technical issues of employing the domain swapping approach will certainly corroborate the merits of biological glycodiversification, and enable the practical and scale-up production of glycoconjugates free of the limitations imposed by nature.

REFERENCES

Demchenko, A.V., ed. (2008). Handbook of Chemical Glycosylation: Advances in Stereoselectivity and Therapeutic Relevance (Weinheim, Germany: Wiley-VCH Verlag GmbH & Co.).

Dwek, R.A. (1996). Chem. Rev. *96*, 683–720.

Langenhan, J.M., Peters, N.R., Guzei, I.A., Hoffmann, F.M., and Thorson, J.S. (2005). Proc. Natl. Acad. Sci. USA *102*, 12305–12310.

Melançon, C.E., Thibodeaux, C.J., and Liu, H.-w. (2006). ACS Chem. Biol. *1*, 499–504.

Thibodeaux, C.J., Melancon, C.E., and Liu, H.-w. (2007). Nature *446*, 1008–1016.

Truman, A.W., Dias, M.V.B., Wu, S., Blundell, T.L., Huang, F., and Spencer, J.B. (2009). Chem. Biol. *16*, this issue, 676–685.

Yang, J., Hoffmeister, D., Liu, L., Fu, X., and Thorson, J.S. (2004). Bioorg. Med. Chem. *12*, 1577–1584.

Zhang, C., Griffith, B.R., Fu, Q., Albermann, C., Fu, X., Lee, I.-K., Li, L., and Thorson, J.S. (2006). Science *313*, 1291–1294.

Oxygenase Catalyzed 5-Methylcytosine Hydroxylation

Christoph Loenarz[1] and Christopher J. Schofield[1,*]
[1]Chemistry Research Laboratory and The Oxford Centre for Integrative Systems Biology, University of Oxford, 12 Mansfield Road, Oxford, OX1 3TA, UK
*Correspondence: christopher.schofield@chem.ox.ac.uk
DOI 10.1016/j.chembiol.2009.06.002

Recent reports identify the oxygenase catalyzed production of 5-hydroxymethylcytosine as a modification to mammalian DNA (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). This discovery has potential far-reaching implications for epigenetic regulation and will stimulate efforts to identify new nucleic acid modifications.

Modifications to genomic DNA include methylation at the 5-position of cytosine (C) bases within cytosine-guanine dinucleotide (CpG) sites. Production of 5-methylcytosine (5mC) alters DNA structure without interfering with base pairing, and is recognized as an important epigenetic modification. Active gene promoters generally have a low abundance of 5mC, because promoter methylation can cause transcriptional repression. The formation of 5mC is catalyzed by DNA methyltransferases (DNMTs) and regulates gene expression patterns and initiates chromatin remodeling.

