Article

Structure of the Epigenetic Oncogene MMSET & Inhibition by N-Alkyl Sinefungin Derivatives

Dominic Tisi, Elisabetta Chiarparin, Emiliano Tamanini, Puja Pathuri, Joseph E. Coyle, Adam Hold, Finn P. Holding, Agnes C. L. Martin, Sharna J. Rich, Valerio Berdini, Jeff Yon, Paul Acklam, Rosemary Burke, Ludovic Drouin, Jenny Elizabeth Harmer, Fiona Jeganathan, Rob L.M. van Montfort, Yvette Newbatt, Marcello Tortorici, Maura Westlake, Amy Wood, Swen Hoelder, and Tom D. Heightman

ACS Chem. Biol., **Just Accepted Manuscript** • DOI: 10.1021/acschembio.6b00308 • Publication Date (Web): 29 Aug 2016

Structure of the Epigenetic Oncogene MMSET & Inhibition by *N*-Alkyl Sinefungin Derivatives

Dominic Tisi[1], *Elisabetta Chiarparin*[1], *Emiliano Tamanini*[1], *Puja Pathuri*[1], *Joseph E. Coyle*[1], *Adam Hold*[1], *Finn P. Holding*[1], *Agnes C.L. Martin*[1], *Sharna J. Rich*[1], *Valerio Berdini*[1], *Jeff Yon*[1], *Paul Acklam*[2], *Rosemary Burke*[2], *Ludovic Drouin*[2], *Jenny E. Harmer*[2], *Fiona Jeganathan*[2], *Rob L.M. van Montfort*[2], *Yvette Newbatt*[2], *Marcello Tortorici*[2], *Maura Westlake*[2], *Amy Wood*[2], *Swen Hoelder*[2], *and Tom D. Heightman*[1]*

[1]Astex Pharmaceuticals, 436 Cambridge Science Park, Milton Road, Cambridge UK CB4 0QA

[2]CRUK Cancer Therapeutics Unit, Institute of Cancer Research, 15 Cotswold Road, Belmont, Sutton, Surrey UK SM2 5NG

Corresponding Author: *To whom correspondence should be addressed.

T.D.H.: Tel +44 1223 226200; Fax: +44 1223 226201 Email: tom.heightman@astx.com.

ABSTRACT

The NSD subfamily of lysine methyl transferases are compelling oncology targets due to the recent characterization of gain-of-function mutations and translocations in several haematological cancers. To date these proteins have proven intractable to small molecule inhibition. Here we present initial efforts to identify inhibitors of MMSET (aka NSD2 or WHSC1) using solution phase and crystal structural methods. Based on 2D NMR experiments comparing NSD1 and MMSET structural mobility, we designed an MMSET construct with five point mutations in the N-terminal helix of its SET domain for crystallisation experiments and elucidated the structure of the mutant MMSET SET domain at 2.1 Å resolution. Both NSD1 and MMSET crystal systems proved resistant to soaking or co-crystallography with inhibitors. However, use of the close homologue SETD2 as a structural surrogate supported the design and characterization of *N*-alkyl sinefungin derivatives which showed low micromolar inhibition against both SETD2 and MMSET.

INTRODUCTION

Over the past decade the large family of chromatin modifying enzymes have emerged as promising targets for new anticancer therapy [1]. Among them, several protein lysine methyltransferases have been characterized as drivers in specific patient subgroups harbouring activating mutations or translocations, spurring concerted drug discovery efforts [2].

Multiple myeloma (MM) is a genetically unstable malignancy of postgerminal center B-lineage cells that frequently results in bone destruction, bone marrow failure, and death [3]. Some 15-20% of multiple myeloma (MM) patients possess a t(4;14) chromosomal rearrangement that is are associated with a poor prognosis [4 5]. This unique translocation results in a simultaneous overexpression of the multiple myeloma SET domain (MMSET, aka NSD2 or WHSC1) and fibroblast growth factor receptor 3 (FGFR3). The expression of FGFR3 does not appear to influence the survival of the t(4;14)+ MM patients [6 7], suggesting that MMSET plays the dominant role in the oncogenic transformation caused by the t(4;14) translocation [8].

MMSET is expressed as several splice variants: MMSET type I (647 aa), MMSET type II (1365 aa), and a third isotype (RE-IIBP) with a 584 aa protein identical to the carboxy-teminal region of MMSET type II [4 9]. Overexpression of MMSET in MM cells results in a switch in methylation of histone H3 from K27 (repressive) to K36 (activating) [10 8], leading to altered expression of a wide range of cell cycle, apoptosis, DNA repair and adhesion regulators [11 12 13 14]. Knockdown of MMSET using shRNA or active site directed mutagenesis in KMS11 cells possessing the t(4;14) translocation caused a significant loss of H3K36 methylation accompanied by antiproliferative effects, which were reversed by rescue transfection with wild-type MMSET II, confirming the driver role of MMSET in these cells [10].

MMSET has been implicated in a range of other cancers beyond MM. Recently a E1099K point mutation in MMSET was shown to enhance its methyltranferase activity and lead to altered global chromatin methylation in lymphoid malignancies including pediatric acute lymphoblastic leukemia [15, 16]. MMSET overexpression has been reported in some solid tumours including lung, prostate, neuroblastoma and bladder [17 18 19].

In addition to MMSET, the closely related homologues NSD1 and NSD3 (aka WHSC1L) are also subject to translocations in acute leukaemia resulting in fusion proteins [3, 20], and NSD3 amplifications occur in breast cancer [21]. Taken together, these findings present a compelling case for inhibitors of MMSET and its close homologues as potential personalized cancer therapies.

Inspired by recent successes in identifying clinical candidates for the lysine methyltransferases EZH2 [22] and DOT1L [23] as well as chemical tool inhibitors for G9a [24], SMYD2 [25] and SET7/9 [26], we set out to discover selective inhibitors of MMSET using a structure-guided design approach. As to date no structure of MMSET has been reported, we initiated efforts to elucidate its structure by X-ray crystallography, while in parallel exploring the potential of recently published crystal structures of the homologues NSD1 and SETD2 as surrogates for MMSET inhibitor design.

The NSD family predominantly mono- and di- methylate H3K36 *in vivo* [27, 28]. The SET domain in which the methyltransferase activity resides is located close to the C-terminus in all three family members, preceded by several Plant Homeodomain (PHD) domains and a Pro-Trp-Trp-Pro (PWWP) domain [29]. Crystal structure analyses of the SET domains of NSD1 and the closely related protein SETD2 have provided valuable insights into the architecture of the cofactor and substrate binding sites, as well as potential conformational changes required for substrate binding [30, 31]. The catalytic core of the SET domain is sandwiched between a

pre-SET region (also referred to as the AWS domain) and a post-SET domain. The post-SET region (~20 residues) is composed of a conserved zinc binding motif and a so-called autoregulatory loop whose conformation regulates accessibility to the substrate binding site, a mechanism also characterized in the methyltransferase ASH1L [32]. The structure of NSD1-SET [30] shows the post-SET loop (PSL) in a closed conformation which occludes the substrate binding site, and suggests a conformational rearrangement is required for substrate binding to occur. The flexibility of the PSL is confirmed by the structural characterisation of SETD2 [31]: depending on the presence of SAM or a derivative of the fungal metabolite sinefungin in the cofactor binding site, the PSL adopt a respective inactive or active conformation. In addition to the design of cofactor binding site inhibitors, the observation that the PSL of SETD2 can adopt open, active conformations upon ligand binding provides an additional opportunity to target the substrate binding site using structure based drug design (SBDD) methods, which is not available using the published NSD1 structure. Whether access to the substrate binding site of other SET domains is also regulated by conformational flexibility of the PSL is still to be determined.

RESULTS AND DISCUSSION

Solution Phase Characterisation and Construct Design

Despite high sequence homology between MMSET and NSD1, our initial efforts to crystallize MMSET using SET domain constructs analogous to the one used to determine the NSD1 structure [30] were unsuccessful. To characterize solution state differences in the behaviour of the two proteins, we carried out a back-bone assignment of the NSD1 and MMSET SET domains (NSD1-SET, residues 1852-2082; and MMSET-SET, residues 973-1203) in the presence of the co-factor SAM [33]. The NMR derived secondary structure assignment of NSD1-SET was consistent with the published crystal structure of NSD1. The backbone assignment of MMSET-SET revealed a secondary structure very similar to that of NSD1, with a core SET domain motif composed of three beta sheets with an alpha-helix adjacent to the second beta sheet (Figure S1) [33]. This core SET domain is flanked by an N-terminal helix adjacent to beta sheet 1 and the AWS domain at the other end, adjacent to beta sheet 3.

^{1}H-NMR and ^{1}H,^{15}N-NMR fingerprints of MMSET-SET demonstrated a large degree of ordering upon addition of co-factor SAM (Figure S3). LC/MS experiments showed that during the purification of the NSD1-SET and MMSET-SET proteins, extensive washing reduced the SAM content in MMSET-SET, but not in NSD1-SET, consistent with a higher degree of conformational mobility in MMSET allowing for a more facile removal of SAM. Moreover, native MS experiments on MMSET SET domain also suggested a highly conformationally mobile structure evidenced by a broad distribution of charge states (Figure S2). The differences in solution phase dynamics between NSD1-SET and MMSET-SET upon co-factor binding prompted us to identify specific regions of conformational mobility which could be modified to generate crystallisable MMSET constructs.

In the search for early tool inhibitors, we conducted a screen of recently published inhibitors of other lysine methyltransferases that might show off-target activity against MMSET. We observed weak inhibition of MMSET (77% inhibition at 300 μM) by the G9a/GLP methyltransferase inhibitor UNC0638 [34], while interestingly no inhibition of NSD1 could be measured. ITC measurements allowed determination of a K_d of 134 μM for UNC0638 *vs* MMSET-SET, while the K_d value *vs* NSD1-SET was very weak (>1mM) and could not be calculated accurately (Fig. 1a). This differential binding of UNC0638 towards MMSET compared with NSD1 prompted us to use further 2D-NMR experiments to explore potential differences in dynamics between the two proteins upon binding of UNC0638.

Analysis of the 2D-NMR spectrum of MMSET-SET in the presence of 1 mM UNC0638 identified regions of the protein which are perturbed upon ligand binding (Figure 1b). In particular, the N-terminal helix of MMSET-SET (residues L974-E982) was observed to undergo significant changes upon UNC0638 binding, as evidenced by the chemical shift changes and/or intensity changes for resonances assigned to Q975, R979, T981, R986, and K987 on the helix itself, and W1075, T1115 and F1139 which are close to the helix based on equivalent residues in NSD1. By contrast, changes in this region were not observed for NSD1-SET on incubation with 1 mM UNC0638 (Figure S4), consistent with the lack of interaction observed by bioassay and ITC experiments. Structural analysis of this region in NSD1 identified a cluster of leucine residues (L-1585, L-1588 and L-1681) which forms a local hydrophobic core stabilising the interaction between the N-terminal helix and the loop comprising residues T1680-W1685 that forms one side of the SAM binding site and precedes the highly conserved Gly-Trp-Gly motif found in many SET domains (Figure 2a). Sequence analysis confirmed that none of the three Leucine residues residues are conserved in MMSET, where the corresponding residues are hydrophilic and not able to form a similar hydrophobic interaction (Figure 2b). In addition in NSD1 two residues Q1682 and R1683 in

the T1680-W1685 loop which engage in polar interactions with the body of the protein are not conserved in MMSET. We hypothesised that these five residues could play a role in the observed differences between NSD1-SET and MMSET-SET in UNC0638 binding and in local conformational flexibility. Accordingly, to reduce conformational flexibility in this region of the protein and encourage crystallization, an MMSET-SET construct incorporating five point mutations was made to replace the wild type MMSET amino acids with the corresponding residues found in NSD1 (MMSET numbering - Q975L, A978L, D1071L, G1072Q, K1073R). This construct was used for crystallisation studies and biophysical characterisation.

