1. Introduction

Rigorous analysis and one-tuning of independent variables has led to an expansion in development of molecular and atom-based descriptors, as well as descriptors derived from quantum chemical calculations and spectroscopy (Livingstone 2000). The improvement in high-through put screening procedures allows for rapid screening of large numbers of compounds under similar test conditions and thus minimizes the risk of combining variable test data from many sources. The formulation of thousands of equations using quantitative structure activity relationships (QSAR) methodology at test to validate its concepts and its utility in the elucidation of the mechanism of action of drugs at the molecular level and a more complete understanding of physicochemical phenomena such as hydrophobicity. It is now possible not only to develop a model for a system but also to compare models from a biological database and to draw analogies with models from a physical organic database (Hansch 1962).

QSAR analysis is a modern methodology that can correlate to the structure of a chemical compound to its biological activity. This can be achieved by deriving empirical models which describe the structural dependence of biological activities either by physicochemical parameters, or by three-dimensional molecular property profiles of the compounds (Kubinyi 1977). The biological activity of compounds used in QSAR, can be expressed as compound concentration that inhibits 50% of e.g. cell growth (IC_{50}), or as compound dose that causes 50% of a certain biological effect (ED_{50}), or compound dose that causes death of 50% of laboratory animals (LD_{50}).

1.1 Historical Development of QSAR

More than a century ago, Crum-Brown and Fraser expressed the idea that the physiological action of a substance was a function of its chemical composition and constitution (Crum-Brown 1968). A few decades later, in 1993, Richet showed that the cytotoxicities of a diverse set of simple organic molecules were inversely related to their corresponding water solubilities. At the turn of the 20th century, Meyer and Overton independently suggested that the narcotic (depressant) action of a group of organic compounds paralleled their olive oil/water partition coefficients (Lorentz 2007). In 1939 Ferguson introduced a thermodynamic generalization to the correlation of depressant action with the relative saturation of volatile compounds in the vehicle in which they were administered (Ferguson 1939). The extensive work of Albert, Bell and Roblin established the importance of ionization of bases and weak acids in bacteriostatic activity (Albert 1945, Bell 1942, Hammett 1935). Meanwhile on the physical organic front, great strides were being made in the delineation of substituent effects on organic reactions, led by the work of Hammett (Hammett 1966). Taft devised a way for separating polar, steric, and resonance effects and introducing the first steric parameter, ES (Taft 1952).

In the past, important qualitative concepts evolved from QSAR studies are; the role of different physicochemical properties being responsible for the drug –receptor interaction; the understanding of influence of lipophilicity and ionization on drug transport and distribution within a biological system; the concept of optimum lipophilicity of a drug (e.g. gastrointestinal absorption or transfer through the blood-brain barrier). Currently, QSAR are being applied in many disciplines; drug design,

environmental risk assessment, environmental regulations (specific modes of action).

The first general formulation development of true QSARs was presented by Crum-Brown and Faster in 1868, (Lorentz 2007), who assumed that biological activity (Φ), is a function of chemical structure (C)

$$\Phi = f(C) \;;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

Hence, the QSAR was still a long way to go, because it was necessary to define proper measures of suitable mathematical formalisms for the function *f* and methods to quantitatively describe chemical structure.

In the 1890's, Hans, noted that the toxicity of organic compounds depended on their lipophilicity (Lipnick 1986). In 1893, Richet showed that the cytotoxicity of a diverse set of simple organic molecules was inversely related to their corresponding water solubility's (Lorentz 2007). At the turn of the 20^{th} century, Meyer and Overton discussed the dependence of the depressant action, of a group of organic compounds, on their olive oil/water partition coefficients (Selassie 2003). Hammett equation relates electronic properties of aromatic organic compounds to their activity and developed the concept of linear free-energy relationships, which states that a combination of correlation factors can be used to describe activity of certain compounds. Taft extended Hammett's idea by introducing the first steric parameter that correlates with hydrolysis rate constants (Taft RW., 1950). The contributions of Hammett and Taft together laid the mechanistic basis for the development of the QSAR paradigm by Hansch and Fujita (Hansch 1964). In 1962 Hansch and Muir published their study on the structure-activity relationships

of plant growth regulators and their dependency on Hammett constants and hydrophobicity (Hansch 1962). Using the octanol/water system, a whole series of partition coefficients was measured, and thus a new hydrophobic scale was introduced. Fujita and Hansch then combined these hydrophobic constants with Hammett's electronic constants and Taft's steric parameter to yield the linear Hansch equation and its many extended forms (Ferguson 1939). Hundreds of equations later, led to the development of the Hansch parabolic equation and the field of QSAR has come a long way. The Kubinyi bilinear model is a refinement of the parabolic model and, in many cases; it has proved to be superior in QSAR (Klopman 1985).

1.2 Tools and techniques of QSAR

1.2.1 Biological parameters

In QSAR analysis, it is imperative that the biological data be both accurate and precise to develop a meaningful model. It must be realized that any resulting QSAR model that is developed is only as valid statistically as the data that led to its development. The equilibrium constants and rate constants that are used extensively in physical organic chemistry and medicinal chemistry are related to free energy values ΔG. Thus for use in QSAR, standard biological equilibrium constants such as ionization constant Ki should be used. Likewise only standard rate constants should be deemed appropriate for a QSAR analysis. Percentage activities (e.g., % inhibition of growth at certain concentrations) are not appropriate biological end points because of the nonlinear characteristic of dose-response relationships. These types of points may be transformed to equieffective molar doses. Only equilibrium and rate constants passmuster in terms of the free-energy

relationships or influence on QSAR studies. Biological data are usually expressed on a logarithmic scale because of the linear relationship between response and log dose in the mid region of the log dose-response curve. Inverse logarithms for activity (log 1/C) are used so that higher values are obtained for more effective analogs. Various types of biological data have been used in QSAR analysis. Biological data should pertain to an aspect of biological /biochemical function that can be measured. The events could be occurring in enzymes, isolated or bound receptors, in cellular systems, or whole animals. Because there is considerable variation in biological responses, test samples should be run in duplicate or preferably triplicate, except in whole animal studies where assay conditions (e.g., plasma concentrations of a drug) preclude such measurements.

It is also important to design a set of molecules that will yield a range of values in terms of biological activities. It is understandable that most medicinal chemists are reluctant to synthesize molecules with poor activity, even though these data points are important in developing a meaningful QSAR. Generally, the larger the range (.2 log units) in activity, the easier it is to generate a predictive QSAR (Donald 2003).

1.2.2. Electronic parameters

Parameters are of critical importance in determining the types of intermolecular forces those underlay drug-receptor interactions. The three major types of parameters that were initially suggested are electronic, hydrophobic, and steric in nature (**Hansch**1969, **Hansch**2001). Extensive studies using electronic parameters reveal that electronic attributes of molecules are intimately related to their chemical reactivities and biological activities. The extent to which a given reaction responds to electronic

perturbation constitutes a measure of the electronic demands of that reaction, which is determined by its mechanism. The introduction of substituent groups into the framework and the subsequent alteration of reaction rates helps delineate the overall mechanism of reaction. Early work examining the electronic role of substituent on rate constants was first tackled by Burckhardt and firmly established by Hammett (Hammett 1935, Hammett 1970). Hammett employed, as a model reaction, the ionization in water of substituted benzoic acids and determined their equilibrium constants K_a. In recent years, there has been a rapid growth in the application of quantum chemical methodology to QSAR, by direct derivation of electronic descriptors from the molecular wave functions (Karelson 1996). The two most popular methods used for the calculation of quantum chemical descriptors are *Ab initio* (Hartree - Fock) and semi empirical methods. As in other electronic parameters, QSAR models incorporating quantum chemical descriptors will include information on the nature of the intermolecular forces involved in the biological response. Unlike other electronic descriptors, there is no statistical error in quantum chemical computations. Quantum chemical descriptors such as net atomic changes, highest occupied molecular orbital/lowest unoccupied molecular orbital (HOMO-LUMO) energies, frontier orbital electron densities, and super delocalizabilities have been shown to correlate well with various biological activities (Gupta1991).

