




## Quantitative structure-activity relationship and molecular docking studies on human proteasome inhibitors for anticancer activity targeting NF- $\kappa$ B signaling pathway

Deepika Yadav, Bhartendu Nath Mishra & Feroz Khan

To cite this article: Deepika Yadav, Bhartendu Nath Mishra & Feroz Khan (2019): Quantitative structure-activity relationship and molecular docking studies on human proteasome inhibitors for anticancer activity targeting NF- $\kappa$ B signaling pathway, Journal of Biomolecular Structure and Dynamics, DOI: [10.1080/07391102.2019.1666743](https://doi.org/10.1080/07391102.2019.1666743)

To link to this article: <https://doi.org/10.1080/07391102.2019.1666743>

 View supplementary material [↗](#)

 Accepted author version posted online: 13 Sep 2019.  
Published online: 23 Sep 2019.

 Submit your article to this journal [↗](#)

 Article views: 12

 View related articles [↗](#)

 View Crossmark data [↗](#)



## Quantitative structure-activity relationship and molecular docking studies on human proteasome inhibitors for anticancer activity targeting NF- $\kappa$ B signaling pathway

Deepika Yadav<sup>a</sup>, Bhartendu Nath Mishra<sup>b</sup> and Feroz Khan<sup>a</sup>

<sup>a</sup>Department of Metabolic and Structural Biology, CSIR–Central Institute of Medicinal and Aromatic Plants, Lucknow, Uttar Pradesh, India;

<sup>b</sup>Department of Biotechnology, Institute of Engineering & Technology (a Constituent Institute of Dr. A.P.J. Abdul Kalam Technical University, Lucknow), Lucknow, Uttar Pradesh, India

Communicated by Ramaswamy H. Sarma

### ABSTRACT

The abnormal ubiquitin-proteasome is found as an important target in various human diseases, especially in cancer, and recently it has received prevalent attention as a challenging therapeutic target. The current work is designed to derive a predictive two-dimensional quantitative structure-activity relationship model for anticancer human proteasome target of NF- $\kappa$ B signaling pathway. The established 2D-QSAR is dependent on multiple linear regression approach and validated through leave-One-Out and external test set prediction method. The robust QSAR model showed the  $r^2$  of 0.83 and  $q^2$  of 0.80 and  $\text{pred}_r^2$  of 0.77. Three chemical properties, electronegativity count, average potential, and T\_2\_N\_6 were identified as significant descriptors to predict the anticancer activities of the proteasome antagonists. Besides, the predicted top hit compounds were considered to check out the compliance with Rule of five and pharmacokinetic parameters for oral bioavailability in the human body. The molecular docking was accomplished to unravel the molecular mode of action of best-predicted compounds which was compatible with the standard drug. Following this approach, lastly two compounds NP and AP were recognized as the best candidates since these top compounds follow all the standard limit point of entire filters and indicated effective and decent docking score. The outcomes of the study sturdily suggested that the developed model and top hit compound's binding conformation are rational in the exploration of unknown antagonist's anticancer activity. The research would be of great support and is supposed to be of immense significance in the development and designing of drug-like candidates in preliminary drug discovery.

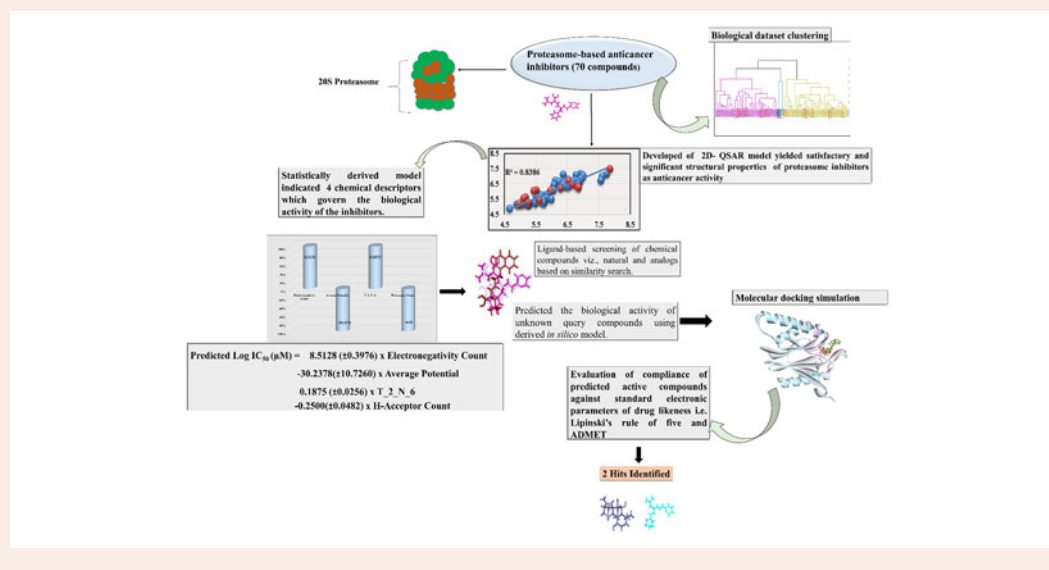
### ARTICLE HISTORY

Received 5 July 2019

Accepted 23 August 2019

### KEYWORDS

Proteasome; NF- $\kappa$ B signaling; QSAR; ADMET; Molecular docking



## Introduction

An ubiquitin-proteasome (UP) emerged as a crucial anticancer protein in current pharmacological research. The UP stimulates NF- $\kappa$ B signaling, which involved commonly in the up-regulation of various antiapoptotic genes, pro-inflammatory cytokines, and cell adhesion molecules, hormonal/growth factors that increase the persistence and vitality of myeloma cells or cancerous cell of plasma (Kravtsova-Ivantsiv & Ciechanover, 2015; Chen & Zhao, 2013; Kubiczkova, Pour, Sedlarikova, Hajek, & Sevcikova, 2014). The ubiquitin-proteasome cascade is primarily considered as an unpretentious destruction process against damaged and insignificant cellular proteins nonetheless at the present time, this pathway has been emerging as a vital mechanism involving in different intracellular signaling cascade regulation, natural breakdown of cyclin-reliant kinases, cyclins, cellular processing, and regulation of apoptosis and angiogenesis in eukaryotes. The UP pathway responsible for degradation and processing of misfolded non-functional proteins like transcription factors through proteolysis (chemical reaction) with the help of enzyme protease and the entire mechanism is guided by a well-organized mode of ubiquitination signaling (Spataro, Norbury, & Harris, 1998; Orłowski & Dees, 2002; Voutsadakis, 2017). The previous research revealed that the UP signaling plays an essential role in the NF- $\kappa$ B signaling. The UP system does not target only I $\kappa$ B protein (NF- $\kappa$ B pathway) degradation nonetheless also participate in the processing of p105 and p100 precursors maturation with the help of proteasome. More remarkably, the NF- $\kappa$ B signaling pathway activates through ubiquitinated activation of cytoplasmic protein kinase by the independent-degradation mechanism which initially activated directly through the phosphorylation of IKK proteins that may be induced by means of various stimuli such as growth factors, the stress hormone, cytokines, and UV rays, etc (Xie et al., 2009; Tu et al., 2012). Conclusively, the phosphorylated IKK protein in the cytoplasm further targeted to the ubiquitination proteasomal degradation of the I $\kappa$ B proteins that results in the liberation of the NF- $\kappa$ B transcription factor and translocation into the nucleus then initiates the transcription process and gene expression (Palombella, Rando, Goldberg, & Maniatis, 1994). Thus, the elevated concentration or high level of proteasome along with the abnormally higher level of proteasome function has been observed in a diverse cancer type and appears to be crucial in the tumor or cancer cell progression. In addition, it supports in shield against apoptotic pathways and eliminating the cell of impaired proteins (Arlt et al., 2009). Dysregulation or frequent activation or high-regulation of the NF- $\kappa$ B signaling establishes a complex microenvironment that is serious towards tumor formation or cancer development (Gilmore, 1999; Pikarsky et al., 2004). Inhibition of the UP pathway has been exhibited to stimulate apoptosis in many cancer types. Previous studies revealed that the tumor or cancer cell is highly sensitive as compared to normal cell towards proteasome inhibition (Adams, 2003; Delic et al., 1998). The previous studies have been widely explored the molecular mechanism of action of proteasome inhibitors that may specifically target cancer cell. The proteasome suppression may selectively induce the apoptosis in various cancer

cells through the ubiquitin-proteasome pathway system (Crawford et al., 2006; Kazi et al., 2009). The proteasome antagonists have also acted as to induce cell cycle arrest. The proteasome prevention in several cases enhance the oxidative stress in tumor or cancer cell and results in the cell death (Wang, Yen, Kaiser, & Huang, 2010).

