ORIGINAL RESEARCH ARTICLES

2D QSAR ANALYSIS OF CARBONITRILE BASED INHIBITORS OF CATHEPSIN S AS POTENTIAL ANTIRHEUMATIC AGENTS

Sneha Kushwaha[a]*, Sarvesh K. Paliwal[b] and Divya Niranjan[c]

(Received 05 March 2022) (Accepted 27 October 2022)

ABSTRACT

2D QSAR has been performed on a series of pyridine carbonitrile and trifluoromethyl phenyl derivatives. 53 compounds were divided into training and test sets out of which 37 compounds generated a final QSAR model. The most significant model with n = 37, r = 0.916, r^2 = 0.762, r^2cv = 0.759, s value = 0.388, f value = 41.76 was developed using MLR analysis. For PLS, the fraction of variance explained = 0.806 was observed. A comparable PLS model with r^2 = 0.806 and Neural model with r^2 = 0.853 indicated good internal predictability of the model. External test set validation provided r^2 values of 0.744 and 0.768 for MLR and PLS analysis, respectively. Dipole moment Z Component, Log P, Shape flexibility index, and Vamp LUMO descriptors proved to be significant for inhibition of Cathepsin S. These findings will be effective in designing more potent and effective Cathepsin S inhibitors.

Keywords: QSAR, Multiple linear regression, Partial least square

INTRODUCTION

The word "Cathepsin" has been derived from the Greek word 'kathepsein', which means "to digest"[1,2]. The enzyme came into light in the 20th century[3]. Eleven human cysteine cathepsins are expressed in the human genome[4]. Cathepsins L, V, S, K, and F are endopeptidases, while cathepsins X, B, C, and H are exopeptidases. Cathepsins O and W are of unknown category[1,5,6].

The gene symbol of cathepsin S is CTSS. It is a non-glycosylated cysteine proteinase. It belongs to the clan C1 (papain family)[7,8]. These enzymes are prominently situated intracellularly in the endolysosomal vesicles[1,4,9]. These are exclusively situated in the dendritic cells, macrophages, spleen, lymph nodes, monocytes and/or thymic cortical epithelial cells[10,11]. The enzyme is majorly involved in antigen processing and presentation[12-14].

All cysteine proteases are made up of a signal peptide, a propeptide, and a catalytic domain[15]. Signal peptides are 10-20 amino acids long. It primarily causes the translocation into the endoplasmic reticulum during mRNA translation. Propeptides are of variable lengths and have three important functions. They act as a scaffold to promote the protein folding of the catalytic domain, as a chaperone to transport the proenzyme to the lysosomal compartment, and as a high-affinity reversible inhibitor to prevent the premature activation of the catalytic domain. The catalytic domain is 214-260 amino acids long. It represents the mature, proteolytically active enzyme. It's exceptionally conserved active site involves cysteine, histidine and asparagine residues[1].

Cathepsin S assembly consists of a single chain monomeric protein of 217 amino acids with a molecular mass of 30kDa. The structure has two domains- left and right. The left domain contains residues 12-111, and 208-211 with helices ranging from residues 25-40, 50- 56, and 68-78. The right domain is grounded on a six-stranded β-barrel motif, residues 1-11, and 112-207, with small helical coiling of residues 119-127, and additional helix from residues 139-143. The cleft of the active site lies in between these two domains containing the residues Cys25 and His159[16-18].

[a] Department of Pharmaceutical Chemistry, Adarsh Vijendra Institute of Pharmaceutical Sciences, Shobhit University, Gangoh, Saharanpur -247 341, Uttar Pradesh, India

[b] Department of Pharmacy, Banasthali University, Banasthali, Tonk - 304 022, Rajasthan, India

[c] Department of Pharmacy, ITM University, Turari, Gwalior- 474 001, Madhya Pradesh, India

*For Correspondence: E-mail: sneha.kush09@gmail.com

https://doi.org/10.53879/id.60.02.13405

Acidic pH is essential for the optimal activity of cathepsin enzyme[19].

Cathepsin S plays a vital role in the various inflammation-associated disorders such as cancer[15,20-25], arthritis[18,26], periodontitis[27], psoriasis[18,28], lung diseases[29-35], cardiovascular disease in patients with chronic kidney disease[36-40], bone[41], Sjögren's syndrome[42,43] and immune disorders[44]. Inhibitors of cathepsin S also act as immunomodulators[45]. Thus, research efforts are necessarily focused on cathepsin S, its use in diagnostics, and as therapeutic targets in diseases[46,47]. Cathepsin S inhibitors of dipeptidyl nitrile are an emerging target for the abolition of tumour[25].

Rheumatoid arthritis (RA) is a chronic systemic autoimmune inflammatory ailment affecting all joints shielded by synovium. The genes encoding the major histocompatibility complex are gathered on a small portion of chromosome 6 in humans. It is also called as MHC complex or human leukocyte antigen complex (HLA) molecule. It significantly plays a central role in the pathology of RA[48]. The antigen is usually consumed by antigen- presenting cells (APCs), typically a macrophage present in the synovium. The antigen is broken into fragments by the peroxide enzyme inside the APCs[49]. The molecular mechanism involves the synthesis of MHC II $\alpha\beta$ heterodimers in the endoplasmic reticulum, followed by the association of a protein, the invariant chain (Li) in the peptide-binding cleft. The $\alpha\beta$Li complex gets relocated to the lysosome. Cathepsin S cleaves a part of the Li leaving a short fragment, CLIP, in the active site which prevents any premature binding of antigenic peptides[50-55].

A different protein, HLA-DM, assists in the dissociation of the CLIP from the MHC protein providing the preferable binding site for the peptide fragments. The complex is transported to the cell surface after binding to the MHC II molecule[50,56]. This complex is presented to T-cells (CD4 cells i.e. T-helper cell). The T-cell receptor (TCR) recognizes it, binds to it, and causes APCs to secrete cytokines like IL-1, IFN-α, IFN-γ, TNF and other factors which activate lymphocytes and other immune cells to respond to the antigens causing inflammation[18,34,48]. Moreover, HLA-DR β1 alleles majorly contribute genetically to RA. This gives a strong confirmation for adaptive immunity noteworthy in the pathogenesis of RA through MHC II-dependent T cell activation[57].

The Quantitative Structure-Activity Relationship (QSAR) method is extensively employed in biological activity modeling, and computing ADME/toxicity properties[58]. A QSAR model correlates the structure/ chemical characteristics of the molecule with their biological activities with the aid of a statistical equation. This data is useful in designing additionally potent compounds. The predictions of the biological activities can be done for new entities[59]. A QSAR analysis has great implications in enzyme inhibition studies, as well as in identifying the significant active sites in the receptor. Thus, QSAR studies have a central role in drug design[60,61].

The present 2D QSAR study used here is simple and relatively less error-prone. It excludes any type of conformational search or structural alignment, thus, it is more valuable over 3D QSAR analysis[62,63]. In 2D QSAR, structural descriptors encode all the chemical information[64]. Thus, 2D is considered superior over 3D QSAR[62,65,66].

MATERIALS AND METHODS

QSAR model was developed using 53 congeneric molecules using Multiple Linear Regression (MLR), Partial Least Squares (PLS), and Artificial Neural Network (ANN)[67].

All the structures of carbonitrile derivatives mentioned in the literature[68,69,70] were sketched using CHEM DRAW ULTRA 12.0 software as listed in Table VII.

Three compounds excluded from the series having undefined activity data were 2Y, 3Y, and 32 Y.

The inhibitors had a suitable pharmacokinetic (ADME) profile. The concept of absorption, distribution, metabolism, and excretion is important to know about the pharmacodynamics and pharmacokinetics of a chemical entity. Thus, the violation of Lipinski's rule of five has been checked. Lipinski's rule of five states that H-bond donors should be less than 5, H-bond acceptors should be less than 10, clog P (calculated log P) should be less than 5, and molecular weight should be less than 500 Da. for good oral absorption of a compound[71,72]. Lipinski's rule of five was applied to the whole data set as shown in Table I.

TSAR 3.3 software was used to calculate the molecular descriptors. Molecular descriptors provided all the valuable information about all the chemical structures and the respective substituents to figure out a good and predictive QSAR model[73,74]. Data reduction was done followed by the model development and validation using Multiple Linear Regression (MLR), Partial Least Squares (PLS), and Artificial Neural Network (ANN)[67].

