International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 02 Issue: 03 | June-2015 www.irjet.net p-ISSN: 2395-0072

Prediction of Median Lethal Dose by QSAR method with their Applications

***Rajendra Kumar Sharma[1], Arun Sikarwar[2], Rajeeev Sharma[3], Pratibha Sharma[4]**

*[*1] Department of Applied Chemistry, SGSITS, Indore- 452003, M.P., India.*

[2] Department of Chemistry, Government Home Science College, Hoshangabad- 461001 (M.P.) India.

[3] Govt. K. P. College (Vigyan Bhawan), Dewas (M.P.) India.

[4] Department of Chemistry, ISLE, IPS Academy, Indore-452012, M.P., India.

*Corresponding author's email address: rksharma210375@gmail.com Mob. No. 91-9039541925

Abstracts: *In the present study 49 molecules belonging to cephalosporin antibiotic drugs were selected. Their 2D and 3D structures were prepared and subjected to Dragon software for the calculation of various indices viz. topological (e.g. wiener, randic, schultz, balban, detour etc.), geometrical [e.g. (T(N-N), T(N-O), T(N-S)], Quantum mechanical (e.g. 3D MoRSE, Kier Hall etc.), electronic (polarizability), simple total atom no. count indices [$(N_C)^R$, $(N_S)^R$] and combinations of various indices like addition, substraction also applied for the study like {$(S_s)^{wox}$ - $(S_s)^{wor}$}, {$(S_s)^{wox}$ + $(S_s)^{wor}$}, {$\sum_{u,am,v,en,p}(MoRSE)wox - \sum_{u,am,v,en,p}(MoRSE)wor$} also applied for the study. QSAR equation/ model derived in the form of a MLR equation for Median lethal dose (LD_{50}) which can be used to minimize toxicity of the particular cephalosporin drug. Further, the derived model for LD_{50} is also useful for modification of particular drug molecule for obtaining most favorable value of other drugs activities like IC_{50}, ED_{50}, MIC_{50} etc for consideration of toxicity change during modification.*

Key Words: *Median lethal dose (LD_{50}), MLR, 3D-MoRSE, Topological Indices and IC_{50}*

Introduction:

Median Lethal Dose (LD_{50})

LD_{50} value is the amount of a solid or liquid material that it takes to kill 50 % of test animals in one dose. It is standard measure of the toxicity [1] of a material that will kill half of the sample population of a specific test animal in a specified period through via inexposure gestion, skin contact, or injection. LD_{50} is measured in micrograms (or milligrams) of the material per kilogram of the test-animal's body weight, lower the amount, more toxic the material. Used in comparison of toxicities, LD_{50} values cannot be directly extrapolated from one specie to the other or to humans. Also called Median Lethal Dose [2], written as LD_{50} also. Value of LD_{50} depends upon the biological types of test animal, route of exposure, little on sex of animal and some times slightly on age of test animal.

Ratio of LD_{50} to ED_{50} is known as Therapatic Index (T.I.) or Therapatic Ratio.

$$\boxed{(TOXICITY)_{DRUG} \ \alpha \ LD_{50}} \quad \rightarrow (1)$$

$$\boxed{\text{Therapatic Index} = LD_{50} / ED_{50}} \quad \rightarrow (2)$$

For safe use of drug T.I. should be high as far as possible.

First time J.W. Trevan attempted in 1927 to find a way to estimate the relative poisoning potency of drugs and medicines used at that time. He developed the LD_{50} test to allow for comparisons between chemicals that poison the body in very different ways. Since Trevan's early work, other scientists have developed different approaches for more direct, faster methods of obtaining the LD_{50}. Recently some other following terms [3] are also applied for toxicity expression of a drug:

LD_{01} = Lethal dose for 1% of the animal test population

LD_{100} = Lethal dose for 100% of animal test population

LD_{LO} = The lowest dose causing lethality

TD_{LO} = The lowest dose causing a toxic effect

In nearly all cases, LD_{50} tests are performed using a pure form of the chemical. Mixtures are rarely studied. Researchers can do the test with any animal species but they use rats or mice most often. Other species include dogs, hamsters, cats, guinea-pigs, rabbits, and monkeys. The chemical may be given to the animals by mouth (Oral), by applying on the skin (dermal or subcutaneous), by injection at sites such as the blood veins, (I.V. - intravenous), muscles (I.M. - intramuscular) or into the abdominal cavity (I.P. – intra-peritoneal). LD_{50} value is expressed as the weight of chemical administered per kilogram body weight of the

animal and it states the test animal used and route of exposure or administration, e.g. LD_{50} (oral, rat) - 8 mg/kg, LD_{50} (skin, mice) - 8 g/kg. So, the example 'LD_{50} (oral, rat) 8 mg/kg' means that 8 milligrams of that chemical for every 1 kilogram body weight of the rat, when administered in one dose by mouth, causes the death of 50% of the test group.

Lethal effects from breathing a compound are also to be tested. In this case LD_{50} is expressed as LC_{50}. In this situation drug chemical (in gas or vapour state) is first mixed in a known concentration in a special air chamber where the test animals will be placed. This concentration is usually quoted as parts per million (ppm) or milligrams per cubic metre (mg/L). In these experiments, the concentration that kills 50% of the animals is called an LC_{50} (Lethal Concentration 50) rather than an LD_{50}. When an LC_{50} value is reported, it should also state the kind of test animal studied and the duration of the exposure, e.g., LC_{50} (rat) - 1000 ppm/ 4 hr or LC_{50} (mouse) - 8mg/m^3/ 2hr. On the basis of LD_{50} values, drugs or chemicals can be divided [4] in following groups:

Toxicity Rating	Commonly Used Term	Oral LD_{50} (single dose to rats)	Probable Lethal Dose for Man
1	Extremely Toxic	1 **mg/kg** or less	1 grain (a taste, a drop)
2	Highly Toxic	1-50 **mg/kg**	4 ml (1 tsp)
3	Moderately Toxic	50-500 mg/kg	30 ml (1 fl. oz.)
4	Slightly Toxic	500-5000 mg/kg	600 ml (1 pint)
5	Practically Non-toxic	5000-15,000 mg/kg	1 litre (or 1 quart)
6	Relatively Harmless	15,000 mg/kg or more	1 litre (or 1 quart)

In present study QSAR equation derived for LD_{50} by applying 49 molecules series of cephalosporin drugs. The descriptors were calculated by Dragon software and LD_{50} values searched from literature and sources of chem.-informatics. Multiple regression analysis and other statistical analysis was carried by Microsoft software. With the help of derived QSAR model, Cefalexin was modified to gain less IC_{50} by the consideration of derived QSAR model for LD_{50} so that toxicity should be controlled or minimized as far as possible.

Research Methodology:

Requirements to Develope the QSAR/ QSPR model:

A) Data set:
These provides experimental values of biological activity/ property measurements for a group of molecules.

B) Molecular structure/ property data:
These are in terms of descriptors or variables or any other predictors for the selected group of molecules

C) Statistical methods:
To establish relationship between above two data set i.e. (a) and (b).

QSAR/ QSPR mathematical model:

Activity = *f* (Physico-chemical properties/ Structural properties) ±Error

QSAR/ QSPR is generally a simple/ multiple linear equation:

$$\text{Biological Activity/ Physical Property} = \text{Const.} + (C_1 P_1) + (C_2 P_2) + (C_3 P_3) + \ldots \pm \text{Error} \quad \rightarrow (3)$$

Where, parameters P_1, P_2, P_3, ... P_n are independent variables and computed for each molecule in the series and, the coefficients C_1, C_2, C_3, ... C_n are calculated by fitting variations in the parameters and the biological activity.

