QSAR and Molecular Modeling Studies on a Series of Inflenza Neuraminidase Inhibitors with Cyclohexene Scaffold

Neelu Singh

Department of Applied Sciences, Sagar Institute of Research and Technology-Excellence, Bhopal-462041, India

***Abstract*- QSAR (Quantitative structure-activity relationship) and molecular modeling studies have been performed on a series of influenza neuraminidase (NA) inhibitors with cyclohexene scaffold. The QSAR model shows that the activity of compounds would be a function of their molecular size, but very large molecule may create steric problem. The drug-receptor interaction might involve dispersion interaction. Using the QSAR model some new compounds with cyclohexene scaffold actind as NA inhibitors have been predicted. Except a few compounds, all other predicted compounds have as good docking scores as the well known NA inhibitor, Zanamivir. The most active compound among the predicted ones is shown to have almost same number of hydrogen bonding with the enzyme as the Zanamivir.**

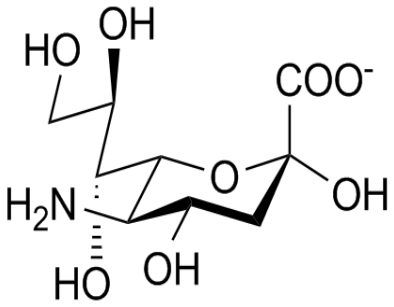
***Keywords*- Neuraminidase inhibitors, cyclohexene analogues, Quantitative structure-activity relationship, Docking studies**

I. INTRODUCTION

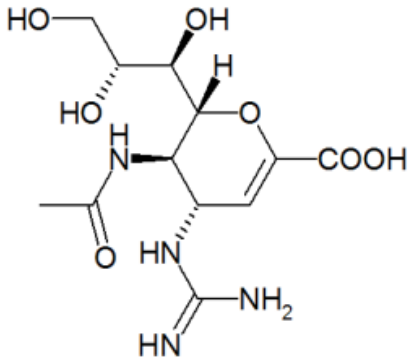
The viral or influenza neuraminidase (NA) is a member of family of neuraminidases which are glycoside hydrolase enzymes (EC 3.2.1.18) that cleave the glycosidic linkages of neuraminic acids. The viral neuraminidase is frequently used as an antigenic determinant found on the surface of the influenza virus. Nine subtypes of influenza neuraminidase are known, many occur only in various species of duck and chicken. Subtypes N1 and N2 have been positively linked to epidemics in man, and strains with N3 or N7 subtype have been identified in number of isolated deaths.

Neuraminidases, also called sialidases, catalyse the termininal neuraminic acid (sialic acid) (**1**) attached to glycoproteins and glycolipids [1,2]. This process is believed to be necessary for the release of newly formed virus from infected cells and for efficient spread of virus in the respiratory tract [3,4]. When influenza virus replicates, it attaches to the cell surface using hemagglutinin, a molecule found on the surface of the virus that binds to sialic acid groups. Sialic acids are found on various glycoproteins at the host cell surface, and the virus exploits these groups to bind at the host cell. In order of the virus to be released from the cell, the sialic acid groups must be enzymatically cleaved from the host glycoproteins. This is done by viral neuraminidase (NA) and thus the NA promotes the release of progeny viruses and the spread of viruses from the host cells. NA also cleaves sialic acid residues from viral proteins preventing aggregation of viruses. Thus because of being a crucial component in the influenza virus replication, the neuraminidase has been found a good target to be exploited to treat the influenza.

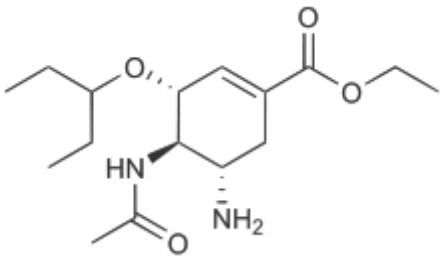
The neuraminidase has been targeted in structure-based design of its inhibitors, resulting in two important drugs, Zanamivir (Relenza, **2**) and Oseltamivir (Tamiflu, **3**), and a new drug, Peramivir (**4**) is under trial. While Zanamivir is administered by inhalation and Oseltamivir orally, Peramivir is being investigated to be given parenterally. The recent emergence of Oseltamivir and Zanamivir resistant human influenza A (H1N1) has, however, necessitated the design and discovery of further NA inhibitors. Attempts are also on to find NA inhibitors which can be administered particularly orally.



