

Structural analysis of human KDM5B guides histone demethylase inhibitor development

Authors: Catrine Johansson^{1,2}, Srikannathasan Velupillai¹, Anthony Tumber^{1,3}, Aleksandra Szykowska¹, Edward Hookway², Radoslaw P Nowak^{1,2}, Claire Strain-Damerell¹, Carina Gileadi¹, Martin Philpott², Nicola Burgess-Brown¹, Na Wu², Jola Kopec¹, Andrea Nuzzi^{1,3}, Holger Steuber⁴, Ursula Egner⁴, Volker Badock⁴, Shonagh Munro⁵, Nick B LaThangue⁵, Sue Westaway⁶, Jack Brown⁶, Nick Athanasou², Rab Prinjha⁶, Paul E Brennan^{1,3} and Udo Oppermann^{1,2}

Affiliations:

¹ Structural Genomics Consortium, University of Oxford, Headington, OX3 7DQ, UK

² Botnar Research Centre, NIHR Oxford Biomedical Research Unit, University of Oxford OX3 7LD, UK

³ Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, UK

⁴ Bayer Healthcare Pharmaceuticals, 13353 Berlin, Germany

⁵ Department of Oncology, University of Oxford, Oxford, OX3 7DQ, UK

⁶ Epinova DPU, Immuno-Inflammation Therapy Area, GlaxoSmithKline R&D, Stevenage, UK SG1 9NY

Corresponding authors: udo.oppermann@sgc.ox.ac.uk, catrine.johansson@chem.ox.ac.uk

ABSTRACT

Members of the *KDM5* (also known as JARID1) family are 2-oxoglutarate and Fe^{2+} dependent oxygenases acting as histone H3K4 demethylases, thereby regulating proliferation, stem cell self-renewal and differentiation. Here we report crystal structures of the catalytic core of the human KDM5B enzyme in complex with three inhibitor chemotypes. These scaffolds exploit several aspects of the KDM5 active site, while their selectivity profiles reflect the hybrid features with the KDM4 and KDM6 families. Whereas GSK-J1, a previously identified KDM6 inhibitor shows about 7-fold less inhibitory activity towards KDM5B, KDM5-C49, displays >25-100 -fold selectivity between KDM5B and KDM6B. The cell-permeable derivative KDM5-C70 provides an anti-proliferative effect in myeloma cells, leading to genome-wide elevated H3K4me3 levels. The selective inhibitor GSK467 exploits unique binding modes, however lacks cellular potency in the myeloma system. Taken together, these structural leads deliver multiple starting points for further rational and selective inhibitor design.

INTRODUCTION

Methylation of lysine residues on histone tails is a reversible epigenetic modification that plays a key role in gene regulation. The transcriptional response is dependent on the specific residue methylated, the number of methyl groups added or removed, as well as on a complex chromatin cross-talk with other chromatin regulators and transcription factors, suggested to fine-tune transcription¹⁻³. Removal of methyl groups from lysine residues is catalysed by two classes of histone lysine (K)-demethylases (KDMs): the flavine adenine dinucleotide (FAD)-dependent amino oxidases⁴ (grouped as KDM1 subfamily)⁵ and the Fe(II) and 2-oxoglutarate (2OG) dependent oxygenases that contain a conserved catalytic Jumonji C (JmjC) -domain (belonging to subfamilies KDM2- KDM7)^{2,3}. In contrast to the KDM1 demethylases which are limited to demethylate mono- and di- methyl lysine residues, the JmjC-type KDMs are able to catalyse demethylation of all 3 possible methylation states of methyl-lysyl residues. Distinctive features between the KDM subfamilies include defining domain organisation patterns as well as site- and methylation state specificity for the different methyl marks^{2,3}. The reaction mechanism for the JmjC-demethylases is dependent on molecular oxygen and proceeds through the oxidation of 2OG and Fe(II), yielding CO₂, succinate and a highly reactive Fe(IV)- oxoferryl intermediate⁶.

Members of the KDM5 subfamily of JmjC-KDMs are transcriptional co-repressors by specifically catalysing the removal of all possible methylation states from lysine 4 of histone H3 (H3K4me3/me2/me1), a histone modification that occurs largely around transcriptional start sites of actively transcribed genes⁷. KDM5 enzymes are often found as components of transcriptional complexes with repressors such as REST, histone deacetylases or histone methyl transferases (HTMs)⁸. In mammals the KDM5 subfamily encompasses four proteins, KDM5A (known as JARID1A or RBP2), KDM5B (known as JARID1B or PLU1), and the KDM5C (JARID1C or SMCX) and KDM5D (JARID1D or SMCY) members, the latter two encoded on the X and Y chromosomes, respectively.

The KDM5 family is conserved from yeast to humans, displaying a similar domain architecture with an N-terminal Jumonji (JmjN) domain, a DNA binding ARID domain (AT-rich interactive domain), a catalytic JmjC-domain, a C5HC2 zinc finger motif located C-terminally to the JmjC-domain, a PLU-1 motif, as well as two to three methyl-lysine or methyl-arginine binding plant homeodomain (PHD) domains denoted PHD1, PHD2 and PHD3 (**Fig. 1**). These additional domains contribute critically to genomic KDM5 target gene

occupation, for example PHD domains bind to modified lysine residues in a sequence-specific manner⁹. However, the complex role of the particular KDM5 PHD domains is not fully understood at present. Whereas in KDM5C the PHD1 domain binds the H3K9me2/me3 methyl mark, the same domain in KDM5B, and to some extent also KDM5A, recognises preferentially unmethylated H3K4 marks, the product of the KDM5 mediated demethylation reaction. Thus, at least in KDM5B and KDM5A¹⁰ the PHD1 domain, sandwiched between the JmjN and C domains, may protect H3K4 from re-methylation¹¹ and moreover could allosterically regulate the catalytic activity of the Jmj domain¹².

The KDM5 enzymes play pivotal roles both during normal development and in pathological conditions^{2,13}, with KDM5A and KDM5B linked to control of cell proliferation, differentiation, and to several cancer types. KDM5C plays a role in neuronal development, whereas the Y-chromosome encoded KDM5D is widely expressed and involved in spermatogenesis.

Development of chemical tools to interrogate KDM5 biology is progressing slowly^{13,14}- this impasse prompted us to determine crystal structures of the catalytic core of human KDM5 enzymes, compare them to related KDM4 and KDM6 enzymes, and to characterise KDM inhibitors, which were used to investigate H3K4me3 epigenomic signatures. The complex structures with distinct chemotypes reveal several distinguishing features and plasticity of the KDM5 cofactor site, which could be utilized in further inhibitor development.

RESULTS

ARID and PHD domains are dispensable for activity - Full-length human KDM5B is a 1544-residue protein that requires the JmjN, ARID, PHD1, JmjC domain as well as the C5HC2 zinc finger for *in vivo* demethylase activity¹⁵. To further understand the effects of domain structure on catalytic activity and to facilitate structural studies of the Jmj catalytic core we first engineered deletion constructs and analysed their enzymatic properties. Assuming that the C-terminal PLU and PHD2 and PHD3 domains were dispensable for enzymatic activity, a KDM5B-c1 construct encompassing the JmjN, ARID, PHD1, JmjC and the zinc finger motif was cloned and found to be active (**Fig. 1a-b, Supplementary Results, Supplementary Table 1, Supplementary Fig. 1-2**). Based on structural analysis^{16,17} of other KDMs that reveal close contacts between the JmjN and JmjC domains, we next designed deletion constructs of KDM5B-c1 where the PHD1 domain and parts of the ARID domain

were deleted (KDM5Bc-2- KDM5Bc4; **Fig. 1a**). All proteins were expressed as N-terminally His-tagged proteins in baculovirus infected SF9 insect cells and purified using Ni-affinity resin followed by size-exclusion chromatography. To probe the effect of the ARID and PHD1 deletion, proteins were analysed for H3K4me3/me2 demethylase activity using either a formaldehyde dehydrogenase (FDH) coupled assay or a direct Rapid-Fire mass spectrometry (RF-MS) assay^{18,19} (**Fig. 1b**, **Supplementary Fig. 1**, **Supplementary Tables 1 and 2**). Activity was detected for the deletion constructs KDM5B-c2 (deletion encompassing residues Asn¹⁰²-Gly³⁶⁹; added –GGGG- linker), KDM5B-c3 (deletion comprising residues Asn¹⁰²-Ala³⁷³) and KDM5C-4 (deletion comprising residues Asn¹⁰²-Ala³⁷³, added –GGGG- linker). We note that cleavage of the His-tag, re-purification and freezing makes the KDM5B-c4 construct amenable to gradual loss of activity, however provides material that can be used for crystallisation. Using the FDH-coupled assay and a histone H3(1-21)K4me2 peptide as substrate, KDM5B-c2 revealed approximately 40% of the catalytic efficiency (k_{cat}/K_m) compared to KDM5B-c1 (**Fig. 1b**, **Supplementary Table 1**), whereas KDM5-c3 and KDM5-c4 retained only 5 and 1.6 %, respectively. The apparent K_m value for a H3(1-21)K4me2 peptide was 0.85 μ M for KDM5B-c1, 2.25 μ M for KDM5B-c2, 12.2 μ M for KDM5B-c3 and 51.1 μ M for KDM5B-c4. These data suggest that the ARID and the PHD1 domains are not per se required for activity but contribute to recognition of the H3(1-21)K4me2 substrate peptide. This is in agreement with results for the human paralog KDM5A¹², showing the allosteric effects of PHD1 on catalytic activity.

