This is a repository copy of *Structure and catalytic mechanism of the human histone methyltransferase SET7/9*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/111/

Article:

28. Ruban, A. V., Lee, P. J., Wentworth, M., Young, A. J. & Horton, P. Determination of the stoichiometry and strength of binding of xanthophylls to the photosystem II light harvesting complexes. *J. Biol. Chem.* **274,** 10458–10465 (1999).

29. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 680–685 (1970).

30. Andersson, J., Walters, R. G., Horton, P. & Jansson, S. Antisense inhibition of the photosynthetic antenna proteins CP29 and CP26: implications for the mechanism of protective energy dissipation. *Plant Cell* **13,** 1193–1204 (2001).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.H. (e-mail: p.horton@sheffield.ac.uk).

....

Structure and catalytic mechanism of the human histone methyltransferase SET7/9

Bing Xiao*, Chun Jing*, Jonathan R. Wilson*, Philip A. Walker*, Nishi Vasisht*, Geoff Kelly*, Steven Howell*, Ian A. Taylor*, G. Michael Blackburn† & Steven J. Gamblin*

* *Structural Biology Group, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK*
† *Department of Chemistry, Krebs Institute, University of Sheffield, Sheffield S3 7HF, UK*

....

Acetylation[1,2], phosphorylation[3] and methylation[4] of the amino-terminal tails of histones are thought to be involved in the regulation of chromatin structure and function[5–7]. With just one exception[8,9], the enzymes identified in the methylation of specific lysine residues on histones (histone methyltransferases) belong to the SET family[10]. The high-resolution crystal structure of a ternary complex of human SET7/9 with a histone peptide and cofactor reveals that the peptide substrate and cofactor bind on opposite surfaces of the enzyme. The target lysine accesses the active site of the enzyme and the *S*-adenosyl-L-methionine (AdoMet) cofactor by inserting its side chain into a narrow channel that runs through the enzyme, connecting the two surfaces. Here we show from the structure and from solution studies that SET7/9, unlike most other SET proteins, is exclusively a mono-methylase. The structure indicates the molecular basis of the specificity of the enzyme for the histone target, and allows us to propose a model for the methylation reaction that accounts for the role of many of the residues that are invariant across the SET family.

Many SET proteins have now been characterized biochemically and several have been the subject of X-ray structure analysis: SET7/9 from human[11] and its complex with the product *S*-adenosyl-L-homocysteine (AdoHcy)[12]; Dim-5 from *Neurospora crassa*[6]; Rubisco large subunit methyltransferase (LSMT) from pea with AdoHcy[13]; and Clr4 from *Schizosaccharomyces pombe*[14]. SET proteins can be classified according to the lysine residues that they target on histones H3, H4 and H2A[4], and it is apparent that methylation at these different sites gives rise to distinct biological effects. An additional level of complexity is that lysine residues may be mono-, di- or tri-methylated and that these distinct species lead

to different signalling events. For example, in *Saccharomyces cerevisiae*, although di-methylation of Lys 4 on histone H3 is present at both active and inactive euchromatic genes, tri-methylation is linked exclusively to active genes[15].

