Mechanism and Control in Biological Amine Methylation

by **G. Michael Blackburn***[a]), **Steven J. Gamblin**[b]), and **Jonathan R. Wilson**[b])

[a]) Krebs Institute, Chemistry Department, Sheffield University, Brook Hill, Sheffield S3 7HF, UK (fax +441142229346; e-mail: g.m.blackburn@shef.ac.uk)

[b]) National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

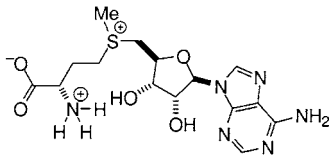
This paper is submitted in appreciation of the great work and warm friendship of Professor *Duilio Arigoni* on the occasion of his 75th Birthday.

Biological methylation is a subject that has fascinated mechanistically minded chemists for over 50 years. While early studies were usually directed at *C*-methylation in natural products, more recent work on *N*-methylation in DNA and proteins is being supported by the results of X-ray crystallography. From this source, significant mechanistic detail can be gleaned and powerful insights gained into the nature of enzyme catalysis and selectivity in methyl-transfer processes. The case of the human histone H3 transmethylase SET7/9 is considered in detail and compared to cognate histone lysine methylases. It provides an analysis of Nature's solution to the task of avoiding over-methylation.

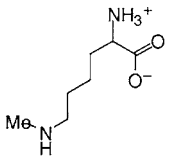
Introduction. – Biological methylation has attracted the attention of mechanistically minded bioorganic chemists for many years. Initially, prime attention was focussed on *C*-methylation processes in secondary metabolism, and these matured eventually into studies on primary metabolites, and especially of vitamin B_{12} [1]. In biochemical terms, such methylation reactions belong to the transferase class of enzymes[1]) with the transfer of methyl (EC 2.1.1), hexosyl (EC 2.4.1), pentosyl (EC 2.4.2), and prenyl (EC 2.5) moieties dominating early examples of substitution patterns at the saturated C-atom. While Me transfers were relatively rare examples of enzyme action in the 1960's, invariably involving *S*-adenosylmethionine (AdoMet; **1**) as the cofactor, the methylation of biological polymers burst on the scene in the late 1960's. *ε-N*-Methyllysine (**2**) was identified in bacterial flagellae in 1966 [2] and *ε-N,N*-dimethyllysine was found soon after [3]. At the same time, *ε-N,N,N*-trimethyllysine (**3**) was found in histones [4] as well as in cytochrome b [5] and skeletal myosin [6]. Even this burgeoning field became swamped by the sheer weight of studies on DNA methylation. At first, such investigations focused on viral DNA methylation as a feature of restricting viral DNA cleavage by bacterial host restriction endonucleases [7]. Today, DNA base-methylation is linked to ageing, cancer, gene expression, nerve development, sexual conflict, and many other normal and abnormal functions of living organisms, with over 2000 reviews on the subject! By contrast, histone methylation is a much smaller research area with only 46 reviews going back to 1974, though recent reports suggest that these two fields are biologically interrelated [8].

---

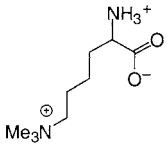
[1]) *Enzyme Commission* classification, 'Recommendations of the IUB on Nomenclature and Classification of Enzymes', Elsevier, Amsterdam, London, New York, 1965



1



2



3

It is not surprising that much attention is now focussed on the mechanism of methylation of C, N, and O in biological species, and that, at best, such studies seek to gain the same degree of understanding that chemists have achieved for nucleophilic substitution reactions in solution.

Substitution reactions were one of the first classes of chemical reactions to be analysed in mechanistic detail by a combination of stereochemical [9], kinetic [10], and structural analysis [11]. The superficial simplicity of substitution reactions of alkyl halides has resulted in their presentation at the beginning of the discussion of heterolytic reaction mechanisms in the majority of organic chemistry student texts as well as in some more-advanced works [12], though there is at least one notable exception to this general pattern [13]. It is perhaps necessary to recall at this juncture that the original *Hughes–Ingold* terminology for nucleophilic substitution reactions [14] was based on kinetic analysis: S_N1 reactions are *first-order* while S_N2 reactions are *second-order* processes. Nonetheless, contemporary discussion has focused rather more on whether such processes are dissociative (*D*, S_N1) or associative (*A*, S_N2) as definitively denoted by a precise nomenclature [15] (not used here).