Table I: Values of the calculated parameters for Lipinski's rule of five

Comp. Name	ADME (Molecular weight)	ADME(H-bond acceptors)	ADME(H-bond donors)	ADME (Log P)	ADME Violations
1X	333.3	5	1	2.838	0
8X	292.28	3	0	4.054	0
9X	288.25	3	1	2.836	0
10X	318.28	4	1	2.583	0
11X	332.31	4	1	2.926	0
12X	348.31	5	2	2.140	0
13X	346.34	4	1	3.394	0
14X	362.34	5	2	2.192	0
15X	394.38	4	1	4.360	0
16X	331.33	3	2	2.609	0
17X	345.31	4	2	1.685	0
19X	445.49	6	0	2.809	0
20X	403.45	5	0	3.151	0
21X	459.52	6	0	2.861	0
22X	471.48	5	0	3.204	0
29X	346.34	4	0	3.172	0
30X	346.34	4	1	2.799	0
31X	360.37	4	1	3.195	0
32X	409.4	5	0	4.159	0
33X	409.4	5	0	3.759	0
34X	409.4	5	0	3.759	0
35X	437.46	5	0	4.395	0
36X	429.44	5	0	2.255	0
37X	401.43	5	1	2.445	0
38X	415.46	5	1	2.868	0
39X	429.49	5	1	3.192	0
42X	389.42	5	0	2.809	0
43X	444.46	6	1	1.555	0
44X	403.45	5	0	2.861	0
45X	429.49	5	0	3.187	0
46X	458.54	6	0	2.662	0
47X	472.57	6	0	3.005	0
1Y	291.3	3	0	4.633	0
4Y	259.25	2	1	2.145	0
5Y	293.69	2	1	2.663	0
9Y	277.24	2	1	2.285	0
10Y	273.28	2	1	2.612	0

Comp. Name	ADME (Molecular weight)	ADME(H-bond acceptors)	ADME(H-bond donors)	ADME (Log P)	ADME Violations
11Y	289.28	3	1	1.892	0
12Y	279.66	2	1	2.195	0
18Y	322.69	3	2	2.279	0
19Y	388.76	4	3	2.233	0
20Y	402.79	4	3	2.247	0
21Y	402.79	4	2	2.479	0
22Y	402.79	4	3	2.701	0
23Y	366.75	4	2	1.368	0
24Y	432.82	5	3	2.323	0
25Y	429.81	5	2	2.355	0
26Y	443.84	5	2	2.823	0
27Y	509.91	6	2	3.555	1
28Y	523.94	6	2	4.024	1
29Y	537.97	6	2	4.427	1
30Y	552	6	2	4.816	1
31Y	443.84	5	2	2.601	0

RESULTS AND DISCUSSION

More than 250 molecular descriptors were generated. After data reduction, four independent molecular descriptors- Dipole moment Z Component (Substituent 1), Log P (Substituent 2), Shape flexibility index (Whole molecule) and Vamp LUMO (Whole molecule) were left with high correlation with the dependent variable i.e. the biological activity. The model generated showed poor predictive ability. MLR was performed with 37 compounds in the training set and 16 compounds in the test set. None of the compounds were removed as outliers. The statistical values of the regression analysis of the whole data set of molecular descriptors are listed in Table II.

The value of $r^2 = 0.762$ means that the MLR equation accounts for 76.2 % variance in the biological activity, illustrating a fairly realistic fit. The cross-validation regression coefficient is greater than 0.6 and the difference between r^2 (0.762) and r^2cv (0.759) is pretty lesser, which indicates the good internal predictive ability of the model.

The value of standard error, s (0.388), is considerably low for the regression to be important. It represents the quality of the fit of the model.

The correlation between parameters used and the biological activity is given in Table III. The statistical significance of the descriptors used in the final QSAR model is given inTable IV. The parameters with t-values greater than 2 indicate their significance in the model.

Table II: Statistical values obtained before data reduction and after performing MLR analysis

Sr. No.	Statistical test	Values before data reduction	Values after MLR
1.	s value	0.49	0.388
2.	f value	29.23	41.76
3.	Regression coefficient, r	0.645	0.916
4.	r^2	0.487	0.762
5.	Cross validation, r^2(cv)	0.354	0.759
6.	Residual sum of squares	26.234	4.813
7.	Predictive sum of squares	31.427	7.193

The four highly correlated descriptors were used to generate the regression equation as shown below and analyzed for their relative impacts on the activity of the compounds.

Original equation (by MLR method)

$$Y = 0.3661 * X1 + 1.8986 * X2 + 0.5152 * X3 - 1.8339 * X4 - 7.6452$$

Standardized equation (by MLR method)

$Y = 2.555 * S1 + 0.5728 * S2 + 0.6482 * S3 - 0.3300 * S4 - 1.5955$

Where, X1 is dipole moment Z component, X2 is log P, X3 is shape flexibility index, X4 is VAMP LUMO and Y is the biological activity.

MLR analysis gave satisfactory results with $r^2 = 0.762$ (training set) and 0.744 (test set). The results suggested good external validation. The MLR graphs for training and test set of compounds are shown in Fig. 1 and Fig. 2

To confirm the liability of the generated model, PLS analysis was performed using the same data set. Both MLR and PLS should have comparable results[75,76].

Table III: Correlation matrix showing the correlation between the biological activity and the molecular descriptors left after data reduction

	X1: Dipole Moment Z Component (Subst. 1)	X2: log P (Subst. 2)	X3: Shape Flexibility Index (Whole Molecule)	X4: VAMP LUMO (Whole Molecule)	log (1/IC_{50}) Values
X1: Dipole Moment Z Component (Subst. 1)	1	-0.26884	-0.18614	-0.24881	0.069073
X2: Log P (Subst. 2)	-0.26884	1	0.3471	0.68148	0.55287
X3: Shape Flexibility Index (Whole Molecule)	-0.18614	0.3471	1	0.29526	0.76985
X4: VAMP LUMO (Whole Molecule)	-0.24881	0.68148	0.29526	1	0.20626
Log (1/IC_{50}) Values	0.069073	0.55287	0.76985	0.20626	1

Table IV: Jacknife SE, Covariance SE, and t-values for the molecular descriptors

Molecular descriptor	Abbreviation	Jacknife SE	Covariance SE	t-value
Dipole moment Z component (subst. 1)	X1	0.1062	0.0970	3.7736
Log P (subst. 2)	X2	0.4577	0.3014	6.2983
Shape flexibility index (whole molecule)	X3	0.0541	0.0552	9.3281
VAMP LUMO (whole molecule)	X4	1.0623	0.4942	-3.7102
Constant	C	1.291		

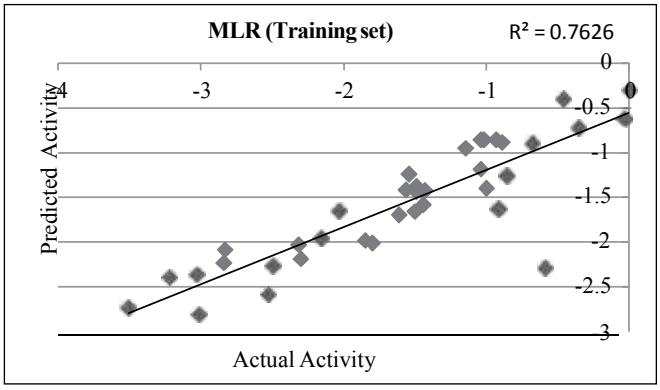


Fig. 1: Actual vs. predicted activity plot for the training set compounds derived from MLR analysis

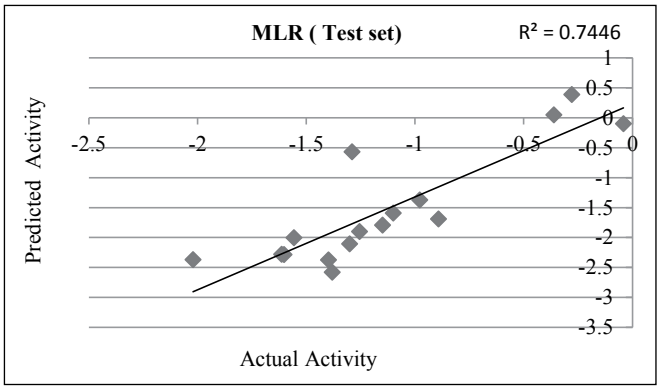


Fig. 2: Actual vs. predicted activity plot for the test set compounds derived from MLR analysis

PLS showed perfect results with r^2 = 0.806 (training set) and 0.768 (test set). The PLS graphs for training and test set of compounds are shown in Fig. 3 and Fig. 4.

