Journal of Computer-Aided Molecular Design
https://doi.org/10.1007/s10822-018-0096-z

Disruptor of telomeric silencing 1-like (DOT1L): disclosing a new class of non-nucleoside inhibitors by means of ligand-based and structure-based approaches

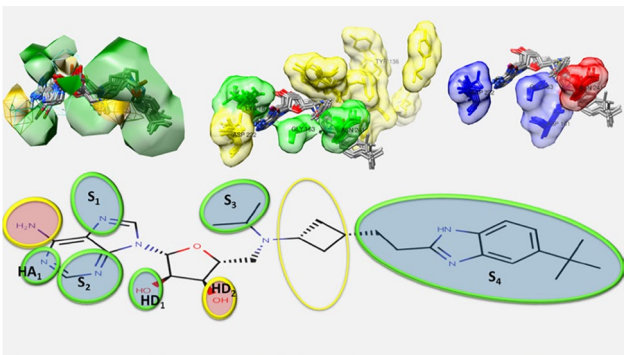
Manuela Sabatino[1,2] · Dante Rotili[2] · Alexandros Patsilinakos[1,2,3] · Mariantonietta Forgione[2] · Daniela Tomaselli[2] · Fréderic Alby[4] · Paola B. Arimondo[5,6] · Antonello Mai[2] · Rino Ragno[1,2,3]

Received: 23 August 2017 / Accepted: 6 January 2018

Abstract

Chemical inhibition of chromatin-mediated signaling involved proteins is an established strategy to drive expression networks and alter disease progression. Protein methyltransferases are among the most studied proteins in epigenetics and, in particular, disruptor of telomeric silencing 1-like (DOT1L) lysine methyltransferase plays a key role in MLL-rearranged acute leukemia Selective inhibition of DOT1L is an established attractive strategy to breakdown aberrant H3K79 methylation and thus overexpression of leukemia genes, and leukemogenesis. Although numerous DOT1L inhibitors have been several structural data published no pronounced computational efforts have been yet reported. In these studies a first tentative of multi-stage and LB/SB combined approach is reported in order to maximize the use of available data. Using co-crystallized ligand/DOT1L complexes, predictive 3-D QSAR and COMBINE models were built through a python implementation of previously reported methodologies. The models, validated by either modeled or experimental external test sets, proved to have good predictive abilities. The application of these models to an internal library led to the selection of two unreported compounds that were found able to inhibit DOT1L at micromolar level. To the best of our knowledge this is the first report of quantitative LB and SB DOT1L inhibitors models and their application to disclose new potential epigenetic modulators.

Graphical Abstract



Keywords Disruptor of telomeric silencing 1-like (DOT1L) · 3-D QSAR · COMBINE · Ligand-Based Drug Design · Structure-Based Drug Design

Abbreviations

DOT1L	Disruptor of telomeric silencing 1-like
SAM	*S*-adenosyl methionine
SAH	*S*-adenosyl homocysteine
3-D QSAR	Three-dimentional quantitative structure–activity relationship

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10822-018-0096-z) contains supplementary material, which is available to authorized users.

Extended author information available on the last page of the article

Published online: 15 January 2018

Py-3-D_QSAR	Python version of 3-D QSAutogrid/R procedure
COMBINE	COMpatative BINding Energy analysis
COMBINEr	Revisited COMBINE
Py-COMBINEr	Python version of COMBINEr
PRMT	Protein arginine methyl transferase
MLL1	Myeloid/lymphoid or mixed-lineage leukemia 1
LB	Ligand-based
LBDD	Ligande-based drug design
SB	Structure-based
SBDD	Structure-based drug design
GRID	Systematic grid spacing variation
VPO	Variable pre-treatment optimization
CV	Cross validation
LOO	Leave one out
L5O	Leave some out with 5 groups
LHO	Leave half out, leave some out with 2 groups
ELE	Per residues electrostatic interaction energies calculated by Autogrid
STE	Per residues steric interaction energies calculated by Autogrid
DRY	Per residues hydrophobic interaction energies calculated by Autogrid
HB	Per residues hydrogen bonding interaction energies calculated by Autogrid
EC	Experimental conformation
RC	Randomized conformation
RD	Re-docking
CD	Cross-docking
DA	Docking accuracy
ECRD	Experimental conformation re-docking
RCRD	Random conformation re-docking
ECCD	Experimental conformation cross-docking
RCCD	Random conformation cross-docking
RMSD	Root mean square deviation
MTS	Modeled test set
CTF	Crystal test set
MPS	Modeled prediction set
MIF	Molecular interaction field
PLS	Partial least square or projection of latent structures
PC	Principal compontent
SDEC	Standard deviation error of calculation
SDEP	Standard deviation error of prediction
r^2	Conventional squared correlation coefficient
q^2	Cross-validated correlation coefficient
COEFs	PLS coefficients
HP	Histogram plot
AC	Activity contribution
AAC	Average activity contribution
MRAC	Molecule–residue activity contribution
MRAAC	Molecule–residue average activity contribution
MRIs	Molecule–residues interactions
AMRIs	Average molecule–residues interactions

Introduction

The switch between heterochromatin and euchromatin is epigenetically controlled by covalent modifications that occur mainly at histone tails lysine and arginine residues. Recently, enzymes that catalyze histone tails' methylation have attracted scientists' attention for their implications in cancer. Histone methylation and related involved enzymes are among the most studied epigenetic mechanisms [1]. These enzymes belong to three distinct protein families: protein arginine methyltransferases (PRMTs), histone SET-domain-containing lysine methyltransferases (SUZ39, SET1, SET2, EZH1/2, RIZ, SMYD, SUZ4-20 and a few orphan members such as SET7/9 and SET8), and the non-SET-domain lysine methyltransferases (DOT1/DOT1L) [2–4]. DOT1L (disruptor of telomeric silencing 1-like), recently renamed KMT4 [5], is a 1537 amino acids enzyme with five α helices and two pairs of short β strand hairpins (Fig. 1). DOT1L transfers from one to three methyl groups on histone H3 lysine 79 (H3K79) through a S_N2 type reaction [1] by means of *S*-(5′-adenosyl)-L-methionine (SAM) as co-substrate, which binds in a pocket defined mainly by Thr139, Asp161, Gln168, Glu186 and Asp222 residues (Fig. 1).

H3K79 methylation has been reported as a biomarker for active gene transcription, and DOT1L was found to play an essential part in transcriptional elongation, DNA repair and cell cycle regulation [6]. DOT1L was also reported to play a functional role in normal cellular and biological processes such as cardiac development and function [7–10], erythropoiesis [11, 12] and leukemia development [13]. Misregulation of DOT1L is observed in the development of MLL-rearranged mixed lineage leukaemia 1. This cancer derives from chromosomal translocations of MLL1 gene. MLL1 is a SET containing histone H3 lysine 4 (H3K4) methyltransferase [14–16] involved in the "on state" of various genes. Indeed, H3K4 methylation represents an active transcription marker for gene expression. When chromosomal 11q23 translocation occurs, there is the *KMT2A* fusion with a subset of partners protein, including AF4, AF9 and AF10.

The resulting oncogenic fusion complex [6, 17, 18] have been reported to recruit active DOT1L. The latter complex activity results in aberrant methylation of H3K79 (the DOT1L substrate) instead of H3K4 as in normally functioning MLL activity [19]. This off-gene epigenetic modification was found to be in high correlation with the mixed

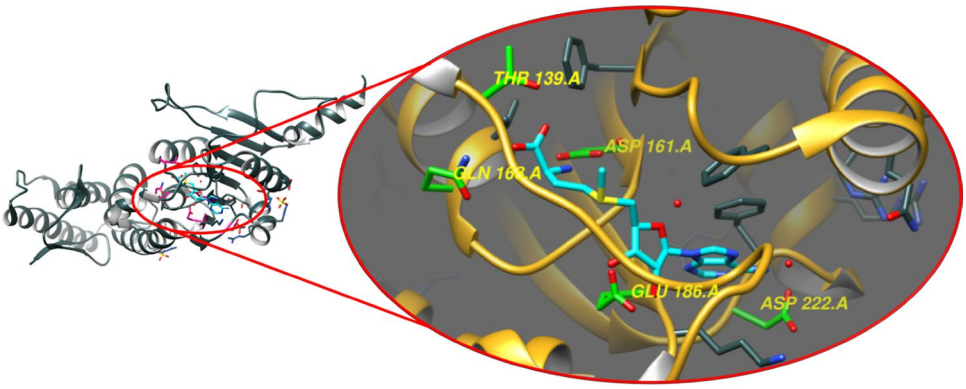


Fig. 1 Full view and binding site zoom of DOT1L crystal structure bound to SAM co-substrate (PDB entry code 1NW3, resolution at 2.5 Å). The residues that interact with the SAM co-substrate (residues 139, 161, 168, 188, 222) are shown in green colored carbon atoms, co-crystallized SAM is depicted in cyan. Water molecules as found in the PDB file are also displayed as red points

lineage leukemia (MLL)-rearranged leukemia insurgence, in which 11q23 translocation was found in > 70% of infants and > 10% of adults [20]. Due to this pathological profile, selective DOT1L inhibition has been proposed as a new approach to treat MLL-type rearranged leukemia-affected patients [21, 22]. Indeed, a survey on ChEMBL [23, 24] (database version 21) revealed hundreds of compounds tested in DOT1L-related assays (https://www.ebi.ac.uk/chembl/target/inspect/CHEMBL1795117), many more are available through patents [25–27] and in not yet listed ChEMBL articles [25, 28–31]. Moreover, several DOT1L co-crystallized with inhibitors are available through the Protein Data Bank (http://www.rcsb.org. PDB) [32]. In addition, molecular docking studies have also been reported [28–30, 33–35]. However, no comprehensive computational analysis has been carried out in order to develop a methodology to design novel inhibitors of DOT1L, yet.

The present study reports a multi-stage and combined approach through ligand-based (LB) and structured-based (SB) computational studies on a series of SAM-competitive DOT1L inhibitors. The study was conducted by maximizing the use of both available structural data present in the PDB and literature available information. The obtained results can be used both to perform virtual screening and to guide the rational design and discovery of new potential DOT1L ligands. To the best of our knowledge, this is the first report of quantitative LB and SB models applied to disclose new DOT1L inhibitors.

Computational strategy

To build predictive LB and SB models a procedure was set up by firstly inspecting the PDB seeking for all possible structural information on DOT1L co-crystallized ligands. The retrieved available complexes were subjected to a cleaning step in which all unresolved residues were built by means of homology modeling software. The completed complexes were then added of missing hydrogens and their geometries optimized to relax steric clashes. The optimized complexes were divided in ligand (key) and protein (lock) and used in a docking assessment protocol to select, from a list of free for academic, the best docking program. In parallel keys and key/lock pairs were used to develop 3-D QSAR and COMBINE models, respectively. The models were internally and externally validated for their reliability as predictive tools to select or prioritized potential DOT1L ligands.

The procedure workflow is depicted in Fig. 2 in which different sections are highlighted: (1) training set structural preparation (dark green paths) and collection of the related biological activities, (2) SB and LB models' building (light and dark red paths), (3) docking assessment (black curved paths), (4) test sets preparation (cyan paths for the crystal test set, dark blue paths for the modeled test sets) for external models' predictive ability evaluation (yellow paths). Each section is described in details in the following paragraphs.

Training set selection and structure preparation

At the beginning of the project, 15 DOT1L/ligand complexes were listed in the PDB [32, 36] (Table 1) and were all retrieved and subjected to a cleaning procedure as recently reported [37]. The procedure led to the preparation of SB-aligned minimized complexes from which three-dimensional quantitative structure–activity relationships (3-D QSAR, LB approach) or comparative molecular binding energy (COMBINE, SB approach) models were built using either extracted DOT1L inhibitor or ligand/protein pairs, respectively. While the manuscript was under preparation further 10 DOT1L crystal structure were released and promptly collected (see below Crystal Test Set preparation). Either crystal structures are all of good resolutions with an average value of 2.43 Å, ranging from 2.05 and 3.15 Å (Supplementary Material Table S1) and thus a good source of structural information.

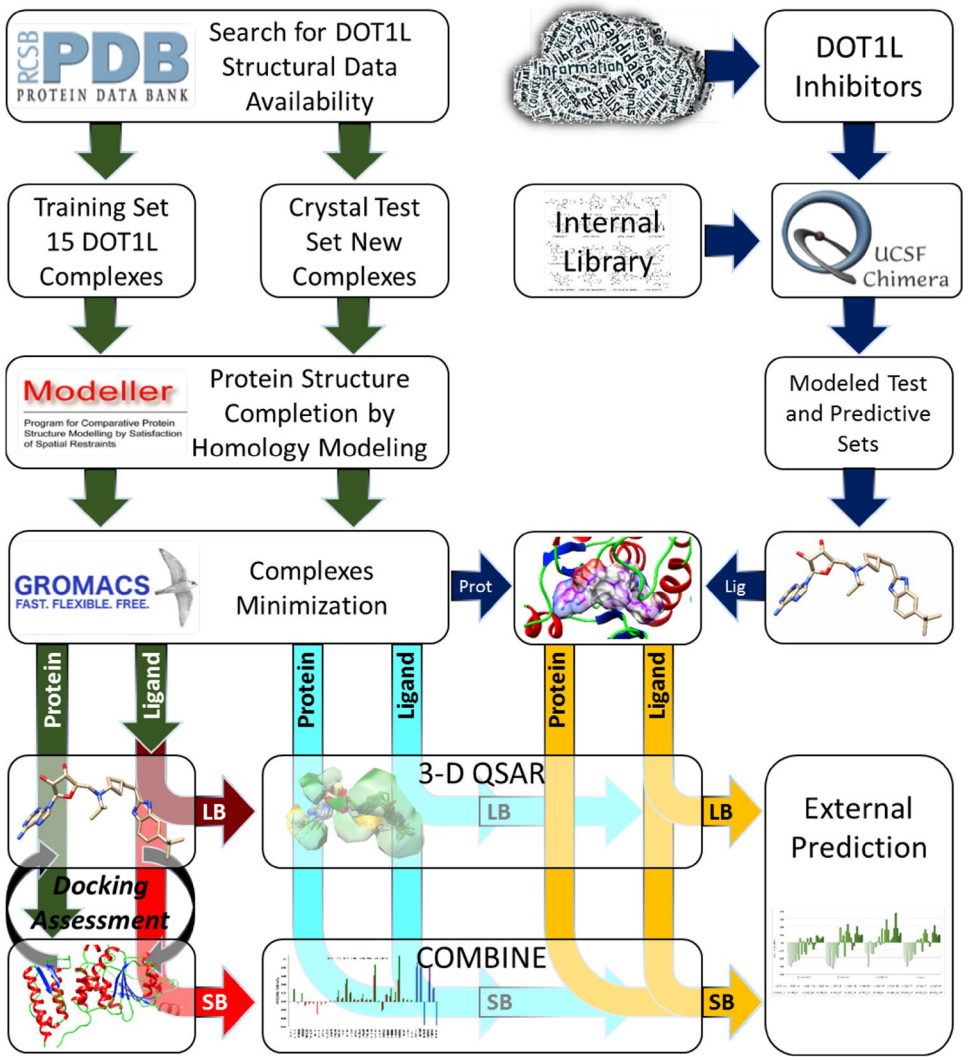


Fig. 2 Scheme of LB and SB computational approach. Dark green lane indicates the SB training and test sets selection and preparation. Black curved arrows represent the alignment rules by docking assessment. Light and dark red arrows designate the core models' building; dark blue routes specify test sets preparation; cyan and yellow paths indicate external models' prediction evaluation

Training set biological activity collection

Training set ligands' biological activities were collected from literature (Table 1). However, available bioactivities were not all homogeneous. In particular, for 10 out of the 15 co-crystallized ligands' *Ki* or K_m values (K-group) were available, while for five IC_{50} values were reported (IC-group) (Table 1). A recent report on a complete kinetic model for PRMT1 [44] displayed for SAM binding $k_2 << k_{-1}$ and comparison of DOT1L and PRMT1 co-substrate binding sites revealed bound DOT1L SAM bound conformation fully superimposed to PRMT bound SAH (and hence SAM) (Supplementary Material Fig. S1). These data support the $K_m \approx Ki$ approximation. To fully homogenate the training set activity type further literature investigation allowed to find for 3SX0, 4EKI and 3QOX bound inhibitors with both *Ki* and IC_{50} values. The three Ki/IC_{50} pairs displayed a squared correlation coefficient r^2 of 0.99 and the related regression equation (Supplementary Material Equation S1), in good agreement with Cheng-Prusoff [45] concept, was use to convert the five pIC_{50}s into the corresponding p*Ki* values (Supplementary Material Table S2).