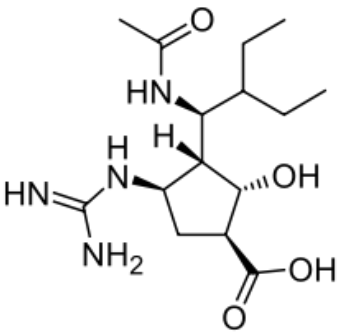
1. Neuraminic acid



2. Zanamivir



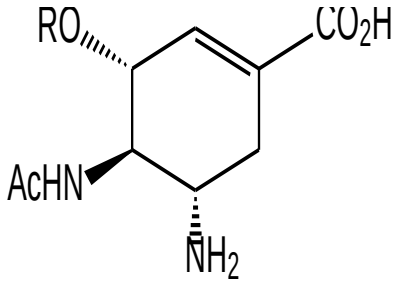
3. Oseltamivir



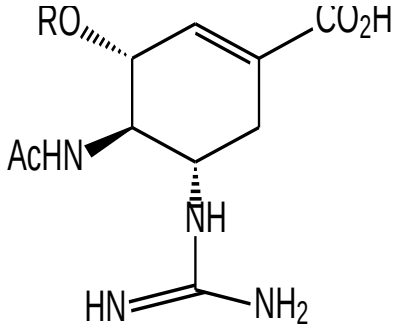
4. Perammivi

It has been proposed that the sialic acid cleavage by NA might proceed via the oxonium cation transition state (Sialosyl cation) [5-7]. Zanamivir is a transition state analogue of this cation which exhibits potent NA inhibitory activity [8]. Taking the cyclohexene ring of this cation as scaffold, some authors [9,10] attempted to design novel NA inhibitors having oral bioavailabilty. The unsaturated sialic acid (N-acetylneuraminic acid [Neu5ac]) derivative, 2-deoxy-2, 3-didehydro-D-N-acetylneuraminic acid (Neu5Ac2en), a sialosyl cation transition-state analogue, is believed to be the most potent inhibitor core template. The structural modifications in Neu5Ac2en have been tried to have more effective inhibitors [11]. Zanamivir, the drug now marketed for treatment of influenza, is a 4-guanidino derivative of Neu5Ac2en. It was designed by Von Itzstein and coworkers [12]. Many other Neu5Ac2en-based compounds have been synthesized and tested for their influenza virus sialidase inhibitory potential. A series of amide-linked C9 modified Neu5Ac2en have been reported by Megesh and colleagues as NEU1 inhibitors [13].

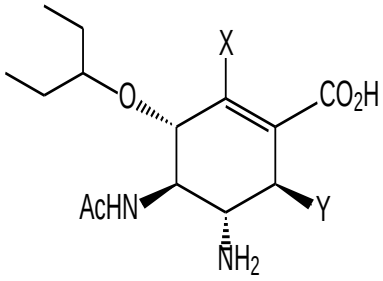
Of the two major proteins on the surface of influenza virus particles, the lectin hemagglutinin protein with three relatively shallow sialic acid-binding sites and the enzyme sialidase with the active site in a pocket, the sialidase constitutes more attractive target than hemagglutinin, as it offers the relatively deeper active site in which low molecular weight inhibitors can make multiple favorable interactions and approachable methods of designing transition state analogues in the hydrolysis of sialosides [14]. After the X-ray crystal structures of several influenza virus sialidases could be available, the structure-based inhibitor design became feasible to discover potent inhibitors of this enzyme [15]. We present here a quantitative structure-activity relationship (QSAR) study on a large series of NA inhibitors (Figure 1) that contain the cyclohexene nucleus of Neu5Ac2en and have been reported by Kim and coworkers in two successive studies [9,10].



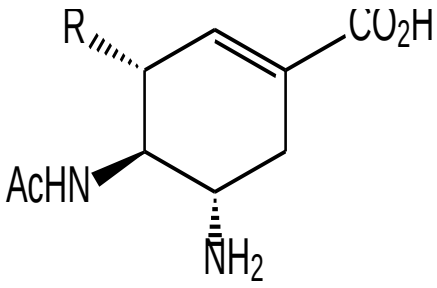
1-26



27-32



33-36



37-50

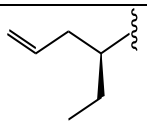
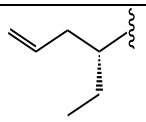
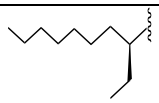
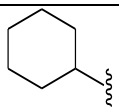
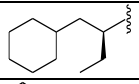
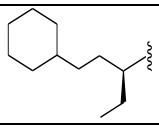
Fig. 1. Compounds with cyclohexene scaffold acting as influenza neuraminidase inhibitors