The different kinds of cancer namely breast cancer, ovarian cancer, non-small cell lung cancer, colon cancer, hepatocellular carcinoma, prostate cancer, and multiple myeloma, etc. develop due to abnormal activation of the ubiquitin-proteasome pathway (Cardoso, Ross, Piccart, Sotiriou, & Durbecq, 2004; Mitra, 2018; Escobar, Velez, Belalcazar, Santos, & Raez, 2011; Jang, 2018; Dawson, 2008; Cao & Mao, 2011).

Beyond the last ten years, proteolytic or proteasome inhibition has become an emerging field of interest in the therapeutic area for medication towards multiple myeloma and different types of lymphomas. In 2003, Bortezomib (BTZ) has reported as the first and principal FDA accepted and permitted drug as a proteasome inhibitor. Hence, the therapies which are centered on Bortezomib drug have become basic intended for multiple myeloma cure at all phases of the disease (Nalepa, Rolfe, & Harper, 2006). Due to the clinical establishment of BTZ, the death rate of MM patients has been reduced. However, it is not possible for all patients to respond to Bortezomib based treatment because of its several limitations. Besides, relapsing arises in some patients who responded initially for medicine. Particularly, solid tumors, have often become resistant to BTZ (Moreau et al., 2012; Dick & Fleming, 2010). Likewise, Carfilzomib, a second-generation U.S. FDA approved (August 2012) proteasome inhibitor stimulates responses in a lesser of multiple myeloma patients. Later, four new second-generation proteasome inhibitors namely Ixazomib, Oprozomib, Marizomib and Delanzomib having wide-range of anticancer and diverse pharmacological properties, also showed several clinical actions in bortezomib-resistant cancers (Chen, Frezza, Schmitt, Kanwar, & P Dou, 2011; Schrader et al., 2016).

Terpenoids signify a wide range of naturally occurring molecules comprising monoterpenoids, diterpenoids, triterpenoids, and tetraterpenoids. Plant-based natural products retain biocompatible characteristic, high chemical diversity, and other crucial molecular properties that make them valuable and advantageous as lead skeleton or scaffold for drug discovery (Harvey, Edrada-Ebel, & Quinn, 2015). These natural compounds can prevent cancerous cell differentiation and have the ability to stimulate cancer cell death by obstructing various specific molecular anticancer targets comprising NF- $\kappa$ B based proteasome complex and a number of antiapoptotic proteins, etc. (Yang and Ping Duo 2010). Moreover, previous studies suggested that Betulinic acid and glycyrrhetic acid have the potential to obstruct and hinder the human 20S proteasome in cancer disease (Huang et al., 2008; Qian et al., 2011).

The present study involved in the understating of physico-chemical properties of molecules and their mechanism which regulates the proteasome pathway activation. To avoid drug resistance and increased cytotoxicity in cancer therapy, the current study focused on the development of more effective, potential, least toxic, safe and more specific agents or

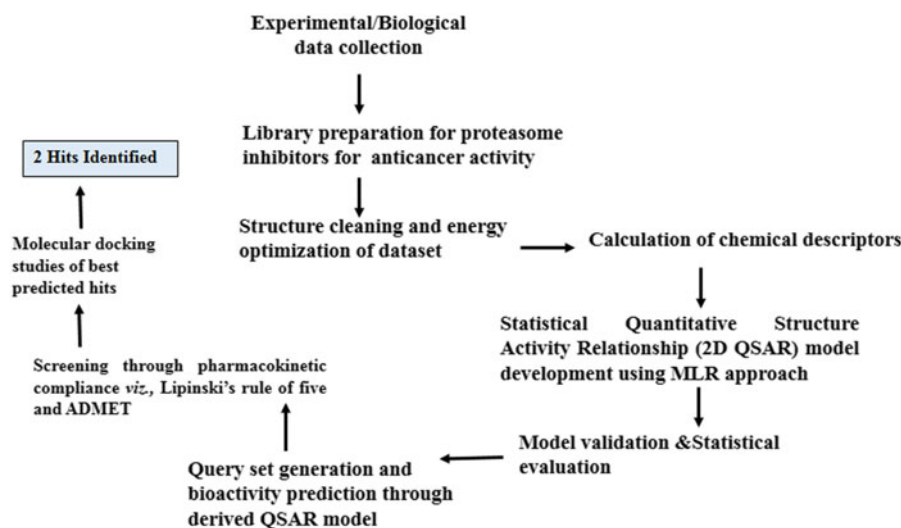


Figure 1. Work flow representation .

compounds against proteasome involved in NF- $\kappa$ B pathway activation. Therefore, the identified compounds could be further modified to express better inhibitory action for proteasome and might be able to offer a fresh and new therapeutic entity to treat a number of cancer types and other inflammatory diseases.

To establish the current study, a total of 71 inhibitors comprising dipeptidyl boronic acid, betulinic acid, and glycyrrhetic acid inhibitors for human proteasome were retrieved through previous literature to develop a QSAR model. Subsequently, best-predicted compounds were screened through a derived model. Later on, predicted best candidates were considered to assess the pharmacokinetics compliances and compared to the FDA approved drug Bortezomib and Carfilzomib against proteasome. Furthermore, the best-screened compounds were prioritized through docking simulation against the human proteasome target of the NF- $\kappa$ B signaling. As far as we know, it is conceivably the first work reported on this subject/topic. Henceforward, the current study would be accountable for deep understanding for identification and optimization of novel drug-like leads in modern drug designing process.

## Materials and Method

### Biological dataset selection

Primarily, a library of 71 known proteasome inhibitors were retrieved and compiled from different works of literature which include terpenoids, dipeptidyl boronic acid analogs (Bortezomib, reference drug derivatives) (Huang et al., 2008; Qian et al., 2011; Milo et al., 2011). The selection of the dataset was centered on the structural diversity and applicability domain of the endpoint ( $IC_{50}$ ) of the dataset. The inhibitory concentration ( $IC_{50}$   $\mu$ M) of proteasome antagonists were transformed into their negative log ( $pIC_{50}$ ) which was considered as the dependent variable and descriptors as independent variables to develop the statistical regression model. Thus, a schematic representation of the methodology adopted in the study is shown in Figure 1.

### Applicability domain of dataset

The applicability domain (AD) is a significant concept in the development of QSAR model that must be quantified before the prediction of a set of molecules. The applicability domain signifies the range of biological activity ( $IC_{50}$   $\mu$ M) of the dataset (Luque Ruiz & Gómez-Nieto, 2018). It allows us to estimate the ambiguity in the prediction set based on similarity to the compounds AD values that involved in the procedure of model generation and improvement. Though, the value of the AD of the training set used to develop QSAR model was persisted between 7.85 and 4.65. Likewise, the range of query set compounds existed as 7.85–5.68.

### Design and execution

The design and execution section define the material and practices employed for (1) structure normalization (2) Molecular descriptors calculation (3) Model generation (4) Statistical parameters evaluation of the generated QSAR model (5) Bioactivity prediction for query set compounds.

### Structure normalization/standardization

The structural geometry optimization and energy calculation of compounds were executed using VLifeMDSv4.5, 2018 (Molecular Design suite) molecular modeling software. Further, entire 2D molecule structures were converted into 3D structures by utilizing the V-Life converter module through consuming Merck Molecular Force Field (MMFF) based on distance-dependent dielectric function and energy gradient of 0.001 kcal/mol Å.

### Chemical descriptors calculation

The molecular descriptors viz., physicochemical, atom-type count and alignment independent were calculated and screened for each compound of the dataset using the QSARplus module of VLifeMDS v4.5, 2018 (VLife Technologies, India) software. Additionally, the highly

correlated chemical descriptors with biological activity were extracted to acquire the superlative and finest subgroup of chemical descriptors.

### QSAR model development

An experimental dataset of 71 compounds with optimized energy and calculated descriptors were further practiced to produce 2D-QSAR model using VLife v4.5, 2018 module. Primarily, the dataset was randomly partitioned as 20% of data into test set (15 compounds) and the remaining 80% of the modeling set was utilized for the training set (56 compounds). As well, the training set was applied to develop a model and to adjust the standard parameters of the model and test set was utilized to evaluate and validate the robustness, predictive/extrapolative ability and competence of the derived 2D-model. The chemical structure of molecules with their actual and observed biological activity against proteasome is presented in Table S1 (Supplementary material). The  $PIC_{50}$  ( $\mu\text{M}$ ) value was taken as the dependent variable along with chemical properties as independent variables. After that, the MLR approach with the stepwise forward-backward manner was employed to generate a reliable QSAR model.

### Prediction set generation and bioactivity prediction

A query set of 3049 compounds were screened through *in silico* approach to identify the best lead. To accomplish the protocol, different query set (prediction set) of 3049 compounds were collected through the PubChem database. The query set compounds have comprised the terpenoids and their derivatives (natural plant-based) and analogs which is based on 80% similarity of parent compound i.e. drug. The prediction set selection was based on Tanimoto score similarity of 80% with that of Bortezomib, Betulinic acid, and Glycyrrhetic acid as well (Bajusz, Rácz, & Héberger, 2015). Further, these query set compounds were screened for bioactivity prediction through the derived 2D-QSAR model.