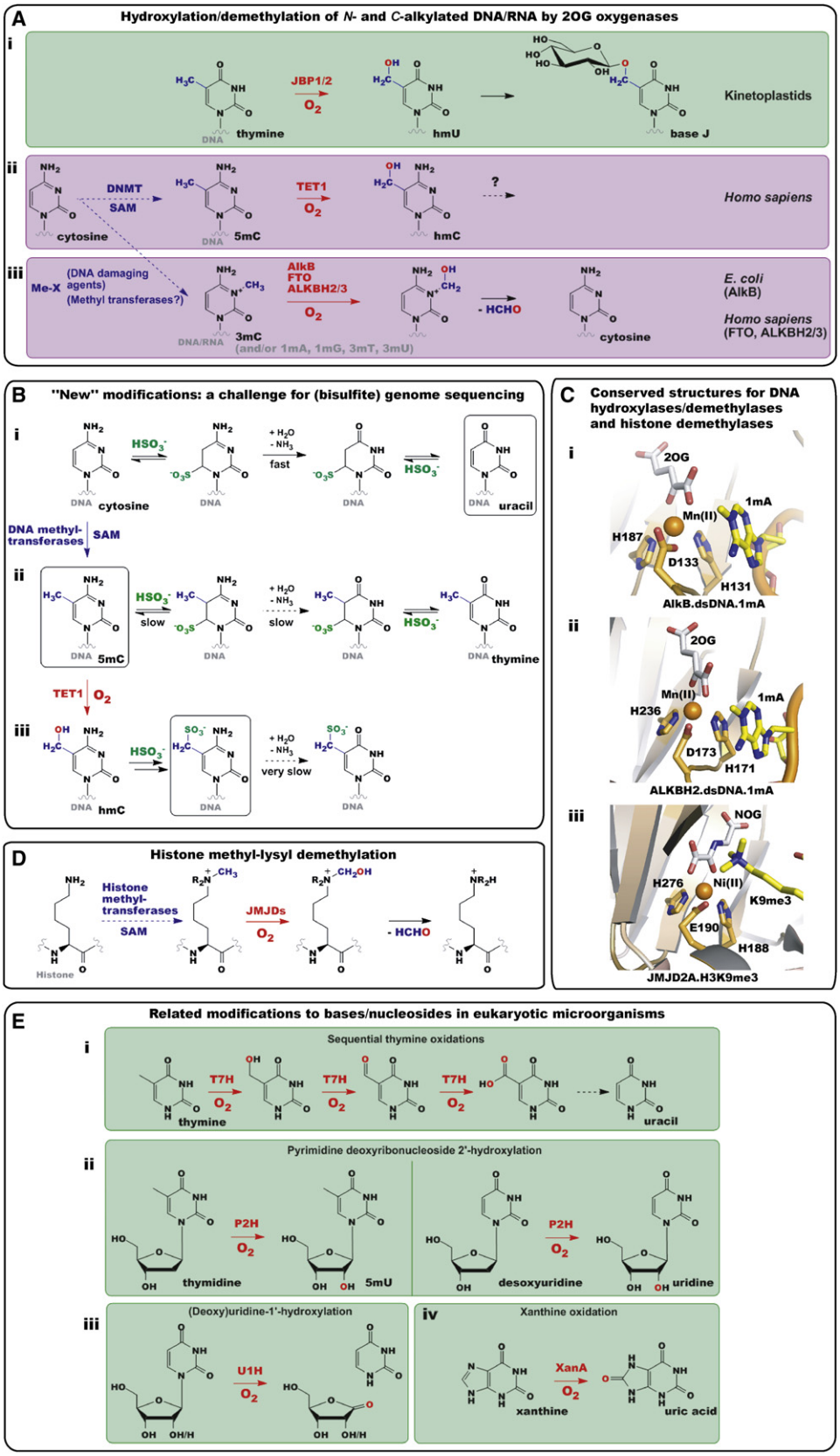
While plants employ pathways involving 5mC glycosylases and base excision repair to reverse cytosine methylation, attempts at identifying enzymes with 5mC reversal or modification activity in metazoans have been extensive but largely fruitless (reviewed in Ooi and Bestor [2008]). However, in recent years, other epigenetic modifications once believed to be static, such as histone lysyl methylation, have been demonstrated to be reversible.

The starting point that led Tahiliani and colleagues (Tahiliani et al., 2009) to the discovery of 5-hydroxymethylcytosine (hmC) as a modification to mammalian DNA was the identification of a previously unknown member of the 2-oxoglutarate (2OG) oxygenase family. 2OG oxygenases couple the two-electron oxidation of substrate to the oxidative decarboxylation of 2OG, and depend on Fe(II) as a cofactor. While bacterial 2OG oxygenases catalyze an extraordinarily wide range of oxidative chemistry including epimerization, ring closure, desaturation, and halogenation reactions, the activities of identified metazoan 2OG oxygenases have so far been limited to hydroxylation and demethylation via hydroxylation (Loenarz and Schofield, 2008).

