ISSN:0975-3583,0976-2833 VOL12,ISSUE06,2021

3D QSAR analysis on substituted Cyanopyrrolidine derivatives as Dipeptidyl peptidase –IV inhibitors

Govind Sharma[1*], Arvind Kumar Jha[2], Ajazuddin[3], Yogesh Vaishnav[2],Shekhar Verma[4], Arpan Kumar Tripathi[2]

1. Rungta College of Pharmaceutical Sciences and Research, Kohka,Bhilai, Chhattisgarh, India.
2. Faculty of Pharmaceutical Sciences, Shri Shankaracharya Technical Campus, JunwaniBhilai, Chhattisgarh, India.
3. Department of Pharmaceutics, School of Pharmacy & Technology Management, SVKM's NMINS, Shirpur 425405, Maharashtra, India.
4. University college of Pharmacy, Pt. Deendayal Upadhyay Memorial Health Sciences and Ayush University of Chhattisgarh, Raipur, India.

Corresponding Author

*Govind Sharma

Email: govindsharma10aug@gmail.com

Abstract:

In the current research work twenty five molecules of substituted cyanopyrrolidine derivatives were subjected to develop best QSAR model. The best model was achieved by means of k nearest neighbor molecular field analysis associated by stepwise forward backward method using partial least square regression method. Random selection method was used to differentiate total number of 25 molecules into 20 of training and 5 of test sets. Total ten trials were run to identify best model in which trial 1 gave significant statistical equation. The best model consist of predictive internal q2 value of 0.9451 & external predictivity (pred_r2 = 0.9402). The best model bearing three steric descriptors at different points such as S_667, S_319, S_916 having values in the range of -3.2599 22.9668, -0.0748665 18.0689, -6.33242 6.35972. The obtained steric descriptors play a key function in determining biological activity. The grid matrix with different steric points helps to understand the relationship of structural function of substituted cyanopyrrolidine derivatives and its biotic action.The obtained statistical data would be considered to design new potent DPP IV inhibitors.

Keywords: Dipeptidyl peptidase IV, Partial least square regression, Drug design, Diabetes, 3D QSAR, Cyanopyrrolidine.

Introduction:

The 9th International Diabetes Federation (IDF) Atlas has experienced a rise of 463 million people (20-79 years) and is expected to reach 700 million by 2045(1). India is the DM epicentre; it affects 77.0 million total adults and in 2020 it reached up to 8.9% for the adult population (2). High levels of blood glucose in the DM can raise the danger to heart and blood vessels, eyes, kidneys, nerves and teeth, and can impact serious illnesses. Furthermore, diabetic individuals are also at greater risk for infection such as COVID 19. The main cause of cardiovascular disease, blindness, renal failure and lower limb amputation in most high-income nations is diabetes (3).

Different kinds of DM described as a result of pancreatic beta cell injury by their pathophysiology, like, Type I DM. The environmental (viral) and autoimmune causes are the reason for this harm (4). The genetically sensitive type I DM has also emerged for the development of type 1 diabetes (5,6). Type II DM is defined by an insulin sensitivity which arises when the body becomes insulin resistant or the pancreas stops generating sufficient insulin or possible beta cell loss. This causes the liver, muscle cells and fat cells to decrease in glucose consumption. Type II DM is the state of insulin resistance to beta cell dysfunction linked with it. In the beginning, insulin production increases compensatorily, maintaining normal levels of glucose (7).

Dipeptidyl-peptidase (DPP) IV is a type II transmembrane protein, which is also cleaved off the membrane and released into the circulation by a process called shedding(8). DPP-IV is a multifunctional enzyme, which serves as a binding partner for numerous peptides, among which are adenosine deaminases (ADA) and extracellular matrix proteins(9). Moreover, as a serine protease, DPP-IV cleaves numerous substrates, which further amplifies its complexity of action. Thus, DPP-IV is involved in signaling processes, immune cell activation, and its dysregulated expression and release is associated with numerous diseases(10). The dipeptidyl peptidase-4 (DPP-IV) inhibitors are

ISSN:0975-3583,0976-2833 VOL12,ISSUE06,2021

a newer class of oral drugs for the treatment of type 2 diabetes. They inhibit the breakdown of glucagon-like peptide-1 (GLP-1) and increase the incretin effect in patients with type 2 diabetes(11,12).