3-D QSAR and COMBINEr models building

3-D QSAR (LB model)

A python implementation of 3-D QSAutogrid/R procedure [46] (Py-3-D_QSAR) was used to develop partial least square (PLS) field-based 3-D QSAR models (8 separated atom probes were used in turn, Supplementary Material Table S1), using the co-crystallized 15 inhibitors listed in Table 1 as training set (M-group). Systematic grid spacing variation (GRID) and variable pre-treatment optimization (VPO) let to determine the optimal GRID-VPO pretreatment parameters and derive the best PLS models. 3-D QSAR models robustness was estimated through different cross-validation (CV) techniques, *leave-one-out*

Table 1 DOT1L 2-D structures of co-crystallized inhibitors retrieved from the PDB and used as training set

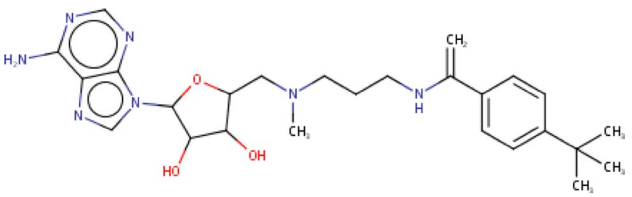
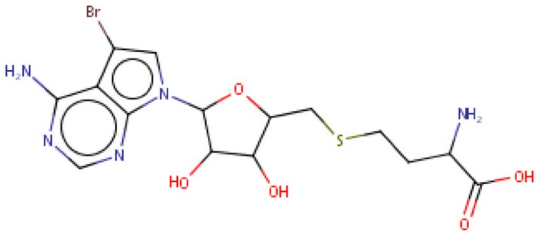
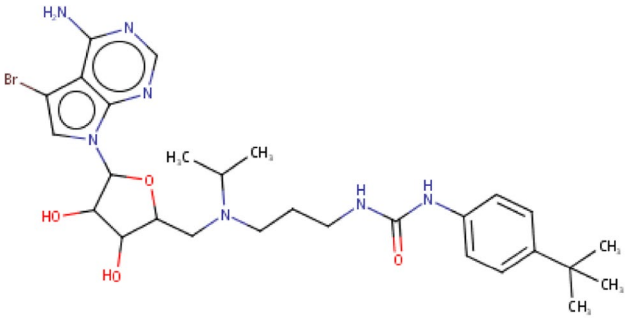
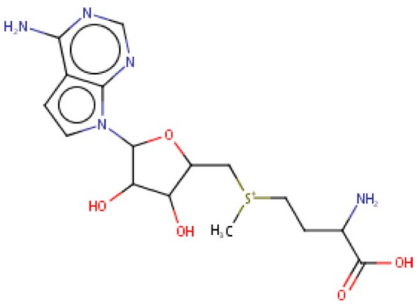
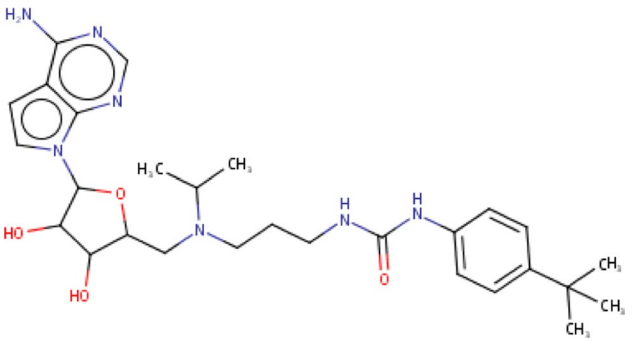
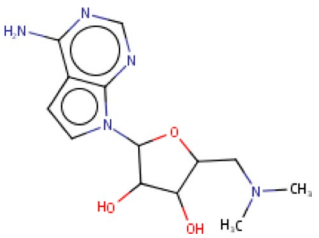
PDB code	2-D chemical structure	pAct[a]
4EKG [38]		7.89[b]
3SX0 [39]		7.11[c]
4ER6 [40]		9.422[c]
3QOW[e] [41] 1NW3[e] [42]		5.35[d]
4EKI[f] [38] 4ER5[f] [40]		9.52[c]
4EK9 [38]		4.42[b]

Table 1 (continued)

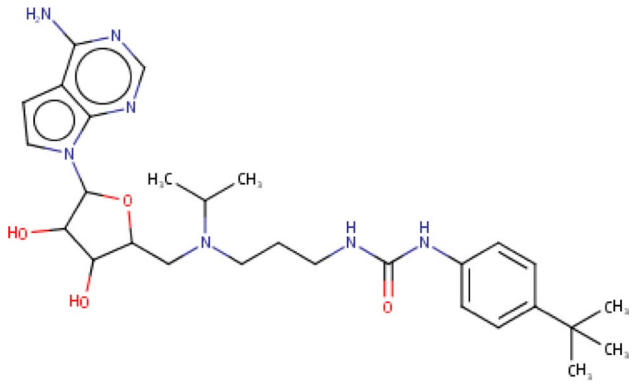
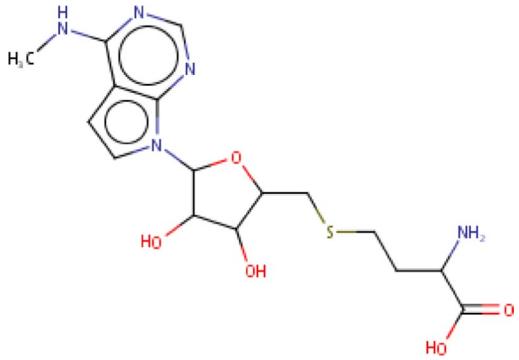
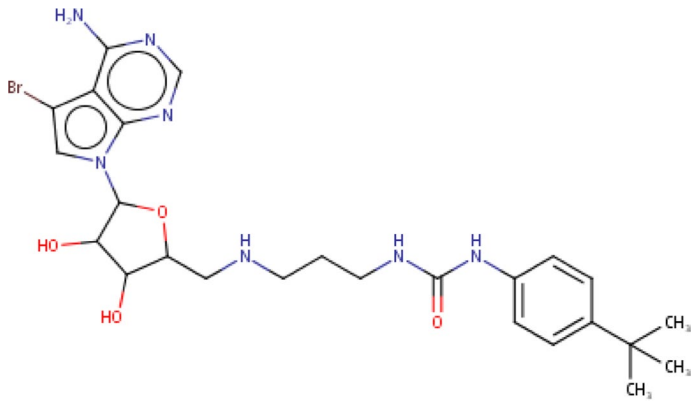
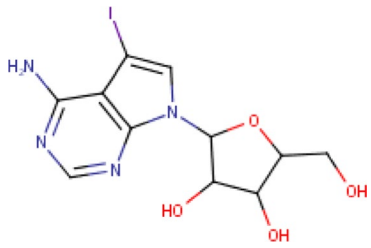
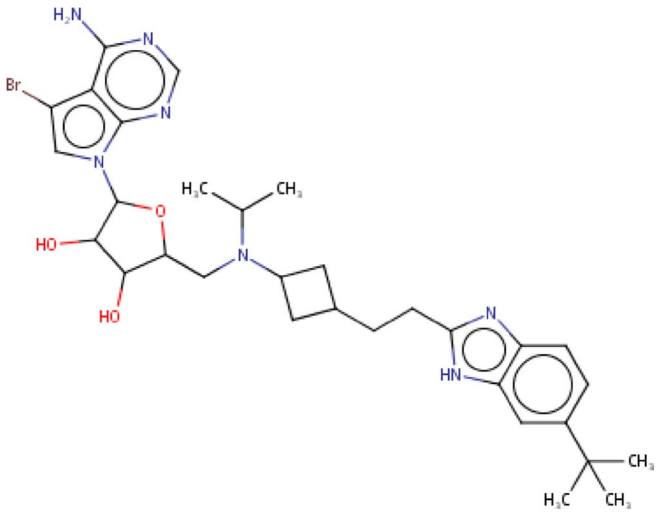
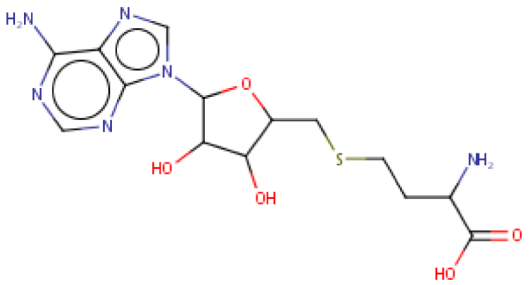
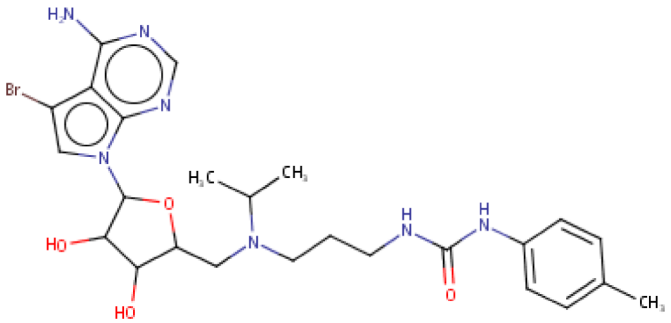
PDB code	2-D chemical structure	pAct[a]
4ER0 [40]		8.29[c]
3SR4 [39]		6.54[c]
4ER7 [40]		8.19[c]
3UWP [40]		3.95[c]

Table 1 (continued)

PDB code	2-D chemical structure	pAct[a]
4HRA [43]		10.10[b]
3QOX [41]		6.22[c]
4EQZ [40]		6.99[c]

The related experimental bioactivities pAct (pIC_{50}, K_i or K_m) and entry codes are also listed

[a]Bioactivities expressed as $pAct = -Log_{10}$ [activity expressed in molarity]

[b]Bioactivities available as pIC_{50}

[c]Bioactivities available as pK_i

[d]Bioactivities available as pK_m

[e]Co-crystals containing the same inhibitor but in different binding conformations

[f]Co-crystals containing the same inhibitor but in different binding conformations

(LOO, simplest CV, minimal perturbation), *leave-five-out* (leave some out with 5 groups, L5O, 20% perturbation), and *leave-half-out* (leave some out with 2 groups LHO, 50% perturbation). Y-scrambling technique in conjunction with CV was used to verify final model's lack of chance correlation [47]. Analysis of the Py-3-D_QSAR models was carried out by means of molecular graphical software (UCSF Chimera) by superimposing 3-D QSAR contour plots on the training set molecules and DOT1L binding site residues.

COMBINE (SB model)

The above-described 15 ligand/protein complexes were used to develop a series of COMBINE models. Herein a python implementation of the COMBINEr procedure [48] was used

(Py-COMBINEr) to investigate the correlation among p*K*is and four per-residue-based energetic interactions, namely electrostatic (ELE), steric (STE), desolvation (DRY) and hydrogen bonding (HB), each featured by a different contribution. Similarly, as in the Py-3-D_QSAR derived models, LOO, L5O and LHO cross-validation coupled with Y-scrambling were used to evaluate Py-COMBINEr models' robustness and lack of chance correlation. Analysis of the Py-COMBINEr models was carried out by means of histogram plots and most important binding site residues were characterized and graphically displayed with molecular graphical software (UCSF Chimera).

Docking assessment

In developing 3-D QSAR models a crucial step is represented by the alignment rules [49]. SB derived 3-D QSARs, like those herein defined, lack predetermined alignment rules to be applied to external test sets. A procedure to align the molecules under prediction (external test sets) need to be carefully carried out and assessed. Of course this apply also to COMBINE models were poses of ligands not yet co-crystallized need to carefully simulated. Therefore, SB alignment was set up by using open source molecular docking programs (AutodockVina [50] [herein just Vina], Surflex-dock [51] [herein just Surflex] and Plants [52]) considering its three different scoring functions, (Chemplp, plp and plp95). Docking algorithms are not yet fully optimized as the proteins are still mainly considered rigid [53] not including protein conformational and domain changes as consequence of ligand and protein binding (induced-fit). Herein cross-docking (or ensemble docking) was used to model MTS, JMCTS and MPS compounds. To this purpose experimental (EC) or randomized (RC) ligand conformation re-docking (RD) and cross-docking (CD) methods were used to assess the software ability in reproducing experimental binding modes of known co-crystallized inhibitors (docking assessment) [37, 54, 55]. As reported [37, 56] RMSD values, calculated between experimental and docked conformations, were used to evaluate docking accuracy (DA) [54, 57]. DA can be used to test how a given molecular docking software is capable to predict a ligand pose as close as possible to the experimentally observed, and can be calculated by the following equation:

$$DA = frmsd \leq a + 0.5(frmsd \leq b - frmsd \leq a)$$

whereas frmsd ≤ a and frmsd ≤ b represent the fraction of aligned ligands showing an RMSD value less than or equal to a and b cutoffs, respectively. The widely accepted standard for small molecule is that the correctly docked conformations are those displaying an RMSD value lower than 2 Å on all heavy atoms from the crystallographic structure of the ligand conformation as found in the inhibitor–enzyme complex. Conformation docked with RMSD between 2 and 3 Å are considered partially docked, whereas those showing RMSD values > 3 were mis-docked and thus not considered in the DA calculation.