As parameters P_1, P_2, P_3, ... P_n are independent variables, various descriptors were considered.

Molecular descriptors

- These are numerical values obtained by the quantificat ion of various structural and physicochemical characteristics of the molecule.
- Molecular descriptors quantify these attributes so as to determine the behavior of the molecule and the way the molecule interacts with a physiological system.

Types of Descriptors:

(a) **Spatial Descriptors**
Describe the molecules 'solvent-accessible' surface areas and their charges.

(b) **Electronic Descriptors** – Describe the electron orientation and charge.

(c) **Topological Descriptors** – Based from graph/structure concepts and geometric features such as shape, size, and branching.
(d) **Thermodynamic Descriptors** – Describes energy of molecules and their conversions
(e) **Quantum Mechanical Descriptors** – descriptors that are calculated using semi-empirical methods that are likely to be more accurate. e.g. 3D MoRSE descriptors.

In research work wiener, randic, balban, schultz, detour, harary, 3D-MoRSE, Keir and Hall valence connectivity indices, electronic, property etc. descriptors were appied. The descriptors who are significant for expression are shown later in Table (2) and (3), others less/non-significant are not shown. Except this followings are also considered:

- Above all descriptors some other indicator indices representing the no. of particular atom in particular group or structure like no. of S atoms in group –X etc.s viz. $(N_S)^X$ are also counted and applied in MLR.
- The combination of 3D Morse and sum of Keir hall topological distances are applied shown in table no. 4 other indices like $(EN_{orb.})^{*O=C<}$, Vander-waal volume, Donar sites and α_{XX} calculated by Dragon considered for MLR analysis. Above indices are defined in the text where they are needed.
- For determining microscopic contribution of group –X the common indices like W, H, X^1, W', T(N-N), T(N-S), T(N-O) also calculated for selected set of molecules by removing –R group.

Concepts of Drug designing

(i) Lipinski's rule of five

This is also known as the Pfizer's rule of five or simply the Rule of five (RO5). This is a rule of thumb to evaluate druglikeness or determine if a chemical compound with a certain pharmacological or biological activity has properties that would make it a likely orally active drug among humans. This rule was formulated by Christopher A. Lipinski in the year 1997, based on the observation that most medication drugs are relatively small and lipophilic molecules [5, 6].

Lipinski's rule states that, in general, an orally active drug has no more than one violation of the following criteria:

- Not more than 5 hydrogen bond donors (nitrogen or oxygen atoms with one or more hydrogen atoms)
- Not more than 10 hydrogen bond acceptors (nitrogen or oxygen atoms)
- A molecular mass less than 500 daltons
- An octanol-water partition coefficient log P not greater than 5

In above all numbers are multiples of five, which is the origin of the rule's name. However, there are many exceptions to Lipinski's Rule. This rule describes molecular properties important for a drug's pharmacokinetics in the human body, including their Absorption, Distribution, Metabolism, and Excretion (ADME). The rule is important to keep in mind during drug discovery when a pharmacologically active lead structure is optimized step-wise to increase the activity and selectivity of the compound as well as to insure drug-like physicochemical properties are maintained as described by Lipinski's rule.

(ii) Extensions of Lipinski's rule of five

In an attempt to improve the predictions of druglikeness, the Lipinski's rule of five have undergoned many extensions, some of them are the following [7]:

- Partition coefficient log P should be in range -0.4 to $+5.6$ range
- Molar refractivity should be ranged from 40 to 130
- Molecular weight from 180 to 500
- Number of atoms present in drug molecule should be ranged from 20 to 70 (includes H-bond donors [e.g. no. of OH's and NH's] and H-bond acceptors [e.g. no. of Nitrogen and Oxygen atoms's])
- Polar surface area should not be greater than 140 $(Å)^2$

Also the 500 molecular weight cutoff has been questioned. Polar surface area and the number of rotatable bonds has been found to better discriminate between compounds that are orally active and those that are not for a large data set of compounds in the rat. In particular, compounds which meet only the two following criteria, are predicted to have good oral bioavailability [160]:

- 10 or fewer rotatable bonds and
- polar surface area equal to or less than 140 $Å^2$

SAR of Cephalosporin:

General structure of cephalosporin can be shown by the following diagram. However, the different side groups imparts effects over the drug in different ways. The figure (1) shows structure activity relationship of the cephalosporin class belonging drugs.

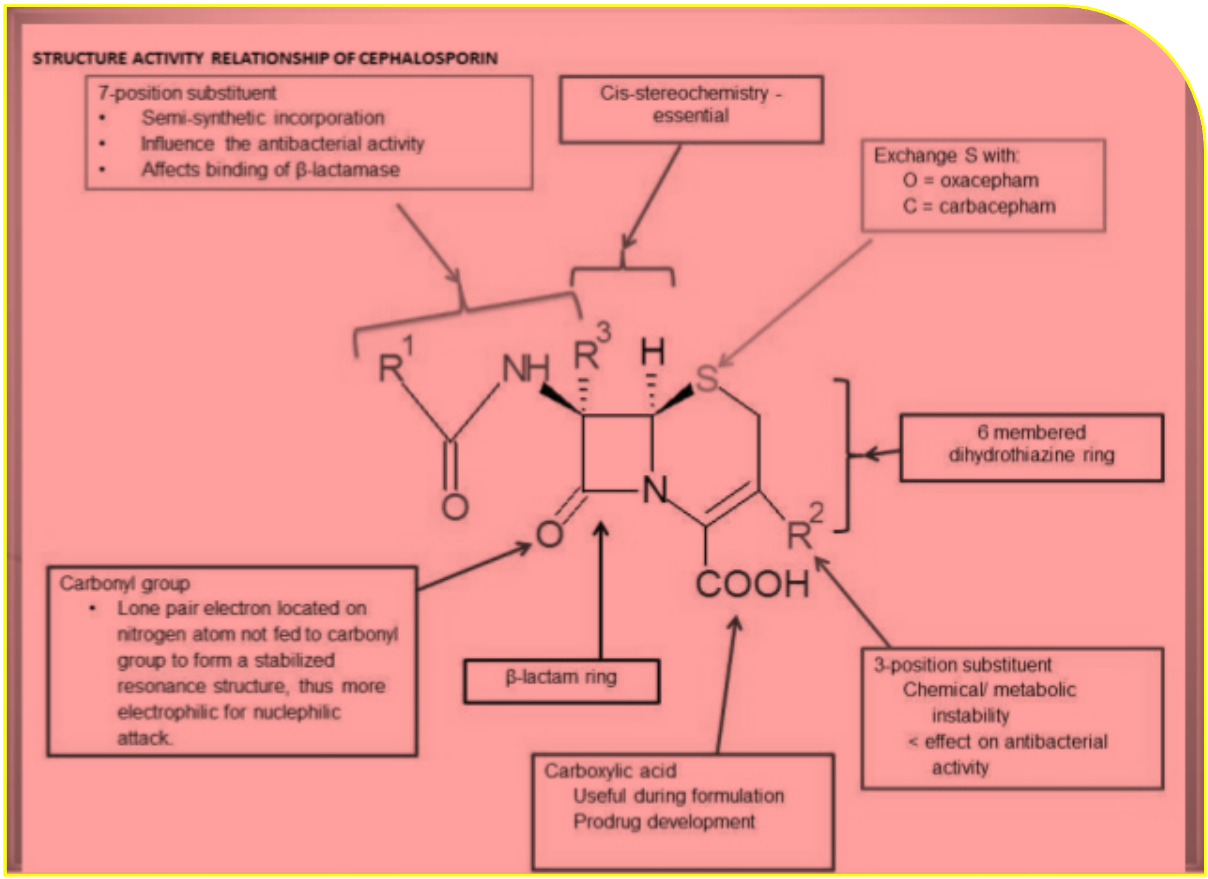
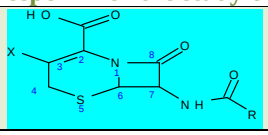
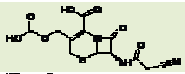
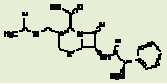
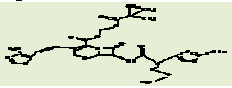
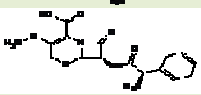
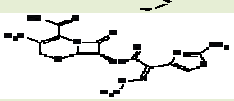
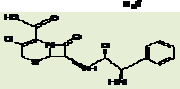
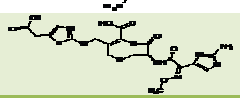