NMR studies (see Supplementary Information) indicated that a histone peptide containing mono-methylated Lys 4 was better ordered in complex with SET7/9 than unmodified peptide. We therefore used the products of the normal histone methyltransferase (HMT) reaction for crystallization experiments (methylated lysine peptide and AdoHcy). In our previous studies[11] we obtained useful SET7/9 crystals only from constructs lacking the small carboxy-terminal segment, which is, nevertheless, essential for the catalytic activity of the enzyme. Here, we have used a catalytically active construct that contains the complete C-terminal segment. We obtained well-ordered crystals of the ternary complex of SET7/9 that diffracted to at least 1.7 Å spacing, and the structure was readily solved by molecular replacement. The C-terminal segment, the AdoHcy cofactor and most of the substrate peptide are well defined in the electron density maps, as are all the important residues around the active site. The overall structure of the ternary complex
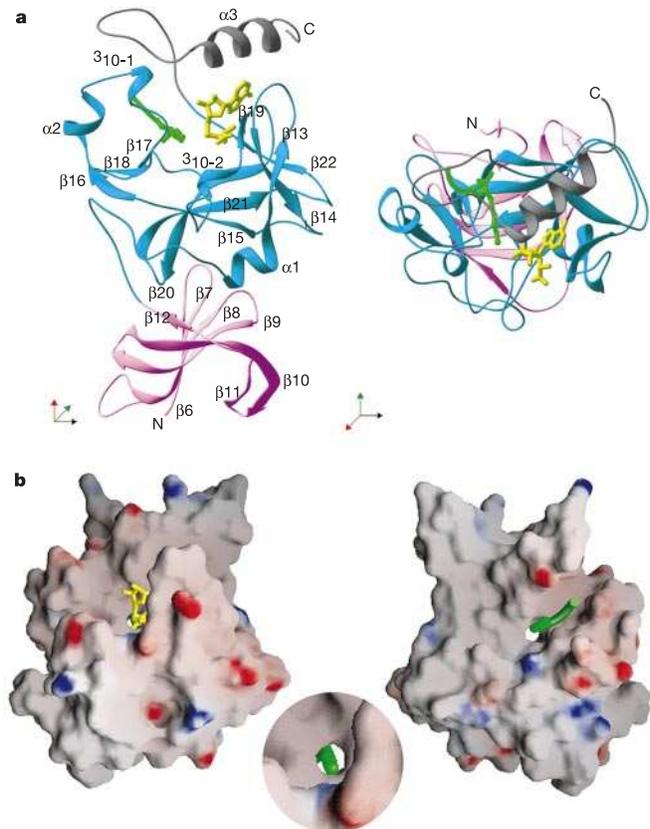

Figure 1 Structure of the SET7/9 ternary complex. **a,** Two orthogonal views of the SET7/9 ternary complex in ribbons representation. The N-terminal domain is coloured pink, the SET domain is blue and the C-terminal segment is grey. The H3 peptide is indicated in green, with the side chain of methylated Lys 4 shown. The *S*-adenosyl-L-homocysteine (AdoHcy) cofactor is coloured yellow. The secondary structure elements are labelled according to our earlier structure. Two small turns of the 3_{10} helix are also labelled. **b,** Two views of the SET domain are shown in a surface representation coloured according to electrostatic potential (the two views are related by a twofold rotation about a vertical axis). The left panel shows AdoHcy coloured yellow; the right panel shows the H3 peptide coloured green. The inset panel shows a close-up view of the lysine access channel containing the methyl lysine side chain as viewed from the *S*-adenosyl-L-methionine (AdoMet)-binding site.

of SET7/9 is shown in Fig. 1 (see Supplementary Information for crystallographic statistics).

The most notable feature of the catalytic structure is that the AdoHcy and the peptide substrate are located on opposite sides of the SET domain and that there is a narrow channel passing through the enzyme that connects the peptide and cofactor binding surfaces. The target lysine residue of the substrate (Lys 4) is inserted into this channel so that its amine can access the methyl donor (AdoMet; see Fig. 1b). The packing of the C-terminal segment against the SET domain is required to form the lysine access channel, which explains