Human 2OG oxygenases are involved in a diverse range of biological functions, including histone demethylation, collagen stabilization, DNA repair, hypoxia sensing, and fatty acid metabolism (Loenarz and Schofield, 2008). Although the catalytic domains of 2OG oxygenases share a common, albeit sometimes distorted, double-stranded β-helix (or jelly-roll) fold (first observed in antibiotic biosynthesis enzymes), they often have little overall sequence conservation, which impedes bioinformatic studies. However, approaches based on structurally informed sequence searches starting from 2OG oxygenases with known functions have led to the discovery of human orthologs, including enzymes with important biomedical functions.

In 1993, Borst and coworkers discovered that the DNA of African trypanosomes, protist parasites that cause African sleeping sickness, contained the unusual base β-D-glucosylhydroxymethyl-uracil, termed base J (Figure 1Ai). Their efforts to understand the nature of this modification and its role in expression led to the proposal of two trypanosomatid J-binding proteins JBP1 and JBP2 as thymidine hydroxylases of the 2OG oxygenase family (Yu et al., 2007). Starting from structurally informed iterative sequence profile searches using the JBP1 and JBP2 oxygenase domains, bioinformatic work from Tahiliani et al. (2009) led to the prediction of the three human enzymes TET1-3 as Fe(II) and 2OG oxygenases that catalyze 5-methylcytosine hydroxylation (Figure 1Aii). TET1-3 appear to be part of the wider TET/JBP fmaily, present minimally in metazoans, fungi, and algae.

CHBIOL 1527_1528


A
Hydroxylation/demethylation of N- and C-alkylated DNA/RNA by 2OG oxygenases
i
thymine
JBP1/2
O2
hmU
base J
Kinetoplastids
ii
cytosine
DNMT
SAM
5mC
TET1
O2
hmC
?
Homo sapiens
iii
Me-X
(DNA damaging agents)
(Methyl transferases?)
3mC
(and/or 1mA, 1mG, 3mT, 3mU)
DNA/RNA
AlkB
FTO
ALKBH2/3
O2
- HCHO
cytosine
E. coli
(AlkB)
Homo sapiens
(FTO, ALKBH2/3)
B
"New" modifications: a challenge for (bisulfite) genome sequencing
i
cytosine
HSO3-
fast
uracil
DNA methyl-transferases
SAM
ii
5mC
slow
thymine
TET1
O2
iii
hmC
very slow
C
Conserved structures for DNA hydroxylases/demethylases and histone demethylases
i
2OG
Mn(II)
1mA
H187
D133
H131
AlkB.dsDNA.1mA
ii
H236
D173
H171
ALKBH2.dsDNA.1mA
iii
NOG
H276
Ni(II)
K9me3
E190
H188
JMJD2A.H3K9me3
D
Histone methyl-lysyl demethylation
Histone
Histone methyl-transferases
SAM
JMJDs
O2
- HCHO
E
Related modifications to bases/nucleosides in eukaryotic microorganisms
i
Sequential thymine oxidations
T7H
O2
thymine
uracil
ii
Pyrimidine deoxyribonucleoside 2'-hydroxylation
P2H
thymidine
5mU
desoxyuridine
uridine
iii
(Deoxy)uridine-1'-hydroxylation
U1H
iv
Xanthine oxidation
XanA
xanthine
uric acid

Transfection of HEK293 cells with HA-tagged TET1 reduced 5mC levels in a statistically significant manner, and this effect was not observed with HA-TET1 mutated to impair the predicted Fe(II) binding sites; however, the antibody staining intensity of 5mC appeared unreliable. To investigate the effect of TET1 overexpression on 5mC levels more accurately, the authors went on to analyze restriction enzyme-cleaved genomic DNA, resolved using thin-layer chromatography (TLC). An additional unidentified species was observed and was associated with a decrease in 5mC abundance in DNA from cells expressing wild-type, but not mutant, TET1. The novel nucleotide was unambiguously identified as hmC by high-resolution mass spectrometry, including MS-MS fragmentation experiments comparing authentic hmC prepared from non-glucosylated DNA of T4 phage. Direct involvement of the TET1 catalytic domain in hmC production was ascertained by insect-cell purified recombinant protein using fully methylated double-stranded DNA oligonucleotides. Similar to other 2OG oxygenases, catalytic activity was absolutely dependent on both Fe(II) and 2OG. Interestingly, thymine conversion to hmU was not detectable in initial experiments, suggesting TET1 may be specific for 5mC. It will be important to analyze how this specificity is achieved.

