

A Regression Analysis and Study of COX-2 Inhibitors Verified by Phylogenetic Tree

Chunduru Madhava Rao
Assistant Professor
Department of ECE
Sir C R Reddy College of Engg
Eluru, Andhra Pradesh
India

Yesu Babu Adimulam, PhD
Professor and HoD,
Department of CSE
Sir C R Reddy College of Engg
Eluru, Andhra Pradesh
India

ABSTRACT

Cyclo-oxygenase-2 (COX-2), a rate-limiting enzyme for prostanoid synthesis, is induced during inflammation and participates in inflammation-mediated cytotoxicity. The discovery of two cyclo-oxygenase (COX) isoenzymes, viz. COX-1 and COX-2, has updated the knowledge of non steroidal anti-inflammatory drugs (NSAIDs). The two COX isoenzymes share structural and enzymatic similarities, but are specifically regulated at the molecular level and may be distinguished apart in their functions, although some physiological overlap between them does occur. The major goal in developing selective COX inhibitors is to improve NSAID tolerability. Conventional non steroidal anti-inflammatory drugs (NSAIDs) nonspecifically inhibit cyclooxygenase-1 (COX-1), an enzyme critical to normal platelet function, and COX-2 which mediates inflammatory response mechanisms. Celecoxib, the 1, 5-diarylpyrazole compounds was the first launched selective COX-2 inhibitor, and has excellent selectivity and potent anti-inflammatory activity. COX-2 is required for both the constitutive and mitogen-induced PGE₂ synthesis. Moreover, over-expression and persistent expression of COX-2 may be influenced by breast tumor hormone status and seem to be a feature of the aggressive, metastatic phenotype type. Recent studies have indicated that the relationships between polyunsaturated fatty acid metabolism and carcinogenesis have led to new targets for the design of mechanism based drugs in cancer chemoprevention research. Selective inhibition of COX-2 provided a new class of anti inflammatory, analgesic, and antipyretic drugs with significantly reduced side effects. It has been reported that inhibiting COX-2 could also be an important strategy for preventing or treating a number of cancers. COX-2 selective inhibitors such as celecoxib, rofecoxib and valdecoxib are currently being used to reduce inflammatory response. However, they lack anti-thrombotic activity and hence lead to cardiovascular and renal liabilities apart from gastrointestinal irritation. Therefore, there is still a need to develop more potent COX-2 inhibitors. One of the keys to developing COX-2 selective drugs is the larger active site of COX-2, which makes it possible to make molecules too large to fit into the COX-1 active site but still able to fit the COX-2.

Keywords

COX-2, Phylogenetic tree, Correlation, Inhibitors.

1. INTRODUCTION

Cyclo-oxygenase (COX) is a rate-limiting enzyme that converts arachidonic acid to prostaglandins (PGs) and exists in two isoforms, COX-1 and COX-2. Despite the molecular and histological similarities between fetal bone development and fracture healing, inflammation is an early phase of

fracture healing that does not occur during development. Cyclo-oxygenase 2 (COX-2) is induced at inflammation sites and produces pro inflammatory prostaglandins. Radiographic, histological, and mechanical testing determined that fracture healing failed in rats treated with COX-2-selective NSAIDs (celecoxib and rofecoxib). Normal fracture healing also failed in mice homozygous for a null mutation in the COX-2 gene. This shows that COX-2 activity is necessary for normal fracture healing and confirms that the effects of COX-2-selective NSAIDs on fracture healing is caused by inhibition of COX-2 activity and not from a drug side effect. Histological observations suggest that COX-2 is required for normal endochondral ossification during fracture healing. Because mice lacking Cox2 form normal skeletons, our observations indicate that fetal bone development and fracture healing are different and that COX-2 function is specifically essential for fracture healing [1].