1.2.3 Hydrophobicity parameters

More than a hundred years ago, Meyer and Overton made their seminal discovery on the correlation between oil/water partition coefficients and the narcotic potencies of small organic molecules (Selassie 2003). Ferguson extended this analysis by placing the relationship between

depressant action and hydrophobicity in a thermodynamic context; the relative saturation of the depressant in the bio phase was a critical determinant of its narcotic potency (Freguson 1939). At this time, the success of the Hammett equation began to permeate structure-activity studies and hydrophobicity as a determinant was relegated to the background. In a landmark study, Hansch and his colleagues devised and used a multi parameter approach that included both electronic and hydrophobic terms, to establish a QSAR for a series of plant growth regulators (Iwamura 1985). Over the last 40 years, no other parameter used in QSAR has generated more interest, excitement, and controversy than hydrophobicity (Hildebrand 1979). Hydrophobic interactions are of critical importance in many areas of chemistry. These include enzyme- ligand interactions, the assembly of lipids in bio membranes, aggregation of surfactants, coagulation, and detergency (Rose t1985, Schneider 1991, Nusselder 1991). The integrity of bio membranes and the tertiary structure of proteins in solution are determined by polar-type interactions. Hydrophobicities of solutes can readily be determined by measuring partition coefficients designated as P. Partition coefficients deal with neutral species, whereas distribution ratios incorporate concentrations of charged and/or polymeric species as well. By convention, P is defined as the ratio of concentration of the solute in octanol to its concentration in water.

$$P = [\text{conc.}]_{\text{octanol}} / [\text{conc.}]_{\text{aqueous}} \quad \text{.....(2)}$$

Generally partition coefficient is used in logarithmic form, Log P. It's value ranging from 0 to 4. (Log P equals zero means that, the solute is equally soluble in two phases). Negative value of Log P represents more solubility of solute in water, while the positive value of Log P represents greater solubility in the organic phase. Octanol is a suitable solvent for the

measurement of partition coefficients for many reasons (Lippold 1972, Leo 1971). It is cheap, relatively nontoxic and chemically unreactive. The hydroxyl group has both hydrogen bond acceptor and hydrogen bond donor features capable of interacting with a large variety of polar groups. Despite its hydrophobic attributes, it is able to dissolve many more organic compounds than can alkanes, cycloalkanes, or aromatic hydrocarbons. It is UV transparent over a large range and has a vapor pressure low enough to allow for reproducible measurements. It is also elevated enough to allow for its removal under mild conditions. In addition, water saturated with octanol contains only 1.023 M octanol at equilibrium, whereas octanol saturated with water contains 2.3 M of water. Thus, polar groups need not be totally dehydrated in transfer from the aqueous phase to the organic phase. Likewise, hydrophobic solutes are not appreciably solvated by the 1023 M octanol in the water phase unless their intrinsic log P is above 6.0. Octanol begins to absorb light below 220 nm and thus solute concentration determinations can be monitored by UV spectroscopy. More important, octanol acts as an excellent mimic for bio membranes, because it shares the traits of amphiphilicity and hydrogen-bonding capability with phospholipids and proteins found in biological membranes. The choice of the octanol/water partitioning system as a standard reference for assessing the compartmental distribution of molecules of biological interest was recently investigated by molecular dynamics simulations (Debolt 1995).

It was determined that pure 1-octanol contains a mix of hydrogen-bonded "polymeric" species, mostly four-, five-, and six-membered ring clusters at 40^0C. These small ring clusters form a central hydroxyl core from which their corresponding alkyl chains radiate outward. On the other hand, water-saturated octanol tends to form well-depended, inverted, micellar

aggregates. Long hydrogen- bonded chains are absent and water molecules congregate around the octanol hydroxyls with the alkyl chains extending outward. Thus, water-saturated octanol has centralized polar cores where polar solutes can localize. Hydrophobic solutes would migrate to the alkyl-rich regions. This is an elegant study that provides insight into the partitioning of benzene and phenol by analyzing the structure of the octanol/water solvation shell and delineating octanol's capability to serve as a surrogate for bio membranes. The shake-flask method, so-called, is most commonly used to measure partition coefficients with great accuracy and precision and with a log P range that extends from 23 to 16 (Leo1991 Klopman1994). The procedure calls for the use of pure, distilled, de-ionized water, high-purity octanol, and pure solutes. At least three concentration levels of solute should be analyzed and the volumes of octanol and water should be varied according to a rough estimate of the log P value. Care should be exercised to ensure that the eventual amounts of the solute in each phase are about the same after equilibrium. Standard concentration curves using three to four known concentrations in water saturated with octanol are usually established. Generally, most methods employ a UV based procedure, although GC and HPLC may also be used to quantitate the concentration of the solute.

Values of Log P can be determined experimentally, using classical shake flask method, filter probe methods and different centrifugal partition chromatographic techniques. Chromatographic methods have frequently been applied in the determination of relative partition coefficients (P) for use in quantitative relationships between chemical structure and biological activity (QSAR) (Bate-Smith 1950). Reverse phase thin layer chromatography and high performance liquid chromatography are used to

evaluate the lipophilicity of a series of organic compounds. These correlations are based on the studies of Dearden (Dearden 1985), Consdem, Martin and Synge on the relationship between chromatographic distribution and the partition coefficient (Maria 2002). Several log *P* calculation methods from chemical structure have been developed. The methods fall into two classes: the group contribution approach; and the whole molecule approach. The group contribution approach includes "atom-based" and "fragment-based" methods, in which a molecule is divided into atoms or fragments and the log *P* values are calculated by summation of the contributions from the fragments or atoms present in the molecule. The whole molecule approach is based on molecular properties such as electrostatic potential, molecular surface area, and molecular volume. Hansch and Fujita (Fujita 1964) developed the first model for calculating log *P* in 1964 based on the π-system. The limitations of the π-system led Rekker to develop the first "fragment-based" methods (Rekker 1977), and later Broto reported the first "atom based" contribution methods for calculating the log *P* values. Since then several other calculation methods have been proposed (Broto). Mannhold listed the addresses of the programs for the calculation of log *P* and they reported a comparative study of log *P* calculation methods (Mannhold 1996). Various computational methods and software's packages are used for calculation of Log P for the octanol/water system e.g. ACD/Log P. Out of 7450 QSAR currently in the Pomona database, the large contain either the term Log P (4150 data) or the hydrophobic parameter for substituent's π (814 data).

1.2.4 Steric parameters

The quantitation of steric effects is complex at best and challenging in all other situations, particularly at the molecular level. An added level of confusion comes into play when attempts are made to delineate size and shape. Nevertheless, steric is of overwhelming importance in ligand-receptor interactions as well as in transport phenomena in cellular systems. The first steric parameter to be quantized and used in QSAR studies was Taft's E_S constant (Taft 1956), ES is defined as;

$$E_S = \log (k_X / k_H)_A \quad \ldots\ldots\ldots\ldots (3)$$

Where k_X and k_H represent the rates of acid hydrolysis of esters XCH_2COOR and CH_3COOR, respectively. One of the most widely used steric parameters is molar refraction (MR), which has been aptly described as a "chameleon" parameter by Tute (Tute 1990). Although it is generally considered to be a crude measure of overall bulk, it does incorporate a polarizability component that may describe cohesion and is related to London dispersion forces as follows: $MR = 4\pi N\alpha / 3$, where N is Avogadro's number and α is the polarizability of the molecule. It contains no information on shape. MR is also defined by the Lorentz-Lorenz equation:

$$MR = \frac{(n^2-1)\ X\ (Mw)}{(n^2+1)\ X\ \text{density}} \quad \ldots\ldots\ldots\ldots (4)$$

MR is generally scaled by 0.1 and used in biological QSAR, where intermolecular effects are of primary importance. The refractive index of the molecule is represented by n. With alkyl substituents, there is a high degree of co-linearity with hydrophobicity; hence, care must be taken in the QSAR analysis of such derivatives. The coefficients with MR terms challenge

interpretation, although extensive experience with this parameter suggests that a negative coefficient implies steric hindrance at that site and a positive coefficient attests to either dipolar interactions in that vicinity or anchoring of a ligand in an opportune position for interaction. Molecular weight (MW) terms have also been used as descriptors, particularly in cellular systems, or in distribution/transport studies where diffusion is the mode of operation. According to the Einstein-Sutherland equation, molecular weight affects the diffusion rate. The Log MW term has been used extensively in some studies (Taft 1956, Tute 1990) Molar refractivity has a broad application in QSAR studies as a physicochemical parameter related to volume and size of a substituent. The larger the polar part of a molecule, the larger is its MR value; The ACD/lab software contains a routine for calculation of MR and MV.