DPP-IV inhibitors of 2-cyano-4-fluoro-1-thiovalylpyrrolidine, utilised in Type II D M therapy. DPP-4 inhibition extends and increases the activity of incretin which plays a major role in insulin secretion and blood glucose management (13,14). Fukushima et al(15) have been reported to have 2-cyano-4-fluoropyrrolidines as powerful DPP-IV inhibitors and to have a more helpful inhibitor by changing 1-position of pyrrolidine.

Qsar is an arithmetic procedure which in homologous series provides a quantitative connection between structure and biological activity. Significant 3D QSAR model designs and validates are the motto of current scientific activity. We have identified a number of cyanopyrrolidine substituted compounds with DPP-IV inhibitory activity (15).

Experimental:

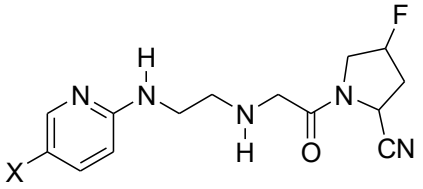
2.1. Workspace:

Present research work was carried out by quantitative structure activity relationship software i.eVLifeMDS of version 4.0.(16). To perform 3 dimensional quantitative structure activity relationships associated with k-nearest neighbor molecular field analysis; molecular design suite software was utilized and was installed in the Dell desktop computer system with configuration of good processor as well as operating system of windows XP.

2.2. Set of molecules:

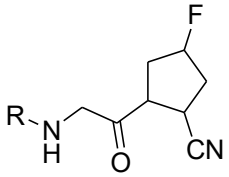
In this present research work we have selected twenty five molecules which were reported earlier as DPP IV inhibitors(15)**.** The reported IC50 values of twenty five molecules have been converted into pIC50 & were shown in Table 1 & 2.

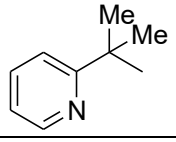
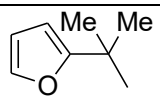
Table-1: Structure of substituted cyanopyrrolidine derivatives and their biotic actions



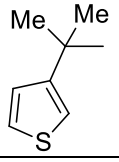
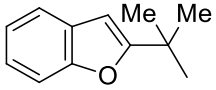
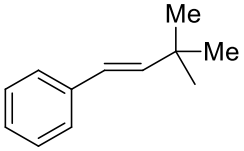
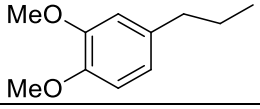
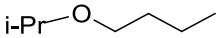
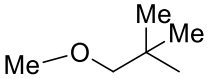
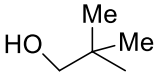
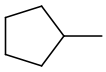
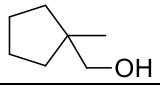
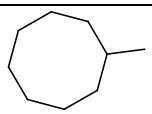
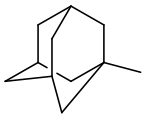
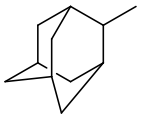
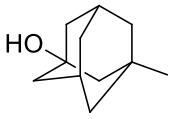
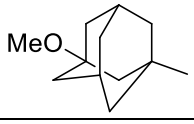
S.NO	MOLECULE	X	IC_{50} (NM)	$log(1/IC_{50})$
01	2a	CN	1.1	8.959
02	2b	H	2.8	8.553
03	2c	Cl	2.7	8.569
04	2d	$CONH_2$	2.4	8.62

Table - 2: Structure of substituted cyanopyrrolidine derivatives and their biotic actions

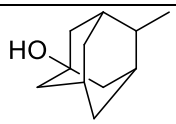


S.NO	MOLECULE	R	IC_{50} (nM)	$log(1/IC_{50})$
05	12a		22	7.658
06	12b		8.6	8.066

ISSN:0975-3583,0976-2833 VOL12,ISSUE06,2021

07	12c		12	7.921
08	12d		88	7.056
09	12e		16	7.796
10	12f		27	7.569
11	12g	t-Bu-	2.9	8.538
12	12h	i-Pr-	7.8	8.108
13	12i		8.7	8.06
14	12j		13	7.886
15	12k		4.6	8.337
16	12m		33	7.481
17	12n		3.1	8.509
18	12o		2.6	8.585
19	12p		3.3	8.481
20	12q		3.3	8.481
21	12r		4.1	8.387
22	12s		3.1	8.509
23	12t		3.1	8.509
24	12u		8.3	8.081

ISSN:0975-3583,0976-2833 VOL12,ISSUE06,2021

25	12v		5.3	8.276

The unicolumn statistics of both sets (test and training) show clear-cut medley of test and training sets. An ideal unicolumn statistics has following criteria i.e, maximum of the training set was extra than that of test set and the slightest of training set was fewer than or equivalent to that of test set. The unicolumn statistics values were shown in Table 3.