Test set selection and preparation

Modeled test set (MTS)

A list of 24 DOT1L inhibitors (Supplementary Material Table S3) with known activities and unknown binding modes were collected from literature. MTS was compiled with molecules structurally related to those of the training set in order to test models' predictive abilities while avoiding any extrapolation. Test set molecules' starting conformations were directly modeled from the closest structurally related co-crystallized DOT1L inhibitor listed in Table 1. Their randomized conformations were generated and cross-docked into all training set of DOT1L apo-proteins. Docked MTS was used to assess 3-D QSAR and COMBINE models' predictive abilities with modeled conformations. During the preparation of the manuscript, Wang et al. [58] reported the application of SB techniques to disclose new non-nucleoside DOT1L inhibitors binding in the SAM pocket. Promptly, the reported active compounds were modeled (herein named *Journal Medicinal Chemistry Test Set*, JMCTS, Supplementary Material Table S4). The JMCTS structures were prepared similarly as for MTS.

Crystal test set (CTS)

While SB and LB models were under interpretation for manuscript preparation further DOT1L crystal structures were released in the PDB database. Among these, ten of them contained an inhibitor co-crystallized and promptly the newly available complexes were treated analogously as training set to prepare an external test set with known bound ligands' conformation (Supplementary Material Table S5). The CTS ligands and ligand/protein pairs were used to assess 3-D QSAR and COMBINE models' predictive abilities with experimental derived conformations.

Modeled prediction set (MPS)

An in-house library of untested compounds (not shown) was modeled similarly as MTS. The most promising MPS compounds were subsequently assayed for enzymatic bioactivities to evaluate real 3-D QSAR and COMBINE models' predictive abilities.

Results and discussion

Training set composition validation by means of 3-D QSAR predictive ability

Preliminary 3-D QSAR models based on eight different probes (Supplementary Material Table S6) were built using only K-group and were found statistically satisfying with high r^2 and cross-validated r^2 (q^2) values (Supplementary Material Table S7). IC-group molecules and associated pIC_{50}s values, used as external test, were correctly predicted with predictive squared correlation coefficients (r^2_{pred}) up to more than 0.7 (Supplementary Material Table S7 and Figs. S2 and S3), indicating IC_{50} as predictive for K_i values. Moreover, direct linear relationship found using three p*K*i/pIC_{50} pairs (Supplementary Material Table S2 and Equation S1) supported the pIC_{50} to p*K*i conversion and the merging all the 15 inhibitor/complexes in a single training set (M-group) to develop chemically wider 3-D QSAR and COMBINE model without including a great source of error.

Py-3-D_QSAR models definition on M-group

Molecular interaction fields (MIFs) were computed using a grid spacing of 1 Å to build eight initial 3-D QSAR models with the training set composed with M-group. Using standard variable pretreatment, initial models indicated a high level of MIF-activity correlation with satisfactory r^2 and q^2_{LOO} values ranging from 0.75 to 0.99 and from 0.45 to 0.66, respectively (Table S8 and Supplementary Material Figs. S4–S11). LOO, L5O and LHO CV statistical coefficients showed the models to be stable to perturbations. On the other hand, the LHO CV (50% perturbation) creates temporary random sub-model with only eight molcules indicating the limits of this CV application with relatively small training sets. Subsequently, GRID-VPO analysis were run with LOO CVs leading to statistically enhanced models without further optimization or variable selection (Table 2 and Supplementary Material Tables S9–S16). Considering r^2, q^2_{LOO} and $SDEP_{LOO}$ values, models 4 and 5, listed in Table 2, obtained with OA and N probes respectively, showed the best results. Therefore, they were further analyzed for either robustness or lack of chance correlation (Fig. 3 and Supplementary Material Figs. S12–S17). Indeed, GRID-VPO optimization lead to more robust OA (model 4) and N (model 5) probe based 3-D QSAR models with up to 21% higher q^2 values (compare q^2 values of Table S8 with those in panel C of Fig. 3). Y-scrambling procedure showed the model was not to be affected by any chance correlation as either r^2_{YS} or q^2_{YS} values were much lower than obtained with unscrambled models (Supplementary Material Figs. S12–S17).

Analysis of CV predicted values clearly indicated that low potent compounds (i.e. 3UWP in Supplementary Material Figs. S12–S17) have higher error of predictions (around 2 p*K*i units) whereas higher potent are predicted with low errors, this reflects a *J* shaped distribution point in the recalculated/predicted versus experimental p*K*is plots (Fig. 3). This is an expected scenario in a 3-D QSAR study, in fact during CV run, low potent compounds sharing similarities with the reduced training set were predicted more potent. In parallel, highly active compounds were predicted less potent but with lower errors of prediction. Models 4 and 5 were selected for the subsequent 3-D QSAR graphical interpretations.

Table 2 GRID-VPO analysis for all considered probes

#	Probe	r^{2a}	q^2_{LOO}[b]	$SDEP_{LOO}$[c]	PC	GRID	Cut off	Zeroing	Min STD
1	A	0.92	0.72	0.99	2	2.30	5	0.009	0.005
2	C	0.92	0.72	0.99	2	2.30	5	0.010	0.005
3	HD	0.95	0.72	0.98	2	2.60	5	0.003	0.040
4	OA	0.97	0.78	0.88	3	2.30	4	0.010	0.025
5	N	0.97	0.80	0.84	3	2.30	5	0.010	0.030
6	NA	0.97	0.77	0.86	3	2.30	5	0.010	0.030
7	e	0.99	0.77	0.82	5	0.85	5	0.008	0.005
8	d	0.85	0.78	0.87	2	1.95	1	0.009	0.015

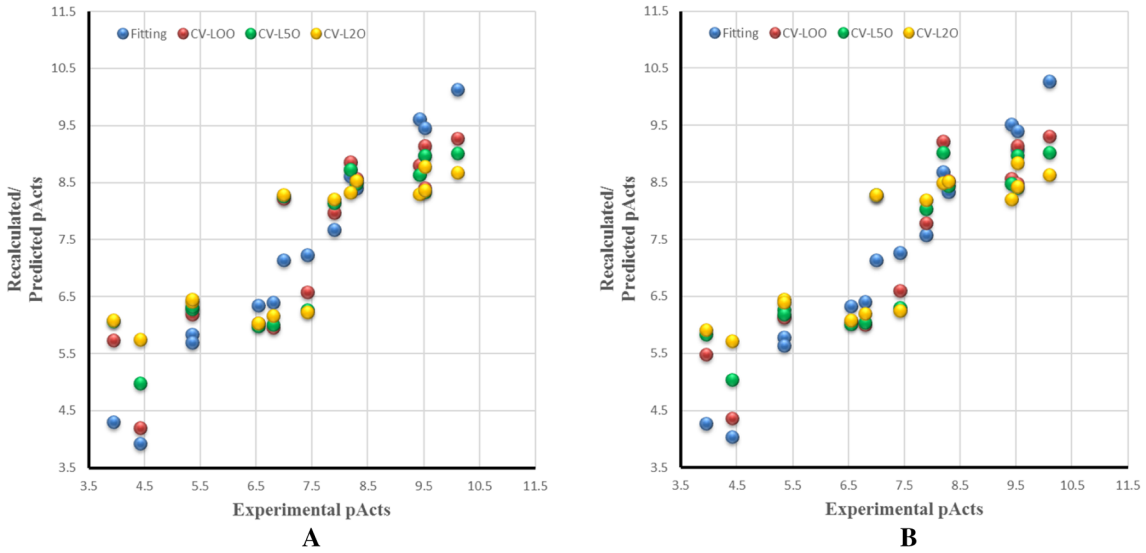
Only best VPO parameters are shown

PC principal components/latent variables, *GRID* grid spacing, *CutOff* maximum and minimum grid energy cutoff value, *Zeroing* zeroing cutoff, *MinStd* minimum standard deviation cutoff

[a]Conventional square-correlation coefficients

[b]Cross-validation correlation coefficients obtained from L5O CV

[c]Standard deviation error of prediction in L5O CV



#[a]	Probe[b]	PC[c]	Fitting		CV-LOO		CV-L5O		CV-LHO	
			r^{2d}	SDEC[e]	q^{2f}	SDEP[g]	q^2	SDEP	q^2	SDEP
4	OA	3	0.97	0.32	0.79	0.85	0.72	0.98	0.68	1.06
5	N	3	0.97	0.32	0.79	0.83	0.74	0.95	0.70	1.03

C

Fig. 3 **a** Recalculated (blue dots), LOO CV predicted (red dots), L5O-CV predicted (green dots) and LHO-CV (yellow dots) p*K*is from OA (panel **a**) and N (panel **b**) probes derived 3-D QSAR models. Models' 4 and 5 statistical PLS results are reported in panel **c**: [a]Model number as in Table 2. [b]Probe atom to calculate MIF. [c]PC: optimal principal components. [d]Conventional square-correlation coefficients. [e]Standard deviation error of calculation. [f]Cross-validation correlation coefficients obtained from CV. [g]Standard deviation error of prediction

Py-3-D_QSAR models graphical analysis

A representative feature of any MIF-based 3-D QSAR is the intrinsic skill to correlate graphically fragments of training set molecules with their biological activities. Consequently, for a given SB-aligned molecule, this correlation can involve the binding site with which the fragments interact. Superimposition of corresponding crystal structures on the classical 3-D QSAR contour plots represents a tool to explore the model's biophysical rationale. Average activity contribution (AAC) contour plots obtained by MIF average values and the PLS coefficients (COEFs) scalar products are of marked use. Differently from most common CoMFA plots, obtained by the scalar product of MIF standard deviation and COEFs values, in case of steric interactions, AACs are of direct use and have only two combinations of values: positive and negative. Positive values indicate regions around training set molecules that explain positive activity contribution, while negative values correlate with negative contributions. At higher resolution, further information can be gathered by means of activity contribution (AC) plots obtained by individual molecules' MIF values and COEFs scalar product. MIF and COEF plots can also be used to aid AAC and AC plots interpretation and all together they are useful to design molecules starting from training set compounds from a LB perspective. Being amide nitrogen atom mainly involved in steric interactions, superimposition of OA and N derived AAC plots (AAC_N and AAC_{OA}) shows that the foremost feature of the MIF/activity relationships (3-D QSAR) is related to the bound conformations' molecular volumes. The only small notable difference is in proximity of the ribose hydroxyl groups (Fig. 4, Supplementary Material Figs. S18, S19 and movies S1–S16).

For the N probe, an average negative contribution is associated with the two hydroxyls (slightly more on sugar position 3′-OH). On the contrary, a positive cyan polyhedron for AAC_{OA} overlaps indicating that a possible compromise would be to maintain hydrogen bonding donator atom (ligand's counter part for OA probe interaction). Interestingly, comparing the most (4HRA) and a less (3RS4) potent ligand conformations, the latter's hydroxyl group is shifted away of more than 1.4 Å from the above described positive polyhedron (Supplementary Material Fig. S19). In fact, smaller MIF_{OA} polyhedron can be observed around 3′-OH (compare Supplementary Material Figs. S18E and S19E) of 3RS4 leading to decreased AC_{OA} associated one

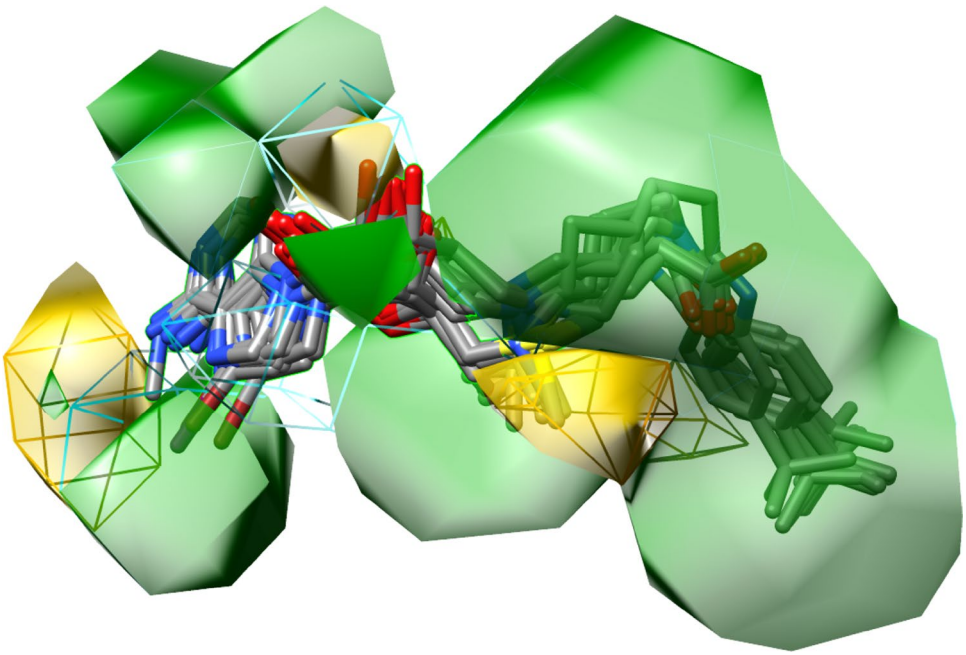


Fig. 4 AAC contour plots. AAC_N are displayed in surface style (positive values in green and negative values in yellow). AAC_{OA} are in mesh style (positive values in cyan lines and negative values in orange lines). Aligned training set is displayed colored by element (grey: carbon, red: oxygen, blue: nitrogen, yellow: sulphur). Hydrogen atoms are not displayed for sake of clarity. AAC surfaces iso-contribution values were not filtered

(compare Figs. S18G and S19G). Regarding the activity-related steric influence, by comparison of the training set most (4HRA) and low (3SR4) potent molecules-associated MIF and AC plots, it is possible to note the presence of big positive AC areas (either N or OA probes) overlapping the 4HRA phenylimidazole moiety which extends further the natural SAM co-substrate methionine group. This area is commonly occupied by all most potent compounds sharing bulky groups instead of the SAM methionine as outlined in the AAC countour plots (Fig. 4). A second steric featured difference among 4HRA and 3SR4 AC plots is related to the alkyl side chain (isopropyl for 4HRA) bound to the nitrogen replacing the SAM sulphur atom, which fulfills an important AC green polyhedron (Fig. S18H). Whereas the area is missing for the least potent *S*-adenosyl-L-homocysteine (SAH) ligand analogue 3SR4 (Supplementary Material Fig. S19). Finally, small negative AACs aeras are visible in Fig. 4 around the adenine amino group and, as it results by inspection of the ACs plots, the negative activity contribution is greater for 3SR4 where a methyl group is replacing one of the two hydrogen atoms.