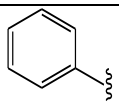
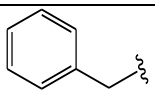
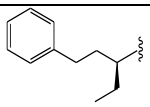
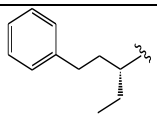
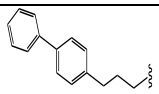
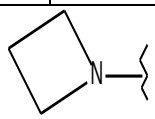
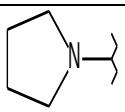
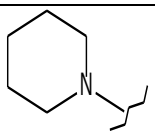
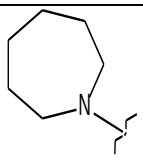
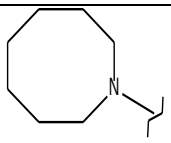
II. MATERIALS AND METHODS

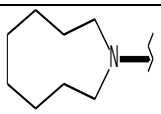
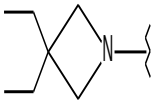
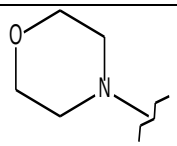
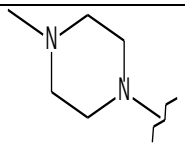
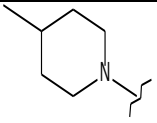
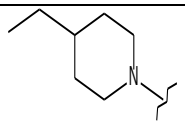
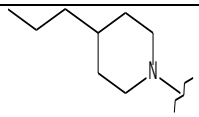
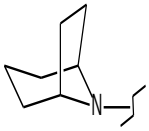
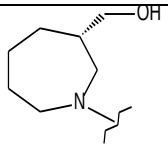
The series of NA inhibitors as shown in Figure 1 were taken from the literature [9,10] and are shown in Table 1 along with their HDAC inhibition activity in terms of log ($1/EC_{50}$), where

EC_{50} refers to molar concentration of compound leading to 50% inhibition of the enzyme. We performed a simple multiple linear regression (MLR) on these compounds. The total 50 compounds of Table 1 were divided into two subsets: the training set comprising of 38 compounds and the test set comprising of 12 compounds. The test set compounds in the table are marked with a superscript 'b' and are given in bold. Compounds for the test set were selected keeping in view the wide variation in structures as well as in activities of compounds. The physicochemical parameters that were found to be significant in multiple linear regression (MLR) were only calculated molar refractivity (CMR) and polarizability (Pol), which were calculated by ACD/ChemSketch (version 11.0). Several other parameters were calculated but found of no use. The values of these parameters for all the compounds are listed in Table 1. In deriving QSAR model, two indicator variables, I_1 and I_2, were also used. I_1 was used with a value of 1 for an R-substituent having a branched group CH_3CH_2 and I_2 was used with a value of 1 for an R-substituent that is a linear chain with more than 6 carbon atoms.

TABLE I. NEURAMINIDSE INHIBITORS (1-50, FIG. 1) and THEIR INHIBITION POTENCY and PHYSICOCHEMICAL PARAMETERS

Compd	R	CMR	Pol	I_1	I_2	log(1/EC_{50})		
						Obsd[a]	Pred,Eq(1)	Pred,LOO
1[b]	H	5.34	20.56	0	0	5.20	5.06	-
2	CH_3	5.80	22.47	0	0	5.43	5.52	5.56
3	CH_3CH_2	6.27	24.31	0	0	5.70	6.01	6.08
4	$CH_3(CH_2)_2$	6.73	26.14	0	0	6.74	6.40	6.37
5[b]	$CH_3(CH_2)_3$	7.20	27.98	0	0	6.52	6.75	-
6	$CH_3(CH_2)_4$	7.66	29.82	0	0	6.70	7.01	7.02
7[b]	$CH_3(CH_2)_5$	8.12	31.65	0	0	6.82	7.21	-
8	$CH_3(CH_2)_6$	8.59	33.49	0	1	6.57	6.49	6.49
9	$CH_3(CH_2)_7$	9.05	35.32	0	1	6.74	6.57	6.54
10	$CH_3(CH_2)_8$	9.51	37.16	0	1	6.68	6.58	6.52
11	$CH_3(CH_2)_9$	9.98	39.00	0	1	6.22	6.56	6.65
12	$(CH_3)_2CHCH_2$	7.20	27.97	0	0	6.70	6.76	6.76
13	$CH_3CH_2(CH_3)CH$*	7.20	27.97	1	0	8.00	8.23	8.22
14	$CH_3CH_2(CH_3)CH$* (S)	7.20	27.97	1	0	8.05	8.23	8.24
15[b]	$(CH_3CH_2)_2CH$	7.66	29.81	1	0	9.00	8.48	-
16		8.10	31.55	1	0	9.00	8.69	8.64
17		8.10	31.55	1	0	8.52	8.69	8.69
18		9.98	38.99	1	0	9.00	8.92	8.94
19		7.95	30.83	0	0	7.22	7.25	7.25
20[b]		9.80	38.17	1	0	7.80	9.01	-
21		10.26	40.00	1	0	9.00	8.94	8.89

