ORIGINAL ARTICLE

# Drug pKa value prediction - Using graph eccentricity

**M. Yamuna*, T. Divya**

*Department of Mathematics, School of Advanced Sciences, VIT, Vellore, Tamil Nadu, India*

**Abstract**

**Aim:** This study aims to predict drug pKa value using graph eccentricity. **Materials and Methods:** pKa values of drugs prescribed for specific disease are used for determining a linear equation using regression analysis. **Results and Discussion:** The pKa value of any new drug for the same disease can be determined using the linear equation and graph eccentricity. **Conclusion:** A study determines that the pKa value can be predicted using graph eccentricity.

**Key words:** Drug graph, drugs, eccentricity, molecular structure of drugs, pKa value, Wiener index

## INTRODUCTION

Harry Wiener has discussed that boiling point of organic compounds as well as all their physical properties depends functionally on the number, kind, and structural arrangement of the atoms in the molecule.[1] He also has discussed that within the group of isomers both the number and kind of atoms are constant and variation in physical properties are due to change in structural interrelationships alone. In this paper, Harry Wiener had defined a topological index relating polarity which later was termed as Wiener polarity index. Later, in[2] Adam Colin has established that there is a strong correlation between boiling point and Wiener index of all alkanes between 2 and 7 carbon atoms. This is a relationship where a quantitative property of a mathematical model of a molecule is closely related to physical property of the molecule. The relationship is established by fitting a power equation between the boiling point B and Wiener index W approximated by $B = 181\ W^{0.1775}$. Calculation of Wiener index involves an adjacency matrix. We know that if a graph G has n vertices, then we have to determine the distance between every pair of vertices, i.e., we have to estimate $n \times n$ values. Can we reduce the number of calculations, but still make it possible to determine the physical properties retaining back the distance property discussed by Wiener. The power equation establishes a solution between boiling point and Wiener index. Is there any statistical to that will help to establish this relation replacing the power formula. This paper attempts to answer these questions.

### Preliminaries

In this section, the basic properties required for developing the formula is provided.

### Drug Class

Drug classes are the groups of related drugs that have similar chemical structures and are used to treat the same disease. Snapshot 1 shows the drugs of similar structure that are used to treat malaria disease i.e., antimalarial drugs.[3]

### pKa value

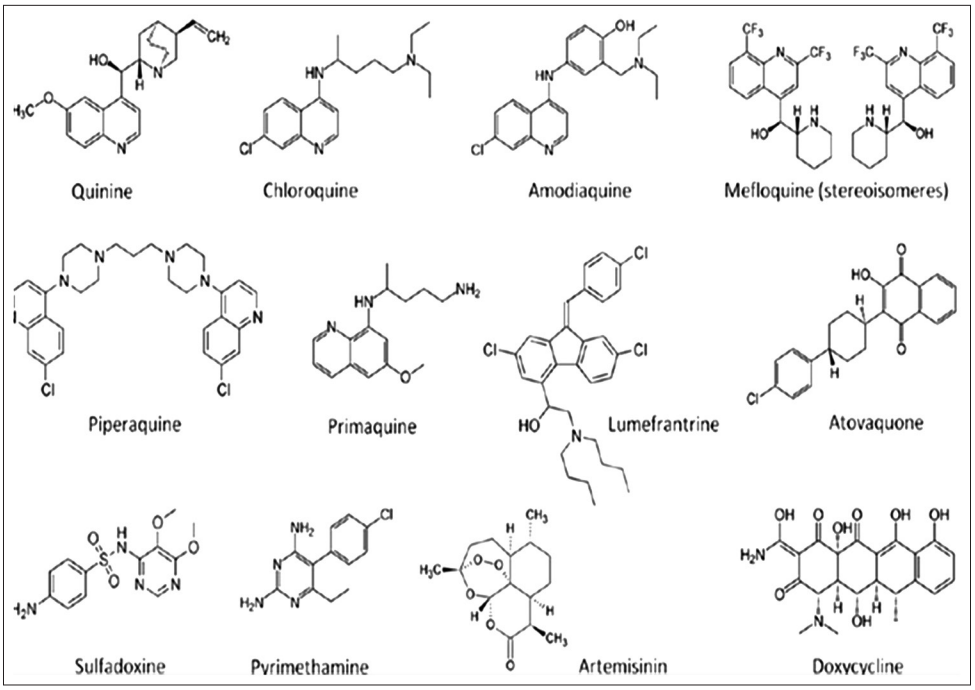
pKa value is defined as $pKa = -\log_{10}K_a$. The lower the pKa value, the stronger the acid. The following Snapshot 2 shows the acidic and basic values of some drugs.[4]