PLS equation (Dimension 2)

Y = 0.4212 * X1 + 1.2180 * X2 + 0.5586 * X3 – 0.7915 * X4 – 6.4590

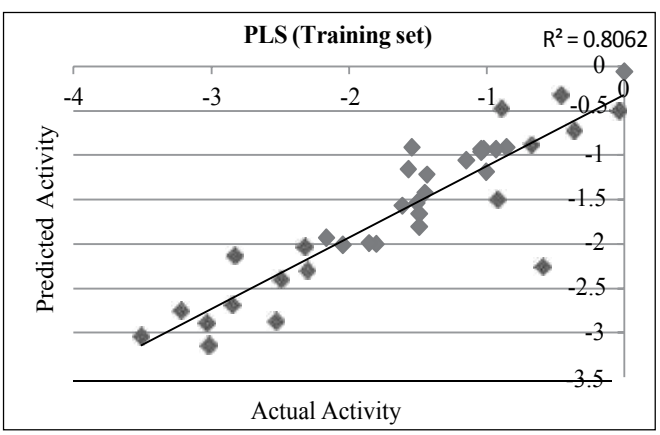


Fig. 3: Actual vs. predicted activity plot for the training set compounds derived from PLS analysis

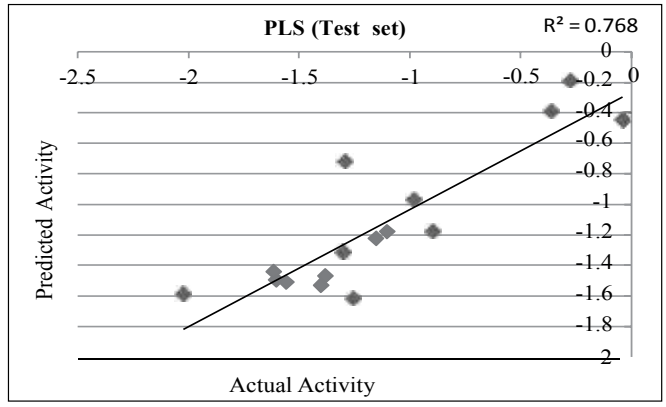


Fig. 4: Actual vs. predicted activity plot for the test set compounds derived from PLS analysis

PLS showed perfect results with r^2 = 0.806 (training set) and 0.768 (test set) whichfurther suggested the good external prediction.

Further validation was done by performing ANN. The ANN graphs for training and test set of compounds are shown in Fig. 5 and Fig. 6. A typical training and validation error curve is shown in Fig. 7.

The best RMS fit was found to be 0.0842 at 429 cycles. Net configuration was 4-1-1 and test RMS fit was 0.09874.

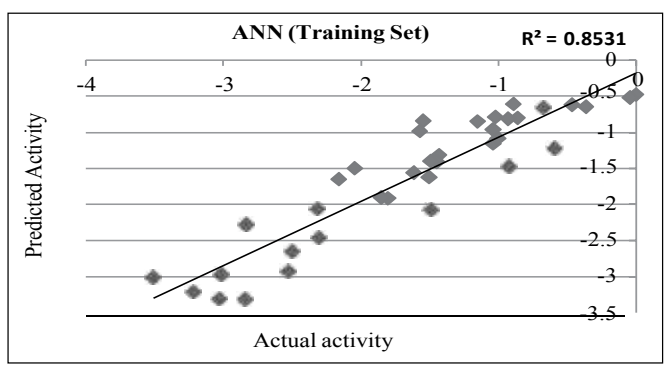


Fig. 5: Actual vs. predicted activity plot for the training set compounds derived from ANN analysis

Dipole moment Z component (subst. 1), Log P (subst. 2), Shape Flexibility Index (whole molecule), and VAMP LUMO (whole molecule) were the inputs and negative log IC50 values were the output for the ANN model.

The actual and predicted values for the training and test set compounds obtained from MLR, PLS, and FFNN analysis are given in Table V and VI, respectively.

DISCUSSION

The first descriptor Dipole moment Z component (subst. 1) explains the charge distribution and orientation

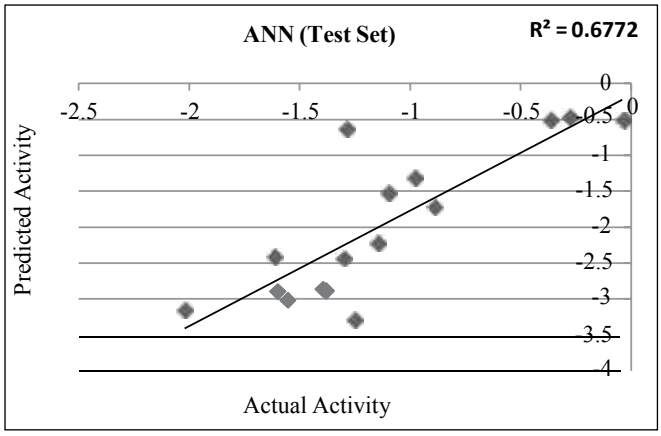


Fig. 6: Actual vs. predicted activity plot for the training set compounds derived from ANN analysis

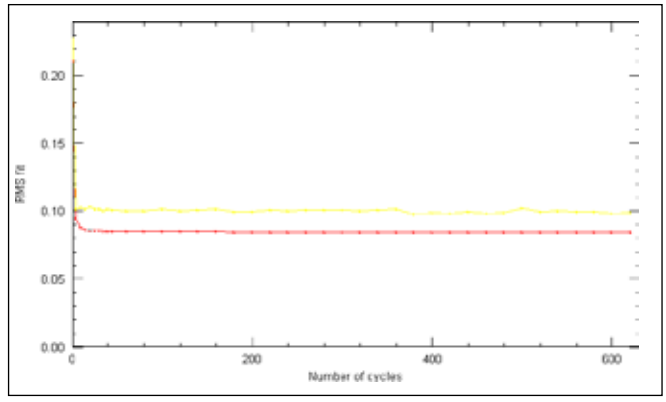


Fig. 7: Typical training and validation error curve

Table V: Actual and predicted values for the training set compounds obtained from MLR, PLS, and FFNN analysis of training set

Sr. No.	Comp. Name	Actual activity	Predicted activity		
			MLR	PLS	FFNN
1.	1X	-0.591	-2.289	-2.247	-1.213
2.	8X	-3.220	-2.388	-2.751	-3.196
3.	9X	-3.019	-2.809	-3.142	-2.965
4.	10X	-2.530	-2.585	-2.858	-2.919
5.	12X	-2.161	-1.954	-1.925	-1.656
6.	13X	-2.301	-2.186	-2.287	-2.447
7.	14X	-1.799	-2.005	-2.000	-1.919
8.	15X	-2.829	-2.078	-2.132	-2.265
9.	17X	-3.510	-2.734	-3.035	-2.992
10.	19X	-1.568	-1.415	-1.159	-0.979
11.	20X	-1.505	-1.650	-1.537	-1.626
12.	21X	-1.544	-1.233	-0.909	-0.844
13.	22X	-1.431	-1.420	-1.222	-1.319
14.	30X	-1.851	-1.978	-1.988	-1.907
15.	34X	-1	-1.394	-1.191	-1.084
16.	35X	-1.041	-1.178	-0.939	-1.153
17.	37X	-2.499	-2.272	-2.401	-2.641
18.	39X	-0.919	-1.626	-1.488	-1.471
19.	42X	-2.320	-2.021	-2.023	-2.054
20.	43X	-0.892	-0.882	-0.458	-0.606
21.	44X	-1.612	-1.686	-1.570	-1.563
22.	45X	-1.447	-1.575	-1.423	-1.411

Sr. No.	Comp. Name	Actual activity	Predicted activity		
			MLR	PLS	FFNN
23.	47X	-0.857	-1.252	-0.911	-0.797
24.	4Y	-3.029	-2.363	-2.887	-3.303
25.	5Y	-1.491	-1.427	-1.811	-2.064
26.	11Y	-2.840	-2.232	-2.674	-3.315
27.	12Y	-2.041	-1.655	-2.014	-1.504
28.	18Y	-1.489	-1.377	-1.667	-1.402
29.	19Y	-1.149	-0.945	-1.054	-0.856
30.	20Y	-1.021	-0.858	-0.931	-0.791
31.	21Y	-0.929	-0.858	-0.935	-0.807
32.	22Y	-0.361	-0.725	-0.721	-0.646
33.	23Y	-1.041	-0.852	-0.965	-0.961
34.	24Y	-0.462	-0.395	-0.317	-0.662
35.	25Y	-0.681	-0.878	-0.880	-0.650
36.	26Y	-0.041	-0.626	-0.501	-0.521
37.	27Y	0	-0.297	-0.062	-0.475

Table VI: Actual and predicted values for the training set compounds obtained from MLR, PLS, and FFNN analysis of test set