1.2.5 Parameters selection

The selection of parameters is an important first step in any QSAR study. QSARs are developed using a variety of parameters as descriptors of the structural properties of molecules. Hammett sigma values are often used for electronic parameters, but quantum mechanically derived electronic parameters such as total energy (E_t), dipole moment (μ) and heat of formation (ΔH_f) may also be used. Other electronic parameters e.g. charge transfer constants, hydrogen-bonding parameters and parameters derived from molecular spectroscopy describing the influence of substituent on electron density distribution, all have been used in QSAR studies (Kubinyi 1977).

1.2.6 Qsar classical parameter calculation methods

1.2.6.1 Quantum mechanics (QM)

Quantum mechanical methods find increasing use in the field of QSAR as a tool to derive physiochemical parameters for drug molecules to be correlated with their biological response data. It is thereof redefendable to include a short expose about these methods herein.

1.2.6.1.1 Schrödinger equation

In 1926, Erwin Schrödinger discovered the basic principles of a new kind of mechanics which provided mathematical techniques competent to deal with the wave-particle duality nature of matter and energy. The Schrödinger equation expressed as

$$\hat{H}\psi = E\psi, \ldots\ldots\ldots\ldots\ldots (4)$$

Where $\hat{H}$, the Hamiltonian operator; which is given in terms of location, mass and potential energy of an electron,. ψ is the wave function which gives the probability of finding an electron at certain region in the space around the atomic nucleus and E stands for total energy. The simplest application of Schrödinger equation is the case of a free particle, but it can be extended to multi electron system; where no exact solution could be given owing to various interactions that happen between electrons with each other and with core particles in atomic nuclei. Approximate solutions had however being given by neglecting some interactions in what is called Hartree-Fock method (Žofka 1970).

1.2.6.1.2 Self consisten field (SCF; Hartree-Fock calculations)

This type of calculations are based on a few principles and do not employ any experimental data. The important concept is to pick one electron and to approximate the interaction between the single and all other electrons by a mean value that is determined from their probability densities. Each result from one calculation is used in subsequent calculation until no change is detected in electrostatic potential; so it is called the self-consistent method (Moore 1972) This method ignores the correlated movement of the electrons as they avoid each other, because of their equal electric charges. This error is called "electron correlation". Despite this deficiency, Hartree-Fock calculations give more accurate picture and are not limited to certain classes of chemical compounds. But, they need considerable computer time. The methods employing Hartree-Fock approach may be divided into two types; *Ab initio* and semi-empirical methods (Sharon H1993).

1.2.6.1.3 *Ab initio* methods

This method; which literally means from the beginning; place a prohibitively high demand upon computing power and memory to be generally useful for the routine modeling of all simplest molecules.

1.2.6.1.4 Semi empirical quantum mechanical method QMM

Semi-empirical quantum chemistry attempts has two limitations, one of them is the slow speed and the other is the low accuracy of the Hartree-Fock calculation; By omitting or parameterizing certain integrals based on experimental data, such as ionization energies of atoms, or dipole moments of molecules. However, semi-empirical methods are very fast, applicable to large molecules, and may give accurate results when applied to the chemical

structure of molecules that are similar to the molecules used for parameterization. Modern models are based on the neglect of diatomic differential overlap (NDDO); in which the overlap matrix ***S*** is replaced by the unit matrix. This allows one to replace the Hartree-Fock secular equation:

$$[H-ES] = 0, \quad (6)$$

With a simpler equation | H-E] = 0, existing semi-empirical models differ by the further approximations that are made when evaluating one and two-electron integrals and by paramertization philosophy.

1.2.6.1.5 Modified-neglect of diatomic overlap (MNDO)

This method was discovered by Michael Dewar and Walter Thiel (Dewar 1977). It is the oldest NDDO-model that parameterizes one-center two-electron integrals depending on spectroscopic data for isolated atoms; and evaluates other two-electron integrals using the idea of multiple-multiple interactions from classical electrostatics. Classical MNDO model uses only *s* and *p* orbital basis sets, while more recent MNDO/d adds *d*-orbitals that are especially important for the description of hypervalent sulfur species and transition metals. MNDO has a number of disadvantages, such as, inability to describe the hydrogen bond, due to a strong intermolecular repulsion. The other disadvantages of MNDO method is the poor reliability in predicting heats of formation. For example highly substituted stereo isomers are predicted to be too unstable compared to linear isomers due to overestimation of repulsion in sterically crowded systems. (Dewar 1985).

1.2.6.1.6 Austin model 1(AM_1):

AM_1 Method takes a similar approach to MNDO in approximating two-electron integrals but uses a modified expression for nuclear-nuclear core repulsion. The modified expression results in non-physical attractive forces that minimize Van der Waals interactions. The modification also necessitated re-parameterization of the model, which was carried out with a particular emphasis on dipole moments, ionization potentials, and geometries of the molecules. While this allows for some description of the hydrogen bond, other deficiencies, such as systematic over-estimates of basicity, remained. However, the lowest energy geometry for the water dimer is predicted incorrectly by the AM_1 model. On the other hand, AM_1 improves nicely some properties, such as heats of formation, when compared to MNDO model (Dewar 1977).

1.2.6.1.7 Parametric method (PM_3)

This method is very similar to the AM_1-Hamiltonian, but the parameterization strategy is different. While AM_1 was parameterized largely, based on a small number of atomic data, PM_3 is parameterized to reproduce a large number of molecular properties. In some sense, chemistry gave way to statistics with the PM_3 model. Different parameterization, and slightly different treatment of nuclear repulsion allow PM_3 to treat hydrogen bonds rather well, but it amplifies non-physical hydrogen-hydrogen attractions in other cases. This results serious problems when analyzing intermolecular interactions, (e.g. methane is predicted to be a strongly-bound dimer) or conformations of flexible molecules (OH is strongly attracted to CH_3 in 1-pentanol). The accuracy of thermochemical prediction (calculation) with PM_3 is slightly better than that of AM_1. The PM_3 model has been

widely used for rapid estimation of molecular properties and has been recently extended to include many elements, including some transition metals (Stewartand James 1989).

1.2.6.1.8 Accuracy and applications of semi empirical methods

The semi empirical models MNDO, AM_1, and PM_3 are available in many computational chemistry programs such as Argus Lab. They are often used in computational chemistry, because they allow study of systems that are out of reach of more accurate methods. For example, modern semi-empirical programs allow study of molecules consisting of 1000 of atoms while *Ab initio* calculations that produce similar thermo chemical accuracy are feasible on molecules consisting of less than 50-70 atoms. Semi-empirical calculations can also be useful in many situations, such as computational modeling of structure-activity relationships to gain insight about reactivity or property trends for a group of similar compounds, design of chemical synthesis, development and testing of new methodologies, for example development of hybrid quantum mechanics/molecular mechanics (QM/MM) methods for modeling of biochemical processes, design of chemical synthesis, preliminary optimization of geometries of unusual molecules and transition states that cannot be optimized with molecular mechanics methods, in many applications where qualitative insight about electronic structure and properties is sufficient.

The limitations of semi-empirical methods should be kept in mind when deciding to use as tools for the task of calculations. For large systems, either molecular mechanics MM or semi-empirical quantum mechanics, could be used for optimization and calculation of conformational energies. MM approach is faster and in most cases produces more accurate

conformational energies and geometries. Some molecular mechanics methods, such as MM_3 and MM_4, can also well predict thermo chemistry of stable species. On the other hand, if there is no suitable force field for the system, semi-empirical methods may be the only tool choice. For small systems, the compromise must be made between the semi-empirical approach and the more reliable ones, but much more time-consuming ab initio calculations. In general, semi-empirical results can be trusted only in situations when they are known to work well (e.g. systems similar to molecules in the parameterization set) and strong caution should be exercised in cases where semi-empirical methods are known to fail (e.g. prediction of activation barriers). Finally, it is not clear to assume that, the semi-empirical methods can be used to model many larger systems.

1.2.5 Variable selection methods

In order to develop regression/classification models, QSAR analysis typically uses molecular descriptors as independent variables. The number of molecular descriptors has hugely increased over time and nowadays thousands of descriptors, able to describe different aspects of a molecule, can be calculated by means of dedicated software. However, when modeling a particular property or biological activity, it is reasonable to assume that only a small number of descriptors is actually correlated to the experimental response and is, therefore, relevant for building the mathematical model of interest.