The catalytic core of human KDM5B is conserved - We initially carried out unsuccessful crystallographic studies on intact KDM5B-c1, prompting us to perform crystallisation trials with the KDM5B deletion constructs encompassing the JmjN-, the JmjC domain and the C5HC2 zinc finger motif. Crystallisation trials using KDM5B-c4 identified two conditions from which several structures were solved in their apo form, in complex with the cofactor 2OG, and with generic as well specific KDM inhibitors (**Supplementary Table 3**). The initial structure of KDM5B-c4, crystallised with inhibitor KDM5-C49 (**1**), was solved by molecular replacement using KDM4A (PDB ID 2bp5) as a search model. Later on, diffracting crystals were also obtained for construct KDM5B-c2, confirming the ligand binding modes observed for KDM5B-c4.

The overall fold of this catalytic core reveals three conserved domains: the JmjN (residues 31-72) and the JmjC domain (residues 375-602) associate tightly and form together with seven residues of the ARID domain (residues 94-100) an extended double-stranded β -helix

(DSBH), characteristic of the 2OG oxygenase superfamily (labelled in pink, blue and yellow, **Fig. 2a**), a C-terminal helical domain (residues 604-671, 737-753, green), and a beta-sheet composed of 3 beta-strands (residues 673-734, cyan) (**Supplementary Fig. 3**) that harbours a C5HC2-zinc finger motif. A Mn^{2+} ion, used in the experiment to replace the active site Fe^{2+} , has an octahedral coordination through His⁴⁹⁹, Glu⁵⁰¹ and His⁵⁸⁷, all part of the conserved HxD/E..H metal chelation motif found in 2-OG oxygenases²⁰ and adjacent to the cofactor-binding site (**Fig. 2b**). In most structures a HEPES buffer molecule is bound in a cavity at the interface between the extended DSBH-fold and a helical domain (**Supplementary Fig. 4**). The C-terminal helical domain is composed of 4 helices, and a Zn-finger C5HC2 –motif is found (**Fig. 2a** and **Fig. 2c**), similar to the GATA-like motif in the KDM6 subfamily^{21,22}. In all structures a Zn^{2+} molecule is coordinated in a tetrahedral geometry by Cys⁶⁹², Cys⁶⁹⁵, Cys⁷¹⁵ and His⁷¹⁸ (**Fig. 2c**, **Supplementary Fig. 5**). Interestingly, this domain appears to be sensitive to the redox environment. In crystals grown and mounted within a 3-6 day period a second Zn^{2+} molecule is observed and coordinated by Cys⁷⁰⁶, Cys⁷⁰⁸, Cys⁷²³ and Cys⁷²⁵ as noted in the ligand structures of KDM5B with 2,4-PDCA (**2**), *N*-oxalylglycine (NOG, **3**), GSK-J1 (**4**) and GSK467 (**5**). In the complex structure of KDM5B with the inhibitor KDM5-C49, determined from a crystal that appeared after three weeks, two intra-molecular disulfide bonds are observed (**Supplementary Fig. 5**). In KDM5 proteins the C5HC2 zinc finger motif is required for its *in vivo* catalytic activity¹⁵ and *in vitro* data show that demethylase activity of KDM5B-c1 is increased in the presence of the reducing agent TCEP (**Supplementary Fig. 6**). This potentially redox-sensitive motif is similar in packing to the Jmj core as observed in the KDM6 proteins with a GATA-like domain^{21,22}, which is involved in substrate binding with a putative induced fit mechanism for substrate recognition. At present it is unknown if this oxidation/reduction mechanism has an impact upon enzymatic activity of KDM5B and further studies will reveal its possible physiological implications.

An overlay of the KDM5B-c4 complex structures provides an overall C_{α} rmsd value of 0.36-0.98Å, with largest differences found in the end of beta strand 5 and the loop connecting $\beta 5$ and $\beta 6$ (**Supplementary Fig. 7**). Alternative conformations are observed for Tyr⁴²⁵, and Asp⁴²⁸ is found on either side of the DSBH-fold, (**Supplementary Fig. 7**), resulting in a cofactor pocket of different size and shape.

Comparison to KDM4 and KDM6 subfamilies - Superimposition of the KDM5B structures with KDM6A (PDB ID 3AVR) or KDM4A (PDB ID 2P5B) shows that the domain architecture and the overall fold is similar to the KDM6 enzymes, however major differences

are found in the loop structures and the JmjC domain (overall backbone rmsd value of C_{α} = 2.3 Å). Instead the KDM5B JmjC domain is more related to its counterpart in KDM4A (**Supplementary Fig. 8**), although the Zn-binding substructure found in KDM4A¹⁷ is missing in KDM5B. The surface charge distribution between the proteins is dissimilar, likely reflecting different substrate specificities and substrate binding sites (**Fig. 3a-c**). Interestingly, β -strand 5 in KDM5B, that is located close to the ARID and PHD1 deletion site, is found at a slightly different position in both KDM6A and KDM4A. In the latter proteins, this β -strand projects 7-8 Å away from the 2OG-binding pocket compared to KDM5B, rendering interpretations and comparisons difficult since the corresponding strands in KDM4 and KDM6 harbour arginine or asparagine residues (Arg¹⁰²⁷ and Asn⁸⁶) in homologous sequence positions to Arg⁹⁸. To explore whether the location of Arg⁹⁸ in the KDM5B-c4 deletion construct affects activity we mutated this residue to a glycine or a lysine residue, which occupies in the structure of KDM4A the same position as Arg⁹⁸, but is located on a distinct secondary structure element. Using the FDH coupled assay and a histone H3(1-21)K4me2 peptide as substrate, both mutants were found to be active, with lower K_m and higher catalytic efficiencies compared to KDM5B-c4 (**Fig. 1b** and **Supplementary Table 1**) indicating indeed a critical role for Arg⁹⁸. These data suggest that despite measurable activity, the engineered KDM5B-c4 construct has limitations for correlating full-length construct activities, however the data also highlight the utility of this construct for analysis of inhibitor binding modes. The complex behaviour of this structural segment around Arg⁹⁸ is further emphasized by the fact that the crystal structure of construct KDM5B-c2, (PDB ID 5f3V) revealed that Arg⁹⁸ is found in an identical position as in construct KDM5-c4 (PDB ID 5a3w). This construct, in terms of kinetic constants more similar to construct KDM5B-c1, shows that Lys¹⁰⁰ has moved approximately 7.8 Å away from the 2OG- binding pocket (**Supplementary Fig. 9**) indicating that this part of the construct indeed contains a flexible segment that has an impact on catalytic behaviour. Taken together, these data suggest that substrate binding is mediated in part by elements between the PHD and the JmjC domain, however significant interpretations of inhibitor binding can be derived from either KDM5B-c2 or KDM5B-c4 ligand structures.

Comparison of the 2OG-binding pocket between KDM5B (PDB ID 5A1F) and KDM6A reveals significant differences (**Fig. 3d**), whereas superimposition of the same structure of KDM5B (PDB ID 5A1F) with KDM4A (PDB ID 2P5B) reveals a more similar shape of the 2OG- binding pocket (**Fig. 3e**) where the most significant differences are Trp⁴⁸⁶ and Ala⁵⁹⁹.

In KDM4A (PDB ID 2OQ6) Asn⁸⁶, Asp³¹¹, Asp¹³⁵, Lys⁴²¹ and Glu¹⁶⁹ interact with the H3K9me3 peptide and the different residues found in KDM5B could possibly reflect the different substrate specificity i.e. methylated H3K4 versus H3K9 as in KDM4A²³.

Domain arrangement of KDM5B determined by SAXS - Building on the high-resolution models of the catalytic core, we next characterised the KDM5B-c1 construct, carrying the additional PHD1 and ARID domains by small-angle X-ray scattering (SAXS) and rigid-body modelling. To study the domain arrangement in this construct we determined the SAXS solution structure of the KDM5B-c1 and KDM5B-c4 constructs (**Fig. 3f, Supplementary Fig. 10**). A possible model derived from these experiments suggests formation of a large surface area by the 270- odd residue insertion containing the ARID-PHD1 section and places the PHD1 domain into close contact with the catalytic core, in line with experimental data suggesting co-operativity between PHD1 and the catalytic core in KDM5 enzymes. This additional surface is possibly suited to make contacts with nucleosomes, the ultimate substrates for KDM5B, and would allow a postulated demethylation spreading across defined chromatin regions¹².

KDM5B is amenable to selective inhibitor development – Generic inhibitors such as *N*-oxalylglycine (NOG) and 2,4-pyridinedicarboxylic acid (2,4-PDCA) coordinate the catalytic metal in a bi-dentate manner and not surprisingly, these inhibitor interactions in KDM5B are largely conserved when compared to other 2-OG oxygenases²⁴. We studied different inhibitor chemotypes, identified from various focussed synthesis and ligand screening campaigns^{19,21,25-27} for KDM5B activity. To understand specificities we profiled these molecules towards a selectivity panel of different Jmj demethylase subfamilies (**Supplementary Tables 4 and 5**). We identified compound KDM5-C49, a 2,4-PDCA analogue, which shows nanomolar inhibitory potencies in enzymatic assays (**Supplementary Table 4**) across several Jmj members and subfamilies. The complex structure with construct KDM5B-c4 reveals occupation of the 2OG-binding site, with the pyridine nitrogen and the aminomethyl nitrogen forming a bidentate interaction with the catalytic metal (**Fig. 4a and Supplementary Fig. 11**). The compound makes extensive polar interactions involving residues Lys⁵¹⁷, Tyr⁴²⁵, Asn⁵⁰⁹, Glu⁵⁰¹, Arg⁹⁸ and Tyr⁴⁸⁸, additionally hydrophobic interactions are mediated by Trp⁴⁸⁷, Val⁹⁹ and Phe⁴⁹⁶. An almost identical coordination is observed in the structure of human KDM5C (PDB ID 5fv3) with the inhibitor, determined from a similar ARID and PHD1 deletion construct as KDM5B-c4 (**Supplementary Fig. 12, Supplementary Table 5**). In contrast to the KDM5B structures however, we do not observe