Compd	R	CMR	Pol	I_1	I_2	log(1/EC_{50})		
						Obsd[a]	Pred,Eq(1)	Pred,LOO
22		7.85	30.43	0	0	6.28	7.20	7.25
23[c]		8.31	32.27	0	0	6.21	-	-
24[b]		10.17	39.60	1	0	9.52	8.98	-
25[b]		10.17	39.60	1	0	7.92	8.98	-
26		11.75	45.74	0	0	7.05	6.92	6.54
27	H	6.36	23.48	0	0	7.00	6.92	6.88
28	$CH_3(CH_2)_2$	7.76	29.10	0	0	8.70	7.85	7.64
29	$CH_3(CH_2)_3$	8.22	30.93	0	0	8.52	8.04	7.94
30	$CH_3CH_2(CH_3)CH^*$ (R)	8.22	30.85	1	0	9.30	9.56	9.65
31	$CH_3CH_2(CH_3)CH^*$ (S)	8.22	30.85	1	0	9.30	9.56	9.65
32	$(CH_3CH_2)_2CH$	8.68	32.68	1	0	9.30	9.69	9.81
33	X: H, Y: H	7.57	29.81	1	0	9.00	8.19	8.03
34[b]	CH_3	8.03	31.55	1	0	5.64	8.46	-
35	F, H	7.59	29.76	1	0	8.52	8.29	8.23
36[b]	H, CH_3	8.12	31.63	1	0	5.82	8.69	-
37[b]		6.77	26.18	0	0	6.00	6.51	-
38		7.23	28.02	0	0	6.51	6.83	6.83
39		7.70	29.86	0	0	7.60	7.11	7.12
40		8.16	31.69	0	0	7.46	7.31	7.31
41		8.62	33.53	0	0	7.59	7.44	7.44

Compd	R	CMR	Pol	I_1	I_2	log(1/EC$_{50}$)		
						Obsd[a]	Pred,Eq(1)	Pred,LOO
42		9.09	35.36	0	0	7.12	7.55	7.55
43		8.62	33.53	0	0	6.58	7.44	7.49
44[b]		7.39	28.71	0	0	6.59	6.88	-
45[c]		8.07	31.39	0	0	5.72	-	-
46		8.16	31.68	0	0	7.29	7.319	7.31
47		8.62	33.52	0	0	7.28	7.45	7.44
48[b]		9.09	35.35	0	0	7.40	7.54	-
49		8.36	32.70	0	0	7.49	7.23	7.16
50		8.78	34.13	0	0	8.10	7.50	7.42

[a] Taken from ref. 9 and 10. [b]Compounds taken for the test set. [c]Not used in regression analysis as they were outliers

III. RESULTS AND DISCUSSION

III.A QSAR 5Results

When a multiple linear regression analysis was performed on the compounds of the training set, the correlation obtained was as:

$$\log(1/EC_{50}) = 5.534(\pm1.697)CMR - 0.145(\pm0.065)CMR^2 - 0.705(\pm0.339)Pol + 1.469(\pm0.328)I_1 - 0.887(\pm0.514)I_2 - 5.857(\pm4.841)$$

$$n = 36,\ r = 0.934,\ r^2_{cv} = 0.83,\ r^2_{pred} = 0.988,\ s = 0.421,\ F_{5,30} = 40.67,$$

$$(CMR)_o = 19.08 \quad (1)$$

In Eq. (1), n is the number of data points, r is the correlation coefficient, r^2_{cv} is the square of the cross-validated correlation coefficient obtained from leave-one-out (LOO) jackknife procedure, s is the standard deviation, F is the Fischer ratio between the variances of calculated and observed activities, and the data within the parentheses with ± sign are 95% confidence intervals. The figures within parenthesis following