As a consequence, a key step is the selection of the optimal subset of variables (i.e. molecular descriptors) for the development of the model. This is precisely the aim of the so-called variable selection methods, which allow improving interpretability (simple models), neglecting not significant

effects, thus reducing noise, increase the model predictive ability and speed up modeling time.

During the years different variable selection methods have been proposed, from relatively simple to more recent ones that took inspiration from different scientific fields, like genetics and ethnology. Furthermore, some methods able to perform both regression and variable selection simultaneously have recently been proposed. There are so many models and here are some.

1.2.5.1 All subset models (ASM)

The All Subset Models (ASM) method is the simplest and computationally consuming. It consists in the generation of all the possible combinations of the *p* variables, from size 1 to p, p being the total number of variables. This method guarantees that the best subset of variables is found, but it's very computationally consuming, being the total number of combinations of *p* variables given by:

$$2^{p} - 1 \ldots\ldots\ldots\ldots\ldots\ldots (7)$$

As a consequence, the method becomes unsuitable for large numbers of variables. If one is interested in developing simple models.

1.2.5.2 Sequential search (SS)

Sequential Search (SS) is a simple method aimed at finding optimal subsets of variables for a specified model size. The basic idea is to replace each variable at a time with all the remaining ones and see whether a better model is obtained. This procedure differs from the All Subset Models method in that in this case not all the possible combinations of the *p*

variables are tested, the method thus being less time consuming and meta-heuristic.

1.2.5.3 Stepwise methods (SW)

Stepwise regression methods are among the most known subset selection methods, although currently quite out of fashion. Step Wise regression is based on two different strategies, namely Forward Selection (FS) and Backward Elimination (BE). Forward Selection method starts with a model of size 0 and proceeds by adding variables that fulfill a defined criterion. Backward Elimination method proceeds in the opposite way, in that it starts from a model of size *p*, *p* being the total number of variables, and eliminates not relevant variables in a step by step procedure (Miller Hastie 2002).

1.2.5.4 Feature selection, regression analysis procedures

Feature subset selection is the process of identifying and removing from a training data set as much irrelevant and redundant features as possible. This reduces the dimensionality of the data and may enable regression algorithms to operate faster and more effectively. In some cases, correlation coefficient can be improved; in others, the result is a more compact, easily interpreted representation of the target concept.

1.3 Current mathematical methods used in QSAR studies

Mathematical regression methods are so important for the QSAR modeling that the choice of the regression method, most of the time, will determine if the resulted model will be successful or not. Fortunately, more and more new methods and algorithms have been applied to the studies of QSAR, including linear and nonlinear, statistics and machine learning. At the same time, the existing methods have been improved. However, it is still

a challenge for the researchers to choose suitable methods for modeling their systems.

The main mathematical methods applied to the regression of QASR models; gene expression programming (GEP), Project Pursuit Regression (PPR) and Local Lazy Regression (LLR) have appeared on the QASR stage. At the same time, the earlier methods, including Multiple Linear Regression (MLR), Partial Least Squares (PLS), Neural Networks (NN), Support Vector Machine (SVM) and so on, are being upgraded to improve their performance in QASR studies. In light of this, these methods are divided as follows: first, the improved methods based on the earlier ones are described and then the new methods are presented.

1.3.1 Multiple linear regressions (MLR)

A statistical tool allows one to examine how multiple independent variables are related to a dependent variable. Once these multiple variables are identified how they relate to the dependent variable, the information about all of the independent variables can be taken and used to make much more powerful and accurate predictions about why things are the way they are. This latter process is called “Multiple regressions”. It is one of the earliest methods used for constructing QSAR/QSPR models, but it is still one of the most commonly used ones to date. The advantage of MLR is its simple form and easily interpretable mathematical expression. Although utilized to great effect, MLR is vulnerable to descriptors which are correlated to one another, making it incapable of deciding which correlated sets may be more significant to the model. Some new methodologies based on MLR have been developed and reported in recent papers aimed at improving this technique. These methods include Best Multiple Linear Regression (BMLR), Heuristic Method (HM) (Karitizky 2001) and Genetic

Algorithm based Multiple Linear Regression (GA-MLR) (Gharagheizi 2009), Stepwise MLR, Factor Analysis MLR and so on. (Du 2008).

1.3.2 Partial least squares (PLS)

The basic concept of PLS regression was originally developed by Wold (Word 1966). As a popular and pragmatic methodology, PLS is used extensively in various fields. In the field of QSAR/QSPR, PLS is famous for its application to comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis CoMSIA. Recently, PLS has evolved by combination with other mathematical methods to give better performance in QSAR analyses. These evolved PLS', such as genetic partial least squares (G/PLS) (Davies MN.), factor analysis partial least squares (FA-PLS) (Mandal) and orthogonal signal correction partial least squares (OSC-PLS) (Priolo 2001).

1.3.3 Neural networks (NN)

As an alternative to the fitting of data to an equation and reporting the coefficients derived there from, neural networks are designed to process input information and generate hidden models of the relationships. One advantage of neural networks is that they are naturally capable of modeling nonlinear systems (Vanyur 2002). Disadvantages include a tendency to over fit the data, and a significant level of difficulty in ascertaining which descriptors are most significant in the resulting model. In the recent QSAR studies, radial basis function neural network (RBFNN) (Xia BB 2008, Xia BB2009) , general regression neural network (GRNN) (Specht 1991 and Szaleniec 2008) were used.

1.3.4 Support vector machine (SVM)

SVM, developed by Vapnik (Cortes 1995) as a novel type of machine learning method, is gaining popularity due to its many attractive features and promising empirical performance. Originally, SVM was developed for pattern recognition problems. After that, SVM was applied to regression by the introduction of an alternative loss function and results appear to be very encouraging (Wang 2003). As a developing method, new types of SVM are coming in on the stage of QSAR, such as least square support vector machine (LS-SVM) (Yuan 2008, Liu 2007, Liu 2005), grid search support vector machine (GS-SVM), potential support vector machine (P-SVM) and genetic algorithms support vector machine (GA-SVM).

1.3.5 Gene expression programming (GEP)

Gene expression programming was invented by Ferreira in 1999 and was developed from genetic algorithms and genetic programming (GP). GEP uses the same kind of diagram representation of GP, but the entities evolved by GEP (expression trees) are the expression of a genome. GEP is more simple than cellular gene progression. It mainly includes two sides: the chromosomes and the expression trees (ETs). The process of information of gene code and translation is very simple, such as a one-to-one relationship between the symbols of the chromosome and the functions or terminals they represent. The rules of GEP determine the spatial organization of the functions and terminals in the ETs and the type of interaction between sub-ETs. Therefore, the language of the genes and the ETs represents the language of GEP. GEP is the newest Chemometrics method, and Si et al (Wang 2008 Li X2007) have applied this method to QSAR studies for the first time. The results from their studies are satisfactory and show a

promising use in the nonlinear structure-activity/property relationship correlation area, but GEP is congenitally defective as far as reproducibility of the predicted results is concerned, and always deduces too complex equations. This means a higher requirement for a user who is involved with GEP.

1.3.6 Project pursuit regression (PPR)

Project pursuit regression, which was developed by Friedman and Stuetzle (Friedman) is a powerful tool for seeking the interesting projections from high-dimensional data into lower dimensional space by means of linear projections. Therefore, it can overcome the curse of dimensionality because it relies on estimation in at most trivariate settings. Friedman and Stuetzle's concept of PPR avoided many difficulties experienced with other existing nonparametric regression procedures. It does not split the predictor space into two regions, thereby allowing, when necessary, more complex models. In addition, interactions of predictor variables are directly considered because linear combinations of the predictors are modeled with general smooth functions. Another significant property of PPR is that the results of each interaction can be depicted graphically. The graphical output can be used to modify the major parameters of the procedure: the average smoother bandwidth and the terminal threshold. In recent QSAR studies (Ren 2007) PPR was employed as a regression method and always resulted in the best final models. This indicates that PPR is a promising regression method in QSAR studies, especially when the correlation between descriptors and activities or properties is nonlinear.