MMSET Crystal Structure

Post SET loop and SAM binding site

The crystal structure of mutant MMSET-SET was solved at 2.1 Å resolution and shares the common SET domain architecture observed in the structures of NSD1 and SETD2 (Fig. 3). The PSL of MMSET is in a closed conformation similar to the auto-inhibited conformation observed in NSD1, albeit with a slight deviation around residues 1180-1182 (Fig. 4(a)). The binding site of the SAM co-factor is largely conserved between all three proteins, with the majority of hydrogen-bonding interactions with SAM mediated by the protein backbone. The adenine ring of SAM interacts via hydrogen bonds between N1 and the main-chain NH of Arg1192, between the 6-NH_2 and the carbonyl of His1142, and between N7 and the main-chain NH of His1142 (Fig. 4 (a)). In addition the face of the adenine ring forms close van der Waals contacts with the side chain of Leu1202. The ribose O2 oxygen interacts with the main chain carbonyl of Thr1115, and with the side chain of His1116, while the ribose O3 oxygen interacts with the main chain C=O of Phe1117 and the side chain of Asn1186, which serves

to anchor the end of the PSL. The amino acid group of SAM interacts with the main chain NH and carbonyl of Trp1075 and with the phenolic OH side chain of the conserved Tyr1118.

Comparison of the SAM binding pocket of MMSET with NSD1 shows a high degree of conservation: His 1116 represents the only change in an amino acid side chain directly contacting the SAM, replaced by Asn1186 in NSD1 which undergoes very similar interaction with the ribose O2 oxygen (Fig. 4c)). The majority of MMSET SAM site interactions are also conserved in SETD2: in this case the only amino acid side chain change is Asn1186, replaced by Ala1675 in SETD2 – here the key interaction with the ribose O3 is mediated by the adjacent Gln1676 side chain.

Comparison of the channel leading from the SAM methyl group to the substrate site also shows high sequence and structural conservation between MMSET and NSD1: all five hydrophobic amino acid side chains lining the channel are identical (Fig. 4e). In SETD2 two amino acid side chains in this channel are different: Leu1120 is replaced by Met1607 in SETD2, while Leu1163 is replaced by the more bulky Phe1650 in SETD2. This feature is somewhat counterintuitive given that unlike MMSET and NSD1, SETD2 is able to methylate H3K36Me_2 to generate the more bulky H3K36Me_3. It is possible that methylation state specificity arises not through steric selection, but due to the two key residue changes in SETD2 being better able to stabilize the cationic quaternary ammonium group of H3K36Me_3 through Π-cation and sulfur-cation interactions with Phe1650 and Met1607 respectively, which are not possible with the two Leu residues present in these positions in MMSET and NSD1. However it is also possible that the static view provided by these structures masks a difference in plasticity which might allow the substrate channel to open more widely in SETD2 to accommodate its larger substrate (see below).

N-terminal helix region

Although the mutation strategy led to successful crystallization of MMSET, and the overall structures of mutant MMSET-SET and NSD1-SET are similar, the N-terminal helix unexpectedly adopts a significantly different orientation (Figure 5a), forming a short helix located ~18Å above the SAM binding site which is not engaged in intra-molecular interactions with the body of the protein. The electron density for this helix is sufficiently resolved (particularly molecule A) to place main chain atoms; however, exact side chain conformations for some residues within the helix are not clearly defined. Analysis of the crystal packing shows that this helix is involved in crystal contacts within the crystal lattice unlike the corresponding regions of NSD1 and SETD2, which make intra-molecular interactions but no crystal contacts within their respective crystal systems. Given the significant differences in the solution phase mobility of this helix between NSD1 and MMSET and its different solid phase orientations for all three proteins, it is tempting to speculate that the helix may play a functional role in the activation of the protein. However it should be noted that the available structures of all three proteins comprise significantly truncated SET-domain constructs which may not represent the endogenous, full length situation. In particular, these constructs lack a number of additional chromatin binding domains which are likely to contribute to the activation mechanism.

SETD2-Inhibitor Co-crystal Structures

Despite the post-SET loop of MMSET appearing unencumbered by crystal contacts, we were unable to obtain co-structures of sinefungin derivatives by co-crystallization or soaking, hampering structure-based design efforts targeting the co-factor binding site. We were also unable to obtain co-structures with the close homologue NSD1 which could be envisaged as a

surrogate for MMSET. The closed conformation of the substrate binding site in the structures of both proteins also prevented structure-based design or in silico docking of potential inhibitors binding in this site. Since SETD2 is the next closest structural homologue of MMSET, we developed conditions to evaluate the use of SETD2 as a surrogate for the design of MMSET inhibitors. SETD2-sinefungin co-crystals were reproduced according to the published conditions [31]. We then solved SETD2 co-structures with several *N*-alkyl sinefungin derivatives by either back-soaking out sinefungin from SETD2-sinefungin co-crystals or by co-crystallizing SETD2 with an *N*-alkyl sinefungin derivative. Back soaking sinefungin co-crystals with an *N*-alkyl sinefungin derivative maintained the same crystal packing observed in the SETD2-sinefungin co-crystal structure. However, co-crystallization of apo SETD2 with *N*-alkyl sinefungin derivatives resulted in a different crystal packing not previously observed with the SETD2 crystal system. We generated the SETD2-*N*-propyl sinefungin derivative crystal structure in both crystal forms and an overlay of the two crystal forms revealed no significant conformational differences in residues lining the SAM binding site. This gave us confidence that either crystal form would be suitable for an in-depth analysis of inhibitors bound in the SAM site of SETD2.

Conformational analysis of the Post SET Loop

High resolution structures of SETD2 with both sinefungin and *N*-propyl sinefungin (**3**) showed a similar overall protein structure; however, key differences between the two protein-inhibitor structures were observed in the PSL. In the SETD2-sinefungin structure the PSL is open and disordered: electron density was particularly weak for residues Gln1667-Glu1674; while in the SETD2-*N*-propyl sinefungin structure the PSL is open and ordered. An overlay of the two structures revealed different orientations for residues Asp1665 and Tyr1666 at the

beginning of the PSL, which may have an impact on the ordering of the loop. Although our SETD2-*N*-propyl sinefungin structure is very similar to that published by Zheng et al., we observed a significant difference in the orientation of Arg1670, a residue shown by mutational analysis to play a pivotal role in enzyme catalysis and substrate recognition [31]. Zheng et al., report Arg1670 pointing into the SAM binding site in the SETD2-SAH co-structure (Fig. S5(a)); in their *N*-propyl sinefungin co-structure Arg1670 is shifted by ~15 Å to point out of the substrate groove, presumably in order to accommodate the *N*-propyl moiety (Fig. S5(b)). However, in our *N*-propyl sinefungin co-structure we observed clear electron density showing Arg1670 some 10 Å away from the Zheng et al. structure, pointing into the substrate binding site and holding the PSL in an open and ordered conformation through interactions with the backbone carbonyls of Asn1601 and Ile1602 (Fig. S5(c)). These observations highlight the dependence of flexibility of the SETD2 PSL on the chemical structure of the ligand bound in the SAM binding site. Additional SETD2 structures in complex with various *N*-alkyl sinefungin derivatives further demonstrate the mobility of the SETD2 PSL (see below).

Sinefungin analogues as MMSET inhibitors

The natural product sinefungin (**2**) has been widely reported as a non-selective inhibitor of SET-domain containing KMTs [35]. Sinefungin is a close structural analogue of the co-factor SAM, in which the sulfonium atom is replaced by a carbon atom and the reactive methyl group by a basic NH_2 group which, when protonated, mimics the incipient positive charge on the methyl group as it is transferred to the lysine ε-nitrogen atom of its substrate. The position of the nitrogen appears important: isosteric analogues in which the nitrogen is in the position of the sulfonium atom of SAM are generally much weaker inhibitors of SET-domain

containing KMTs - see for example [36, 37] (although for protein arginine methyltransferases (PRMTs) and the phylogenetically related KMT DOT1L this isostere appears favourable [23]).

A number of groups have pursued analogues of sinefungin as KMT inhibitors, including Zheng et al who explored N-substituted sinefungin derivatives as inhibitors of SETD2 [31]. These authors proposed that the N-alkyl substituent could mimic the substrate lysine side chain, potentially bridging between the co-factor and substrate binding sites on opposite sides of the protein. This led to the characterization of inhibitors such as N-propyl sinefungin (**3**) and N-benzyl sinefungin (**9**) which showed sub-micromolar inhibition of SETD2 and, somewhat surprisingly, promising selectivityagainst a panel of 13 other KMTs.

Noting the relatively close homology between SETD2 and MMSET, we explored the potential of N-substituted sinefungin analogues to inhibit MMSET. This required establishment of a synthetic route allowing late-stage functionalization of the amino group with a manageable number of steps. An assessment of published syntheses of sinefungin suggested that of Geze et al [38] to be most suited for adaptation to N-substitution (Scheme 1). The readily available starting material adenosine acetonide (**21**) was protected by N,N-dibenzoylation, followed by conversion to the 5-iodo-5-deoxy derivative **22** *via* the tosylate. Displacement of the iodide by diethyl cyanomethylphosphonate led to a mixture of the monobenzoylated phosphonate ester **24** and its dibenzoylated analogue – both could be transformed to the desired monobenzoylated alkene **27** by Wittig reaction with the aldehyde **26**, itself prepared by Dess-Martin oxidation of N-Boc homoserine methyl ester (**25)**. Magnesium reduction of the conjugated alkene in **27** followed by partial hydrolysis of the nitrile afforded the primary carboxamide **28**, which was transformed to the primary amine **29** using iodosobenzene bis-trifluoroacetate. Acidic deprotection of **29** afforded sinefungin as a mixture of 6'-epimers; alternatively, reductive amination with aldehydes and $NaBH_3CN$ followed by deprotection gave the *N*-substituted sinefungin derivatives **3**-**9** as mixtures of 6'-

epimers which were used for SAR and structural studies. The corresponding analogues of N-propyl sinefungin **3** lacking either the terminal carboxyl (**10**) or amino groups (**11**) were prepared by condensation of the cyanophosphonate **24** with the corresponding aldehydes, followed by a similar sequence of nitrile hydrolysis, degradation to the primary amine, reductive amination and deprotection (Scheme 2).