why this feature has not been observed in previous HMTase structures (although it was noted in the context of Rubisco LSMT, as was the presence of a molecule of HEPES in the channel)[13]. The C-terminal segment (residues 345–366) is organized into two structural features. Residues 337–349, belonging mainly to the SET domain, form an approximate β-hairpin structure that protrudes at a right angle to the surface of the enzyme. This is followed by three residues that accommodate a sharp bend in the polypeptide chain before the final stretch of the protein that adopts an α-helical conformation (Fig. 1a). Tyr 335 and Tyr 337, located
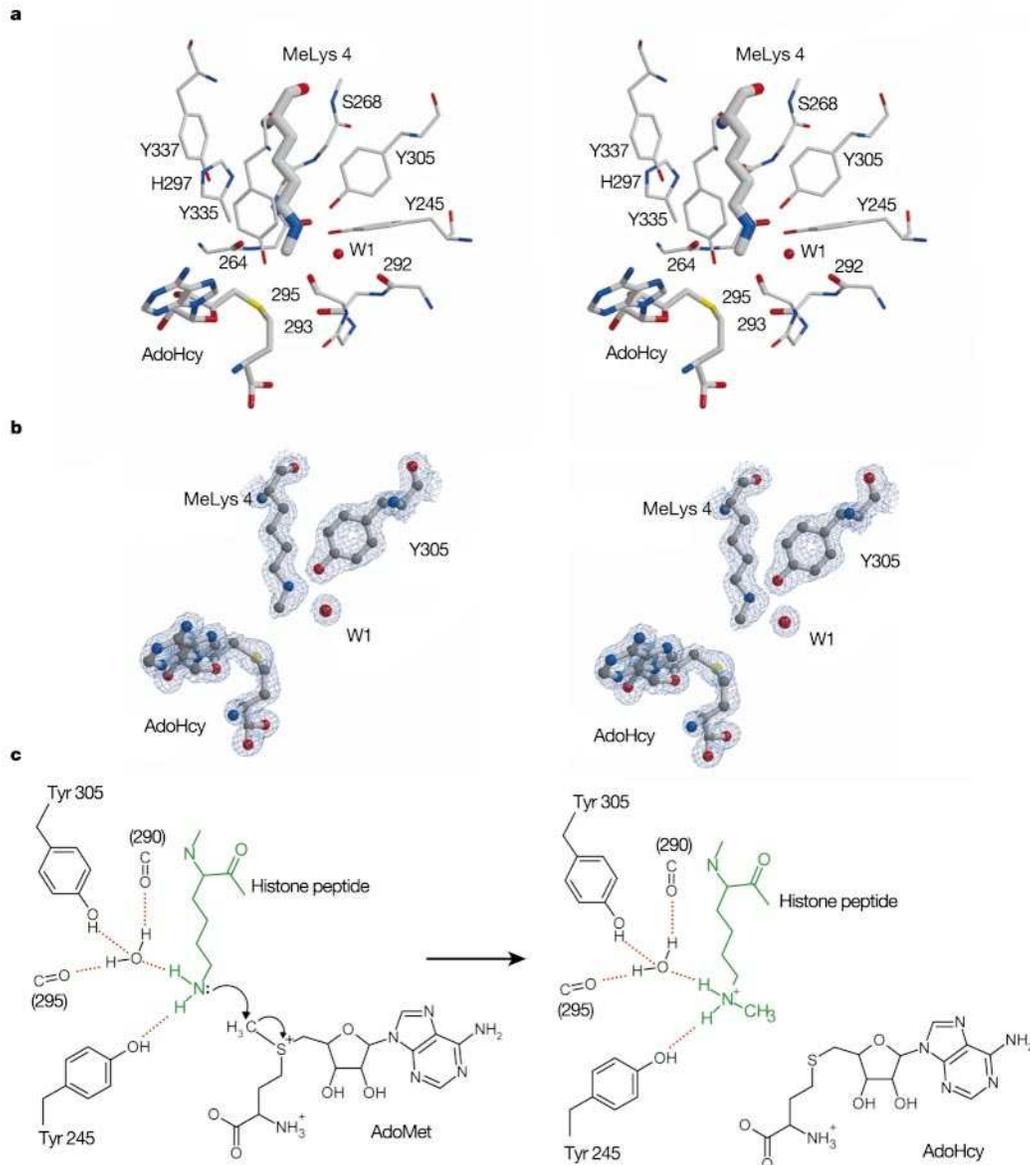

Figure 2 Active site of SET7/9. **a**, Stereographic representation of selected active-site residues together with methylated Lys 4 (MeLys 4), AdoHcy and several main-chain carbonyl groups oriented towards the lysine amine. **b**, Electron density ($2F_o - F_c$) covering part of the active site; the view is in the same orientation as in **a**. **c**, Schematic diagram of the reaction. The structure shows that the lysine has been stripped of all solvent molecules except for the one water used as a hydrogen-bond acceptor to orient the amino group. This desolvation will lower the pK_a of the lysine amino group and also enhance its nucleophilicity. At the same time, the local orientation of the dipoles of main-chain carbonyl groups towards the nitrogen will stabilize the developing positive charge on

that atom as the methylation reaction proceeds. There is no proximate proton acceptor to provide general base catalysis for the nucleophilic nitrogen. Mechanistically, it seems probable that the lysine side chain enters the active site with difficulty in its protonated form, the passage of this cation through the channel being facilitated by the faces of the flanking tyrosines. In the active site, the desolvated lysine is deprotonated, possibly to one of the flanking tyrosine oxygens. Thereafter the methylation reaction proceeds without general base catalysis, facilitated simply by alignment of orbitals, by desolvation, and by stabilization of charge reorganization.