FIGURE (1): SAR of Cephalosporin molecules

Table (1) : Molecular set of cephalosporin for the study of LD_{50}

General core structure of molecular set 

Training Set	Name	Structure of molecule	Training Set	Name	Structure of molecule
Cep 01	Cefacetrile		Cep 28	Cefdinir	
Cep 03	Cefalexin		Cep 30	Ceftobiprole	
Cep 04	Cefaloglycin		Cep 31	Cefditoren	
Cep 05	Cefroxadine		Cep 32	Cefatamet	
Cep 06	Cefaclore		Cep 34	Cefpodizime	

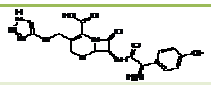
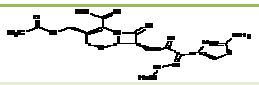
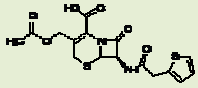
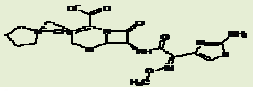
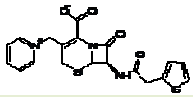
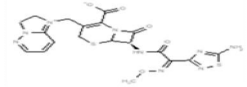
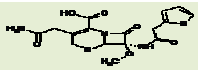
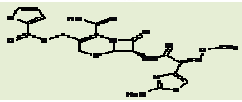
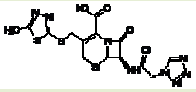
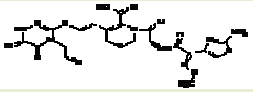
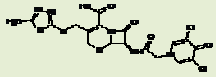
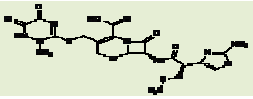
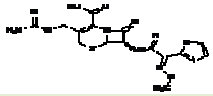
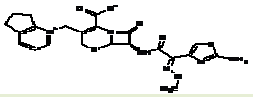
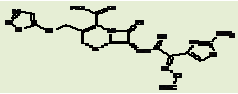
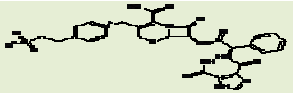
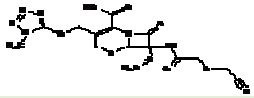
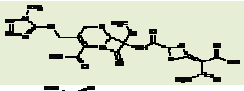
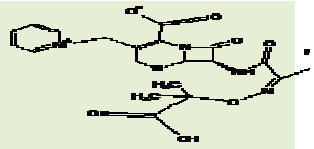
Cep 10	Cefatrizine		Cep 35	Cefotaxime	
Cep 11	Cefalothin		Cep 38	Cefepime	
Cep 13	Cefaloridine		Cep 39	Cefozopran	
Cep 14	Cefoxitin		Cep 40	Ceftiofur	
Cep 15	Cefazoline		Cep 41	Ceftiolene	
Cep 18	Cefazedone		Cep 43	Ceftriaxone	
Cep 20	Cefuroxime		Cep 44	Cefpirome	
Cep 21	Cefuzonam		Cep 46	Cefpimizole	
Cep 22	Cefmetazole		Cep 48	Cefoperazone	
Cep 23	Cefotatam		Cep 49	Ceftazidime	
Cep 25	Cefminoxime				

Table (2) : LD_{50} values and descripters generated by combination of Keir-Hall electro topological state and 3D MoRSE indices

Molecular Set	LD_{50} (in mg/kg/day) [For mice through oral Dose]	LD_{50} (in mg/kg/day) [For mice through I.V. Dose]	References	$(S_s)^{wox} - (S_s)^{wor}$	$(S_s)^{wox}$	$(S_s)^{wox} + (S_s)^{wor}$	$(Mor_{01en})^{wox}$	$\sum_{u,am,v,en,p}(MoRSE)wox$	$\sum_{u,am,v,en,p}(MoRSE)wox - \sum_{u,am,v,en,p}(MoRSE)wor$
C_1	1900	3700	[8]	-5.66	56.17	118.0	548.52	1752.02	-1256.0
C_2	3400	3200	[8,9]	20.66	68.83	117.0	791.69	2550.27	-1809.3
C_3	3100	3000	[10]	16.67	63.17	109.67	739.94	2435.05	-1732.1
C_4	4100	3000	[8]	1.340	63.17	125.0	739.94	2435.05	-1732.1
C_5	13500	13500	[8, 11]	10.50	62.17	113.84	739.94	2435.05	-1732.1
C_6	3000	2900	-"-	12.89	63.17	113.45	739.94	2435.05	-1732.1