### Oral bioavailability (drug-likeness) and ADMET risk screening

A thumb rule of five and pharmacokinetic compliance were calculated to screen out best compounds which were having the physiochemical properties like a drug. It is very crucial to recognize the various physiochemical properties that would responsible to prepare and produce a compound as an oral bioavailable in the physiological system i.e. human body. In addition, ADMET risk parameters and drug-likeness signify the absorption, distribution, metabolism, excretion, and toxicity of compounds as well. The ADMET descriptor evaluation allows us to eliminate the compounds with poor features of ADMET to elude expensive reformulation in early drug discovery. The druggability or pharmacokinetics compliance of predicted compounds were evaluated using the United States Food and Drug Administration (USFDA), a standard TOPKAT, Discovery Studio, Accelrys, USA software (Shukla et al., 2014).

**Table 1.** Statistical parameters of QSAR model.

Parameters	Value
N (Number of components)	56
Degree of freedom	51
$r^2$ (Regression coefficient)	0.83
$q^2$ (Cross-validated correlation)	0.80
$F$ -test (Fischer's value)	65.34
$r^2$ _se (Standard error)	0.30
$q^2$ _se (Standard error)	0.33
Pred_ $r^2$	0.77
Pred_ $r^2$ _se	0.39
<b>Z-Score</b>	
$r^2$	7.67
$q^2$	6.89

### Target identification and molecular docking studies

The proteasome was reported as a unique and novel prominent target against cancer treatment (Adams, 2004). The ubiquitination-proteasome showed a vital role at the time of NF- $\kappa$ B activation. Moreover, proteasome was involved in the polyubiquitination and subsequent proteolytic degradation of cytoplasmic inhibitory protein kinase i.e. I $\kappa$ B kinase. The proteasome released the NF- $\kappa$ B transcription factor bounded with I $\kappa$ B which stimulate the NF- $\kappa$ B canonical or non-canonical pathway activation through IKK complex phosphorylation. Therefore, ubiquitin-proteasome/proteolysis cascade showed a dynamic and vital role during protein degradation in the cytosol or nucleus at various steps of the signaling (Xiao, Harhaj, & Sun, 2001; Pickart, 2004). Hence, docking analysis was performed by employing BioSolveIT, FlexX v2.1.8. and docking poses visualization was done through Discover Studio v3.5 (Accelrys, San Diego, CA, 2013) to examine the binding active pose of the best antagonist within a bioactive pocket of a protein/target. The docking simulations may offer a deep and broad understanding of the ligand and target/protein interaction.

### Protein preparation

To accomplish the molecular docking simulation, the 3D crystallographic structure (X-Ray) of the target protein, proteasome bound with drug Bortezomib (resolution of 2.1 Å) was retrieved through the RCSB Protein database (<http://www.rcsb.org/>). To execute the docking simulation, the human 20S proteasome, **PDB ID: 5IF3** with bound drug Bortezomib as a complex was considered. After that, the protocol of protein preparation was applied to accomplish the tasks including the addition of missing atoms in incomplete residues, modeling required or absent loop regions, deletion of extra or substitute conformations, deletion of heteroatoms and water molecules, protonation of titratable residues, a negative log measure of the acid dissociation constant and addition of hydrogen atoms.

### Ligand preparation

The best-predicted compounds including drugs along with its analogs and natural compounds were prepared before performing the docking study. The geometry optimization of modeling compounds were completed by applying the



algorithmic force field monitored by the software protocol. The three-dimensional conformation generation and prepared ligands were saved as MOL2 files.

### Terpenoids molecules parameterization

The natural molecules were standardized prior to docking studies. The geometry optimization and energy minimization of natural compounds were performed using the Chem3Dv15.0.0.106. The natural compound structures were subjected to energy minimization by following the two-phase process. In the first phase, the Molecular Mechanics-2 or MM2 force field was employed for minimization and executed until the Root Mean Square (RMS) gradient turn into 0.1 kcal/mol Å then the for the second phase, MM2 minimized molecules were again considered to re-minimize by the molecular orbital package (MOPAC) approach and executed until the RMS gradient achieved the value of 0.0001 kcal/mol or less. Afterward, the molecular docking studies were performed with the standard default parameters of the software.

### In silico protein-ligand docking studies

Furthermore, the best predicted active compounds were exposed to docking studies in order to investigate and explore their probable interacting poses and affinity towards the target. Furthermore, molecular docking study was accomplished by employing BioSolveIT, FlexX v2.1.8, as per hardware standard (Pagadala, Syed, & Tuszynski, 2017). Additionally, standard parameters of FlexX were used for iterative growing and successive scoring of different docking poses. Henceforth, in order to implement the docking experiments for all best predicted and screened antagonists, the active and vigorous site atoms within a protein/target was demonstrated as atoms within a radius of 6 Å with the particular co-crystallized ligand of the receptor. The files of receptor description were automatically created from the (receptor) PDB coordinates which were used by FlexX. Moreover, 2D interaction was also performed to ravel out the definite interactive residues of the targeted protein with bound ligand (Hit).

### Reference drug

The First U.S. FDA approved drug was considered for comparative study. Thus, Bortezomib expressed a noteworthy breakthrough as the leading and principal drugs in the treatment of malignant disease by functioning as proteasome inhibitors. This drug was used as a control drug against human proteasome complex in molecular docking and pharmacokinetic compliance assessment for comparative analysis with respect to predicted best compounds.

### In silico druggability and pharmacokinetics compliance evaluation

Human druggability and pharmacokinetic assessment perform a vital role in evaluating the quality of potential clinical hit compounds where the precise and accurate valuation of

clearance, distribution volume, drug bioavailability, and the plasma protein- concentration-time figures are the coveted endpoints. The pharmacokinetic properties specify to the Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) of the compounds. Drug-likeness of the candidate compounds were evaluated through Lipinski's Rule of Five to understand the physicochemical properties of the candidate compounds which would be responsible to make it orally active inside the human body. While primary drug discovery, optimization of pharmacokinetics compliance and druggability are crucial for minimizing the later failure in the drug designing process.

## Results and Discussion

### Biological dataset clustering

Clustering is a computational method to understand and visualize the structure of complex data and make a grouping of the data points based on similarity. The clustering analysis of proteasome inhibitors were performed using Molsoft L.L.C., San Diego, CA software and represented in Figure S1 (Abagyan, 2018). This approach enables us to create biological intrusion or interference for additional experiments. As far as, clustering is an unsubstantiated and unsupervised splitting of the dataset into two different groups so that the compounds in each group would be more similar and comparable to each other (Ronan, Qi, & Naegle, 2016). Hierarchical clustering analysis of entire dataset was done based on Tanimoto structure similarity distance of 0–0.7 which indicated that the training and test set compounds fall under the applicability domain of the developed model.

### QSAR model and its validation

Previously, it has been reported that the Bortezomib; a dipeptidyl boronic acid and its derivatives, Glycyrrhizin acid, Betulinic acid, and their derivatives are strong, potent, effective, and reversible proteasome antagonists. Furthermore, experimental *in vitro* antagonist activity ( $IC_{50}$ ) of known proteasome inhibitors were implemented to develop a statistically significant model using a multiple linear regression approach. The structure-activity relationship model yielded satisfactory and significant correlation between biological activity-chemical descriptors in the statistical form of regression coefficient  $r^2 = 0.83$ , the reliability and the robustness for the derived linear model was assessed from the cross-validated correlation coefficient i.e.  $q^2 = 0.80$ , represented in a regression plot Figure 2. The regression plot between observed vs predicted activity provides knowledge about the dataset fitness and the dataset activity accuracy for the external test set. The external extrapolative prediction strength of the 2D-model was (pred\_ $r^2$ ) 0.77 with a degree of freedom 51. The regression graph of the modeling set including training and test set shows a simple relationship through a straight regression line between two variables i.e. dependent and independent variable. The observed straight line found to have best and proper fitness of the entire