1.1 Differential functions of the COX proteins

If the only basis for the differences between the COX isoforms was their differential gene expression, then replacing the gene for COX-2 with that for COX-1 should produce no noticeable phenotype. However, “knockin” of the COX-1 gene into the COX-2 locus in mice only partially replenishes the deficit in PGI₂ synthesis and fails to fully ameliorate defects in reproductive and renal function associated with COX-2 deletion [2]. These results clearly indicate that COX-1 and COX-2 are not functionally interchangeable at the protein level.

One basis for the results of the COX-1 knock in study may lie in differential coupling between the two COX proteins and downstream syntheses. For example, lack of full restoration of PGI₂ synthesis by COX-1 knock in may be due to a failure of coupling between COX-1 with PGI synthases. Numerous studies support selective isoform association [3 and 4], but much of this work has been done with cells over expressing the relevant enzymes, and no basis for the differential coupling has been advanced. Therefore, confirmation of this hypothesis awaits further investigation.

Differences in substrate specificity between the COX isoforms are not limited to neutral derivatives of 20:4. Indeed, COX-2 has greater capacity to oxygenate a number of polyunsaturated free fatty acids that are poor substrates for COX-1 (Fig. 1) [5]. Since these include v-3 fatty acids, it is possible that the differential use of these lipids by the COX enzymes may help to explain the health benefits of dietary v-3 fatty acids, a hypothesis that remains an intriguing subject for future work [6].

A final difference in COX isoform function that may be of physiologic significance lies in the ability of the aspirin-treated enzymes to oxygenate AA. Aspirin inhibits PG formation by both isoforms via covalent modification of Ser-530 in the cyclo oxygenase active site. However, COX-2 retains the capacity to oxygenate AA, resulting in the formation of 15R-hydroperoxy-eicosatetraenoic acid instead of PGG₂ (34). Through this lipoxygenase-type reaction, aspirin-treated COX-2, together with other lipoxygenases, forms poly-hydroxylated lipids known as aspirin-triggered lipoxins and resolvins. Since these lipids have anti-inflammatory activity, their production may help to explain some of the clinical benefits of aspirin [7].

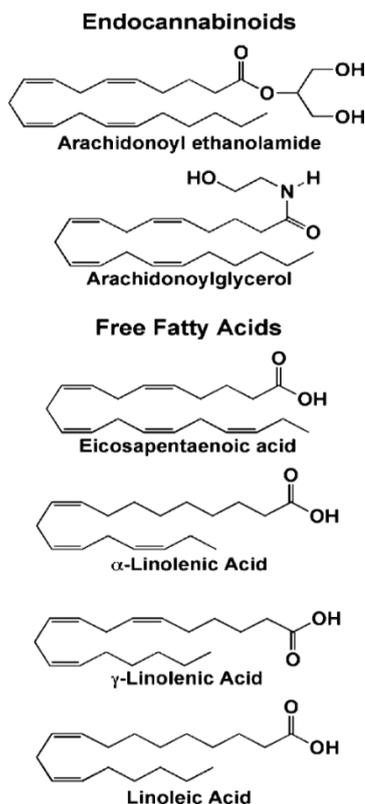


Fig.1: Selective COX-2 substrates. The structures of endocannabinoids and free fatty acids that are metabolized more efficiently by COX-2 than by COX-1 are shown.

2. MATERIALS AND METHODS

2.1 ISIS Draw

ISIS Draw is a chemical drawing program similar to Chem raw. When ISIS Draw is started, a window appears with a list of tools down the left hand side of the window, and a list of templates along the top. There are two buttons towards the top left of the window: Molecule and Sketch. To draw chemical structures, Molecule is selected. Using Sketch, one can draw other shapes which may be useful to annotate and beautify diagrams. An image showing 2D structure of a compound is given in Fig.2.