Table 3. Unicolumn statistics for model 3D-RAN-80%-KNN-SWFB-PLS-TRIAL1

Model no.	Column name	Average	Maximum	Minimun	StdDev	Sum
MODEL 1	**Training Set**	8.2138	8.9590	7.4810	0.3989	164.2750
	Test set	8.1440	8.5850	7.0560	0.6431	40.7200

2.3. Methodology:

Two dimensional structures of all twenty five molecules were sketch by means of 2 dimensional draw tool which was available in the software. Once when all the structures were drawn in 2D then it was converted into 3 dimensional structures by the aid of 3D converting tool. After conversion of 2 dimensional molecules into 3 dimensional, all the converted molecules were saved as dot (.)Mol2 file format. In the next step, we had opted MMFF method to minimize the energies.To minimize the energy and to correct the geometry of the molecules methods like energy minimization and geometry optimization was done. The method applied to minimize the energy and to optimize the geometry was MMFF (Merck Molecular Force Field)(17).

Other important parameters like highest number of cycle were set to 1000, dielectric constant value was 1.0, root mean square gradient value at 0.01 also the values of steric and electrostatic energy within the cutoff range of 30 and 10kcal/mol.

2.4. Molecular molding and organization:

Out of twenty five molecules, the most active molecule was chosen to align other twenty four molecules and this was done by the help of template based alignment method in which the most active molecule (2a) was considered as template. All twenty five aligned molecules were shown in Figure 1.

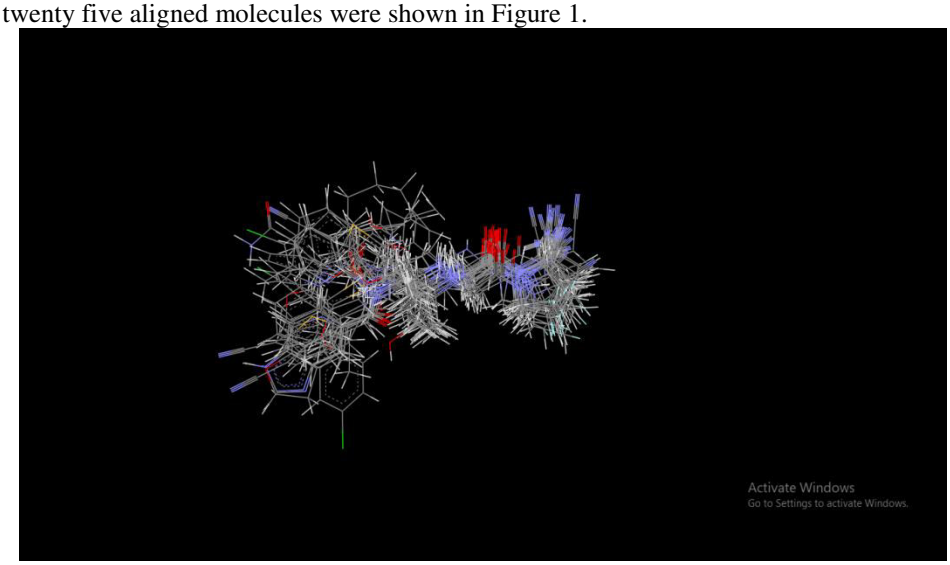


Figure 1: 3D alignment of twenty five molecules.

2.5. Computation of field descriptors:

Calculation of field descriptors was done in the worksheet page which was present in the 3D QSAR window. After all calculations of descriptors, various values of different descriptors were obtained in different columns. Before doing analysis of 3d qsar, invariable columns were eliminated from the worksheet and final worksheet was saved. Following framework has been chosen for descriptors calculations. The charge type was selected as gasteiger-

ISSN:0975-3583,0976-2833 VOL12,ISSUE06,2021

marsili/delre, carbon probe atom was set as charge 1. Other important field descriptors like steric and electrostatic were in the cut off limit of 30 & 10kcal/mol respectively.