### Graph

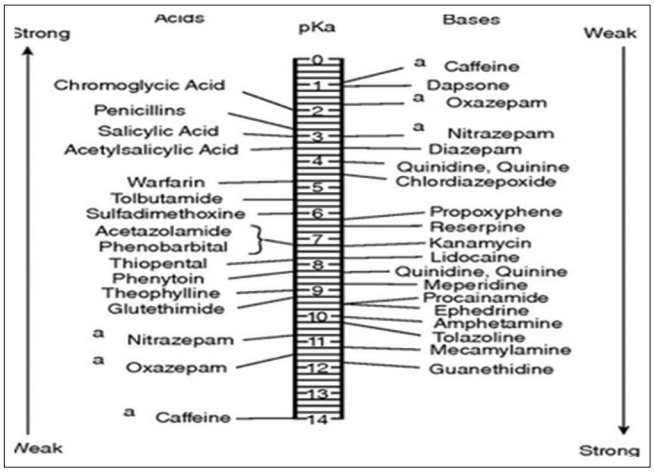
A graph G=(V,E) consists of a set V={v_1,v_2,…} called vertices and another set E={e_1,e_2,…}, whose elements are called edges, such that each edge e_k is identified with an unordered pair (v_i,v_j) of vertices. The number of vertices is called the order of graph and denoted by |V|=n. The number

**Address for correspondence:**
M. Yamuna, Department of Mathematics, School of Advanced Sciences, VIT, Vellore, Tamil Nadu, India.
E-mail: myamuna@vit.ac.in

**Received:** 27-03-2018
**Revised:** 17-07-2018
**Accepted:** 29-07-2018




**Snapshot 1:** The drugs of similar structure that are used to treat malaria disease i.e., antimalarial drugs[3]




**Snapshot 2:** The acidic and basic values of some drugs[4]

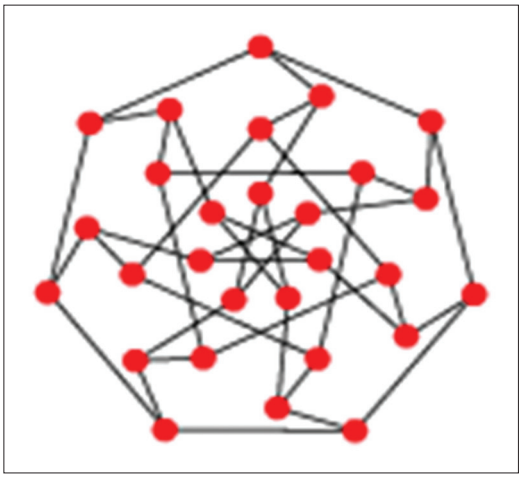


**Figure 1:** Example of a graph[6]

of edges is called the size of graph and denoted by |E|=m [5] Figure. 1 provides an example of graph.[6]

## Molecular Graph

In chemical graph theory and mathematical chemistry, a molecular graph or chemical graph is a representation of the structural formula of a chemical compound in terms of graph theory, whose vertices correspond to the atoms of the compound and edges correspond to chemical bonds. Figure 2 shows an example of the molecular graph of chemical compound[7] and the corresponding graph structure.

## Path

An open walk in which no vertex appears more than once is called a path and is denoted by $P_n$. The number of edges in a path is called the length of the path.[5] A path with n-vertices is denoted by $P_n$. Examples of paths with $n$ = 1, 2, 3, 4 vertices are shown in Figure 3.

Length of a path is the number of edges in the path. The length of a path with n-vertices is n – 1.

## Eccentricity

The eccentricity of any vertex v in a graph G is defined as the length of the longest possible path from the vertex to any other vertex in the graph.