the F-value is the standard F-value at 99% level. The values of these statistical parameters exhibit that the correlation obtained is quite significant. The internal validity of the correlation is judged by the value of its r^2_{cv} which is calculated as:

$$r^2_{cv} = 1 - [\Sigma_i (y_{i,\ obsd} - y_{i,\ pred})^2 / \Sigma_i (y_{i,obsd} - y_{av,obsd})^2] \quad (2)$$

where $y_{i,obsd}$ and $y_{i,pred}$ are the observed and predicted (from LOO) activity values of compound *i*, respectively, and $y_{av,obsd}$, is the average of the observed activities of all compounds used in the correlation. The correlation is supposed to be valid if $r^2_{cv} > 0.60$. From this point of view, the correlation expressed by Eq. (1) seems to be quite valid. However, the predictive ability of any correlation equation is judged by predicting the activity of the compounds in the test set using it and calculating the value of r^2_{pred}, which is defined as:

$$r^2_{pred} = 1 - [\Sigma_i (y_{i,obsd} - y_{i,pred})^2 / \Sigma_i (y_{i,obsd} - y_{av,obsd})^2] \quad (3)$$

where $y_{i,obsd}$ and $y_{i,pred}$ refer to the observed and predicted (from eq. obtained) activity of compound *i* in the test set and $y_{av,obsd}$ is same as in Eq.(2). A value of r^2_{pred} equal to 0.988, signifies a good predictive ability of the correlation. The activity values predicted from this equation for the test set compounds are given (in bold) in Table **1**. A comparison shows that these predicted values are in very good agreement with the corresponding observed ones. In the training set also, the calculated values are found to be in excellent agreement with the observed ones. All these observations can be better visualized in the graphs drawn between the predicted and observed activities (Figure 2). It is also to be noted that all the five parameters of the Eq.(1) are statistically quite significant in the correlation. If they are removed one by one, a significant successive decrease in the significance of the correlation is observed (Eqs. 4-6).

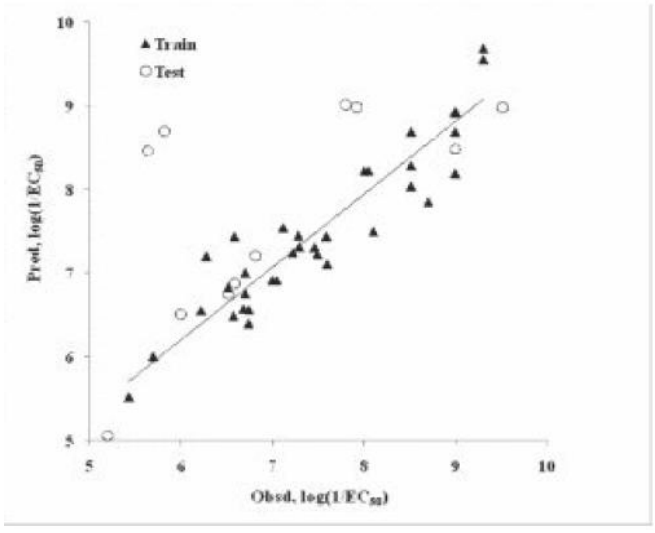


Fig. 2. A Plot Berween observed and predicted activities for training and test set compounds

$$\log(1/EC_{50}) = 5.535(\pm1.982)CMR - 0.127(\pm0.075)CMR^2 - 0.805(\pm0.390)Pol + 1.623(\pm0.369)I_1 - 4.096(\pm5.527)$$

$$n = 36,\ r = 0.905,\ s = 0.49,\ F_{4,31} = 34.89 \quad (4)$$

$$\log(1/EC_{50}) = 6.977(\pm3.647)CMR - 0.170(\pm0.140)CMR^2 - 0.974(\pm0.724)Pol - 7.097(\pm10.233)$$

$$n = 36,\ r = 0.588,\ s = 0.92,\ F_{3,32} = 5.63 \quad (5)$$

$$\log(1/EC_{50}) = 3.282(\pm2.622)CMR - 0.180(\pm0.152)CMR^2 - 7.018(\pm11.183)$$