just before the C segment, are both important for the formation of the lysine access channel. The arrangement of the β-hairpin is such that it stabilizes the conformation of these two tyrosine residues while also contributing to one of the sides of the groove into which the peptide binds. The second side of the peptide-binding groove is made up by residues 255–268 (including β-17). The α-helix at the end of the C-terminal segment packs against β-19, particularly Phe 299 (located just beyond the conserved NHS signature motif)[16], and makes hydrophobic packing interactions with the AdoHcy cofactor through Trp 352.

From the high-resolution structure presented here, and from the previous co-crystal structures of Rubsico LSMT/AdoHcy[13] and SET7/AdoHcy[12], it is apparent that in our earlier studies the orientation of AdoMet, which we modelled into a low-resolution difference map, was incorrect. The mode of cofactor binding in our present structure is essentially identical to that described previously[12,13]. In this alignment, the methyl group to be transferred from the AdoMet to the amine is pointing into the lysine-binding channel (see Fig. 2a). At the peptide-binding site, the channel surface is largely made up by the side chains of Leu 267 and Tyr residues 305, 335 and 337 (Fig. 2a). The alkyl component of the lysine side-chain therefore inhabits a hydrophobic environment. The other end of the channel, where it opens onto the cofactor-binding surface, is dominated by four tyrosine OH groups (Tyr 245, 305, 335, 337) and five main-chain carbonyl groups, all approximately oriented towards the lysine amine group. There are also several well-defined water molecules in the vicinity of the active site, although only one is close to the lysine amine. The structure, with its catalytic and binding surfaces, explains the role of a number of invariant residues whose mutation to alanine has been shown to abolish HMTase activity without significantly affecting the affinity for substrate or cofactor binding. The aromatic ring of Tyr 335 forms a significant part of the binding cavity for the alkyl portion of the Lys 4_{PEP} residue (histone peptide substrate) while its side-chain oxygen makes a single hydrogen bond to the main-chain carbonyl of residue 295. The importance of the correct positioning of this tyrosine residue is underscored by the observation that His 297 (of the NHS signature) acts to stabilize its conformation by acting as a hydrogen-bond acceptor to its main-chain amide (Fig. 2a).

The electron density for the methyl Lys 4_{PEP} is very well defined and clearly shows the location of the single methyl group (Fig. 2b). Examination of the active site also reveals that the lysine amine
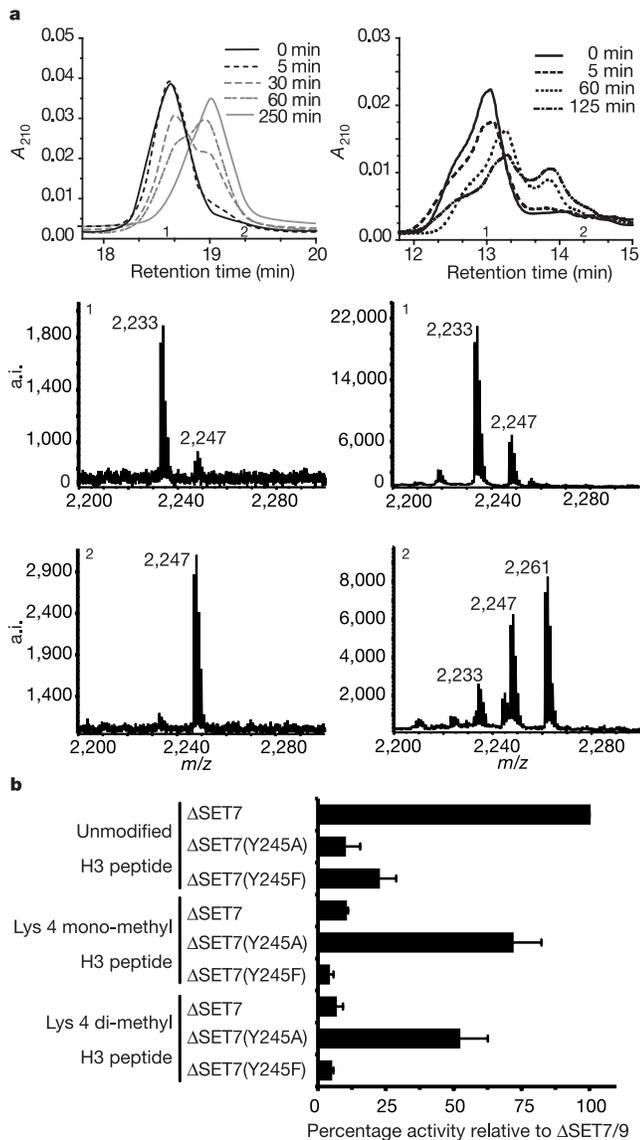