For example, for the graph in Figure 4, the length of the longest possible path is to vertex 5, namely 1 to 2 to 3 to 4 to 5. Hence, the eccentricity of vertex 1 is 3. Similarly, the

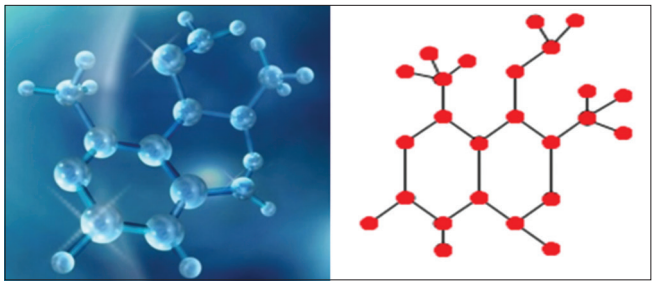


**Figure 2:** Example of a molecular graph and its corresponding graph structure[7]

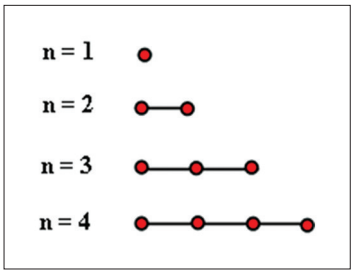



**Figure 3:** Examples of paths with n = 1, 2, 3, 4 vertices[5]

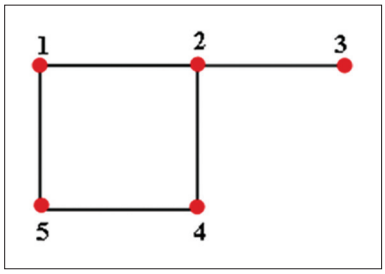



**Figure 4:**Example of a graph with eccentricity 3, 3, 4, 3, 3

eccentricity of the vertices 2, 3, 4, and 5 is 3, 4, 3, and 3, respectively.

## Drug Graph Determination

We know that any medicine has the chemical formula and hence a basic molecular structure. We draw a chemical graph for this molecular structure, where vertices represent atoms and edges represent the bonds between them. Figure 5 shows the molecular structure, chemical graph, and of a drug tetracaine.[8]

## Proposed Method in Determining the pKa Values

In [2] a power formula to determine the boiling point from Wiener index of a graph is provided. To determine the Wiener index of a graph, the distance between every pair of vertices should be determined. Eccentricity can be determined for any vertex in a graph. If instead of determining Wiener index we determine only the eccentricity for every vertex, then instead of $n^2$ distance calculation for any graph G with n vertices we need to determine only n values. Hence, if the physicochemical properties can be determined using eccentricity only, then the number of calculations would be reduced by $n^2$-n.

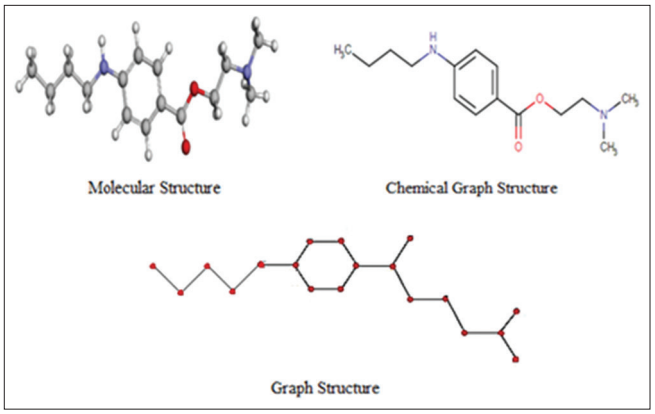



**Figure 5:** The molecular structure, chemical graph structure of a drug Tetracaine[8]

## Graph Eccentricity

Medicines prescribed for a particular kind of disease share some common molecular structure, and hence, their eccentricity values can be expected to be closer. For discussion, we choose 15 drugs prescribed as local anesthetics. Consider any drug graph. Determine the eccentricity of all the vertices of the drug graph. Sum of the eccentricity of all the vertices is defined as the eccentricity of the graph. Figure 6 shows the graph structure of the local anesthesia drug tetracaine. The vertices are labeled with their respective eccentricity values. Here, we get the sum of the eccentricities is 212.

Table 1 provides the molecular structure and graph structure of the randomly chosen local anesthetic drugs and their sum of the eccentricities.

Table 2 provides the Wiener index and graph eccentricity value for these 15 drugs. An observation of Table 2 shows that Wiener index is larger.