In stereochemical terms, S_N2 reactions are characterised by *inversion* of configuration at C while S_N1 processes generally show *racemisation*. Building on the pioneering studies of *Kenyon* and *Phillips* [9], stereochemistry came to be applied to biological methylation processes as a result of the creation of the synthesis and stereochemical analysis of the chiral Me group by *Arigoni* and co-workers [16] and *Cornforth et al.* [17]. Much later, the stereochemistry of enzyme-catalysed Me transfer to C [18], N [19], and O [20] was shown by *Floss* and co-workers to involve inversion of configuration.

Kinetic criteria, media effects, *etc.* used to discriminate between associative and dissociative substitution reactions at the saturated C-atom in solution can only rarely be applied to enzymic reactions largely because the association of substrates in the protein catalytic site usually overrides such analysis. Therefore, structural studies on enzymes have been pursued as a means of understanding reaction mechanisms, with varying degrees of success. At their best, insights gained from protein crystallography can be backed up by studies with chemically modified substrates and quantum-mechanical methods to provide a definitive mechanism for some processes. From our own work, the hydrolytic cleavage of the glycosylic bond in an aberrant deoxyuridine residue in a DNA duplex by the repair enzyme uracil DNA glycosylase is a case in point. Both structural and computational analysis identify the cleavage as being dissociative in character, involving the transient formation of an oxocarbenium ion and a uracil anion [21] that is driven by remarkable electrostatic stabilisation by the substrate itself [22]. That analysis has been amply endorsed by *Stivers* and *Liang* [23] by using primary and

secondary kinetic isotope effects and through the device of phosphate replacement by uncharged methylphosphonates.

Biological Methylation in Nucleic Acids. – In nucleic acids, *N*-methylation occurs at both sp^2- and sp^3-N-atoms in the heterocyclic bases, as well as at C(5) of cytosine (**4**). In DNA, this involves a base-flipping process [24] to allow enzyme access to the substrate base. *S*-Adenosyl-L-methionine (AdoMet; **1**) is the standard source of the Me group and, for N^6-methylation of adenine, the reaction proceeds with inversion of configuration [25]. While several enzyme structures have been described, there is as yet no ternary complex involving two reactants or two products. A structure of a T250G mutant of the restriction methylase HhaI with a DNA duplex and AdoHcy (**5**) gives a near-complete picture of the *C*-methylation reaction (*Fig. 1*) [26]. Methylation of cytosine (**4**) at C(5) clearly identifies suprafacial interaction with the cofactor **1** suggested to involve reversible covalent addition at C(6) with the eventual release of *S*-adenosyl-L-homocysteine (AdoHcy; **5**) [27]. The 4-Å separation of the S-atom from cytosine leaves just the right space to insert an *S*-Me group in *Van der Waals* contact with C(5). Nonetheless, the nature of the transition state for methylation at the N-atom in adenine and cytosine remains unresolved and key questions such as '*Is methylation directed at in-plane sp^2 lone pairs or out-of-plane π-electrons?*' or '*Is general acid-base or covalent catalysis operative?*' remain unanswered.