As anticipated, overlap of DOT1L corresponding binding sites with AAC or AC contour plots allows implementing the LB nature of the 3-D QSAR models with SB information, thus allowing deeper analysis and consideration (Supplementary Material Fig. S21 and movies S17–S32). Actually, AAC maps overlapped to DOT1L SAM binding sites residues display a good complementarity. In particular, referring to 4HRA ligand/protein complex, the important positive contributing steric interactions (AAC and AC maps) around the *tert*-butylphenylimidazole fulfill a deep pocket delimited mainly by Pro133, Leu143, Met147, Val169, Phe239, Val 267 and Tyr312 side-chains; the isopropyl group-associated AAC and AC contours fit a small cleft formed by Asn241 and Phe245 side chains. Positive contributing AAC small polyhedrons close to 2′- and 3′-ribose hydroxyl groups correctly describe the interfaces with Lys187 amide nitrogen and Glu186 γ-carboxylate, respectively. Small positive AAC and AC plots are also observable in a pocket defined by Val185, Gly221 and Asp222 main chains and side chains indicating adenine C2 position located in pyrimidine ring as a possible modification point to potentially increase the activity. The negatively contributing polyhedron around the adenine amine group designates that deamination could be favorable for the activity. To complete the description, the two small areas around adenine nitrogen N7 account for 3SX0 and 3UWP steric interaction of bromine and iodine atoms, respectively. Regarding molecule 3RS4, in general the associated AC plots are not fulfilling the SAM binding site correctly indicating the lack of activity contribution. On the other hand, greater negative contribution maps around adenine amino group specify some steric hindrance exerted by the methyl substituents, likely due to negative interactions with Leu224 side chain.

The above observation suggested that a linear correlation could exist between p*K*is and classical QSAR calculated parameters. Indeed, mono parametric linear regressions (LR) performed with molecular weight (MW), molecular refractivity (MR), logarithm of *n*-octanol/water partition coefficients (LogP), total polar Surface Area (TPSA), number of heavy atom (HAC), number of hydrogen acceptor atoms (HA), number of hydrogen donor (HD), number of rotatable bonds (RB) and number of rings (RC) returned r^2 values up to 0.86 being the higher those obtained with MW,

MR and HAC (Supplementary Material Table S17). The simple QSARs thus confirmed that actually a correlation exists between inhibitory activities and steric parameters supporting the above 3-D QSAR analysis.

COMBINE models definitions

Through application of the Py-COMBINEr protocol, four ligand/residues energetic fields were investigated: electrostatic (ELE), steric (STE), desolvation (DRY) and hydrogen bond (HB), and all the possible 11 field combinations (Supplementary Material Table S18). For each single field and their combination, the corresponding COMBINE model's robustness and lack of chance correlation were evaluated by means of LOO, L5O and LHO cross-validations. High level of models' robustness was achieved without the need of any optimization as evinced by the most restrictive q^2_{LHO} values ranging from 0.36 to 0.66 (Supplementary Material Table S18). Furthermore, the models developed with block scaled data showed to be slightly better than those obtained with unscaled data.

Among the COMBINE models, the one endowed with the best statistical values ($r^2 = 0.94$ and $q^2 = 0.66$ at three principal components and LHO CV) was the one obtained with STE and HB field combination (Fig. 5). Optimum number of PC was calculated by checking percentage decrease of standard deviation error of prediction (SDEP) values, setting as significant the next component with a SDEP decrease ($SDEP_{\%}$) higher than 5% (Supplementary Material Table S19 and Figs. S22–S66), this procedure is used in the GOLPE program, a software used to develop either 3-D QSAR or COMBINE models [59]. Consequently, all the subsequent analyses were conducted on the STE + HB field-derived COMBINE model at 3 PCs.

COMBINE model analysis

Similarly as for the above analyzed 3-D QSAR models, representative features for COMBINE models are the graphical correlations of training set molecules with their biological activities by means of histogram plots (HPs) that aid to individuate protein fragments more involved in modulating

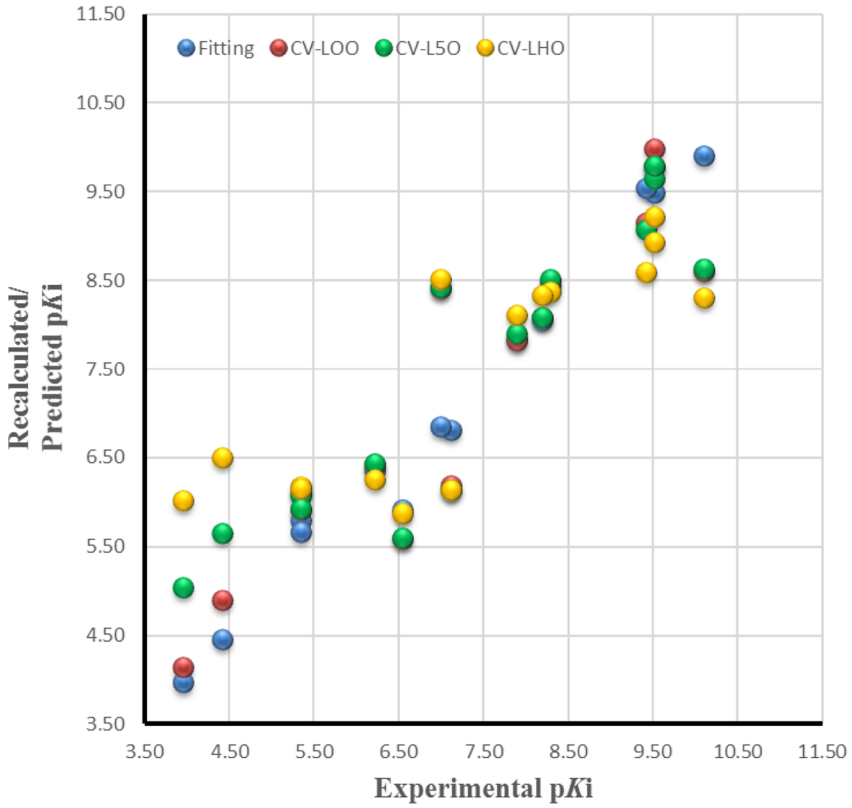


Fig. 5 Recalculated (blue dots), LOO CV predicted (red dots), L5O-CV predicted (green dots) and LHO-CV (yellow dots) p*K*is from STE-HB derived COMBINE model. Statistical PLS results for model 7 s are also reported: [a]model number as in Table S12. [b]Field combination. [c]PC: optimal principal components. [d]Conventional square-correlation coefficients. [e]Standard deviation error of calculation. [f]Cross-validation correlation coefficients obtained from CV. [g]Standard deviation error of prediction

#[a]	Molecule-Residues Interaction Field Combination[b]	PC[c]	Fitting		LOO		L5O		LHO	
			r^{2d}	SDEC[e]	q^{2f}	SDEP[g]	q^2	SDEP	q^2	SDEP
7s	STE.HB	3	0.95	0.41	0.86	0.71	0.81	0.81	0.66	1.098

inhibitors' activities. Thus, molecule–residue average activity contribution (MRAAC) HPs obtained by average molecule–residues interactions (AMRIs) values and the PLS coefficients (COEFs) scalar product are used to deeply analyze COMBINE models (Supplementary Material Fig. S67). As AACs, MRAACs have only two combination of values: positive and negative. Positive values indicate protein residues that are in close proximity to the training set molecules, while negative values indicate negative contributions due to the protein residues with which the interaction is penalizing. In particular, from the histogram plot reported in Fig. S67 residues Gly163, Lys187, Asn241 are those more involved in positively modulating the ligand/residues steric interactions with average MRAAC values of about 0.54, 0.81 and 0.59 p*K*is units, respectively. Lys187 is involved in maintaining positive contacts with the SAM's adenine moiety, while Gly163 and Asn241 that form part of the SAM binding site hosting the modified methionine sulphur atom (changed to an alkylated nitrogen) of the more potent DOT1L inhibitor (Fig. 6a). Only a few of residue displayed slightly negative steric MRAAC values. Among these it is interesting to point that more of them are located in a DOT1L flexible loop (Pro130–Thr139) which is unstructured in the complexes containing potent inhibitors (Supplementary Material Fig. S68). The coordinates of this loop are not always present in the PDB structures, therefore were modeled by homology modelling. COMBINE data connected with this modeled loop may contain source of errors; in any case, the model recorded the residues of this flexible loop as important residue to which an inhibitor should not set up non-bonding interactions. Indeed, regarding HB ligand/residues interactions Asp161, Gly163 and Asp222 are those mainly responsible for positive average activity contributions (Fig. 6b). Inspecting the most potent DOT1L inhibitor, at least four hydrogen bonds (Fig. S69) mostly account for its picomolar activity (p*K*i = 10.10, Table 1). Although engaged in hydrogen bonding to compounds bearing a carboxylate group similarly to the SAM, Asn241 seems detrimental for the activity with an MRAAC value lower than − 0.6 p*K*is (Supplementary Material Fig. S67). The negative activity contribution associated with a hydrogen bond seems paradoxical, but comparing 4HRA and 3UWP ligands (Fig. 7 and Supplementary Material Fig. S67) the loss of a hydrogen bond (3UWP) is fulfilled by three new ones (4HRA). Therefore, the PLS algorithm correctly assigned a positive COEF value to turn into negative the MRAAC cost.

At higher resolution, further information can be gathered with molecule–residue activity contribution (MRAC) histogram plots obtained by individual molecules' molecule/residue interactions (MRIs) and COEFs related values. Focusing on the most and least potent ligands, 4HRA and 3UWP, and comparing their MRACs values (Fig. 7) it is possible to get further details on the ligand/residues interactions that account for almost six orders of magnitude in potency spectrum (see Table 1, Δp*K*i = 6.15). Focusing to absolute MRACs values > 0.15 p*K*i units, only 5 and 12 residues for 3UWP and 4HRA, respectively, delineate the most important sterical ligand/protein per residues interactions accounting for a Δp*K*i of about 2.7. While for hydrogen bonding for the most active ligand (4HRA) a high positive contribution is associated with Gly163 and on the contrary for the least one (3UWP) there is a high negative contribution related to Glu186 (Fig. 7). A straightforward graphical picture is displayed by plotting the residues associated with positive and negative activity contributions (Supplementary Material Figs. S70–S73). For the most potent compound, a wider positive area promptly delineates the positive residues. In particular, the highest activity of 4HRA is mainly due to

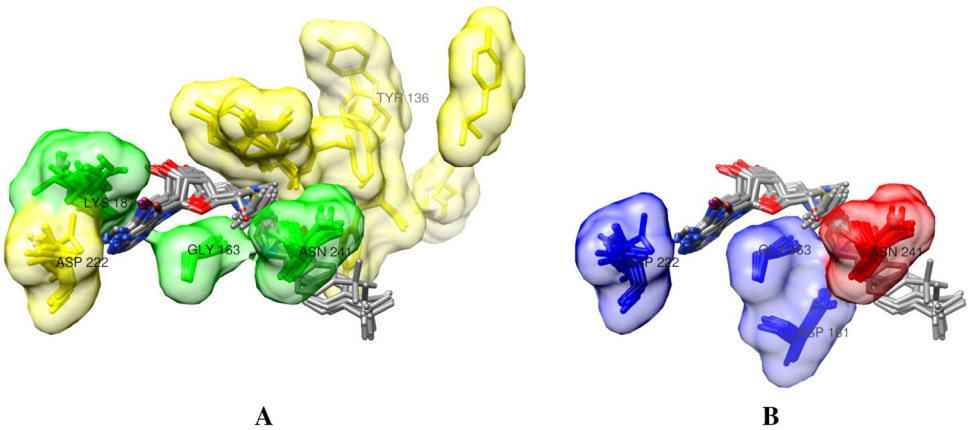


Fig. 6 Representation of positively and negatively contribution residues as depicted from COMBINE model 7s (Table S12, Supplementary Material) analysis (see text). STE MRAACs (**a**) are depicted with surfaces colored in green (positive) and yellow (negative). HB MRAACs (**b**) are depicted with surfaces colored in blue (positive) and red (negative). Figure are made to directly compare the STE and HB surfaces. All training set DOT1Lis are depicted in atom type colored atoms

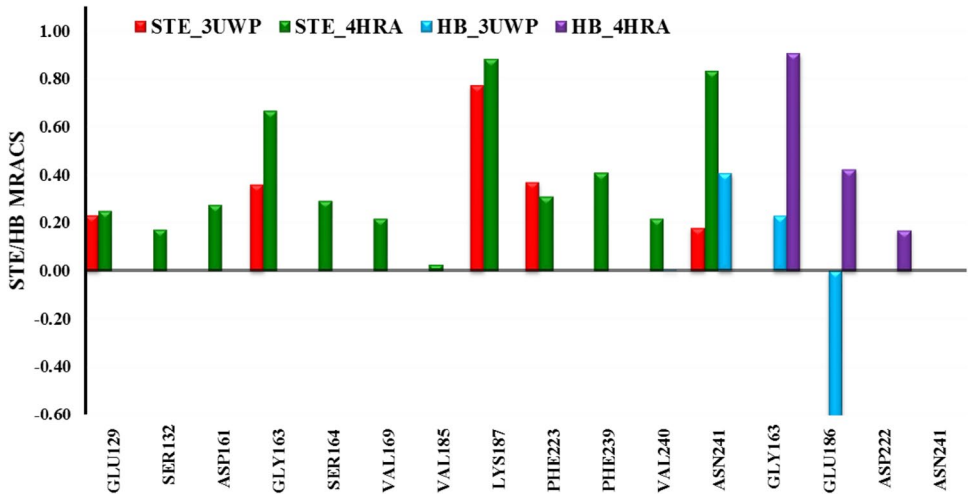


Fig. 7 Representation of positively and negatively contribution residues as depicted from COMBINE model 7s (Table S12, Supplementary Material)

a greater extent of positive ligand/residues STE interactions, while the 3UWP lower potency can be ascribed to both a sterical and hydrogen bonding unfavorable complementarity (Fig. S74).