$$n = 36,\ r = 0.438,\ s = 1.01,\ F_{2,33} = 3.91 \quad (6)$$

$$\log(1/EC_{50}) = 0.215(\pm0.327)CMR + 5.663(\pm2.703)$$

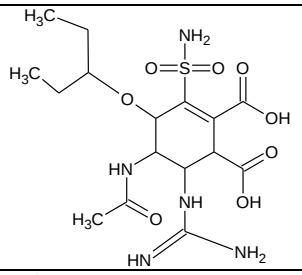
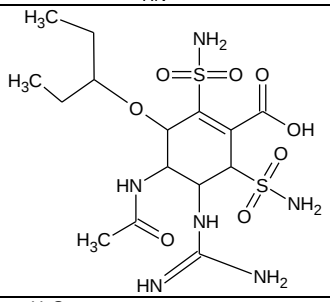
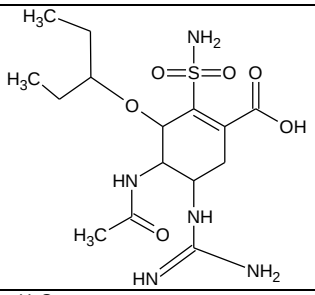
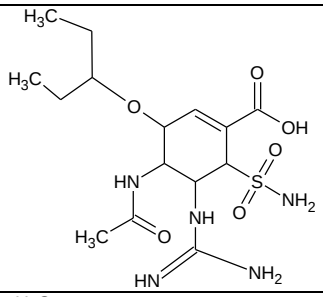
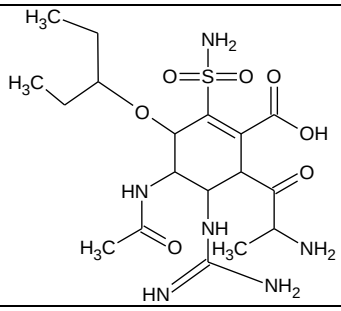
$$n = 36,\ r = 0.224,\ s = 1.112,\ F = 1.795 \quad (7)$$

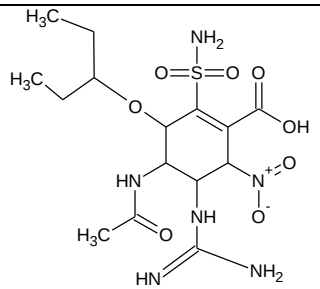
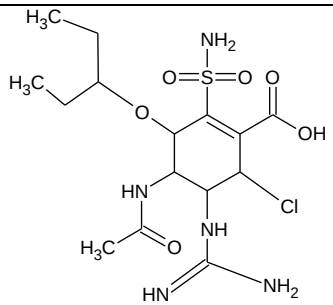
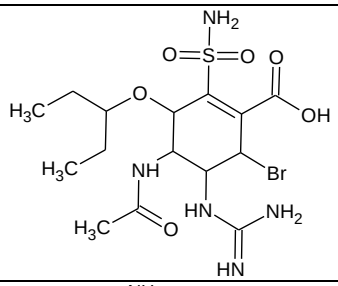
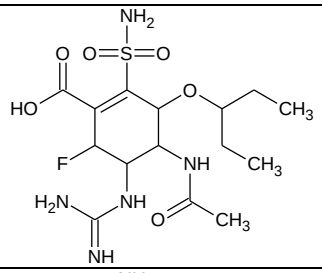
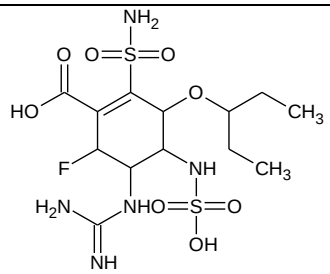
From mechanistic point of view, Eq. (1) suggests that activity would be controlled by molecular refractivity (MR) of the molecule. The MR refers to the size of the molecule and its positive coefficient indicates the dispersion interaction with the receptor. However, since activity is shown to be parabolically correlated to CMR, CMR attains an optimum value equal to 19.08, beyond which the activity will start decreasing. This suggests that very bulky molecule will not be conducive to activity because of steric problem. Equation (1) also shows the negative dependence of activity on polarizability of the molecule, suggesting that highly polarized molecule might not be preferred due to probably unwanted electrostatic interaction between the molecule and the receptor.

Out of the two indicator parameters I_1 and I_2, the positive coefficient of the former suggests that a branched chain R-substituent may be beneficial to activity, while the negative coefficient of the latter suggest that an R-substituent with long linear chain may be detrimental to the activity. This may be due to some steric problem created by long linear chain.

However, using Eq.(1), we have predicted the activity of some new prospective compounds with high potency (Table 2). The activities of these compounds are higher than any compound in the present series (Table 1). On these predicted compounds we then performed docking studies to see their bindings with the protein (NA).