C_7	8500	3000	[12]	15.0	62.17	111.34	739.94	2439.73	-1736.7
C_8	17000	4100	[13]	-17.91	65.17	148.25	710.27	2344.74	-1678.7
C_9	2000	2000	[14]	16.66	68.83	121.0	791.69	2586.27	-1845.3
C_{10}	19000	4100	[12, 13]	7.16	68.83	130.50	791.69	2586.27	-1845.3
C_{11}	2000	4800	[16,17,18,19]	-4.66	57.17	119.0	562.99	1949.90	-1421.9
C_{12}	9300	1400	[20, 12, 21]	-14.83	57.17	129.17	562.99	1949.90	-1421.9
C_{13}	16000	4100	[12, 22]	-2.30	58.17	118.64	531.78	1868.73	-1373.0
C_{14}	10000	4970	[12, 22]	-3.17	62.58	128.33	713.12	2410.38	-1744.4
C_{15}	9000	4500	-"-	-2.0	60.67	123.34	519.15	1711.97	-1247.0
C_{16}	15000	4700	[12]	-0.33	60.67	121.67	519.15	1711.97	-1247.0
C_{17}	19000	4100	[20]	-1.16	60.67	122.50	570.23	2000.71	-1472.7
C_{18}	6000	3900	[12, 23]	11.22	73.89	136.56	704.70	2516.19	-1886.2
C_{19}	10000	2900	-"-	10.58	74.75	138.92	513.89	1723.10	-1288.1
C_{20}	10000	10400	[12]	3.67	67.50	131.33	730.66	2395.06	-1729.1
C_{21}	8000	4500	[24, 25]	9.50	70.50	131.50	772.13	2611.33	-1908.3
C_{22}	3700	4200	-	-4.66	64.92	134.50	647.24	2166.76	-1571.8
C_{23}	4990	4990	[12]	17.67	87.25	156.83	925.65	3194.12	-2374.1
C_{24}	10000	100000	[25, 26,27]	28.34	98.42	168.50	1361.21	4255.05	-3030.1
C_{25}	5000	5000	[28]	10.34	79.42	149.0	822.93	2712.72	-1971.7
C_{26}	2000	2000	[29]	-2.33	67.50	137.33	867.22	2915.36	-2095.4
C_{27}	5000	5000	-	17.83	71.0	124.17	662.81	2274.93	-1679.9
C_{28}	5600	2000	[12, 30]	19.83	71.0	122.17	662.81	2274.93	-1679.9
C_{29}	4000	2500	[31]	7.50	71.0	134.50	662.81	2274.93	-1679.9
C_{30}	9600	2200	-	0.0	71.83	143.66	700.69	2291.44	-1661.4
C_{31}	5100	5100	[12]	8.17	70.0	131.83	889.76	2972.03	-2152.0
C_{32}	2000	1500	-"-	22.38	70.55	118.72	772.13	2611.31	-1908.3
C_{33}	17540	7830	[12, 40, 41]	6.38	70.55	134.72	772.13	2611.31	-1908.3
C_{34}	15500	6570	[32]	-5.28	70.55	146.38	772.13	2611.31	-1908.3
C_{35}	20000	7000	[20]	8.72	70.55	132.38	772.13	2611.31	-1908.3
C_{36}	19000	300	[33]	17.38	70.55	123.72	772.13	2611.30	-1908.3
C_{37}	6000	5090	[12, 15]	8.22	70.55	132.88	772.13	2611.31	-1908.3
C_{38}	1500	1500	[12, 39]	13.33	71.50	129.67	735.50	2517.60	-1851.6
C_{39}	15000	3820	[12]	5.17	71.50	137.83	735.50	2517.60	-1851.6
C_{40}	7760	3840	[20]	1.17	70.50	139.83	772.13	2611.31	-1908.3
C_{41}	10000	3800	-"-	-18.50	70.50	159.50	772.13	2611.31	-1908.3
C_{42}	6000	6000	[36]	24.05	70.55	117.05	772.13	2611.31	-1908.3
C_{43}	10000	2800	[37]	-7.45	70.55	148.55	772.13	2611.31	-1908.3
C_{44}	16200	2400	-"-	6.05	70.55	135.05	772.13	2611.31	-1908.3
C_{45}	4420	4420	[34]	33.50	84.67	135.84	922.12	3104.48	-2284.5
C_{46}	15000	2700	[12]	11.58	95.83	180.08	1333.63	4472.80	-3197.8
C_{47}	10000	6000	[35]	33.67	80.17	126.67	899.16	3043.24	-2223.2

C_{48}	15000	15000	[20]	22.33	86.50	150.67	1311.07	4266.44	-3041.4
C_{49}	17000	4200	[38]	28.75	89.42	150.09	1141.25	3794.64	-2759.6

Computer softwares used in present research work:

- All calculations were run on Acer personal computer (laptop, model Acer ASPIRE 4530) with a Pentium IV configuration and windows XP as operating system and AMD Athlon-X_2 processer.
- All of indices were calculated by using by software. DRAGON version 5.5- 2007
- DRAGON software acceptable molecular structures were prepared by Chem sketch version 12.0 as MDL file formate.
- The molecular structures of data set were sketched using Chem sketch as MDL files.
- This can calculate chemical properties also. This can draw Chemical structure and graphical images.
- For collecting various properties like half life ($t_{1/2}$), log P values, MIC values, LD_{50} etc., the journals as well as different sources of Chem informatics were applied.
- Some of the properties and activities searched from the merck index.

For statistical analysis like correlation, regression (MLR) and validation etc. Microsoft Office Excel, software applied during research work.

MLR can be done in two ways:

(a) By forward selection method

- Initially all variable indices are taken in regression analysis.
- In this each variable index is added in various steps one by one.
- In each step variable index is selected on the basis of lowest P-value.
- Regression ansalysis is carried out till the step at which all variable appears significant ($P < 0.05$) in any one or more regression equation.

(b) By backward selection method:

- Initially all variable indices are taken in regression analysis.
- Variable indices are removed in step by step.
- Removal is done on the basis of P-value.
- In each step the variable index, whose removal results in highest significant regression equation, is removed.

Linear regression and Multiple regression is to be calculated by MS excel 2007

Half maximal Inhibitory Concentration (IC_{50}):

Definition- "It is a measure of the effectiveness of a compound (drug) in inhibiting biological or biochemical function by half."

- IC_{50} of a drug can be determined by constructing a 'dose response curve'
- IC_{50} values can be calculated for a given antagonist by determining the concentration needed to inhibit half of the maximum biological response of the agonist.
- IC_{50} can be calculated for a given antagonist by determining the conc. needed to inhibit half of the maximum biological response of the antagonist.

The various descriptors were calculated through 'Dragon' software, steps of operation in this software [42] are given in the following figure:

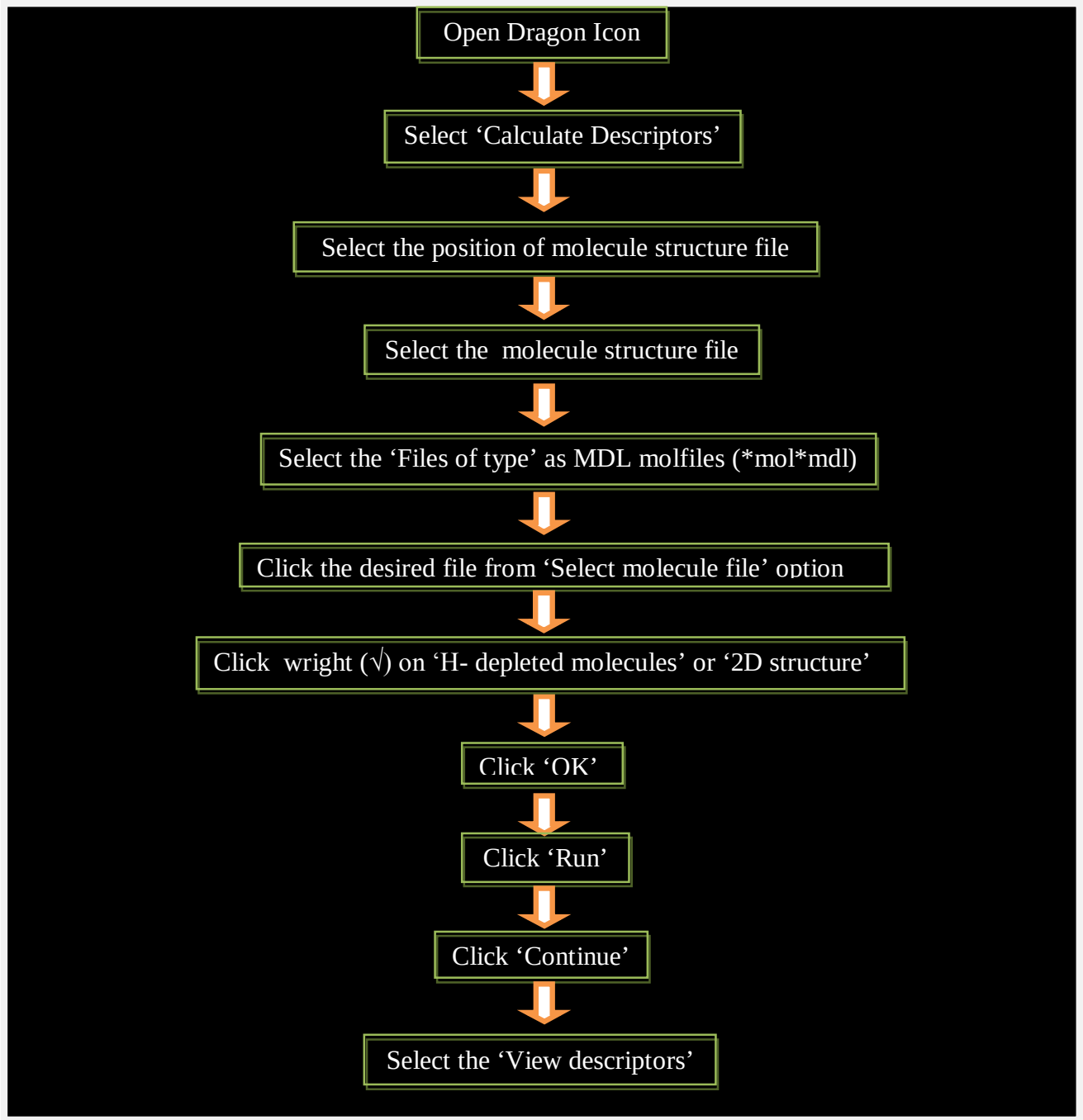


FIGURE (2): Sketch showing the major steps for calculation of descriptors through Dragon software

Table (3) : Values of various topological, topological distance, atom count no. and electronic descriptors.

Molecular Set	$(N_S)^R$	$(N_C)^R$	$(N_S)^X$	α_{xx}	W	T(N-N)	T(N-S)	X^1	J
C_1	0	2	0	29.79	1073	14	10	10.362	2.02
C_2	0	7	0	36.22	1570	12	9	11.807	1.57
C_3	0	7	0	34.24	1242	12	9	11.003	1.56
C_4	0	7	0	59.81	2172	12	9	13.307	1.58
C_5	0	7	0	53.15	1548	12	9	11.952	1.59
C_6	0	7	0	50.13	1383	12	9	11.414	1.58
C_7	0	7	0	50.28	1383	12	9	11.414	1.58
C_8	0	7	2	74.23	3766	67	69	16.091	1.33
C_9	0	7	0	60.01	1954	12	9	12.845	1.58
C_{10}	0	7	1	43.14	2945	99	56	14.863	1.31
C_{11}	1	5	0	37.42	1790	3	14	12.38	1.56
C_{12}	1	5	0	45.68	2975	40	54	14.846	1.31
C_{13}	1	5	0	40.09	2180	14	29	13.542	1.33
C_{14}	1	5	0	42.16	1837	16	31	12.797	1.70
C_{15}	0	2	2	70.44	2468	191	172	13.936	1.30
C_{16}	0	2	3	65.84	2223	191	172	13.542	1.31
C_{17}	1	6	0	37.92	2008	18	21	12.88	1.54
C_{18}	3	6	2	49.45	3579	69	73	15.651	1.29
C_{19}	1	1	1	37.63	2470	67	105	13.581	1.54
C_{20}	0	6	0	57.71	2349	39	15	13.845	1.62
C_{21}	1	5	2	64.17	3133	166	192	15.401	1.35
C_{22}	1	3	1	48.45	2591	144	130	14.351	1.64
C_{23}	2	5	1	590.65	4350	150	196	17.012	1.37
C_{24}	0	9	1	66.72	6176	294	108	19.311	1.38
C_{25}	1	4	1	45.73	3492	144	130	15.618	1.59
C_{26}	1	8	0	61.64	2584	72	52	14.239	1.62
C_{27}	1	4	0	59.96	1903	46	42	12.883	1.62

C_{28}	1	4	0	54.02	1701	46	42	12.383	1.62
C_{29}	1	4	2	76.72	3225	236	222	15.401	1.31
C_{30}	0	4	0	49.49	4333	222	48	17.295	1.19
C_{31}	1	4	1	89.72	6137	106	119	19.417	1.41
C_{32}	1	4	0	52.28	1693	46	42	12.345	1.63
C_{33}	1	4	1	74.58	3413	276	177	15.812	1.35
C_{34}	1	4	2	88.36	4530	101	162	17.189	1.31
C_{35}	1	4	0	62.52	2584	46	42	14.239	1.62
C_{36}	1	4	0	60.02	2096	46	42	13.383	1.63
C_{37}	1	4	0	65.53	3081	246	114	15.295	1.37
C_{38}	1	4	0	70.34	3029	91	58	15.237	1.40
C_{39}	1	4	0	50.58	3932	264	107	16.884	1.21
C_{40}	1	4	1	55.39	3753	46	87	16.312	1.34
C_{41}	1	4	1	84.59	5623	234	163	18.671	1.32
C_{42}	1	4	0	48.65	1531	46	42	11.935	1.61
C_{43}	1	4	1	80.56	4386	219	152	17.116	1.35
C_{44}	1	5	0	80.26	3906	91	58	16.884	1.22
C_{45}	1	5	0	59.11	2560	46	42	14.239	1.63
C_{46}	3	12	1	70.77	8803	99	116	21.835	1.17
C_{47}	1	6	0	43.67	1897	29	32	12.791	1.62
C_{48}	0	7	1	73.29	7383	294	108	21.01	1.23
C_{49}	1	8	0	57.98	4573	91	58	17.491	1.40

* Where, meaning of symbols/ abbreviations used in Table (2) & (3) are as follows:
W= Weiner index; $\chi 1$ = Randic index; J = Balban index; H = Harary index; α_{XX} = Molecular polarizability in x- direction; T(N-N) = Topological distances between N and N; T(N-S) = Topological distances between N and S; T(N-O) = Topological distances between N and O.

$(S_s)^{wox}$, $(S_s)^{wor}$ = Keir-Hall electro topological state for selected set of molecules with out -x group and without –R group respectively.

Sum of all 3D MoRSE descripters unweighted and weighted atomic by mass, vander waal volume, electronegativity and polarizability without – x group = $\sum_{u,am,v,en,p}$ (MoRSE)wox

Sum of all 3D MoRSE descripters unweighted and weighted atomic by mass, vander waal volume, electronegativity, polarizibility for molecule without – R group = $\sum_{u,am,v,en,p}$ (MoRSE)wor

$(Mor_{1en})^{wox}$ = 3D MoRSE descriptors for selected set of molecules with out -x group weighted by electronegativity.
(Superscript 'wox' and 'wor' shows values for structures with out –X group and with -R group respectively.)
$(N_S)^R$ = No. of S atoms in group –R, $(N_S)^X$ = No. of S atoms in group X
$(N_C)^R$ = No. of C atoms in group –R

QSAR study of LD_{50} for Cephalosporin derivatives

{A} Data generation:

For LD_{50} study 49 molecules were selected. Molecular set no. of these selected molecules are - C_1 to C_{49}. The values of LD_{50} were collected from literature as well as sources of cheminformatics like ACD i- lab and others [42] and given in Table (1) The indices values of corresponding selected set of molecules as given in Table (2), (3), (4) were used for further data processing i.e. stepwise MLR analysis etc. However, for this study properties/ activities given in Table (2) and (3), (4) also gave significant contribution in prediction of LD_{50} when taken in consideration during MLR analysis.

{B} Data processing and outcome with their statistical validation:

Similar to other methods first distribution correlation [given in Table (4)] of various indices with LD_{50} were observed in which this was observed that no single index is sufficient enough to predict the selected property LD_{50} significantly and satisfactorily. So, stepwise MLR analysis were needed and hence carried-out. Summary of stepwise MLR analysis is given in Table (5).