Figure 3 Activity of SET7/9. **a**, Analysis of methylation reaction products. The left-hand panels are for SET7/9; the right-hand panels are for G9a. The upper panel shows methylation of an H3 peptide at various time intervals as resolved by high-performance liquid chromatography (HPLC). The middle and lower panels show representative matrix-assisted laser desorption/ionization–time-of-flight (MALDI–TOF) analysis of fractions collected from the HPLC. Examples are from fraction points 1 and 2, respectively. A_{210}, absorbance at 210 nm; a.i., arbitrary intensity. **b**, Histone methyltransferase assay of SET7/9 and the point mutations Y245A and Y245F, with H3 peptide as substrate either unmodified, mono- or di-methylated at Lys 4.

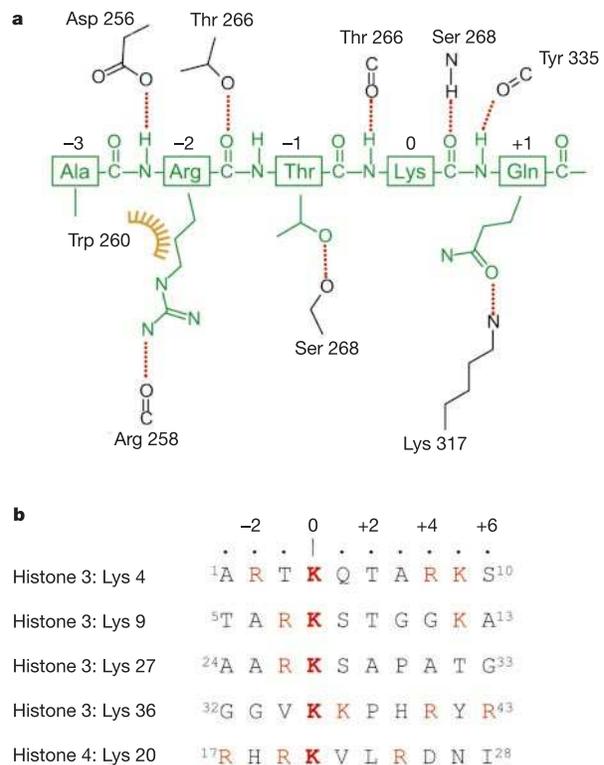Figure 4 Peptide interactions with SET7/9. **a**, Schematic representation of interactions made by the histone H3 peptide in complex with SET7/9. The positions of residues beyond the +1 position are not shown because they are affected by lattice contacts. **b**, Sequences of histones subject to methylation, aligned according to the position of their target lysine residues. Basic residues are highlighted in red.

donates hydrogen bonds to both the invariant Tyr 245 and to a tightly bound water molecule (W1). We are confident that the lysine is acting as a hydrogen-bond donor in both cases because of the nature of the other hydrogen bonds made by Tyr 245 and W1 (Fig. 2a). The orientation of the lysine amine group is such that the amine-methyl bond is aligned towards the sulphur atom of the AdoHcy. Moreover, it is directed at sulphur along the tetrahedral vector corresponding to the (S)-location of the methyl group in AdoMet[17]. In other words, the geometry is exactly that expected for the trajectory of the methylation reaction and this feature will provide an important entropic contribution towards enzyme-mediated catalysis (Fig. 2c).