The crucial question of whether hmC is present in mammalian DNA at physiologically relevant levels was addressed by Tahiliani et al. (2009) in different cell types. While no significant levels of hmC were found in some cell types, hmC was clearly present in human embryonic stem (ES) cells. Important evidence for the physiological relevance of hmC comes from a simultaneously published report. The work of Kriaucionis and Heintz (2009) provides important insights into the potential significance of hmC. Their work is a nice example of how recognition of the unexpected can lead to an important discovery. When investigating levels of 5-methylcytosine in the nuclei of cerebellar cells (Purkinje and granule cells, the latter making up > 50% of all neurons in the central nervous system), they detected the presence of an unexpected nucleotide. Careful chromatographic and mass-spectrometric analyses identified the nucleotide as hmC. Importantly, both studies find, at least in some cell types, hmC as a significant fraction of the total mC content. In DNA from Purkinje cells, hmC is present at ~40% of the abundance of mC (Kriaucionis and Heintz, 2009) and as 4% to 6% of all observable cytosine species within CpG sites in ES cells (Tahiliani et al., 2009). Tahiliani et al. (2009) estimate the frequency of hmC at ~1 in every 3000 nucleotides, a significant fraction, especially if it is a regulatory modification.

The presence of hmC as a stable base in genomic DNA also raises questions as to the validity of the widely used "bisulfite" method for the determination of 5mC levels during DNA sequencing. Bisulfite treatment enables 5mC to be distinguished from cytosine, because 5mC produces an intermediate that is more stable with respect to deamination, whereas cytosine is rapidly converted to uracil (Figure 1B). However, there is evidence that the analogous intermediate formed by reaction of hmC with bisulfite is more stable than that formed from 5mC (Hayatsu and Shiragami, 1979). Further, the product of the reaction with hmC, cytosine 5-methylenesulfonate (or related modifications), may directly interfere with sequencing. Thus, the results of the numerous studies based on bisulfite sequencing may need to be reassessed. The development of efficient sequencing methodology that simultaneously distinguishes hmC, 5mC, and C is now an important challenge. Tahiliani et al. (2009) also point out that the presence of hmC affects the selectivity of methylation-specific restriction endonucleases, and they argue that the results obtained using these reagents need to be reassessed. An analogous limitation applies to the use of proteases in proteomic analyses; for example, trypsin-mediated cleavage at lysyl residues is hindered by acetylation.

In recent years, advances in analytical techniques, coupled with an enhanced realization of the biomedical importance of post-translational modifications, have led to the identification of many new modifications to proteins. Recent work has suggested that 2OG oxygenases catalyze many post-translational modifications. Following the identification of their role in metazoan oxygen sensing,

Figure 1. Involvement of 2OG Oxygenases in Nucleic Acid and Nucleoside Modifications

(A) 2OG oxygenase catalyzed hydroxylations (enzyme acronyms in red) are coupled to the conversion of 2OG and O_2 to CO_2 and succinate. Both *N*-and *C*-alkylated DNA/RNA are hydroxylated by 2OG oxygenases: (i) thymidine 5-hydroxylation is catalyzed by JBP1/2 in kinetoplastids during base J biosynthesis (Yu et al., 2007); (ii) 5-hydroxymethylcytosine (hmC) is formed as a stable base by human TET1 and might be subject to subsequent modifications; (iii) reversal of *N*-alkylation damage to DNA/RNA by hydroxylation proceeds via fragmentation of the unstable intermediate. FTO (3mT, 1mA, 3mC and 3mU RNA) and ALKBH3 (1mA and 3mC) prefer ssDNA, while ALKBH2 (1mA and 3mC) prefers dsDNA; AlkB (1mA, 3mC, 1mG, 3mT) prefers ssDNA over dsDNA and RNA; it can also remove ethyl- and etheno-DNA lesions (Gerken et al., 2007; Sedgwick et al., 2007).