Combined 3-D QSAR and COMBINE analysis

The use of two computational approaches (LB and SB) can complement each to other and compensate respective weaknesses, realizing a synergism to increase the SBDD potentialities. Combining 3-D QSAR graphical results with those of COMBINE has double advantage: first, the overlapping models can be used to confirm goodness of each other; second, as an ultimate analysis, it can give a broader scenario to design new molecules. Effectively, as shown in Fig. S75 (see also associated movies in Supplementary Material), the LB and SB approaches display a high level of agreement and with respect to 4HRA ligand sections (adenine, ribose and replaced methionine moiety). Thus some structure–activity relationship rules about potential new derivatives can be derived (Fig. 8). Concerning the adenine ring, the C6 amino group can be deleted to reduce negative steric interaction with Asp222 and Leu224 side chains; N1 is important for a hydrogen bond with Phe223 amide NH which can be replaced by a carbonyl group to strengthen the binding; N7 can be replaced with a carbon atom while inserting a bulky branching; C2 and N3 should be maintained or at least replaced with a two carbon atoms moiety. About the ribose,

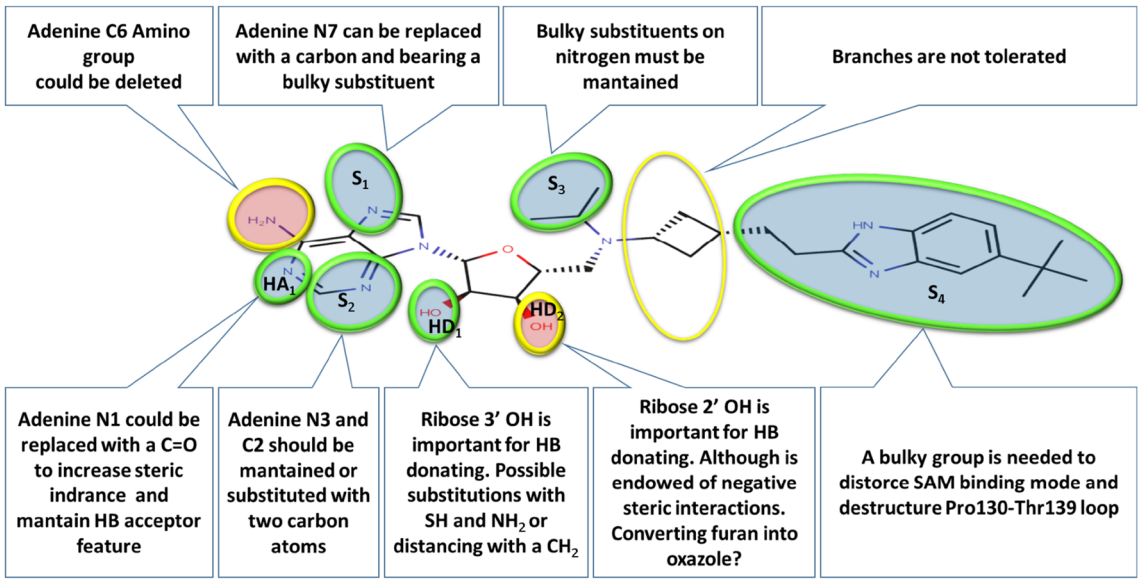


Fig. 8 4HRA ligand and potential modification points are suggested. Steric, hydrogen bonding acceptor and hydrogen binding donating features are indicated by S_{1-4}, HA and HD_{1-2}, respectively

the 3′-hydroxy group is important for a hydrogen bond with Glu186 carboxylate as it could be replaced with SH or NH_2, and also a methylene can be inserted. The 2′-hydroxy group is less prone to establish hydrogen bond but C–OH can be replaced by a NH to convert the tetrahydrofuran ring into tetrahydrooxazole. Regarding the replaced SAM methionine group, the models confirmed that bulky groups are needed to destructure the Pro130–Thr139 DOT1L loop also supported by a good correlation with classical steric QSAR paramenter such as MW and MR; finally, branching is not tolerated in the methionine binding site.

SB alignment assessment

The validity of any applied docking methodology needs to be established (docking assessment) by evaluating the binding modes reproduction of the ligands for which X-ray structures are available. As previously reported [54–56], docking assessment was used to evaluate the most suitable docking program among a list of five program/scoring function combinations (Vina, Plants [with three scoring function, chemplp, plp and plp95] and Surflex) through experimental conformation re-docking (ECRD), random conformation re-docking (RCRD), experimental conformation cross-docking (ECCD), random conformation cross-docking (RCCD) procedures [54]. Analysis of calculated docking accuracy percentages (DA%) revealed Surflex as the program displaying the highest DA% value in all four procedures (ECRD-DA% = 100, ECCD-DA% = 86.7, RCRD-DA% = 93.3 and RCRD-DA% = 90, Table 3 and Supplementary Material Tables S20–S23). These results indicate that in case of an effective DOT1L inhibitor Surflex is expected to predict its binding mode with a reasonable low error.

Table 3 The docking results obtained from the random conformation cross-docking software used

Statistical data	Plants			Surflex	Vina
	Chemplp[a]	plp[a]	plp95[a]		
Min[b]	0.53	0.61	1.32	0.46	1.21
Max[c]	11.93	11.74	12.41	10.32	18.56
Average[d]	3.56	6.25	4.76	1.73	7.24
StDev[e]	3.01	3.91	3.49	2.42	4.96
DA%[f]	50.00	23.33	40.00	90.00	20.00

[a]The scoring function names as implemented in the plants docking programs

[b]Minimum value of RMSD obtained for each docking software used

[c]Maximum value of RMSD obtained for each docking software used

[d]Average value of RMSD obtained for each docking software used

[e]Standard deviation value

[f]Docking accuracy as defined in the experimental section

3-D QSAR and COMBINE models' predictive ability evaluation through external test sets

Modeled test sets (MTS)

In view of the above docking assessments, Surflex was selected (SB alignment rules) to dock the MTS molecules. Due to high structural similarity with training set molecules, the obtained MTS docked conformation were found to overlap quite well those of the training set (Supplementary Material Fig. S76). The MTS conformations were then directly used to evaluate the predictive abilities of either 3-D QSAR or COMBINE models for compounds with unknown experimental binding modes. In general, the model over predicted low potent compounds and under predicted high potent ones. Nevertheless, the models proved to have predictive abilities at different extents (Supplementary Material Table S24), STE.HB COMBINE model based showed the lowest predictive SDEP ($SDEP_{pred}$ = 1.58), whereas OA and N-based 3-D QSAR models had a lower predictive power with higher $SDEP_{pred}$ values of 1.78 and 1.80, respectively.

For a prospective consensus scoring function application combining LB and SB techniques, the average $SDEP_{pred}$ was also calculated to be 1.67 Apart the $SDEP_{pred}$ values, the models were also evaluated for their propensity to discriminate high from low potent compounds: in general, the models displayed a high capability in identifying highly potent compounds with low propensity to false negatives (highly potent compounds predicted with low p*K*i values). The 3-D $QSAR_{OA}$ model displayed a higher accuracy prediction for more potent compounds (smaller under-prediction), whereas the COMBINE one had a balanced over and under-prediction profile (Supplementary Material Fig. S77).

During the preparation of the manuscript, Wang et al. reported the application of SB techniques to disclose new non-nucleoside DOT1L inhibitors binding in the SAM pocket [58]. Promptly the reported active compounds were modeled (herein named JMCTS), cross-docked and used as a further external modeled test set. The 3-D QSAR and COMBINE models correctly predicted the micromolar range potencies with SDEP values as low as 0.91 for the consensus scoring (Table 4).

Crystal test set (CTS)

As already above cited, during the preparation of this manuscript a new series of DOT1L complexes were released (Supplementary Material Tables S1, S5 and S25). Thus, the 3-D QSAR and COMBINE models were promptly applied to these complexes to evaluate their predictive ability with experimental data. As expected, the models showed low errors of prediction (Supplementary Material Table S25 and Fig. S78). As a support to the above consensus scoring

Table 4 Experimental and predicted pIC_{50} for JMCTS

#[a]	Exp pK_i[b]	Predicted pKis			
		3-D QSAR		COMBINE	Average[c]
		Probe OA	Probe N		
6	4.34	5.77	6.44	4.98	5.37
9	4.63	3.82	4.14	3.96	3.89
10	4.46	5.74	5.91	4.53	5.13
15	4.70	5.94	6.11	4.72	5.33
16	4.33	6.01	5.97	4.41	5.21
17	4.64	4.32	4.43	4.57	4.44
21	3.91	5.32	5.43	3.88	4.60
39	4.69	6.11	6.07	4.81	5.46
40	3.97	6.78	7.31	4.69	5.73
SDEP	–	1.52	1.73	0.40	0.91

[a]Compound enumeration as in the original article

[b]pKi derived from the experimental pIC_{50} using Supplementary Material Equation S1

[c]Average of predicted activities by the three models

function application, the average $SDEP_{pred}$ values was lower between 7 and 24% than the single models' $SDEP_{pred}$ values. Nevertheless, most of the CTS ligands bound conformations were only partially overlapped to those of the training set (Supplementary Material Fig. S78), thus explaining the non-perfect predictions.

4ER5 and 5JUW ligands are those more closely related to the training set structures and were predicted with low errors by the models; whereas 4WVL although strictly related to 4ER0, bears an acetyllysine-like side-chain on the adenine amino group extending the interactions not fully represented in the 3-D QSAR and COMBINE models was slightly over-predicted. Ligands contained in 5DRT, 5DRY, 5DSX, 5DT2, 5DTM, 5DTQ and 5DTR complexes were only partially superimposed to the ligand training set and although were all over-predicted by either models SDEP values were still not too high (Table S25).

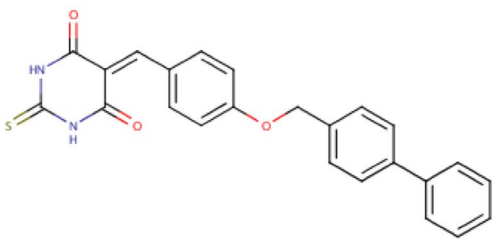
Modeled predictive set (MPS)

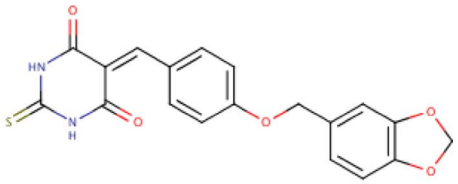
A series of 87 compounds taken from our in-house library (not shown) was modeled and subjected to both Py_3-D_QSAR and Py-COMBINEr models. Among the virtually screened compounds, only two of them ($\mathbf{1}_{MPS}$ and $\mathbf{2}_{MPS}$) showed predicted p*Ki* in the low micromolar or even sub-micromolar range and were therefore promptly assayed as DOT1L inhibitor by using AlphaLISA technology (PerkinElmer) and then confirmed by radioactive assay (Table 5 and Supplementary Material Fig. S80).

Chemistry

The synthetic route for the preparation of the final thiobarbiturates $\mathbf{1}_{MPS}$ and $\mathbf{2}_{MPS}$ is depicted in Scheme 1. On one hand, treatment of commercially available 4-(chloromethyl)-1,1′-biphenyl with commercial 4-hydroxybenzaldehyde in the presence of anhydrous potassium carbonate and sodium iodide in dry acetonitrile at reflux brought to the

Table 5 Structures of $\mathbf{1}_{MPS}$ and $\mathbf{2}_{MPS}$ and their 3-D QSAR and COMBINE models' predictions for DOT1L inhibition



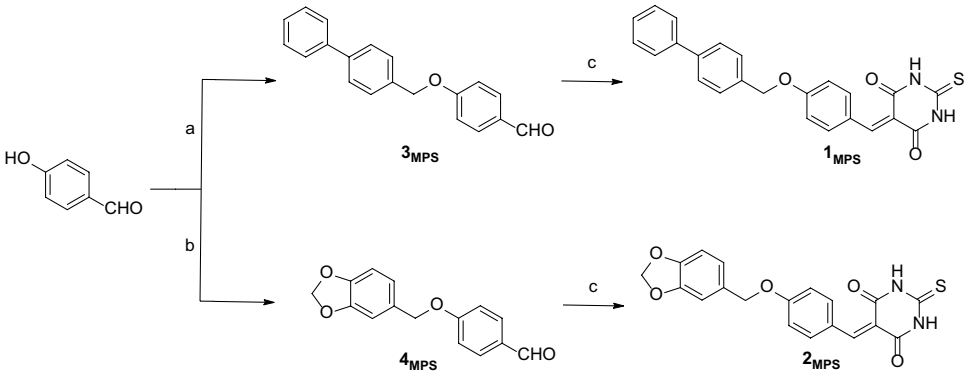


1_{MPS}					2_{MPS}	
#	Predicted p*Ki*s				Experimental pIC_{50} (p*Ki*)[b]	
	3-D QSAR		COMBINE	Average[a]	AlphaLISA	Radioactive assay
	Probe OA	Probe N				
1_{MPS}	6.52	6.42	5.54	6.16	5.14 (4.40)	5.02 (4.27)
2_{MPS}	6.81	7.05	5.04	6.3	5.10 (4.35)	4.11 (3.22)

Experimental data are also reported for comparison

[a]Average of predicted activities by the three models

[b]p*Ki* derived from the experimental pIC_{50} using Supplementary Material Equation S1



Scheme 1 Synthesis of Compounds 1_{MPS} and 2_{MPS}: Reagents and conditions: (*a*) 4-(chloromethyl)-1,1′-biphenyl, K_2CO_3, NaI, dry CH_3CN, reflux, 4 h; (*b*) benzo[*d*] [1,3] dioxol-5-yl-methanol, PPh_3, DIAD, dry THF, N_2, 0 °C→rt, 16 h; (*c*) 2-thiobarbituric acid, EtOH, 80 °C, 6.5 h

Table 6 Selectivity of $\mathbf{1_{MPS}}$ and $\mathbf{2_{MPS}}$ towards DOT1L inhibition

Compd.	$IC_{50,}$ μM			
	DOT1L	G9a	SET7/9	PRMT1
1_{MPS}	7.906 ± 0.012	No inhibition	No inhibition	No inhibition
2_{MPS}	7.211 ± 0.011	No inhibition	No inhibition	No inhibition

4-([1,1′-biphenyl]-4-ylmethoxy)benzaldehyde intermediate $\mathbf{3_{MPS}}$. On the other hand, the 4-(benzo[*d*] [1,3] dioxol-5-ylmethoxy)benzaldehyde intermediate $\mathbf{4_{MPS}}$ was achieved by a Mitsunobu reaction performed between the commercially available benzo[*d*] [1,3] dioxol-5-ylmethanol and the 4-hydroxybenzaldehyde in the presence of triphenylphosphine and diisopropyl azodicarboxylate (DIAD) in dry THF. The two aldehydes were then condensed with commercial 2-thiobarbituric acid in ethanol at reflux to give the required final products $\mathbf{1_{MPS}}$ and $\mathbf{2_{MPS}}$.

Enzyme assays

To test $\mathbf{1_{MPS}}$ and $\mathbf{2_{MPS}}$, human recombinant DOT1L (residues 1-416) containing an *N*-terminal GST-tag was used as enzyme source, and purified oligonucleosomes from HeLa cells and SAM were used as substrate and co-substrate, respectively (AlphaLISA technology) (Tables 5, 6 and Fig. S80). To confirm the inhibition data, $\mathbf{1_{MPS}}$ and $\mathbf{2_{MPS}}$ were also tested in radioactive assay with human DOT1L and oligonucleosomes/3H-SAM, in a 10-dose IC50 mode with twofold serial dilution starting from 100 μM (Tables 4, 5). In the two enzyme assays,$\mathbf{1_{MPS}}$ displayed similar IC50 values (7.91 μM in AlphaLISA and 9.48 μM in radioactive assay) while $\mathbf{2_{MPS}}$ was 10-fold less effective in the radioactive assay (77.7 μM) respect to the AlphaLISA test (7.22 μM). The relative dose-dependent inhibition curves for the two inhibitors obtained with the both methods are reported in Fig. 9. These values confirmed the 3-D QSAR and COMBINE models as effective tools to design new inhibitors (compare prediction and actual values in Table 4). Again, as for the MTS and CTS, MPS compounds were overpredicted by the 3-D QSAR models and underpredicted by Py-COMBINEr ones, the average predictions being the most accurate.