To determine inhibitory structure-activity relationships for compounds against MMSET and SETD2, we used assays in which the catalytic methylation of H3K36 by each enzyme is measured by selective antibody detection. For SETD2 a short peptide containing the H3K36Me_2 was used as substrate with H3K36Me_3 product detection in ELISA format. The assay for MMSET was performed using a protein construct containing additional C-terminal residues (aa 941-1240), with recombinant nucleosomes containing unmodified H3K36 as substrate and a H3K36Me_2 antibody for product detection by AlphaScreen. Inhibition and binding constant data are summarized in Table 1. Sinefungin (**2**) shows a modest IC_{50} of 26 μM vs MMSET, consistent with the previously published value of 30 μM using a ^{3}H-SAM turnover assay [35]; a similar level of inhibition was measured against SETD2. The *N*-propyl derivative **3** showed more potent inhibition of both enzymes: an IC_{50} of 3.3 μM was determined in the MMSET bioassay, consistent with a K_d of 1.6 μM as determined by ITC (Table 1; Figure 6). Encouraged by these results, we explored the effects of growing further into the substrate binding channel. The structure of **3** in complex with SETD2 suggested opportunities for modification of the propyl group in several ways. The propyl group inserts into the channel formed by hydrophobic side chain residues from the PSL (Y1666, F1664) and the body of the enzyme (residues M1607, Y1579, F1650) (Figure 7a). The presence of a small pocket adjacent to the propyl group suggested branching to give the sec-butyl derivative **4**: this resulted in a modest increase in affinity against both enzymes (Table 1). The structure of **4** in complex with SETD2 shows good steric complementarity as the small

pocket is filled by the additional methyl group (Figure 7b-c). The inhibitors were further elaborated to target the hydroxyl groups of the nearby Y1579 and Y1666. Substitution of the n-propyl with a hydroxyl group at the 2-position to give the secondary alcohol **5** resulted in a significant loss in potency, despite a clear hydrogen bond to the hydroxyl group of Y1579 in the SETD2 complex structure (Figure 7d). Similarly the 3-substituted primary alcohol **6** incurred a significant potency loss despite three new H-bonds observed between the inhibitor hydroxyl group and the hydroxyl groups of Y1579 and Y1666 and the backbone carbonyl of F1606 (Figure 7e). Incorporation of a basic group in this region of the molecule to give secondary amine **7** also resulted in a significant loss of potency. These results suggested that this region of the pocket prefers hydrophobic groups. Indeed, replacement of the *n*-propyl with *n*-pentyl (**8**) was tolerated, with only a modest loss of potency against both enzymes. The SETD2 structure in complex with **8** shows the extended alkyl chain tucked under the aryl ring of Tyr1579 in a similar region to the branched alkyl group of the *sec*-butyl derivative **4**. In contrast, replacement of the propyl with the more bulky benzyl group (**9**) was tolerated in SETD2 but not in MMSET. The structure of **9** in complex with SETD2 shows the benzyl group filling the lipophilic pocket created by the side chains of Tyr1579 and 1666 and by Phe1650 and 1664, with the side chain of Met1607 forming a lid (Figure 7f). This complex shows only modest structural rearrangements compared with the SETD2 structure in complex with the n-propyl derivative **3**. By contrast, in the SAM-bound MMSET structure, the side chain of Tyr1092, the corresponding residue of Tyr1579 in SETD2, is shifted to partially occupy this pocket, reflecting differences between the two proteins in the positioning of the PSL and the β-sheet comprising residues 1575-1585 (SETD2 numbering; 1088-1098 in MMSET). These findings suggest that the SETD2 substrate channel overall is more conformationally flexible, provide a rationale for its ability to accommodate this more bulky derivative compared with the less flexible channel in MMSET.

Finally we explored the effects of changes to the amino acid terminus of **3** which mimics the methionine moiety of SAM/SAH. Removing either the carboxy group to give **10** or the amino group to give **11** resulted in a dramatic loss in affinity vs MMSET, underscoring the importance of these groups in the molecular recognition of the co-factor.

CONCLUSIONS

The crystallization of the MMSET SET domain proved highly challenging despite the precedence of structures of the closely related homologues NSD1 and SETD2. By comparing NSD1 and MMSET SET domain solution structures and dynamics using 2D-NMR techniques, we identified regions of increased conformational mobility in MMSET. We developed a hypothesis for a cluster of amino acid residues underlying these conformational differences, and expressed mutant constructs designed to restrict conformational mobility in MMSET. This work led to the identification of a construct of MMSET SET domain which crystallized and allowed structure determination at 2.1 Å resolution. In parallel, we used biochemical and biophysical assays to characterize *N*-propyl sinefungin as a low micromolar inhibitor of MMSET. Despite the post-SET loop appearing unencumbered by crystal contacts, we were unable to obtain a co-structure of *N*-propyl sinefungin by co-crystallization or soaking. The similarity of the MMSET structure with that of SETD2 indicated that the latter protein might be a valid surrogate for structure-based design of inhibitors. We developed robust conditions which allowed soaking of *N*-alkyl sinefungin derivatives into SETD2 crystals, allowing the structural characterization of a number of derivatives which inhibited both SETD2 and MMSET with varying degrees of selectivity. The data indicated that there is some useful overlap between the two proteins in terms of inhibitor design, but also some divergence arising both from amino acid differences in the substrate binding channel and also apparent differences in the plasticity of the post-SET loop and SET domains

of the two proteins, such that SETD2 can accommodate larger groups in this region. Nonetheless we believe our findings provide a helpful foundation for further hypothesis-driven design of selective inhibitors of MMSET and/or SETD2, to allow deeper exploration of the roles of these epigenetic modulatory proteins in diseases including cancer.

EXPERIMENTAL

Protein expression and purification of MMSET

The gene encoding the SET domain of MMSET (residues 973-1203) was cloned into pET28b (Clontech). Site directed mutagenesis was performed using the QuikChange site-directed mutagenesis system (Agilent Technologies) to incorporate point mutations at the following positions - Q975L, A978L, D1071L, G1072Q, K1073R. For protein expression, transformed BL21 DE3 *E.coli* cells were grown in 2XYT media at 37°C until mid log phase. The temperature was reduced to 18°C and expression was induced by the addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Cells were incubated at 18°C overnight. Cell pellets were harvested by centrifugation at 4000rpm. Cell pellets were resuspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl) and lysed by sonication. Lysate was clarified by centrifugation at 20000 rpm for 30 mins. Clarified lysate was applied to a 5 ml HisTrap FF Crude column (GE Healthcare) attached to an AKTA PURE chromatography system (GE Healthcare). The column was washed with lysis buffer + 50 mM imidazole and fractions containing MMSET were eluted with lysis buffer + 250 mM imidazole. To generate SAM free MMSET for biophysical characterisation, MMSET bound to the HisTrap column was washed with 1L of lysis buffer + 50 mM imidazole at a flow rate of 1ml/min prior to elution. The vector derived N-terminal hexahis tag was removed by incubating the protein with Thrombin (Sigma) overnight at 4°C whilst dialysing against 50 mM Tris pH 8.0, 300 mM NaCl, 5 mM DTT. The protein sample was then applied to a Superdex75 16/60 gel filtration column pre-equilibrated in 50 mM Tris pH 8.0, 300 mM NaCl, 5 mM DTT. Fractions containing MMSET were pooled and used for crystallisation studies.

Isothermal titration calorimetry

ITC experiments were performed on a Microcal VP-ITC (Malvern Instruments Ltd) at 25°C in a buffer comprising 50 mM Hepes, 100 mM NaCl, 1 mM TCEP and 5% DMSO at pH 7.5. The protein construct used corresponded to human MMSET residues 973 – 1203 or human NSD1 residues 1852-2082. For all experiments compound was in the syringe and protein was in the sample cell. Data analysis was performed using Origin 7.0 software. Heats of dilution were estimated using the final injections of each individual titration and subtracted before data analysis. 1mM UNC0638 was titrated into 34 μM MMSET in the presence and absence of a 10-fold molar excess of SAM. In the absence of SAM the binding isotherm revealed that UNC0638 was binding but the data quality prevented unambiguous interpretation of the UNC0638 binding affinity. The ITC data quality was improved in the presence of SAM allowing an estimate of the UNC0638 binding affinity for the MMSET:SAM complex. Identical conditions were used for titration of 1mM UNC638 into 40 μM NSD1. The titration for n-propyl sinefungin **3** was performed with 400μM **3** in the syringe and 20μM MMSET was in the cell.

Expression and purification of ^{15}N labelled MMSET

^{15}N labelled MMSET (residues 973-1203) was produced using the isotope labelling method described by Li et al [39]. Several colonies of transformed BL21 DE3 *E.coli* were transferred into 50 ml of non-inducing media and incubated overnight at 30°C. The overnight culture was used to inoculate the inducing media with a starting OD600 of 0.02, incubated at 25°C for 4-5 hr and left at 18°C for another 48 hr. The final OD600 reached 14. Cells were then harvested, resuspended in lysis buffer (50 mM Hepes, 300 mM NaCl, 200 μM PMSF at pH 8) and lysed in a cell disruptor at 30 psi (2 passes). Lysate was centrifuged at 100,000 g and

the supernatant was supplemented with 10mM imidazole and batch bound to NiNTA resin overnight at 4°C. After extensively washing the resin in lysis buffer containing 20 mM imidazole, MMSET was eluted in lysis buffer containing 250 mM imidazole. MMSET fractions were pooled and dialysed overnight at 4°C into gel filtration buffer (50 mM Hepes, 150 mM NaCl, 2 mM DTT at pH 7.5) in the presence of thrombin (Sigma) to remove the N-terminal histidine tag. Protein was then applied to a Superdex75 26/60 gel filtration column equilibrated in gel filtration buffer. MMSET containing fractions were pooled and concentrated to ~ 10 mg/ml. The protein was then applied to a desalting column prior to loading onto a Q-sepharose column equilibrated in 50 mM sodium phosphate and 10 mM NaCl at pH 7.5. The Q-sepharose bound protein was extensively washed in the running buffer to facilitate the removal of any residual SAM associated with MMSET. The protein was eluted from the Q-sepharose column by applying a 10-500 mM NaCl gradient. MMSET containing fractions were pooled, concentrated and snap frozen in liquid nitrogen.

NMR spectroscopy

All NMR experiments were carried out at 310 K, using a Bruker AvanceIII 500 MHz spectrometer equipped with a cryoprobe. NMR spectra were processed using TOPSPIN and analysed using TOPSPIN 3.1. Each NMR sample at a protein concentration of 200 μM and saturating concentration of the co-factor SAM of 2 mM was made up to a volume of 160 μl in a 2.5 mm capillary which was placed in a standard 5 mm NMR tube. The NMR buffer was sodium phosphate 50 mM pH 7, NaCl 450 mM, TCEP 2 mM, SAM 2 mM, NaN_3 0.01% w/v, D_2O 10% v/v. Binding of UNC0638 was measured by monitoring chemical shift and shape changes of ^{1}H and ^{15}N resonances as a function of compound concentration. Each sample

contained 200 μM MMSET-SET protein and increasing concentrations of UNC0638 up to 1000 μM.

Crystallisation and structure determination of MMSET

Prior to crystallisation, MMSET protein was incubated with 1mM SAM (Sigma) dissolved in gel filtration buffer and the pH readjusted to pH7.5. Protein was then concentrated to ~7mg/ml. Broad crystallisation screens were set up including the JCSG Core Suite (Qiagen). Crystals of MMSET grew from conditions containing 16-20% PEG3350, 0.2M ammonium chloride at 20°C. Plate-like crystals grew overnight and were cryo-protected in 30% PEG3350 prior to being flash frozen in liquid nitrogen. Data were collected at 100K on beamline ID29 at ESRF and processed using XDS [40]. Data was scaled and merged using SCALA [41] from the CCP4i suite of programs [42]. The structure of MMSET was solved by molecular replacement using AMORE (CCP4i) [43] using NSD1 as the template model (PDB: 3OOI). The protein structure was rebuilt and refined using COOT [43] and REFMAC [44] respectively.

MMSET crystallised in space group $P2_12_12_1$ with two molecules in the asymmetric unit. All residues present in the protein construct were clearly resolved in electron density for molecule A. In molecule B, there are breaks in electron density for residues Gln982-Arg986 and Cys1018-Pro1025. Side chains where specific rotamer conformations were ambiguous due to poor electron density were trimmed back to the carbon-β atom position.

Protein expression and purification of SETD2

The gene encoding the SET domain of SETD2 (residues 1433-1711) was a gift from Structural Genomics Consortium, Toronto. Protein expression and purification of SETD2 broadly followed the published methods [31].

