INTRODUCTION TO QSAR

QSAR (Quantitative Structure Activity Relationship)

Most molecular discoveries today are the results of an iterative, three-phase cycle of design, synthesis and test. Analysis of the results from one iteration provides information and knowledge that enables the next cycle of discovery to be initiated and further improvement to be achieved. A common feature of this analysis stage is the construction of some form of model which enables the observed activity or properties to be related to the molecular structure. Such models are often referred to as Quantitative Structure Activity Relationships. Quantitative structure-activity relationships (QSARs) studies unquestionably are of great importance in modern chemistry and biochemistry. The concept of QSAR is to transform searches for compounds with desired properties using chemical intuition and experience into a mathematically quantified and computerized form. QSAR methods are characterized by two assumptions with respect to the relationship between chemical structure and the biological potency of compounds. The first is that one can derive a quantitative measure from the structural properties significant to the biological activity of a compound. The properties assumed to be physicochemical such as partition coefficient or sub structural such as presence or absence of certain chemical features. The other assumption is that one can mathematically describe the relationship between biological property one wishes to optimize and the molecular property calculated from the structure. QSAR's general mathematical form is represented by the following equation.

Biological Activity = f (Physicochemical Property)

Objective of QSAR

QSAR attempts to correlate structural, chemical, statistical and physical properties with biological activity by various approaches. QSAR models are scientific credible tools for predicting and classifying biological activities of untested chemicals.

QSAR is an essential tool for lead development (optimization), a growing trend is to use QSAR early in drug discovery process as a screening and enrichment tool to eliminate from further development those chemicals lacking "drug like" properties or those chemicals predicted to elicit a toxic response.

THE HISTORY

It is widely accepted that modern QSAR began in the early 1960s. However, as long ago as 1816 scientists were making predictions about physical and chemical properties. The first investigations into the correlation of biological activities with physicochemical properties such as molecular weight and aqueous solubility began in 1841, almost 60 years before the important work of Overton and Meyer linking aquatic toxicity to lipid-water partitioning. Throughout the 20th century QSAR progressed, though there were many lean years. In 1962 came the seminal work of Corwin Hansch and co-workers, which stimulated a huge interest in the prediction of biological activities. Initially that interest lay largely within medicinal chemistry and drug design, but in the 1970s and 1980s, with increasing ecotoxicological concerns, QSAR modelling of environmental toxicities began to grow, especially once regulatory authorities became involved. Since then QSAR has continued to expand, with over 1400 publications annually from 2011 onwards.

INTRODUCTION: WHAT IS A QSAR?

Humans are inherently inquisitive. Even small children persistently ask "Why?". So it is no surprise that for many years scientists have asked why some substances have a beneficial effect on the body, whilst others are toxic, and why some are more beneficial, or more toxic, than are others. That led Crum Brown and Fraser (1868-1869) to postulate that "there can be no reasonable doubt but that a relation exists between the physiologic action of a substance (Φ) and its chemical composition and constitution (C)". Hence $\Phi = f$C. They did not go on to suggest what functions of composition and constitution might be important. Nevertheless, their equation is a valid generic quantitative structure-activity relationship (QSAR). They also pointed out that "to discover *f* we produce a known change on the constitution by which it becomes C + ΔC, and examine the corresponding change of physiological action which has become $\Phi + \Delta\Phi$. We thus obtain the relation between ΔC and $\Delta\Phi$, and by sufficiently varying C and ΔC, we may hope to get at all events an approximate solution of the problem". That was a remarkably prescient statement, for it is exactly how QSAR modelling is performed (Kubinyi, 1993).

Hence, in Crum Brown and Fraser's terminology, a QSAR equation would be:

$$\Phi = c_1C_1 + c_2C_2 + c_3C_3 + \ldots c_nC_n \quad (1)$$

where C_n represents one constitutional (structural) property, and c_n is its coefficient.

A QSAR is now defined as a mathematical relationship linking chemical structure and pharmacological activity or other property in a quantitative manner for a series of compounds. It should be noted that a correlation between a physico-chemical property (such as aqueous solubility) and some function(s) of chemical composition and constitution is usually called a quantitative structure-property relationship (QSPR), although strictly the term QSPR covers both structure-biological activity and structure-physico-chemical property relationships, as the title of this journal indicates.