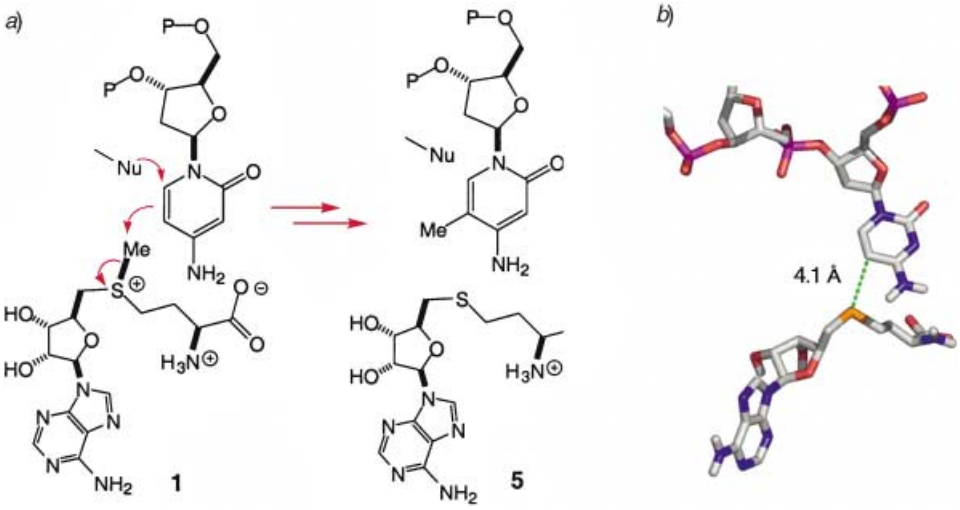
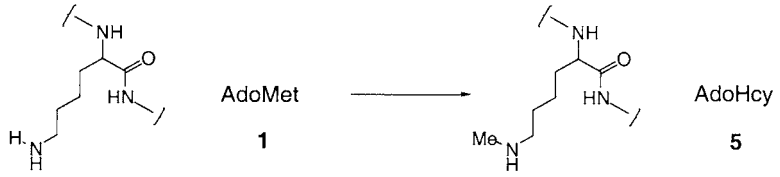


Fig. 1. a) *Mechanism of the* C(5)*-methylation of cytosine in DNA by HhaI with AdoMet* (**1**) *as cofactor*. b) *Part structure of a ternary complex of a mutant HhaI with AdoHcy* (**5**) *and a 'flipped-out' cytosine from a decanucleotide substrate* [26].

Protein *N*-methylation. – The major target for protein methylation is the ε-amino group of lysine with lesser occurrence of arginine methylation. Such lysine methylation has been shown to involve inversion of configuration for the transfer of the Me group from AdoMet (**1**) to the lysine ε-amino group [19] (*Scheme*). After 37 years since the

discovery of mono-*N*′-methyllysine in proteins [2], there is a current flurry of activity because of the profound significance of ε-*N*,*N*,*N*-trimethyllysine (**3**) in signalling and in the structure of chromatin [28]. Throughout the eukaryotes, chromatin structure is regulated through the *N*-methylation of specific lysines in the N-terminal tails of histones H3 and H4, while plants regulate the function of *Rubisco* by methylation of Lys^{14} in the N-terminus of the large subunit. It is claimed that certain inherited neurological disorders are linked to mutations in genes that regulate DNA methylation and alterations in DNA; while protein methylation and/or acetylation has been documented in studies of age-related neurodegenerative disorders including *Alzheimer*'s disease, *Parkinson*'s disease, and *Huntington*'s chorea [29]. Such mono-, di-, and trimethylation is carried out by specific enzymes, whose mechanism of action is now becoming fascinatingly clear as a result of recent protein structure determination.

Scheme



The best insight to date emerges from a study of the human histone H3 methyltransferase SET7/9. This protein transfers a single Me group from AdoMet (**1**) to Lys^4 of histone H3. The target lysine reaches the AdoMet (**1**) by inserting its lysyl side chain into a narrow channel that connects the two faces (*Fig. 2*). The structure of a ternary complex of this protein with AdoHcy (**5**) and a histone peptide containing an ε-*N*-methyllysine has been solved to 1.7-Å resolution [30]. This shows that the peptide substrate and the cofactor bind on opposite faces of the enzyme. The molecular detail reveals clearly how this enzyme catalyses the S_N2 methylation process and achieves the control necessary to limit reaction to a mono-methylation process. First, the lysine Me group is 3.52 Å from the AdoHcy S-atom, putting it at *Van der Waals* separation. It is exactly in line between the N- and S-atoms (177°) and directed at the *si*-face of the S-atom. Finally, the H-atoms at ε-N are H-bonded to a H_2O molecule W1 and the O-atom of Tyr^{245} in a network of unambiguous pairings (*Fig. 3*, *right*). This relative geometry should be much the same for the reactants since a binary complex of SET7/9 with AdoMet (**1**) [31] maps almost directly onto the position of AdoHcy (**5**) in the ternary complex and with the *S*-Me group pointing to the site occupied by the lysine-ε-N-atom in the ternary complex [30]. This necessarily means that the Lys^4 ε-N-atom lone-pair must be directed at the Me group of AdoMet in the reactant complex. Finally, the whole reaction site is dominated by four tyrosine OH groups (Tyr^{245}, Tyr^{305}, Tyr^{335}, Tyr^{337}) and five main-chain carbonyl groups, all approximately oriented towards the lysine amine group. The reaction is evidently carried out in a largely dipolar aprotic medium!