2.6. K-Nearest Neighbor Molecular Field Analysis (KNNMFA):

A kNN MFA (k Nearest Neighbor Molecular Field Analysis) is an innovative methodology for correlating molecular field descriptors with biotic activity. This method was applied after training and test sets of molecules were generated. Once when training and test sets were generated KNN methodology was applied to different descriptors shaped over the lattice. At outline spot of the lattice important interaction energies like steric as well as electrostatic energies are figured out by the aid of methyl probe with one charge. The interaction energy values of steric as well as electrostatic helps to construct bond formation.

2.7. KNN-MFA associated with Stepwise Forwardbackward (SW-FB) Variable Selection Method:

In present research work best model was generated by means of k-nearest neighbor molecular field analysis annealed with variable selection method named as stepwise forward backward (SWFB) method in which following framework has been set such as 0.5 for cross connection edge, q2 should be criteria for term selection, in & out F-test values were in the range of 4 & 3.99 respectively. Other important criteria such as cutoff variance set to 2 kcal/mol Å, auto scaling technique was used and distance based weighted average method was used as estimate method.

3. Results and discussion:

To develop most significant 3D QSAR models, twenty five molecules bearing substituted cyanopyrrolidine nucleus was subjected to develop models. Ten trials were done to construct 3D QSAR models by the aid of different selection method like random and manual selection method. After carryout ten trials Model 1 was chosen as significant as it consist of very good values of q2 & predicted r2. Model 1 consists of following parameters.

Optimum component = 4, n = 20, Degree of freedom = 15, q2= 0.9451, r2= 0.9689, q2_se = 0.1052, Predr2 = 0.9402, pred_r2se = 0.1584. Table-4 consists of different parameters.

Table-4: Results of 3D-QSAR analysis using KNN method (k nearest neighbor) by random selection 80%. with PLS (partial least square) regression analysis used as variable selection.

3D QSAR Model 1- Random Selection-80%-SWFB-PLS-KNN-Trial 1.

S.No	TRIAL	TEST SET MOLECULES	3D-RAN-80%-SWFB-PLS
01	01	12 d 12 g 12 i 12 o 12 q	Optimum Components = 4 n = 20 Degree of freedom = 15 r2 = 0.9689 q2 = 0.9451 F test = 116.7912 r2_se = 0.0792 q2_se = 0.1052 pred_r2 = 0.9402 pred_r2se = 0.1584

The most significant 3D QSAR model was achieved by the aid of random selection method in which the whole series of molecules were estranged into sets of training and test molecules followed by 80%. The significant model consist of 20 molecules of training set and the remaining 05 molecules were comes under test set. The obtained arithmetical equation proves that the achieved model has good extrapolative capability of 95%. The achieved model in 3D grid matrix consist of following steric descriptors at different points such as S_667, S_319, S_916 having values in the range of -3.2599 22.9668, -0.0748665 18.0689, -6.33242 6.35972. The obtained steric descriptors play a key function in determining biological activity. Figure 2 shows the contribution of different steric descriptors.

ISSN:0975-3583,0976-2833 VOL12,ISSUE06,2021

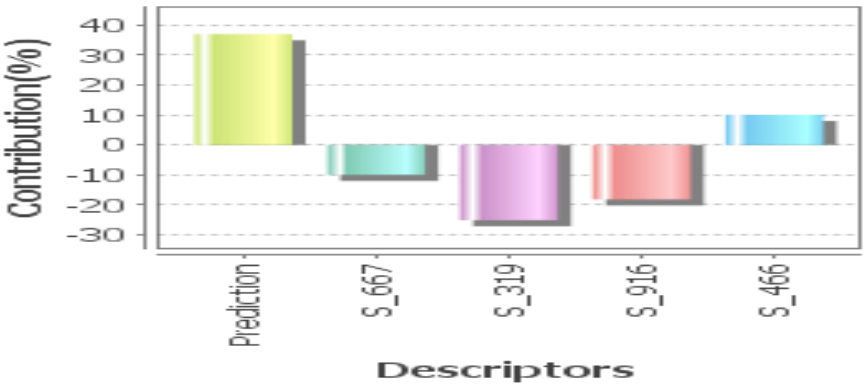


Figure 2: Contribution chart of steric descriptors

The real and forecast activities of all the twenty five molecules in training as well as test set were in good harmony with their personal biotic activities. The actual & predicted values of all the twenty five molecules were shown in Table 5.