in the KDM5C structure electron density for Arg⁹⁸ or the beta strands that correspond to β 5, β 12, or the loop connecting β 5 and β 6, again suggesting that this region is highly flexible. The structural relationships between KDM5B and the KDM4 and KDM6 subfamilies are also reflected by the selectivity profile of compound KDM5-C49. Calculation of K_i values from the IC₅₀ determinations^{28,29} reveals 25-150 -fold differences between KDM5 and KDM6 members, whereas the close relationship to the KDM4 family is indicated by 2-30 fold differences in K_i (**Supplementary Table 4**). To investigate this aspect further, the inhibitor was co-crystallised with the KDM6-related member UTY²² as well as with KDM4A¹⁷ (**Supplementary Table 5**). The structures reveal several decisive similarities and differences across these Jmj subfamilies (**Supplementary Fig. 13**): In both KDM4 and KDM6 structures, the KDM5-C49 inhibitor occupies the cofactor site and coordinates the metal in a similar manner as KDM5B. In KDM4A (PDB ID 5FPV), the interaction with the inhibitor KDM5-C49 is almost identical to KDM5B, explaining the somewhat low selectivity (7-8 -fold between KDM5B and KDM4C) of the inhibitor, compared to the KDM6 members (ca. 150-fold difference in K_i , KDM5B vs KDM6B) where lesser interactions are found (**Supplementary Fig. 13**). Comparison of UTY·KDM5-C49 with the structure of KDM6A (PDB ID 3AVR), in complex with NOG and H3K27me₃ peptide (**Supplementary Fig. 13**), reveals that the pyridine ring overlaps as expected with the NOG position, whereas the dimethyl amino ethyl portion in KDM5-C49 is located within the methylated H3K27me₃ binding pocket. This is further supported by an overlay of the KDM5B·KDM5-C49 complex structure with the related plant demethylase JMJ703 with a substrate H3K4me₃ peptide (PDB ID 4IGQ) (**Supplementary Fig. 13**). Assuming similar interactions are formed with the highly conserved KDM6B and KDM4C enzymes, we hypothesize that the somewhat fewer interactions detected in the KDM5-C49 UTY complex may explain the lower inhibitory activity, whereas the similar interactions observed with KDM4A could account for the low selectivity between KDM4C and KDM5B. However, in the absence of structural data for KDM5-c1 constructs that contain the ARID and PHD1 domains, which were used for the inhibitor assays, we cannot rule out that further stabilising interactions may be formed between the ethylamide portion and the additional elements present in the KDM5B-c1 and other KDM5 protein constructs. With this caveat further interpretations about structure-function relationships appear difficult.

In order to further map the relationships between KDM4, 5 and 6 subfamilies we used the previously identified inhibitor GSK-J1²¹, a KDM6 and KDM5 subfamily specific inhibitor

(**Fig. 4b** and **Supplementary Fig. 11**). In KDM5B-c4 GSK-J1 is bound to the 2OG- binding site in an almost identical mode as previously observed with UTY•GSK-J1²² and JmjD3/KDM6B•GSK-J1²¹. The heterocyclic biaryl ring system of GSK-J1 makes bidentate interactions with the catalytic metal, facilitated by a side-chain movement of Arg⁹⁸ to accommodate the inhibitor molecule. As also observed in the KDM6 enzymes, the Mn²⁺ ion translocates ~2.0 Å away from the HXE ...H metal chelation triad with GSK-J1 as compared to other structures, and this movement results in an indirect water-bridge interaction between His⁵⁸⁷ and the Mn²⁺ ion, confirming the previously observed elasticity of the active site metal position in 2OG oxygenases^{21,22}. The tetrahydrobenzazepine part of GSK-J1 has a bent conformation and is located in a cavity between β -strands 5 and 9, and this part of GSK-J1 has previously been shown to occupy the peptide-binding site in KDM6B²¹ and UTY (KDM6C)²² potentially explaining the higher selectivity towards the KDM6 enzymes.

Whereas the ligand structures of KDM5B with KDM5-C49 and GSK-J1 highlight the overlapping selectivity of the inhibitors with the KDM4 and KDM6 families, the structure of KDM5B with GSK467 (compound **5**) provides a possible template for selective KDM5B inhibitor development. GSK467 has been disclosed by GlaxoSmithKline (GSK) as part of a sub-micromolar inhibitor series for the KDM4 family and KDM5C with cellular activity (IC-50 <10 μ M) in cellular imaging assays³⁰. In our AlphaScreen assay GSK467 shows a calculated KI value of 10 nM for KDM5B (**Supplementary Table 4**) with an apparent 180-fold selectivity to KDM4C and no measurable inhibitory effects towards KDM6 or other JmJ members (**Supplementary Fig. 14**). GSK467A was co-crystallised with KDM5B revealing that GSK467 (**Fig. 4c** and **Supplementary Fig. 11**) is found in the 2OG cofactor-binding pocket, where the inhibitor makes a mono-dentate interaction with the catalytic metal via its pyrido-nitrogen, with the two remaining coordination sites occupied by water molecules. The pyrido[3,4-d]pyrimidine-4(3H)-one forms hydrophobic interactions with Trp⁵¹⁹, Phe⁴⁹⁶ and Tyr⁴⁸⁸, and the pyrimidin-4(3H)-one oxygen forms a polar interaction with Lys⁵¹⁷ (**Fig. 4c**). The benzyl ring is located in a cavity formed by Tyr⁴²⁵, Gln⁸⁸, Ala⁴²⁶ and Arg⁹⁸, and induces a movement of Arg⁹⁸ and the loop containing the Leu¹⁰¹-Ala³⁷³ fusion site away from the DSBH-fold (**Supplementary Fig. 7**). Thus, it is possible that further interactions are formed between this part of the molecule and the additional domains present in the KDM5B-c1 construct.

The inhibitor KDM5-C70 increases global H3K4me3 levels - Having identified KDM5-C49 and GSK467 as potent inhibitor scaffolds for KDM5 enzymes, we next set out to explore their utility in interrogating KDM5 biology. Given the role of KDM5 in cancer biology¹³ we studied the possible anti-proliferative roles of GSK467 and KDM5-C49 in the human multiple myeloma tumour cell line MM-1S. Treatment with GSK467 and KDM5-C70 (compound **6**, an ethyl ester derivative of KDM5-C49 to facilitate cellular permeability) shows anti-proliferative effects after 7 days of treatment (**Fig. 5a-b; Supplementary Fig. 15**) at elevated concentrations (estimated 50% reduction of viability/proliferation for KDM5-C70 ca. 20 μ M; for GSK467 >50 μ M). Neither KDM5-C70 nor GSK467 increased the rate of apoptosis in myeloma cells (**Supplementary Fig. 15**) but treatment with KDM5-C70 decreased the level of phosphorylation of retinoblastoma protein (pRb) compared with both the vehicle control and GSK467 (**Fig. 5c and Supplementary Fig. 16**) whilst leaving the total level of pRb unchanged, indicating impairment of cell cycle progression, and in line with the anti-proliferative effects observed. Chromatin immunoprecipitation followed by next generation sequencing (CHIPseq) showed an increase in H3K4me3 levels around transcription start sites with KDM5-C70 but little change with GSK467A (**Fig. 6a**) at 50 μ M inhibitor concentrations. When all H3K4me3 peaks are considered, which includes promoters and enhancers distant from the transcription start site, the increase seen with KDM5-C70 is more pronounced (**Fig. 6b, Supplementary Fig. 17**). Analysis between KDM5-C70 and the DMSO control reveals 2872 differentially bound peaks at a false discovery rate of 0.1 (**Supplementary Fig. 18, Supplementary Dataset 1**), of which 2728 show increased H3K4me3 with KDM5-C70. Examination of the genomic location of differentially bound peaks shows a modest number of peaks close to transcription start sites with a larger number being intergenic (**Fig. 6c**). Gene ontology (GO) analysis of the differential bound peaks increased with KDM5-C70 shows enrichment of terms associated with B cell malignancies whereas the much smaller number of peaks associated with decreased binding on addition of KDM5-C70 are associated with cell cycle progression (**Supplementary Table 6**). Examples of the transcription start sites of three genes, associated with myeloma biology, are given (**Fig. 6d-f**).

DISCUSSION

Although their precise physiological roles are still incompletely understood, the KDM5 family members have attracted attention due to their important roles in stem cell biology,

development and oncology¹³. For example, elevated levels of KDM5A and KDM5B are associated with poor prognosis in several cancer types, resulting in an increasing interest in identifying selective inhibitors. However, only partially selective KDM5 inhibitor chemotypes have been described^{31,32} and are based -like KDM5-C49- on cofactor site binding modes. To facilitate further development of novel and selective KDM5 inhibitors we set out to determine crystal structures of the catalytic core in human KDM5 enzymes.

