC with 5hmC in the genomic DNA of *E. coli* using metabolic engineering (Figure 1B). We further engineered the glucose metabolic pathway to determine whether 5hmC in the *E. coli* genomic DNA can be efficiently converted to glucosyl-5-hydroxymethyl-2′-deoxycytidine (5-gmC).

To incorporate 5hmC into the *E. coli* genome, we engineered the *E. coli* DNA biosynthetic pathway using the following bacteriophage T4 genes: *gp 56*, phosphatase; *gp 42*, hydroxymethylase; *gp 1*, kinase; and *cd*, deaminase.[5,16−18] This set of genes along with *E. coli ndk* was expected to convert

Received: September 14, 2016
Published: October 20, 2016

© 2016 American Chemical Society

DOI: 10.1021/jacs.6b09661
J. Am. Chem. Soc. 2016, 138, 14230−14233

2′-deoxycytidine diphosphate (dCDP), an intermediate in 2′-deoxycytidine triphosphate (dCTP) biosynthesis in *E. coli*, to 2′-deoxy-5-hydroxymethylcytidine triphosphate (5hm-dCTP) (Figure 2). The resulting 5hm-dCTP was expected to compete with dCTP as endogenous DNA polymerase substrates for incorporation into the *E. coli* genome.

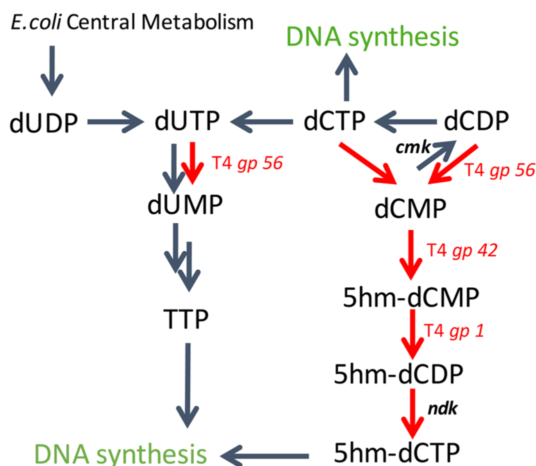

Figure 2. Engineered metabolic network for 5hmC incorporation in the genomic DNA.

The bacteriophage T4 genes *gp 56*, *gp 42*, *gp 1*, and *cd* were codon-optimized for expression in *E. coli* and were introduced as a part of a synthetic operon under control of the pBAD promoter on a plasmid imparting tetracycline resistance (pAM38). Defined spacers containing ribosomal binding sites (rbs's) were introduced between each of the genes for optimal expression (see Supporting Information (SI) for details of plasmid construction of pAM38). Plasmid pAM38 was transformed into *E. coli* DH10B (hmC-1 strain, see Table 1), and individual colonies were picked and inoculated into LB medium containing tetracycline. Cells were harvested after 18–20 h of growth, and the genomic DNA was isolated and digested with phosphodiesterase, calf intestinal phosphatase, and benzonase.[8]

Table 1. hmC and g-hmC Strain Details

strain name	parent strain of *E. coli*	plasmids transformed
hmC-1	DH10B	pAM38
hmC-2	DH10B	pAM37, pAM38
hmC-3	DH10B	pAM38, pAM39
hmC-4	DH10B	pAM38, pAM52
hmC-5	Δ*pyrF-E. coli* BW25113	pAM38, pAM39
g-hmC-1	DH10B	pAM52, pAM53
g-hmC-2	DH10B	pAM38, pAM39, pAM76
g-hmC-3	DH10B	pAM38, pAM39, pAM92

The resulting pool was analyzed by high-performance liquid chromatography (HPLC, 270 nm). Along with the standard genomic nucleosides, 5hmC was detected, and the 5hmC levels were calculated by comparison of the HPLC peak areas at 270 nm. Under these conditions around 25% of dC in the genomic DNA was replaced by 5hmC [5hmC/(5hmC + dC) ≈ 26%]. Similarly, we observed 5hmC in the plasmid DNA isolated from these strains [5hmC/(5hmC + dC) ≈ 20%, Figure S8].