1.3.7 Local lazy regression (LLR)

Most QSAR models often capture the global structure-activity trends which are present in a whole dataset. In many cases, there may be groups of molecules which exhibit a specific set of features which relate to their activity or property. Such a major feature can be said to represent a local structure activity relationship. Traditional models may not recognize such local relationships. LLR is an excellent approach which extracts a prediction by locally interpolating the neighboring examples of the query which are considered relevant according to a distance measure, rather than considering the whole dataset. That will cause the basic core of this approach which is a simple assumption that similar compounds have similar activities or properties; that is, the activities or properties of molecules will change concurrently with the changes in the chemical structure. However, as a new arising method in the field of QSAR, LLR is not used extensively, with only a few relevant studies (Du 2008 and Guha 2006). It is expected that more application studies involving LLR will appear in future QSAR analyses.

1.4 Applications of QSAR

QSAR has been applied extensively and successfully over several decades to find predictive models for activity of bioactive agents. It has also been applied to areas related to discovery and subsequent development of bioactive agents: toxicity prediction (Bashir 2000), physicochemical properties prediction (Gombar 1996) and recent reviews (Karitizky 2002) summarize work in a number of these areas. The Journal of quantitative structure activity relationships contains abstracts of QSAR studies in other journals in each issue.

1.5 Anthrapyrazole anticancer Agent

Anthraquinone, also called anthracenedione or dioxoanthracene is a compound having the formula $C_{14}H_8O_2$, molar mass: 208.21 g mol−1, melting point 286 °C, boiling point 379.8°C that can be viewed as a diketone derivative of anthracene, with loss of two double bonds in the latter. The term usually refers to one specific isomer, 9,10-anthraquinone or 9,10-dioxoanthracene, whose ketone groups are on the central ring. This compound is an important member of the Quinone family (Fig 1.1). Several other anthraquinone isomers are 1,2-, 1,4-, and 2,6-anthraquinone, but they are of comparatively minor importance.

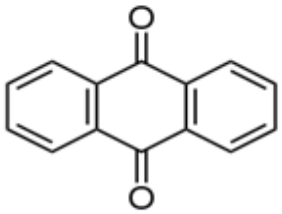


Fig (1.1): 9, 10-anthraquinone

9, 10-Anthraquinone can be synthetized industrially by the oxidation of anthracene, a reaction that is localized at the central ring. Chromium (VI) is the typical oxidant. It is also prepared by the Friedel-Crafts reaction of benzene and phthalic anhydride in presence of $AlCl_3$. The resulting o-benzoylbenzoic acid then undergoes cyclization, forming anthraquinone. It is also prepared by Diels-Alder reaction by the reaction of naphthoquinone and butadiene followed by oxidative dehydrogenation. Lastly, Basf has developed a process that proceeds via the acid-catalyzed dimerization of styrene to give a 1, 3-diphenylbutene, which then can be transformed to the anthraquinone.

1.5.1 Applications and natural occurrence of quinones

Synthetic dyes are often derived from 9,10-anthraquinone, such as alizarin. Important derivatives are 1-nitroanthraquinone, anthraquinone-1-sulfonic acid, and the dinitro anthraquinone. Natural pigments that are derivatives of anthraquinone are found in *Aloe latex*, senna, rhubarb, and *Cascara buckthorn*, fungi, lichens, and some insects.

9,10-Anthraquinone is used as a catalyst in production of an alkaline pulping by the Kraft and alkaline sulfite processes. The anthraquinone goes through a redox cycle and it gives a catalytic effect. The anthraquinone oxidizes cellulose and thereby protects it from alkaline degradation. It is also reduced to hydro anthraquinone which then can react with lignin. The lignin is degraded and becomes more water-soluble and the anthraquinone is regenerated. A large industrial application of anthraquinone is the production of peroxide, where 2-Ethyl- 9, 10-anthraquinone or related alkyl derivatives is used, rather anthraquinone itself. Derivatives of 9,10-anthraquinone, include many important drugs. Their medical uses are as laxatives like Dantron, Emodin, Aloe Emodin and some of the Senna glycosides. Others are anti malaria's like Rufigallol or Antineoplastic.

1.5.2 Clinical importance of anthraquinone

A cytotoxic anti leukemic agent with anthraquinone structure has been isolated from Morindaparvifolia.

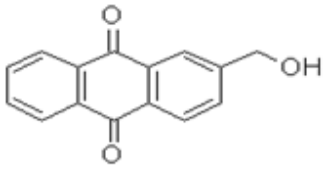


Fig 1.2, a: anthracene-9, 10-diones

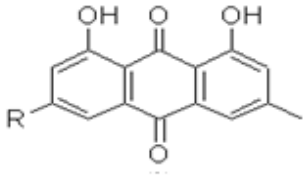


Fig 1.2, b: 1, 5-dihydroxy anthracene-9, 10-diones

The cytotoxic 1, 8-dihydroxy substituted anthracene-9, 10-diones (1.2, a) (R= β-D-glucose) has been isolated from the roots of an Indonesian medicinal plant Kelembak. The cytotoxic 1,5-dihydroxy analogue (Fig 1.2, b) have been isolated from Cassia Italica found in Pakistan. They are most used in the treatment of cancer, like Mitoxantrone (fig 1.3) and Pixantrone (fig 1.4).

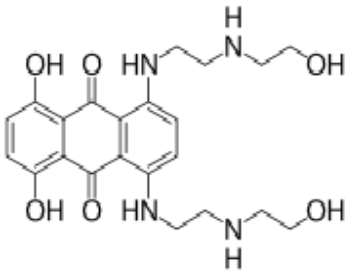


Fig 1.3: Mitoxantrone

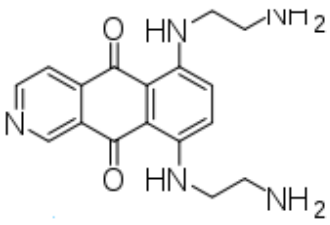


Fig 1.4: Pixantrone

Mitoxantrone (anthracene-9,10-diones analogue); which was developed as a potential blue ballpoint ink in USA, has been used for the treatment of breast cancer, acute leukemia and as a good anti neoplastic

agent. Daunorubicin (fig 1.5) and doxorubicin (fig 1.6) have well established rules in the treatment of human cancer.

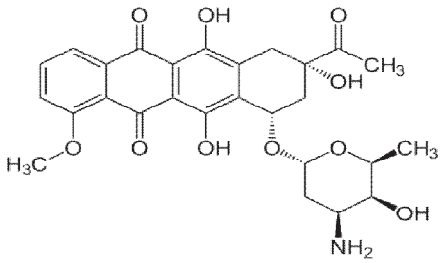


Fig 1.5: Daunorubicin

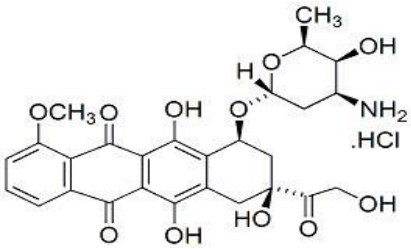


Fig 1.6: doxorubicin

In recent years, intensive research efforts were directed toward finding new anthracycline analogues with better efficacy and less toxicity than daunorubicin and doxorubicin. In this regard, Farmorubicin (fig 1.7) is reported to have lower cumulative cardio toxicity than doxorubicin. Biosynthesis yielded over 300 new compounds whereas more than 2000 analogues were obtained via semi-synthesis or total synthesis. Carminomycin (fig 1.8) is a 4-hydroxy analogue of daunorubicin produced by Actinomaduracarminata.

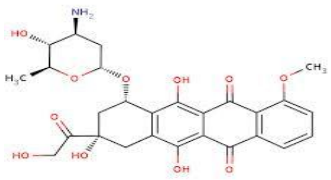


Fig 1.7: Farmorubicin

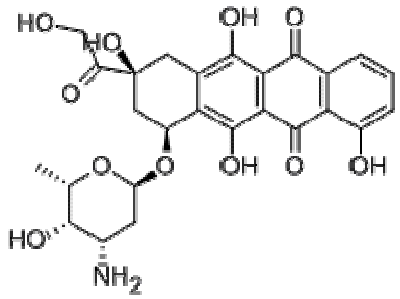


Fig 1.8: Carminomycin

It was originally developed in the Soviet Union in 1980s and once marketed in Europe. A clacinomycin A (fig 1.9) (Aclarubicin) is the only representative of class II anthracycline which has been once marketed in Europe and Japan for treatment of leukemia. It is believed to interfere with the RNA synthesis more than DNA synthesis. It also lacks the cardiotoxicity shown by daunorubicin and doxorubicin.