These structural studies of human KDM5B and KDM5C, further supported by a recent structure of KDM5A³³, highlight several important aspects of Jmj-type histone demethylases, namely (i) the identification of structural relationships between the KDM4, KDM5 and KDM6 subfamilies, further extended to include domain organisation affecting features of the active sites and (ii) the suitability to guide identification of KDM5-selective chemotypes. First, the KDM5B structures reveal hybrid features when compared with the KDM4 and KDM6 families. Despite the conserved nature of the 2OG cofactor site, the residues building this part of the active site reveal the conformational plasticity as well as elasticity of the metal coordination as critical elements for cofactor and, importantly, inhibitor ligand binding, which when combined indeed allow identification of apparently selective inhibitor chemotypes such as the pyrido-pyrimidinone scaffold of GSK467. Furthermore, our study extends significantly other data^{12,17,21,34}, which have shown the importance of adjacent packing domains to the JmjN/C subdomains. These arrangements appear to be important for substrate binding (as observed in the KDM6 and KDM7 structures^{21,34}, or could serve as dimerization domain (as observed in the ribosomal hydroxylase MINA53³⁵). Although no full-length structures of the KDM4, 5 or 6 subfamilies have been solved, the partial structures obtained so far highlight the importance of adjacent domains not only for activity and substrate selectivity but possibly also for the stability of the catalytic Jmj domain itself. In this respect, the SAXS data provide additional clues on domain arrangement and a step forward towards understanding enzyme-chromatin interactions.

Lastly, we show that selective inhibitor compounds such as GSK467 or KDM5-C49/KDM5-C70 can be used as starting points for further chemical tool development and optimization. The CHIPseq experiments with the cell-permeable KDM5-C49 derivative (KDM5-C70) demonstrate a genome-wide increase of H3K4me3 marks, in line with the postulated role of KDM5 enzymes as regulators of this chromatin modification^{12,13}. Although the apparent lack of increasing H3K4me3 levels using GSK467 cannot be explained at present, it is however correlated to a weak anti-proliferative effect. The specific perturbation of the ethyl-ester pro-

drug KDM5-C70 leading to the observed cell cycle arrest in the multiple myeloma cell system is likely to be related to the observed dysregulation of cell cycle and metabolic genes. Although the cellular potency of KDM5-C70 is weak and about one to two orders of magnitude beyond a desired cellular potency characterising high-quality chemical probes³⁶, the correlated anti-proliferative phenotype and the favourable in vitro selectivity profile suggest that further optimisation of this scaffold could potentially lead to valuable tools to interrogate KDM5 phenotypes in human biology.

ACCESSION CODES: The macromolecular structures were deposited with ProteinDataBank under following accession numbers 5a3p (apo), 5fup (2'OG), 5a3w ((2,4 PDCA), 5a1f (NOG), 5a3t (KDM5-C49), 5fpu (GSK-J1), 5fun (GSK-467) (all KDM5B-c4); 5fv3 (KDM5B-c2, NOG), 5fwj (KDM5C, KDM5C-49), 5fpv (KDM4A, KDM5-C49) and 4uf0 (UTY, KDM5-C49). CHIPseq data were deposited with the GEO database with accession number [GSE78206](#).

ACKNOWLEDGEMENTS

Research in our laboratories is supported through funding from Arthritis Research UK (program grant number 20522), the NIHR Oxford Biomedical Research Unit, Sarcoma UK, the Bone Cancer Research Trust and the Rosetrees Trust, and Cancer Research UK (Grant number C8717/A18245). The Structural Genomics Consortium is a registered charity (number 1097737) that receives funds from AbbVie, Bayer Pharma AG, Boehringer Ingelheim, the Canada Foundation for Innovation, Genome Canada, GlaxoSmithKline, Janssen, Lilly Canada, Merck & Co., the Novartis Research Foundation, the Ontario Ministry of Economic Development and Innovation, Pfizer, São Paulo Research Foundation-FAPESP, Takeda, and the Wellcome Trust [092809/Z/10/Z]. We would like to thank Diamond Light Source for beamtime (proposal mx10619), and the staff of beamlines I02 and I03 for assistance with crystal testing and data collection. We are grateful to Dr. R. Rambo for help with SAXS data collection and analysis at the Diamond Light Source beamline B21.

AUTHOR CONTRIBUTIONS: CJ and UO designed, analysed data, supervised the study and wrote the manuscript. Data collection and structure refinements were done by SV, RN, JK, construct design was by CJ, HS, UE, VB, cloning, mutagenesis and expression trials were done by CSD, AS, MP, NBB, purification, crystallisation and optimisation by CJ, SV, AS and CG, enzymology by AT, cell culture experiments by NW, EH, SM, NA, NLT and

compounds were provided by RP, SW, AN, JB and PB. All authors approved the final version of the manuscript.

REFERENCES

- 1 Ng, S. S., Yue, W. W., Oppermann, U. & Klose, R. J. Dynamic protein methylation in chromatin biology. *Cell Mol Life Sci* **66**, 407-422 (2009).
- 2 Johansson, C. *et al.* The roles of Jumonji-type oxygenases in human disease. *Epigenomics* **6**, 89-120 (2014).
- 3 Kooistra, S. M. & Helin, K. Molecular mechanisms and potential functions of histone demethylases. *Nat Rev Mol Cell Biol* **13**, 297-311 (2012).
- 4 Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941-953 (2004).
- 5 Allis, C. D. *et al.* New nomenclature for chromatin-modifying enzymes. *Cell* **131**, 633-636 (2007).
- 6 Walport, L. J., Hopkinson, R. J. & Schofield, C. J. Mechanisms of human histone and nucleic acid demethylases. *Curr Opin Chem Biol* **16**, 525-534 (2012).
- 7 ENCODE Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
- 8 Pasini, D. *et al.* Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. *Genes Dev* **22**, 1345-1355 (2008).
- 9 Sanchez, R. & Zhou, M. M. The PHD finger: a versatile epigenome reader. *Trends Biochem Sci* **36**, 364-372 (2011).
- 10 Tahiliani, M. *et al.* The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. *Nature* **447**, 601-605 (2007).
- 11 Zhang, Y. *et al.* The PHD1 finger of KDM5B recognizes unmodified H3K4 during the demethylation of histone H3K4me_{2/3} by KDM5B. *Protein Cell* **5**, 837-850 (2014).
- 12 Torres, I. O. *et al.* Histone demethylase KDM5A is regulated by its reader domain through a positive-feedback mechanism. *Nat Commun* **6**, 6204 (2015).
- 13 Rasmussen, P. B. & Staller, P. The KDM5 family of histone demethylases as targets in oncology drug discovery. *Epigenomics* **6**, 277-286 (2014).
- 14 Pilka, E. S., James, T. & Lisztwan, J. H. Structural definitions of Jumonji family demethylase selectivity. *Drug Discov Today* **20**, 743-749 (2015).
- 15 Yamane, K. *et al.* PLU-1 is an H3K4 demethylase involved in transcriptional repression and breast cancer cell proliferation. *Mol Cell* **25**, 801-812 (2007).
- 16 Chang, K. H. *et al.* Inhibition of histone demethylases by 4-carboxy-2,2'-bipyridyl compounds. *ChemMedChem* **6**, 759-764 (2011).
- 17 Ng, S. S. *et al.* Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* **448**, 87-91 (2007).
- 18 Hutchinson, S. E. *et al.* Enabling lead discovery for histone lysine demethylases by high-throughput RapidFire mass spectrometry. *J Biomol Screen* **17**, 39-48 (2012).

- 19 Sakurai, M. *et al.* A miniaturized screen for inhibitors of Jumonji histone demethylases. *Mol Biosyst* **6**, 357-364 (2010).
- 20 Clifton, I. J. *et al.* Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J Inorg Biochem* **100**, 644-669 (2006).
- 21 Kruidenier, L. *et al.* A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. *Nature* **488**, 404-408 (2012).
- 22 Walport, L. J. *et al.* Human UTY(KDM6C) is a male-specific N-methyl lysyl demethylase. *J Biol Chem* **289**, 18302-18313 (2014).
- 23 Hillringhaus, L. *et al.* Structural and evolutionary basis for the dual substrate selectivity of human KDM4 histone demethylase family. *J Biol Chem* **286**, 41616-41625 (2011).
- 24 McDonough, M. A., Loenarz, C., Chowdhury, R., Clifton, I. J. & Schofield, C. J. Structural studies on human 2-oxoglutarate dependent oxygenases. *Curr Opin Struct Biol* **20**, 659-672 (2010).
- 25 King, O. N. *et al.* Quantitative high-throughput screening identifies 8-hydroxyquinolines as cell-active histone demethylase inhibitors. *PLoS One* **5**, e15535 (2010).
- 26 Rose, N. R. *et al.* Selective inhibitors of the JMJD2 histone demethylases: combined nondenaturing mass spectrometric screening and crystallographic approaches. *J Med Chem* **53**, 1810-1818 (2010).
- 27 Woon, E. C. *et al.* Linking of 2-oxoglutarate and substrate binding sites enables potent and highly selective inhibition of JmjC histone demethylases. *Angew Chem Int Ed Engl* **51**, 1631-1634 (2012).
- 28 Cheng, Y. & Prusoff, W. H. Relationship between the inhibition constant (K₁) and the concentration of inhibitor which causes 50 per cent inhibition (I₅₀) of an enzymatic reaction. *Biochem Pharmacol* **22**, 3099-3108 (1973).
- 29 Copeland, R. A. *Enzymes*. Wiley-VCH, Inc, 2nd edition, 266-304 (2000).
- 30 Westaway, S. M. *et al.* Cell Penetrant Inhibitors of the KDM4 and KDM5 Families of Histone Lysine Demethylases. 2. Pyrido[3,4-d]pyrimidin-4(3H)-one Derivatives. *J Med Chem* **59**, 1370-1387 (2016).
- 31 Heinemann, B. *et al.* Inhibition of demethylases by GSK-J1/J4. *Nature* **514**, E1-2 (2014).
- 32 Sayegh, J. *et al.* Identification of small molecule inhibitors of Jumonji AT-rich interactive domain 1B (JARID1B) histone demethylase by a sensitive high throughput screen. *J Biol Chem* **288**, 9408-9417 (2013).
- 33 Horton, J. R. *et al.* Characterization of a Linked Jumonji Domain of the KDM5/JARID1 Family of Histone H3 Lysine 4 Demethylases. *J Biol Chem* **291**, 2631-2646 (2016).
- 34 Horton, J. R. *et al.* Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nat Struct Mol Biol* **17**, 38-43 (2010).
- 35 Chowdhury, R. *et al.* Ribosomal oxygenases are structurally conserved from prokaryotes to humans. *Nature* **510**, 422-426 (2014).
- 36 Frye, S. V. The art of the chemical probe. *Nat Chem Biol* **6**, 159-161 (2010).