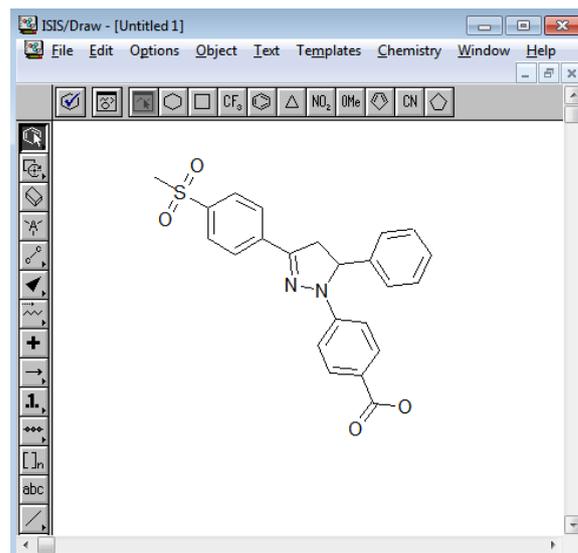


Fig. 2: Screen-shot image of ISIS Draw v3.2 showing compound 8a

2.2 Tsar Software

Tsar (Tools for Structure Activity Relationships) is an integrated analysis package used to investigate Quantitative Structure-Activity Relationships. It is intended to provide all the functions required to carry out any investigation, whether in pharmaceutical, agrochemical, toxicological, or any other area of application. It uses a spreadsheet to store and manipulate different types of data, including:

- Molecular descriptions
- 3D structures
- Activity data
- Computed data.

2.3 Correlation

The Correlation of the dependent and independent variables measures the strength of the relationship between the two. Since our unit of measure may influence the slope of our regression line, we want our correlation to be dimensionless. In general, a positive correlation indicates a strong positive relationship, where an increase in one variable implies an increase in the other variable. A negative correlation indicates that an increase in the first variable implies a decrease in the second variable. A correlation of zero does not necessarily imply there is no relationship between the variables, if the data is non-linear. Often, the Pearson Correlation Coefficient is used to measure correlation, as denoted r . Here is the formula:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

2.4 Python code: k-means

The complete python code implementing k-means algorithm is given here.

```
import math
import random
```

```
import time
from Tkinter import *
import matplotlib.pyplot as plt
t0 = time.clock()
def fileopen(fileName):
    fileHandler = open(fileName, "rt")
    lines = fileHandler.readlines()
    fileHandler.close()
    del lines[0]
    dataset = []
    for line in lines:
        obj = lineToTuple(line)
        dataset.append(obj)
    return dataset
def lineToTuple(line):
    clearTrace = line.strip()
    clearTrace = clearTrace.replace("'", "")
    lineClear = clearTrace.split(",")
    convertString(lineClear)
    lineTuple = tuple(lineClear)
    return lineTuple
def convertString(string):
    for i in range(len(string)):
        if (ValidString(string[i])):
            string[i] = float(string[i])
def ValidString(s):
    if len(s) == 0:
        return False
    if len(s) > 1 and s[0] == "-":
        s = s[1:]
    for c in s:
        if c not in "0123456789.":
            return False
    return True
def dist(obj1, obj2):
    if obj1 == None or obj2 == None:
        return float("inf")
    sumofsquares = 0
    for i in range(1, len(obj1)):
        sumofsquares += (obj1[i] - obj2[i])**2
    return sumofsquares
def meanobj(name, objList):
    numobjs = len(objList)
    if (numobjs == 0):
        return
    numAttributes = len(objList[0])
    means = [name] + [0] * (numAttributes-1)
    for obj in objList:
        for i in range(1, numAttributes):
            means[i] += obj[i]
    for i in range(1, numAttributes):
        means[i] /= float(numobjs)
    return tuple(means)
def assign(obj, centroids):
    mindist = dist(obj, centroids[0])
    mindistIndex = 0
    for i in range(1, len(centroids)):
        d = dist(obj, centroids[i])
        if (d < mindist):
            mindist = d
            mindistIndex = i
    return mindistIndex
def createEmptyListOfLists(numSubLists):
    string = []
    for i in range(numSubLists):
        string.append([])
    return string
def assignAll(objs, centroids):
    clusters = createEmptyListOfLists(len(centroids))
    for obj in objs:
        clusterIndex = assign(obj, centroids)
        clusters[clusterIndex].append(obj)
    return clusters
def computeCentroids(clusters):
    centroids = []
    for i in range(len(clusters)):
        name = "centroid" + str(i)
        centroid = meanobj(name, clusters[i])
        centroids.append(centroid)
    return centroids
def kmeans(objs, k, animation=False, initCentroids=None):
    result = {}
    if (initCentroids == None or len(initCentroids) < k):
        random.seed(time.time())
        centroids = random.sample(objs, k)
    else:
```

```
centroids = initCentroids
prevCentroids = []
iteration = 0
while (centroids != prevCentroids):
    iteration += 1
    clusters = assignAll(objs, centroids)
    prevCentroids = centroids
    centroids = computeCentroids(clusters)
result["clusters"] = clusters
result["centroids"] = centroids
print "clusters", clusters
print "centroids", centroids
return result
def output(objs):
    for obj in objs:
        if obj != None:
            line = obj[0] + "\t"
            for i in range(1, len(obj)):
                line += "%.2f " % obj[i]
            print line
dataset = fileopen("kmeans.csv")
clustering = kmeans(dataset, 3, True)
output(clustering["centroids"])
t1 = time.clock()
print("The search took {0:.2f} seconds.".format(t1-t0))
```