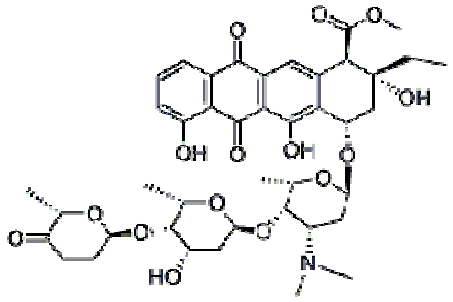


Fig 1.9: Carminomycin

Another anthracycline of significant anti-tumor activity is Nogalamycin which was isolated and characterized by Wiley et al. from S. nogalater. It differs from other anthracyclines in that the amino-sugar is joined to the nucleus by carbon-carbon bond and a cyclic acetal linkage in ring D. The usual place of attachment of the sugar residue at C-7 is, however occupied by a non-amino sugar nogalose. Nogalamycin, (fig 1.10) with its ring-D bicyclic sugar substituent actually intercalates with nucleic acids, and

the DNA bases buckle even more than other anthracyclines. This has been proved by X-ray analysis of drug-DNA hexamers.

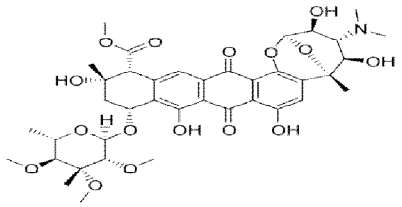


Fig 1.10 : Nogalamycin

1.5.3 Anthrapyrazoles and their Aza-bioisosteres

In medicinal chemistry, bioisosteres are substituents or groups with similar physical or chemical properties which produce broadly similar biological properties to a chemical compound. In drug design,(Nathan B) the purpose of exchanging one bioisostere for another is to enhance the desired biological or physical properties of a compound without making significant changes in chemical structure. The main use of this term and its techniques are related to pharmaceutical sciences. Bioisosterism is used to reduce toxicity or modify the activity of the lead compound, and may alter the metabolism of the lead. Several studies have suggested that anthra cycline cardio toxicity may be associated with the formation of reactive oxygen species and subsequent intracellular lipid peroxidation from enzymatic reduction of the quinine chromophore to a semiquinone radical species.

In an attempt to provide agents with diminished or no cardio toxicity, chromophore- modified anthraquinone (anthracenediones) related to mitoxantrone, the anthrapyrazoles (anthra [1,9-cd]pyrazol-6(2H)-one) ring system (fig 1.11) thus became an initial target. In these compounds, not only is the central quinone moiety being modified to quasi-iminoquinone but also, in contrast to the chromophore- modified anthracenediones, another ring was

incorporated into the chromophore. Several members of the above class of compounds were synthesized and biologically evaluated.

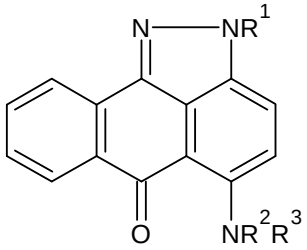


Fig 1.11: 1, 9-cd] pyrazol-6(2H)-one

Losoxantrone [(Fig 1.12) emerged as the most promising candidate, it has demonstrated good clinical efficacy in treatment of breast cancer but cardio toxicity was still observed during the clinical trials.

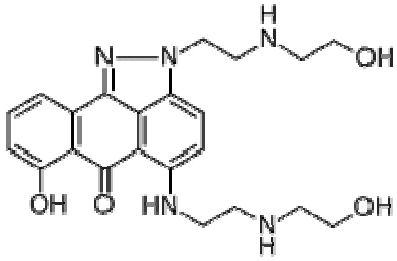


Fig 1.12: Losoxantrone

Aza-analogues of anthrapyrazoles synthesized by incorporation of N-atoms either in position 5 or position 9 of the anthrapyrazole ring. The 9-aza-anthrapyrazoles, built in anticipation of improving the properties of their carbocyclic congeners, seem to be very promising as antitumor agent with no cardio toxicity.

In vitro evaluation of these compounds against both systemic P388 murine leukemia and human mammary carcinoma, demonstrated high cytotoxic potency and the ability to overcome multidrug resistance induced in a doxorubicin-resistant cell line. Of these, two derivatives BBR 3434 (fig 1.12) and BBR 3576 (fig 1.14) have surfaced as potential clinical candidates.

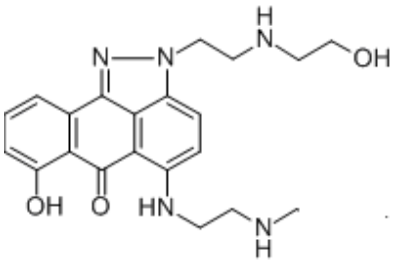


Fig 1.13: BBR 3434

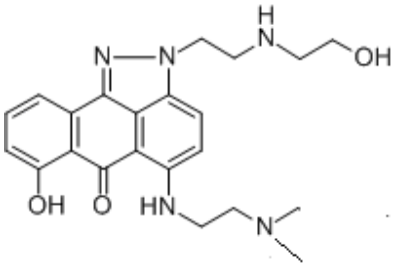


Fig 1.14: BBR 3576

Both of the above compounds could be considered as bioisosteric analogues of Losoxantrone.

1.5.4 Mode of action of anthrapyrazoles anticancer agents

Anthrapyrazoles have been investigated as cancer chemotherapeutic agents. The mechanism of action of these compounds is thought to involve inhibition of DNA topoisomerase II. A structure-activity study was carried out to determine the *in vitro* cytotoxic activity of nine novel anthrapyrazoles against human breast carcinoma, head and neck squamous cell carcinoma and leukemia cells, and against Chinese hamster ovary cells. The activity of these anthrapyrazole analogues was compared with that of two clinically tested anthrapyrazoles, losoxantrone and piroxantrone. Inhibition of topoisomerase II as a mechanism of action for the analogues was also investigated. The cytotoxic activity of the analogues was determined in vitro by MTT cell growth inhibition assay and inhibition of catalytic topoisomerase II activity by each compound was measured using a

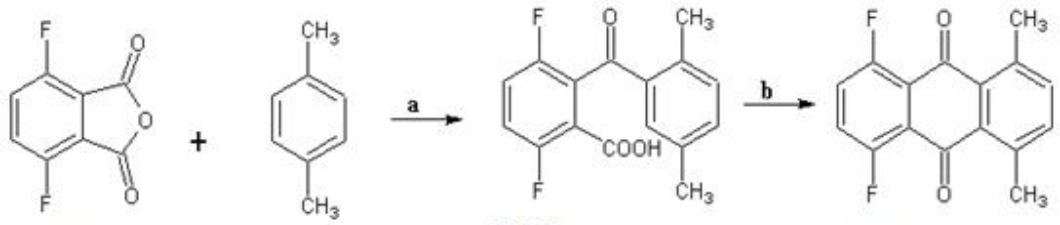
fluorometric DNA decatenation assay. All of the anthrapyrazole analogues inhibited the growth of the four cell lines with IC_{50} values that ranged from 0.1 to 45.2 μ M. Losoxantrone was the most potent of the anthrapyrazole analogues studied. A tertiary amine in the basic side chain at N_2 increased the cytotoxic activity compared with a secondary amine in this side chain for many of the analogues, but not if there was a basic side chain at the C_5 position. A chlorine substituent on the basic side chain at N_2 did not have a consistent effect on activity. Moving the position of a chlorine substituent from C_5 to C_7 or introducing a basic side chain at C_5 did not have a consistent effect on cytotoxic activity. Anthrapyrazole analogues showed a broad range of activity for inhibiting topoisomerase II decatenation activity. Losoxantrone and Pixantrone were the most potent inhibitors of topoisomerase II activity. QSAR correlation and 3D-QSAR analyses of 13 anthrapyrazole analogues of Losoxantrone and Pixantrone, showed the importance of anthrapyrazole-DNA van der Waals interactions, while 3D-QSAR CoMFA showed that hydrogen-bond donor interactions and electrostatic interactions with the protonated amino side chains of the anthrapyrazoles led to high cell growth inhibitory activity

1.5.5 Main synthetic pathways of anthracenediones

1.5.5.1 Friedle-Craft acylation

A great deal of effort has been expended in the field of anthracenediones synthesis. Various reactions and synthetic routes were tried. This reaction has been extensively used to construct the chromophore part of various anthracenedione derivatives. For instance, 1, 4-diflourophthalic anhydride (1.15) reacted with p-xylene to give the