To further increase 5hmC levels, we sought to increase the 5hm-dCTP pool. To this end, T4 bacteriophage genes *gp 56*, *gp 42*, and *gp 1* were introduced in a high copy number vector (CDF origin and chloramphenicol resistance, pAM37). pAM37 was co-transformed with pAM38 into *E. coli* DH10B (hmC-2 strain, see Table 1), which resulted in significantly fewer colonies; this may be due to excessive buildup of non-native nucleoside analogues (e.g., 5hm-dCMP) resulting in toxicity. The genomic DNA isolated from the surviving colonies had 5hmC/(5hmC + dC) ≈ 30%. Alternatively, lowering the dCTP levels may allow 5hmC to better compete with dC. Therefore, we designed a new construct on a high copy number vector with the native T4 bacteriophage genes *gp 56* and T4 *gp 1* (again with defined spacers and rbs's between the two genes), but lacking the T4 *gp 42* gene (pAM39, CDF origin). T4 gp 56 converts dCDP to dCMP and was expected to decrease the dCTP pools. Simultaneously, T4 gp 56 was also expected to increase dCMP pools, the substrate for T4 gp 42. T4 gp 1 was expected to drive the conversion of 5hm-dCMP to 5hm-dCTP for incorporation into genomic DNA. We co-transformed pAM38 and pAM39 into *E. coli* DH10B (hmC-3 strain), which resulted in significantly less toxicity as compared with pAM38-pAM37 co-transformation in *E. coli* DH10B. We observed colony to colony variation in the 5hmC content: 5hmC/(5hmC + dC) levels ranged from 48% to 53% (50 ± 2%) in genomic DNA (Figures 3, S9, and S10) and from 50% to 54% (52 ± 2%) in plasmid DNA (Figures S11–S13).

To further increase 5hmC levels in the genomic DNA, we sought to introduce selective pressure for 5hmC incorporation. T4 bacteriophage has endonucleases (DenB, endonuclease IV) that break down C-containing DNA.[19] We codon-optimized the gene for the major endonuclease DenB for *E. coli* expression and incorporated it into pAM39 to afford pAM52. We co-transformed pAM38 and pAM52 into *E. coli* DH10B (hmC-4 strain) and analyzed genomic and plasmid DNA for 5hmC content. 5hmC/(5hmC + dC) levels ranged from 51% to 57% (54 ± 3%) in genomic DNA (Figures 3, S14, and 15) and from 62% to 71% (65 ± 5%) in plasmid DNA (Figures 3, S16, and S17).

At this stage, we hypothesized that a steady-state may have been reached between dCTP biosynthesis and engineered 5hm-dCTP biosynthesis. We therefore asked whether further perturbation of dCTP biosynthesis at the genomic level could increase 5hmC content. Because CTP and CDP are the substrates for dCTP biosynthesis, we hypothesized that decreasing the CDP and CTP levels would enable us to lower the dCTP pools. To test this notion, we incorporated the 5hmC pathway into Δ*pyrF-E. coli* BW25113, which is auxotrophic for pyrimidine ribonucleosides (U or C). We initially transformed pAM38-pAM52 into Δ*pyrF-E. coli* BW25113, which resulted in severe toxicity; the surviving colonies showed low 5hmC incorporation (5hmC/(5hmC + dC) ≈ 25%). We then transformed pAM38-pAM39 into Δ*pyrF-E. coli* BW25113 (hmC-5 strains), which resulted in much lower toxicity, possibly due to less stringent selection pressure. Individual colonies were grown in minimal medium with low levels of U (0.5–5 mg/L). However, under these growth conditions we observed very low levels of 5hmC (5hmC/(5hmC + dC) ≈ 10% (Figure S18). Colonies of the hmC-5 strain were then grown in LB medium in the absence of exogenous pyrimidine nucleosides. An increase was observed in the genomic DNA 5hmC content, and 5hmC/(5hmC+dC) levels ranged from 54% to 63% (59 ± 4%) (Figures 3, S19, and S20). However, the 5hmC content in the plasmids dropped to 46% to 50% (47 ± 2%) (Figures S21 and S22). It appears that,