TABLE II. SOME PREDICTED COMPOUNDS and THEIR ACTIVITY PREDICTED from EQ. (1)

Compd	Structure	CMR	CMR^2	Pol	I_1	I_2	Pred, Eq. (1)
1		10.58	111.86	39.81	1	0	9.86
2		11.17	124.66	41.93	1	0	9.76
3		9.92	98.48	37.31	1	0	9.95
4		9.92	98.48	37.31	1	0	9.95
5		11.7196	137.349	43.81	1	0	9.67

Compd	Structure	CMR	CMR^2	Pol	I_1	I_2	Pred, Eq. (1)
6		10.5353	110.993	39.55	1	0	9.94
7		10.4152	108.476	39.13	1	0	9.93
8		10.7008	114.507	40.3	1	0	9.82
9		9.9393	98.7897	37.26	1	0	10.02
10		10.00	100.03	37.54	1	0	9.99

III.B Docking results

The docking study was performed for all the compounds predicted. For this the FlexX software was used and the X-ray structure of neuraminidase bound with Zanamivir was taken from protein data bank (PDB entry 3b7e). The docking score of each compound is given in Table 3

TABLE III DOCKING SCORES of PREDICTED COMPOUNDS as COMPARED to that of ZENAMIVIR

Compd	Zenamivir	1	2	3	4	5	6	7	8	9	10
Score	-59.84	-54.84	-50.45	-58.20	-41.36	-48.06	-53.90	-58.14	-45.20	-55.96	-56.10

Along with them is given the docking score of Zanamivir also for comparison. One can see that except for compound 4, 5, and 8, all other compounds have their docking score comparable to that of Zanamivir. The docking score (given here in kJ/mole) gives the stability of binding of compounds with the enzyme. Table 2 shows that of all the predicted compounds, compound 9 has the highest potency. Therefore, as a representative compound, its docked structure within the enzyme NA is shown in Figure 3 and that of Zanamivir in Figure 4 for comparison. In Figure 3, compound 9 is shown to have 13 hydrogen bondings with NA, while in Figure 4 Zanamivir is shown to have only 14 hydrogen bondings, just 1 more than compound 9.

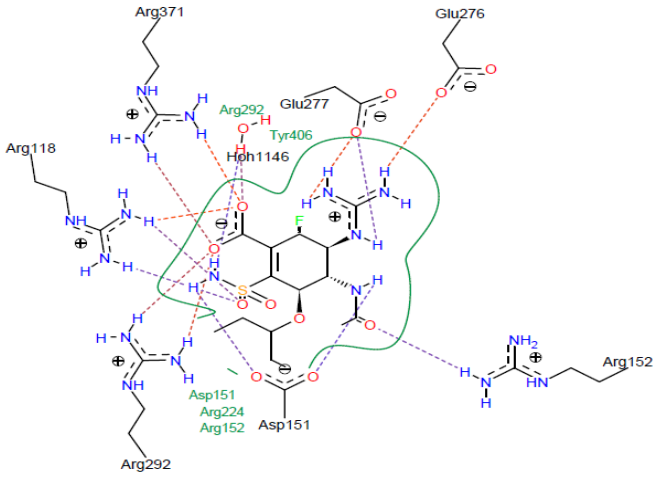


Fig. 3. Binding of Zanamivir with neuraminidase. Figure shows the hydrogen bonding of the compound with various amino acid residues in the enzyme

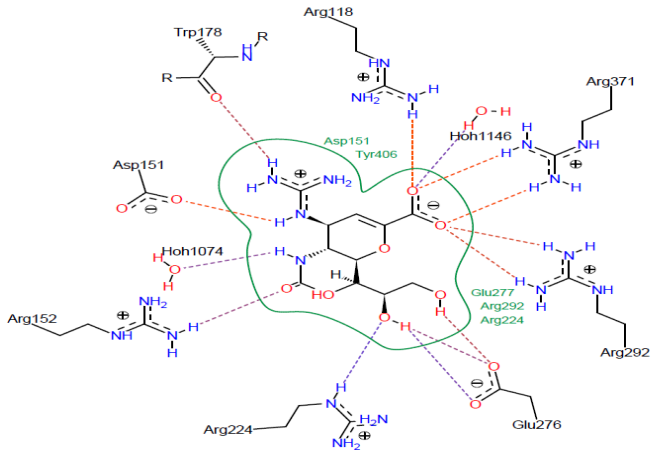


Fig. 4. Binding of Zanamivir with neuraminidase. Figure shows the hydrogen bonding of the compound with various amino acid residues in the enzyme.