Typically, protein expression was performed using transformed BL21 DE3 Codon plus RIL *E.coli* cells (Agilent Technologies) grown at 37°C in Terrific Broth (Sigma) until mid log phase. The temperature was lowered to 18°C and expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Cells were incubated over night at 18°C prior to harvesting by centrifugation at 4000 rpm. Cells were resuspended in Lysis buffer (50 mM Hepes, pH 8.0, 500 mM NaCl, 1.6 mM β-mercaptoethanol) and lysed by sonication. Lysate was clarified by centrifugation at 14,000 rpm for 30 mins. The supernatant was incubated with ~10 ml of Nickel NTA resin (Qiagen) and incubated at 4°C for 1 hr. The resin was washed with ~200 ml of lysis buffer and SETD2 protein was eluted in ~50 ml of lysis buffer + 200 mM imidazole. The N-terminal hexahis tag was removed by incubating SETD2 with rTEV protease (Sigma) whilst dialysing against 20 mM Hepes pH7.3, 5 mM DTT. This sample was then applied to a 5 ml HiTrap SP FF ion exchange column (GE Healthcare) equilibrated in 20 mM Hepes pH 7.3, 5 mM DTT. A linear salt gradient was applied over 20 column volumes using 20 mM Hepes pH 7.3, 500 mM NaCl, 5 mM DTT. Fractions containing SETD2 were pooled and applied to a Hiprep S-200 26/60 gel filtration column (GE Healthcare) equilibrated in 20 mM Hepes pH 7.3, 150 mM NaCl, 5 mM DTT. Fractions containing SETD2 were pooled and concentrated to 1 mg/ml.

Crystallisation of SETD2

For co-crystallisation of SETD2, protein at 1 mg/ml was incubated with sinefungin or the compound of interest to a final concentration of 1.5 mM. The protein was then concentrated to ~10 mg/ml for crystallisation. 2 µl of protein solution was mixed with 2 µl of reservoir solution containing 0.1 M Hepes pH 7.3, 0.1 M KSCN, 25-30% MPEG2000 at 20°C. Crystals grew using the hanging drop vapour diffusion method, appeared overnight and took ~1 week to grow to full size. For the SETD2–compound complexes obtained via compound exchange, SETD2-sinefungin co-crystals were soaked in a solution containing 5 mM of the compound of interest for 24 h at 20°C to exchange sinefungin for the compound of interest. All crystals were cryoprotected in a solution containing 35% PEG3350, 0.1 M KSCN, 0.1 M Hepes pH 7.3 prior to flash freezing in liquid nitrogen for data collection.

MMSET Bioassay

MMSET activity was measured using an AlphaScreen assay. Each well contained a final concentration of 10 nM MMSET enzyme (MMSET, aa 953-1240), 12.5 µM nucleosomes (New England Biolabs, NEB), 0.5 µM S-adenosyl-methionine (NEB) and 3% v/v DMSO in reaction buffer (50 mM Tris pH 8.5, 5 mM $MgCl_2$, 1 mM DTT, 0.01% Tween-20). Initial reactions were performed for 60 mins at room temperature, protected from light. 1 nM Bio-antiH3 (PerkinElmer), 20 µg/ml mouse IgG beads, 1 nM $H3K36Me_2$ antibody (Active Motif) and 20 µg/ml Strepavidin donor beads (PerkinElmer anti-mouse IgG kit) in detection buffer (50 mM Tris pH 7.5, 900 mM NaCl, 0.01% Tween-20 and 0.001% Poly-L-lysine) was then added and incubated overnight in the dark at room temperature before reading using the Pherastar FS (BMG), Ex680 nm/Em615 nm. IC_{50} values were determined by fitting the data to a four parameter logistic fit using GraphPad Prism 6.0.

SETD2 Bioassay

SETD2 activity was measured using an AlphaScreen assay. Reactions were set up using a SETD2 chemiluminescent kit (BPS Bioscience). The method was as described in the product protocol with the exception that SAM from NEB was used. Luminescence was measured using a Pherastar FS (BMG). IC_{50} values were determined by fitting the data to a four parameter logistic fit using GraphPad Prism 6.0.

SAM quantitation by LC/MS

SAM content in MMSET was quantified by LC/MS using an HP1100 (Agilent Technologies) with diode array UV detection connected to a Bruker Esquire 3000 plus Ion Trap MS (Bruker Corporation). LC conditions: column - Hypercarb 100x3 mm, 3 µm (Thermo Scientific, part number 35005-103030) at 40 °C, aqueous solvent (A) - 100% water / 0.1% formic acid, organic solvent (B) - 95% acetonitrile / 5% water / 0.1% formic acid. Solvent gradient – 0 min (5% B), 0.5 min (5%B), 2 min (95% B), 4 min (95% B), 4.1 min (5% B), 7 min (5% B), flow rate = 1 ml/min with all flow to the UV detector then to the MS. SAM levels were quantitated from the UV data (summed 265 nm to 350 nm) and the extracted ion chromatogram (XIC) for the $(M+H)^+$ of SAM (m/z 399.1). SAM standards for UV and MS calibration curves were prepared from 100 mM SAM in 100 mM ammonium acetate, pH 7.4 and serially diluted in 80:20 1 M KOH:TFA to give 250 µM, 12.5 µM and 0.625 µM stock solutions. The average peak areas from duplicate injections of the standard solutions for m/z 399.1 and the corresponding UV peak areas were plotted against amount injected in nMoles to give calibration plots; from the MS data of 6.25 pM (LOQ) to 0.125 nM ($R^2 = 0.9995$) and from UV absorption data of 25 pM (LOQ) to 0.5 nM ($R^2 = 0.9995$). Preparation of SAM from MMSET samples was according to the following procedure: 5 µL of MMSET protein

was diluted with 10 µL of solvent A and 2 µL of 100% TFA was added to precipitate the protein, the sample was aspirated a number of times and centrifuged for 5 minutes at 13000 rpm to pellet the protein. The supernatant containing SAM was removed (approximately 17 µL) and 8 µL of 1 M KOH added giving a 1 in 5 dilution of the original sample at a final volume of 25 µL. Injection volumes were between 0.5 µL to 10 µL depending on the original protein concentration.

Accession Codes

MMSET-SAM: tbc; SETD2-sinefungin: tbc; SETD2-3: tbc; SETD2-4: tbc; SETD2-5: tbc; SETD2-6: tbc; SETD2-9: tbc.

ASSOCIATED CONTENT

Supporting Information. Figure S1 – Secondary structure assignment of NSD1 and MMSET-SET domains. Native MS procedures for MMSET-SET. Figure S2 – Native MS for MMSET-SET domain. Figure S3 – NMR observation of MMSET ordering upon addition of co-factor SAM. Figure S4 - 15N-1H HSQC 2D-NMR spectrum of labelled NSD1-SET domain showing NSD1 signals in presence (blue) or absence (red) of UNC0638. Figure S5 – Post-SET loop conformations for SETD2 in complex with SAM analogues. Table S1 – X-ray collection statistics for MMSET and SETD2 structures. Synthetic procedures for *N*-propyl sinefungin (**3**) and analogues. This material is available free of charge via the Internet at http://pubs.acs.org.

PDB ID Codes

Coordinates and structure factors for the MMSET and SETD2 ligand complex have been deposited with PDB with accession codes: *tbc*.

AUTHOR INFORMATION

Corresponding Author

T.D.H.: Astex Pharmaceuticals, 436 Cambridge Science Park, Cambridge CB4 0QA, tom.heightman@astx.com.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

ACKNOWLEDGMENTS

We gratefully acknowledge H. Wu, Structural Genomics Consortium, University of Toronto for helpful discussions on the crystallography of SETD2. This work was supported by Cancer Research UK grant number C309/A11566. We acknowledge NHS funding to the NIHR Biomedical Research Centre at The Institute of Cancer Research and the Royal Marsden Hospital. We thank the staff of the European Synchrotron Radiation facility for their support during data collection.

ABBREVIATIONS

Boc_2O, di-tert-butyl dicarbonate; BzCl, benzoyl chloride; DCM, dichloromethane; DMF, dimethylformamide; DMSO, dimethylsulfoxide; DOT1L, DOT1-like histone H3K79 methyltransferase; DTT, dithiothreitol; EZH2, enhancer of zeste 2; IPTG, isopropyl β-D-1-thiogalactopyranoside; ITC, isothermal calorimetry; KMT, protein lysine methyltransferase; MM, multiple myeloma; MMSET, multiple myeloma SET-domain containing; NSD1, nuclear receptor binding SET domain protein 1; PHD, plant homeodomain; PMSF, phenylmethanesulfonyl fluoride; PRMT, protein arginine methyltransferase; PSL, post-SET loop; SAM, S-adenosyl methionine; SBDD, structure-based drug design; SET, Su(var)3-9 Enhancer of zeste Trithorax; SETD2, SET domain containing 2; SET7/9, SET domain containing 7/9; SMYD2, SET and MYND domain containing; TCEP, tris(2-carboxyethyl)phosphine; TEA, trimethylamine; TFA, trifluoroacetic acid; TLC, thin layer chromatography; TMSCl, trimethylsilyl chloride; TsCl, tosyl chloride.