FIGURE LEGENDS

Figure 1: Domain organisation of full-length human KDM5B, construct design and substrate kinetics for KDM5B deletion constructs - (a) Schematic overview of KDM5 domain organisation and summary of deletion and mutagenesis constructs used in this study. The * indicates insertion of a four glycine linker **(b)** H3(1-21)K4me2 demethylase activity measured by a coupled formaldehyde dehydrogenase enzyme assay. Initial velocities plotted against various peptide concentrations and fitted with the Michaelis-Menten equation using GraphPad Prism 5.0. Error bars indicate standard error of the mean value from three independent experiments performed in duplicate technical replicates.

Figure 2: Structure determination of human KDM5B - (a) Ribbon representation of human KDM5B-c4 in complex with Zn, Mn and cofactor analog- N-oxalylglycine (NOG, PDB ID 5aif). Colour coding of individual domains is shown as in Figure 1A. Red arrow indicates the PHD1-ARID deletion site. **(b)** Details and plasticity of cofactor binding site residues with overlaid structures of NOG (yellow) and 2,4 PDCA (purple). Side-chain residues positions are shown in blue (NOG) and silver (2,4 PDCA). Mn (II) used to replace the active site iron is shown as a turquoise sphere and water molecules are shown as red spheres. Hydrogen bonds and electrostatic interactions are depicted as black dashed lines. **(c)** Zn coordination in KDM5B, with two Zn (II) ions (shown as orange spheres).

Figure 3: The KDM5B structures reveal hybrid features with the KDM4 and KDM6 families: (a-c) Electrostatic surface potential (left) and domain packing (right) of KDM5B (PDB ID 5A1F) **(a)**, KDM6A (PDB ID 3AVR) **(b)**, and KDM4A (PDB ID 2P5B) **(c)**. KDM5B superimposed with KDM6A **(d)** and KDM4A **(e)** showing differences within the cofactor binding pocket. Side-chain residues in KDM5B are shown in blue sticks, in KDM6A in green and in KDM4A in salmon colouring. NOG and the H3K27me3 **(d)** and H3K36me3 peptides **(e)** are presented as yellow sticks. **(f)** Low resolution solution-structure derived from SAXS experiments using constructs KDM5B-c1 and KDM5B-c4. The experimental structures of KDM5B-c4 and the ARID (PDB ID 2EQY) and PHD1 (PDB ID 2MNY) domains were fitted into the envelope obtained by the SAXS data for KDM5B-c1.

Figure 4: The active site of human KDM5B is amenable to selective inhibitor development and can accommodate distinct inhibitor chemotypes. (a-c) KDM5 inhibitor

binding sites with KDM5-C49 (**a**, yellow), GSK-J1 (**b**, magenta) and GSK467 (**c**, orange). For display of electron density and stereoviews for ligands see **Supplementary Fig.11**). Side-chains are displayed in cyan, water molecules as red spheres, and the metal centre as green sphere.

Figure 5: KDM5-C70 is active within MM1 myeloma cells - Dose response curves showing viability after 7 days treatment with either KDM5-C70 (**a**) or GSK467 (**b**) compared to DMSO control. Data represents the mean and standard error of three independent experiments with 4 technical replicates per independent experiment. (**c**) Phosphorylation of retinoblastoma protein after 7 days treatment with either 50 μ M GSK467 or 50 μ M KDM5-C70 (n=3 biological replicates for all conditions). The original western blots are displayed in **Supplementary Fig. 15**.

Figure 6: KDM5-C70 increases H3K4me3 levels within myeloma cells (**a**) Normalised coverage and (**b**) peak-level coverage of H3K4me3 marks for all peaks identified (n=3 independent experiments for DMSO and KDM5-C70 and n=2 for GSK467). To account for different peak width the central portion of each plot is scaled to show the coverage at each position as a proportion of the width of the peak. (**c**) Number of peaks categorised by their genomic location. All peaks detected are shown in black, and peaks showing significant increased binding (blue) and decreased binding (red) with KDM5-C70. A list of peaks changed by inhibitor treatment is available as **Supplementary Dataset 1**. (**d-f**) Representative tracks of three genes associated with myeloma biology.

ONLINE METHODS

Protein constructs - A KDM5B-c1 construct was amplified from an Origene cDNA clone, and cloned into a pFastBac-derived vector (pFB-LIC-Bse), which contains a tobacco etch virus (TEV) protease cleavable N-terminal 6x-histidine tag. The same cDNA clone was also used to generate ARID and PHD1 domain deletion mutants with and without insertion of a four-residue glycine linker (KDM5B-c2, KDM5B-c3 and KDM5B-c4, see Figure 1 and Supplementary Results, Supplementary Table 1 for details) by overlap extension PCR; these were then cloned into the pFB-LIC-Bse vector. KDM5B-c4 R98G and R98K mutants were generated by megaprimer PCR using KDM5B-c4 as template.

Human KDM5A (residue Met¹-Leu⁸⁰¹) was amplified from a construct kindly received from Rob Klose (Department of Chemistry, Oxford University, UK) and cloned into pFB-LIC-Bse, KDM5D (residue Met¹-Asp⁷⁷⁵) was amplified from a cDNA clone ordered from Open Biosys (IMAGE clone 40146743) and cloned into pFB-LIC-Bse, whereas KDM5C (residue Met¹-Val⁷⁶⁵) was amplified from an MGC cDNA clone (IMAGE clone 5492114) and cloned into a pFastBac-derived vector (pFB-CTHF-LIC) which contains a tobacco etch virus (TEV) protease cleavable C-terminal 10x-histidine tag (KDM5C). The KDM5C deletion constructs were generated as described for KDM5B-c4 (residue Phe⁸-Thr⁷⁷², Leu⁸³-Gly³⁸⁴ deleted with insertion of a 4x- glycine linker). All constructs were confirmed by Sanger DNA sequencing.

Protein expression and purification - Recombinant KDM5B constructs were expressed in Sf9 cells and generation of recombinant baculo viruses, insect cell culture, and infections were performed according to the manufacturer's instructions (Invitrogen). The cells were collected 72 hours post infection and suspended in a buffer containing 50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM imidazole, 5% glycerol, 0.5 mM TCEP, and a protease inhibitor mix (Calbiochem). All KDM5B variants were purified using nickel affinity chromatography using a stepwise gradient of imidazole. The eluted protein was then incubated with TEV protease at 4 °C overnight followed by size-exclusion chromatography (Superdex 200). The TEV protease and uncleaved protein were removed using nickel affinity chromatography and the mass was verified by electrospray ionization time-of-flight mass spectrometry (ESI-TOF, Agilent LC/MSD). KDM5 proteins for biochemical assays were purified as above but without cleavage of the histidine tag. UTY/KDM6C and JMJD2/KDM4 proteins were expressed and purified as described previously^{17,22} whereas KDM2, 3, and 6B are described in detail elsewhere³⁷.

Activity assays- reagents and conditions: Assay specifics including final enzyme concentrations, co-factor final concentrations, assay incubation times and antibody final concentrations are detailed in Supplementary Results, Supplementary Table 2A-C. All enzymes used for IC₅₀ determinations contained the catalytic JmjC domain (**Supplementary Table 2c**). All IC₅₀ determinations were performed using 2-OG concentrations at or near the K_m value for the respective enzyme and incubation times with substrate were determined from the linear range of enzyme progress curves for respective enzymes (**Supplementary Fig. 6**). All reagents were from Sigma Aldrich unless otherwise stated and were of the highest purity. Bovine serum albumin (BSA) used in the AlphaScreen assays was fatty acid and globulin free (Sigma A7030). Hydroxy ethyl piperazine ethane sulfonic acid (HEPES) buffer was from Life Technologies. Ferrous Ammonium Sulphate (FAS) was dissolved in 20 mM HCl to a concentration of 400 mM and diluted to 1 mM in deionized water. L-Ascorbic Acid (50 mM stock concentration in deionized water) and 2-oxoglutarate (10 mM stock concentration in deionized water) were prepared fresh each day. Biotinylated peptides were purchased from Anaspec (**Supplementary Table 2a**) and antibodies to product methyl marks were purchased from Abcam, Cell Signalling Technology or Millipore (Supplementary Table 2B).

Demethylase AlphaScreen - The demethylase AlphaScreen assay was performed in 384-well plate format using white proxiplates (Perkin Elmer) and transfer of compound (100 nl) was performed using an ECHO 550 Acoustic Dispenser (Labcyte). After establishment of suitable purification conditions (see above), enzyme samples showed normal distribution of their activities. All subsequent steps were carried out in assay buffer (50 mM HEPES pH 7.5, 0.1% (w/v) bovine serum albumin and 0.01 % (v/v) Tween-20). In brief 5 µl of assay buffer containing demethylase enzyme at 2X final assay concentration (see **Supplementary Table 3c** for assay specifics) was pre-incubated for 15 minutes with dilutions of compound. The enzyme reaction was initiated by addition of substrate (5 µl) consisting of L-ascorbic Acid (200 µM), 2-oxoglutarate at 2X final assay concentration (KDM5A-D (10 µM), KDM4C (20 µM), KDM3A (10 µM), KDM6B (20 µM), KDM2A (20 µM)), ferrous ammonium sulphate (2X final assay concentration) and histone H3 substrate peptide (2X final assay concentration). The enzyme reaction was allowed to proceed for the required time (**Supplementary Table 2c**) and was stopped by addition of 5 µl assay buffer containing EDTA (30 mM) and NaCl (800 mM). The final concentration of DMSO was 1%. Streptavidin Donor beads (0.08 mg/ml) and Protein-A conjugated acceptor beads (0.08

mg/ml) were pre-incubated for 1 hour with anti-methyl mark antibody (4X final assay concentration) and the presence of histone H3 product methyl mark was detected by addition of the pre-incubated AlphaScreen beads (5 μ l). Detection was allowed to proceed for 2 hour at room temperature and the assay plates were read in a BMG Pherastar FS plate reader (Excitation 680 nM / Emission 570 nM). Data were normalized to the (no enzyme) control and the IC50 values were determined from the nonlinear regression curve fit using GraphPad Prism 5. Using the kinetic parameters for substrate and cofactor, apparent Ki values were extrapolated from IC50 values using the described relationships between IC50 and Ki values assuming competitive inhibition^{28,29}.

