

## RESEARCH ARTICLE

# Novel DOT1L Receptor Natural Inhibitors Involved in Mixed Lineage Leukemia: a Virtual Screening, Molecular Docking and Dynamics Simulation Study

Utkarsh Raj, Himansu Kumar, Saurabh Gupta, Pritish Kumar Varadwaj\*

### Abstract

**Background:** The human protein methyl-transferase DOT1L catalyzes the methylation of histone H3 on lysine 79 (H3K79) at homeobox genes and is also involved in a number of significant processes ranging from gene expression to DNA-damage response and cell cycle progression. Inhibition of DOT1L activity by shRNA or small-molecule inhibitors has been established to prevent proliferation of various MLL-rearranged leukemia cells *in vitro*, establishing DOT1L an attractive therapeutic target for mixed lineage leukemia (MLL). Most of the drugs currently in use for the MLL treatment are reported to have low efficacy, hence this study focused on various natural compounds which exhibit minimal toxic effects and high efficacy for the target receptor. **Materials and Methods:** Structures of human protein methyl-transferase DOT1L and natural compound databases were downloaded from various sources. Virtual screening, molecular docking, dynamics simulation and drug likeness studies were performed for those natural compounds to evaluate and analyze their anti-cancer activity. **Results:** The top five screened compounds possessing good binding affinity were identified as potential high affinity inhibitors against DOT1L's active site. The top ranking molecule amongst the screened ligands had a Glide g-score of -10.940 kcal/mol and Glide e-model score of -86.011 with 5 hydrogen bonds and 12 hydrophobic contacts. This ligand's behaviour also showed consistency during the simulation of protein-ligand complex for 20000 ps, which is indicative of its stability in the receptor pocket. **Conclusions:** The ligand obtained out of this screening study can be considered as a potential inhibitor for DOT1L and further can be treated as a lead for the drug designing pipeline.

**Keywords:** DOT1L - mixed lineage leukemia - virtual screening - docking - dynamics

*Asian Pac J Cancer Prev*, 16 (9), 3817-3825

### Introduction

The Mixed-Lineage Leukemia (MLL) is an extremely fatal blood cancer that predominantly happens in pediatric patients. The modifications of the MLL gene which are present in approx. 75% of infants and about 10% of adults (Daser et al., 2005; Krivtsov et al., 2007; Liedtke et al., 2009) leads to MLL. Regardless of the phenotypic distinction, the gene expression profiles for MLL-rearranged leukemias of both acute lymphocytic leukemia and acute myeloid leukemia lineage overlaps (Ross et al., 2004). This sort of intense leukemia has an especially poor prognosis (Chen et al., 1993; Felix et al., 1995), with 5-year occasion free survival with MLL-rearranged acute lymphocytic leukemia being just ~40%. Hence, there occurs a pressing need to discover new medication

The human protein methyl-transferase DOT1L has been discovered to be a focus for leukemia with MLL translocations. ENL, AF4, AF9 and AF10 are the four most basic MLL fusion partners which were reportedly

have the capacity to bind the DOT1L protein as a feature of the transcription protein complexes (Okada et al., 2005; Zhang et al., 2006; Mueller et al., 2007; Bitoun et al., 2007; Krivtsov et al., 2008; Mueller et al., 2009). These fusion partners as directed by the AT hook domain of the MLL-rearranged gene, were proposed to structure a large complex which binds to the promoter regions of certain MLL directed genes viz. Hoxa9 (Bitoun et al., 2007). Methylation of H3K4 doesn't occur if the C-terminal of these MLL targeted fusion partners don't possess a SET domain. Rather, the recruited DOT1L methylates H3K79 that can also initiate and maintain transcriptions of the gene as shown in Figure 1. This unusual epigenetic event leads to the dysregulation of the several MLL targeted genes viz. Meis1, Hoxa7 and Hoxa9, which if overexpressed are known to cause leukemia (Okada et al., 2005).

Since DOT1L is the main H3k79 methyltransferase whose contribution in MLL leukemogenesis is clear. Moreover, the transformation of normal progenitor cells

into leukemic stem cells by an artificial fused MLL-DOT1L gene was also reported. However, the failure occurs in the case when MLL fused with a mutant DOT1L which didn't possess methyltransferase activity. This further resulted in establishing the role of the activity of methyltransferase in causing this type of leukemia (Okada et al., 2005). In addition, a high level of the activity of the methyltransferase DOT1L is required for the transformation and maintenance of MLL leukemia. It was also found that the transforming ability of MLL-AF4 and MLL-AF10 got nullified due to the 50%-70% declination in the expression of DOT1L caused by the RNA interference (Okada et al., 2005; Krivtsov et al., 2008). An alternate fascinating application focusing on DOT1L is its role in the programming of cells. Recent studies on the inhibition of DOT1L by utilizing small molecule inhibitor as well as RNA interference have also been reported to significantly improve the reprogramming efficacy to develop induced pluripotent stem cells (iPSC) from somatic cells (Onder et al., 2012). These outcomes have validated the key role of DOT1L in leukemogenesis of MLL-rearranged leukemia and thus signifies a target for therapeutic intervention (Mc Lean et al., 2014).