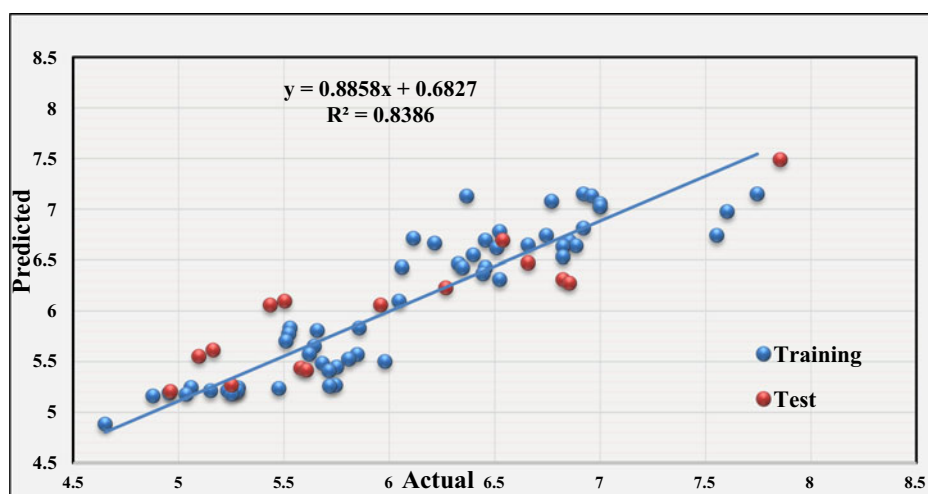


Figure 2. Regression plot between actual and predicted activity of compounds.

biological dataset. In the fitness plot, the value of  $q^2$  (cross-validation) with 0.80 signifies that the compounds of the training set, shown by blue dot have not exhibited any statistical dissonance. As well, the regression  $r^2$  of 0.83 (test set as a red dot) indicates the decent predictability of the derived model towards untested or unknown molecules along with small standard error that illustrates the comprehensiveness of the biological dataset. Moreover, the  $\text{pred}_r^2 = 0.77$ , Fisher values = 65.34 and Z-score values of 7.67 ( $r^2$ ) & 6.89 ( $q^2$ ) statistical potentials established the good and robust quality of the derived model.

### Stepwise forward-backward

The stepwise forward-backward is a potent, magnificent, and long exhausted method in the QSAR model development. In this method, descriptors as independent variables have a strong correlation with the biological activity or experimental  $\text{IC}_{50}$  values were included step by step to a regression equation and importance of each incorporated descriptors was analyzed to remain in the final multivariate equation.

### Model statistical evaluation

Model validation is a precise and crucial phase in model development. Consequently, different kinds of validations were accomplished and the best and robust 2D-model was obtained employing the various statistical limits of parameters (Cramer, Bunce, Patterson, & Frank, 1988; Golbraikh & Tropsha, 2002). Validations were accomplished for the tested alignment which is listed as, e.g., (i) Coefficient of determination ( $r^2$ ) must be  $> 0.7$ , (ii) Leave-one-out cross-validation (LOOcv) or correlation coefficient ( $q^2$ ) must be  $> 0.5$  (iii)  $\text{pred}_r^2$ , a Correlation coefficient of external prediction set must be  $> 0.5$  (Tropsha, Gramatica, & Gombar, 2003). This is essential to validate the reliability and ability of the developed QSAR model to predict the bioactivity of somehow similar compounds. (iv) degree of freedom must be higher. (vi) higher F-test value for statistical importance of model (Same set of compounds and descriptors). (vii) standard error of estimate i.e.  $r^2_{\text{se}}$  and  $q^2_{\text{se}}$  should be

smaller. (viii)  $\alpha_{\text{test}}$ , a statistical parameter evaluated by randomization must be  $< 0.1$ . (ix) Z-score calculated by randomization must be higher. The predictive ability and strength of the derived model are analyzed by calculating the cross-validated regression coefficient ( $q^2$ ) using the following mathematical equation 1 (Shen et al., 2002), where  $Y_{\text{act}}$  and  $Y_{\text{pred}}$  denote the actual and predicted  $\text{pIC}_{50}$  values for  $i^{\text{th}}$  compound respectively, however  $Y_{\text{mean}}$  signifies the average of actual activity ( $\text{pIC}_{50}$ ) value of entire compounds of training set.

$$q^2 = 1 - \frac{\sum (Y_{\text{act}} - Y_{\text{pred}})^2}{\sum (Y_{\text{act}} - Y_{\text{mean}})^2} \quad (1)$$

The regression coefficient ( $\text{pred}_r^2$ ) for external test set was calculated by employing the equation (2) (Kier & Hall, 1977; Cramer et al., 1988). The  $\text{pred}_r^2$  validate the predictive strength and ability of the derived QSAR model for external query set and verify the robustness and reliability of the predicted result.

$$\text{Pred}_r^2 = 1 - \frac{\sum (Y_{\text{act}} - Y_{\text{pred}})^2}{\sum (Y_{\text{act}} - Y_{\text{pred}})^2} \quad (2)$$

The observed satisfactory standard error of the QSAR model for  $r^2$  and  $q^2$  were 0.30 and 0.33 respectively along with  $\text{pred}_r^2_{\text{se}}$  of 0.39. Hence, the minimum standard error of  $r^2_{\text{se}}$  and  $q^2_{\text{se}}$  revealed the overall features and qualities of the QSAR model. Additional statistical evaluation of the model was observed through Z-scores which was indicated 7.67 and 6.89 for  $r^2$  and  $q^2$  respectively and Fischer's value ( $F_{\text{test}}$ ) i.e. 65.34 was found to have statistically significant values for the derived model and presented by Table 1. The random distribution pattern indicated that the developed QSAR model provided a decent and proper fitness of the data (Figure 3). The plot indicated the random distribution of compounds to analyze the regression value of the derived MLR model. The graphical representation of the residual values (training and test set) explains the degree of correlation between actual and predicted values of the developed model. The distribution of the residuals on both sides of the middle X-axis line (Error line range  $-1$  to  $1$ ) which is indicating actual values of the dataset. The standard error of the

derived model range from 0.3 to  $-0.243$ , that indicates the acceptable error of the derived MLR model.

### Outlier's recognition

In the developed QSAR model, the compounds which have unpredicted and unexpected biological activity, do not follow the applicability domain range and are not able to fit in the model are recognized as outliers. Therefore, the compounds that are found as outliers in the modeling set were removed from the dataset and query set compounds. The resultant outliers may have a different mode of action for a particular receptor/target and act by a different mechanism. Removal of outliers are valuable in signifying the experimental limits of the biological dataset and also define the common mechanism of action of compounds which is modeled by more than one descriptors (Verma & Hansch, 2005).

### Contribution of molecular descriptors

The contribution of chemical descriptors correlated to anti-cancer activity was recognized in the resultant QSAR model. The model revealed that four molecular descriptors, namely Electronegativity count, Average

Potential, T<sub>2</sub>N<sub>6</sub>, and H-Acceptor Count have shown a significant correlation with the biological activity. So far, descriptors involved to govern the anti-cancer property is presented by equation 3.

$$\begin{aligned} \text{Predicted Log pIC}_{50}(\mu\text{M}) &= 8.5128 (\pm 0.3976) \\ &\times \text{Electronegativity Count} \\ &- 30.2378 (\pm 10.7260) \\ &\times \text{Average Potential} \\ &0.1875 (\pm 0.0256) \times \text{T}_2\text{N}_6 \\ &- 0.2500 (\pm 0.0482) \\ &\times \text{H-Acceptor Count} \\ &+ 0.002 \end{aligned} \quad [3]$$

The statistically derived linear model significantly showed a strong positive correlation with the Electronegativity Count and T<sub>2</sub>N<sub>6</sub> and negatively correlation with Average Potential and H-Acceptor Count. Moreover, T<sub>2</sub>N<sub>6</sub> is topological chemical descriptor which defined that the number of double-bonded atoms (any double-bonded atom, T<sub>2</sub>) separated from nitrogen atom by six bonds in a molecule. The descriptors contribution of the model is represented in Figure 4. Moreover, the graph (Figure 5A and B) demonstrates the fitness and pertinence of actual versus predicted values of entire compounds for biological activity.

### Ligand-based drug likeness screening (analogs)

Conclusively, to identify the best and safe antagonist, a library of 3049 (Query set) compounds were virtually screened through PubChem database with 80% similarity to the drug and terpenoids for proteasome targeting NF- $\kappa$ B pathway inhibition.

Additionally, these predicted compounds were screened for bio-activity prediction using the derived QSAR model.

### Prediction set screening through thumb rule of five and ADMET risk parameters

Poor or undesirable pharmacokinetic properties of drugs or compounds are a major cause of the cessation of a compound's progression in the drug development pipeline. Successively, these query set were again filtered through Lipinski's rule of five directed to assess the chemical properties concerning oral bioavailability. Out of 3049 prediction set compounds, only twenty compounds were found to follow the standard limit of the drug-likeness features. Later, top screened compounds were analyzed through the pharmacokinetic compliance: Absorption, Distribution, Metabolism, Excretion, and Toxicity to find a drug-like lead compound with good pharmacokinetic properties. The top-most hit compounds have been found to compatible with that of standard drug Bortezomib (Yadav, Nath Mishra, & Khan, 2018). To evade the late-phase failure, calculation of pharmacokinetic parameters of the predicted compounds are essential in the early drug development practice.