(B) Bisulfite sequencing (boxed: modification present at the end of a typical bisulfite reaction) converts cytosine to uracil (i) while retaining 5mC (ii), thus enabling identification of 5mC sites by sequencing. (iii) Under these conditions, the newly identified base hmC would react to cytosine 5-methylenesulfonate, which is likely to interfere with subsequent analyses. Investigation of the stability of the different bases towards deamination resulted in approximate half-lives of 6 min, 8 hr, and 28 hr for C, 5mC, and hmC, respectively (Hayatsu and Shiragami, 1979).

(C) Active sites of 2OG oxygenases involved in DNA and histone hydroxylation/demethylation: (i) AlkB, the DNA demethylase from *E. coli* involved in DNA repair (Protein Data Bank (PDB) code 3BIE); (ii) human ALKBH2, involved in reversal of DNA *N*-alkylation (PDB code 3BUC); (iii) human jumonji domain-containing protein 2A (JMJD2A), involved in demethylation of methyl-lysyl histone residues (PDB code 2OQ6). Color code: grey-brown, double-stranded β-helix fold; yellow, substrate; orange, Fe(II)-binding residues; white, 2OG or *N*–oxalylglycine (NOG). Note in these structures Ni(II) and Mn(II) substitute for Fe(II).

(D) A large subfamily of 2OG oxygenases is involved in the reversal of histone lysyl-methylation using chemistry analogous to that of DNA demethylases.

(E) Other base/nucleoside modifications known to be catalyzed by 2OG oxygenases in eukaryotes (Simmons et al., 2008): (i) thymine 7-hydroxylase (T7H) from *Neurospora crassa* catalyzes sequential thymine oxidation as part of the pyrimidine salvage pathway; (ii) pyrimidine deoxyribonucleoside 2′-hydroxylation by P2H has been shown in *N. crassa, Aspergillus nidulans,* and *Rhodotorula glutinis*; (iii) 1′-hydroxlation of both deoxyuridine and uridine by U1H in *R. glutinis*; (iv) xanthine hydroxylation by XanA in *A. nidulans*. The requirement by oxygenases for 2OG, Fe(II), and the coproducts succinate and CO_2, is omitted for clarity. Color codes: red, hydroxylation reactions and incorporated oxygen; blue, methylation reactions and methyl residue acted upon. Abbreviations: dsDNA, double-stranded DNA; K9me3, histone trimethyllysyl-9; 1mA, 1-methyladenine, 1mG, 1-methylguanine; 3mC, 3-methylcytosine; 3mT, 3-methylthymine; 3mU, 3-methyluracil; 5mU, 5-methyluracil; 5mC, 5-methylcytosine; hmC, 5-hydroxymethylcytosine; hmU, 5-hydroxymethyluracil; SAM, *S*-adenosylmethionine; ssDNA, single-stranded DNA.

CHBIOL 1527_1528

2OG oxygenases have been shown to catalyze the post-translational hydroxylation of proteins containing the ubiquitous ankyrin repeat domain, providing evidence that intracellular protein hydroxylation is common (Cockman et al., 2006). 2OG oxygenases (the Jmj subfamily) have also emerged as the largest family of histone methyl-lysyl demethylases (Figures 1C and 1D). Presently, it is unclear whether the plethora of post-translational modifications will be reflected in modifications to nucleic acids. However, the recent reports on hmC will both further stimulate systematic searches for such modifications and alert the scientific community to their potential importance.