3. RESULTS AND DISCUSSION

Multivariate regression analysis using python program resulted in few influential parameters displayed significant positive and negative contribution towards biological activity of COX-2 inhibitors.

3.1 Clustering Analysis

The k-means clustering algorithm is popular because it can be applied to relatively large sets of data. The user specifies the number of clusters to be found. The algorithm then separates the data into spherical clusters by finding a set of cluster centers, assigning each observation to a cluster, determining new cluster centers, and repeating this process.

A modified k-means python program was written and implemented to evaluate the clusters and assessing the probability of designed compounds to appear in a particular cluster thereby confirming the similarity of such compounds to appear in a cluster of approved drugs.

The data file is formatted as a comma delimited file .csv and the file was called from python. Further clustering code was implemented to perform k-means clusters on the dataset by calculating number of clusters for given centroids. The number of centroids can be changed depending on the data being clustered. In this case, 3 centroids are considered. Now, the k-means algorithm was implemented by finding distances between objects taken for the study. Once distances between

objects are calculated, the sum of squares for the two objects from the centroid is calculated. Following this, means are calculated for all objects which are nearer to each other making it to study all possible objects and the List of such objects will be created by the program to iterate the process. Centroids are assigned and the distance is calculated from each centroid to the objects and clusters are identified. To find out better clusters, a minimum distance is calculated from each object to the nearest assigned centroid. Once each object is assigned to its nearest centroid, then the distances are recalculated by adjusting the centroids in such a way that the objects should have the shortest possible distance with the centroid. Likewise, a cluster index is created and appended to the objects. Finally, the k in k-means is calculated by randomly selecting k initial centroids.

Algorithm: K-Means

1. Initialization: choose k initial centroids arbitrarily (or randomly).
2. Assign each data point to the centroid that is closer to it.
3. Compute the distance between the centroids and objects using the Euclidean Distance equation
4. Update all the centroids. The new centroid of a cluster is the mean of all the points within that cluster.
5. Repeat points 2 and 3 until the new centroids are the same as the previous centroids.

K-means suffers from drawback on the number of clusters k as an input parameter. This is because of an inappropriate choice of k which might yield spurious results. Hence, it is always an important task to run diagnostic checks when using k-means clustering to determine the number of clusters in the given dataset. Moreover, applying k-means with values ranging from k=2, 3, 4 or 5 depends on the number of objects in the dataset and to avoid expected clusters of similar size so that the assignment of objects to the nearest cluster center or centroid will result in correct clusters. Therefore, an attempt has been made to cluster dataset using different k value ranges. The first step in k-means algorithm is to divide the given data set into user defined number of clusters. The initial choice of k in k-means is an interpretive decision and successive runs should be performed to obtain an optimized division of data for any chosen k value. A prior knowledge on the data structure would result in more appropriate clusters. However, as the dimensionality of the data increases, it becomes increasingly difficult to determine a proper value for k [8].