### Assessment of drug-likeness for oral bioavailability

Furthermore, Lipinski's rule of five was applied with the aim of calculation of drug-like physiochemical properties of best candidates. As far as, drug-likeness is defined as a qualitative principle used in the drug design to understand how molecules or candidate compounds can show drug-like properties concerning bioavailability. It is crucial to recognize the different physiochemical properties that are responsible to make a compound biologically active and that would work as an orally stable and active molecule in the biological system. So far, every predicted hit compound cannot be treated as a therapeutic compound before any confirmation and validation of some pharmacokinetic parameters and toxicity using *in silico* approach. A traditional approach to calculate the drug-like properties is to analyze the amenability of the rule of thumb i.e. Lipinski's Rule of Five (Lipinski, Lombardo, Dominy, & Feeney, 2001).

To complete this purpose, the aqueous solubility, intestinal absorption, hepatotoxicity, blood-brain penetration (BBB) and plasma protein binding (PPB) along with CYP2D6 binding of the potential compounds (predicted) were calculated simultaneously. Though, the thumb rule could not predict whether a compound is pharmacologically active or not. This thumb rule is significant in the drug designing and development process wherever an active and potential lead could be optimized in a step-wise manner for enhanced biological activity and specificity against target/protein. The crucial role of the rule mainly focus on the information that the orally managed and controlled drugs must contain a molecular weight less than 500 Da, AlogP i.e. an octanol-water partition coefficient less than 5, H-bond acceptor and donor of 10 and 5 respectively (Table 2). Also, the oral bioavailability of all predicted compounds and drugs were measured through the polar surface area (PSA) that was calculated



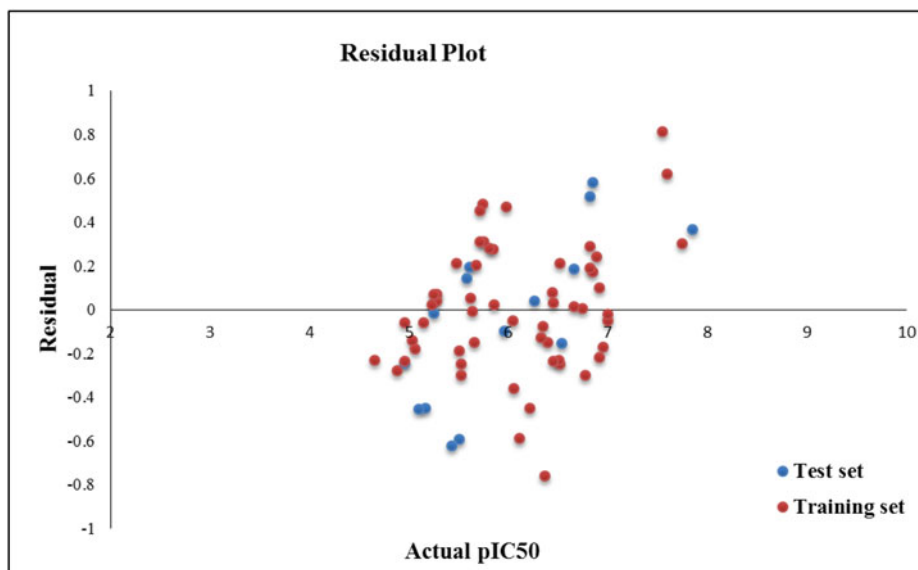


Figure 3. Plot of residuals for both training and test sets against the experimental pIC50 values.

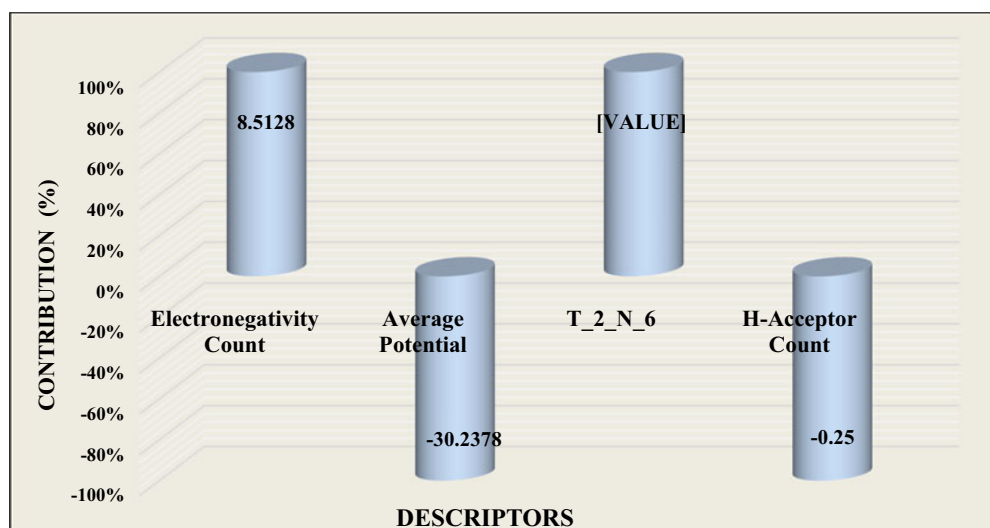


Figure 4. Contribution of chemical descriptors correlated well with anticancer biological activity showing a positive correlation with the Electronegativity Count and T\_2\_N\_6 and negatively correlation with Average Potential and H-Acceptor Count.

using a methodology, centered on the addition tabularized surface assistances of polar fragments viz. topological polar surface (TPSA). The PSA was found to link substantially with unrecaptive and passive transport across the cell membranes and hence, permits the calculation and extrapolation of transport properties of compounds or drugs and was related to drug bioavailability. Usually, it was observed that the passively/inactively absorbed compounds with less than 140 Å of PSA value have shown to have low oral bioavailability (Ertl, Rohde, & Selzer, 2000; Maurya, Khan, Bawankule, Yadav, & Srivastava, 2012). Therefore, predicted compounds were subjected to screened out by utilizing the filtering parameters of Lipinski's rule of five which is principal and key screening practice as well as pharmacokinetic evaluation; a secondary screening process. Results indicated that NP and AP predicted hits of the query set were found to settle within the satisfactory limits that possess all the physiochemical properties liable for the oral bioavailability. Moreover, it was noticed that NP and AP were not violated any

rule of Lipinski as similar to the standard drug of the target. In the perspective of pharmacokinetics, the ADME parameters (Absorption, Distribution, Metabolism, and Excretion) such as aqueous solubility, serum protein binding, blood-brain barrier penetration, hepatotoxicity and intestinal absorption of best candidates were calculated and compared with the control drug, presented in Table 3. As per the study, NP (natural) and AP (analog) showed good intestinal absorption as compared to bortezomib. The control drug indicated moderate intestinal absorption and undefined blood-brain barrier penetration.

#### **In silico toxicity risk valuation**

As a final point, the toxicity risk calculation is important in the screening of best and effective drug or compound in the early drug development process in respects to the safety concern. *In silico* compliance with toxicity, the calculation is not only rapid but it may similarly reduce the cases of animal

experiments. To capture this, the different parameters of toxicity such as TD50 (tumorigenic dose), LD50 (lethal dose), mutagenicity, skin irritancy, aerobic biodegradability, and ocular irritancy were calculated (Table S2). Results indicated that the potential hits namely AP (analog) and NP (natural) possess lethal dose (LD<sub>50</sub>) values of 4.74789 and 3.97152 (g/kg body weight) respectively. As per toxicity assessment, NP and AP have mild skin irritancy, weak skin sensitization, and non-mutagenicity as similar to control drugs Carfilzomib and Bortezomib. In the context of aerobic biodegradability, NP and AP were found to possess biodegradability rather than control drugs that have not shown the aerobic biodegradability

property. Additionally, a rat concentrated tolerated dosage for NP and AP are 0.0755808 & 0.104032 g/kg body weight, however, Carfilzomib and Bortezomib indicated 0.0746945 and 0.440114 respectively. The observed *Fathead Minnow* LC50 (lethal concentration) values of NP and AP are 0.964428 and 0.00758264 g/l whereas in case of drugs it is 0.000111043 (Carfilzomib) and 0.0431359 (Bortezomib). Likewise, *Daphnia magna* EC50 (Effective concentration) is 12.0665 (NP) and 5.62132 (AP) but in the drugs, it has found 0.0533036 and 0.414224 (mg/l). Therefore, results of toxicity prediction suggested that the NP and AP retain non-mutagenicity of chromosomal abnormalities and aerobic biodegradability, hence these two predicted best compounds are non-persistent, non-toxic and eco-friendly. As a result, the computationally calculated toxicity parameters of the topmost candidates could be examined and observed with the assistance of dose limited studies.