It has been realized that numerous illnesses including tumor are produced by abnormal epigenetic alterations. By taking into account the vital role of DOT1L or H3K79 methylation in the regulation of gene, it is expected that more infections or therapeutic applications may be associated with aberrant behavior of DOT1L. On the other hand, with the exception of MLL-rearranged leukemia, there is not any convincing proof indicating potential therapeutic applications of focusing on DOT1L. The recent demonstration of DOT1L enzymatic activity as a driver of MLL-rearranged leukemia (Bernt et al., 2011; Daigle et al., 2011) provides a convincing basis for the development of effective, selective DOT1L inhibitors as therapeutic agents for the cure of MLL-rearranged leukemia patients. Till now, that standard inhibitors reported for histone methyltransferases mainly DOT1L receptor are CHAETOCIN (Papovic et al., 2012), DZNep (Helin et al., 2013), ENTACAPONE (Forsberg et al., 2003), EPZ-004777 (Yu et al., 2012), EPZ-5676 (Daigle et al., 2013), EPZ-6438 (Knutson et al., 2013), EPZ-005687 (Tummino et al., 2014), GSK-126 (Verma, 2015), OMEGA-3 FATTYACID (Dimri et al., 2010), P35 (Cacabelos, 2014), TANSHINDIOLS (Woo et al., 2014), CURCUMIN (Lu et al., 2013), EPZ000004 (Basavapathruni et al., 2012), EPZ003696 (Basavapathruni et al., 2012), 5-IODOTUBERCIDIN 004777 (Yu et al., 2012), FED1 and FED2 004777 (Yu et al., 2012), SAH and SAM 004777 (Yu et al., 2012). But these current treatment options are of limited effectiveness; hence, the development of novel small molecule inhibitors of DOT1L is necessary. Not only can these compounds be further developed to become clinically useful therapeutics for MLL leukemia (Chen, 2014), they may also serve as powerful chemical probes to investigate the biology of DOT1L in cancer and other diseases (Chen et al., 2014).

In this study, we employed exhaustive virtual screening and molecular docking studies to report natural inhibitors for DOT1L which more potent than the existing reference

inhibitors (Raj et al., 2015). The critical residues in DOT1L methyltransferase were identified and validated by docking and simulation results of the screened inhibitor in order to elucidate their roles in causing this chronic disease. Further, the Absorption, Distribution, Metabolism, Excretion (ADME) and Toxicity properties were calculated in order to check the potency of these compounds. These results provides a clear demonstration of the potential for potent, selective small molecule inhibitors of DOT1L to affect selective killing of MLL-rearranged leukemias.

## Materials and Methods

### Protein preparation

The three dimensional crystal structure of Human methyl-transferase with PDB-ID: 3UWP (Yu et al., 2012) in complex with the ligand (2R, 3R, 4S, 5R)-2-(4-AMINO-5-IODO-7H-PYRROLO [2, 3-D] PYRIMIDIN-7-YL)-5-(HYDROXYMETHYL) TETRAHYDROFURAN-3, 4-DIOL was retrieved from the Protein Data Bank (<http://www.rcsb.org/pdb>) (Berman et al., 2002). The complexes bound to the receptor molecule, all the heteroatoms and the non-essential water molecules were removed and the PDB structure of this protein was prepared by Protein Preparation Wizard of Maestro, v9.7, Schrodinger, LLC, New York, NY, 2014 (Sastry et al., 2013).

### Ligand selection and preparation

Different databases of compounds were downloaded for the structure based drug design namely: 1) Natural product library with 400 compounds [<http://www.specs.net/snpage.php?spageid=home>], 3) ND (Natural derivatives library) database, Flavonoids library and extended flavonoids library of 3040, 500 and 4000 compounds respectively [<http://www.timtec.net>] (Radha M et al., 2014), 4) NATx (Semi-Synthetic screening compounds) library, MACROx (Next Generation Macrocycles) library, MEGx\_new (Natural products) and FRGx (Fragments from nature) library of 1307, 52, 652 and 59 compounds respectively [<http://www.ac.discovery.com>], 5) Diversity Set IV library and Natural Set II library of 1596 and 120 compounds respectively [<http://www.dtp.nci.nih.gov/repositories.html>], 7) IBS Natural compounds library, Natural compounds scaffold library, bioactive compounds library of 55000, 503 and 761 compounds respectively [<http://www.ibscreen.com>], 8) Ten Specialized databases containing 31805 natural products and metabolites available at zinc [<http://zinc.docking.org/browse/catalogs/natural-products>], 9) 20 Well-known DOT1L inhibitors. All the ligand structures in the libraries were in 2D SDF format which were first converted to 3D for docking. Geometry minimizations were performed on all the ligands using the OPLS 2005 force fields and Truncated Newton Conjugate gradient (TNCG). The ligands were first formed by ligprep with Epik to expand protonation and tautomeric states at 7.0+2.0 Ph units (Shaw et al., 2005; Kresten et al., 2010). For each ligand five low energy stereoisomers were generated and a single low-energy 3D structure with correct chiralities was retained. Qikprop was used for the screening of the ligands and further to

calculate the absorption, distribution, metabolism, and excretion (ADME) properties (Kresten et al., 2010). Both, physicochemically significant descriptors and pharmacokinetically relevant properties were predicted for all the ligands from the compound libraries. Lipinski's rule of five was one of the factors used to assure the drug like (oral) pharmacokinetic profile of the ligands.

#### Active site identification of DOT1L

The catalytic site of human protein methyl-transferase DOT1L was identified by using SiteMap, version 3.0, Schrodinger, LLC, New York, NY, 2014 (Halgren et al., 2009). Active sites with best site scores (top ranked potential receptor binding cavity) were taken as a prerequisite for receptor grid generation. The active sites identified by this program have also been validated with the active site reported in literature available for this receptor. After preparation of protein and identification of active site, receptor grid was generated for the protein by Grid Generation panel of Glide module of version 6.2, Schrodinger, LLC, New York, NY, 2014. Grid dimension for DOT1L protein was as follows: Inner box: X=10, Y=10, Z=12, Outer box: X=22.618795, Y=22.618795, Z=22.618795, Grid center X= 18.042143, Y=62.62294, Z= -9.56231.