3.2 FIT Kubinyi Function

All the three models passed the conditions for validation sets. Further, to define the statistical quality of activity prediction, the number of variables that enter in a regression equation are compared by using FIT Kubinyi function (Eq. 10), a criteria closely related to F value was proven to be useful [9].

Eq. 10:

$$FIT = R^2 (n - k - 1) / (n + k^2) (1 - R^2)$$

Where n is the number of compounds in training set and k is the number of variables in the regression equation.

The main feature of the F value is its sensitivity to changes in k, if k is small and its lower sensitivity if k is large. The FIT criterion has a low sensitivity towards changes in k values, as

long as they are small numbers, and a substantially increasing sensitivity for large k values [10]. The best model will be the one that possess a high value of this function. Hence, regression models with 12, 9 and 11 variables are generated (Table 1) to choose the best among them.

Table 1: Statistical parameters of the regression models obtained for 12, 9 and 11 variables.

Var ^a	r ²	k	FIT	Eq No.
12	0.782	12	0.799	10
9	0.645	9	0.627	11
11	0.741	11	0.735	12

3.3 Screening non-tested compounds from literature

Based on the above, it can be stated that the proposed model due to its high predictive ability can be utilized as an alternative aid to the costly and time consuming experiments for recognizing and determining compounds with high COX-2 binding affinity. The method can also be used to screen similar repertoire of compounds reported in various literature sources or chemical compound libraries to identify new potentially active compounds.

Hence, a group of new derivatives from literature are subjected to screening utilizing the produced model. We, therefore, searched for compounds with pyrazoles, pyrazolines, or other pyrazole derivatives in Archives of organic chemistry journal [11]. Therefore, a set of 32 compounds with pyrazole ring as main nucleus was selected from a published review paper [12]. The structures are devoid of biological activity against COX-2, however, these pyrazole compounds are known to inhibit certain disease causing protein targets, for example, Factor Xa, IXa, platelet aggregation inhibition, antidepressant, voltage-gated sodium channel blockers etc. All compounds screened for this study are provided with compound numbers as given in reference article (Table 2).

Table 1: Compounds along with predicted activity data after applying regression model on this 32 compound dataset

Molecule No.	Predicted Activity (log ₁ /IC ₅₀)	Calculated IC ₅₀ (uM)
28.mol	-0.6172	4.1420
35.mol	-1.1829	15.2383
37.mol	-2.0143	103.3463
38.mol	0.5860	0.2594
42.mol	-2.2016	159.0788
52.mol	-2.2864	193.3873
57.mol	-1.0800	12.0230
58.mol	0.7703	0.1697
62.mol	0.1038	0.7874
64.mol	-2.3705	234.7178
65.mol	-2.1758	149.9144
66.mol	-1.5239	33.4115
67.mol	0.9942	0.1013

68.mol	1.9191	0.0120
69.mol	2.1406	0.0072
70.mol	0.6351	0.2317
72.mol	1.8572	0.0139
73a.mol	1.1774	0.0665
73b.mol	2.8803	0.0013
73c.mol	1.8022	0.0158
74a.mol	1.2406	0.0575
74b.mol	-0.1164	1.3074
74c.mol	1.9006	0.0126
75.mol	-0.5985	3.9675
76a.mol	-0.3190	2.0843
76b.mol	-1.7174	52.1630
77.mol	1.4140	0.0385
78.mol	-1.5899	38.8932
79.mol	-2.0934	123.9857
80.mol	-3.6364	4329.3848
81.mol	-2.8340	682.3371
90.mol	-0.8823	7.6259
Celecoxib	0.2472	0.5660

4. CONCLUSIONS

Cyclooxygenase-2 is an enzyme which is mainly used in the treatment of various disorders such as breast cancer, atherosclerosis etc. Cyclooxygenase-2 (COX-2) enzyme binds to arachidonic acid and releases metabolites that are used to induce pain and inflammation. COX-2 selective inhibitors such as celecoxib, rofecoxib and valdecoxib are currently being used to reduce inflammatory response. However, they lack anti-thrombotic activity and hence lead to cardiovascular and renal liabilities apart from gastrointestinal irritation. Therefore, there is still a need to develop more potent COX-2 inhibitors.