Table 5: Actual & predicted value of twenty five molecules.

Molecules	Actual	Predicted
2a	8.959	8.9816
2b	8.553	8.48166
2c	8.569	8.57101
2d	8.62	8.62187
12a	7.658	7.82588
12b	8.066	8.04841
12c	7.921	7.83912
12d*(Test set)	7.056	6.96545
12e	7.796	7.79194
12f	7.569	7.63006
12g*(Test set)	8.538	8.75913
12h	8.108	7.94943
12i*(Test set)	8.06	7.88057
12j	7.886	7.87613
12k	8.337	8.39664
12m	7.481	7.51339
12n	8.509	8.4997
12o*(Test set)	8.585	8.48535
12p	8.481	8.43808
12q*(Test set)	8.481	8.44752
12r	8.387	8.46076
12s	8.509	8.46076
12t	8.509	8.59734
12u	8.081	8.04867
12v	8.276	8.24257

The real and forecast biological activity of training and test set was depicted in Figure 3

ISSN:0975-3583,0976-2833 VOL12,ISSUE06,2021

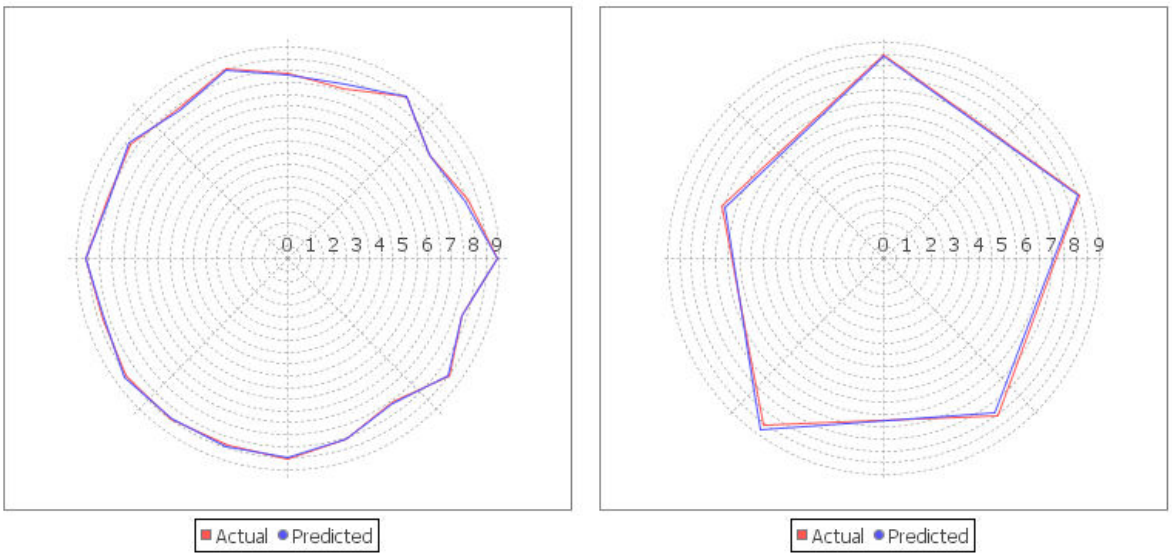


Figure 3:Real and Forecast activity of training & test set molecules.

The green colored ball represents the steric field descriptors and the negative values of all the three steric descriptors indicate the need of less bulky group to enhance the biotic activity & were depicted in figure 4.