DOI: 10.1021/jacs.6b09661
J. Am. Chem. Soc. 2016, 138, 14230–14233

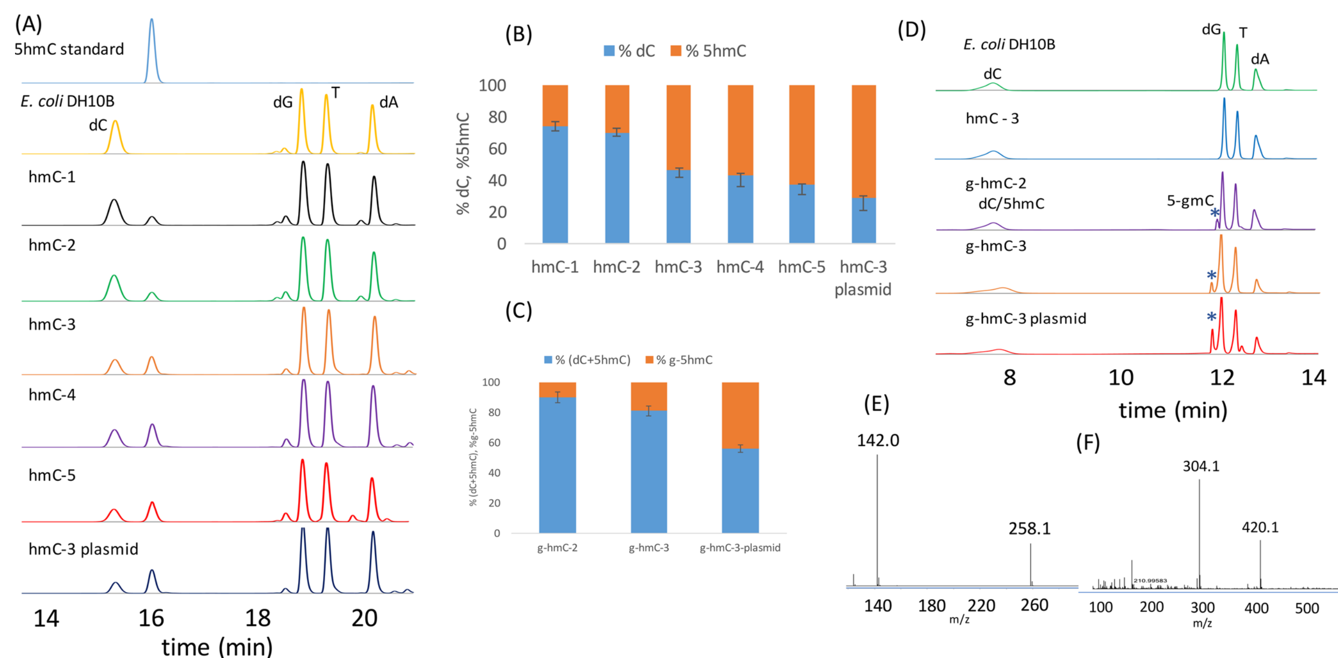

Figure 3. (A) HPLC (270 nm) traces showing standard 5hmC and genomic digestion products (single nucleosides) for *E. coli* DH10B and g-hmc-1 to g-hmC-3 strains. The bottom trace represents the digestion products for the plasmid isolated from hmC-3 strain, which shows maximum 5hmC content. (B) Comparison of 5hmC and dC levels, showing the maximal 5hmC content observed for each hmC strain. (C) Comparison of 5-gmC and (dC + 5hmC) levels, showing the maximal 5-gmC content observed for each hmC strain. (D) HPLC (270 nm) traces showing genomic digestion products (single nucleosides) for *E. coli* DH10B and hmC-1 to hmC-5 strains. The bottom trace represents the digestion products for the plasmid isolated from g-hmC-3 strain showing maximum 5-gmC content. (E) ESI-MS positive ion mode for 5hmC in the genomic digests (F) ESI-MS positive ion mode for 5-gmC in the genomic digests.

under these experimental conditions, we have reached a maximum level of 5hmC incorporation in *E. coli* genomic DNA (5hmC/(5hmC + dC) = 59 ± 4%) and plasmid DNA (5hmC/(5hmC + dC) = 65 ± 5%). It may be possible that this is the maximum 5hmC content in the genome that *E. coli* can tolerate without changing system level processes related to replication, transcription, and repair.