To check the selectivity of $\mathbf{1_{MPS}}$ and $\mathbf{2_{MPS}}$ towards DOT1L respect to other methylatransferases, the two compounds were further against G9a and SET7/9 (two other histone lysine methyltransferases) PRMT1(an arginine methyltranferase), but it was not possible to determine any IC_{50} value for $\mathbf{1_{MPS}}$ and $\mathbf{2_{MPS}}$ against these enzymes (Table 6).

Binding mode of 1_{MPS} and 2_{MPS} into the DOT1L structure

Inspection of $\mathbf{1_{MPS}}$ and $\mathbf{2_{MPS}}$ Surflex-proposed bound conformations revealed some differences in their binding modes. Although to any trained classical medicinal chemist the two compounds share visible molecular resemblance, although on the basis of calculated molecular fingerprints as implemented in openbabel [60], $\mathbf{1_{MPS}}$ and $\mathbf{2_{MPS}}$ showed Tanimoto coefficient T ranging from 0.38 to 1.00 (Supplementary Material Table S26). The values obtained with fp3 and fp4 methods are based on small number of bits (55 and 307 for fp3 and fp4, respectively) that normally should not be used to describe drug like molecules [61]. Considering that two structures can be considered similar at T > 0.85 [62] the two molecules for 9 out of 10 methods have lower values and therefore can be considered to be of very low molecular similarity in agreement with the performed molecular docking simulations.

Although not superimposed, $\mathbf{1_{MPS}}$ and $\mathbf{2_{MPS}}$ thiobarbituric moieties are both buried into the SAM binding site to fill part of the methionine pocket (Fig. 9), even if to a different extent. In particular, $\mathbf{1_{MPS}}$ sulfur atom points towards a sub-pocket formed by the Asp161–Gln169 loop making a weak hydrogen bond with Val169 NH (S···N distance 3.97 Å), while one of the thiobarbituric carbonyl oxygen is engaged in a hydrogen bond with Asn241 NH (O···N distance 2.93 Å). Differently, the $\mathbf{2_{MPS}}$ thiobarbituric portion is placed in the distal SAM methionine carboxylic group, almost superimposed to the ureidic portion of 4ER5 co-crystallized ligand, with the sulfur atom

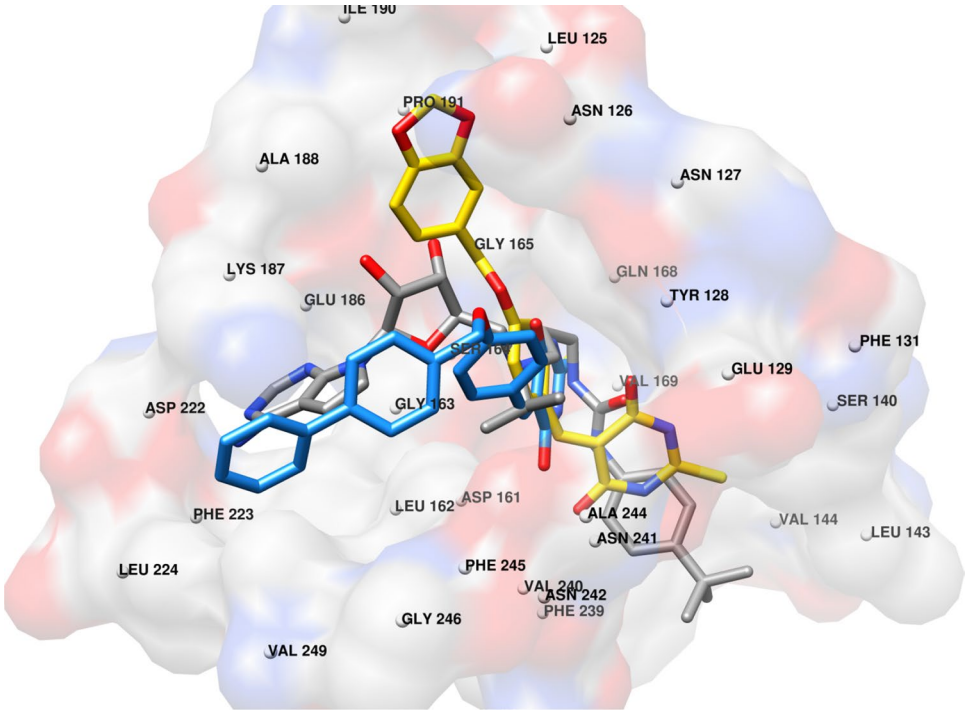


Fig. 9 Binding mode of $\mathbf{1}_{\mathbf{MPS}}$ (carbon atoms colored in blue) and $\mathbf{2}_{\mathbf{MPS}}$ (carbon atoms colored in yellow). For comparison purpose, 4ER5 ligand (carbon atoms colored in grey) and the surface of 4ER5 protein (colored by atom type) are also represented

pointing on the same space occupied by the 4ER5 ligand *tert*-butylphenyl moiety delimitated by Ser140, Leu143 and Val144 main- and side-chains. This position allows one of the thiobarbituric carbonyl oxygen to be involved in a hydrogen bond with Asn241 NH (O···N distance 2.77 Å).

The $\mathbf{1}_{\mathbf{MPS}}$ and $\mathbf{2}_{\mathbf{MPS}}$ central 1,4-disubstituted phenyl rings are shifted of about 3.6 Å and rotated of about 65 degrees so working as to two different bridges connecting the diphenyl ($\mathbf{1}_{\mathbf{MPS}}$) or the methylendioxyphenyl ($\mathbf{2}_{\mathbf{MPS}}$) groups. The distal phenyl group of $\mathbf{1}_{\mathbf{MPS}}$ diphenyl is placed in a sub-pocked formed by that is Lys187, Leu224 and Val249 side-chains. Interestingly, in this area bromine (3SX0, 4ER6 and 4ER7) and iodine atoms (3UWP) were also found, bound to experimentally co-crystallized ligands adenine mimetics. In this scenario, noteworthy the diphenyl moiety fully overlaps the 4WVL acetyllysine-like side-chain. With this binding mode, $\mathbf{1}_{\mathbf{MPS}}$ closely binds in a SAM- or nucleoside DOT1L inhibitor-like mode. Differently, $\mathbf{2}_{\mathbf{MPS}}$ methylendioxyphenyl group is placed in a different sub-pocket (Lys124, Leu125, Asn126, Glu186, Ala188 and Pro191), always free in all co-crystallized complexes, thus highlighting a possible new anchor point to be filled during the design of new non-nucleoside DOT1Li. Binding modes of $\mathbf{1}_{\mathbf{MPS}}$ and $\mathbf{2}_{\mathbf{MPS}}$ were obtained from cross-docking all 25 experimental DOT1L binding sites (Training Set + CTS locks) and are the lowest energy docked conformations of two ensembles of 25 low energy poses (Supplementary Material Fig. S81).

Conclusion

In this study, 3-D QSAR, COMBINE and molecular docking procedures were applied to a series of DOT1L co-crystallized SAM-competitive inhibitors, obtaining models with remarkable statistical results and predictive capabilities. Furthermore, it has been demonstrated that computational technologies such as 3-D QSAR and COMBINE can benefit from accurately conducted docking studies as they crucially rely on how molecular conformations are chosen and overlaid. Through external tests sets the combined methodologies displayed good descriptive and predictive capabilities. Analysis of 3-D QSAR contour plots (AAC, COEFs, MIFs, ACs) together with COMBINE histograms (MRAACs, COEFs, AMRIs, MRIs and MRACs) allowed a deep description of chemical features leading to an overall LB/SB derived pharmacophore model (Fig. 8), potentially useful to design new molecules by maximizing ligand/protein interactions. In a real drug design application, the 24 compounds listed in Table S2 (MTS) would have represented a predictive test set. Arbitrarily selecting the top 20% predicted compounds, two active compounds at nanomolar ($\mathbf{23}_{\mathbf{MTS}}$ and $\mathbf{24}_{\mathbf{TS}}$) ranges would have been disclosed. Only one false positive ($\mathbf{3}_{\mathbf{MTS}}$) would have resulted thus leading to a success rate of 80%, a much higher value than those reported in SB virtual screening applications [63]. This predictive trend was also observed in the case of the CTS, an experimentally aligned derived test set. As pointed above, compound displaying bound conformations comprised in the training set space the 3-D QSAR and COMBINE models

showed to perfectly predict the experimental potency (check experimental and predicted p*K*is for 4ER3 and 5JUW in Supplementary Material Table S21). On the other hand, greater errors of prediction were found associated to those complexes containing bound inhibitors displaying some differences in either structures or binding modes (see values in Supplementary Material Table S21 for 4WVL, 5DRT, 5DRY, 5DSX, 5DT2, 5DTM, 5DTQ and 5DTR).

Finally, the models' predictive ability was experimentally evaluated through their application to an internal library of 87 available compounds with unknown DOT1L inhibition aptitude. The model indicated compounds $\mathbf{1}_{\mathbf{MPS}}$ and $\mathbf{2}_{\mathbf{MPS}}$ as promising DOT1L inhibitors. Enzymatic assay confirmed their inhibitory potency at low micromolar level (Tables 4, 5). Interestingly, the models revealed that $\mathbf{2}_{\mathbf{MPS}}$ binds DOT1L differently from all the other ligands (training set, MTS and CTS), and nonetheless the low IC_{50} it can be considered a potential hit from which draw out some ideas to design new DOT1L inhibitors. This additional testing fully qualified the models as effective tools for the design of DOT1L inhibitors. Application of the models to wider chemical libraries for virtual screenings is in due course.

Experimental section

Computational procedures

All calculations were done on a 6 blades (8 Intel-Xeon E5520 2.27 GHz CPU and 24 GB DDR3 RAM each) cluster (48 CPU total) running Debian GNU/Linux "Jessie" 8.5 64 bit operating system.

Training set preparation

The 15 DOT1L/inhibitors complexes (Table 1) were submitted to a similar previously reported [56] molecular modeling protocol. At the time we started this study only SAM competitive compounds co-crystallized with DOT1L were available. During the analysis of the results and manuscript preparation other complexes were released. Nevertheless, no new SAM competitive co-cristallized inhibitor were present at the time of submission. The complexes were loaded through UCSF Chimera v1.10.1 and visually inspected. Some of these complexes showed missing residues, therefore the incomplete complexes were first processed with Modeller [64], to fill the gaps. The completed DOT1L complexes were SB aligned using the alpha carbon atoms (UCSF Chimera MatchMaker [65] module) using **1NW3** as template (the one with the most complete structure and best resolution combination, 416 residues, 2.50 Å resolution). After hydrogen addition (UCSF Chimera addh command) the complexes were geometrically optimized by means of GROMACS 5.0.5 in a single point minimizarion. The AMBER99 sb-ildn force filed was used to create the protein topology, then the complexes were collocated in a cubic box and solvated with water molecules (TIP4P models) and neutralized with either Na^+ or Cl^- ions. For the single point minimization was used the steepest descendent algorithm with maximum 50,000 steps. The minimized complex were SB realigned (**1NW3** as template) and separated into proteins (locks) and ligands (keys), the latter were directly used to compose the SB aligned training set for the subsequent 3-D QSAR and COMBINEr models.

External test set selection and preparation

All of the MTS and JCTS compounds were modeled beginning from the co-crystalized inhibitors though UCSF Chimera (Build structures module). MPS compounds were modeled starting from SMILES format and treated with openbabel [66] to generate a randomized initial conformation. The new DOT1L complexes were subjected to the same procedure of the training set described above.

3-D QSAR modeling

A python implementation of 3-D QSAutogrid/R procedure [46] was used. All details will be given elsewhere. Briefly the autogrid module of the latest autodock suites [67, 68] was used to calculate the molecular interaction fields (MIFs). All statistical analysis and validations were implemented by means of the scikit-learn python module [69].

For the preliminary 3-D QSAR models, MIFs for each probe were computed using a grid spacing of 1 Å. The xyz coordinates (in angstroms) of the cuboid grid box used for the computation were Xmin/Xmax = − 10.979/11.021, Ymin/Ymax = 12.857/50.857, Zmin/Zmax = 6.523/28.523 to embrace all the minimized inhibitors spanning 10 Å in all three dimensions. Data for graphical analysis were stored in cube format and read in UCSF Chimera.

COMBINE modeling

A python implementation of a previously reported COMBINEr procedure [48, 70] was used. All details will be given elsewhere. Briefly AutoDockTools4 [68] python code was imported in python and modified to calculate the per-residue ligand–protein interactions. All statistical analysis and validations were implemented by means of the scikit-learn [69] python library. Data for model analysis were stored directly in xlsx format to be read in Microsoft Excel.

Autodock Vina settings

Intermediary steps, such as pdbqt files for protein and ligand preparation, were completed using different AutoDock Tools (ADT) Scripts. The grid size was expanded 10 Å beyond any external ligand atoms with grid spacing of 0.375 Å and centered at the mean molecules' center of mass. For each calculation, twenty poses were obtained and ranked according to the scoring-functions.

Plants settings

The docking of the target protein with the ligand was performed using Plants v1.2 version with three different scoring functions at default speed (SPEED1). The docking tools generated 20 conformations for each docked ligand. The docking binding site was centered at the molecules' mean center and enlarged to a radius of 15 Å. Docking was performed using three different scoring functions: Chemplp, Plp and Plp95.

Surflex-dock settings

Version 2.6 of the program was used; the input file was built using the mol2 prepared protein structure. The protomol was generated using all the ligands structures with a threshold of 0.50 and bloat set to 0 (default settings). Ligand were prepared as described above and docked as mol2 files.

Ligand random conformation generation

SMILES formatted molecules were treated with openbabel [71] to generate unbiased and mass centered to zero Cartesian coordinates. Starting from SMILES formatted molecules, openbabel was used to generate unbiased and mass centred to zero Cartesian coordinates.

Synthetic procedure for the preparation of compounds 1_{MPS} and 2_{MPS}

Chemistry

Melting points were determined on a Buchi 530 melting point apparatus. ^{1}H NMR spectra were recorded at 400 MHz using a Bruker AC 400 spectrometer; chemical shifts are reported in δ (ppm) units relative to the internal reference tetramethylsilane (Me_4Si). Mass spectra were recorded on a API-TOF Mariner by Perspective Biosystem (Stratford, Texas, USA), samples were injected by an Harvard pump using a flow rate of 5–10 µL/min, infused in the Electrospray system. All compounds were routinely checked by TLC and ^{1}H-NMR. TLC was performed on aluminum-backed silica gel plates (Merck DC, Alufolien Kieselgel 60 F254) with spots visualized by UV light or using a $KMnO_4$ alkaline solution. All solvents were reagent grade and, when necessary, were purified and dried by standard methods. Concentration of solutions after reactions and extractions involved the use of a rotary evaporator operating at reduced pressure of ~ 20 Torr. Organic solutions were dried over anhydrous sodium sulfate. Elemental analysis has been used to determine purity of the described compounds, that is > 95%. Analytical results are within 0.40% of the theoretical values. All chemicals were purchased from Sigma Aldrich srl, Milan (Italy) or from TCI Europe NV, Zwijndrecht (Belgium), and were of the highest purity. As a rule, samples prepared for physical and biological studies were dried in high vacuum over P_2O_5 for 20 h at temperatures ranging from 25 to 40 °C, depending on the sample melting point.