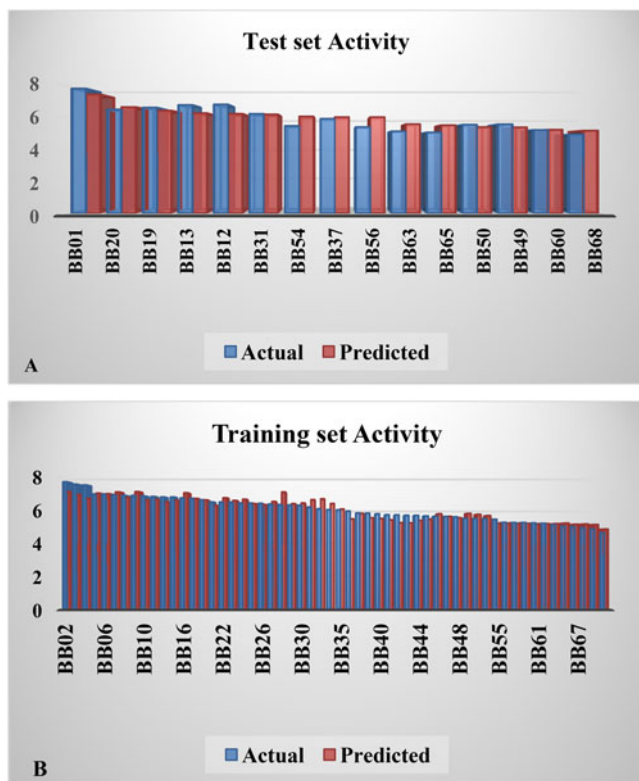


Figure 5. (A) Actual versus Predicted activity for the Training set (B) Actual versus Predicted activity for the Test set.

### Molecular docking studies for potential target

Molecular docking studies of most potent anticancer compounds were performed to explore their respective binding mode of action and binding affinity with the target protein. Hence, all predicted compounds through the derived QSAR model were further screened through molecular docking studies. Conclusively, the crucial objective of docking is to light up the interaction of best effective hit which governs the anticancer activity against the specified target i.e. human 20S proteasome. The docking procedure was standardized through the re-docking study of the co-crystallized drug (Bortezomib) with the proteasome target. The optimized molecules were shown the various interactions like hydrogen bonding, Pi-Pi interaction and van der Waal's interaction. The hydrogen bonds and pi-pi interaction between the target and hit compounds may be considered as a complex stabilizer. Thus far, the docking results revealed that the predicted compounds NP and AP indicate the most remarkable molecular interaction between inhibitors and target. The docking simulation suggested that the predicted compounds namely NP and AP signify the highest negative docking score of  $-23.85$  and  $-25.14$  respectively among all the

Table 2. *In silico* Lipinski's Rule compliance of top hits for oral bioavailability.

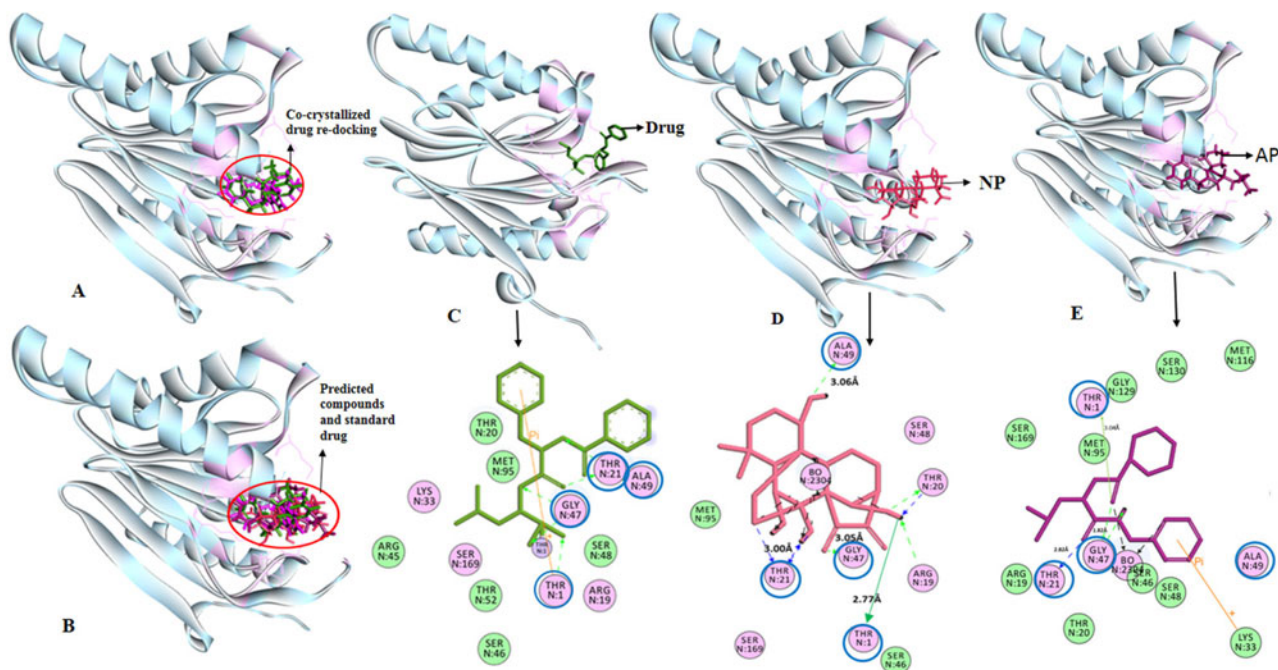
Compound	Molecular weight ( $\leq 500$ Da)	Oral Bioavailability: TPSA ( $\leq 140$ Å)	H- Bond Donor ( $\leq 5$ )	H- Bond Acceptor ( $\leq 10$ )	AlogP ( $\leq 5$ )	Rotatable Bond ( $\leq 10$ )	Rule of 5 violation
Bortezomib (Reference drug)	384.247	124	4	6	2.58	9	0
Carfilzomib	719.93	159	4	8	3.79	20	3
NP	364.189	109.42	4	6	0.2	0	0
AP	331.462	71.483	2	3	2.92	7	0

Table 3. *In silico* pharmacokinetic assessment of predicted hit compounds (terpenoids and analogs).AQ4

Compounds	Aqueous solubility	Blood-brain barrier (BBB) penetration	CYP2D6 Prediction	Hepatotoxic Prediction	Intestinal absorption	Plasma protein binding
Bortezomib (Reference drug)	3 (Good)	4 (undefined)	FALSE (non-inhibitor)	FALSE (non-toxic)	1 (moderate)	FALSE (poorly bounded)
Carfilzomib	3	4	FALSE	FALSE	3 (very poor)	FALSE
NP	4 (Very good)	3 (low)	FALSE	FALSE	0 (Good)	TRUE (highly bounded)
AP	3	2 (Medium)	FALSE	FALSE	0	TRUE

**Table 4.** Comparison of binding affinity of control drug and predicted natural and analog in terms of docking score and binding site residues against anticancer proteasome target (PDB: 5LF3).

S.No.	Compound	Docking Score	Residues involved in H-bonding	Residues interacted and H-bond length	Binding pocket residues within 4Å region
1	AP	-25.14	Thr1, Gly47, Thr21	Gly47:O-1.82Å Thr1:O-3.04Å Thr21:O2.82Å	Gly129, Sr130, Met 116, Ser48, Thr20, Arg19, Ala49, Ser169, Ser48, Met95, Ser169, Ser46
2	NP	-23.85	Thr1, Thr21, Thr20, Arg19, Gly47, Ala49	Gly47:O-3.05Å Thr21:O-3.00Å Thr1:O-2.77Å Ala49:O-3.06Å	Thr20, Ser48, Arg19, Thr52, Ser46, Lys33, Arg45, Ser169
3	Bortezomib (Control drug)	-28.2	Thr1, Thr21, Gly47, Ala49	-----	Thr20, Ser48, Arg19, Thr52, Ser46, Lys33, Arg45, Ser169

**Figure 6.** Docking pose and binding pocket residues with their best favorable conformation of top hit compounds and control drug against human proteasome complex (PDB: 5LF3). (A) Docking protocol standardization by re-docking of co-crystallized ligand (Bortezomib drug) against human proteasome complex. (B) Binding pocket site residues with the best fit conformation of predicted and control compounds. (C) Bortezomib (standard drug) docked with binding pocket site residues (D) Docked conformation of NP (natural) with binding pocket residues. (E) Docked conformation of AP (analog) with binding pocket residues.

predicted compounds as compared to that of reference drug bortezomib i.e.  $-28.2$ . Noteworthy, molecular interaction of these potential hits along with standard drug revealed that Thr20, Thr21, Ala 49, Gly47 and Thr1 are common crucial key residues located within the active site (binding pocket) of the protein/target. The natural predicted compound NP showed the four hydrogen bonds with a measured distance of 2.77Å, 3.06Å, and 3.05Å with Thr20, Thr21, Ala49, and Gly47 respectively. The highest negative docking score showed that the target protein was docked effectively with predicted hit compounds (Table 4). The promising binding mode of action of best natural compounds and analog at active sites is demonstrated in Figure 6. The common crucial residues of target protein involved in the formation of hydrogen bonding are Gly47, Ala49, Thr21, and Thr1. The information showed that the proteasome active pocket is occupied by all the predicted compounds but NP and AP show highest docking score and similarity of the active site with that of the complex molecule (Drug with target protein). Moreover, result signified that the predicted best hits showed the most positive and favorable environment inside a defined and specific receptor cavity that might develop a pharmacological reaction or response. So far, the two-

dimensional interaction part was also observed in order to identify particular amino acid residues of the target and atom of the antagonist participated in the molecular interaction. Though, a 2D graph examination of docking insight indicated the numerous molecular communications through hydrogen bonding, Pi-sigma interaction, etc. between predicted hits and the surrounding residues of the target in order to reveal the mode of action which is liable for the pharmacological response and defined receptor binding.