REFERENCES

1. Dawson, Mark A., and Kouzarides, T. (2012) Cancer Epigenetics: From Mechanism to Therapy, *Cell 150*, 12-27.
2. Copeland, R. A. (2013) Molecular Pathways: Protein Methyltransferases in Cancer, *Clin. Cancer Res. 19*, 6344-6350.
3. Chesi, M., and Bergsagel, P. L. (2011) Many Multiple Myelomas: Making More of the Molecular Mayhem, *ASH Education Program Book 2011*, 344-353.
4. Chesi, M., Nardini, E., Lim, R. S. C., Smith, K. D., Kuehl, W. M., and Bergsagel, P. L. (1998) The t(4;14) Translocation in Myeloma Dysregulates Both FGFR3and a Novel Gene, MMSET, Resulting in IgH/MMSET Hybrid Transcripts, *Blood 92*, 3025-34.
5. Keats, J. J., Reiman, T., Maxwell, C. A., Taylor, B. J., Larratt, L. M., Mant, M. J., Belch, A. R., and Pilarski, L. M. (2003) In multiple myeloma, t(4;14)(p16;q32) is an adverse prognostic factor irrespective of FGFR3 expression, *Blood 101*, 1520-9.
6. Rasmussen, T., Hudlebusch, H. R., Knudsen, L. M., and Johnsen, H. E. (2002) FGFR3 dysregulation in multiple myeloma: frequency and prognostic relevance, *Br. J. Haematol. 117*, 626-628.
7. Keats, J. J., Maxwell, C. A., Taylor, B. J., Hendzel, M. J., Chesi, M., Bergsagel, P. L., Larratt, L. M., Mant, M. J., Reiman, T., Belch, A. R., and Pilarski, L. M. (2005) Overexpression of transcripts originating from the MMSET locus characterizes all t(4;14)(p16;q32)-positive multiple myeloma patients, *Blood 105*, 4060–4069.
8. Mirabella, F., Wu, P., Wardell, C. P., Kaiser, M. F., Walker, B. A., Johnson, D. C., and Morgan, G. J. (2013) MMSET is the key molecular target in t(4;14) myeloma, *Blood Cancer J. 3*, e114.
9. Garlisi, C. G., Uss, A. S., Xiao, H., Tian, F., Sheridan, K. E., Wang, L., Motasim Billah, M., Egan, R. W., Stranick, K. S., and Umland, S. P. (2001) A Unique mRNA Initiated within a Middle Intron of WHSC1/MMSET Encodes a DNA Binding Protein That Suppresses Human IL-5 Transcription, *Am. J. Respir. Cell Mol. Biol. 24*, 90-98.
10. Martinez-Garcia, E., Popovic, R., Min, D.-J., Sweet, S. M. M., Thomas, P. M., Zamdborg, L., Heffner, A., Will, C., Lamy, L., Staudt, L. M., Levens, D. L., Kelleher, N. L., and Licht, J. D. (2011) The MMSET histone methyl transferase switches global histone methylation and alters gene expression in t(4;14) multiple myeloma cells, *Blood 117*, 211-20.
11. Lauring, J., Abukhdeir, A. M., Konishi, H., Garay, J. P., Gustin, J. P., Wang, Q., Arceci, R. J., Matsui, W., and Park, B. H. (2008) The multiple myeloma–associated MMSET gene contributes to cellular adhesion, clonogenic growth, and tumorigenicity, *Blood 111*, 856-64.
12. Brito, J. L. R., Walker, B., Jenner, M., Dickens, N. J., Brown, N. J. M., Ross, F. M., Avramidou, A., Irving, J. A. E., Gonzalez, D., Davies, F. E., and Morgan, G. J. (2009) MMSET deregulation affects cell cycle progression and adhesion regulons in t(4;14) myeloma plasma cells, *Haematologica 94*, 78-86.
13. Kuo, Alex J., Cheung, P., Chen, K., Zee, Barry M., Kioi, M., Lauring, J., Xi, Y., Park, Ben H., Shi, X., Garcia, Benjamin A., Li, W., and Gozani, O. (2011) NSD2 Links Dimethylation of Histone H3 at Lysine 36 to Oncogenic Programming, *Mol. Cell 44*, 609-620.
14. Walker, B. A., Wardell, C. P., Chiecchio, L., Smith, E. M., Boyd, K. D., Neri, A., Davies, F. E., Ross, F. M., and Morgan, G. J. (2011) Aberrant global methylation patterns affect the molecular pathogenesis and prognosis of multiple myeloma, *Blood 117*, 553-62.
15. Jaffe, J. D., Wang, Y., Chan, H. M., Zhang, J., Huether, R., Kryukov, G. V., Bhang, H.-e. C., Taylor, J. E., Hu, M., Englund, N. P., Yan, F., Wang, Z., Robert McDonald Iii, E., Wei, L., Ma, J., Easton, J., Yu, Z., deBeaumount, R., Gibaja, V., Venkatesan, K., Schlegel, R., Sellers, W. R., Keen, N., Liu, J., Caponigro, G., Barretina, J., Cooke, V. G., Mullighan, C., Carr, S. A., Downing, J. R., Garraway, L. A., and Stegmeier, F. (2013) Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia, *Nat. Genet. 45*, 1386-1391.
16. Oyer, J. A., Huang, X., Zheng, Y., Shim, J., Ezponda, T., Carpenter, Z., Allegretta, M., Okot-Kotber, C. I., Patel, J. P., Melnick, A., Levine, R. L., Ferrando, A., MacKerell, A. D.,

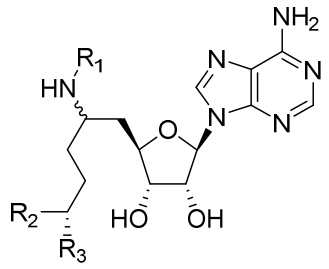
Kelleher, N. L., Licht, J. D., and Popovic, R. (2014) Point mutation E1099K in MMSET/NSD2 enhances its methyltranferase activity and leads to altered global chromatin methylation in lymphoid malignancies, *Leukemia 28*, 198-201.

17. Hudlebusch, H. R., Santoni-Rugiu, E., Simon, R., Ralfkiær, E., Rossing, H. H., Johansen, J. V., Jørgensen, M., Sauter, G., and Helin, K. (2011) The Histone Methyltransferase and Putative Oncoprotein MMSET Is Overexpressed in a Large Variety of Human Tumors, *Clin. Cancer Res. 17*, 2919-2933.
18. Hudlebusch, H. R., Skotte, J., Santoni-Rugiu, E., Zimling, Z. G., Lees, M. J., Simon, R., Sauter, G., Rota, R., De Ioris, M. A., Quarto, M., Johansen, J. V., Jørgensen, M., Rechnitzer, C., Maroun, L. L., Schrøder, H., Petersen, B. L., and Helin, K. (2011) MMSET Is Highly Expressed and Associated with Aggressiveness in Neuroblastoma, *Cancer Res. 71*, 4226-4235.
19. Asangani, Irfan A., Ateeq, B., Cao, Q., Dodson, L., Pandhi, M., Kunju, Lakshmi P., Mehra, R., Lonigro, Robert J., Siddiqui, J., Palanisamy, N., Wu, Y.-M., Cao, X., Kim, Jung H., Zhao, M., Qin, Zhaohui S., Iyer, Mathew K., Maher, Christopher A., Kumar-Sinha, C., Varambally, S., and Chinnaiyan, Arul M. (2013) Characterization of the EZH2-MMSET Histone Methyltransferase Regulatory Axis in Cancer, *Mol. Cell 49*, 80-93.
20. Rosati, R., La Starza, R., Veronese, A., Aventin, A., Schwienbacher, C., Vallespi, T., Negrini, M., Martelli, M. F., and Mecucci, C. (2002) NUP98 is fused to the NSD3 gene in acute myeloid leukemia associated with t(8;11)(p11.2;p15), *Blood 99*, 3857-3860.
21. Angrand, P.-O., Apiou, F., Stewart, A. F., Dutrillaux, B., Losson, R., and Chambon, P. (2001) NSD3, a New SET Domain-Containing Gene, Maps to 8p12 and Is Amplified in Human Breast Cancer Cell Lines, *Genomics 74*, 79-88.
22. Knutson, S. K., Kawano, S., Minoshima, Y., Warholic, N. M., Huang, K.-C., Xiao, Y., Kadowaki, T., Uesugi, M., Kuznetsov, G., Kumar, N., Wigle, T. J., Klaus, C. R., Allain, C. J., Raimondi, A., Waters, N. J., Smith, J. J., Porter-Scott, M., Chesworth, R., Moyer, M. P., Copeland, R. A., Richon, V. M., Uenaka, T., Pollock, R. M., Kuntz, K. W., Yokoi, A., and Keilhack, H. (2014) Selective Inhibition of EZH2 by EPZ-6438 Leads to Potent Antitumor Activity in EZH2-Mutant Non-Hodgkin Lymphoma, *Mol. Cancer Ther. 13*, 842-854.
23. Daigle, S. R., Olhava, E. J., Therkelsen, C. A., Basavapathruni, A., Jin, L., Boriack-Sjodin, P. A., Allain, C. J., Klaus, C. R., Raimondi, A., Scott, M. P., Waters, N. J., Chesworth, R., Moyer, M. P., Copeland, R. A., Richon, V. M., and Pollock, R. M. (2013) Potent inhibition of DOT1L as treatment of MLL-fusion leukemia, *Blood 122*, 1017-25.
24. Liu, F., Barsyte-Lovejoy, D., Li, F., Xiong, Y., Korboukh, V., Huang, X.-P., Allali-Hassani, A., Janzen, W. P., Roth, B. L., Frye, S. V., Arrowsmith, C. H., Brown, P. J., Vedadi, M., and Jin, J. (2013) Discovery of an in Vivo Chemical Probe of the Lysine Methyltransferases G9a and GLP, *J. Med. Chem. 56*, 8931-8942.
25. Ferguson, Andrew D., Larsen, Nicholas A., Howard, T., Pollard, H., Green, I., Grande, C., Cheung, T., Garcia-Arenas, R., Cowen, S., Wu, J., Godin, R., Chen, H., and Keen, N. (2011) Structural Basis of Substrate Methylation and Inhibition of SMYD2, *Structure 19*, 1262-1273.
26. Barsyte-Lovejoy, D., Li, F., Oudhoff, M. J., Tatlock, J. H., Dong, A., Zeng, H., Wu, H., Freeman, S. A., Schapira, M., Senisterra, G. A., Kuznetsova, E., Marcellus, R., Allali-Hassani, A., Kennedy, S., Lambert, J.-P., Couzens, A. L., Aman, A., Gingras, A.-C., Al-Awar, R., Fish, P. V., Gerstenberger, B. S., Roberts, L., Benn, C. L., Grimley, R. L., Braam, M. J. S., Rossi, F. M. V., Sudol, M., Brown, P. J., Bunnage, M. E., Owen, D. R., Zaph, C., Vedadi, M., and Arrowsmith, C. H. (2014) (R)-PFI-2 is a potent and selective inhibitor of SETD7 methyltransferase activity in cells, *Proc. Natl. Acad. Sci. USA 111*, 12853-12858.
27. Wagner, E. J., and Carpenter, P. B. (2012) Understanding the language of Lys36 methylation at histone H3, *Nat. Rev. Mol. Cell Biol. 13*, 115-126.
28. Lucio-Eterovic, A. K., Singh, M. M., Gardner, J. E., Veerappan, C. S., Rice, J. C., and Carpenter, P. B. (2010) Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function, *Proc. Natl. Acad. Sci. USA 107*, 16952-16957.

29. Morishita, M., and di Luccio, E. (2011) Structural insights into the regulation and the recognition of histone marks by the SET domain of NSD1, *Biochem. Biophys. Res. Commun. 412*, 214-219.
30. Qiao, Q., Li, Y., Chen, Z., Wang, M., Reinberg, D., and Xu, R. M. (2011) The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation, *J. Biol. Chem. 286*, 8361-8368.
31. Zheng, W., Ibáñez, G., Wu, H., Blum, G., Zeng, H., Dong, A., Li, F., Hajian, T., Allali-Hassani, A., Amaya, M. F., Siarheyeva, A., Yu, W., Brown, P. J., Schapira, M., Vedadi, M., Min, J., and Luo, M. (2012) Sinefungin Derivatives as Inhibitors and Structure Probes of Protein Lysine Methyltransferase SETD2, *J. Am. Chem. Soc. 134*, 18004-18014.
32. An, S., Yeo, K. J., Jeon, Y. H., and Song, J.-J. (2011) Crystal Structure of the Human Histone Methyltransferase ASH1L Catalytic Domain and Its Implications for the Regulatory Mechanism, *J. Biol. Chem. 286*, 8369-8374.
33. Amin, N., Nietlispach, D., Qamar, S., Coyle, J., Chiarparin, E., and Williams, G. (2016) Backbone resonance assignment and secondary structure determination of human NSD1 and NSD2, *Biomol NMR Assign (2016)*. doi:10.1007/s12104-016-9691-x.
34. Vedadi, M., Barsyte-Lovejoy, D., Liu, F., Rival-Gervier, S., Allali-Hassani, A., Labrie, V., Wigle, T. J., DiMaggio, P. A., Wasney, G. A., Siarheyeva, A., Dong, A., Tempel, W., Wang, S.-C., Chen, X., Chau, I., Mangano, T. J., Huang, X.-p., Simpson, C. D., Pattenden, S. G., Norris, J. L., Kireev, D. B., Tripathy, A., Edwards, A., Roth, B. L., Janzen, W. P., Garcia, B. A., Petronis, A., Ellis, J., Brown, P. J., Frye, S. V., Arrowsmith, C. H., and Jin, J. (2011) A chemical probe selectively inhibits G9a and GLP methyltransferase activity in cells, *Nat. Chem. Biol. 7*, 566-574.
35. Richon, V. M., Johnston, D., Sneeringer, C. J., Jin, L., Majer, C. R., Elliston, K., Jerva, L. F., Scott, M. P., and Copeland, R. A. (2011) Chemogenetic Analysis of Human Protein Methyltransferases, *Chem. Biol. Drug Des. 78*, 199-210.
36. Mori, S., Iwase, K., Iwanami, N., Tanaka, Y., Kagechika, H., and Hirano, T. (2010) Development of novel bisubstrate-type inhibitors of histone methyltransferase SET7/9, *Bioorg. Med. Chem. 18*, 8158-8166.
37. Kung, P.-P., Huang, B., Zehnder, L., Tatlock, J., Bingham, P., Krivacic, C., Gajiwala, K., Diehl, W., Yu, X., and Maegley, K. A. (2015) SAH derived potent and selective EZH2 inhibitors, *Bioorg. Med. Chem. Lett. 25*, 1532-1537.
38. Geze, M., Blanchard, P., Fourrey, J. L., and Robert-Gero, M. (1983) Synthesis of sinefungin and its C-6' epimer, *J. Am. Chem. Soc. 105*, 7638-7640.
39. Li, Z., Nimtz, M., and Rinas, U. (2011) Optimized procedure to generate heavy isotope and selenomethionine-labeled proteins for structure determination using Escherichia coli-based expression systems, *Appl. Microbiol. Biotechnol. 92*, 823-833.
40. Kabsch, W. (2010) XDS, *Acta Crystallogr., Sect. D: Struct. Biol. 66*, 125-132.
41. Evans, P. (2006) Scaling and assessment of data quality, *Acta Crystallogr., Sect. D: Struct. Biol. 62*, 72-82.
42. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments, *Acta Crystallogr., Sect. D: Struct. Biol. 67*, 235-242.
43. Navaza, J. (1994) AMoRe: an automated package for molecular replacement, *Acta Crystallogr., Sect. A 50*, 157-163.
44. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method, *Acta Crystallogr., Sect. D: Struct. Biol. 53*, 240-255.