#### In silico screening and molecular docking

The virtual screening and docking analysis of Human methyl-transferase DOT1L with 99795 (excluding 20 reference inhibitors) compounds was carried out by GLIDE v6.2, Schrodinger, LLC, New York, NY, 2014. Virtual screening is a technique by which small molecules are identified from the large dataset and which are most likely to bind the targeted protein (Rajamani et al., 2007). Generated dataset was screened at the active site of SH2 domains of both proteins. Screening process of Glide includes three steps High Throughput Virtual Screening (HTVS), Standard Precision (SP) and Extra precision (XP). Through this rigorous docking based in-silico filtering, we obtained some compounds with significant glide and docking score. Some of the well-known inhibitors like CHAETOCIN, DZNep, ENTACAPONE, EPZ-004777, EPZ-5676 etc., were collected as reference molecules. Molecular docking was also performed with these molecules for the same receptor grid for comparative analysis by using GLIDE v6.2, Schrodinger, LLC, New York, NY, 2014.

#### Molecular dynamics simulation for the top ranked complex

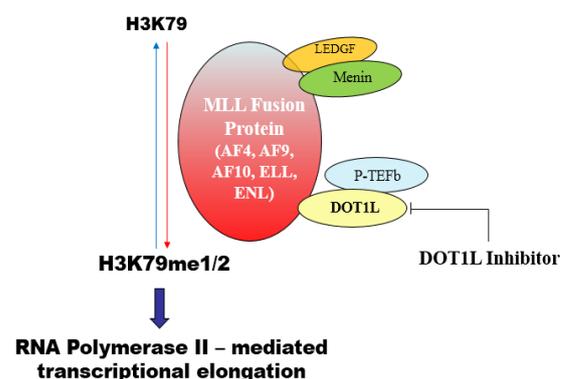
Desmond was used to carry out molecular dynamics (MD) simulation on the complex structure (DOT1L receptor with the top ranked ligand), which utilizes a definite "neutral territory" midpoint method to make efficient utilization of a high degree of computational parallelism (Bowers et al., 2006). The Optimized Potentials for Liquid Simulations (OPLS) 2005 force-field was used to model the amino acid interactions in the modeled protein, and the SPC (simple point charge) method was used for the water model. Default parameters and protocol of Desmond were used to equilibrate the system which consists of steps of restrained minimizations

and simulations which are used for relaxing the system with no significant deviation from the initial coordinates of the modeled protein. The initial coordinates for the molecular dynamics simulation were taken from the modeled protein. The SPC water molecules were then added (the orthorhombic dimensions of each water box were 9Å x 11Å x 12 Å approximately, which ensured that the whole surfaces of the complexes were covered). 11 Na<sup>+</sup> counter ions were added to the system for neutralization in order to balance the net charge of the system. Before equilibration and the long production MD simulations, the systems were minimized for 5000 iterations on a convergence threshold of 1kcal/mol/Å and pre-equilibrated by means of the inbuilt relaxation protocol implemented in Desmond. The whole system was subjected to 300 K for 20000 ps at NPT ensemble of MD simulation with a recording interval of 100 ps for total energy. The structural changes and dynamic behaviour of the modeled protein were analyzed by calculating the root mean square deviation (RMSD) and potential energy.

#### Calculation of ADME and Toxicity

QikProp, v3.9, Schrodinger, LLC, New York, NY, 2014 was used for calculation of ADME properties of the top five inhibitors such as Molecular Weight (MW), hydrophobicity, hydrophilicity, solvent accessible surface area, molecular volume, number of rotatable bond, donor-hydrogen bonds, acceptor hydrogen bonds etc.

OSIRIS property explorer was used for toxicity prediction of top five inhibitors of DOT1L protein (<http://www.organic-chemistry.org/prog/peo/>). OSIRIS property explorer is an online toxicity prediction tool for the analysis of drug relevant property like Mutagenic, Tumorigenic, Irritant, Reproductive effective, cLogP, Solubility, Molecular Weight etc. All the five screened top ranked inhibitors were also selected for the analysis of druglikeness (predicted bioactivity) properties via



**Figure 1. . Model of DOT1L mechanism of action in MLL-rearranged leukemias. MLL fusion proteins, including fusions with MLL-AF4, MLL-AF9, MLL-AF10, MLL-ELL and MLL-ENL, recruit DOT1L, which catalyzes mono- and dimethylation of H3K79, a crucial step in RNA Pol II-mediated transcriptional elongation. A DOT1L inhibitor inhibits H3K79 methylation, which inhibits MLL-fusion protein mediated transcription. Adapted from J Clin Invest (Tummino et al., 2014)**

## Results

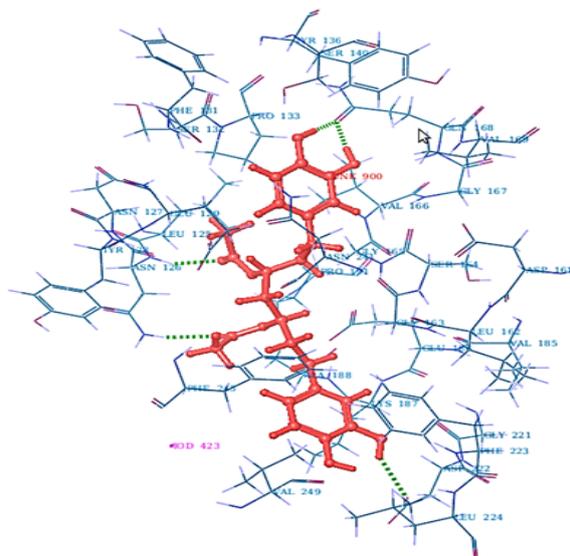
### Structure of the target protein

Human methyl-transferase (DOT1L) has been

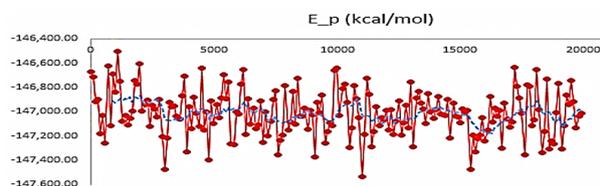
exploited as a main therapeutic target for Mixed-Lineage Leukemia. The three dimensional structure of Human methyl-transferase DOT1L retrieved from the Protein Data Bank with PDB ID: 3UWP determined by X-Ray crystallography at a resolution of 2.05 (Å) was visualized in Pymol. It contains 438 amino acids residues and 98.6% of them are present in the favored regions of