Based on this docking study, it could be anticipated that both the active hits NP and AP adopt a satisfactory conformation within the active binding pocket of the protein and significant molecular interactions were observed very well from the 2D and 3D figures of the interaction. These inferences would support in direction of rational designing of novel and more potential analogs or derivatives against cancer.

## Conclusion

In the current study, a QSAR study was done on 71 anti-cancer compounds against human proteasome target. The MLR QSAR model would help as a tool for virtual screening,

understanding of the anticancer lead identification against cancer. The established model indicated a strong correlation between actual and predicted endpoints of the biological activity/anticancer activity that specified the robustness and validity of the derived linear model. Likewise, distinguished and defined the crucial role of significant chemical descriptors which govern the biological activity. *In silico* methodologies were adopted in this study to screen out top hit best, effective and safe compounds namely NP and AP. Besides, the top hit candidates were validated through evaluation of compliance with pharmacokinetic parameters simultaneously with Lipinski's rule of five for drug-likeness and oral bioavailability. Finally, the predicted hit compounds were further directed to molecular docking studies in order to explore the molecular mode of action and putative binding active site for the proteasome target of NF- $\kappa$ B pathway. Also, the binding affinity and conformation of the candidate compounds, NP and AP were found to have decent docking score of  $-23.85$  and  $-25.14$  respectively and found as good as the standard drug, Bortezomib with a score of  $-28.2$ . Noteworthy, NP, AP and control drug revealed the promising binding mode of action and four hydrogen bonds with common key residues of Gly47, Ala49, Thr21, and Thr1 in the putative active pocket of the protein. This information also highlighted that the proteasome binding pocket was captured by entire predicted active compounds but NP and AP showed the highest docking score and similarity of the active site with that of the standard drug. Hence, the two predicted hit candidates *viz.*, NP and AP were anticipated to be the favorable and promising compound against proteasome anticancer target. As far as we know, it is conceivably the first work reported on this subject/topic using *in silico* approaches. Henceforward, the employed studies could be applied and explored as a pattern work for modern drug designing process in the future. Moreover, the results obtained would be of great support in advance and new anticancer drug discovery procedure, and drug-like lead identification and optimization based on the natural active scaffold.

### Acknowledgment

The authors are thankful to the Director, CSIR-CIMAP, Lucknow India, for providing required research facilities and support. Author DY is thankful to Indian Council of Medical Research (ICMR), New Delhi for financial support through Senior Research Fellowship (SRF) (Ref. No.: 5/3/8/10/ITR dated 09/04/2018) at CSIR-CIMAP, Lucknow, India. The CSIR-CIMAP communication no. is **CIMAP/2019/05/43**.

### Disclosure statement

The authors declare no conflict of interest.

### Funding

This research work was financially supported by the Indian Council of Medical Research (ICMR), New Delhi [Ref. No.: 5/3/8/10/ITR-F/2018-ITR] at CSIR-CIMAP, Lucknow, India.

### References

- Abagyan, R. (2018). Retrieved from <http://www.molsoft.com/icm-chemist-pro.html>.
- Adams, J. (2004). The development of proteasome inhibitors as anti-cancer drugs. *Cancer Cell*, 5(5), 417–421. doi:10.1016/S1535-6108(04)00120-5
- Adams, J. (2003). The proteasome: Structure, function, and role in the cell. *Cancer Treatment Reviews*, 29, 3–9. doi:10.1016/S0305-7372(03)00081-1
- Arlt, A., Bauer, I., Schafmayer, C., Tepel, J., Mürköster, S. S., Brosch, M., ... Schäfer, H. (2009). Increased proteasome subunit protein expression and proteasome activity in colon cancer relate to an enhanced activation of nuclear factor E2-related factor 2 (Nrf2). *Oncogene*, 28(45), 3983. doi:10.1038/onc.2009.264
- Bajusz, D., Rácz, A., & Héberger, K. (2015). Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of Cheminformatics*, 7(1), 20. doi:10.1186/s13321-015-0069-3
- Cao, B., & Mao, X. (2011). The ubiquitin-proteasomal system is critical for multiple myeloma: Implications in drug discovery. *American Journal of Blood Research*, 1, 46.
- Cardoso, F., Ross, J. S., Piccart, M. J., Sotiriou, C., & Durbecq, V. (2004). Targeting the Ubiquitin-Proteasome Pathway in Breast Cancer. *Clinical Breast Cancer*, 5(2), 148–157. doi:10.3816/CBC.2004.n.020
- Chen, D., Frezza, M., Schmitt, S., Kanwar, J., & P. Dou, Q. (2011). Bortezomib as the first proteasome inhibitor anticancer drug: Current status and future perspectives. *Current Cancer Drug Targets*, 11(3), 239–253. doi:10.2174/156800911794519752
- Chen, F. Z., & Zhao, X. K. (2013). Ubiquitin-proteasome pathway and prostate cancer. *Oncology Research and Treatment*, 36, 592–596. doi:10.1159/000355166
- Cramer, III, R. D., Bunce, J. D., Patterson, D. E., & Frank, I. E. (1988). Cross-validation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quantitative Structure-Activity Relationships*, 7(1), 18–25. doi:10.1002/qsar.19880070105
- Crawford, L. J., Walker, B., Ovaa, H., Chauhan, D., Anderson, K. C., Morris, T. C., & Irvine, A. E. (2006). Comparative selectivity and specificity of the proteasome inhibitors, PS-341, and MG-132. *Cancer Research*, 66(12), 6379–6386. doi:10.1158/0008-5472.CAN-06-0605
- Dawson, S. P. (2008). Hepatocellular carcinoma and the ubiquitin-proteasome system. *Biochimica et Biophysica Acta (Bba) - Molecular Basis of Disease*, 1782(12), 775–784. doi:10.1016/j.bbdis.2008.08.003
- Delic, J., Masdehors, P., Ömura, S., Cosset, J.-M., Dumont, J., Binet, J.-L., & Magdelénat, H. (1998). The proteasome inhibitor lactacystin induces apoptosis and sensitizes chemo- and radioresistant human chronic lymphocytic leukaemia lymphocytes to TNF- $\alpha$ -initiated apoptosis. *British Journal of Cancer*, 77(7), 1103. doi:10.1038/bjc.1998.183
- Dick, L. R., & Fleming, P. E. (2010). Building on bortezomib: Second-generation proteasome inhibitors as an anti-cancer therapy. *Drug Discovery Today*, 15(5–6), 243–249. doi:10.1016/j.drudis.2010.01.008
- Escobar, M., Velez, M., Belalcazar, A., Santos, E. S., & Raez, L. E. (2011). The role of proteasome inhibition in nonsmall cell lung cancer. *Journal of Biomedicine and Biotechnology*, 2011, 1–10. doi:10.1155/2011/806506
- Ertl, P., Rohde, B., & Selzer, P. (2000). Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *Journal of Medicinal Chemistry*, 43(20), 3714–3717. doi:10.1021/jm000942e
- Jang, H. H. (2018). Regulation of Protein Degradation by Proteasomes in Cancer. *Journal of Cancer Prevention*, 23(4), 153. doi:10.15430/JCP.2018.23.4.153
- Gilmore, T. D. (1999). Multiple mutations contribute to the oncogenicity of the retroviral oncoprotein v-Rel. *Oncogene*, 18(49), 6925. doi:10.1038/sj.onc.1203222
- Golbraikh, A., & Tropsha, A. (2002). Beware of q<sup>2</sup>!. *Journal of Molecular Graphics and Modelling*, 20, 269–276. doi:10.1016/S1093-3263(01)00123-1