Formaldehyde Dehydrogenase (FDH) Coupled Enzyme Assay - Peptide Km values were determined using the FDH coupled enzyme assay. The formaldehyde dehydrogenase expression clone was obtained from Rob Klose (University of Oxford) as a 1320 bp fragment in the vector pNIC-28 Bsa4 and transformed into the Rosetta strain of *E.coli*. Histone H3 peptides (amino acid 1-21) methylated at lysine 4 were synthesized by Peptide Protein Research Ltd. All other reagents were from Sigma Aldrich.

The coupled enzyme assay was performed in black 384-well Non-Binding Surface (NBS) microplates (Corning, Code 3650) and all steps were carried out in assay buffer (50 mM HEPES pH 7.5, 0.1% BSA, 0.01% v/v Tween-20). For peptide Km determinations the concentration of 2-OG was kept at 10X the Km concentration. Assay buffer (25 μ l) containing substrate at 2X final concentration (200 μ M L-Ascorbic Acid, 100 μ M FAS, 100 μ M 2-OG, 500 μ M NAD, 0.1 – 20 μ M H3-K4-(me2) peptide) was transferred to wells of a 384-well plate in triplicate for each peptide concentration. The enzyme reaction was initiated by plate reader injection of 25 μ l of assay buffer containing enzymes (600 nM KDM5B, 1.0 μ M FDH) and production of NADH was measured in a fluorescent plate reader (BMG Labtech Pherastar FS, excitation 355 nM, emission 450 nM). Readings were taken every 60 sec over a time course of 30 minutes. The initial slope for enzyme progress curves was taken for each peptide concentration and data fitted to the Michaelis-Menten equation in GraphPad Prism 5.

KDM5B RapidFire Mass Spectrometry (RF-MS) assay - Activity of KDM5B variants were assessed by RF-MS. The KDM5B H3-K4 dimethyl peptide substrate ARTK(me2)QTARKSTGGKAPRKQLA was synthesized by Peptide Protein Research Ltd. All steps were performed in assay buffer (50 mM MES pH7.0, 50 mM NaCl, 1 mM TCEP).

0.5 ml of assay buffer containing 1.25x KDM5B enzymes (188 nM) was transferred into wells of a 96-deep well polypropylene block and the enzyme reaction was initiated by addition of 100 μ l of 6X substrate in assay buffer (60 μ M FAS, 600 μ M L-Ascorbic Acid, 60 μ M μ KG, 60 μ M H3-K4 dimethyl peptide). The deep well block was transferred to a RapidFire RF360 high throughput sampling robot connected to a 6530 Accurate-Mass Quadrupole Time-of-Flight (Q-TOF) mass spectrometer (Agilent) operated in positive ion mode. Samples were aspirated under vacuum for 400 ms and applied to a C4 solid phase extraction (SPE) cartridge. The SPE was washed to remove non-volatile buffer salts with water containing 0.1% (v/v) formic acid applied at a flow rate of 1.5 ml / min for 6.0 sec and peptides were eluted onto the Q-TOF with 85% acetonitrile, 25% water containing 0.1% formic acid at a flow rate of 1.25 ml / min for 6.0 sec. The cartridge was re-equilibrated with water for 500 ms. A cycle of aspiration, aqueous wash, organic elution and re-equilibration takes approximately 15 sec, for establishment of enzyme progress curves a sample was aspirated every 3.5 minutes. Ion chromatogram data was extracted for the +6 charge state for the dimethyl substrate and the mono-methyl product and peak area data for extracted ion chromatograms were integrated using RapidFire Integrator software (Agilent). The fractional conversion of dimethyl substrate to monomethyl product was calculated using the equation:

$$\% \text{ Conversion} = 100 \times \text{Monomethyl} / (\text{Monomethyl} + \text{Dimethyl}).$$

Active site titration - Active site titration of KDM5B-c1 was performed in triplicate at KDM5-C49 concentrations of 9.986, 3.495, 0.998, 0.7, 0.5, 0.3, 0.1, 0.05, 0.025, 0.0083 and 0.0028 μ M, adding in a constant volume of 70 nl of DMSO and 70 nl of DMSO control. The KDM5B-c1 construct was diluted to 880 nM in FDH buffer containing 1 μ M of FDH enzyme. Following a 15min preincubation of the compound within the 25 μ l of enzyme solution the reaction was started by injection of 25 μ l of peptide solution (200 μ M LAA, 20 μ M FAS, 10 μ M 2OG, 500 μ M NAD⁺ and 10 μ M H3K4me2 peptide in FDH assay buffer) using the Pherastar injection system on a Labcyte robotic system. The reading for each of the 30 min progress curves was taken every 60 seconds and the velocity was calculated as the slope of the progress curve in the first 4 min after the start of the reaction. The calculated velocities were scaled between 0 and 1 based on the DMSO control (V_{DMSO}) and fully inhibited enzyme (V_{min} at 9.986 μ M compound concentration):

$$\frac{V}{V_0} = \frac{V - V_{min}}{V_{DMSO} - V_{min}}.$$

The resulting data was used to fit the Morrison equation ²⁹ in GraphPad Prism.

Crystallisation and data collection – Protein preparations were concentrated in concentrators (Amicon) to about 8 mg/ml, and were subjected to crystallisation experiments at 4 °C using the sitting drop vapour diffusion method. The proteins were pre-incubated with 1 mM of the specific inhibitors and 2-4 mM MnCl₂ before the protein-compound mixture was transferred to crystallisation plates. KDM5B-c4 was crystallised with the inhibitor KDM5-C49 (2-(((2-((2-(dimethylamino)ethyl)(ethyl)amino)-2-oxoethyl)amino)methyl) isonicotinic acid, in a drop consisting of 50 nL protein-compound mix (8.1 mg/ml) and 100 nL of a precipitant consisting of 60 % MPD, 0.1 M Hepes pH 7.5. The apo-form of KDM5B and soaked co-crystals of KDM5B in complex with either 2,4 PDCA, NOG, and GSK467 were obtained in drops consisting of 100 nL protein-compound mix (8.1 mg/ml), 200 nL of a precipitant consisting of 1.6 M Na/K phosphate, 0.1 M Hepes pH 7.5 and 20 nl of KDM5B seeds of crystals obtained from the same condition. GSK-J1(3-{[2-(pyridin-2-yl)-6-(2,3,4,5-tetrahydro-1H-3-benzazepin-3-yl)pyrimidin-4yl]amino} propan- oic acid) was soaked into apo crystals of KDM5B for ten minutes at a concentration of 1 mM. Crystals of KDM5C (encompassing residues Phe8-Thr772 with internal deletion of residues Leu⁸³-Gly³⁸⁴ and insertion of a 4x-glycine linker) in complex with KDM5-C49 was obtained by incubating the protein at 7.9 mg/ml with 4mM MnCl₂ and 500 μM KDM5-C49. Crystals were obtained at 4° C in a drop consisting of 100 nl protein-compound mix and 50 nl precipitant containing 28% PEG3350, 0.25M MgCl₂, 0.1M Bis-Tris pH 5.5. Crystals of UTY in complex with the KDM5-C49 inhibitor were obtained at 4° C in a drop consisting of 140 nL protein-compound (11 mg/ml) and 70 nL of a precipitant consisting of 25% (w/v) PEG 3350, and 0.1 M bis-tris pH 6.5. Crystals of KDM4A in complex with KDM5-C49 were obtained by incubating the protein at 3mg/ml with 1mM MnCl₂ and 200 μM KDM5-C49 for 45 minutes on ice. The protein-compound mixture was then concentrated to 26 mg/ml and crystals were obtained at 4° C in a drop consisting of 105 nl protein-compound mix and 105 nl precipitant containing 0.1M Bis-Tris pH 6.5, 0.25 M ammonium sulphate and 30% PEG3350.

All crystals were cryo-protected with mother liquor supplemented with 25% ethylene glycol before they were flash frozen in liquid nitrogen. Data sets were collected on beamlines, I02, I03 or I04-1 at the Diamond Light source UK. Metal contents of crystals were investigated from a KDM5B crystal using X-ray fluorescence scanning on beamline I02 or I03 at Diamond Light Source, UK, on a Vortex-EX fluorescence detector (Hitachi High-Technologies Science). A peak for Zn (observed peak at 9667.14 eV, expected peak at 9658.6

eV) and a peak for Mn (observed peak at 6548.69 eV, expected peak at 6539.0eV) were observed as expected based on the structural work.