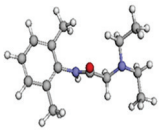
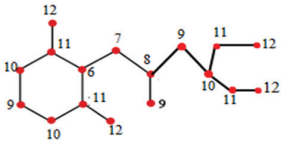
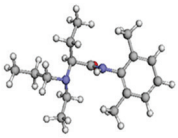
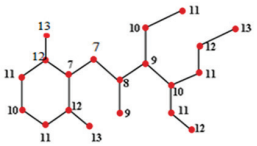
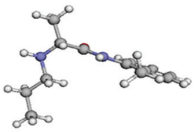
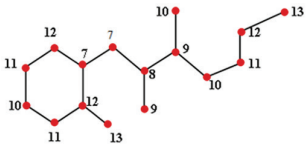
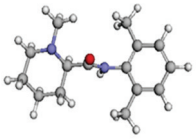
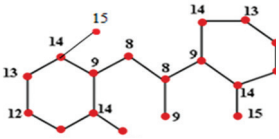
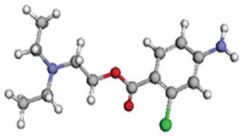
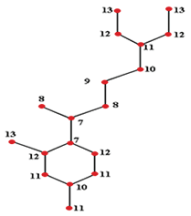
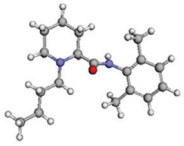
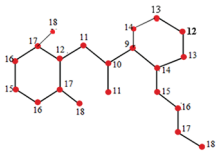
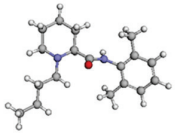
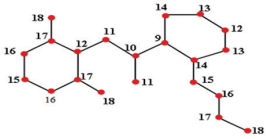
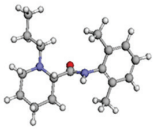
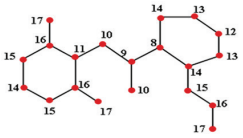
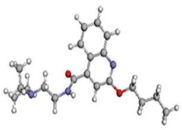
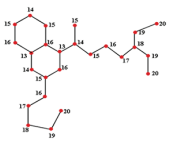
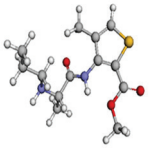
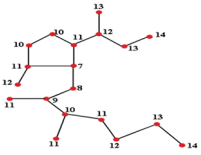
Graph 1 provides the correlation between the Wiener index and graph eccentricity for these 15 drugs.

From these graphs (a) and (b), we understand that as the Wiener index increases the graph eccentricity increases and as the Wiener index decreases the graph eccentricity also decreases. Hence, we can use graph eccentricity instead of Wiener index.

## Regression Analysis

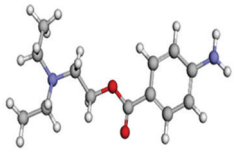
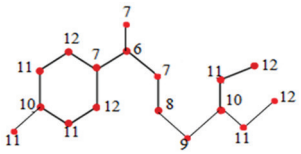
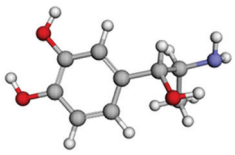
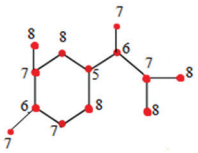
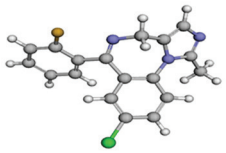
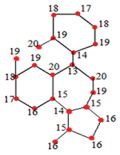
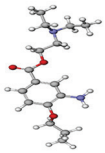
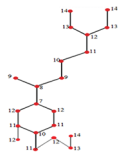
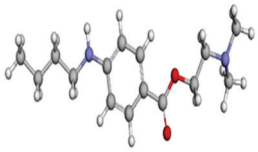
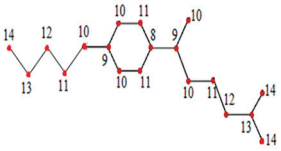
We know that in statistical modeling, regression analysis is a statistical process to estimate the relation among variables. When the focus is on the relationship between a dependent variable and one or more independent variables, this technique is generally used. More specifically regression analysis helps us understand the changes in the typical value of the dependent variable when any one of the independent variables is varied while the other independent variables are held fixed.[9] A simple linear regression is a linear