Our previous data[11] suggested that SET7/9 could not use mono-methyl Lys 4 peptide as a substrate for further methylation. Thus we have further analysed the methylation reaction. Reaction mixtures were set up with unmodified peptide and AdoMet, and conditions found where the reaction was driven essentially to completion. As shown in Fig. 3a, high-performance liquid chromatography (HPLC) separation of the reaction products resulted in a peak that eluted at a different position to the unmodified peptide, and matrix-assisted laser desorption/ionization (MALDI) analysis of this material revealed only mono-methylated peptide. Taken with our previous data, these results support the conclusion that SET7/9 catalyses the addition of a single methyl group to its target lysine. The reason for SET7/9 acting as a mono-methyltransferase is apparent from our crystal structure. The peptide used for crystallization was synthesized using mono-methylated lysine at position 4. In the crystal structure there is well-defined electron density for this methyl group in just one position. The arrangement of the two hydrogen-bond acceptor groups (for the lysine amine) provides an immediate explanation for the lack of rotation about the CE–NZ bond; Tyr 245 and W1 not only make favourable interactions that stabilize the observed rotamer, but they also preclude, on steric grounds, a methyl group in either of the two other positions that would orient a lone pair of electrons on the lysine nitrogen towards the sulphur of the AdoMet. Thus the structure shows that the arrangement of protein side chains and, indirectly, water molecules at the active site of SET7/9 is such that it can only catalyse the addition of a single methyl group to the lysine amine.

Inspection of the sequence of other SET proteins[4] near the active site suggests why many of these enzymes catalyse di- or tri-methylation of their target lysines. Only Tyr 245 and Tyr 335 are considered invariant across the SET family, many other residues are highly variable. So, for example, Tyr 305 is substituted for a valine residue in SUV39H1 (a tri-methylase) and a proline in G9a (a di-methylase). Both of these substitutions would seem likely to produce a cavity at the active site that could accommodate a methyl substituent on the lysine amine. We have previously reported that the Y245A mutation in SET7/9 leads to a marked reduction in HMTase activity[11]. We now show that although this mutant has greatly reduced activity with respect to an unmodified lysine substrate, it has substantial activity if assayed with mono- or with di-methylated Lys 4 substrate (Fig. 3b). Thus, Y245A is able to convert mono-methylated substrate to tri-methylated product. The structure readily provides a rationale for these observations. The chemistry of methyl transfer dictates that for di- and then for tri-substitution to occur, the existing methyl groups on the lysine would first have to be positioned approximately either at the OH of Tyr 245 or W1 sites, and finally in both of these positions. Removal of the aromatic ring of Tyr 245 creates a site for the first attached methyl group, thus enabling a second addition. The fact that the mutant enzyme can generate tri-methylated lysine further implies that the Y245A mutation also disrupts the local structure, so that W1 is displaced, allowing di-methylated lysine to bind appropriately for a third methylation.

The histone peptide binds in a largely extended conformation into a shallow groove, as shown in Fig. 1b. The binding is mediated by a network of hydrogen bonds and salt bridges, involving both the main chain and side chains of the peptide (Fig. 4a). The target Lys 4_{PEP} residue is located approximately at the centre of the defined peptide, and the interactions of non-conserved SET7/9 residues with the peptide seems to account for the enzyme's specificity. Arg($-2)_{PEP}$ (residues are numbered relative to the target lysine) forms a hydrogen bond with the main-chain carbonyl of residue 258, such that the alkyl component of its side chain packs against the face of Trp 260. Thr($-1)_{PEP}$ forms a hydrogen bond with Ser 268, and Gln($+1)_{PEP}$ forms a hydrogen bond with the side chain of Lys 317. Arg($-2)_{PEP}$ therefore seems to have a particularly important role in mediating substrate specificity. Moreover, inspection of the various histone target sequences (Fig. 4b) shows that only the Lys 4 site of histone H3 contains a basic residue in the -2 position. The other histone target sites all have a basic residue either at the -1 or $+1$ position. Both activity measurements[11] and peptide affinity measurements, using a peptide in which Arg($-2)_{PEP}$ is mutated to an alanine (data not shown), support the hypothesis that this arginine residue contributes significantly to peptide binding to SET7/9. It seems that the structure of the ternary complex of SET7/9 presented here will provide a good model for assessing the probable substrate specificity of other SET proteins from their primary sequence. □