Further, applying FIT Kubinyi function on the models resulted in model with 12 variables as the best model. Hence, designing or screening compound libraries for new compounds or analogs with possible H-bond acceptor and donor substitutions, and increase in number of Fluorine atoms and decrease in molecular weight, logp and rotatable bonds on the molecule with concomitant increase in dipole. Lipole and molar refractivity would enhance inhibitory activity against COX-2.

Finally, a phylogenetic tree was presented to know the branches on which these novel compounds appear and hence can be used to further extend the study towards identifying more statistical parameters that would substantiate the identification of entirely novel, potent COX-2 inhibitors.

5. REFERENCES

- [1] Developmental expression of the cyclo-oxygenase-1 and cyclo-oxygenase-2 genes in the peri-implantation mouse uterus and their differential regulation by the blastocyst and ovarian steroids {Chakraborty, SK Das, J Wang, and SK Dey,}, Journal of Molecular Endocrinology, Vol 16, Issue 2, 107-122.
- [2] Yu, Y., J. Fan, Y. Hui, C. A. Rouzer, L. J. Marnett, A. J. Klein-Szanto, A. FitzGerald, and C. D. Funk. 2007. Targeted cyclooxygenase gene (ptgs) exchange reveals discriminant isoform functionality. Biol. Chem. 282: 1498–1506.
- [3] Ueno, N., Y. Takegoshi, D. Kamei, I. Kudo, and M. Murakami. 2005. Coupling between cyclooxygenases

- and terminal prostanoid syn-thases. *Biochem. Biophys. Res. Commun.* 338: 70–76.
- [4] Smith, W. L., and R. Langenbach. 2001. Why there are two cyclo-oxygenase isozymes. *J. Clin. Invest.* 107: 1491–1495.
- [5] Laneuville, O., D. K. Breuer, N. Xu, Z. H. Huang, D. A. Gage, J. T. Watson, M. Lagarde, D. L. DeWitt, and W. L. Smith. 1995. Fatty acid substrate specificities of human prostaglandin-endoperoxide H synthase-1 and -2. Formation of 12-hydroxy-(9Z,13E/Z,15Z)-octadeca - trienoic acids from α -linolenic acid. *J. Biol. Chem.* 270: 19330–19336.
- [6] Smith, W. L. 2005. Cyclooxygenases, peroxide tone and the allure of fish oil. *Curr. Opin. Cell Biol.* 17: 174–182.
- [7] Serhan, C. N., N. Chiang, and T. E. Van Dyke. 2008. Resolving inflammation: dual anti-inflammatory and pro-resolution lipid mediators. *Nat. Rev. Immunol.* 8: 349–361.
- [8] Jonathan Baarsch and M. Emre Celebi "Investigation of Internal Validity Measures for K-Means Clustering" in *Proceedings of the International Multi conference of Engineers and Computer Scientists 2012 Vol 1, IMCES 2012, March 14-16 2012, HongKong.*
- [9] Kubinyi H. Variable selection in QSAR studies. II. A Highly Efficient Combination of Systematic Search and Evolution. *Quant. Struct. Act. Relat.* 13, 1994, 393-401.
- [10] Kubinyi H. Variable selection in QSAR studies. I. An evolutionary algorithm. *Quant. Struct. Act. Relat.* 13, 1994, 285-294.
- [11] www.arkat-usa.org
- [12] Ruth Pérez-Fernández, Pilar Goya, and José Elguero. A review of recent progress (2002-2012) on the biological activities of pyrazoles. *ARKIVOC* 2014 (ii) 233-293.