**Table 1:** The eccentricity value of the molecular structure of local anesthetic drugs

| Drug | Ball and stick representation | Drug graph with eccentricity values | Sum of the eccentricities |
|---|---|---|---|
| Lidocaine |  |  | 170 |
| Etidocaine |  |  | 212 |
| Prilocaine |  |  | 165 |
| Mepivacaine |  |  | 220 |
| Chloroprocaine |  |  | 190 |
| Bupivacaine |  |  | 302 |
| Levobupivacaine |  |  | 302 |
| Ropivacaine |  |  | 272 |
| Cinchocaine |  |  | 410 |
| Articaine |  |  | 211 |

*(Contd...)*

**Table 1:** *Continued*

| Drug | Ball and stick representation | Drug graph with eccentricity values | Sum of the eccentricities |
|---|---|---|---|
| Procaine |  |  | 167 |
| Levonordefrin |  |  | 92 |
| Midazolam |  |  | 393 |
| Proparacaine |  |  | 238 |
| Tetracaine |  |  | 212 |

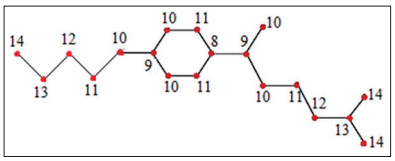
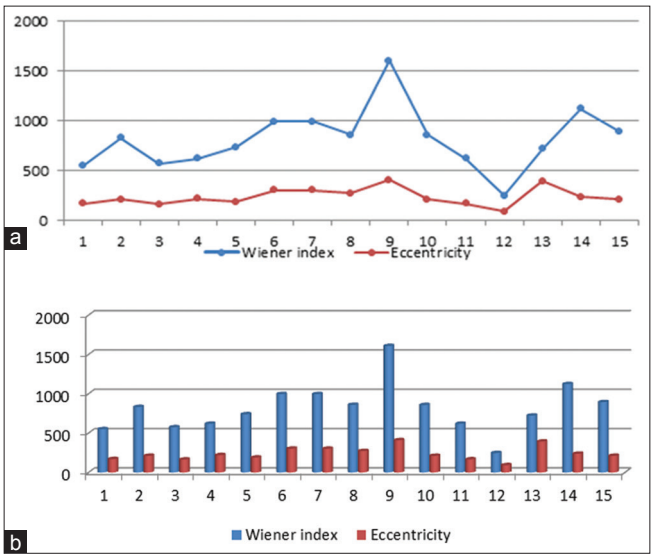


**Figure 6:** The graph structure of the local anesthesia drug Tetracaine

regression model which concerns two-dimensional sample points with one independent variable and one dependent variable. This determines a non-vertical straight line that predicts the dependent variable values as a function of the independent variables as accurately as possible. Hence, for any given set of data X and Y, a linear regression model is an equation of the form Y = a + b (X) where a and b are constants.

When medicines are prescribed for a particular kind of disease, the pKa value for the existing medicines is readily available. We have a procedure of determining the graph eccentricity. Whenever a new medicine for the same disease is discovered, then the basic molecular formula will be known. Hence, an approximate pKa value of this new medicine can be determined using regression analysis by fitting in a linear equation of a straight line. Let us randomly




**Graph 1: Correlation between Wiener index and eccentricity of drugs. (a) Line graph. (b) Column graph**

choose pKa value which is one of the physicochemical properties for discussion.

Table 3 lists the random 15 local anesthetics drugs, their pKa values, and the corresponding graph eccentricities.

**Table 2:** Value of Wiener index and eccentricity of drugs

| Name | Wiener index | Eccentricity |
|---|---|---|
| Lidocaine | 550 | 170 |
| Etidocaine | 832 | 212 |
| Prilocaine | 574 | 165 |
| Mepivacaine | 619 | 220 |
| Chloroprocaine | 739 | 190 |
| Bupivacaine | 996 | 302 |
| Levobupivacaine | 996 | 302 |
| Ropivacaine | 859 | 272 |
| Cinchocaine | 1608 | 410 |
| Articaine | 858 | 211 |
| Procaine | 619 | 167 |
| Levonordefrin | 246 | 92 |
| Midazolam | 722 | 393 |
| Proparacaine | 1123 | 238 |
| Tetracaine | 894 | 212 |