The incorporation of 5hmC into the genome provided a chemical handle for further modification of the genomic template. For example, in T4 bacteriophage, most 5hmC bases are *O*-glucosylated to give glucosyl-5-hydroxymethylcytosine (5-gmC).[5,20,21] Similarly, we sought to investigate whether 5hmC in the engineered *E. coli* genome could be *O*-glucosylated. The T4 bacteriophage gene *bgt* was codon-optimized for *E. coli* expression and was initially expressed in pAM38 to give pAM53. Plasmids pAM52 and pAM53 were co-transformed into *E. coli* DH10B (g-hmC-1 strain). Individual colonies were picked, the genome was digested, and the resulting pool was analyzed by LCMS for 5-gmC incorporation. Trace levels of 5-gmC were detected in the extracted ion chromatogram (EIC, 420.0 Da, Figure S23). We hypothesized that having T4 *bgt* induced along with 5hmC genes may lead to toxicity, and *bgt* should be induced at a later stage of growth. Hence, we designed a new construct where T4 *bgt* was incorporated on a vector with a T5 promoter/lac operon system (pAM76) to allow independent induction of T4 *bgt.* pAM76, pAM38, and pAM39 were co-transformed into *E. coli* DH10B (g-hmC-2 strain). The strain 5-gmC-2 was grown in LB medium followed by IPTG induction of *bgt* after 16 h of growth. Genomic digests followed by LCMS indicated the presence of 5-gmC by UV detection (270 nm) and corresponding EIC, 420.0 Da. Approximately 10% of dC was replaced by 5-gmC [5-gmC/(5-gmC + 5hmC+dC) ≈ 10% (9 ± 2%) for genomic templates (Figure 3D) and around 20% (18 ± 3%) for the plasmids Figures 3D and S25]. The presence of 5-gmC was confirmed by mass spectrometry analysis and LCMS co-injection/co-elution experiments (Figure S24). We speculated that the low levels of 5-gmC were due to the fact that UDP-glucose, the substrate for T4 bgt, is limiting. To increase the pools of UDP-glucose, we incorporated *E. coli* genes *pgm* and *galU* into pAM76 to afford pAM92. Plasmids pAM92, pAM38, and pAM39 were co-transformed into *E. coli* DH10B (g-hmC-3 strain), and then single colonies were grown in LB medium and induced with IPTG after 16−18 h of growth. Cells were harvested 8 h after induction, and the genomic and plasmid DNA digests were analyzed by LCMS. In the genomic templates 5-gmC/(5-gmC + 5hmC + dC) ranged from 17% to 21% (19 ± 2%) (Figures 3D, S26, and S27), whereas in the plasmids this ratio ranged from 38% to 45% (42 ± 3%) (Figures 3D, S28, and S29). LCMS analysis also revealed only trace levels of 5hmC in the genomic and plasmid templates, indicating that most of the 5hmC residues were *O*-glucosylated. This result suggests that although the engineered 5hmC metabolic network is able to incorporate much higher 5hmC levels (∼60%), the introduction of 5hmC glucosylation pathway may result in toxicity as 5-gmC incorporation may interfere with *E. coli* transcription and replication processes. This may limit 5hmC incorporation to about 21% in the genome.

To summarize, we have used a metabolic engineering approach to incorporate 5hmC and 5-gmC into the *E. coli* genome. Maximum 5hmC levels detected were 59 ± 4% and 65 ± 5% in the genomic templates and plasmids, respectively; maximum 5-gmC levels were 19 ± 2% in the genome and 42 ± 3% in the plasmids. At present we do not know if these are the maximum levels of 5hmC and 5-gmC that can be realized. We

DOI: 10.1021/jacs.6b09661
J. Am. Chem. Soc. 2016, 138, 14230−14233

are currently developing more stringent selection platforms (for example, using a combination of *denB-denA* nucleases along with T4 gene *alc*) along with random mutagenesis to further increase 5hmC levels. These studies may provide insights into the response of physiological regulators in *E. coli* due to cytosine modification on genomic templates that limit the replacement of dC with dC analogues. This work also creates a platform to study the effects of 5hmC on *E. coli* replication, transcription, and repair.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b09661.