TABLES

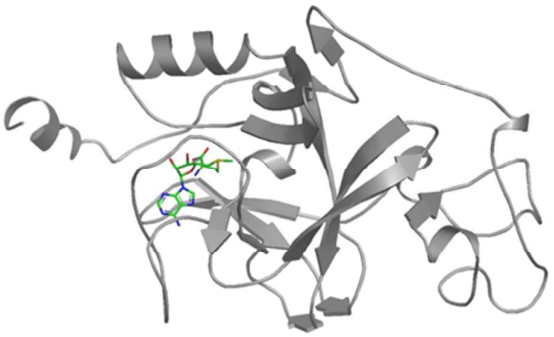
Table 1. Structure-activity relationships of sinefungin analogues *vs* MMSET and SETD2 catalytic SET domain constructs. IC_{50} values were generated using AlphaScreen assays for both proteins. For MMSET recombinant nucleosomes were used as substrate with $H3K36Me_2$ antibody detection of product formation. For SETD2 a $H3K36Me_2$ peptide was used as substrate with $H3K36Me_3$ antibody detection of product. K_d values for MMSET were generated by isothermal titration calorimetry.

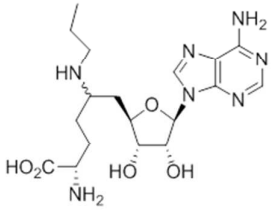


Compound	R1	R2	R3	IC_{50} / µM (*n*)*		Kd / µM
				MMSET	SETD2	MMSET
SAH	-	-	-	54% @ 30 (*1*)	33% @ 1 (*1*)	-
Sinefungin (2)	H	CO_2	NH_2	26±4.5 (*3*)	71% @ 30 (*1*)	35
3	nPr	CO_2	NH_2	3.3±1.0 (*4*)	0.49 (*1*)	1.6
4	sBu	CO_2	NH_2	1.8±0.4 (*2*)	0.29 (*1*)	-
5	$CH_2CH(OH)Me$	CO_2	NH_2	48% @ 300 (*1*)	46% @ 30 (*1*)	-
6	$(CH_2)_3OH$	CO_2	NH_2	69 (*1*)	36% @ 30 (*1*)	-
7	$(CH_2)_2NHMe$	CO_2	NH_2	63% @ 1000 (*1*)	-	-
8	nPent	CO_2	NH_2	20±11 (*4*)	3.9±0.2 (*2*)	-
9	$PhCH_2$	CO_2	NH_2	52% @ 1000 (*1*)	1.2 (*1*)	-
10	nPr	H	NH_2	55% @ 300 (*1*)	-	-
11	nPr	CO_2	H	9% @ 1000 (*1*)	-	-

*IC_{50} values are quoted as geometric means of the number of replicates *n* in parentheses. Where an IC_{50} curve could not be fitted, the concentration giving closest to 50% inhibition is indicated for comparative purposes.

Table of Contents graphic



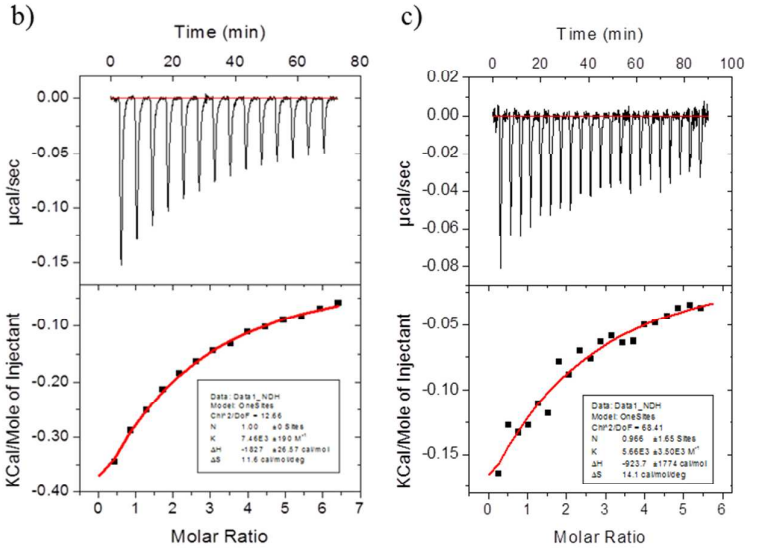


3: MMSET IC_{50} 3.3 μM
MMSET K_d 1.6 μM (ITC)

FIGURES

Figure 1. a) Structure of UNC0638. b)-c) Isothermal calorimetry traces for UNC0638 *vs* MMSET and NSD1 SET domains respectively: a K_d of 134 µM was determined vs MMSET, while binding to NSD1 was too weak to determine a K_d. d) ^{15}N-^{1}H HSQC 2D-NMR spectrum of labelled MMSET-SET domain showing chemical shift perturbations in MMSET signals (blue) observed by upon incubation with 1mM UNC0638 (red). Residues showing significant changes in chemical shift or intensity are highlighted.

a)



d)

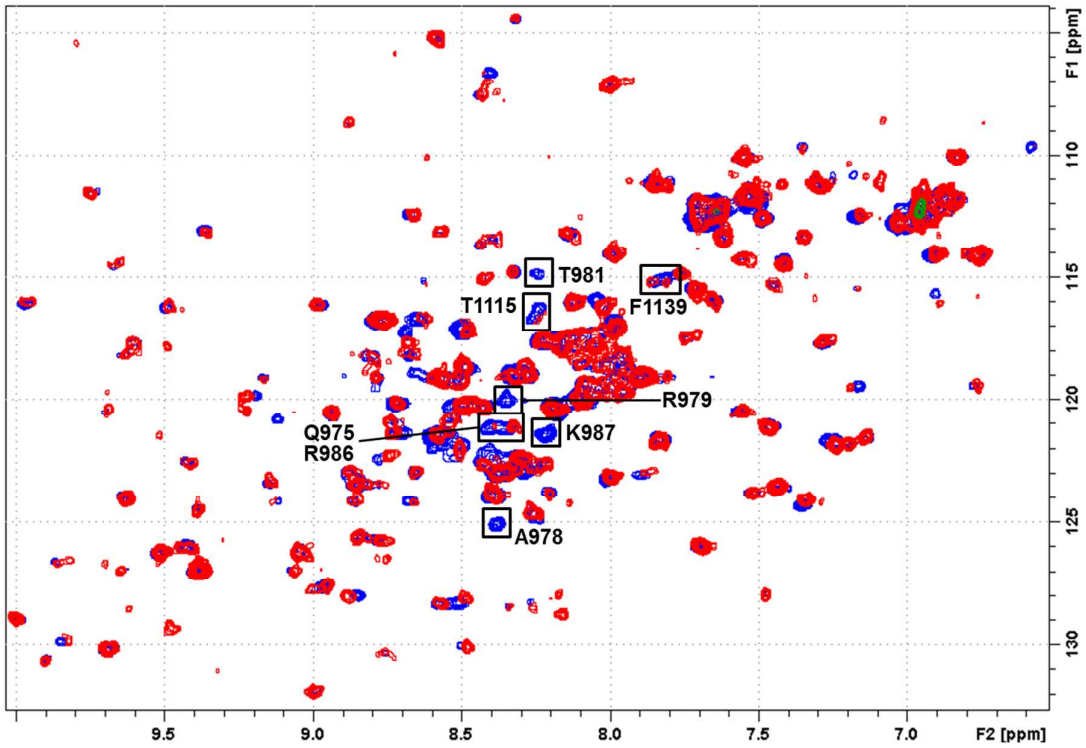
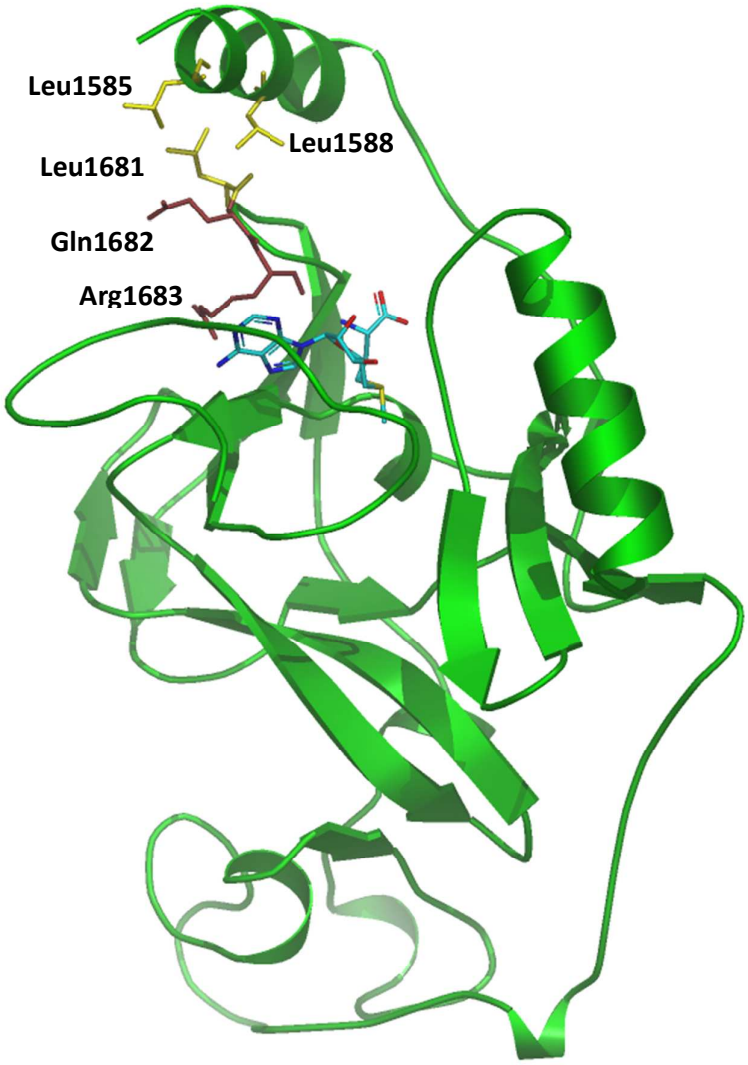