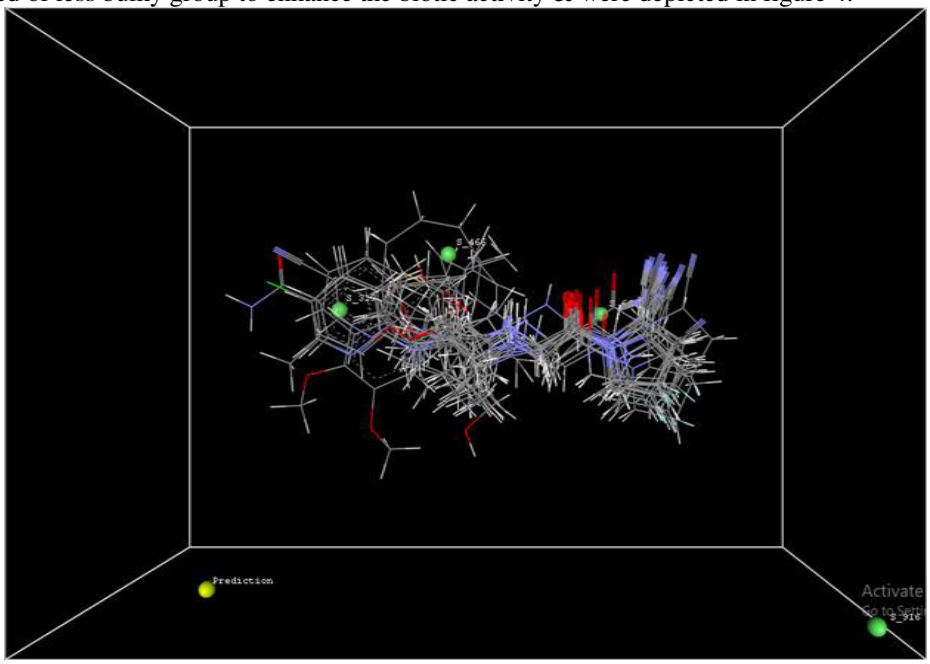


Figure 4: Grid matrix showing points of steric descriptors

The fitness plot of real verses forecast biological activity of model 1 was depicted in figure 5

ISSN:0975-3583,0976-2833 VOL12,ISSUE06,2021

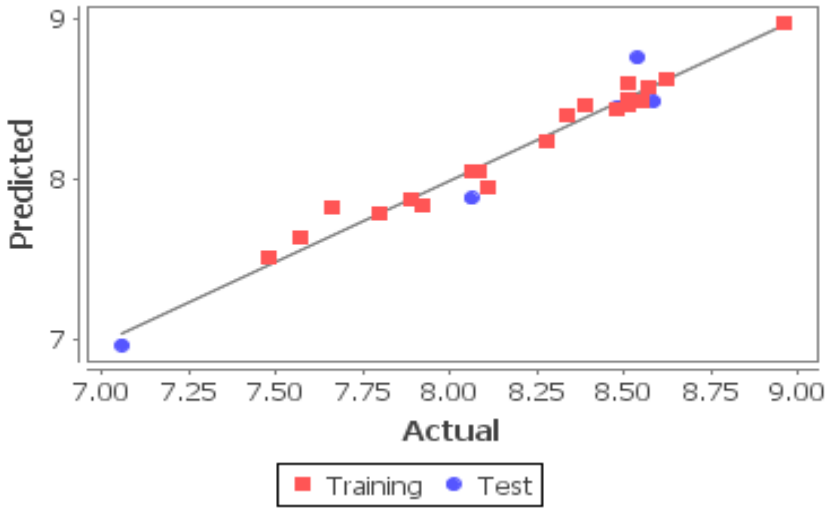


Figure 5: Fitness plot of developed Model

Conclusion:

In present research work a significant 3D QSAR model was developed by the aid of QSAR software which has ability to forecast the important requirement for good anti-diabetic activity. The model which has been developed by the aid of k nearest neighbor (k-NN) molecular field analysis annealed with variable selection methods like stepwise forward backward using partial least square regression analysis. The arithmetic data reveals that the model has good q2 and pred_r2 values of 0.9451 &0.9402 respectively. Three descriptors were involved in the model which plays a significant role in determining the DPP IV inhibitory activity. The three steric descriptors were in negative range representing the need of less bulky group in the substituted cyanopyrrolidine moiety. The k nearest neighbor method outline suggestan auxiliary perception of the relationship between structural features of substituted cyanopyrrolidine derivatives and their activities, which should be applicable to design newer potential DPP IV inhibitors.

Conflict of interest:

The authors confirm no conflict of interest

Acknowledgement:

The authors would like to thank Rungta College of Pharmaceutical Sciences and Research, Kohka, Shri Shankaracharya Technical Campus, Bhilai, Chhattisgarh and VLife Sciences Pvt.Ltd, Pune for providing necessary facility.