Methods

Protein constructs

ΔSET7/9 (residues 52–366) and G9a (621–1000) were expressed as glutathione S-transferase (GST) fusions in pGEX 6P1 in *Escherichia coli* BL21. The GST was removed by overnight treatment with PreScission Protease (Amersham) before gel filtration. Preparation of ΔSET7/9 in D_2O for NMR studies resulted in a series of N-terminal degradation products that remained catalytically active. Subsequently, ΔΔSET7/9 (108–366) was prepared as above and found to be stable for growth in D_2O, and was consequently used for further NMR and crystallography experiments. Site-directed alanine mutations were introduced using the Stratagene Quikchange mutagenesis kit, and mutations were confirmed by DNA sequencing and electrospray mass spectrometry. Synthetic peptides were prepared by W. Mawby, and AdoHcy was obtained from Fluka.

Crystallography

Protein stock solution was prepared at $100 \, mg \, ml^{-1}$ in 50 mM Tris, pH 7.0, 100 mM NaCl, and then incubated with a twofold molar excess of mono-methylated Lys 4 ten-residue peptide and AdoHcy. Crystals were grown by vapour diffusion at room temperature as hanging drops. Drops were prepared by mixing equal volumes of protein complex with reservoir solution containing 0.1 M Tris, pH 7.8, and 22% PEG 3350. Crystals were first transferred into mother liquor augmented with an additional 5% PEG 400, before plunging into liquid nitrogen. Data were collected from flash-cooled crystals at 100K on a Raxis-II detector mounted on a Rigaku RU200 generator. Diffraction data were integrated and scaled using DENZO and SCALEPACK[18]. The structure was solved by molecular replacement using our previous model (1H3I.brk) with AMORE. Subsequent refinement was carried out using REFMAC5[19] and manual model building in O[20].

NMR

NMR spectra were recorded at 25 °C on Varian Inova spectrometers operating at ^1H frequencies of 600 MHz and 800 MHz. Protein samples (approximately 0.5 mM) were prepared in 50 mM Tris-HCl, 0.5 mM Tris(2-carboxyethyl)phosphine hydrochloride, 10% D_2O, pH 6.5.

HMTase activity measurements

The methyltransferase activity of SET7/9 and mutant constructs described in the text were determined in a reaction volume of 20 μl containing 3 μM AdoMet supplemented with [methyl-^3H]AdoMet (4 μCi) (Amersham) and 750 nM purified methylase in reaction buffer (50 mM Tris, pH 8.5, 100 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol (DTT)) with 50 μM histone peptide. After incubation at 37 °C for 60 min the reaction was vacuum-blotted onto membrane (Hybond-C; Amersham), washed, and the activity was measured by scintillation counting.

Analytical analysis of histone methylation

The histone methyltransferase assay was carried out at 37 °C in 50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM DTT, with 300 μM AdoMet (Fluka), 100 μM H3 peptide (ARTKQTARKSTGGKAPRKQY), 1.5 μM enzyme. At the time intervals indicated, an aliquot of the reaction was removed and quenched in 8 M urea and acidified with glacial acetic acid. The reaction products was separated by reverse-phase HPLC (Jasco) on a Zorbax 300SB-C18 column (Rockland Technologies) using a gradient from 0% to 40% acetonitrile in the presence of 0.05% trifluoroacetic acid at 55 °C. Fractions from the peptide peak were analysed using a Reflex III MALDI–time-of-flight mass spectrometer (Bruker Daltonik) to obtain positive ion mass spectra.