Table (4): Correlation of LD_{50} with selected indices

Indices	Corre-Lation	Indices	Corre-Lation
W	0.305	T(N-O)	0.448
$\chi 1$	0.281	SMTI	0.298
J	- 0.072	$(S_s)^{wox}$	0.452
W'	0.288	$(Mor_{1u})^{wox}$	0.469
H	0.297	$(Mor_{1am})^{wox}$	0.357
T(N-N)	0.350	$(Mor_{1v})^{wox}$	0.380
T(N-S)	0.086	$(Mor_{1en})^{wox}$	0.499
$(Mor_{1p})_{wox}$	0.360	$(Mor_{1p})^{wor}$	0.028
$(S_s)^{wor}$	0.106	$\sum_{u,am,v,en,p}(MoRSE)wox$	0.447
$(Mor_{1u})^{wor}$	0.023	$\{\sum_{u,am,v,en,p}(MoRSE)wor - \sum_{u,am,v,en,p}(MoRSE)wox\}$	- 0.434
$(Mor_{1am})^{wor}$	0.111		
$(Mor_{1v})^{wor}$	0.034	$(N_C)^R$	0.291
$(Mor_{1en})^{wor}$	0.038	$(N_S)^R$	-0.200

Table (5) : Summary of stepwise MLR regression for LD_{50}

S. No.	Step no.	Developed MLR equation	R^2	PRESS	SE
1	Step-(1)	LD_{50}= -22325.83 - 37.15(±9.38) $(Mor_{1en})^{wox}$	0.249	4.2X 10^9	12188
2	Step-(2)	LD_{50}= - 7822.25 + 382.80(±75.65) $(Mor_{1en})^{wox}$ - 108.82 (±23.69) $\sum_{u,am,v,en,p}(MoRSE)wox$	0.486	4.7 X 10^9	10200
3	Step-(3)	LD_{50}= -10352.31 + 652.44(±106.18) $(Mor_{1en})^{wox}$ - 383.54 (±85.37) $\sum_{u,am,v,en,p}(MoRSE)wox$ – 268.82(±80.86) $\{\sum_{u,am,v,en,p}(MoRSE)wox\}$	0.587	3.8 X 10^9	9240
4	Step-(4)	LD_{50}= -11552.98 + 835.29(±122.26) $(Mor_{1en})^{wox}$ 598.37(±115.31) $\sum_{u,am,v,en,p}(MoRSE)wox$ – 81.23(±111.71) $\{\sum_{u,am,v,en,p}(MoRSE)wox$ - $\sum_{u,am,v,en,p}(MoRSE)wor\}$ + 3239.44(±1246.63) $(N_C)^R$	0.588	3.8 X 10^9	9332.83
5	Step-(5)	LD_{50}= -15974.84 – 901.73(±126.94 $(Mor_{1en})^{wox}$ - 664.82 (±120.50) $\sum_{u,am,v,en,p}(MoRSE)wox$ – 547.66(±11 7.18) $\{\sum_{u,am,v,en,p}(MoRSE)wox$ - $\sum_{u,am,v,en,p}(MoRSE)wor\}$ +3602.40 (±1244.92) $(N_C)^R$ – 203.80(±126.26) $(S_s)^{wox}$	0.667	3.14 X 10^9	8546.36
6	Step-(6)	LD_{50}= 55167.23 – 1519.56(±191.53 $(Mor_{1en})^{wox}$ -1063.33(±158.15) $\sum_{u,am,v,en,p}(MoRSE)wox$ – 860.46(±154.10) $\{\sum_{u,am,v,en,p}(MoRSE)wox$ - $\sum_{u,am,v,en,p}(MoRSE)wor\}$ + 4206.24(±1 114.92) $(N_C)^R$ – 692.55(±467.32)	0.866	2.3 X 10^9	7433.49

		$(S_s)^{wox}$ + 1236 5.31(±35 48.06) $(N_S)^R$			
7	Step - (7)	LD_{50}= 56942.16 - 1503.32 (±185.44) $(Mor_{1en})^{wox}$ -1062. 35(±152.96) $\sum_{u,am,v,en,p}$(MoRSE)wox – 862.25(±149.60) {$\sum_{u,am,v,en,p}$(MoRSE)wox - $\sum_{u,am,v,en,p}$(MoRSE)wor} + 507 0.88(±1164.08) $(N_C)^R$ – 172 9.42(±452.40) $(S_s)^{wox}$ + 1252 9.93 (±3432.91) $(N_S)^R$ + 29.1 1(±14.76) T(N-N) ± 7190	0.772	2.12 X10^9	7190.18
8	Step - (8) (Final)	No variable can be added satisfactory.	-	-	-

Results and discussions:

Outcome of stepwise MLR analysis is following regression equation model:

$$\mathbf{LD_{50}} = 56942.16 - 1503.32(\pm185.44)\,(Mor_{1en})^{wox} - 1062.35\,(\pm152.97)\,\sum_{u,am,v,en,p}\mathbf{(MoRSE)wox} - 862.25\,(\pm149.06\,\{\sum_{u,am,v,en,p}\mathbf{(MoRSE)wox} - \sum_{u,am,v,en,p}\mathbf{(MoRSE)wor}\} + 5070.88(\pm1164.08)\,\mathbf{(N_C)^R} - 1729.42(\pm452.40)\,\mathbf{(S_s)^{wox}} + 12529.93(\pm3432.91)\,\mathbf{(N_S)^R} + 29.1140(\pm14.76)\,T(N\text{-}N) \pm 7190 \quad \rightarrow (3)$$

$$\mathbf{(LD_{50})_{pred.}} = 0.772\,\mathbf{(LD_{50})_{obs.}} + 1442 \quad \rightarrow (4)$$

Statistics of the developed regression equation is as follows

N	R^2-Adjusted	R^2	Pearson's r	F-ratio	Overall significance-F	SE	PRESS
49	0.733	0.772	0.879	19.868	2.55 X 10-^{11}	71.90	2.12 X 10^{-9}

→ (5)

All the stastistical data are validated by the stastical parameter that for all the regression results for QSAR given in Eq. (5) and satisfy the following validation conditions:

- n / p > = 4, where, n = no. of molecules taken for modeling. p = no. of descriptors used in model for prediction.
- $R^2 > 0.6$, where R^2 = correlation coefficient.
- $q^2 > 0.6$, where q^2 = cross validated correlation coefficient.
- $R^2 - q^2 < 0.3$
- Standard error is least in most of the regression models.

Table (6) : Observed and predicted LD_{50}

Molecular Set	Observed LD_{50} (gm/ kg)	Predicted LD_{50} (gm/ kg)	Molecular Set	Observed LD_{50} (gm/ kg)	Predicted LD_{50} (gm/ kg)
C_1	3.7	16.7	C_{26}	2.0	8.7
C_2	3.2	14.6	C_{27}	5.0	3.5
C_3	3.0	2.4	C_{28}	2.0	3.
C_4	3.0	2.4	C_{29}	2.5	1.9
C_5	13.5	4.2	C_{30}	2.2	11.1
C_6	2.9	2.5	C_{31}	5.1	7.6
C_7	3.0	1.5	C_{32}	1.5	1.2
C_8	4.1	5.9	C_{33}	7.8	7.8
C_9	2.0	7.4	C_{34}	6.6	2.7
C_{10}	4.1	10.0	C_{35}	7.0	1.2
C_{11}	4.8	3.0	C_{36}	3.0	1.1
C_{12}	1.4	1.9	C_{37}	5.1	6.9
C_{13}	4.1	7.3	C_{38}	1.5	3.6
C_{14}	4.9	2.5	C_{39}	3.8	1.4
C_{15}	4.5	4.6	C_{40}	3.8	1.2
C_{16}	4.7	4.6	C_{41}	3.8	6.7
C_{17}	4.1	2.8	C_{42}	6.0	1.2
C_{18}	3.9	11.8	C_{43}	2.8	6.2
C_{19}	2.9	1.14	C_{44}	2.4	7.5
C_{20}	10.4	16.7	C_{45}	4.4	7.7
C_{21}	4.5	9.8	C_{46}	2.7	3.0
C_{22}	4.2	2.9	C_{47}	6.0	2.2
C_{23}	4.9	6.2	C_{48}	15.0	12.4
C_{24}	100.0	79.6	C_{49}	4.2	21.9
C_{25}	5.0	11.9			