Figure 2: a) Three Leu residues in NSD1 (yellow) that form a hydrophobic interface between the N-terminal helix and a loop region (Gln1682 and Arg1683 coloured in red) adjacent to the SAM binding site. b) Sequence alignment of NSD1 and MMSET in these two regions showing that these residues are not conserved in MMSET. Dashed blue line - region of N-term helix of NSD1, red triangles – residues mutated in MMSET to corresponding residues in NSD1.

a)



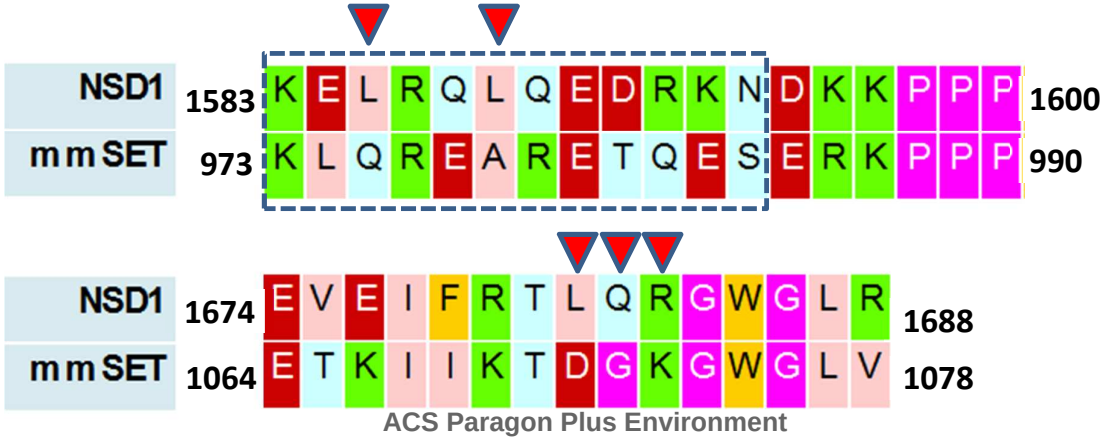


Figure 3. Overlay of novel MMSET structure (magenta) with previously published NSD1 structure (green)[30] showing the difference in orientation of the N-terminal α-helix.

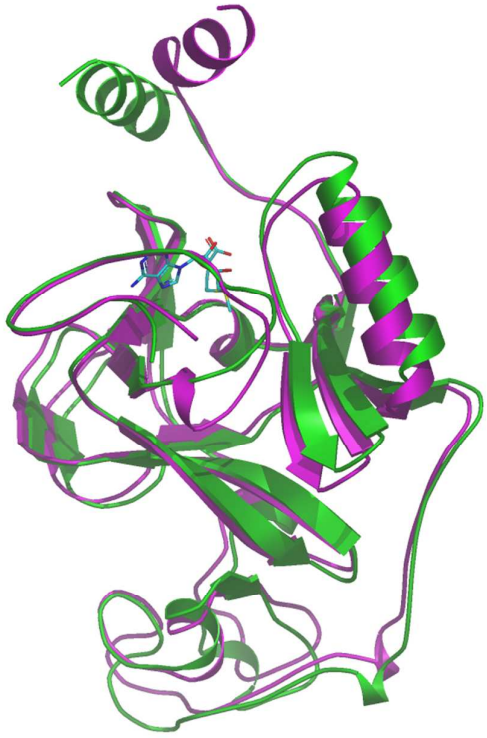


Figure 4. a) Analysis of the SAM binding site of MMSET. H-bonding interactions are observed between: SAM N1 and R1192 main chain NH; SAM 6-NH_2 and H1142 main chain C=O; SAM N7 and H1142 main chain NH; SAM 2'-OH and T1115 main chain C=O and H1116 side chain; SAM 3'-OH and H1116 main chain NH and N1186 side chain NH_2. The SAM-methionine amino group interacts with the W1075 main chain C=O and the N1141 side chain C=O; and the SAM-methionine carboxy group interacts with the W1075 main chain NH and Y1118 side chain OH. The SAM purine ring system makes good van der Waals contacts with the side chain of L1202. b) Despite being unoccluded by crystal contacts, the post-SET loop of MMSET (magenta) is in a closed conformation similar to the auto-inhibited conformation observed in NSD1 (green), albeit with a slight deviation around residues Asn1180-Cys1183 (residues coloured brown). c)-f) Superposition of cofactor and substrate channel for MMSET (magenta) with NSD1 (green) or SETD2 (yellow): c) SAM site, MMSET vs NSD1: Key amino acids with side chains contacting the SAM cofactor are conserved with the notable exception of H1116 (N in NSD1). d) SAM site, MMSET vs SETD2: Key amino acids with side chains contacting the SAM cofactor are conserved with the notable exception of N1186 (A in SETD2, interaction with SAM ribose 3'OH is functionally replaced by Q1676). e) Channel leading from the SAM methyl group to the substrate site, MMSET vs NSD1: all five amino acid side chains are identical. f) Channel leading from the SAM methyl group to the substrate site, MMSET vs SETD2: Y1092, F1177 and Y1179 are identical; L1120 is replaced by the sterically similar M1607 in SETD2, and F1163 is replaced by the less bulky F1650 in SETD2, creating a larger channel consistent with its ability to accommodate its H3K36Me_3 product, while MMSET (and NSD1) are not able to methylate beyond H3K36Me_2.

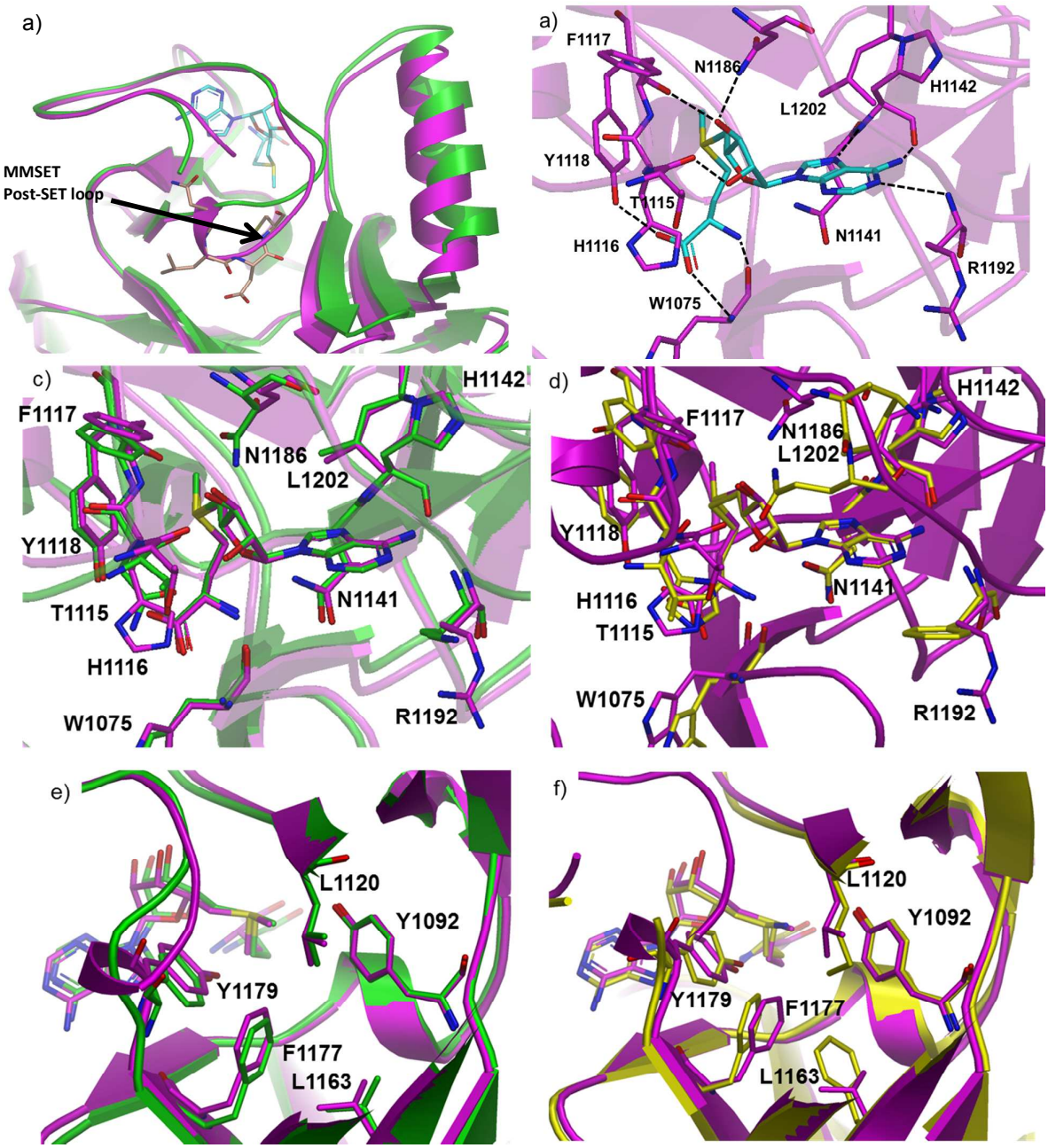

a)
MMSET
Post-SET loop
a)
F1117
N1186
H1142
L1202
Y1118
T1115
H1116
N1141
R1192
W1075
c)
H1142
F1117
N1186
L1202
Y1118
N1141
T1115
H1116
W1075
R1192
d)
H1142
F1117
N1186
L1202
Y1118
N1141
H1116
T1115
W1075
R1192
e)
L1120
Y1092
Y1179
F1177
L1163
f)
L1120
Y1092
Y1179
F1177
L1163

Figure 5. (a-b) The orientation of the N-terminal helix of MMSET-SET (magenta) differs considerably from those of NSD1 (green) and SETD2 (yellow). This helix is involved in crystal contacts within the crystal lattice unlike the corresponding regions of NSD1 (3ooi) and SETD2 (3h6l), which make intra-molecular interactions but no crystal contacts within their respective crystal systems. The molecular surface for SAM bound to MMSET is shown in blue gridlines for reference.

(a) (b)

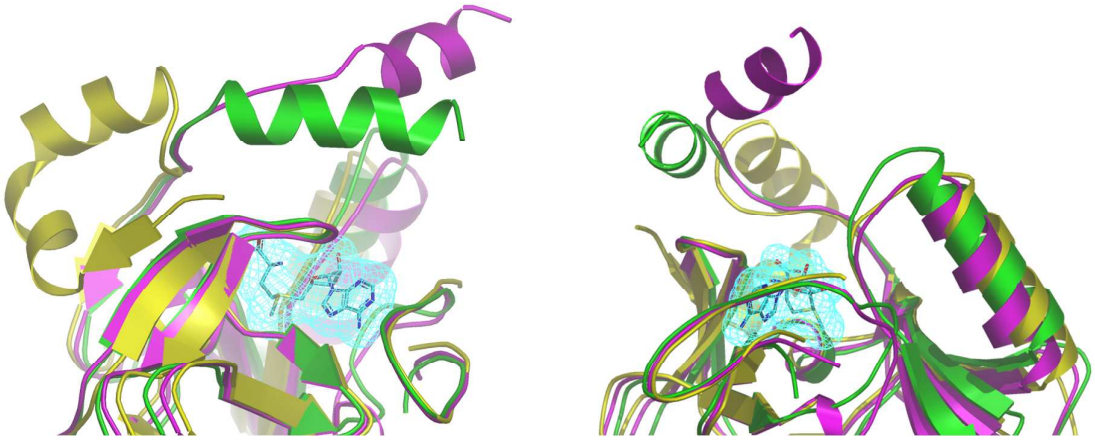


Figure 6. a) IC_{50} determination for n-propyl sinefungin (**3**) vs MMSET-SET domain. Data were generated using an AlphaScreen assay with recombinant nucleosomes as substrate and $H3K36Me_2$ antibody detection of product formation. b)-c) K_d determinations by ITC for sinefungin (**1**) and n-propyl sinefungin (**3**) respectively *vs* MMSET-SET domain.