The overall extent to which DNA/RNA-modifying 2OG oxygenases, such as TET1, play a role in nucleic acid modifications is unclear (as is whether or not they are involved in oxygen sensing). Members of the human AlkB homolog family (ALKBH1-8) are believed to enable repair of DNA/RNA damage by base *N*-methylation via a hydroxylation fragmentation mechanism that is analogous to that of the histone demethylases (Figures 1Aiii, 1C, and 1D) (Sedgwick et al., 2007). The fat mass and obesity associated protein (FTO) is related to the ALKBHs, and catalyzes the *N*-demethylation of 3-methylthymine oligonucleotides (Figure 1A) (Gerken et al., 2007). Mutations associated with FTO are linked to an increased risk of obesity revealing that at least one nucleic acid modifying 2OG oxygenase is important from a pathophysiological perspective.

Identifying the roles of the hmC modification at molecular, cellular, and physiological levels is now an important objective. One possibility is that 5-hydroxylation enables further modifications, as occurs in base J production (Figure 1Ai); intriguingly, the hmC hydroxyl group is known to be exposed and accessible to external reagents, even within double-stranded DNA (Hayatsu and Shiragami, 1979). A recent study reports the direct reversible conversion of hmC within double-stranded DNA to unmodified cytosine and free formaldehyde, via covalent catalysis by a bacterial DNMT lacking cofactors (Liutkeviciute et al., 2009). This work thus suggests a possible (reversible) mechanism for removal of the hmC hydroxylation mark. Another possibility is that further oxidation at the cytosine 5-position occurs, either to enable chemical demethylation or to reverse the function of cytosine 5-methylation. Precedents for this chemistry are modifications to bases/nucleosides catalyzed by 2OG oxygenases in micro-organisms (Figure 1E). It may be that some of the other base/nucleoside modifications catalyzed by 2OG oxygenases in micro-organisms, including deoxyribose 2′-hydroxylation, also occur in animals.

REFERENCES

Cockman, M.E., Lancaster, D.E., Stolze, I.P., Hewitson, K.S., McDonough, M.A., Coleman, M.L., Coles, C.H., Yu, X., Hay, R.T., Ley, S.C., et al. (2006). Proc. Natl. Acad. Sci. USA *103*, 14767–14772.

Gerken, T., Girard, C.A., Tung, Y.-C.L., Webby, C.J., Saudek, V., Hewitson, K.S., Yeo, G.S.H., McDonough, M.A., Cunliffe, S., McNeill, L.A., et al. (2007). Science *318*, 1469–1472.

Hayatsu, H., and Shiragami, M. (1979). Biochemistry *18*, 632–637.

Kriaucionis, S., and Heintz, N. (2009). Science *324*, 929–930.

Liutkeviciute, Z., Lukinavicius, G., Masevicius, V., Daujotyte, D., and Klimasauskas, S. (2009). Nat. Chem. Biol. *5*, 400–402.

Loenarz, C., and Schofield, C.J. (2008). Nat. Chem. Biol. *4*, 152–156.

Ooi, S.K.T., and Bestor, T.H. (2008). Cell *133*, 1145–1148.

Sedgwick, B., Bates, P.A., Paik, J., Jacobs, S.C., and Lindahl, T. (2007). DNA Repair (Amst.) *6*, 429–442.

Simmons, J.M., Muller, T.A., and Hausinger, R.P. (2008). Dalton Trans. *38*, 5132–5142.

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). Science *324*, 930–935.

Yu, Z., Genest, P.-A., ter Riet, B., Sweeney, K., DiPaolo, C., Kieft, R., Christodoulou, E., Perrakis, A., Simmons, J.M., Hausinger, R.P., et al. (2007). Nucleic Acids Res. *35*, 2107–2115.

CHBIOL 1527_1528