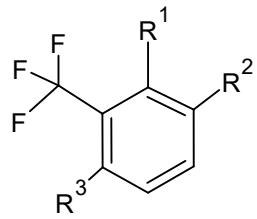
Sr. No.	Comp. Name	Actual activity	Predicted activity		
			MLR	PLS	FFNN
1.	11X	-1.397	-2.375	-1.528	-2.874
2.	16X	-2.021	-2.370	-1.586	-3.151
3.	29X	-1.602	-2.285	-1.495	-2.906
4.	31X	-1.380	-2.577	-1.471	-2.897
5.	32X	-1.301	-2.105	-1.307	-2.424
6.	33X	-1.100	-1.591	-1.178	-1.523
7.	36X	-0.892	-1.691	-1.173	-1.733
8.	38X	-1.149	-1.793	-1.222	-2.211
9.	46X	-0.977	-1.374	-0.964	-1.327
10.	1Y	-1.612	-2.279	-1.439	-2.422
11.	9Y	-1.556	-2.001	-1.507	-1.017
12.	10Y	-1.255	-1.901	-1.609	-3.297
13.	28Y	-0.041	-0.100	-0.448	-0.529
14.	29Y	-0.361	0.047	-0.385	-0.511
15.	30Y	-0.278	0.390	-0.184	-0.484
16.	31Y	-1.290	-0.565	-0.716	-0.634

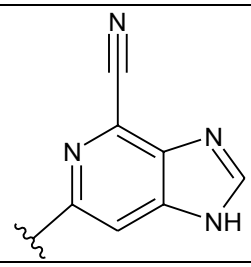
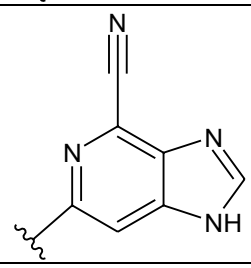
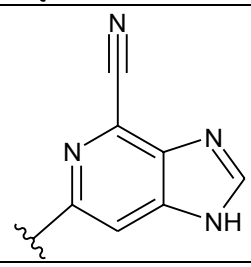
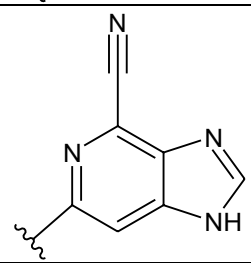
behavior of the molecule. It is negatively correlated with biological activity. This indicates that adding such groups in a molecule or a lead compound will lead to the increased polarity of the molecule, and thus decrease the biological activity. This clearly shows that the active site of cathepsin S enzyme will show some hydrophobic pockets to have hydrophobic interactions. It also provides the fact that the active site of cathepsin S enzyme is lipophilic in nature.

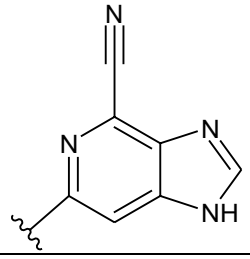
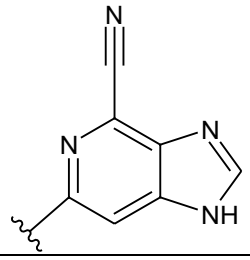
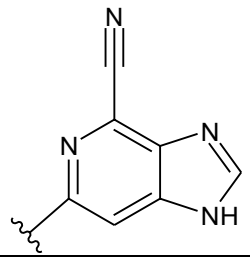
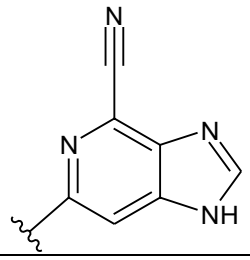
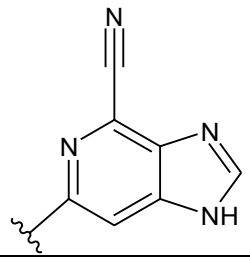
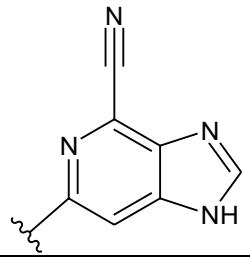
The second descriptor log P (subst. 1) explains the lipophilic character of the molecule. The descriptor is positively correlated with the biological activity. The less polar groups when introduced will tend to increase the biological activity.

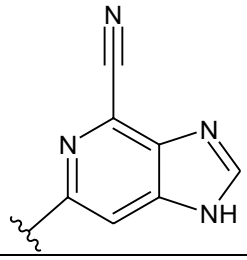
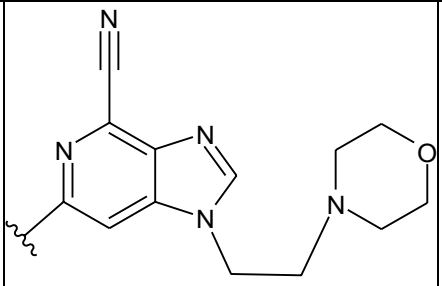
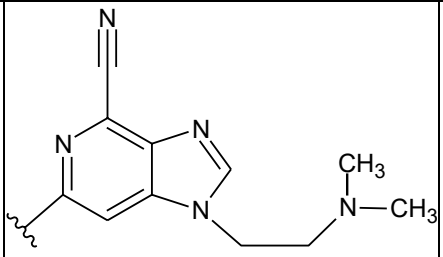
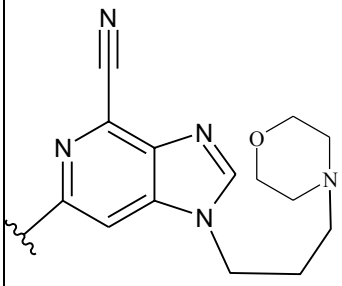
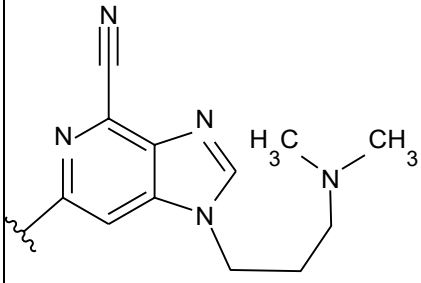
Thus, the nature of both the descriptors clearly explains the hydrophobic nature of the active site of the target- cathepsin S enzyme.

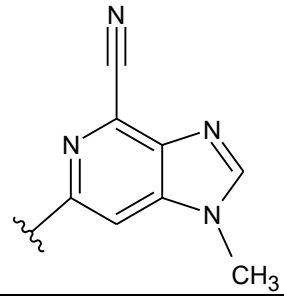
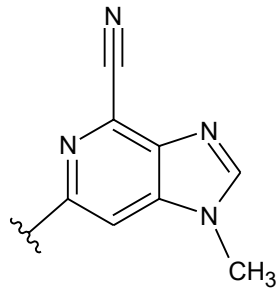
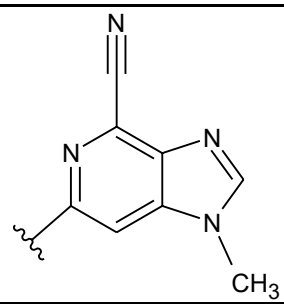
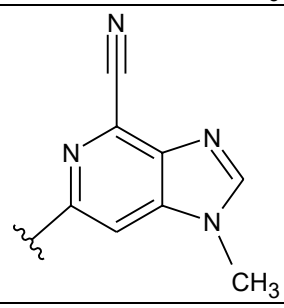
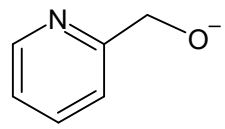
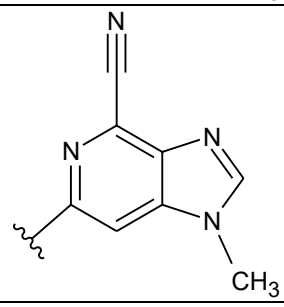
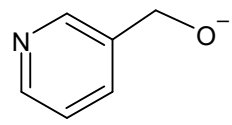
Table VII: Chemical data of carbonitrile derivatives