Structure determination of human KDM5B - Datasets were collected and structures were solved with statistics shown in Supplementary Tables 3 and 5. Data were processed, scaled and merged with either: (i) Mosflm³⁸ and SCALA or (iii) XDS^{39,40} or xia2⁴¹ and AIMLESS. The resolution cut-off for the final dataset was based on CC1/2. A monomer of KDM4A (PDB ID 2bp5) was used to solve the structures of the KDM5B·KDM5-C49 and KDM4A·KDM5-C49 complexes, and UTY was solved using PDB ID 3ZLI as a search model. The rest of the KDM5B complex structures were solved by molecular replacement⁴² using the structure of KDM5B·KDM5-C49 (PDB ID 5A3T). All structures were refined in an iterative process combining REFMAC5 and PHENIX⁴³ with electron-density and difference-density map inspections and model manipulations in Coot⁴⁴. Water molecules and the ligands were added during the refinement process. Where appropriate, dual rotamer side-chain conformations were also included. Refinements were terminated when there were no significant changes in the R_{work} and R_{free} values and inspection of the difference density maps suggested that no further corrections or additions were justified.

Small-Angle X-ray Scattering (SAXS) – Synchrotron radiation X-ray scattering data were collected at the B21 beamline at the Diamond Light Source (Didcot, UK). The protein samples (KDM5B-c1, 7.1mg/ml and KDM5B-c4, 15.2mg/ml) were subjected to size-exclusion on a Shodex KW404 column and the data were collected in in-line mode, where the UV signal on the HPLC detector triggered small-angle x-ray scattering (SAXS) data collection. The protein sample was flowing through an in-vacuum quartz capillary of 1.6 mm diameter at 0.16 mL/min. Data were collected using a Pilatus2M detector (Dectris) at sample-detector distance of 3914 mm and a wavelength of $\lambda=1 \text{ \AA}$ and 1 sec exposure. The range of momentum transfer $0.1 < s < 5 \text{ nm}^{-1}$ was covered ($s = 4\pi\sin\theta/\lambda$, where θ is the scattering angle). The data were radially averaged and the scattering of the buffer was subtracted. The forward scattering $I(0)$, radius of gyration R_g , pair distribution of the particle $p(r)$, and maximum dimension D_{max} were calculated using ScAtter [Unpublished (Diamond Light Source, Didcot, UK)]. *Ab initio* models were computed using the DAMMIF pipeline embedded in ScAtter. We used the structures of the KDM5B-c4 catalytic domain (PDB 5A1F), the murine KDM5 ARID domain (PDB 2EQY) and the PHD1 structure (PDB

2MNY¹¹ (both determined by NMR) to generate a model of KDM5B-c1 that fits the SAXS data.

Small-Molecule KDM5 Inhibitors - KDM5-C49 and KDM-C70 were identified from a patent application (WO 2014053491 A1) and were purchased from Xcessbio or synthesized as described in Supplementary Notes. A sample of GSK467 was kindly provided by GSK and details of its synthesis and profiling are described elsewhere³⁰.

Assessment of cellular activity for GSK467 and KDM5-C70 - Anti-proliferative effects of compounds were tested in several multiple myeloma cell lines, among these MM1S, a B-lymphoblastic cell line from a multiple myeloma patient (obtained from ATCC, tested by PCR and found negative for mycoplasma infection) was found to be sensitive. Cells were cultured in RPMI (Sigma) supplemented with 10%FCS and 2 mmol/L glutamine at 37°C in 5% CO₂ at 37C. Cell viability assays were conducted with cells cultured in a 96 well plate at a density of 3000 cells per well. Viability was assessed using PrestoBlue reagent (Thermo Fisher) according to the manufacturer's instructions using a FluroStar Optima (BMG) plate reader. Data was normalised to a vehicle control and plotted using Prism version 5.03 (GraphPad).

Western Blot analysis of Rb phosphorylation - MM1S cell pellets were resuspended and lysed in a buffer containing 50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1% Igepal CA-630/NP-40, 1 mM EDTA, 1 mM NaF, 1 mM Na₃VO₃, protease inhibitor cocktail and 1 mM AEBSF. Following 30 min incubation on ice, the cell suspension was centrifuged at 13,000 rpm for 10 minutes to remove cell debris. Protein concentrations of the extracts were determined using Bradford protein assay (BioRad). 10µg of each lysate was subjected to SDS-PAGE followed by immunoblotting with the following antibodies (for dilutions see Supplementary Table 3): anti-pRb monoclonal (4H1), anti-phospho-pRb (S807/S811) polyclonal, anti-phospho-pRb (S608) polyclonal, anti-phospho-pRb (S780) polyclonal, anti-phospho-pRb (S795), anti-phospho-pRb (T356) polyclonal antibodies were from Cell Signalling. After primary antibody incubation, the membrane was washed three times and incubated with the appropriate HRP conjugated secondary antibody (Cell Signalling) for 1h. Excess secondary antibody was then rinsed off by washing the membrane three times. The membrane was incubated with SuperSignal West Dura Extended Duration substrate (ThermoScientific) for 5 minutes and luminescence from the membrane was detected using Fuji Medical X-ray film (Fujifilm).

Chromatin immunoprecipitation (ChIPSeq) followed by sequencing – CHIPseq was carried out to assess the genome-wide effect of GSK467 and KDM5-C70 treatment on H3K4me3 levels. A total of 6 samples treated with DMSO, 3 with KDM5-C70 and 2 with GSK467 were used. For immunoprecipitation, 10^7 cells per condition were fixed in 1% formaldehyde and sheared by sonication using a Bioruptor Pico (Diagenode). The lysate was pre-cleared with sepharose beads pre-blocked with bovine serum albumin and incubated overnight with 1 μ g of anti-H3K4me3 antibody (Millipore Cat 07-473 Lot 2207275). Sequencing libraries were prepared using the NebNEXT Ultra DNA library prep kit for Illumina and sequenced on an Illumina NextSeq 500 platform. Reads were aligned to GRCh37 using bowtie (v 1.1.1)⁴⁵, sorted and deduplicated with Picard Tools. Peaks for each sample separately were called using macs2⁴⁶ with the “broad” option and the gapped peaks file used for downstream analysis. Differential binding was assessed in R using DiffBind⁴⁷ and the EdgeR⁴⁸ with a false discovery rate of 0.1. An MA plot was generated in using the DiffBind package with the function dba.plotMA⁴⁹ plotting the log-fold change in KDM5-C70 treated versus control samples against the mean normalized RPKM between groups. Average profiles were generated using ngs.plto.r on a concatenated input of each biological replicate. Peak annotation and GO analysis was performed using Homer (Version 4.8).

Statistical analysis of data - Statistical analysis of the structural and the sequencing data was automatically performed in all of the programs used as noted above. Other statistical analyses were performed using the GraphPad Prism 5 software.

REFERENCES FOR ONLINE METHODS

- 37 Rose, N. R. *et al.* Plant growth regulator daminozide is a selective inhibitor of human KDM2/7 histone demethylases. *J Med Chem* **55**, 6639-6643 (2012).
- 38 Leslie, A. G. The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* **62**, 48-57 (2006).
- 39 Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* **66**, 133-144 (2010).
- 40 Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132 (2010).
- 41 Winter, G., Lobley, C. M. & Prince, S. M. Decision making in xia2. *Acta Crystallogr D Biol Crystallogr* **69**, 1260-1273 (2013).
- 42 McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).

- 43 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010).
- 44 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
- 45 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
- 46 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
- 47 Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481** (2012).
- 48 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
- 49 Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15**, 284 (2014).

COMPETING FINANCIAL INTERESTS - HS, UE, VB, SW, JB and RP are pharmaceutical industry employees and shareholders.