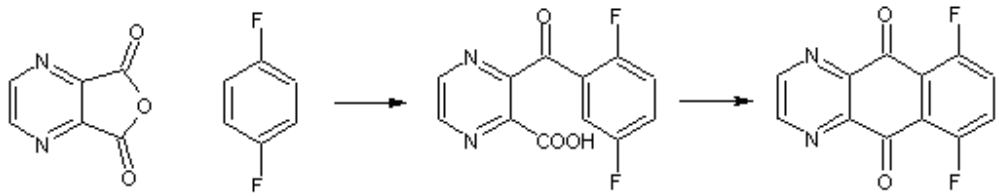
substituted anthracenediones (1.16), which was used in the synthesis of the Mitoxantrone analogue (1.17). (Scheme, 1. 1)



1.15 **1.16** **1.17**

Scheme 1-1

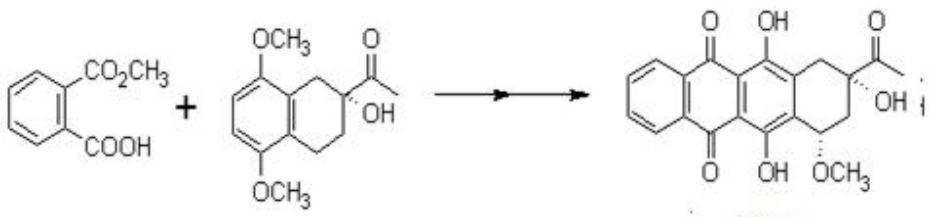
This reaction also successfully applied to diaza-anthracenediones to synthesize the precursor (1.18) used to obtain some aza-bioisosteres of amentantrone (scheme, 1- 2) (Krapcho A.P., *et al*, 1993).



1.18

Scheme 1-2

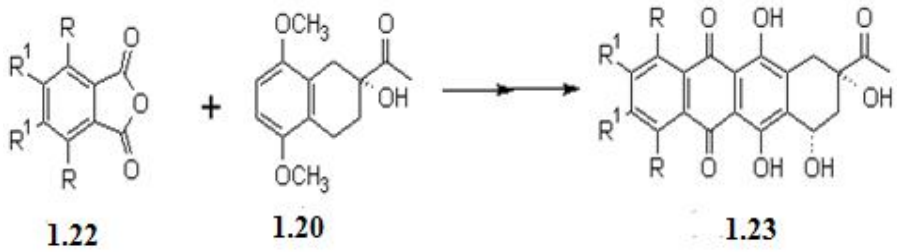
In the anthracycline series this reaction was used to obtain various precursors useful in constructing tetracyclic backbone. 7-O-methyl-4-demethoxydaunomycinone (1.21) was first synthesized by Wong et al. through the reaction of phthalic acid half ester (1.19). DC-building block] and AB- building block](1.20) to give (1.21) in a stepwise Friedel-Craft reaction, scheme 1-3A (Wong 1971),



1.19 **1.20** **1.21**

Scheme 1-3A

The Friedel-Craft reaction with symmetrically substituted phthalic anhydrides (1.22) can also occur in one-pot procedure. 4-Dimethoxydaunomycinone , R =R` = H] is now manufactured in large scale using this way, in addition to other derivatives, scheme1-3B (Arcamone, F., *et al*, 1987).



1.22 **1.20** **1.23**

Scheme 1-3B

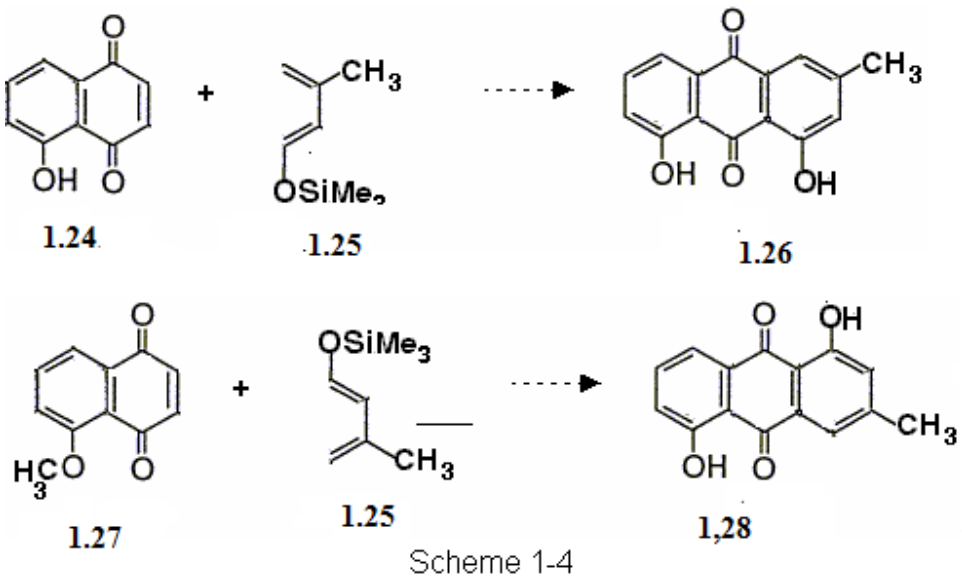
Basically, the acylation using phthalic anhydride derivatives proceed in two steps whereby the first deactivate the arene for a second electrophilic attack; this hold even if in practice one pot procedures (e.g. $AlCl_3$/ NaCl melt) are used. The drastic condition of the Friedel-Craft reaction limit this methods to the daunomycinone- type anthracyclinones and do not allow the use of dihydroxylated building blocks. Another limitation of Friedel-Craft reactions is the lack of regioselectivity when using non-symmetrically substituted phthalic acid derivatives.

1.5.5.2 Diels-Alder reaction

The very mild conditions of this reaction made it a very attractive option to choose in the anthracenediones synthesis. (2+4) cycloaddition offers ideal possibilities for the construction of linearly annelated ring systems containing six-membered rings. Indeed, this reaction is increasingly

used when regioselectivity or stereospecifity is the question. The regioselectivity of Diels-Alder reaction with junglone derivatives has been thoroughly investigated experimentally and is in good agreement of Frontier orbital considerations (Muxfeldt 1962).

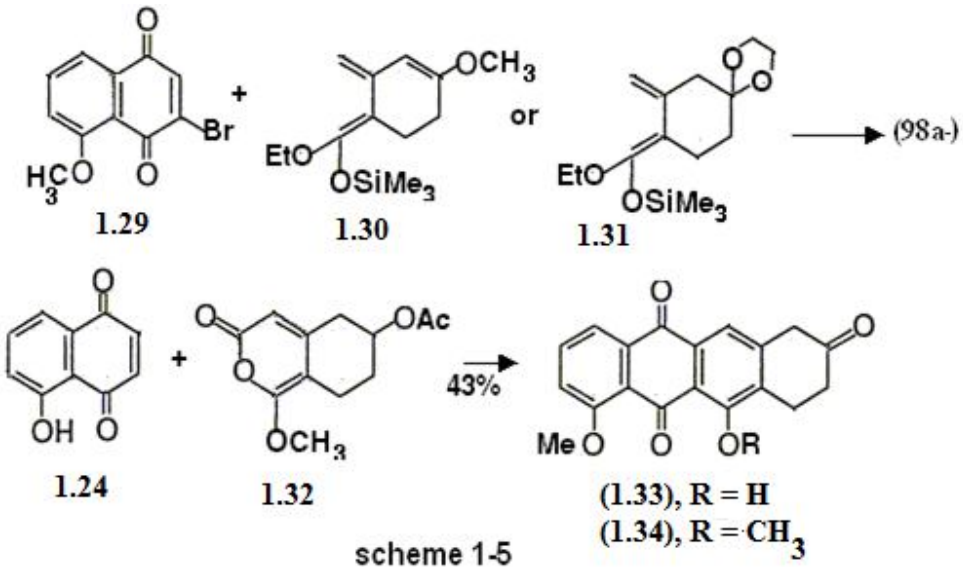
The strong hydrogen bridge in junglone (1.24) is responsible for the extremely different orbital coefficient in the LUMO of the Quinone, which consequently leads to the formation of (1.26) in the reaction of with the diene (1.25). In contrast, the methyl ether yields the adduct (1.27) with inverse orientation of the methoxy group, scheme 1- 4, (Rozeboom, M.D., *et al*, 1981).