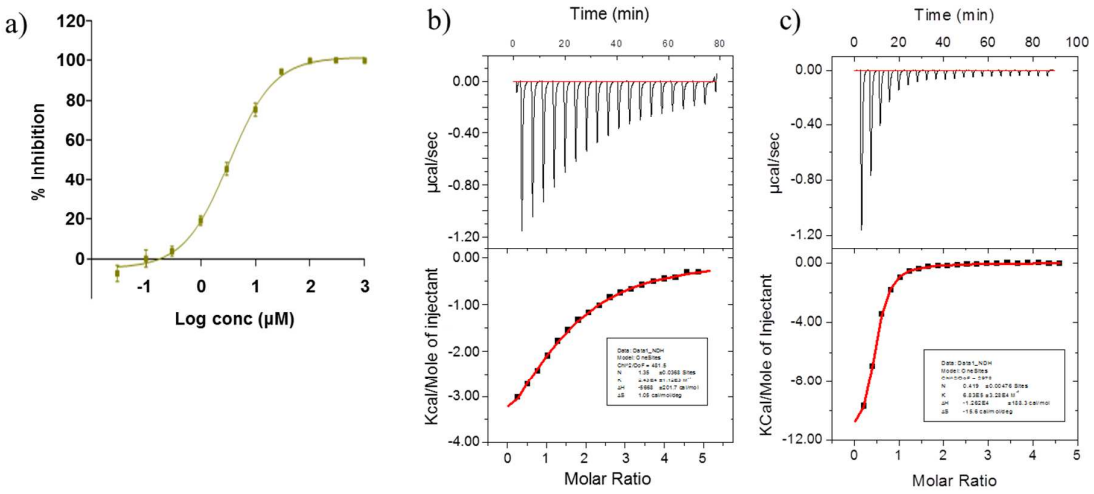
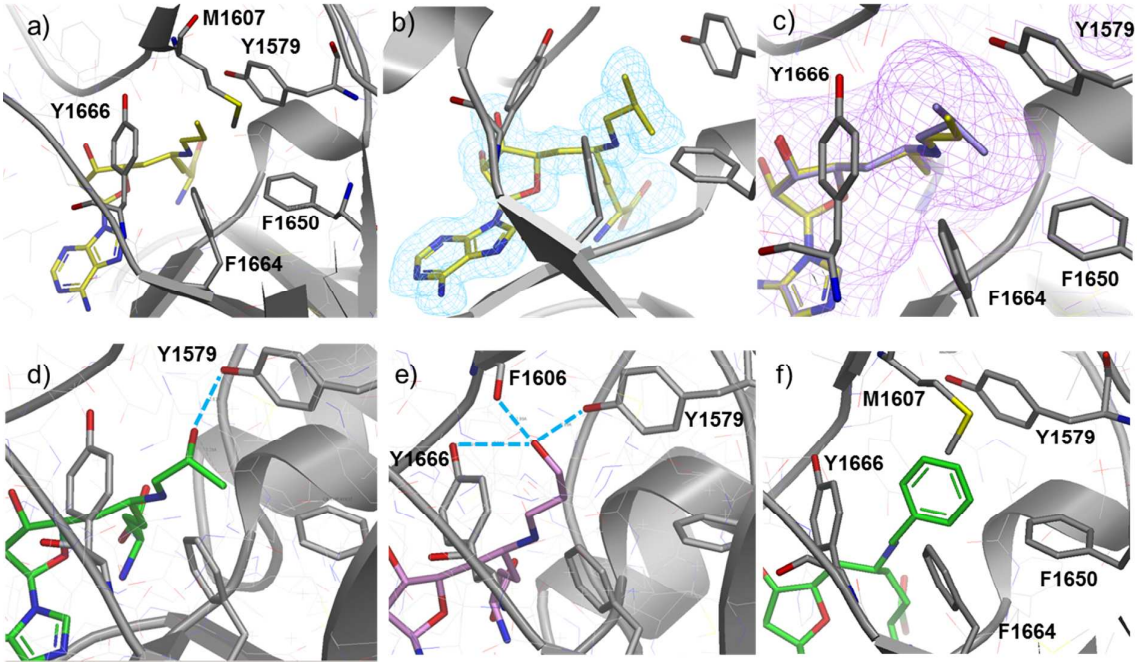
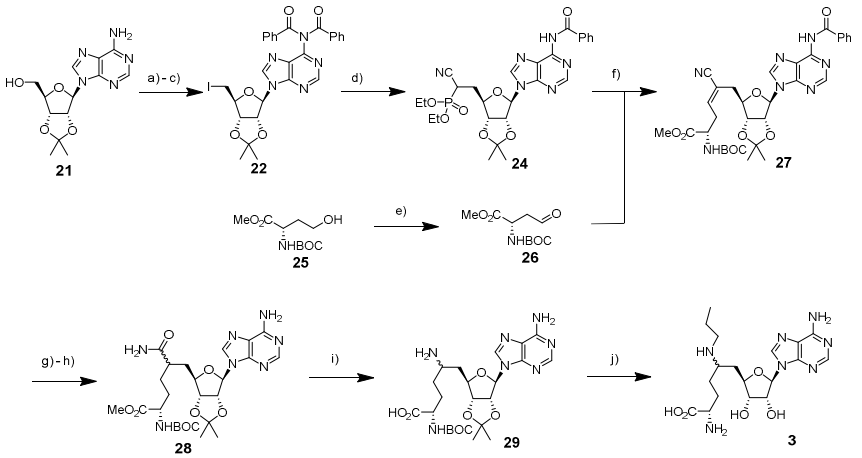


Figure 7. Crystal structures of SETD2 – *N*-alkyl sinefungin complexes. (a) The propyl group of **3** inserts into a channel formed by hydrophobic side chain residues from the post-SET loop (Y1666, Y1664) and body of the enzyme (M1607, Y1579, F1650). (b) Electron density map for **4** in complex with SETD2. (c) The structure of **4** in complex with SETD2 shows good steric complementarity as the small hydrophobic pocket formed by Y1579, F1650 and F1664 is filled by the additional methyl group (protein surface shown in magenta). (d) The secondary alcohol **5** engages in a hydrogen bond to the hydroxyl group of Y1579. (e) The 3-substituted primary alcohol **6** interacts *via* three new H-bonds between the inhibitor hydroxyl group and the hydroxyl groups of Y1579 and Y1666 and the backbone carbonyl of F1606. (f) The benzyl group of **9** fills the lipophilic pocket created by the side chains of Y1579 and 1666 and by F1650 and 1664, with the side chain of M1607 forming a lid.



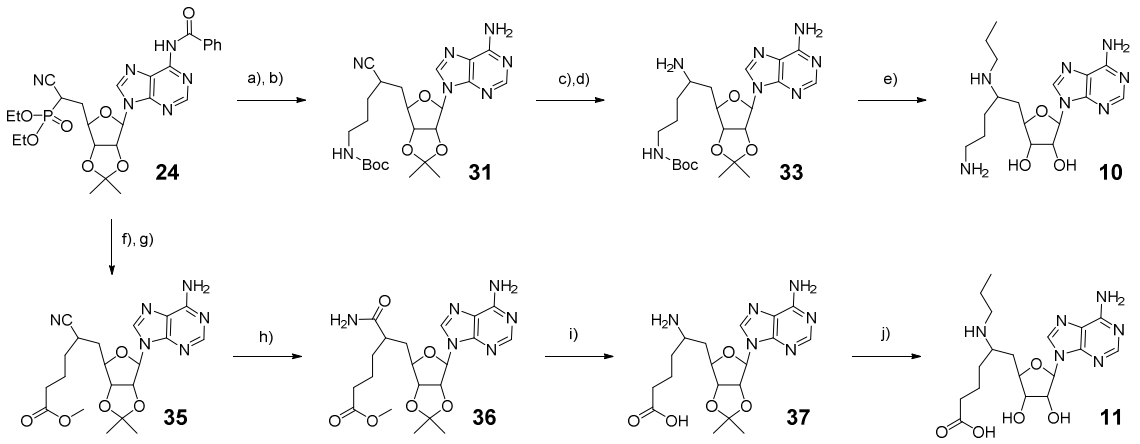
SCHEMES

Scheme 1. Synthetic route to *N*-substituted sinefungin derivatives exemplified by *N*-propyl sinefungin (3).



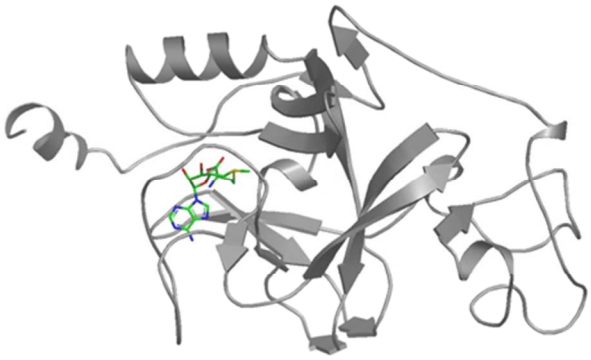
Reagents and conditions: a) TMSCl, py, 0°C, then BzCl, 68%; b) TsCl, DCM, py, 0°C, 76%; c) NaI, acetone, 75%; d) $(EtO)_2P(O)CH_2CN$, NaH, DMSO, 81%; e) Dess-Martin periodinane, DCM, 74%; f) $Mg(OH)_2$, MeOH, 0°C, 59%; g) Mg, MeOH, 69%; h) H_2O_2, NaOH, DMSO, MeOH, 55°C, 47%; i) $PhI(CF_3COO)_2$, H_2O, DMF, py, 29%; j) EtCHO, $NaCNBH_3$, MeOH, then TFA, H_2O, 16%.

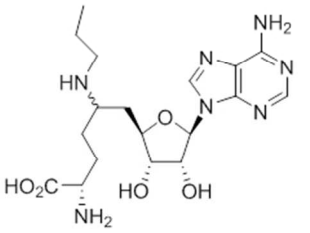
Scheme 2. Synthesis of *N*-propyl sinefungin derivatives 10-11.



Reagents and conditions:

a) $Mg(OH)_2$, MeOH, 0°C; b) Mg, MeOH, 36% over 2 steps; c) H_2O_2, NaOH, DMSO, MeOH, 55°C, 84%; d) $PhI(CF_3COO)_2$, H_2O, DMF, py, 48%; e) EtCHO, $NaCNBH_3$, MeOH, then TFA, H_2O, 3%; f) $Mg(OH)_2$, MeOH, 0°C; g) Mg, MeOH, 35% over 2 steps; h) H_2O_2, NaOH, DMSO, MeOH, 55°C, 29%; i) $PhI(CF_3COO)_2$, H_2O, DMF, py, 15%; j) EtCHO, $NaCNBH_3$, MeOH, then TFA, H_2O, 30%.





3: MMSET IC_{50} 3.3 μM
MMSET K_d 1.6 μM (ITC)

Table of contents graphic

205x77mm (96 x 96 DPI)