**Table 1. Comparison of top five screened natural inhibitors and well-known inhibitors for DOT1L protein with its Docking score, GLIDE g-score, and GLIDE e-model**

S. No.	Compound IDs	Receptor DOT1L		
		Docking Score	GLIDE g-score	GLIDE e-model
1	(C1)	-10.94	-10.94	-86.011
	(3R,5R)-5-(acetyloxy)-1,7-bis(3,4-dihydroxyphenyl)heptan-3-yl acetate			
2	(C2)			
	({6,7-dihydroxy-4-oxo-1H,2H,3H,4H-cyclopenta[c]chromen-8-yl)methyl) dimethylazanium			
3	(C3)			
	(1R,2S,3R,4R)-N-(2H-1,3-benzodioxol-4-ylmethyl)-2,3-dihydroxy-4-(hydroxymethyl) cyclopentan-1-aminium	-10.015	-10.02	-53.907
4	(C4)	-9.93	-9.956	-57.631
	[(2R)-2-hydroxy-3-[(1R)-6-hydroxy-7-methoxy-1-methyl-1,2,3,4-tetrahydroisoquinolin-2-yl]propyl] (2-hydroxyethyl)azanium			
5	(C5)	-9.429	-9.431	-66.381
	(1R,2R,3S,5S)-5-acetyl-2,3,5-trihydroxycyclohexyl (2E)-3-(3,4-dihydroxyphenyl) prop-2-enoate			
6	CHAETOCIN			
7	DZNep			
8	ENTACAPONE			
9	EPZ-004777	-6.987	-6.987	-65.2
10	EPZ-5676	-8.381	-8.824	-74.187
11	EPZ-6438	-3.36	-3.484	-58.261
12	EPZ-005687	-5.101	-5.224	-68.939
13	GSK-126	-4.675	-4.702	-65.993
14	OMEGA-3 FATTYACID	-4.408	-4.411	-50.701
15	P35	-4.928	-4.95	-64.213
16	TANSHINDIOL B	-5.128	-5.128	-41.237
17	TANSHINDIOL C	-5.12	-5.12	-41.499
18	CURCUMIN	-6.533	-6.533	-60.495
19	EPZ000004	-8.633	-8.68	-58.562
20	EPZ003696	-6.948	-6.968	-83.38
21	5-IODOTUBERCIDIN	-8.811	-8.811	-57.622
22	FED1	-7.099	-7.119	-86.19
23	FED2	-6.427	-6.447	-58.194
24	SAH	-8.992	-8.991	-78.611
25	SAM	-7.582	-7.609	-80.645



**Figure 2. Interaction of the Top Ranked Compound C1 (red in color) with the Active Site Residues of the DOT1L Protein, H-bonds are Shown in Green Color**



**Figure 3. Potential Energy Plot of the Top Ranked Complex with Respect to Time**

Ramachandran plot.

#### *Analysis of the catalytic site of the DOT1L*

Human methyl-transferase (DOT1L) catalytic site predictions were analyzed using SiteMap, version 3.0, Schrodinger, LLC, New York, NY, 2014. SiteMap has detected the largest volume of 527.8 Å<sup>3</sup> in the top most active site of ABL protein on the basis of site score. The active sites identified by this program have been in accordance with the active site reported in literature available for this receptor.

#### *In-silico screening and molecular docking analysis*

The complete set of 99795 compounds (excluding 20 reference inhibitors) were subjected to three levels of virtual screening HTVS, SP and XP which resulted in 11525 compounds, 1212 compounds and 125 compounds respectively. Initially the compounds with high molecular weight (MW >500) and with reactive functional groups were filtered out. The final compounds were then docked into the active site of DOT1L protein, using earlier mentioned three stages of screening and docking. The twenty reported inhibitors were docked on the active site of DOT1L receptor directly utilizing XP-docking technique.

The Glide score was selected as the scoring function and the E-model (Kcal/mol) was used to rank the poses of the ligands as the E-model combines the Glide score, non-bonded interaction energy and excess internal energy of generated ligand conformation for flexible docking. Similarly, docking results for the reported inhibitors are

also reported in Table 1 for the comparative purpose.

The docking studies indicated that the top ranked compounds showed strong hydrogen bonding interactions with the receptor. On the basis of docking score, glide g-score and glide e-model value, compound id's named C1-C5, are more effective against all the human methyl-transferase (DOT1L) receptor as compared to the reported inhibitors as depicted in Table 1. These top five screened compounds could be useful for identification and development of new preventive and therapeutic drug against Mixed-Lineage Leukemia.

Top ranked compound C1 forms two H-bonds with the ASN 122 of the receptor (One with the backbone and other with the side chain of the amino acid residue) as shown in Figure 2. C1 also forms two hydrogen bond interactions with the side chain of GLN 168 and one backbone H-bond with the ASP 122. There is also one  $\Pi$ - $\Pi$  stacking interaction with the PHE 223 amino acid residue. Polar residues ASN 127, SER 132, SER 140, SER 164 and ASN 241 also plays an important role in binding of the C1. C1 also forms two 'vely' charged interactions with GLU 129 and GLU 186 and one '+vely' charged interaction with LYS 187 amino acid residue as indicated in Table 2. Since the screened ligand binds to the same active site residues as reported (Yu et al., 2012), hence it can be safely assumed that C1 also inhibits the methylation of H3K79, which in-turn inhibits MLL-fusion protein mediated transcription.