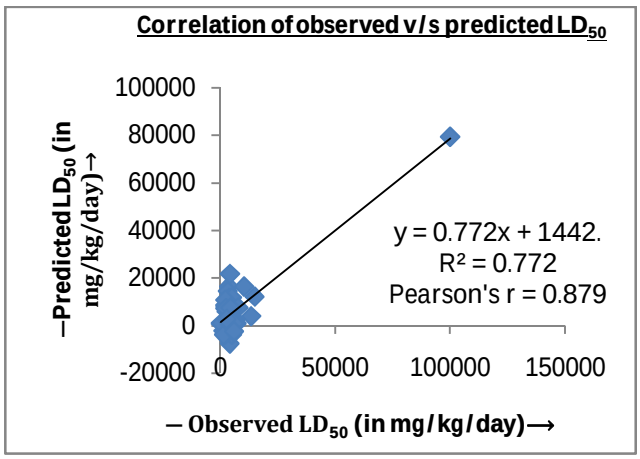


CHART (1) : Predicted and observed values of LD_{50}

Molecular Modeling based on LD_{50}:

Multilinear regression study of LD_{50} correspond to regression equation (3) was derived significantly. Regression equation was derived with the help of six variable parameter indices. Most dominant indiex is 'total number of sulphur atoms in group –X', since it's coefficient is larger value than others. Total no. of carbon atoms is the second most effective index for prediction. The value of predicted against observed for LD_{50} shown in Table (6) and correlation graph is shown in Figure (3) under LD_{50} is very important during modeling of drug molecule. During the modeling this is taken in consideration that there should be no decrease in LD_{50}. For this purpose the developed regression eq. is helpful. The developed regression eq. (5)
is applied during the modification of various parental molecules to corresponding proposed molecules with improved particular activity/ property viz., P_5, molecules. The corresponding data for LD_{50} are shown in Table (2), (3), (4). The above developed model are example for LD_{50} importance. Developed regression equation gives satisfactory value of correlation between predicted and observed, predicted residual squares sum, Overall significance. These are shown in Table (6)

Modelling of minimum IC_{50} with in consideration of maximum LD_{50} :

By using the same method R. Sharma derived the model [43, 44] for expression of IC_{50} in terms of pIC_{50} as follows:

$$\mathbf{pIC_{50}} = 5.41 - 2.19(\pm 0.42)\,\mathbf{(N_S)^X} + 0.02(\pm 0.01)\,\mathbf{T(N\text{-}S)} - 0.04(\pm 0.02)\,\mathbf{\{(S_s)^{wox} - (S_s)^{wor}\}} - 0.01(\pm 0.001)\,\boldsymbol{\alpha_{XX}} \quad \rightarrow (6)$$

Since, $pIC_{50} = -\log IC_{50}$. For a good cephalosporin drug IC_{50} should be low i.e. pIC_{50} should be high. In prediction of pIC_{50}, the + ive factor which increases it's value is T(N-S). So, this can be suggested to introduce more –N and –S atom in a molecule but this is limited by two important situations

- In reg. eq. (5) $(N_S)^X$ term decreases pIC_{50}, so insertion of –S atom in group- X should be ignored and insertion preferred in –R group which increases LD_{50} also.
- $\{(S_s)^{wox} - (S_s)^{wor}\}$ appears in –ive term, so high pIC_{50} this term should be –ive so which is possible only when $(S_s)^{wor} > (S_s)^{wox}$. For this –X and –R group should be comparable size with more cyclic structure in –X and less cyclic structurs in -R group should be preffered.

On the basis of above facts following molecule can be proposed by modification of Cefalexin:

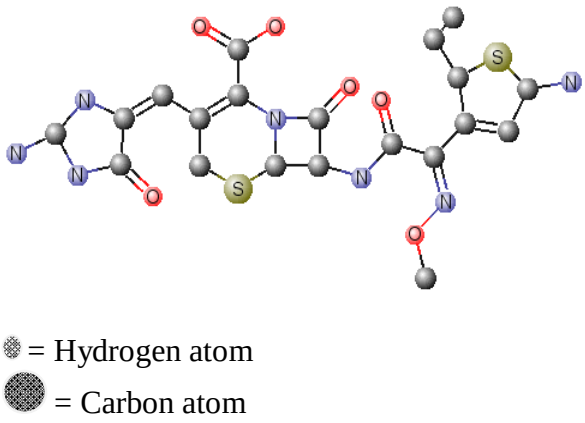


FIGURE (3) : Modified molecule P_5 through eq. (3) & (6)

Features of proposed molecule, P_5: This molecule is supposed to more pIC_{50} i.e less IC_{50} than parental. Further this molecule is less toxic than parental as shown by data in Table (5.12) and (5.13)
IUPAC Name = 7-[(2Z)-2-(5-amino-2-ethyl-2,5-dihydrothiophen-3-yl)-2-ethoxyimino)acetamido]- -{[(4E)-2-amino-5-oxoimidazolidin-4-ylidene]methyl}-8-oxo-5-thia-1-azabicyclo[4.2.0]oct-2-ene-2-carboxylic acid.
Particulars of this proposed molecule, P_5 are as follows:
Molecular formula: $C_{20}H_{25}N_7O_6S_2$,
Molecular weight: 523.58,
Molar refractivity = 125.42 cm^3, {which enable this compound to satisfy the one condition of 'Lipinski's rule of five's extension' [discussed in earlier section] as showing suitability for application through oral dose.} Molar volume = 283.4 cm^3,
Parachore value = 866.3 cm^3,
Surface tension = 87.2 dyne /cm.
Polarizability 49.72,
Density = 1.84 g/ cm^3

Table (6) : Comparative data of parental and newer proposed molecules with improved pIC_{50}

S. No.	Name of molecule	(+) ive terms	(-) ive terms			Inference
		T(N-S)	$(N_S)^X$	$\{(S_s)^{wox} - (S_s)^{wor}\}$	α_{XX}	
1	Cefalexin (C_3) (Parantal molecule)	9	0	16.67	34.24	$(pIC_{50})_2 > (pIC_{50})_1$
2	Proposed molecule, P_5	95	0	16.02	33.20	

Table (7) : Comparative data of parental and newer proposed molecule, P_5 for toxicity (LD_{50})

S. No.	Name of molecule	(+) ive terms		(-) ive terms					Inference
		$(N_C)^R$	$(N_S)^R$	$\sum_{u,am,v,en,p}(MORSE)^{wox}$	$\{\sum_{u,am,v,en,p}(MORSE)^{wox} - \sum_{u,am,v,en,p}(MORSE)^{wor}\}$	$(Mor_{1en})^{wox}$	T(N-N)	$(S_s)^{wox}$	
1	Cefalexin (C_3) (Parantal molecule)	7	0	2435.05	-1809.3	739.94	12	63.17	$(LD_{50})_2 > (LD_{50})_1$
2	Proposed molecule, P_5	7	1	2754.55	-1294.19	804.30	158	70.50	

Conclusions:

Insertion of –N atom in –R group always increases toxicity while Insertion of –C and –S atom decreases toxicity so insertion of –N atom should be minimized or ignored during modification of molecule till the other parameter do not require/ support it.