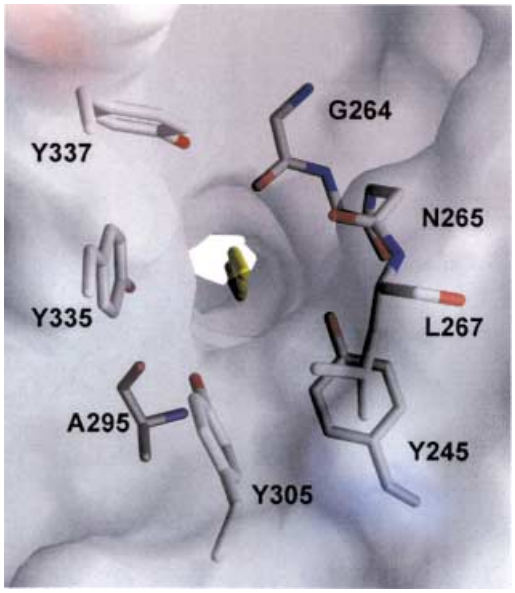


Fig. 2. *Surface of the ST7/9 complex with a decapeptide from human histone H3 showing the side chain of Lys*^{4} *inserted into the narrow cleft from the rear face of the enzyme*

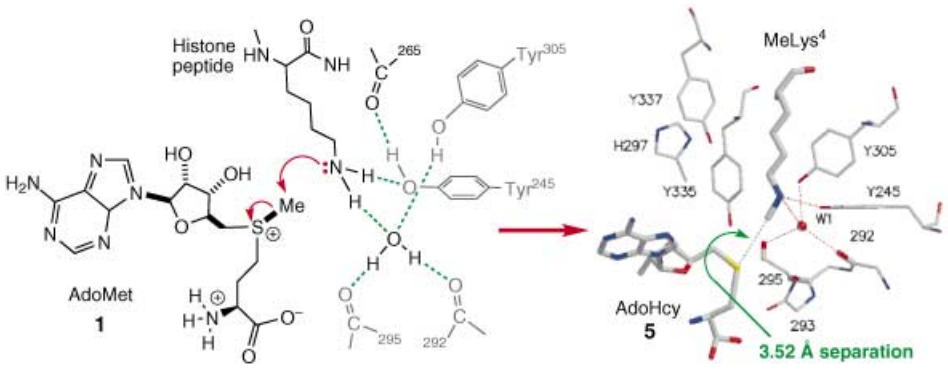


Fig. 3. *Mechanism of methylation of human histone H3 Lys*^{4} *by SET7/9. Right:* Structure of the ternary product complex of SET7/9 with AdoHcy (**5**) and the ε-MeLys^4 peptide showing key amino acid residues and water molecule W1. *Left:* Mechanistic scheme for reaction of Lys^4 with AdoMet (**1**) showing the same interactions with significant side chains and W1.

This Me transfer, thus, fulfils the requirements for a concerted S_N2 process with 'in line' geometry at the Me C-atom. It is essentially an example of the 'orbital steering' concept of *Storm* and *Koshland* [32]. It also meets the *Jencks* conjecture [33] that enzymes might stabilize the transition state for simple displacement reactions by induced intramolecularity, ionic and non-polar interactions, and possibly compression. There is no structural feature indicative of general acid-base catalysis. While *Dietze* and *Jencks* [33] identified general-base catalysis of a nucleophilic displacement on a

benzylsulfonium species by trifluoroethanol, their analysis suggested that such catalysis would not be anticipated for a very weakly acidic nucleophile. So it would appear that the reaction catalysed by SET7/9 involves the *neutral* Lys^4 and delivers the product initially as the *N*-methylammonium species.