- Harvey, A. L., Edrada-Ebel, R., & Quinn, R. J. (2015). The re-emergence of natural products for drug discovery in the genomics era. *Nature Reviews Drug Discovery*, 14(2), 111. doi:10.1038/nrd4510
- Huang, L., Yu, D., Ho, P., Qian, K., Lee, K. H., & Chen, C. H. (2008). Synthesis and proteasome inhibition of glycyrrhetic acid derivatives. *Bioorganic & Medicinal Chemistry*, 16, 6696–6701. doi:10.1016/j.bmc.2008.05.078
- Kazi, A., Lawrence, H., Guida, W. C., McLaughlin, M. L., Springett, G. M., Berndt, N., ... Sebti, S. M. (2009). Discovery of a novel proteasome inhibitor selective for cancer cells over non-transformed cells. *Cell Cycle*, 8(12), 1940–1951. doi:10.4161/cc.8.12.8798
- Kier, L. B., & Hall, L. H. (1977). Nature of structure-activity-relationships and their relation to molecular connectivity. *Journal of Medicinal Chemistry*, 12, 307–312. doi:10.1021/jm00222a019
- Kravtsova-Ivantsiv, Y., & Ciechanover, A. (2015). The ubiquitin-proteasome system and activation of NF- $\kappa$ B: Involvement of the ubiquitin ligase KPC1 in p105 processing and tumor suppression. *Molecular & Cellular Oncology*, 2, 1054552. doi:10.1080/23723556.2015.1054552
- Kubiczkova, L., Pour, L., Sedlarikova, L., Hajek, R., & Sevcikova, S. (2014). Proteasome inhibitors—molecular basis and current perspectives in multiple myeloma. *Journal of Cellular and Molecular Medicine*, 18(6), 947–961. doi:10.1111/jcmm.12279
- Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 46(1-3), 3–26. doi:10.1016/S0169-409X(00)00129-0
- Luque Ruiz, I., & Gómez-Nieto, M. (2018). Study of the Applicability Domain of the QSAR Classification Models by Means of the Rivality and Modelability Indexes. *Molecules*, 23(11), 2756. doi:10.3390/molecules23112756
- Milo, L. J., Jr, Lai, J. H., Wu, W., Liu, Y., Maw, H., Li, Y., ... Sanford, D. G. (2011). Chemical and biological evaluation of dipeptidyl boronic acid proteasome inhibitors for use in prodrugs and pro-soft drugs targeting solid tumors. *Journal of Medicinal Chemistry*, 54(13), 4365–4377. doi:10.1021/jm200460q
- Maurya, A., Khan, F., Bawankule, D. U., Yadav, D. K., & Srivastava, S. K. (2012). QSAR, docking and in vivo studies for the immunomodulatory activity of isolated triterpenoids from *Eucalyptus tereticornis* and *Gentiana kurroo*. *European Journal of Pharmaceutical Sciences*, 47(1), 152–161. doi:10.1016/j.ejps.2012.05.009
- Mitra, S. (2018). 'Ubiquitin Signaling in Ovarian Cancer: From Potential to Challenges' in Deveja, O. (ed.) *Ovarian Cancer: From Pathogenesis to Treatment*, Intechopen Limited, London, U.K., Chapter 6, pp. 135–154. doi:10.5772/intechopen.75485
- Moreau, P., Richardson, P. G., Cavo, M., Orlowski, R. Z., San Miguel, J. F., Palumbo, A., & Housseau, J. L. (2012). Proteasome inhibitors in multiple myeloma: 10 years later. *Blood*, 120(5), 947–959. doi:10.1182/blood-2012-04-403733
- Nalepa, G., Rolfe, M., & Harper, J. W. (2006). Drug discovery in the ubiquitin-proteasome system. *Nature Reviews Drug Discovery*, 5(7), 596. doi:10.1038/nrd2056
- Orlowski, R. Z., & Dees, E. C. (2002). The role of the ubiquitination-proteasome pathway in breast cancer: Applying drugs that affect the ubiquitin-proteasome pathway to the therapy of breast cancer. *Breast Cancer Research*, 5(1), 1–7. doi:10.1186/bcr460
- Pagadala, N. S., Syed, K., & Tuszynski, J. (2017). Software for molecular docking: A review. *Biophysical Reviews*, 9(2), 91–102. doi:10.1007/s12551-016-0247-1
- Palombella, V. J., Rando, O. J., Goldberg, A. L., & Maniatis, T. (1994). The ubiquitin-proteasome pathway is required for processing the NF- $\kappa$ B1 precursor protein and the activation of NF- $\kappa$ B. *Cell*, 78(5), 773–785. doi:10.1016/S0092-8674(94)90482-0
- Pickart, C. M. (2004). Back to the future with ubiquitin. *Cell*, 116(2), 181–190. doi:10.1016/S0092-8674(03)01074-2
- Pikarsky, E., Porat, R. M., Stein, I., Abramovitch, R., Amit, S., Kasem, S., ... Ben-Neriah, Y. (2004). NF- $\kappa$ B functions as a tumour promoter in inflammation-associated cancer. *Nature*, 431(7007), 461. doi:10.1038/nature02924
- Qian, K., Kim, S. Y., Hung, H. Y., Huang, L., Chen, C. H., & Lee, K. H. (2011). New betulonic acid derivatives as potent proteasome inhibitors. *Bioorganic & Medicinal Chemistry Letters*, 21, 5944–5947. doi:10.1016/j.bmcl.2011.07.072
- Ronan, T., Qi, Z., & Naegle, K. M. (2016). Avoiding common pitfalls when clustering biological data. *Sci. Signal*, 9, 432. doi:10.1126/scisignal.aad1932
- Schrader, J., Henneberg, F., Mata, R. A., Tittmann, K., Schneider, T. R., Stark, H., ... Chari, A. (2016). The inhibition mechanism of human 20S proteasomes enables next-generation inhibitor design. *Science*, 353(6299), 594–598. doi:10.1126/science.aaf8993
- Shukla, A., Sharma, P., Prakash, O., Singh, M., Kalani, K., Khan, F., ... Srivastava, S. K. (2014). QSAR and docking studies on capsazepine derivatives for immunomodulatory and anti-inflammatory activity. *PLoS One*, 9(7), e100797. doi:10.1371/journal.pone.0100797
- Shen, M., LeTiran, A., Xiao, Y., Golbraikh, A., Kohn, H., & Tropsha, A. (2002). Quantitative structure-activity relationship analysis of functionalized amino acid anticonvulsant agents using k nearest neighbor and simulated annealing PLS methods. *Journal of Medicinal Chemistry*, 45(13), 2811–2823. doi:10.1021/jm010488u
- Spataro, V., Norbury, C., & Harris, A. L. (1998). The ubiquitin-proteasome pathway in cancer. *British Journal of Cancer*, 77(3), 448. doi:10.1038/bjc.1998.71
- Tropsha, A., Gramatica, P., & Gombar, V. K. (2003). The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *Qsar & Combinatorial Science*, 22, 69–77. doi:10.1002/qsar.200390007
- Tu, Y., Chen, C., Pan, J., Xu, J., Zhou, Z. G., & Wang, C. Y. (2012). The Ubiquitin-Proteasome Pathway (UPP) in the regulation of cell cycle control and DNA damage repair and its implication in tumorigenesis. *International Journal of Clinical and Experimental Pathology*, 5(8), 726.
- Verma, R. P., & Hansch, C. (2005). An approach toward the problem of outliers in QSAR. *Bioorganic & Medicinal Chemistry*, 13, 4597–4621. doi:10.1016/j.bmc.2005.05.002
- Voutsadakis, I. A. (2017). Proteasome expression and activity in cancer and cancer stem cells. *Tumor Biology*, 39(3), 101042831769224. doi:10.1177/1010428317692248
- Wang, X., Yen, J., Kaiser, P., & Huang, L. (2010). Regulation of the 26S proteasome complex during oxidative stress. *Science Signaling*, 3(151), ra88. doi:10.1126/scisignal.2001232
- Xiao, G., Harhaj, E. W., & Sun, S. C. (2001). NF- $\kappa$ B-inducing kinase regulates the processing of NF- $\kappa$ B2 p100. *Molecular Cell*, 7(2), 401–409. doi:10.1016/S1097-2765(01)00187-3
- Xie, P., Fan, Y., Zhang, H., Zhang, Y., She, M., Gu, D., ... Li, H. (2009). CHIP represses myocardin-induced smooth muscle cell differentiation via ubiquitin-mediated proteasomal degradation. *Molecular and Cellular Biology*, 29(9), 2398–2408. doi:10.1128/MCB.01737-08
- Yadav, D., Nath Mishra, B., & Khan, F. (2018). 3D-QSAR and docking studies on ursolic acid derivatives for anticancer activity based on bladder cell line T24 targeting NF- $\kappa$ B pathway inhibition. *Journal of Biomolecular Structure and Dynamics*, 37(14), 3822–3837. doi:10.1080/07391102.2018.1528888
- Yang, H., & Ping Dou, Q. (2010). Targeting apoptosis pathway with natural terpenoids: Implications for treatment of breast and prostate cancer. *Current Drug Targets*, 11(6), 733–737. doi:10.2174/138945010791170842