IV. CONCLUSION

The compounds with cyclohexene scaffolds acting as NA inhibitors are shown to have dispersion interaction with the neuraminidase enzyme and new analogues predicted by QSAR model are found to have similar interaction with NA as the FDA approved NA inhibitor, Zanamivir.

ACKNOWLEDGMENT

This research work is funded by Madhya Pradesh Council for Science and Technology (MPCOST), Bhopal under the research project grant, section letter no.

REFERENCES

[1]. Colman, P. M.(1989) In *The Influenza Viruses: Influenza Virus Neuraminidase, Enzyme and Antigen*; Krug, R.M.,Ed.;Plenum Press: New York,; pp175-218.

[2]. Colman, P. M.(1994) Influenza virus neuraminidase: structure, antibodies, and inhibitors. *Protein Science*. 1994,3,1687-1696.

[3]. Palese, P.; Tobita, K.; Ueda, M.; Compans, R. W. (1974) Characterization of temperature sensitive influenza virus mutants defective in neuraminidase. *Virology*, *61*, 397-410.

[4]. Liu, C.; Air, G. M.(1993) Selection and characterization of a neuraminidase-minus mutant of influenza virus and its rescue by cloned neuraminidase genes. *Virilogy*, *193*, 1-5.

[5]. Janakiraman, M. N.; White, C. L.; Laver, W. G.; Air, G. M.; Luo, M. (1994) Structure of Influenza virus neuraminidase B/Lee/40 complexed with sialic acid and a dihydro analogue at 1.8-Å resolution: Implication for the catalytic mechanism. *Biochemistry*, *33*, 8172-8179.

[6]. Chong, A. K.; Pegg, M. S.; Taylor, N. R.; Von Itzstein, M (1992).; Evidence for a sialosyl cation transition-state complex in the reaction of sialidase from influenza virus. *Eur. J. Biochem.*, *207*, 335-343.

[7]. Taylor, N. R.; Von Itzstein, M. (1994) Molecular modeling studies on ligand biding to sialidase from influenza virus and the mechanism of catalysis. *J. Med. Chem.*, *37*, 616-624.

[8]. Von Itzstein, M.; Wu, W.-Y.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B.; Phan, T. B.; Smyth, M. L.; White, H. F.; Oliver, S. W.; Colman, P. M.; Varghese, J. N.; Ryan, D. M.; Would, J. M.; Bethell, R. C.; Hotham, V. J.; Cameron, J. M.; Penn, C. R. (1993) Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature (London)*, *363*, 418-423.

[9]. Kim, C. U.; Lew, W.; Williams, M. A.; Wu, H.; Zhang, L.; Chen, X.; Escarpe, P. A.; Mendel, D. B.; Laver, W. G.; Stevens, R. C. (1998) Strucure-activity relationship studies of novel carbocyclic influenza neuraminidase inhibitors. *J. Med. Chem.*, *41*, 2451-2460.

[10]. Lew, W.; Wu, H.; Chen, X.; Graves, B. J.; Escarpe, P. A.; MacArthur, H. L.; Mendel, D. B.; Kim, C. U. (2000) Carbocyclic influenza neuraminidase inhibitors possessing a C_3-cyclic amine side chain: synthesis and inhibitory activity. *Bioorg. Med. Chem. Lett.*, *10*, 1257-1260.

[11]. Fgedi, P. (2006) *The Organic Chemistry of Sugars*, , Washington, DC: Taylor & Francis. pp. 822– 823.

[12]. Von Itzstein, M.; Wu W.-Y.; Jin, B. (1994) The synthesis of 2,3-didehydro-2,4-dideoxy-4-guanidinyl-N-acetylneuraminic acid: a potent influenza virus sialidase inhibitor. *Carbohyd. Res.*, *259*, 301-305.

[13]. Magesh, S.; Moriya, S.; Suzuki, T.; Miyagi, T.; Ishida, H.; Kiso, M. (2008) Design, synthesis, and biological evaluation of human sialidase inhibitors. Part 1: selective inhibitors of lysosomal sialidase (NEU1). *Bioorg. Med. Chem. Lett.*, *18*, 532-537.

[14]. Drickamer, K.; Taylor, M. P.(2006) *Introduction to Glycobiology*. Oxford [Oxfordshire]: Oxford University Press., pp. 177–178.

[15]. Dyason, J. C.; Von Itzstein, M.(2001) Anti-influenza virus drug design: sialidase inhibitors. *Aust. J. Chem*, *54*, 663–670.