What of the specificity of the enzyme as a mono-methylase [30]? It is abundantly clear from the structure of the product complex that there is only one possible location for the ε-*N*-Me group of Lys^4. Tyr^{245} and W1 not only make favourable interactions that stabilize the observed rotamer, but they also sterically preclude a Me group in either of the two other positions that could orientate a lone pair of electrons on the lysine ε-N-atom towards the S-atom of the AdoMet. This analysis suggests that protein mutations that create free space in the region of the Lys^4 ε-N-atom and/or disrupt the H-bond network should lead to further methylation. Such is the case! The Y245A mutation in SET7/9 leads to a marked reduction in HMTase activity as a mono-methylase. However, it has substantial activity when assayed with a mono- or with dimethylated Lys^4 substrate showing that Y245A can convert monomethylated substrate to trimethylated product. The structure of the SET7/9 ternary complex (*Fig. 3*, *right*) provides a rationale for these observations: the first Me group on Lys^4 could be positioned either at the Tyr^{245} OH or W1 site, and finally both of these positions can be occupied when ε-*N*,*N*-dimethyllysine-4 is the substrate.

Finally, how does this analysis fit with the recently published structures of two other lysine methylases? The narrow lysine access channel of SET7/9 is a common feature of Dim-5 [34] and the *Rubisco* LSMT methylase in which AdoHcy is bound in the same position with the same orientation [28]. However, both these proteins are trimethylases. The Tyr^{245} and Tyr^{305} residues in SET7/9 are absent in LSMT where their place is taken approximately by Phe^{224}, His^{252}, and Ile^{285} with a more spacious lysine binding-pocket. LSMT is thus capable of carrying out multiple lysine methylations because its active site is less constricted and H-bonding groups are available to position the methylated ε-amino group in a reactive conformation [28]. In the case of DIM-5, the product is trimethyl-lysine with little accumulation of mono- and dimethyl intermediates. It has a pH optimum of 10 which makes not only the substrate H3 Lys^9 readily deprotonated but Tyr^{178} and Tyr^{283} may exist as anions. This protein also lacks the key tyrosine residues in the active site but, critically, the F281Y mutation site changed the product specificity of DIM-5 from a trimethylase to a mono- and dimethylase without affecting overall catalytic activity [34].

In conclusion, it appears that evolution has elegantly and efficiently solved one of the long-standing problems of secondary amine synthesis: '*How to arrest the direct alkylation reaction*.' Surely, lessons are here to be learned for host–guest catalyst design!

REFERENCES

[1] M. Imfeld, C. A. Townsend, D. Arigoni, *J. Chem. Soc.* **1976**, 541.
[2] D. Kerridge, *J. Gen. Microbiol.* **1966**, *42*, 71.
[3] A. N. Glazer, R. J. De Lange, R. J. Martinez, *Biochim. Biophys. Acta* **1969**, *188*, 164.
[4] K. Hempel, H. W. Lange, L. Birkofer, *Naturwissenschaften* **1968**, *55*, 37.
[5] R. J. DeLange, A. N. Glazer, E. L. Smith, *J. Biol. Chem.* **1969**, *244*, 1385.
[6] W. M. Kuehlm, R. S. Adelstein, *Biochem. Biophys. Res. Commun.* **1969**, *37*, 59.