HPLC data, experimental procedure, and design of the constructs, including Figures S1−S29 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*schultz@scripps.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge Kristen Williams for her assistance in manuscript preparation. This work is supported by NIH R01 GM062159-14.

■ REFERENCES

(1) Nabel, C. S.; Manning, S. A.; Kohli, R. M. *ACS Chem. Biol.* **2012**, *7*, 20.
(2) Olinski, R.; Starczak, M.; Gackowski, D. *Mutat. Res., Rev. Mutat. Res.* **2016**, *767*, 59.
(3) Sanchez-Romero, M. A.; Cota, I.; Casadesus, J. *Curr. Opin. Microbiol.* **2015**, *25*, 9.
(4) Münzel, M.; Globisch, D.; Carell, T. *Angew. Chem., Int. Ed.* **2011**, *50*, 6460.
(5) Miller, E. S.; Kutter, E.; Mosig, G.; Arisaka, F.; Kunisawa, T.; Ruger, W. *Microbiol Mol. Biol. Rev.* **2003**, *67*, 86.
(6) Kutter, E. M.; Wiberg, J. S. *J. Virol.* **1969**, *4*, 439.
(7) Kutter, E.; Gachechiladze, K.; Poglazov, A.; Marusich, E.; Shneider, M.; Aronsson, P.; Napuli, A.; Porter, D.; Mesyanzhinov, V. *Virus Genes* **1995**, *11*, 285.
(8) Mehta, A. P.; Li, H.; Reed, S. A.; Supekova, L.; Javahishvili, T.; Schultz, P. G. *J. Am. Chem. Soc.* **2016**, *138*, 7272.
(9) Hanawalt, P. C. *Methods in Enzymology*; Academic Press: San Diego, 1967; Vol. *12*, Part A, p 702.
(10) Marlière, P.; Patrouix, J.; Döring, V.; Herdewijn, P.; Tricot, S.; Cruveiller, S.; Bouzon, M.; Mutzel, R. *Angew. Chem., Int. Ed.* **2011**, *50*, 7109.
(11) Jackson-Grusby, L.; Laird, P. W.; Magge, S. N.; Moeller, B. J.; Jaenisch, R. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 4681.
(12) Michalowsky, L. A.; Jones, P. A. *Mol. Cell. Biol.* **1987**, *7*, 3076.
(13) Negishi, K.; Loakes, D.; Schaaper, R. M. *Genetics* **2002**, *161*, 1363.
(14) Negishi, K.; Williams, D. M.; Inoue, Y.; Moriyama, K.; Brown, D. M.; Hayatsu, H. *Nucleic Acids Res.* **1997**, *25*, 1548.
(15) de Kort, M.; de Visser, P. C.; Kurzeck, J.; Meeuwenoord, N. J.; van der Marel, G. A.; Rüger, W.; van Boom, J. H. *Eur. J. Org. Chem.* **2001**, *2001*, 2075.
(16) Graves, K. L.; Butler, M. M.; Hardy, L. W. *Biochemistry* **1992**, *31*, 10315.
(17) Graves, K. L.; Hardy, L. W. *Biochemistry* **1994**, *33*, 13049.
(18) Song, H. K.; Sohn, S. H.; Suh, S. W. *EMBO J.* **1999**, *18*, 1104.
(19) Carlson, K.; Oevervatn, A. *Genetics* **1986**, *114*, 669.
(20) Moréra, S.; Imberty, A.; Aschke-Sonnenborn, U.; Rüger, W.; Freemont, P. S. *J. Mol. Biol.* **1999**, *292*, 717.
(21) Kornberg, J. J. A. *J. Biol. Chem.* **1962**, *237*, 1968.

DOI: 10.1021/jacs.6b09661
J. Am. Chem. Soc. 2016, 138, 14230−14233