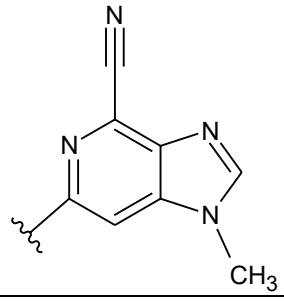
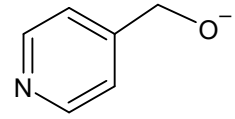
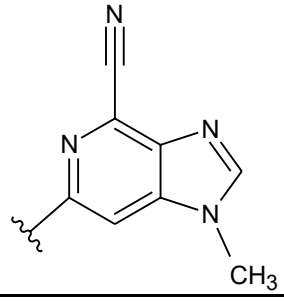
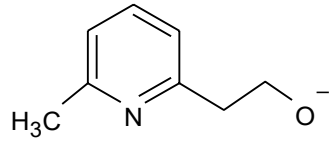
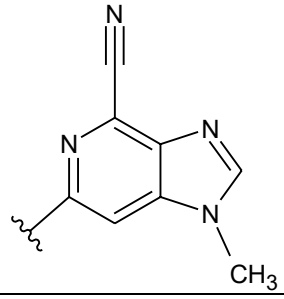
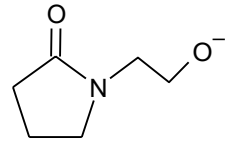
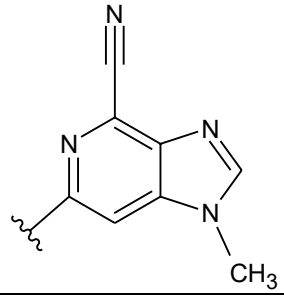
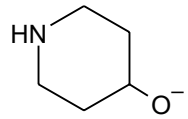
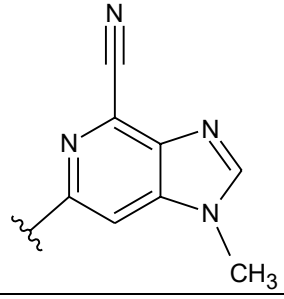
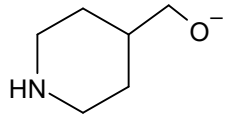


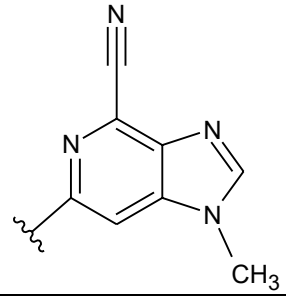
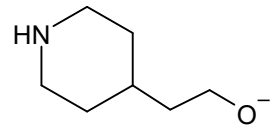
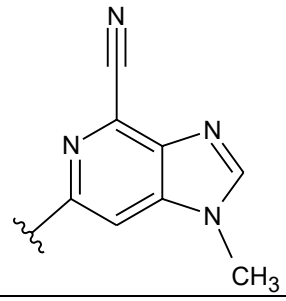
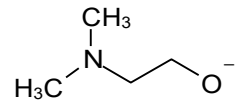
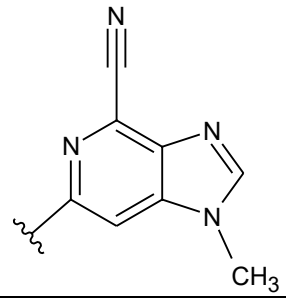
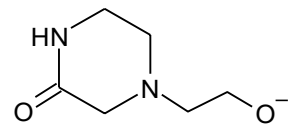
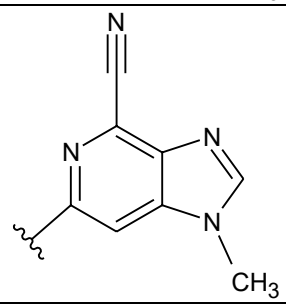
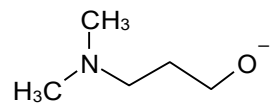
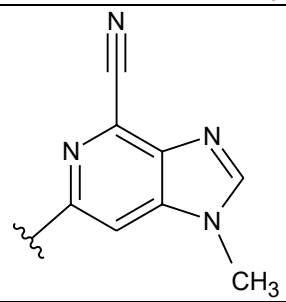
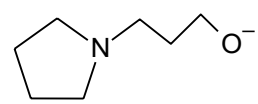
Sr. No.	Comp. Name	R1	R2	R3
1	1	H		EtO
2	8	H		EtO
3	9	H		H
4	10	H		MeO

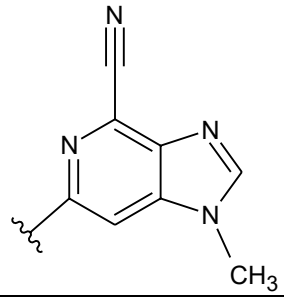
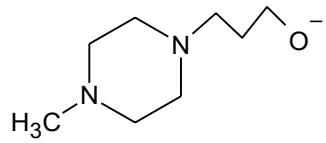
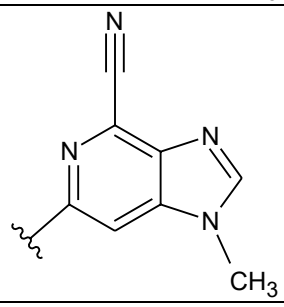
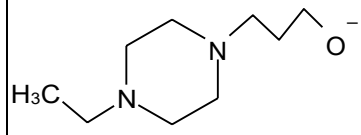
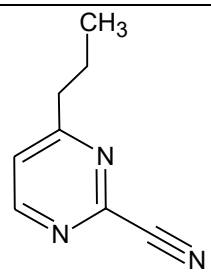
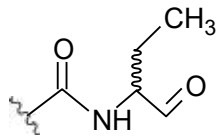
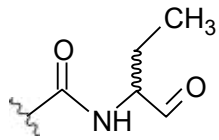
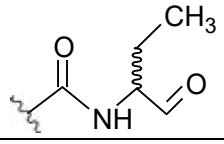
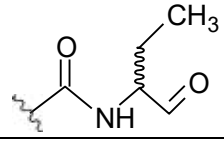
Sr. No.	Comp. Name	R1	R2	R3
5	11	H		EtO
6	12	H		$HO(CH_2)_2O$
7	13	H		n-Pro
8	14	H		$HO(CH_2)_3O$
9	15	H		BenzylO
10	16	H		EtNH

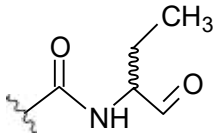
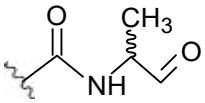
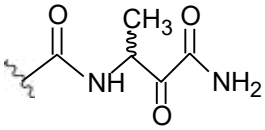
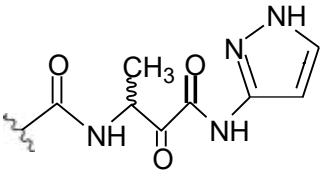
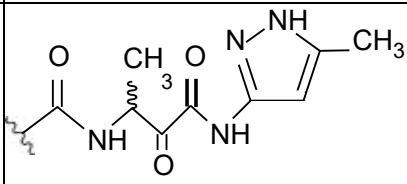
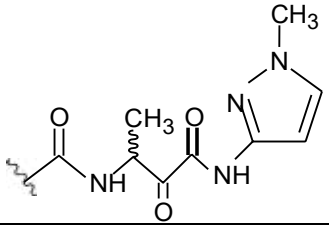
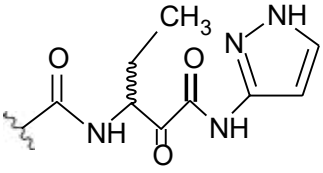
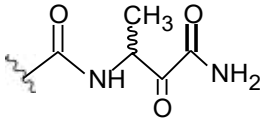
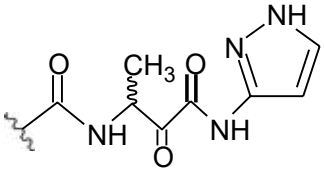
Sr. No.	Comp. Name	R1	R2	R3
11	17	H		AcNH
12	19	H		EtO
13	20	H		EtO
14	21	H		EtO
15	22	H		EtO