letters to nature

1. Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature* **389**, 349–352 (1997).
2. Turner, B. M. Histone acetylation and an epigenetic code. *Bioessays* **22**, 836–845 (2000).
3. Berger, S. L. Histone modifications in transcriptional regulation. *Curr. Opin. Genet. Dev.* **12**, 142–148 (2002).
4. Lachner, M. & Jenuwein, T. The many faces of histone lysine methylation. *Curr. Opin. Cell Biol.* **14**, 286–298 (2002).
5. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
6. Zhang, X. *et al.* Structure of the neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase. *Cell* **111**, 117–127 (2002).
7. Spotswood, H. T. & Turner, B. M. An increasingly complex code. *J. Clin. Invest.* **110**, 577–582 (2002).
8. Feng, Q. *et al.* Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr. Biol.* **12**, 1052–1058 (2002).
9. van Leeuwen, F., Gafken, P. R. & Gottschling, D. E. Dit1p modulates silencing in yeast by methylation of the nucleosome core. *Cell* **109**, 745–756 (2002).
10. Jenuwein, T. Re-SET-ting heterochromatin by histone methyltransferases. *Trends Cell Biol.* **11**, 266–273 (2001).
11. Wilson, J. R. *et al.* Crystal structure and functional analysis of the histone methyltransferase SET7/9. *Cell* **111**, 105–115 (2002).
12. Jacobs, S. A. *et al.* The active site of the SET domain is constructed on a knot. *Nature Struct. Biol.* **9**, 833–838 (2002).
13. Trievel, R. C., Beach, B. M., Dirk, L. M. A., Houtz, R. L. & Hurley, J. H. Structure and catalytic mechanism of a SET domain protein methyltransferase. *Cell* **111**, 91–103 (2002).
14. Min, J., Zhang, X., Cheng, X., Grewal, S. I. & Xu, R. M. Structure of the SET domain histone lysine methyltransferase Clr4. *Nature Struct. Biol.* **9**, 828–832 (2002).
15. Santos-Rosa, H. *et al.* Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407–411 (2002).
16. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
17. Hofmann, J. L. Chromatographic analysis of the chiral and covalent instability of *S*-adenosyl-L-methionine. *Biochemistry* **25**, 4444–4449 (1986).
18. Otwinowski, Z. & Minor, W. in *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 556–562 (SERC Daresbury Laboratory, Warrington, 1993).
19. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
20. Jones, T. A., Zhou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).

....

corrigendum

Contemporary fisherian life-history evolution in small salmonid populations

Mikko T. Koskinen, Thrond O. Haugen & Craig R. Primmer

Nature **419**, 826–830 (2002).

The neutrality test equation applied in this Letter ($F = (N_e\sigma^2_{GB})/(h^2\sigma^2_{GW}t)$) was incorrect: the correct equation is $F = (N_e\sigma^2_{GB})/(\sigma^2_{GW}t)$. Consequently, the numbers in the right three columns of the original Table 1 are wrong (corrected below). In addition, some variance estimates in Table 1 and some values of Fig. 2 were reported incorrectly (the entire correct Table 1 and the correct Fig. 2 are available from the authors). These errors do not affect our conclusion that the populations evolved predominantly as a result of natural selection. We thank W. G. Hill for bringing these errors to our attention. □

Table 1

Trait	Pairwise comparison	$F_{1,\infty}$	P	N_e (sign)
Length at termination (mm)	Les–Ht	37.7	***	2.5
	Les–Aur	0.78	0.38	272
	Ht–Aur	8.63	**	17.7
Yolk-sac volume (mm^3)	Les–Ht	29.0	***	3.21
	Les–Aur	59.2	***	3.59
	Ht–Aur	3.17	0.07	48.3
Growth rate (mm/ΔD)	Les–Ht	20.5	***	4.55
	Les–Aur	6.31	**	33.7
	Ht–Aur	41.5	***	3.68
Survival (%)	Les–Ht	5.90	*	15.8
	Les–Aur	3.13	0.08	68.0
	Ht–Aur	1.20	0.27	128
Incubation time (days)	Les–Ht	4.02	*	23.1
	Les–Aur	1.56×10^{-7}	>0.99	1.36×10^9
	Ht–Aur	7.82	**	19.5
Swim-up length (mm)	Les–Ht	0.26	0.61	363
	Les–Aur	0.04	0.84	5.79×10^3
	Ht–Aur	2.73×10^{-8}	>0.99	5.60×10^9
Hatching length (mm)	Les–Ht	1.46×10^{-8}	>0.99	6.38×10^9
	Les–Aur	0.78	0.38	274
	Ht–Aur	2.16×10^{-8}	>0.99	7.07×10^9