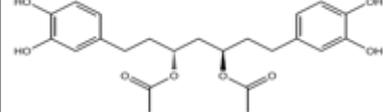
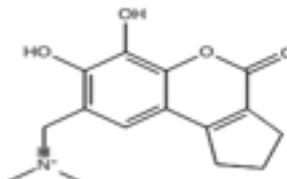
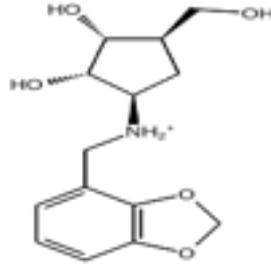
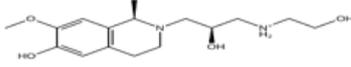
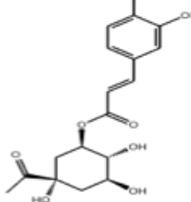
#### *Molecular dynamics simulation results*

Molecular dynamics (MD) simulation was used to explore dynamic perturbations in the conformation of the top ranked complex structure. Here, we have shown the detailed analysis of MD simulation results of the top ranked compound (C1) with DOT1L receptor (GLIDE g-score=-10.940 and GLIDE e-model=-86.011). The initial energy of the complex was -146668kcal/mol. During the course of 20 ns simulation the potential energy shows a decrease and after 10 ns, it was found that the graph shows not much significant differences which is indicative of stabilization of the system as shown in Figure 3.

The conformations obtained during the 20000 ps simulation for the top ranked complex were analyzed using the simulation interaction diagram option of the Desmond program. Root Mean Square Deviation (RMSD) is used to measure the average change in displacement of a selection of atoms for a particular frame with respect to a reference frame. It is calculated for all frames in the trajectory. RMSD was calculated for the protein C-alpha and side chain atoms during the course of 20000 ps MD simulation and the ligand RMSD plot was also fitted on the protein RMSD to check the overall stability of the system as shown in Figure 4.

It is clear from the RMSD graph of the top ranked DOT1L-C1 complex that there was a slight change in conformation by 1.5 Å at very beginning of simulation, after which stability was attained, and the range of backbone atoms varied from 1.000 to 3.963 Å throughout the MD simulation as shown in Figure 4. The RMSD values of the C-alpha and side chain atoms in the system

**Table 2. Top Five Screened Compounds with their Interacting Active Site Residues and their no. of H-Bonds Formed**

Sl No.	Compounds	No. of H-Bonds	H-Bond Interacting Residues	Other non-bonded interactions
1		5	ASN 126(2), GLN 168(2), ASP 222	$\Pi$ - $\Pi$ stacking (PHE 223); Polar (ASN 127, SER 132, SER 140, SER 164, ASN 241); Hydrophobic (LEU 125, TYR 128, PHE 131, PRO 133, TYR 136, VAL 166, VAL 169, ALA 188, PRO 191, LEU 224, PHE 245, VAL 249); Charged +ve (LYS 187); Charged -ve (GLU 129, GLU 186)
2		3	GLU 186(2), PHE 223	Salt-bridge (GLU 129); Polar (ASN 126); Hydrophobic (TYR 128, ALA 188, LEU 224, ALA 244, PHE 245, VAL 249); Charged +ve (LYS 187); Charged -ve (ASP 122, GLU 186)
3		5	GLU 129(2), GLU 186, LYS 187, PHE 223	Polar (ASN 126, SER 164, ASN 241); Hydrophobic (TYR 128, VAL 185, ALA 188, PRO 191, LEU 224, ALA 244, PHE 245, VAL 249); Charged -ve (ASP 222)
4		5	GLU 129, GLY 163, ASP 222, PHE 223, ASN 241	$\Pi$ - $\Pi$ stacking (PHE 223); Salt-bridge (GLU 129); Polar (ASN 126, SER 164, ASN 242); Hydrophobic (TYR 128, LEU 162, ALA 188, PRO 191, LEU 224, VAL 240, PHE 245, VAL 249); Charged -ve (ASP 161, GLU 186)
5		5	ASN 126(2), ASN 126, GLU 163, ASP 222(2)	Polar (ASN 126, SER 164, ASN 241); Hydrophobic (LEU 125, TYR 128, PRO 133, TYR 136, VAL 166, PRO 191, PHE 223, LEU 224, PHE 245, VAL 249); Charged -ve (GLU 186)

are likely to converge after 15000 ps, showing deviation of around 0.5 Å but later on it reaches a plateau with a decrease in deviation to 0.292Å. In the similar manner, the ligand fit protein graph also showed similar behaviour during the simulation process.

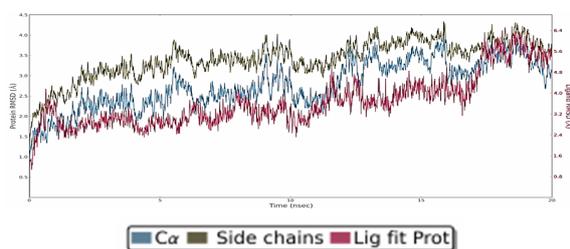
Likewise Root Mean Square Fluctuation (RMSF) for the protein was also calculated to decipher the perturbation in the amino acid residues of the protein. Root Mean Square Fluctuation (RMSF) is useful for characterizing local changes along the protein chain in order to elucidate which residues are more important in the activation of the

receptor. For most residues, RMSF values were within the range of 2 Å and few residues exceeded the 3 Å value. Generally the tails (N- and C-terminal) fluctuate more than any other part of the protein. Secondary structure elements like alpha helices and beta strands are usually more rigid than the unstructured part of the protein, and thus fluctuate less than the loop regions. In this case also, the same result was obtained which is indicative of the stability of the system. The amino acid residues viz. ASN 126, GLU 129, ASP 161, ASP 222 and PHE 223 play a key role in the regulation of DOT1L receptor as more than