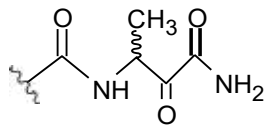
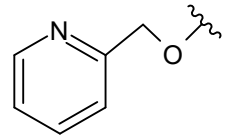
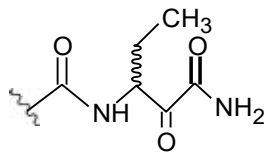
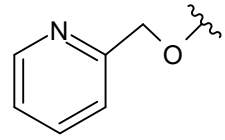
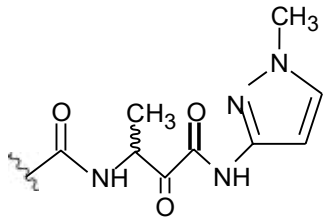
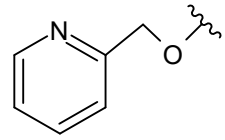
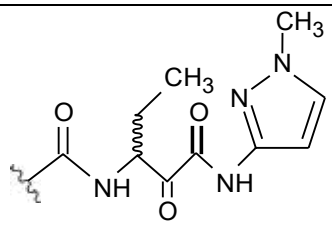
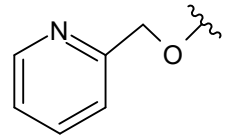
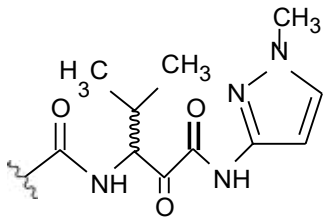
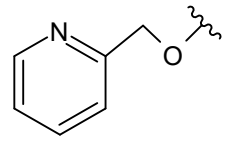
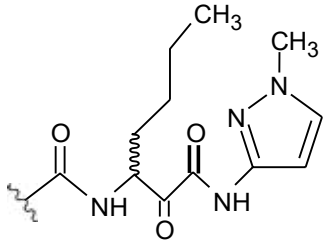
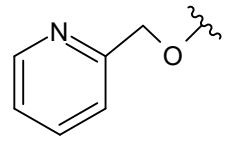
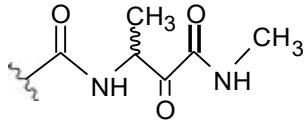
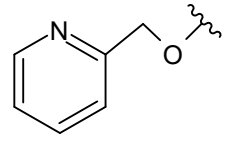
Sr. No.	Comp. Name	R1	R2	R3
16	29	H		EtO
17	30	H		$HO(CH_2)_2$
18	31	H		$HO(CH_2)_3$
19	32	H		
20	33	H		

Sr. No.	Comp. Name	R1	R2	R3
21	34	H		
22	35	H		
23	36	H		
24	37	H		
25	38	H		

Sr. No.	Comp. Name	R1	R2	R3
26	39	H		
27	42	H		
28	43	H		
29	44	H		
30	45	H		

Sr. No.	Comp. Name	R1	R2	R3
31	46	H		
32	47	H		
33	1B	H		H
34	4B	H		H
35	5B	Cl		H
36	9B	F		H
37	10B	Me		H

Sr. No.	Comp. Name	R1	R2	R3
38	11B	MeO		H
39	12B	Cl		H
40	18B	Cl		H
41	19B	Cl		H
42	20B	Cl		H
43	21B	Cl		H
44	22B	Cl		H
45	23B	Cl		EtO
46	24B	Cl		EtO

Sr. No.	Comp. Name	R1	R2	R3
47	25B	Cl		
48	26B	Cl		
49	27B	Cl		
50	28B	Cl		
51	29B	Cl		
52	30B	Cl		
53	31B	Cl		

The third descriptor is the shape flexibility index (whole molecule) and it defines the flexible nature of the substituent which aids in the favorable drug-receptor interactions. The descriptor is positively correlated with the biological activity, thus adding such groups will cause an increase in the biological activity.

The fourth descriptor is the VAMP LUMO (whole molecule). VAMP LUMO is the energy of the lowest occupied molecular orbital. This energy is directly related to the electron affinity and characterizes the susceptibility of the molecule towards the attack by a nucleophile. It is negatively correlated with biological activity. Thus, the lesser the electron- donating groups added to the nucleus, the lesser is the electrostatic nature of the substituent and the greater is the increase in the biological activity.

CONCLUSION

QSAR study was successfully performed on a series of 6-phenyl-1*H*-imidazo [4, 5-c] pyridine-4-carbonitrile and trifluoromethylphenyl derivatives. Significant statistical values of MLR, PLS, and FFNN indicated the robustness of the model. Value of r^2 of 0.762, 0.806, and 0.853 for MLR, PLS, and FFNN (training set), respectively, indicated the soundness of the model. Value of r^2 of 0.744, 0.768, and 0.677 for MLR, PLS and FFNN (test set), respectively, indicated better results. According to the classical QSAR models presented in the present work, the remaining four molecular descriptors- dipole moment, log P, shape flexibility index, and VAMP LUMO encoding the polarity, lipophilicity, shape and electrophilicity property of molecules gives the predictive information about the overall behavior of the molecules and are considered to be the important contributors to their biological properties. These findings will be effective in designing more potent and effective cathepsin S inhibitors.

REFERENCES

1. Bromme D. and Wilson, S.: Role of Cysteine Cathepsins in Extracellular Proteolysis, In: Extracellular matrix degradation, Vol. 2, Springer Link, Canada, 2011, pp. 23-51.
2. Reiser J., Adair B. and Reinheckel T.: Specialized roles for cysteine cathepsins in health and disease, **J. Clin. Invest**., 2010, 120, 3421-3431.
3. Guha S. and Padh H.: Cathepsins: Fundamental Effectors of Endolysosomal Proteolysis, **Indian J. Biochem. Biophys.**, 2008, 45, 75-90.
4. Mohamed M. M. and Sloane B. F.: Cysteine cathepsins: multifunctional enzymes in cancer, **Nat. Rev. Cancer**, 2006, 6, 764-775.
5. Li Y. Y., Fang J. and Ao G. Z.: Cathepsin B and L inhibitors: a patent review, **Expert Opin. Ther. Pat.**, 2016, 27, 643-656.
6. Sena B. F., Figueiredo J. L. and Aikawa E.: Cathepsin S as an inhibitor of Cardiovascular inflammation and Calcification in Chronic Kidney Disease, **Front. Cardiovasc. Med.**, 2018, 4, 1-7.
7. Chang W. S. W., Wu H. R., Yeh C. T., Wu C. W. and Chang J. Y.: Lysosomal Cysteine Proteinase Cathepsin S as a Potential Target for Anti-cancer Therapy, **J. Cancer Mol.**, 2007, 3, 5-14.
8. Pauly T. A., Sulea T., Ammirati M., Sivaraman J., Danley D. E., Griffor M. C., Kamath A. V., Wang I-K., Laird E. R., Seddon A. P., Ménard R., Cygler M. and Rath V. L.: Specificity Determinants of Human Cathepsin S revealed by Crystal Structures of Complexes, **Biochem**., 2003, 42, 3203-3213.
9. Hsing L. C. and Rudensky A. Y.: The lysosomal cysteine proteases in the MHC class II antigen presentation, **Immunol. Rev**., 2005, 207, 229-241.
10. Shi G. P., Chapman H. A., Webb A. C., Foster K. A., Knoll J. H. M., Lemere C. A. and Munger J. S.: Cathepsin S: Chromosomal Localization, Gene Structure and Tissue Distribution, **J. Biol. Chem.**, 1994, 269, 11530-11536.
11. Nakagawa T. Y. and Rudensky A. Y.: The role of lysosomal proteinases in MHC class- II mediated antigen processing and presentation, **Immunol. Rev**., 1999, 172, 121-129.
12. Riese R. J. and Chapman H. A.: Cathepsins and compartmentalization in antigen presentation. **Curr. Opin., Immunol**., 2000, 12, 107-113.
13. Shi G. P., Bryant R., Riese R., Verhelst S., Driessen C., Li Z., Bromme D., Ploegh H. L. and Chapman H.: A.Role of cathepsin S in invariant chain processing and major histocompatibility complex class II peptide loading by macrophages, **J. Exp. Med.,** 2000, 191, 1177-1186.
14. Tolosa E., Li W., Yasuda Y., Wienhold W., Denzin L. K., Lautwein A., Driessen C., Schnorrer P., Weber E., Stevanovic S., Kurek R., Melms A. and Bromme D.: Cathepsin V involved in the degradation of invariant chain in human thymus and is overexpessed in myasthenia gravis, **J. Clin. Invest.**, 2003, 112, 517-526.
15. Huang C. C., Lee C. C., Lin H. H. and Chang J. Y.: Cathepsin S attenuates endosomal EGFR signalling: A mechanical rationale for the combination of cathepsin S and EGFR tyrosine kinase inhibitors, **Sci. Rep**., 2016, 6, 1-12.
16. McGrath M. E., Palmer J. T., Bromme D. and Somoza J. R.: Crystal structure of human cathepsin S, **Protein Sci**., 1998, 7, 1294-1302.
17. Vasiljeva O., Reinheckel T., Peters C., Turk D., Turk V. and Turk B.: Emerging Roles of Cysteine Cathepsins in Disease and their Potential as Drug Targets, **Curr. Pharm. Des**., 2007, 13, 385-401.
18. Löser R. and Pietzsch J.: Cysteine cathepsins: their role in tumor progression and recent trends in the development of imaging probes, **Front. Chem**., 2015, 3, 1-36.