**Table 3:** pKa values and corresponding graph eccentricity of drugs

| Drugs | Original pKa (basic) values | Eccentricity (E) |
|---|---|---|
| Lidocaine | 7.75 | 170 |
| Etidocaine | 9.4 | 212 |
| Prilocaine | 8.82 | 165 |
| Mepivacaine | 7.25 | 220 |
| Chloroprocaine | 8.96 | 190 |
| Bupivacaine | 8.1 | 302 |
| Levobupivacaine | 8.1 | 302 |
| Ropivacaine | 7.82 | 272 |
| Cinchocaine | 9.04 | 410 |
| Articaine | 8.91 | 211 |
| Procaine | 8.96 | 167 |
| Levonordefrin | 8.96 | 92 |
| Midazolam | 6.57 | 393 |
| Proparacaine | 8.96 | 238 |
| Tetracaine | 8.42 | 212 |

## Linear Regression Fit for pKa Value Determination

Table 3 provides two sets of data, namely pKa and graph eccentricity. We can fit a linear regression for these two sets of data. We shall determine this equation using R software. We fit a regression equation for determining the pKa value of drug classes from the eccentricity of these drug graphs. Using the R software, we find the correlation and the regression equation between the eccentricity and pKa value of the above local anesthetic drugs. Thus, by knowing the linear regression

```
> Eccentricity=c(170,212,165,220,190,302,302,272,410,211,167,92,393,238,212)
> pKa=c(7.75,9.4,8.82,7.25,8.96,8.1,8.1,7.82,9.04,8.91,8.96,8.96,6.57,8.96,8.42)
> fit=lm(pKa~Eccentricity)
> fit

Call:
lm(formula = pKa ~ Eccentricity)

Coefficients:
 (Intercept)  Eccentricity
    9.281806     -0.003714
```