Figure 1

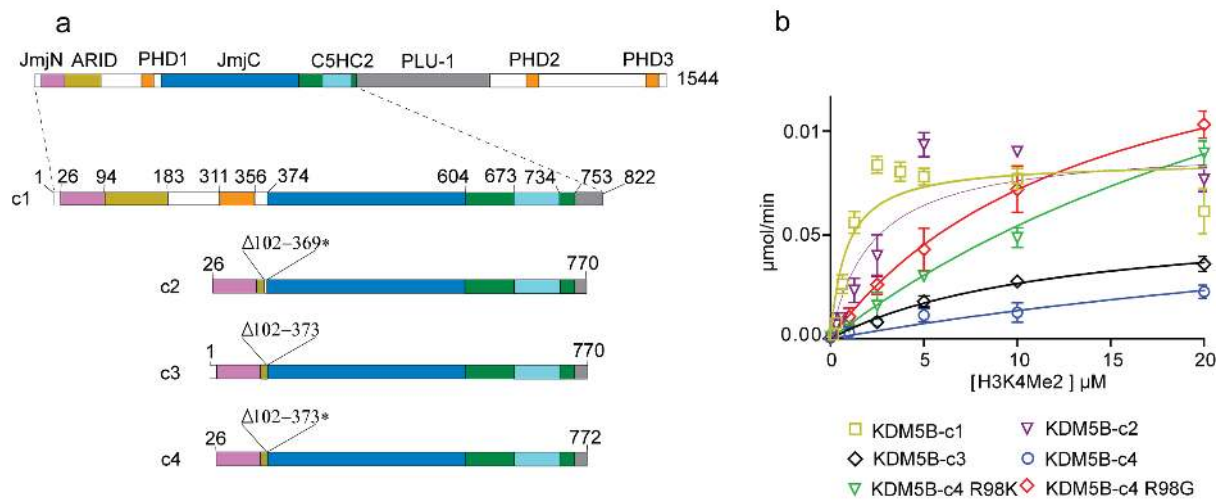


Figure 2

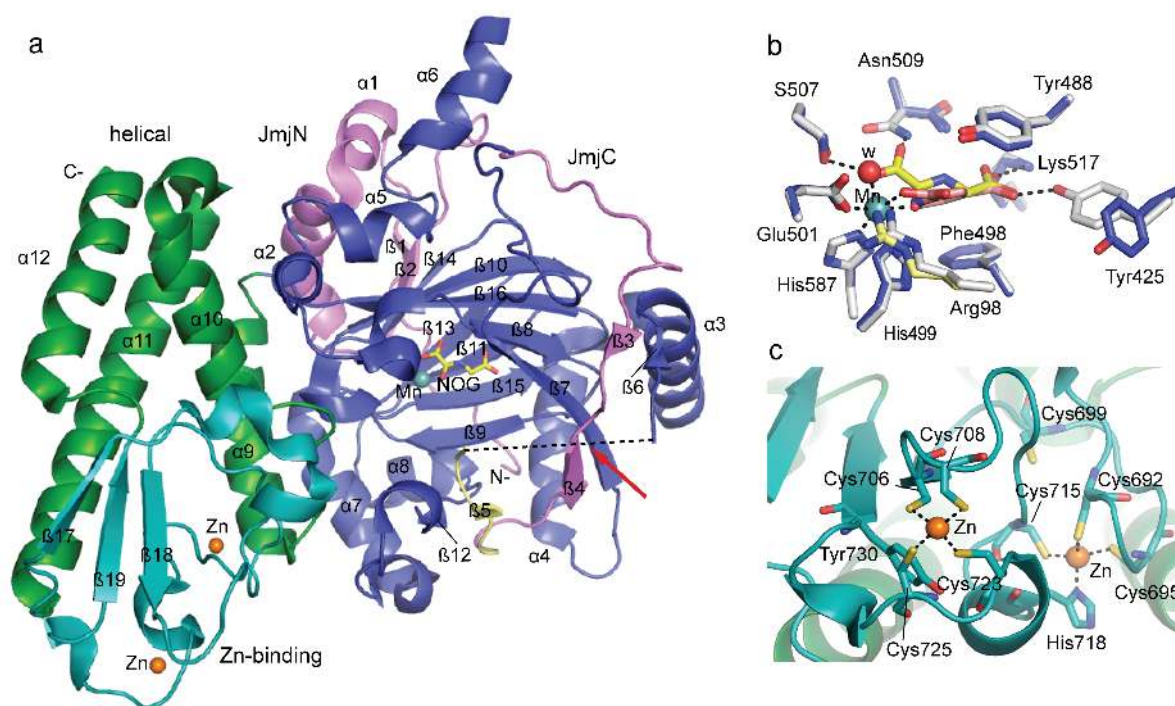


Figure 3

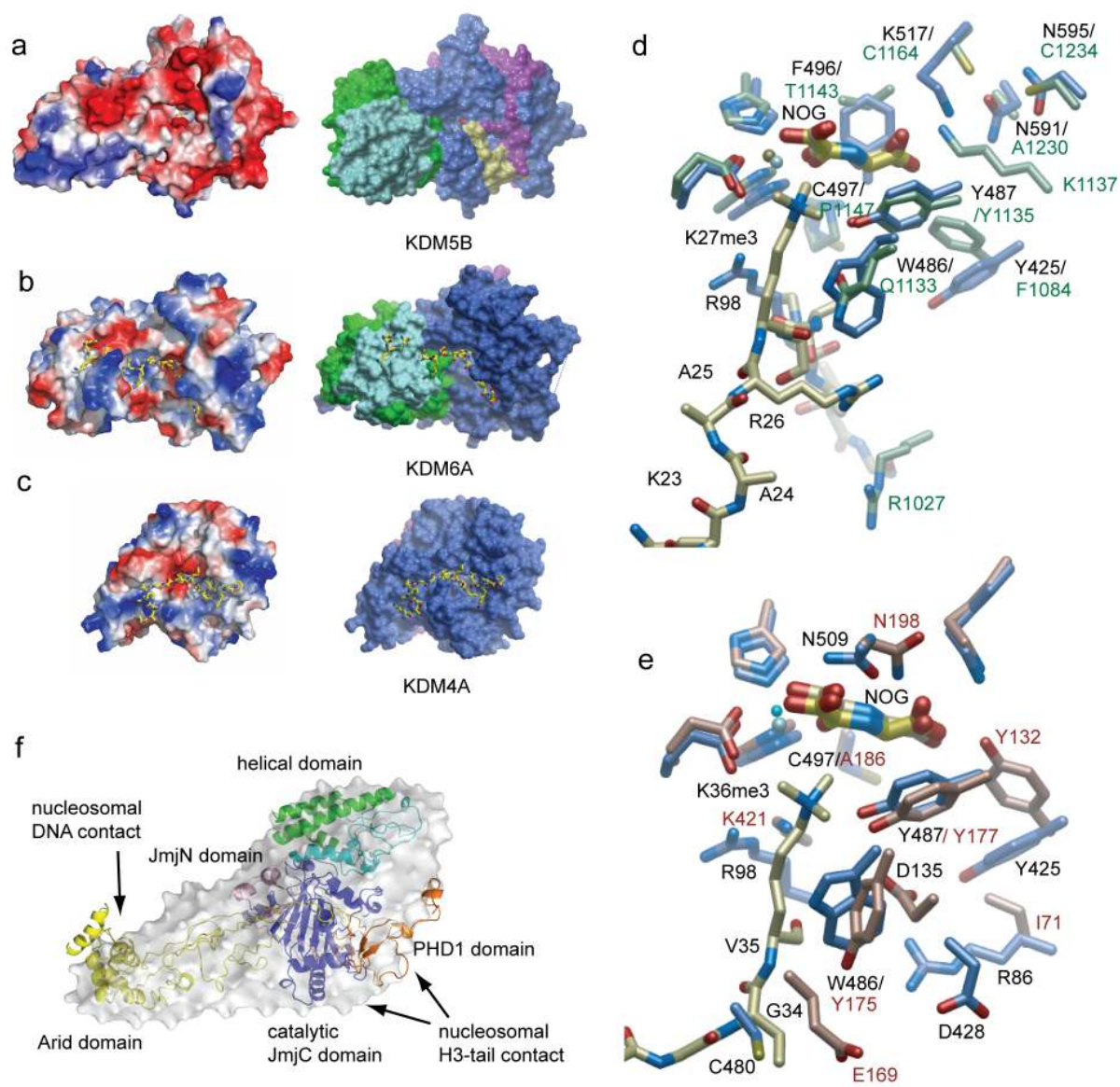


Figure 4

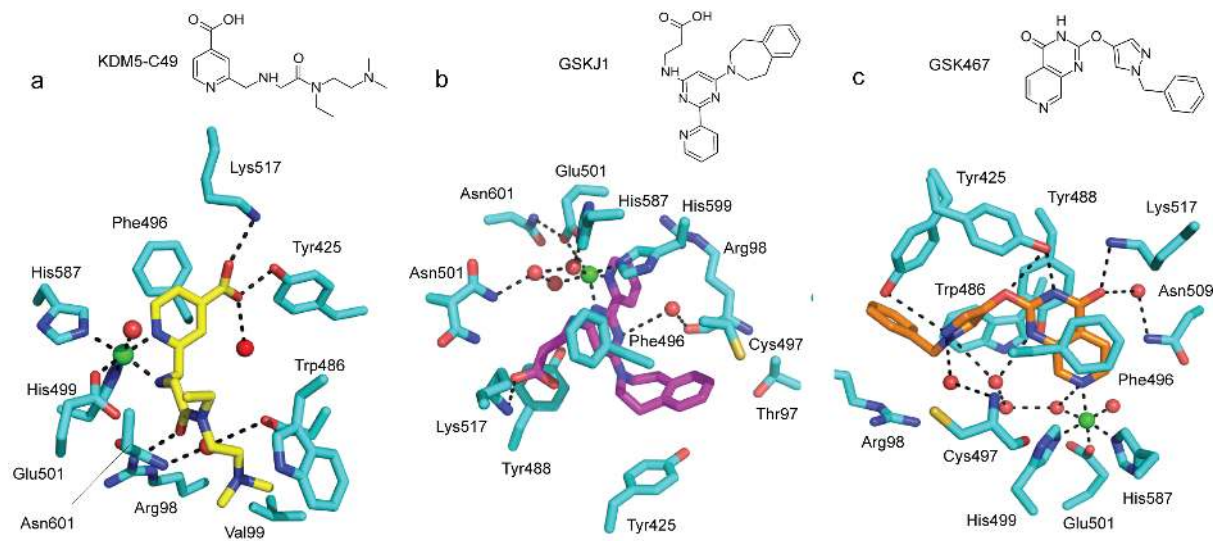


Figure 5

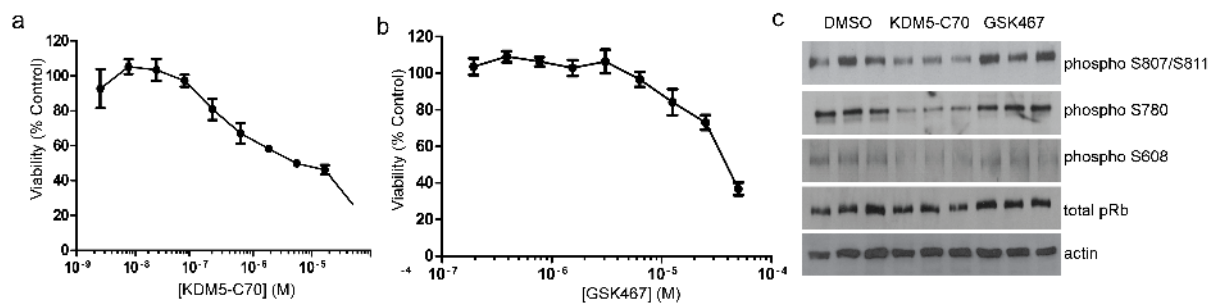


Figure 6