However, by the developed reg. eq's for predictions of some other activities viz. , MIC_{50}, IC_{50}, MIC_{50}, ED_{50}, this is possible (although with lengthy calculations) to design a **'Dream cephalosporin molecule'** with all high activity paramaters i.e. minimum value of half life, MIC_{50}, IC_{50}, MIC_{50}, ED_{50} but more value of LD_{50} and other physical properties with desirable range. Although this will be time consuming but if recent software enable to perform such related calculations are applied then this becomes quickly.

The ratio of LD_{50} to ED_{50} is known **Therapatic Index (T.I.**) For a good drug T.I. should be of more value as far as possible. So, by the present study this is possible to modify a particular cephalosporin drug molecule towards high activity with high T.I. (i.e. vast safety range of dose) by considering reg. eq. (3) and (5) collectively with individual activity/ property reg. eq. For example a new cephalosporin molecule with more activity (low IC_{50}) can be developed with high safty range (i.e. high T.I.) by applying eq. (3) & (5) collectively during the modeling of drug.

3D-MoRSE and Sum of Keir-hall topological indices are relatively more useful for prediction of activities than properties. Among 3D-MoRSE indices $(Mor_{1u})^{wox}$ and sum of Keir-Hall topological distances (S_s) with out –X group proved to be more valuable than others. Since, indices without –X group has more significant role than indices without –R group.

References:

1. E. Hodgson, In: A Textbook of Modern Toxicology, Wiley-Interscience, 2004, pp 3.
2. http://www.inchem.org/documents/pims/chemical/pim507.htm, (Chemical Safety Information from Intergovernmental Organizations)
3. D. O. Fleming, D. L. Hunt, In: Biological Safety: principles and practices, ASM Press Washington, DC, 2000, pp 267.
4. R. L. Lipnick, J. A. Cotruvo, R. N. Hill, R. D. Bruce, K. A. Stitzel, A. P. Walker, I. Chu, M. Goddard, Food and chemical toxicology**,** 1995, 33 (3), pp 223–31.
5. C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, Adv. Drug Deliv. Rev, 2001, 46, 1 (3), pp 3–26.
6. C. A. Lipinski, Drug Discovery Today: Technologies, 2004, 1 (4), pp 337–341.
7. A. K. Ghose, V. N. Viswanadhan, J. J. Wendoloski, J. Comb. Chem., 1999, 1 (1), pp 55- 68.
8. http://www.lookchem.com/Cefacetrile
9. I. Kenkyu, Study of Medical Supplies, 1982, 13, pp 1128.
10. Veterinary Substance database, University of Hertfordshire, May, 2011.
11. http://www.chemcas.com/msds/cas/msds130/82219-81-6.asp
12. http://www.lookchem.com/Ceftiofur/
13. Material safety data sheet for cefonicid, Clearsynth Labs Pvt. Ltd., Mumbai, pp 1-3.
14. K. Shinagawa, Japanese Journal of Antibiotics. 1990, 43, pp 1238.
15. http://www.druglead.com
16. D. Eleftherioy, Thesis, Kalamazod college, Winter, 1988.
17. K. Shinagawa, Japanese Journal of Antibiotics, 1974, 27, pp 746.
18. http://www.drugfuture.com/
19. American Soc. for Microbiology, Antimicrobial Agents and Chemotherapy,1972, 1, pp 67
20. Veterinary Substance database, University of Hertfordshire, May, 2011.
21. I. Kenkyu, Study of Medical Supplies, 1982, 13, pp 1128.
22. Y. Doi *et.al*, Antimicrobial agents and Chemotherapy, July 2004, pp 2652-2658.
23. http://www.druglead.com/cds/cefuzonam.html
24. M. C. Nahata *et. al.,* Antimicrob Agents Chemother, 1995, 39 (10), pp 2348-2349.
25. http://www.chemcas.com/msds/cas/msds130/82219-81-6.asp
26. Kamiosaki, Shinagawa, NKRZAZ (Nippon Kagaku Ryoho Gakkai) Chemotherapy (Tokyo), 1982, 30(3), pp 232.
27. http://www.hiintermediates.com/cas-766/76648-01-6.html
28. Y. Azuma, H. Tanimura , Y. Aoki, T. Isii , H. Goi, Nihon Geka Hokan, Sept. 1989, 58 (5), pp 445-51.
29. I. Kenkyu, Study of Medical Supplies, 1993, 24, pp 186.
30. I. I. Kenkyu, Study of Medical Supplies, 1992, 23, pp 93.
31. I. Kato, K. Nishimura, M. Ueno, S. Inoue, A. Harihara, K. Yabuuchi, K. Sato, H. Miyauchi, M. Hirata, Y. Kimura, H. Furukawa, The Journal of Toxicological Sciences, 2001, 26 (6-1), 5-12.
32. Drugs in Japan, 1995, pp 667.
33. Clearsynth., Material safty data sheet for cefpodoxime, 2003, pp 1- 3.

34. K. Rinsho. Clinical Report. 1986, 20, pp 3509.
35. http://www.guidechem.com/reference/dic-29148.html
36. K. Fukuhara, Y. Emi , K. Iwanami , T. Furukawa, T. Fujii , N. Itoh , S. Katsuki, H. Yoshimitsu, Y. Ii, N. Watanabe, Arzneimittelforschung. 1980, 30(10), pp 1669-79.
37. Material safety data sheet for cefriaxone, Clearsynth Labs Pvt. Ltd., Mumbai, pp 1-3.
38. G. Yakuji, Pharmaceuticals Monthly, 1997, 39, pp 2605.
39. Phyzer, Material safty data sheet for cefpodoxime proxetil, 2004, 1, pp 1-6.
40. http://www.hiintermediates.com
41. G. Yakuji, Pharmaceuticals Monthly, 1980, 22, pp 1605.
42. R. K. Sharma, Ph.D. thesis, Barkatullah University, Bhopal, 2014.
43. R. K. Sharma, A. Sikarwar, R. Sharma, P. Sharma, AJBPR, 2012, 2(4), pp 180-194.
44. R. K. Sharma, A. Sikarwar, IJRASET, 2013, 1(3), pp 85-95

-*-

BIOGRAPHIES

Dr. Rajendra Kumar Sharma has been teaching Engineering Chemistry at B.E. and Applied Chemistry at M.Sc. level since more than 8.5 years in SGSITS, Indore (M.P.) India. He also taught B.Sc. and M.Sc. ~ 3 years in Govt. Colleges of M.P. (India)

Professor, Dr. Arun Sikarwar is senior faculty at Govt. Home Sc. P.G. College Hoshangabad under M.P. Higher Education Department, India. He supervised and has been supervising many students for Ph.D. with U.G. and P.G. teaching more than 22 years.

Dr. Rajeev Sharma

Dr. Rajeev Sharma has been teaching in Govt. K. P. College (Vigyan Bhawan), Dewas. He has more than 12 years experience to teach B.Sc. and M.Sc.

Mm,

Dr. Pratibha Sharma is Assistant Professor in Department of Chemistry, ISLE, IPS Academy, Indore (M.P.) India. She has been teaching more than 7 years at the level of B.Sc. (Chemistry), M.Sc. (Chemistry and Medicinal Chemistry).

-#-