Preparation of 4-([1,1′-biphenyl]-4-ylmethoxy)benzaldehyde (3_{MPS})

The 4-hydroxybenzaldehyde (500 mg, 4.09 mmol), 4-(chloromethyl)-1,1′-biphenyl (1.24 g, 6.14 mmol), K_2CO_3 (848.8 mg, 6.14 mmol), NaI (675.1 mg, 4.5 mmol) and dry CH_3CN (20 mL) were mixed in a round bottom flask and stirred at reflux for 4 h. After the completion of the reaction, the solvent was evaporated, the residue quenched with water (30 mL) and extracted first with AcOEt (4 × 50 mL) and then with a mixture $CHCl_3$/*i*PrOH 4:1 (1 × 30 mL). The combined organic layers were washed with brine, dried over anhydrous Na_2SO_4, filtered, and evaporated under reduced pressure to give a crude that was purified by silica gel column chromatography eluting with a mixture ethyl acetate/*n*-hexane 1:13 to give the expected product as a white powder. Yield: 77%. mp: 161–163 °C. ^{1}H-NMR (DMSO): δ 5.29 (s, 2H, OC*H*$_2$), 7.23–7.25 (d, 2H, C*H* aromatic ring), 7.38–7.39 (t, 1H, C*H* aromatic ring), 7.45–7.49 (t, 2H, C*H* aromatic ring), 7.55–7.57 (d, 2H, C*H* aromatic ring), 7.67–7.71 (t, 4H, C*H* aromatic ring), 7.88–7.90 (d, 2H, C*H* aromatic ring), 9.88 (s, 1H, C*H*O). MS-ESI m/z: 289 $[M+H]^+$.

Preparation of 4-(benzo[*d*] [1,3] dioxol-5-ylmethoxy)benzaldehyde (4_{MPS})

To a mixture of commercially available benzo[*d*] [1,3] dioxol-5-ylmethanol (292 mg, 1.92 mmol) and 4-hydroxybenzaldehyde (234 mg, 1.92 mmol) in dry THF (6.5 mL) were added in sequence while cooling at 0–5 °C under a nitrogen atmosphere PPh_3 (629 mg, 2.4 mmol) and, dropwise, a DIAD solution (485 mg, 0.47 mL, 2.4 mmol) in dry THF (4 mL). The resulting mixture was then allowed to warm to room temperature and stir overnight. After the completion of the reaction the solvent was evaporated and the resulting crude purified by silica gel column chromatography eluting with a mixture ethyl acetate/petroleum ether

1:12, to afford the pure $\mathbf{4_{MPS}}$ as a white powder. Yield: 47%. mp: 61–63 °C. ^{1}H-NMR (DMSO): δ 5.11 (s, 2H, OCH_2-benzodioxole ring), 5.95 (s, 2H, OCH_2O), 6.91–6.99 (m, 2H, C*H* benzodioxole ring), 7.04 (s, 1H, C*H* benzodioxole ring), 7.18–7.20 (d, 2H, C*H* benzene ring), 7.86–7.88 (d, 2H, C*H* benzene ring), 9.86 (s, 1H, C*H*O). MS-ESI m/z: 257 $[M+H]^+$.

General procedure for the preparation of 5-(benzylidene)-2-thioxodihydropyrimidine-4,6(1H,5H)-diones $\mathbf{1_{MPS}}$ and $\mathbf{2_{MPS}}$. Example: 5-(4-(benzo[d] [1,3] dioxol-5-ylmethoxy)benzylidene)-2-thioxodihydropyrimidine-4,6(1H,5H)-dione ($\mathbf{2_{MPS}}$)

4-(Benzo[*d*] [1,3] dioxol-5-ylmethoxy)benzaldehyde $\mathbf{4_{MPS}}$ (112 mg, 0.44 mmol) and 2-thiobarbituric acid (66 mg, 0.46 mmol) were dissolved in EtOH (15 mL) and then refluxed for 6 h 30 min. When TLC showed the disappearance of the starting materials, the yellow solid in suspension was filtered from hot EtOH and then washed in sequence over filter with hot EtOH and Et_2O, to give the pure $\mathbf{2_{MPS}}$ as a yellow powder. Yield: 61%. mp: 248–249 °C. ^{1}H-NMR (DMSO): δ 5.14 (s, 2H, OCH_2-benzodioxole ring), 5.95 (s, 2H, OCH_2O), 6.92–6.98 (m, 2H, C*H* benzodioxole ring), 7.01 (s, 1H, C*H* benzodioxole ring), 7.06–7.10 (d, 2H, C*H* benzene ring), 8.26 (s, 1H, =C*H*), 8.33–8.34 (d, 2H, C*H* benzene ring), 12.38 (s, 1H, N*H*), 12.40 (s, 1H, N*H*). MS-ESI m/z: 383 $[M+H]^+$.

$\mathbf{1_{MPS}}$

Yield: 89%. mp: > 299 °C. ^{1}H-NMR (DMSO): δ 5.40 (s, 2H, OCH_2), 6.83 (d, 2H, C*H* aromatic ring), 7.33–7.58 (m, 5H, C*H* aromatic ring), 7.62–7.90 (d, 4H, C*H* aromatic ring), 8.27 (s, 1H, =C*H*), 8.31–8.47 (d, 2H, C*H* aromatic ring), 12.31 (s, 1H, N*H*), 12.38 (s, 1H, N*H*). MS-ESI m/z: 415 $[M+H]^+$.

Enzyme assays

AlphaLISA method DOT1L inhibition assay

Histone H3 lysine-*N*-methyltransferase assay was performed in 384-well plates (Corning, # 3673) with human recombinant DOT1L (residues 1-416) containing an *N*-terminal GST-tag (Reaction Biology, # HMT-11-101). Purified oligonucleosomes (ON) from HeLa cells (Reaction Biology, # HMT-35-130) were used as substrate and reactions were performed in assay buffer (AB: 50 mM Tris–HCl pH 8.0, 150 mM NaCl, 3 mM $MgCl_2$, 0.1% BSA). DOT1L enzyme, SAM at 4 µM (0.5 µM final), inhibitors and oligonucleosomes at 0.4 ng/µL (0.5 ng/well final) were diluted in the assay buffer just before use. The assay reactions were in a final volume of 10 µL including 5 µL of diluted compounds in AB, 2.5 µL of diluted DOT1L and 2.5 µL of SAM/ON mix. Control reactions were also performed with or without enzyme and without compounds (replaced by 5 µL of DMSO-AB mix). Reactions were run at room temperature taking into consideration the incubation time for a complete methylation. Reactions were then stopped by adding 5 µL of high salt buffer (50 mM Tris–HCl pH 7.4, 1 M NaCl, 0.1% Tween-20, 0.3% poly-L-lysine) and incubated at room temperature for 15 min. A mix of anti-Histone H3 (C-ter) AlphaLISA acceptor beads (PerkinElmer, #AL147) (0.1 mg/mL final) and AlphaLISA biotinylated anti-dimethyl-Histone H3 Lysine 79 (H3K79) antibody (PerkinElmer, #AL148) (5 nM final) in detection buffer (DB) (AlphaLISA 5X Epigenetics Buffer 1 + AlphaLISA 30X Epigenetics Buffer Supplement, PerkinElmer, #AL008C1 & #AL008C2) were prepared and 5 µL of this mix were added for 1 h incubation at room temperature. For detection 5 µL Alpha Streptavidin Donor beads (PerkinElmer, #6760002) (0.1 mg/mL final) in DB was added and incubated for 30 min at room temperature. Finally, the plates were read using an EnVision 2103 multilabel plate reader (PerkinElmer) in AlphaLISA mode. Each point/concentration of compounds was evaluated in triplicates per assay and the percentage of inhibition was calculated as the mean of at least three experiments. The percentage inhibition was calculated using the following equation:

$$\%\ \text{of inhibition} = \left(1 - \left(\frac{Xi - Xm}{XM - Xm}\right)\right) \times 100;$$

where Xi, Xm, XM are the average signal at the considered concentration, minimal signal response (without enzyme and compound) and maximum signal response (without compound), respectively. Data analysis was performed using the GraphPad Prism 5 software. IC_{50} values were determined using the nonlinear regression fittings with sigmoidal dose–response (variable slope) function and the displayed EC_{50} are the mean of three independent experiments with associated standard deviations.

DOT1L, G9a, SET7/9, and PRMT1 inhibition radioactive assays

The appropriate histone methyltransferase (HMT) substrate [0.05 mg/mL oligonucleosomes for DOT1L, 5 µM histone H3 (1–21) peptide for G9a, 0.05 mg/mL core histone for SET7/9, and 5 µM histone H4 for PRMT1] was added in freshly prepared reaction buffer (50 mM Tris-HCl (pH 8.5), 5 mM $MgCl_2$, 50 mM NaCl, 0.01% Brij35, 1 mM DTT, 1% DMSO). The HMT enzyme [human recombinant DOT1L (residues 1-416; Genbank Accession No. NM_032482), MW = 80.0 kDa, expressed as an N-terminal GST fusion protein in E. coli, 50 nM in the reaction; human

G9a (GenBank Accession No. NM_006709) (amino-acids 785–1210), with *N*-terminal GST tag, MW = 79.6 kDa, expressed in an *E. coli* expression system, 10 nM in the reaction; human SET7/9 (SETD7, GenBank Accession No. NM_030648) (amino-acids 2-366), with *N*-terminal GST tag and C-terminal His tag, MW = 68.5 kDa, expressed in an *E. coli* expression system, 8 nM in the reaction;r human PRMT1 (GenBank Accession No. NM_001536) (amino-acids 2-371), with *N*-terminal GST tag, MW = 68.3 kDa, expressed in an *E. coli* expression system, 20 nM in the reaction] was delivered into the substrate solution and the mixture was mixed gently. Afterwards, the tested compounds dissolved in DMSO were delivered into the enzyme/substrate reaction mixture by using Acoustic Technology (Echo 550, LabCyte Inc. Sunnyvale, CA) in nanoliter range, and 1 μM ^{3}H-SAM was also added into the reaction mixture to initiate the reaction. The reaction mixture was incubated for 1 h at 30 °C and then it was delivered to filter-paper for detection. The data were analysed using Excel 2016 and GraphPad Prism ver. 6 softwares for IC_{50} curve fits.

Acknowledgements M.S. acknowledges FIRB Grant RBFR10ZJQT for a 1 year fellowship and Sapienza University for "Progetto di Avvio alla Ricerca". This work was supported by grants from PRIN 2016 (prot. 20152TE5PK) (A.M.), AIRC 2016 (n. 19162) (A.M.), AIRC Fondazione Cariplo TRIDEO Id. 17515 (D.R.), NIH (n. R01GM114306) (A.M.), Italian Ministry of Health Grant PE-2013-02355271 (A.M.) and Progetti di Ricerca di Università (Prot. C26A15988X) (R.R.). Many thanks to the reviewers that with their suggestions helped to improve the manuscript quality.

References

1. Copeland RA, Solomon ME, Richon VM (2009) Protein methyltransferases as a target class for drug discovery. Nat Rev Drug Discov 8(9):724–732. https://doi.org/10.1038/nrd2974
2. Bannister AJ, Kouzarides T (2005) Reversing histone methylation. Nature 436(7054):1103–1106. https://doi.org/10.1038/nature04048
3. Lachner M, O'Sullivan RJ, Jenuwein T (2003) An epigenetic road map for histone lysine methylation. J Cell Sci 116(Pt 11):2117–2124. https://doi.org/10.1242/jcs.00493
4. Margueron R, Trojer P, Reinberg D (2005) The key to development: interpreting the histone code? Curr Opin Genet Dev 15(2):163–176
5. Allis CD, Berger SL, Cote J, Dent S, Jenuwien T, Kouzarides T, Pillus L, Reinberg D, Shi Y, Shiekhattar R, Shilatifard A, Workman J, Zhang Y (2007) New nomenclature for chromatin-modifying enzymes. Cell 131(4):633–636. https://doi.org/10.1016/j.cell.2007.10.039
6. Nguyen AT, Zhang Y (2011) The diverse functions of Dot1 and H3K79 methylation. Genes Dev 25(13):1345–1358. https://doi.org/10.1101/gad.2057811
7. Illi B, Scopece A, Nanni S, Farsetti A, Morgante L, Biglioli P, Capogrossi MC, Gaetano C (2005) Epigenetic histone modification and cardiovascular lineage programming in mouse embryonic stem cells exposed to laminar shear stress. Circ Res 96(5):501–508. https://doi.org/10.1161/01.RES.0000159181.06379.63
8. Jones B, Su H, Bhat A, Lei H, Bajko J, Hevi S, Baltus GA, Kadam S, Zhai H, Valdez R, Gonzalo S, Zhang Y, Li E, Chen T (2008) The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure. PLoS Genet 4(9):e1000190. https://doi.org/10.1371/journal.pgen.1000190
9. Kimura A (2008) Molecular etiology and pathogenesis of hereditary cardiomyopathy. Circ J 72(Suppl A):A38–A48
10. Nguyen AT, Xiao B, Neppl RL, Kallin EM, Li J, Chen T, Wang DZ, Xiao X, Zhang Y (2011) DOT1L regulates dystrophin expression and is critical for cardiac function. Genes Dev 25(3):263–274. https://doi.org/10.1101/gad.2018511
11. Feng Y, Yang Y, Ortega MM, Copeland JN, Zhang M, Jacob JB, Fields TA, Vivian JL, Fields PE (2010) Early mammalian erythropoiesis requires the Dot1L methyltransferase. Blood 116(22):4483–4491. https://doi.org/10.1182/blood-2010-03-276501
12. Im H, Park C, Feng Q, Johnson KD, Kiekhaefer CM, Choi K, Zhang Y, Bresnick EH (2003) Dynamic regulation of histone H3 methylated at lysine 79 within a tissue-specific chromatin domain. J Biol Chem 278(20):18346–18352. https://doi.org/10.1074/jbc.M300890200
13. Slany RK (2016) The molecular mechanics of mixed lineage leukemia. Oncogene. https://doi.org/10.1038/onc.2016.30
14. Butler LH, Slany R, Cui X, Cleary ML, Mason DY (1997) The HRX proto-oncogene product is widely expressed in human tissues and localizes to nuclear structures. Blood 89(9):3361–3370
15. Gu Y, Nakamura T, Alder H, Prasad R, Canaani O, Cimino G, Croce CM, Canaani E (1992) The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to Drosophila trithorax, to the AF-4 gene. Cell 71(4):701–708
16. Tkachuk DC, Kohler S, Cleary ML (1992) Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocations in acute leukemias. Cell 71(4):691–700
17. Okada Y, Feng Q, Lin Y, Jiang Q, Li Y, Coffield VM, Su L, Xu G, Zhang Y (2005) hDOT1L links histone methylation to leukemogenesis. Cell 121(2):167–178. https://doi.org/10.1016/j.cell.2005.02.020
18. Daigle SR, Olhava EJ, Therkelsen CA, Basavapathruni A, Jin L, Boriack-Sjodin PA, Allain CJ, Klaus CR, Raimondi A, Scott MP, Waters NJ, Chesworth R, Moyer MP, Copeland RA, Richon VM, Pollock RM (2013) Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. Blood 122(6):1017–1025. https://doi.org/10.1182/blood-2013-04-497644
19. Bitoun E, Oliver PL, Davies KE (2007) The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling. Hum Mol Genet 16(1):92–106. https://doi.org/10.1093/hmg/ddl444
20. Bernt KM, Armstrong SA (2011) Targeting epigenetic programs in MLL-rearranged leukemias. Hematol Am Soc Hematol Educ Program 2011:354–360. https://doi.org/10.1182/asheducation-2011.1.354
21. Yao Y, Chen P, Diao J, Cheng G, Deng L, Anglin JL, Prasad BV, Song Y (2011) Selective inhibitors of histone methyltransferase DOT1L: design, synthesis, and crystallographic studies. J Am Chem Soc 133(42):16746–16749. https://doi.org/10.1021/ja206312b
22. Yu W, Smil D, Li F, Tempel W, Fedorov O, Nguyen KT, Bolshan Y, Al-Awar R, Knapp S, Arrowsmith CH, Vedadi M, Brown PJ, Schapira M (2013) Bromo-deaza-SAH: a potent and selective DOT1L inhibitor. Bioorg Med Chem 21(7):1787–1794. https://doi.org/10.1016/j.bmc.2013.01.049
23. Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, Kruger FA, Light Y, Mak L, McGlinchey S, Nowotka M, Papadatos G, Santos R, Overington JP (2014) The ChEMBL