19. Brocklehurst K.: A sound basis for pH-dependent kinetic studies on enzymes, **Protein Eng.**, 1994, 7, 291-299.
20. Huang C. C., Lee C. C., Lin H. H., Chen M. C., Lin C. C. and Chang J. Y.: Autophagy- Regulated ROS from Xanthine Oxidase Acts as an Early Effector for Triggering Late Mitochondria-Dependent Apoptosis in Cathepsin S-Targeted TumorCells, **PLoS One**, 2015, 10, 1-20.
21. Hsieh M. J., Lin C. W., Chen M. K., Chien S. Y., Lo Y. S., Chuang Y. C., Hsi Y-T., Lin C-C., Chen J-C. and Yang S-F.: Inhibition of cathepsin S confers sensitivity to methyl protodioscin in oral cancer cells via activation of p38 MAPK/JNK signaling pathways, **Sci. Rep.**, 2017, 7, 1-11.
22. Zhang L., Wang H. and XU J.: Cathepsin S as a cancer target, **Neoplasma**, 2015, 62, 16- 28.
23. Da Costa A. C., SantaCruz F., Mattos Jr L. A. R., Aquino M. A. R., Martins C. R. and Ferraz A. A. B.: Cathepsin S as a target in gastric cancer, **Mol. Clin. Oncol.**, 2020, 12, 99-103.
24. Liu W. L., Liu D., Cheng K., Liu Y. J., Xing S., Chi P., Xiao-Hua Liu X-H., Xue N., Lai Y-Z., Ling Guo L. and Zhang G.: Evaluating the diagnostic and prognostic value of circulating cathepsin S in gastric cancer, **Oncotarget,** 2016, 7, 28124-28138.
25. Wilkinson R. D. A., Young A., Burden R. E., Williams R. and Scott C..J.: A bioavailable cathepsin S nitrile inhibitor, **Mol. Cancer**, 2016, 15, 2-11.
26. Weitoft T., Larsson A., Manivel V. A., Lysholm J., Knight A. and Ronnelid J.: Cathepsin S and cathepsin L in serum and synovial fluid in rheumatoid arthritis with and without autoantibodies arthritis, **Rheumatology**, 2015, 54, 1923-1928.
27. Memmert S., Damanaki A., Nogueira A. V. B., Eick S., Nokhbehsaim M., Papadopoulou A. K., Till A., Rath B., Jepsen S., Götz W., Piperi C., Basdra E. K., Cirelli J. A., Jäger A. and Deschner J.: Role of Cathepsin S in Periodontal Inflammation and Infection, **Mediators Inflamm.**, 2017, 1-11.
28. Ainscougha J. S., Macleoda T., McGonagleb D., Brakefielda R., Baron J. M., Alase A., Wittmann M. and Stacey M.: Cathepsin S is the major activator of the psoriasis associated proinflammatory cytokine IL-36, **Proc. Natl. Acad. Sci.** U.S.A., 2017, 114, 1-10.
29. Andrault P. M., Samsonov S. A., Weber G., Coquet L., Nazmi K., Bolscher J. G. M., Lalmanach A-C., Jouenne T., Brömme D., Pisabarro M. T., Lalmanach G. and Lecaille F.: Antimicrobial Peptide LL-37 Is Both a Substrate of Cathepsins S and K and a Selective Inhibitor of Cathepsin L, **Biochem.**, 2015, 54, 2785–2798.
30. Brown R., Nath S., Lora A., Samaha G., Elgamal Z., Kaiser R., Taggart C., Weldon S. and Geraghty P.: Cathepsin S: investigating an old player in lung disease pathogenesis, comorbidities, and potential therapeutics, **Respir. Res.**, 2020, 21, 1-17.
31. Wartenberg M., Andrault P. M., Saidi A., Bigot P., Nadal-Desbarats L., Lecaille F. and Lalmanach G.: Oxidation of cathepsin S by major chemicals of cigarette smoke, **Free Radic. Biol. Med.**, 2020, 150, 53-65.
32. Andrault P. M., Schamberger A. C., Chazeirat T., Sizaret D., Renault J., Staab- Weijnitz C. A., Hennen E., Petit-Courty A., Wartenberg M., Saidi A., Baranek T., Guyetant S., Courty Y., Eickelberg O., Lalmanach G. and Lecaille F.: Cigarette smoke induces overexpression of active human cathepsin S in lungs from current smokers with or without COPD, **Am. J. Physiol. Lung Cell Mol.**, 2019, 317, L625- L638.
33. Nakajima T., Nakamura H., Owen C. A., Yoshida S., Tsuduki K., Chubachi S., Shirahata T., Mashimo S., Nakamura M., Takahashi S., Minematsu N., Tateno H., Fujishima S., Asano K., Celli B. R. and Betsuyaku T.: Plasma Cathepsin S and Cathepsin S/Cystatin C Ratios Are Potential Biomarkers for COPD, **Dis. Markers**, 2016, 1-10.
34. Zhou P. P., Zhang W. Y., Li Z. F., Chen Y. R., Kang X. C. and Jiang Y. X.: Association between SNPs in the promoter region in cathepsin S and risk of asthma in Chinese Han population, **Eur. Rev. Med. Pharmacol. Sci.**, 2016, 20, 2070-2076.
35. Small D. M., Brown R. R., Doherty D. F., Abladey A., Zhou-Suckow Z., Delaney R. J., Kerrigan L., Dougan C. M., Borensztajn K. S., Holsinger L., Booth R., Scott C. J., López-Campos G., Elborn J. S., Mall M. A., Weldon S. and Clifford C.: Taggart. Targeting of cathepsin S reduces cystic fibrosis-like lung disease, **Eur. Respir. J.**, 2019, 53, 1-11.
36. Sena B. F., Figueiredo J. L. and Aikawa E.: Cathepsin S as an inhibitor of Cardiovascular inflammation and Calcification in Chronic Kidney Disease, **Front. Cardiovasc. Med.**, 2018, 4, 1-7.
37. Figueiredo J. L., Aikawa M., Zheng C., Aaron J., Lax L., Libby P., Filho J. L. L., Gruener S., Fingerle J., Haap W., Hartmann G. and Aikawa E.: Selective Cathepsin S Inhibition Attenuates Atherosclerosis in Apolipoprotein E-Deficient Mice with Chronic Renal Disease, **Am. J. Pathol.**, 2015, 185, 1155-1166.
38. Ahmad S. and Siddiqi M. I.: Insights from molecular modeling into the selective inhibition of cathepsin S by its inhibitor, **J. Mol. Model.**, 2017, 23, 3255-3256.
39. Figueiredo J. L., Aikawa M. and Sena B. F.: Cathepsin S as an Inhibitor of Cardiovascular Inflammation and Calcification in Chronic Kidney Disease, **Front. Cardiovasc. Med.**, 2018, 4, 88.
40. Wu H., Qiuna D., Dai Q., Ge J. and Cheng X.: Cysteine protease Cathepsins in Atherosclerotic Cardiovascular Diseases, **J. Atheroscler. Thromb.**, 2017, 24, 1-13.
41. Repnik U., Starr A. E., Overall C. M. and Turk B.: Cysteine Cathepsins Activate ELR Chemokines and Inactivate Non-ELR Chemokines, **J. Biol. Chem.**, 2015, 290, 13800–13811.
42. Hargreaves P., Daoudlarian D., Theron M., Kolb F. A., Young M. M., Reis B., Tiaden A., Bannert B., Kyburz D. and Manigold T.: Differential effects of specific cathepsin S inhibition in biocompartments from patients

with primary Sjögren syndrome, **Arthritis Res. Ther**., 2019, 21, 1-11.