References

1. International Diabetes Federation. IDF Diabetes Atlas 2019 [Internet]. International Diabetes Federation. 2019. 1 p. Available from: http://www.idf.org/about-diabetes/facts-figures
2. Members [Internet]. [cited 2021 Aug 4]. Available from: https://idf.org/our-network/regions-members/south-east-asia/members/94-india.html
3. Scheen AJ, Charbonnel B. Effects of glucose-lowering agents on vascular outcomes in type 2 diabetes: A critical reappraisal. Vol. 40, Diabetes and Metabolism. Elsevier Masson; 2014. p. 176–85.
4. Deshmukh CD, Jain A. Diabetes Mellitus : A Review Diabetes Mellitus : A Review. Int J Pure App Biosci. 2015;3(3):224–30.
5. Robertson RP. Antagonist: diabetes and insulin resistance--philosophy, science, and the multiplier hypothesis. Vol. 125, Journal of Laboratory and Clinical Medicine. 1995. p. 560–4; discussion 565.
6. Chatterjee S, Khunti K, Davies MJ. Type 2 diabetes. Vol. 389, The Lancet. Elsevier; 2017. p. 2239–51.
7. Goyal R, Jialal I, Castano M. Diabetes Mellitus Type 2 (Nursing). StatPearls. 2021 May 20.
8. Lambeir AM, Durinx C, Scharpé S, De Meester I. Dipeptidyl-peptidase IV from bench to bedside: An update on structural properties, functions, and clinical aspects of the enzyme DPP IV. Vol. 40, Critical Reviews in Clinical Laboratory Sciences. Taylor & Francis; 2003. p. 209–94.
9. Röhrborn D, Eckel J, Sell H. Shedding of dipeptidyl peptidase 4 is mediated by metalloproteases and up-regulated by hypoxia in human adipocytes and smooth muscle cells. FEBS Lett. 2014 Nov 3;588(21):3870–

ISSN:0975-3583,0976-2833 VOL12,ISSUE06,2021

7.
10. Mulvihill EE, Drucker DJ. Pharmacology, physiology, and mechanisms of action of dipeptidyl peptidase-4 inhibitors [Internet]. Vol. 35, Endocrine Reviews. Endocr Rev; 2014. p. 992–1019.
11. Cordero OJ, Salgado FJ, Nogueira M. On the origin of serum CD26 and its altered concentration in cancer patients. Vol. 58, Cancer Immunology, Immunotherapy. Cancer Immunol Immunother; 2009. p. 1725–49.
12. Röhrborn D, Wronkowitz N, Eckel J. DPP4 in diabetes. Vol. 6, Frontiers in Immunology. Frontiers; 2015. p. 386.
13. Green BD, Flatt PR, Bailey CJ. Dipeptidyl peptidase IV (DPP IV) inhibitors: A newly emerging drug class for the treatment of type 2 diabetes. Vol. 3, Diabetes and Vascular Disease Research. SAGE PublicationsSage UK: London, England; 2006. p. 159–65.
14. Hulin B, Cabral S, Lopaze MG, Van Volkenburg MA, Andrews KM, Parker JC. New fluorinated pyrrolidine and azetidine amides as dipeptidyl peptidase IV inhibitors. Bioorganic Med Chem Lett. 2005 Nov 1;15(21):4770–3.
15. Fukushima H, Hiratate A, Takahashi M, Saito-Hori M, Munetomo E, Kitano K, et al. Synthesis and structure-activity relationships of potent 1-(2-substituted-aminoacetyl)-4-fluoro-2-cyanopyrrolidine dipeptidyl peptidase IV inhibitors. Chem Pharm Bull. 2008 Aug 1;56(8):1110–7.
16. Molecular Design Suite [Internet]. Vlife Sciences Technologies Pvt. Ltd, Pune, India. 2004 [cited 2021 Aug 4]. Available from: https://www.vlifesciences.com/
17. Vaishnav Y, Kashyap P, Kaur C. 3D QSAR Analysis on 2piperidinopiperidinethiadiazole Derivatives as Histamine H3 Receptor Inhibitors. Curr Nanomedicine. 2017;7(1):59–66.