**Table 3. ADMET descriptors and Predicted Bioactivity of top scoring screened natural inhibitors**

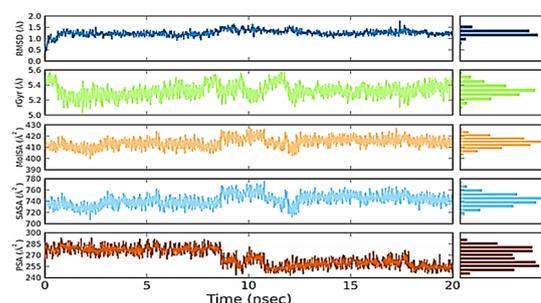
S.No.	ADME Descriptors	(C1)	(C2)	(C3)	(C4)	(C5)
1	% Human Oral Absorption	70.022	66.044	65.265	53.754	50.375
2	logP(o/w)a	3.096	0.761	0.016	0.173	-0.041
3	QPlogSb	-5.591	-1.904	-0.906	-0.384	-2.938
4	QP PCacoc	24.778	86.206	136.735	27.628	21.008
5	logHERGd	-6.496	-4.797	-5.14	-6.113	-5.316
6	PMDCKe	9.089	38.696	63.711	12.514	7.604
7	Polrzf	43.889	27.318	25.954	30.384	32.579
8	logPC16g	16.14	8.336	9.241	10.434	12.51
9	logPocth	24.924	16.108	18.728	19.914	23.529
10	Metabi	6	6	5	8	5
11	logKhsaj	0.289	-0.24	-0.662	-0.524	-0.662
12	Donor HBk	4	2	4	4	5
13	Acceptor HBl	7	6	8.1	8.4	9.65
14	Pot. energy OPLS2005m	30.589	116.004	196.918	178.059	105.668
15	dipolen	7.978	9.411	5.46	3.614	3.305
16	Mutagenic	NO	NO	NO	NO	NO
17	Tumorigenic	NO	NO	NO	NO	NO
18	Irritant	NO	NO	NO	NO	NO
19	Reproductive Effect	NO	NO	NO	NO	NO
20	TPSA	133.522	75.103	95.757	89.761	144.516
21	Mol.wt	432.522	276.312	282.316	311.402	352.339
22	Volume	394.246	250.007	282.316	311.402	304.811
23	GPCR ligand	0.2	-0.52	0.5	0.24	0.14
	Ion Channel Modulator	0.21	-0.62	-0.05	0.02	0.09
	Kinase Inhibitor	-0.06	-0.39	-0.01	-0.16	-0.21
	Nuclear Receptor Ligand	0.33	-0.32	-0.15	-0.47	0.63
	Protease Inhibitor	0.24	-0.57	0.06	0.07	0.28
	Enzyme Inhibitor	0.3	0.07	0.4	-0.04	0.48

\*alog P for octanol/water (logPo/w); blogS for aqueous solubility (QPlogS); capparent Caco-2 permeability in nm/sec (QP PCaco); dHERG K+ channel blockage: log IC50 (logHERG); eapparent MDCK permeability in nm/sec (PMDCK); f Polarizability in cubic angstroms (Polrz); glogP for hexadecane/gas(logPC16); hlogP for octanol/gas (logPoct); iNo. of primary metabolites (Metab); jlog Khsa serum protein binding (logKhsa); kDonor HB; lAcceptor HB; mPot. Energy OPLS2005; ndipole

**Figure 4. Protein-Ligand RMSD Aligned on Protein w.r.t Time for 20 ns Simulation**

96% of the simulation time, these specific interactions are maintained.

Five properties were analyzed to explain the stability of the top ranked ligand in the DOT1L receptor during the simulation of 20 ns as shown in Figure 5: (1) Ligand RMSD - Root mean square deviation of a ligand with respect to the reference conformation (typically the first frame is used as the reference and it is regarded as time t=0);

**Figure 5. Variation of the Properties of the Ligand w.r.t Time During 20 ns Simulation**

(2) Radius of Gyration (rGyr) - It is used for measuring the 'extendedness' of a ligand, and is equivalent to its principal moment of inertia; (3) Molecular Surface Area (MolSA) - Molecular surface calculation with 1.4 Å probe radius; (5) Solvent Accessible Surface Area (SASA) - Surface area of a molecule accessible by a water molecule; (6) Polar Surface Area (PSA) - Solvent accessible surface area in a

molecule contributed only by oxygen and nitrogen atoms.

From the graph, it is evident that Ligand RMSD remains constant during the simulation process. The overall RMSD of the ligand C1 was below 2 Å. The ligand C1 in the initial stage shows fluctuation from 0 ps to 1,300 ps but later maintains constant RMSD during the whole simulation. This analysis indicates clearly that both compounds maintained constant RMSD during the simulation. Initially there was a fluctuation in ROG for the ligand C1 until 12 ns and then a stable conformation was acquired throughout the simulation. The radius of gyration throughout the 20 ns simulation for the ligand C1 ranged from 5.0 Å to 5.6 Å. MolSA, SASA and PSA plots also indicated the consistency of the ligand mainly after 11-12 ns during the simulation process.