43. Edman M. C., Janga S. R., Meng Z., Bechtold M., Chen A. F., Kim C., Naman L., Sarma A., Teekappanavar N., Kim A. Y., Madrigal S., Singh S., Elizabeth Ortiz E., Christianakis S., Arkfeld D. G., Mack W. J., Heur M., Stohl W. and Hamm-Alvarez S. F.: Increased Cathepsin S activity associated with decreased protease inhibitory capacity contributes to altered tear proteins in Sjögren's Syndrome patients, **Sci. Rep.,** 2018, 8, 1-12.
44. Vizovišek M., Fonovi M. and Turk B.: Cysteine cathepsins in extracellular matrix remodeling: Extracellular matrix degradation and beyond, **Matrix Biol**., 2019, 75-76, 149-159.
45. Gupta S., Singh R. K., Dastidar S. and Ray A.: Cysteine cathepsin S as an immunomodulatory target: present and future trends, **Expert Opin. Ther. Targets**, 2008, 12, 291-299.
46. Vizovišek M., Vidak E., Javoršek U., Mikhaylov G., Bratovš A. and Turk B.: Cysteine cathepsins as therapeutic targets in inflammatory diseases, **Expert Opin. Ther. Targets**, 2020, 24, 573-588.
47. Wilkinson R. D. A., Williams R., Scott C. J. and Burden R. E.: Cathepsin S: therapeutic, diagnostic, and prognostic potential, **Biol. Chem**., 2015, 396, 867–882.
48. Kumar V., Abbas A. K., Fausto N., and Aster J. C.: Pathologic Basis of Disease, 8th ed., Sauders Elsevier, Philadelphia, 2008, 190-192.
49. Riese R., Villadangos J. A., Shi G. P., Ploegh H. L., Chapman H. A., Bryant R. R., Wroth W., Saftig P., Peters C., Deussing J., Driessen C. and A M Lennon-Duménil A. M.: Proteases involved in MHC class II antigen presentation, **Immunol. Rev**., 1999, 172, 109-120.
50. Riese R. J., Mitchell R. N., Villadangos J. A., Shi G. P., Palmer J. T., Karp E. R., De Sanctis G. T., Ploegh H. L. and Chapman H. A.: Cathepsin S Activity Regulates Antigen Presentation and Immunity, **J. Clin. Invest**., 1998, 101, 2351-2363.
51. Riese R., Villadangos J. A., Shi G. P., Ploegh H. L., Chapman H. A., Dranoff G., Small C., Gu L. and Haley K. J: Cathepsin S Required for Normal Class II Peptide Loading and Germinal Center Development, **Immunity**, 1999, 10, 197-206.
52. Nakagawa T. Y., Bressette W. H., Lira P. D., Griffiths R. J., Petrushova N., Stock J., McNeish J. D., Eastman S. E., Howard E. D., Clarke S. R., Rosloniec E. F., Elliott E. A. and Rudensky A. Y.: Impaired Invariant Chain Degradation and Antigen Presentation and Diminished Collagen-Induced Arthritis in Cathepsin S Null Mice, **Immunity**, 1999, 10, 207-217.
53. Lutzner N. and Kalbacher H.: Quantifying Cathepsin S Activity in Antigen Presenting Cells Using a Novel specific Substrate, **J. Biol. Chem**., 2008, 283(52), 36185-36194.
54. Costantino C. M., Ploegh H. L. and Hafler D. A.: Cathepsin S Regulates Class II MHC Processing in Human CD4 + HLA-DR+ T Cells, **J. Immunol**., 2009, 183, 945- 952.
55. Mcgrath M. E.: The Lysosomal Cysteine Proteases, **Annu. Rev. Biophys. Biomol. Struct**., 1999, 28, 181-204.
56. Kumar P. and Mina U.: Life Sciences, Fundamentals and Practices, Pathfinder Publication, New Delhi, 2013, Vol. 1, pp. 460-515.
57. Snir O. Specific autoimmunity in rheumatoid arthritis-T cells, antibodies and genetic regulation (Ph.D. Thesis), Karolinska Institutet: Stockholm, May, 2011.
58. Hansch C. and Fujita T.: p-δ-ϖ analysis, a method for the correlation of biological activity and chemical structure, **J. Am. Chem. Soc**., 1964, 86, 1616-1626.
59. Jamloki A., Karthikeyan C., Sharma S. K., Moorthy N. S. H. N. and Trivedi P.: QSAR studies on some GSK-3α Inhibitory 6-aryl-pyrazolo(3,4-b)pyridines, **Asian J. Biochem**., 2006, 1, 236-243.
60. Thomas G.: Medicinal Chemistry: An Introduction, 2nd edition, Wiley, UK, 2008, pp. 71-91.
61. Gupta S. P. and Kumaran S.: Quantitative structure-activity relationship studies on benzodiazepine hydroxamic acid inhibitors of matrix metalloproteinases and tumour necrosis factor-α converting enzyme, **Asian J. Biochem**., 2006, 1, 47-56.
62. Hoffman B., Cho S. J., Zheng W., Wyrick S., Nichols D. E., Mailman R. B. and Tropsha A.: Quantitative structure-activity relationship modeling of dopamine D1 antagonists using comparative molecular field analysis, genetic algorithms-partial least- squares and K nearest neighbor methods, **J. Med. Chem**., 1999, 42, 3217-3226.
63. Shen M., LeTiran A., Xiao Y., Golbraikh A., Kohn H. and Tropsha A.: Quantitative structure-activity-relationship analysis of functionalized amino acid anticonvulsant agents using k nearest neighbour and simulated annealing PLS methods, **J. Med. Chem**., 2002, 45, 2811-2823.
64. Rogers D. and Hopfinger A. J.: Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships, **J. Chem. Inf. Comput. Sci**., 1994, 34, 854-866.
65. Golbraikh A., Bonchev D. and Tropsha A.: Novel chirality descriptors derived from molecular topology, **J. Chem. Inf. Comput. Sci**., 2001, 41, 147-158.
66. Zheng W. and Tropsha A.: A novel variable selection quantitative structure-property relationship approach based on the k-nearest-neighbour principle, **J. Chem. Inf. Comput. Sci.**, 2000, 40, 185-194.
67. Sen S. and Sarker K.: 2D QSAR Study of Some Inhibitors of Ehrlich Ascites Carcinoma, **Int. J. Pharm. Sci. Rev. Res**., 2012, 15, 102-107.
68. Cai J., Baugh M., Black D., Long C., Bennette D. J., Dempster M., Fradera X., Gillespie J., Andrews F., Boucharens S., Bruin J., Cameron K. S., Cumming I., Hamilton W., Jones P. S., Kaptein A., Kinghorn E., Maidment M., Martin I., Mitchell A., Rankovic Z., Robinson J., Scullion P., Uitdehaag J. C. M., Vink P., Westwood P., Zeeland M., Berkom L., Bastiani M. and Meulemans

T.: 6-Phenyl-1*H*-imidazo[4,5- *c*]pyridine-4-carbonitrile as cathepsin S inhibitors, **Bioorg. Med. Chem. Lett.**, 2010, 20(15), 4350-4354.

69. Cai J., Baugh M., Bruin J., Robinson J., Bennette D. J., Rankovic Z., Dempster M., Fradera X., Gillespie J., Cumming I., Finlay W., Sylviane Boucharens S., Cameron K. S., Hamilton W., Kerr J., Kinghorn E., McGarry G., Scullion P., Uitdehaag J. C. M., Zeeland M., Potin D., Saniere L., Fouquet A., Chevallier F., Deronzier H., Dorleans C. and Nicolai E.: 2-Phenyl-9*H*-purine-6-carbonitrile derivatives as selective cathepsin S inhibitors, **Bioorg. Med. Chem. Lett**., 2010, 20(15), 4447-4450.
70. Cai J., Baugh M., Bruin J., Robinson J., Bennette D. J., Fradera X., Zeeland M., Dempster M., Cameron K. S., Popplestone L., Westwood P., Hamilton W., Kinghorn E., Long C. and Uitdehaag J. C. M.: 4-(3-Trifluoromethylphenyl)-pyrimidine-2- carbonitrile as cathepsin S inhibitors: N3, not N1 is critically important, **Bioorg. Med. Chem. Lett.**, 2010, 20(15), 4507-4510.
71. Paliwal S. K., Pandey A. and Paliwal S.: Quantitative Structure Activity Relationship Analysis of *N*-(mercapto-alkanoyal)-[(acylthio)alkanoyal] glycine derivatives as ACE inhibitors, **Am. J. Drug Discov. Dev.**, 2011, 1, 85-104.
72. Kushwaha S. and Paliwal S. K.: 2D QSAR analysis of dipeptide nitrile based cathepsin S inhibitors, **Int. J. Pharm. Sci. Res**., 2021, 12(6), 3391-3302.
73. Klocker J., Wailzer B., Buchbauer G. and Wolschann P.: Bayesian neural networks for aroma classification, **J. Chem. Inf. Comput. Sci**., 2002, 42, 1443-1449.
74. Kovatcheva A., Buchbauer G., Golbraikh A. and Wolschann P.: QSAR modelling of alpha campholenic derivatives with sadalwood odor, **J. Chem. Inf. Comput. Sci**., 2003, 43, 259-266.
75. Paliwal S., Narayan A. and Paliwal S.: Quantitative structure activity relationship analysis of dicationic diphenylisoxazole as potent anti-trypanosomal agents, QSAR **Comb. Sci**., 2009, 28, 1367-1375.
76. Paliwal S. K., Pal M. and Siddiqui A. A.: Quantitative structure activity relationship analysis of angiotensin II AT1 receptor antagonists, **Med. Chem. Res**., 2010, 19, 475- 489.

● ● ●

● ● ●