bioactivity database: an update. Nucleic Acids Res 42(Database issue):D1083–D1090. https://doi.org/10.1093/nar/gkt1031
24. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP (2012) ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 40(Database issue):D1100–D1107. https://doi.org/10.1093/nar/gkr777
25. Song Y, Lisheng DE, Yao Y, Zhang L, Wang C, Redell MS, Shuo DO (2016) Non-ribose containing inhibitors of histone methyltransferase dot1l for cancer treatment. Google Patents
26. Song Y, Chen P, DIAO J, Cheng G, DENG L, ANGLIN JL, PRASAD VBV, Yao Y (2014) Selective inhibitors of histone methyltransferase dot1l. Google Patents
27. Bradner J (2015) Dot1l probes. Google Patents
28. Chen C, Zhu H, Stauffer F, Caravatti G, Vollmer S, Machauer R, Holzer P, Möbitz H, Scheufler C, Klumpp M, Tiedt R, Beyer KS, Calkins K, Guthy D, Kiffe M, Zhang J, Gaul C (2016) Discovery of novel Dot1L Inhibitors through a structure-based fragmentation approach. ACS Med Chem Lett 7(8):735–740. https://doi.org/10.1021/acsmedchemlett.6b00167
29. Chen S, Li L, Chen Y, Hu J, Liu J, Liu YC, Liu R, Zhang Y, Meng F, Zhu K, Lu J, Zheng M, Chen K, Zhang J, Jiang H, Yao Z, Luo C (2016) Identification of novel disruptor of telomeric silencing 1-like (DOT1L) inhibitors through structure-based virtual screening and biological assays. J Chem Inf Model 56(3):527–534. https://doi.org/10.1021/acs.jcim.5b00738
30. Luo M, Wang H, Zou Y, Zhang S, Xiao J, Jiang G, Zhang Y, Lai Y (2016) Identification of phenoxyacetamide derivatives as novel DOT1L inhibitors via docking screening and molecular dynamics simulation. J Mol Graph Model 68:128–139. https://doi.org/10.1016/j.jmgm.2016.06.011
31. Scheufler C, Möbitz H, Gaul C, Ragot C, Be C, Fernández C, Beyer KS, Tiedt R, Stauffer F (2016) Optimization of a fragment-based screening hit toward potent DOT1L inhibitors interacting in an induced binding pocket. ACS Med Chem Lett 7(8):730–734. https://doi.org/10.1021/acsmedchemlett.6b00168
32. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28(1):235–242
33. Anglin JL, Song Y (2013) A medicinal chemistry perspective for targeting histone H3 lysine-79 methyltransferase DOT1L. J Med Chem 56(22):8972–8983. https://doi.org/10.1021/jm4007752
34. Meng F, Cheng S, Ding H, Liu S, Liu Y, Zhu K, Chen S, Lu J, Xie Y, Li L, Liu R, Shi Z, Zhou Y, Liu YC, Zheng M, Jiang H, Lu W, Liu H, Luo C (2015) Discovery and optimization of novel, selective histone methyltransferase SET7 inhibitors by pharmacophore- and docking-based virtual screening. J Med Chem 58(20):8166–8181. https://doi.org/10.1021/acs.jmedchem.5b01154
35. Raj U, Kumar H, Gupta S, Varadwaj PK (2015) Novel DOT1L receptornatural inhibitors involved in mixed lineage leukemia: a virtual screening, molecular docking and dynamics simulation study. Asian Pac J Cancer Prev 16(9):3817–3825. https://doi.org/10.7314/APJCP.2015.16.9.3817
36. Kouranov A, Xie L, de la Cruz J, Chen L, Westbrook J, Bourne PE, Berman HM (2006) The RCSB PDB information portal for structural genomics. Nucleic Acids Res 34(Database issue):D302–D305. https://doi.org/10.1093/nar/gkj120
37. Ragno R, Ballante F, Pirolli A, Wickersham RB 3rd, Patsilinakos A, Hesse S, Perspicace E, Kirsch G (2015) Vascular endothelial growth factor receptor-2 (VEGFR-2) inhibitors: development and validation of predictive 3-D QSAR models through extensive ligand- and structure-based approaches. J Comput Aided Mol Des 29(8):757–776. https://doi.org/10.1007/s10822-015-9859-y
38. Basavapathruni A, Jin L, Daigle SR, Majer CR, Therkelsen CA, Wigle TJ, Kuntz KW, Chesworth R, Pollock RM, Scott MP, Moyer MP, Richon VM, Copeland RA, Olhava EJ (2012) Conformational adaptation drives potent, selective and durable inhibition of the human protein methyltransferase DOT1L. Chem Biol Drug Des 80(6):971–980. https://doi.org/10.1111/cbdd.12050
39. Yao Y, Chen P, Diao J, Cheng G, Deng L, Anglin JL, Prasad BVV, Song Y (2011) Selective inhibitors of histone methyltransferase DOT1L: design, synthesis, and crystallographic studies. J Am Chem Soc 133(42):16746–16749. https://doi.org/10.1021/ja206312b
40. Yu W, Chory EJ, Wernimont AK, Tempel W, Scopton A, Federation A, Marineau JJ, Qi J, Barsyte-Lovejoy D, Yi J, Marcellus R, Iacob RE, Engen JR, Griffin C, Aman A, Wienholds E, Li F, Pineda J, Estiu G, Shatseva T, Hajian T, Al-Awar R, Dick JE, Vedadi M, Brown PJ, Arrowsmith CH, Bradner JE, Schapira M (2012) Catalytic site remodelling of the DOT1L methyltransferase by selective inhibitors. Nat Commun 3:1288. https://doi.org/10.1038/ncomms2304
41. Richon VM, Johnston D, Sneeringer CJ, Jin L, Majer CR, Elliston K, Jerva LF, Scott MP, Copeland RA (2011) Chemogenetic analysis of human protein methyltransferases. Chem Biol Drug Des 78(2):199–210. https://doi.org/10.1111/j.1747-0285.2011.01135.x
42. Min J, Feng Q, Li Z, Zhang Y, Xu RM (2003) Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase. Cell 112(5):711–723
43. Daigle SR, Olhava EJ, Therkelsen CA, Basavapathruni A, Jin L, Boriack-Sjodin PA, Allain CJ, Klaus CR, Raimondi A, Scott MP, Waters NJ, Chesworth R, Moyer MP, Copeland RA, Richon VM, Pollock RM (2013) Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. Blood 122(6):1017
44. Hu H, Luo C, Zheng YG (2016) Transient kinetics define a complete kinetic model for protein arginine methyltransferase 1. J Biol Chem 291(52):26722–26738. https://doi.org/10.1074/jbc.M116.757625
45. Cheng Y, Prusoff WH (1973) 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(23):3099–3108
46. Ballante F, Ragno R (2012) 3-D QSAutogrid/R: an alternative procedure to build 3-D QSAR models. Methodologies and applications. J Chem Inf Model 52(6):1674–1685. https://doi.org/10.1021/ci300123x
47. Verma J, Khedkar VM, Coutinho EC (2010) 3D-QSAR in drug design—a review. Curr Top Med Chem 10(1):95–115
48. Ballante F, Musmuca I, Marshall GR, Ragno R (2012) Comprehensive model of wild-type and mutant HIV-1 reverse transciptases. J Comput Aided Mol Des 26(8):907–919. https://doi.org/10.1007/s10822-012-9586-6
49. Melo-Filho CC, Braga RC, Andrade CH (2014) 3D-QSAR approaches in drug design: perspectives to generate reliable CoMFA models. Curr Comput Aided Drug Des 10(2):148–159
50. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31(2):455–461. https://doi.org/10.1002/jcc.21334
51. Cleves AE, Jain AN (2015) Knowledge-guided docking: accurate prospective prediction of bound configurations of novel ligands using Surflex-Dock. J Comput Aided Mol Des 29(6):485–509. https://doi.org/10.1007/s10822-015-9846-3
52. Korb O, Stutzle T, Exner TE (2009) Empirical scoring functions for advanced protein-ligand docking with PLANTS. J Chem Inf Model 49(1):84–96. https://doi.org/10.1021/ci800298z
53. Yuriev E, Ramsland PA (2013) Latest developments in molecular docking: 2010–2011 in review. J Mol Recognit 26(5):215–239. https://doi.org/10.1002/jmr.2266
54. Caroli A, Ballante F, Wickersham RB 3rd, Corelli F, Ragno R (2014) Hsp90 inhibitors, part 2: combining ligand-based and structure-based approaches for virtual screening application.

J Chem Inf Model 54(3):970–977. https://doi.org/10.1021/ci400760a
55. Consalvi S, Alfonso S, Di Capua A, Poce G, Pirolli A, Sabatino M, Ragno R, Anzini M, Sartini S, La Motta C, Di Cesare Mannelli L, Ghelardini C, Biava M (2015) Synthesis, biological evaluation and docking analysis of a new series of methylsulfonyl and sulfamoyl acetamides and ethyl acetates as potent COX-2 inhibitors. Bioorg Med Chem 23(4):810–820. https://doi.org/10.1016/j.bmc.2014.12.041
56. Musmuca I, Caroli A, Mai A, Kaushik-Basu N, Arora P, Ragno R (2010) Combining 3-D quantitative structure-activity relationship with ligand based and structure based alignment procedures for in silico screening of new hepatitis C virus NS5B polymerase inhibitors. J Chem Inf Model 50(4):662–676. https://doi.org/10.1021/ci9004749
57. Ballante F, Caroli A, Wickersham RB 3rd, Ragno R (2014) Hsp90 inhibitors, part 1: definition of 3-D QSAutogrid/R models as a tool for virtual screening. J Chem Inf Model 54(3):956–969. https://doi.org/10.1021/ci400759t
58. Wang Y, Li L, Zhang B, Xing J, Chen S, Wan W, Song Y, Jiang H, Jiang H, Luo C, Zheng M (2017) Discovery of novel disruptor of silencing telomeric 1-like (DOT1L) Inhibitors using a target-specific scoring function for the *S*-adenosyl-L-methionine (SAM)-dependent methyltransferase family. J Med Chem. https://doi.org/10.1021/acs.jmedchem.6b01785
59. Lozano JJ, Pastor M, Cruciani G, Gaedt K, Centeno NB, Gago F, Sanz F (2000) 3D-QSAR methods on the basis of ligand-receptor complexes. Application of COMBINE and GRID/GOLPE methodologies to a series of CYP1A2 ligands. J Comput Aided Mol Des 14(4):341–353
60. O'Boyle NM, Morley C, Hutchison GR (2008) Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit. Chem Cent J 2:5. https://doi.org/10.1186/1752-153X-2-5
61. O'Boyle NM, Sayle RA (2016) Comparing structural fingerprints using a literature-based similarity benchmark. J Cheminform 8(1):36. https://doi.org/10.1186/s13321-016-0148-0
62. Maggiora G, Vogt M, Stumpfe D, Bajorath J (2014) Molecular similarity in medicinal chemistry. J Med Chem 57(8):3186–3204. https://doi.org/10.1021/jm401411z
63. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 3(11):935–949. https://doi.org/10.1038/nrd1549
64. Fiser A, Sali A (2003) Modeller: generation and refinement of homology-based protein structure models. Methods Enzymol 374:461–491. https://doi.org/10.1016/S0076-6879(03)74020-8
65. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612. https://doi.org/10.1002/jcc.20084
66. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) Open Babel: an open chemical toolbox. J Cheminform 3:33–33. https://doi.org/10.1186/1758-2946-3-33
67. Morris GM, Huey R, Olson AJ (2008) Using AutoDock for ligand-receptor docking. Curr Protoc Bioinformatics. https://doi.org/10.1002/0471250953.bi0814s24
68. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 30(16):2785–2791. https://doi.org/10.1002/jcc.21256
69. Pedregosa F, Gaël V, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830
70. Silvestri L, Ballante F, Mai A, Marshall GR, Ragno R (2012) Histone deacetylase inhibitors: structure-based modeling and isoform-selectivity prediction. J Chem Inf Model 52(8):2215–2235. https://doi.org/10.1021/ci300160y
71. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) Open Babel: an open chemical toolbox. J Cheminform 3:33. https://doi.org/10.1186/1758-2946-3-33

Affiliations

Manuela Sabatino[1,2] · Dante Rotili[2] · Alexandros Patsilinakos[1,2,3] · Mariantonietta Forgione[2] · Daniela Tomaselli[2] · Fréderic Alby[4] · Paola B. Arimondo[5,6] · Antonello Mai[2] · Rino Ragno[1,2,3]

✉ Antonello Mai
antonello.mai@uniroma1.it

✉ Rino Ragno
rino.ragno@uniroma1.it

[1] Rome Center for Molecular Design, Dipartimento di Chimica e Tecnologie del Farmaco, Sapienza Università di Roma, P. le A. Moro 5, 00185 Roma, Italy

[2] Dipartimento di Chimica e Tecnologie del Farmaco, Istituto Pasteur - Fondazione Cenci Bolognetti, Sapienza Università di Roma, P. le A. Moro 5, 00185 Roma, Italy

[3] Alchemical Dynamics s.r.l., 00125 Roma, Italy

[4] Laboratoire Pierre Fabre, 3 Avenue H. Curien, 31100 Toulouse, France

[5] CNRS FRE3600 ETaC, Bât IBCG, Toulouse, France

[6] Oversea Fellow, Churchill College, Cambridge CB3 0DS, UK