[7] H. O. Smith, T. J. Kelly, P. H. Roy, *Methods Enzymol.* **1974**, *29*, 282.
[8] T. M. Geiman, K. D. Robertson, *J. Cell. Biochem.* **2000**, *87*, 117; A. Bird, *Science* **2001**, *294*, 2113; T. Stokes, *Trends Plant Sci.* **2003**, *8*, 53.
[9] H. Phillips, J. Kenyon, H. D. Turley, *J. Chem. Soc.* **1923**, 44; **1923**, 64; **1925**, 399, **1925**, 2552, and later papers.
[10] E. D. Hughes, K. C. Ingold, S. Patel, *J. Chem. Soc.* **1933**, 526, and later papers.
[11] S. S. Shaik, H. B. Schlegel, S. Wolfe, 'Theoretical Aspects of Organic Chemistry, the S_N2 Mechanism', John Wiley & Sons, New York, 1992.
[12] E. S. Gould, 'Mechanism and Structure in Organic Chemistry', Henry Holt, New York, 1959, Chapt 8.
[13] J. Clayden, N. Greeves, S.Warren, P. Wothers, 'Organic Chemistry'. Oxford University Press, Oxford, 2001, Chapt 17.
[14] W. A. Cowdrey, E. D. Hughes, C. K. Ingold, S. Masterman, A. D. Scott, *J. Chem. Soc.* **1937**, 1252.
[15] W. P. Jencks, *J. Phys. Chem.* **1996**, *9*, 337.
[16] J. Luethy, J. Retey, D. Arigoni, *Nature (London)* **1969**, *221*, 1213.
[17] J. W. Cornforth, J. W. Redmond, H. Eggerer, W. Buckel, C. Gutschow, *Nature (London)* **1969**, *221*, 1212.
[18] J. T. Kealey, S. Lee, H. G. Floss, D. V. Santi, *Nucleic Acids Res.* **1991**, *19*, 6465.
[19] Y. Asano, R. W. Woodard, D. R. Houck, H. G. Floss, *Arch. Biochem. Biophys.* **1984**, *231*, 253.
[20] R. W. Woodard, M. D. Tsai, H. G. Floss, P. A. Crooks, J. K. Coward, *J. Biol. Chem.* **1980**, *255*, 9124.
[21] S. S. Parikh, G. Walcher, G. D. Jones, G. Slupphaug, H. E. Krokan, G. M. Blackburn, J. A. Tainer, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5083.
[22] A. R. Dinner, G. M. Blackburn, M. Karplus, *Nature (London)* **2001**, *413*, 752.
[23] J. T. Stivers, Y. L. Jiang, *Chem. Rev.* **2003**, *103*, 2729.
[24] X. Cheng, R. J. Roberts, *Nucleic Acids Res.* **2001**, *29*, 3784.
[25] D. K. Ho, J. C. Wu, D. V. Santi, H. G. Floss, *Arch. Biochem. Biophys.* **1991**, *284*, 264.
[26] G.Vilkaitis, A. Dong, E. Weinhold, X. Cheng, S. Klimasauskas, *J. Biol. Chem.* **2000**, *275*, 38722.
[27] G. Vilkaitis, E. Merkiene, S. Serva, E. Weinhold, S. Klimasauskas, *J. Biol. Chem.* **2001**, *276*, 20924.
[28] R. C. Trievel, E. M. Flynn, R. L. Houtz, J. H. Hurley, *Nat. Struct. Biol.* **2003**, *10*, 545.
[29] M. P. Mattson, *Ageing Res. Rev.* **2003**, *2*, 329.
[30] B. Xiao, C. Jing, J. R. Wilson, P. A. Walker, N. Vasisht, G. Kelly, S. Howell, I. A. Taylor, G. M. Blackburn, S. J. Gamblin, *Nature (London)* **2003**, *421*, 652.
[31] J. R. Wilson, C. Jing, P. A. Walker, S. R. Martin, S. A. Howell, G. M. Blackburn, S. J. Gamblin, B. Xiao, *Cell* **2002**, *111*, 105.
[32] D. R. Storm, D. E. Koshland, *Proc. Natl. Acad. Sci. U.S.A.* **1970**, *66*, 445.
[33] P. E. Dietze, W. P. Jencks, *J. Am. Chem. Soc.* **1989**, 111, 340.
[34] X. Zhang, Z. Yang, S. I. Khan, J. R. Horton, H. Tamaru, E. U. Selker, X. Cheng, *Mol. Cells.* **2003**, *12*, 177.

Received September 16, 2003