**Snapshot 3:** R program to fit a linear regression line

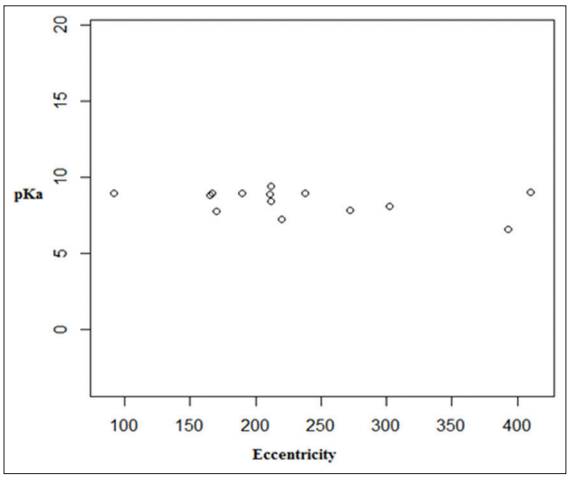



**Figure 7:** The Scatter diagram for drug graph eccentricities and pKa values of the drugs

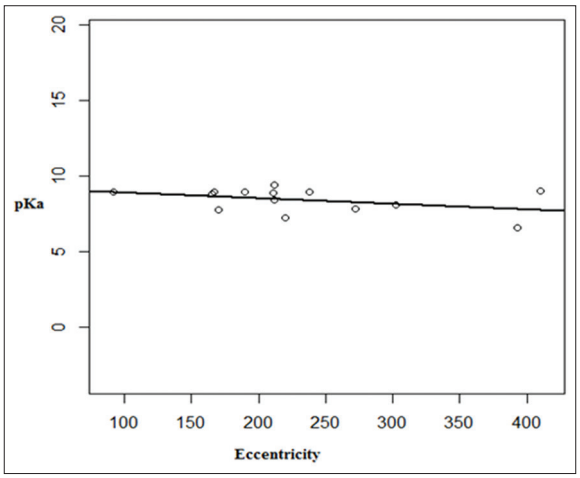



**Figure 8:** The linear regression line for the 15 local anesthetic drugs

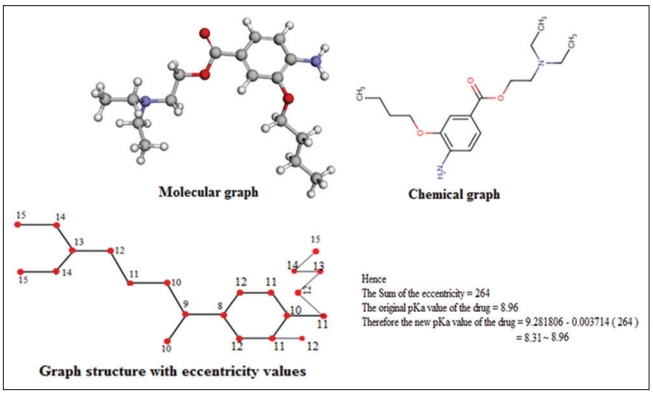



**Figure 9:** pKa value of Oxybuprocaine[10]

line for the randomly chosen local anesthetic drugs, we can find the pKa value of any other local anesthetic drugs which are having more or less same molecular structures. Snapshot 3 provides the R program to fit a linear regression line for the data in Table 3.

Hence, the linear regression equation for the randomly chosen 15 local anesthetic drugs is determined as pKa = 9.281806 – 0.003714 (eccentricity).

Figure 7 shows the Scatter diagram for drug graph eccentricities and pKa values of the drugs listed in Table 2 using the R program.

Figure 8 shows the linear regression fit line for the above listed 15 local anesthetic drugs.

## RESULTS AND DISCUSSION

Let us now consider a new local anaesthetic drug Oxybuprocaine. The drug graph and hence the graph eccentricity can be determined as discussed. The pKa value can be determined using the regression formula determined from the 15 local anaesthetic drugs. Hence, we get a new pKa value which is approximately equal to the original pKa value of the local anesthesia oxybuprocaine [Figure 9].[10]

## CONCLUSION

The proposed method is used to find the pKa value of drugs using the linear regression equation. This regression fit is the best fit to estimate the future values. Thus, by determining the regression line from randomly chosen 15 local anesthetic drugs with their graph eccentricity, the new physicochemical properties can be estimated which is approximate to the original values. Here, we found out that the regression fit can be used to find the pKa value which is one of the physicochemical properties of chemical compounds or drugs. This method can be used to find the other physicochemical properties such as boiling point and melting point using the graph property eccentricity.

## REFERENCES


1. Wiener H. Structural determination of paraffin boiling points. J Am Chem Soc 1947;69:17-20.
2. Adams C, Franzosa R. Introduction to Topology. New York: Pearson Education; 2008. p. 420-1.
3. Aminake MN, Prade G. Antimalarial drugs resistance in Plasmodium falciparum and the current strategies to overcome them. In: Méndez-Vilas, A, editor. Microbial Pathogens and Strategies for Combating Them. Spain: Science, Technology and Education; 2015. Available from: http://www.formatex.info/microbiology4/vol1/269-282.pdf.
4. Google Trends-Hot Keyword Searches. Pka Chart Medication Related Keywords. Available from: https://www.researchgate.net/publication/227579847/figure/fig7/AS:273283803250689@1442167344101/Figure-5-The-pKa-values-of-acidic-and-basic-drugs-vary-widelySome-of-drugs-are.png.
5. Deo N. Graph Theory with Applications to Engineering and Computer Science. Englewood Cliffs, N.J: Prentice-Hall, Inc.; 2010.
6. Eric WW. Coxeter Graph, From Math World--A Wolfram Web Resource; 2001. Available from: http://www.mathworld.wolfram.com/CoxeterGraph.html.
7. Adams W. Great Things from Small Things – Searching for Global Energy. Water and Food Solutions; 2014. Available from: http://www.genesisnanotech.com/wp-content/uploads/2014/06/Nano-Model-3-1018x1024.jpg.
8. National Center for Biotechnology Information. PubChem Compound Database, CID=5411. Available from: https://www.pubchem.ncbi.nlm.nih.gov/compound/5411. [Last accessed on 2018 Aug 30].
9. Wikipedia Contributors, Regression Analysis, Wikipedia, The Free Encyclopedia. Available from: https://www.en.wikipedia.org/w/index.php?title=Regression_analysis&oldid=850022021. [Last accessed on 2018 Jul 13].
10. Drug Bank CA. Available from: https://www.drugbank.ca/drugs/DB00892. [Last accessed on 2005 Jun 13].


**Source of Support:** Nil. **Conflict of Interest:** None declared.