#### ADMET predictions

ADMET properties are considered to be one of the most important criteria in order to determine true potential of the compound (Puzyn et al., 2010). About 50% of drugs are found to be unsuccessful due to the toxicity of compound during the clinical trials. Therefore, in order to save the expenses at the clinical trial stage the in silico toxicity prediction tools are used. Physically critical descriptors and pharmaceutically significant properties of these five lead compounds were investigated utilizing Qikprop program as depicted in Table 3. All the five screened compounds were in the acceptable range of Lipinski's rule of five. For these five screened lead compounds, the partition coefficient (QP logP (o/w)) and the water solubility (QP log S), which are essential when measuring the absorption and distribution of drugs within the body, ranged between -0.041 to 3.096 and -0.384 to -5.591, respectively, while the cell permeability (QP PCaco), a crucial factor governing drug metabolism and its access to biological membranes, ranged from 21.008 to 136.735. Overall, the percentage human oral absorptions for these five compounds ranged from 50% to 75%.

Toxicity Risk Assessment panel of the Osiris Property Explorer (OPE) was used to predict toxic property of selected natural inhibitors as summarized in Table 3. OPE use to predict toxic property by analyzing available molecular toxic property in the Registry of Toxic Effects of Chemical Substances (RTECS) database. RTECS, a Accelrys database which is a repository of chemical structures and their toxic property.

By analyzing the results of toxicity predictor, all five screened compounds were non mutagenic, non-tumorigenic, nonirritant and non-reproductive. The predicted bioactivities of these screened compounds for acting as GPCR ligand Ion, Channel Modulator, Kinase Inhibitor, Nuclear Receptor Ligand, Protease Inhibitor and Enzyme Inhibitor calculated through MOLINSPIRATION server were also depicted in Table 3. As a general thumb rule, larger is the bioactivity score, higher is the probability that investigated compound will be active. Therefore, a molecule having bioactivity score more than 0.00 is most likely to possess considerable biological activities, while values -0.50 to 0.00 are expected to be moderately active and if score is less than -0.50 it is

presumed to be inactive (Verma et al., 2012). The results of the current study revealed that the screened natural compounds are biologically active molecules and will produce the physiological actions by multiple mechanisms after interacting with GPCR ligands, nuclear receptor ligands, and inhibit protease and other enzymes. Though bioactivity score for Enzyme inhibition is found to be >0.00 for all tested compounds except compound C4, but the highest score (0.48) was observed for compound C5 closely followed by compound C3 (0.40) and compound C1 (0.30).

## Discussion

In conclusion, in our study we have employed virtual screening, molecular docking and dynamics simulation studies in order to find the natural inhibitors for DOT1L receptor which can be more potent in curing MLL. It was observed that the top five screened natural compounds shows minimum binding energy and better docking score with the human methyl-transferase (DOT1L) receptor. The molecular dynamics study of the best ligand-receptor complex (Compound C1 with DOT1L) also exhibit the stability of the compound during the 20 ns simulation process which validate the virtual screening and docking results. Interactions with the amino acid residues ASN 126, GLU 129, ASP 161, ASP 222 and PHE 223 remains consistent during the MD simulation process. The screened ligand acts apparently by inhibiting methylation of H3K79 thereby preventing the aberrant RNA poly2 mediated transcriptional elongation. For further validation of these five natural screened compounds, predicted bioactivities, ADME properties and their toxicity results indicates their drug likeliness. Thus, we can conclude that these five natural compounds especially compound C1 ((3R,5R)-5-(acetyloxy)-1,7-bis(3,4-dihydroxyphenyl)heptan-3-yl acetate) can be used either as potential lead to synthesize new drugs for DOT1L receptor involved in MLL or could be introduced into clinical trials as a fire new drug for the treatment of mixed lineage leukemia.

## Acknowledgements

Authors are thankful to the department of Bioinformatics, Indian Institute of Information Technology, Allahabad for providing computing facility.

## References

- Basavapathruni A, Jin L, Daigle SR, et al (2012). Conformational adaptation drives potent, selective and durable inhibition of the human protein methyltransferase DOT1L. *Chemical Biology Drug Design*, **80**, 971-80.
- Berman, HM, Westbrook J, Feng Z, et al (2002). The protein data bank. *Biological Crystallography*, **58**, 899-907.
- Bernt KM, Zhu N, Sinha AU, et al (2011). MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. *Cancer Cell*, **20**, 66-78
- 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**, 92-106.