Scheme 1-4

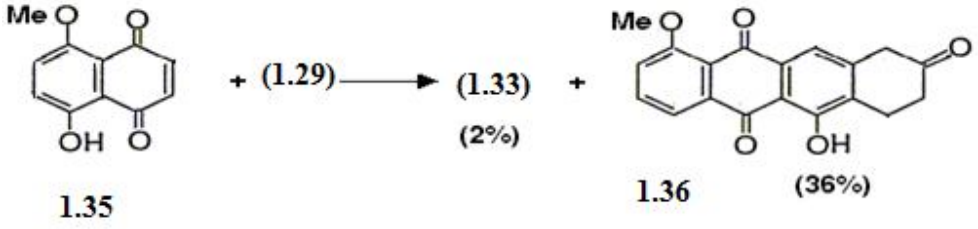
The reaction of the Quinone of bromojunglone methyl ether with the ketenacetal (1.30) (Gesson 1980) or (1.31) to give the tetracycle (1.33) also proceed with formation of ring B. In this case the bromine atom direct the regioselectivity. The presence of the bromine atom facilitates the aromatization of ring B with elimination of HBr and alcohol and

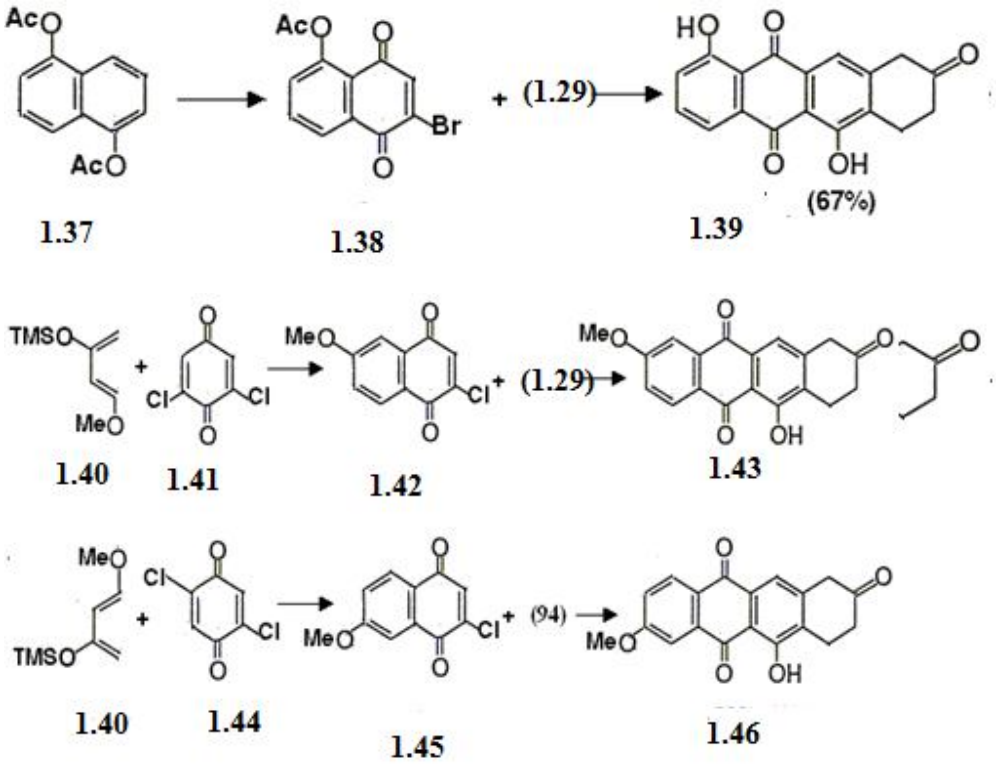
aromatization with CO_2 elimination follows the addition of junglone to the bicyclic 6-alkoxypyranone (1.32) to afford (1.34), scheme 1-5, (Jung 1982).



scheme 1-5

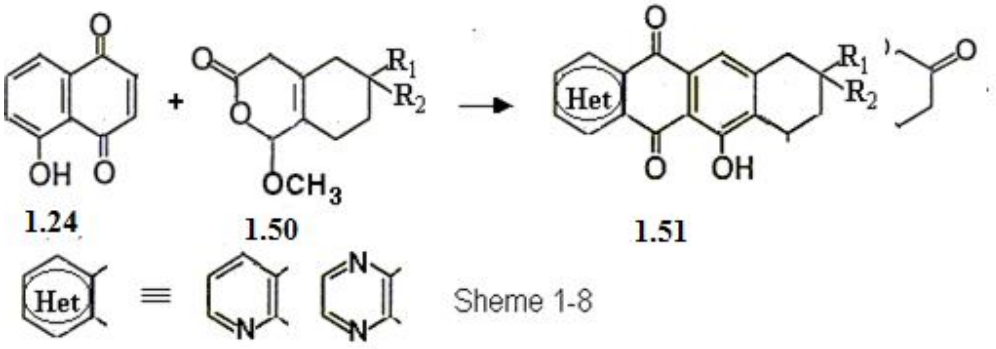
Abdallah et al. used the intermediate (1.30) to synthesized ring-D modified 11-deoxydaunomycinone by reacting it with various AB building blocks as depicted in scheme 1-6. The tetracyclic ketones (1.33), (1.36), (1.39), (1.43) and (1.46) were then subjected to ring A functionalization. The desired stereochemistry of C-7 and C-9 hydroxy groups (cis- configuration) was thus achieved, scheme 1- 6.





Scheme 1-6

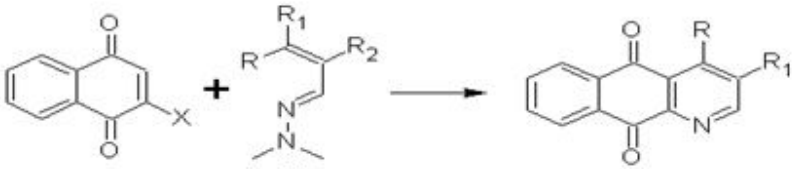
Kita, *et.al* use the precursor (1.50) to synthesize D-ring aza-analogue of 11-deoxyanthracyclinone (1.51) utilized in the synthesis of the two heterocyclic Daunorubicin analogues, scheme 1 – 8, (Kita Y., et al).



Sheme 1-8

The 1-aza-anthracene-5, 10-diones nucleus was synthesized smoothly and in good yields by reacting the azadiene (1.53) with 1, 4-naphthaquinone

(1.52) in the synthesis of Cleistopholine (1.54) and its analogues, scheme, 1-9.



1.52 **1.53** **1.54**

Scheme 1-9

Cleistopholine is a naturally occurring benzo[g]quinoline-5; which exhibits some antimicrobial activity 5,6 but weak in in vitro activity against several cancerous cell lines (Paul 2000) These analogues were employed in the synthesis of the anti-fungal Sampagine (1.55) and also in the synthesis of the experimental anti-tumor agent 3-alkenyl-1-azaanthracen-5,10-dione (1.56) and its analogue (1.57), (Vanelle 2002).

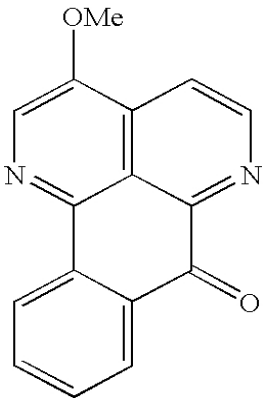
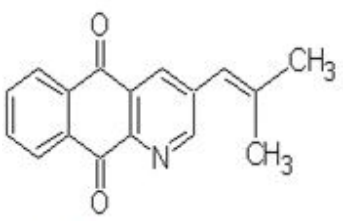
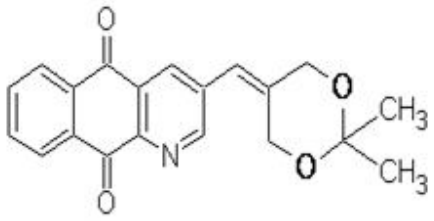


Fig. 1.55



1.56



1.57

1.6 Research objectives

- To study the discerning activity trend in the known Anthrapyrazoles using a QSAR.
- To inspect the possibility of interplay between $N_2 - C_5$ substituent in determining the activity of individual anthrapyrazoles.
- To define the parameter that encode inter-substituent interaction that define activity profile of anthrapyrazoles .
- To build a mathematical model that could be used to examine importance of position in carbon to nitrogen bio isomerism, in case of desoxy anthrapyrazole.