- Bowers KJ, Dror RO, Shaw DE (2006). The midpoint method for parallelization of particle simulations. *J Chem Phys*, **124**, 184109-11.
- Cacabelos R (2014). Epigenomic networking in drug development: from pathogenic mechanisms to pharmacogenomics. *Drug Development Res*, **75**, 348-365.
- Chen CS, Sorensen PH, Domer PH, et al (1993). Molecular rearrangements on chromosome 11q23 predominate in infant acute lymphoblastic leukemia and are associated with specific biologic variables and poor outcome. *Blood*, **81**, 2386-93.
- Chen SJ (2014). A Potential target of tanshinone iia for acute promyelocytic leukemia revealed by inverse docking and drug repurposing. *Asian Pac J Cancer Prev*, **15**, 4301.
- Chen, SJ, Ren JL (2014). Identification of a potential anticancer target of danshensu by inverse docking. *Asian Pacific J Cancer Prev*, **15**, 111-116.
- Daigle SR, Olhava EJ, Therkelsen CA, et al (2011). Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor. *Cancer Cell*, **20**, 53-65.
- Daigle SR, Olhava EJ, Therkelsen CA, et al (2013). Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. *Blood*, **122**, 1017-25.
- Daser A, Rabbitts TH (2005). The versatile mixed lineage leukaemia gene MLL and its many associations in leukaemogenesis. *Semin Cancer Biol*, **15**, 175-188.
- Dimri M, Bommi PV, Sahasrabudhe AA, et al (2010). Dietary omega-3 polyunsaturated fatty acids suppress expression of EZH2 in breast cancer cells. *Carcinogenesis*, **31**, 489-495.
- Felix CA, Hosler MR, Winick NJ, et al (1995). ALL-1 gene rearrangements in DNA topoisomerase II inhibitor-related leukemia in children. *Blood*, **85**, 3250-6.
- Forsberg M, Lehtonen M, Heikkinen M, et al (2003). Pharmacokinetics and pharmacodynamics of entacapone and tolcapone after acute and repeated administration: a comparative study in the rat. *J Pharmacol Exp Ther*, **304**, 498-506.
- Halgren T (2009). Identifying and characterizing binding sites and assessing druggability. *J Chem. Inf. Model*, **49**, 377-389.
- Helin, K, Dhanak D (2013). Chromatin proteins and modifications as drug targets. *Nature*, **502**, 480-488.
- Knutson SK, Warholc NM, Wigle TJ, et al (2013). Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. *Proceed National Academy Scien*, **110**, 7922-27.
- Krivtsov AV, Armstrong SA (2007). MLL translocations, histone modifications and leukaemia stem-cell development. *Nat Rev Cancer*, **7**, 823-833.
- Krivtsov AV, Feng Z, Lemieux ME, et al (2008). H3K79 methylation profiles define murine and human MLL-AF4 leukemias. *Cancer Cell*, **14**, 355-68.
- Liedtke M, Cleary ML (2009). Therapeutic targeting of MLL. *Blood*, **113**, 6061-8.
- Lu D (2013). Epigenetic modification enzymes: catalytic mechanisms and inhibitors. *Acta Pharmaceutica Sinica B*, **3**, 141-9.
- McLean CM, Karemaker ID, van Leeuwen F (2014). The emerging roles of DOT1L in leukemia and normal development. *Leukemia*, **28**, 2131-8.
- Mrozek K, Heinonen K, Lawrence D, et al (1997). Adult patients with de novo acute myeloid leukemia and t(9; 11)(p22; q23) have a superior outcome to patients with other translocations involving band 11q23: a cancer and leukemia group B study. *Blood*, **90**, 4532-8.
- Mueller D, Bach C, Zeisig D, et al (2007). A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification. *Blood*, **110**, 4445-54.
- Mueller D, Garcia-Cuellar MP, Bach C, et al (2009). Misguided transcriptional elongation causes mixed lineage leukemia. *PLoS Biol*, **7**, 1000249
- Okada Y, Feng Q, Lin Y, et al (2005). hDOT1L links histone methylation to leukemogenesis. *Cell*, **121**, 167-178.
- Onder TT, Kara N, Cherry A, et al (2012). Chromatin-modifying enzymes as modulators of reprogramming. *Nature*, **483**, 598-602.
- Popovic R, Licht JD (2012). Emerging epigenetic targets and therapies in cancer medicine. *Cancer discovery*, **2**, 405-413.
- Puzyn T, Leszczynski J, Cronin MTD (2010). In silico approaches for predicting ADME properties. *Recent Advances in QSAR Studies*, **8**, 283-304.
- Radha M, Suganya J, Naorem DL, Nishandhini M (2014). In silico docking studies of selected flavonoids-natural healing agents against breast cancer. *Asian Pac J Cancer Prev*, **15**, 8155-9.
- Rajamani R, Good AC (2007). Ranking poses in structure-based lead discovery and optimization: current trends in scoring function development. *Curr Opin Drug Discov Devel*, **10**, 308-15.
- Raj U, Varadwaj PK (2015). Flavonoids as multi-target inhibitors for proteins associated with ebola virus: in-silico discovery using virtual screening and molecular docking studies. *interdisciplinary sciences: computational life sciences*, 1-10. DOI: 10.1007/s12539-014-0246-5.
- Ross ME, Mahfouz R, Onciu M, et al (2004). Gene expression profiling of pediatric acute myelogenous leukemia. *Blood*, **104**, 3679-3687.
- Sastry, GM, Adzhigirey M, Day T, Annabhimoju, R, Sherman W (2013). Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J Comput Aid Mol Des*, **27**, 221-34.
- Shaw David E, Kresten Lindorff-Larsen, Stefano Piana, et al (2010). Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Structure Function Bioinformatics*, **78**, 1950-8.
- Shaw DE, Yibing S, John LK, Michael PE, Ron OD (2005). Gaussian split ewald: a fast ewald mesh method for molecular simulation. *J Chem Phys*, **122**, 1-13.
- Tummino PJ, Campbell RM (2014). Cancer epigenetics drug discovery and development: the challenge of hitting the mark. *J Clin Invest*, **124**, 64-9.
- Verma SK (2015). Recent progress in the discovery of epigenetic inhibitors for the treatment of cancer. *Cancer Epigenetics*, **1238**, 677-88.
- Verma A (2012). Lead finding from *Phyllanthus debelis* with hepatoprotective potentials. *Asian Pac J Trop Biomedicine*, **2**, 1735-7.
- Woo J, Kim HY, Byun BJ, et al (2014). Biological evaluation of tanshindols as EZH2 histone methyltransferase inhibitors. *Bioorganic Med Chem Letters*, **24**, 2486-92.
- Yu W, Chory EJ, Wernimont AK, et al (2012). Catalytic site remodelling of the DOT1L methyltransferase by selective inhibitors. *Nature Communications*, **3**, 1288.
- Yu W, Tempel W, Smil D, et al (2012). Crystal structure of Dot1l in complex with 5-iodotubercidin. DOI:10.2210/pdb3uwpp/pdb.
- Zhang W, Xia X, Reisenauer MR, Hemenway CS, Kone BC (2006). Dot1a-AF9 complex mediates histone H3 Lys-79 hypermethylation and repression of ENaCalpha in an aldosterone-sensitive manner. *J Biol Chem*, **281**, 18059-18068.