As will be seen, the growth of QSAR since the 1970s has been huge. Consequently, it has not been possible in this brief review to provide more than a glimpse into modern aspects and achievements of QSAR. For those wishing to know more of those approaches, the recent excellent review by Cherkasov et al. (2014) is recommended.

EARLY APPROACHES

It is significant that for centuries there has been recognition that quantitation is an essential part of science. Leonardo da Vinci (1452-1519) commented that "there is no certainty in sciences where one of the mathematical sciences cannot be applied" (da Vinci). According to Galileo (1564-1642) "to study a given phenomenon, it was necessary to measure quantities, identify regularities, and obtain relationships representing mathematical descriptions as simply as possible" (Ponte, 1992). Gay-Lussac (1778-1850) optimistically said that "we are perhaps not far from the epoch when we will be able to submit to calculation the majority of chemical phenomena" (Gay-Lussac, 1809). Charles Babbage (1791-1871), the father of computing, could almost be thought to have had QSAR in mind when he stated that "all of chemistry…would become a branch of mathematical analysis which, like astronomy, taking its constants from observation, would enable us to predict the character of any new compound" (Babbage, 1837). A dissenting voice was that of Auguste Comte (1798-1857), who wrote that "every attempt to employ mathematical methods in the study of chemical questions must be considered profoundly irrational and contrary to the spirit of chemistry. If mathematical analysis should ever hold a prominent place in chemistry – an aberration which is happily almost impossible – it would occasion a rapid and widespread degeneration of that science" (Liang, Kvalheim, & Manne, 1993).

Sir William Thomson (1824-1907), later Lord Kelvin, was very forthright in his views on the importance of mathematics in science; "I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the state of *Science*, whatever the matter may be" (Thomson, 1884).

THE BEGINNINGS OF CORRELATION (1816-1900)

Physicochemical Properties: The Periodic Table

So far as is known, the earliest work on property prediction involved the elements, or as many of them as were known in the early nineteenth century. The first to publish on this topic appears to have been Döbereiner, who in a letter to Goethe in 1816 (Kauffman, 1999) mentioned what was to evolve into his *Dreiheit* (rule of triads); "The mineral coelestine [strontium sulphate] shows remarkable relationships: its specific weight is the mean of that of [calcium sulphate] and [barium sulphate], namely $(2.95 + 4.47)/2 = 3.71$". The rule of triads is in effect a read-across technique (Van Leeuwen, Schultz, Henry, Diderich, & Veith, 2009) whereby a property value of a chemical is predicted from known values of that property from one or more similar chemicals. Nonetheless, it is a valid predictive approach, and thus can be included as an historical use of QSPR.

Applications of QSAR Study in Drug Design

Quantitative structure-activity relationship (QSAR) and quantitative structure-property relationship (QSPR) studies are important in silico methods in rational drug design.
The aim of this methods is to optimize the existing leads in order to improve their biological activities and physico-chemical properties. Also, to predict the biological activities of

untested and sometimes yet unavailable compounds. This article is a general review of different QSAR/QSPR studies in different previous researches. R2 and Q2 parameters are used in some studies to predict the predictability and robustness of the constructed models. In all mentioned articles QSAR study were good prediction tool for investigation drug activity or binding mode on specific receptors.

Drug discovery and development is a process aims to design safe and effective medications to improve life's quality and to reduce suffering to minimum. However, the process is very complex, time consuming, and resource intensive, requiring multi-disciplinary expertise and innovative approaches. Technology in medicine and health care have rapidly changed over the past decades. Biomedical Engineering development has an essential rule in solving medical problems. Over the past ten to twenty years, there is an increased effort to apply computational abilities to the combined chemical and biological space to simplify drug discovery, and designing processes. Rational drug design methods minimize the time and cost needed in drug designing process in comparison to traditional drug discovery methods.

QSAR/QSPR studies can be used to design and identify new inhibitors de novo or to optimize absorption, distribution, metabolism, excretion and toxicity profile of identified molecules from various sources. Advances in computational techniques and hardware have eased the application of in silico methods in the designing process. Drug design can be divided in two groups: Structure based drug design (SBDD) and Ligand based drug design (LBDD). SBDD is the approach applying the structural information of the drug target to develop its inhibitor. While LBDD is used in the absence of the receptor 3D information and it relies on molecules bind to the biological target of interest. Figure I explain all different groups and types of drug designing techniques.

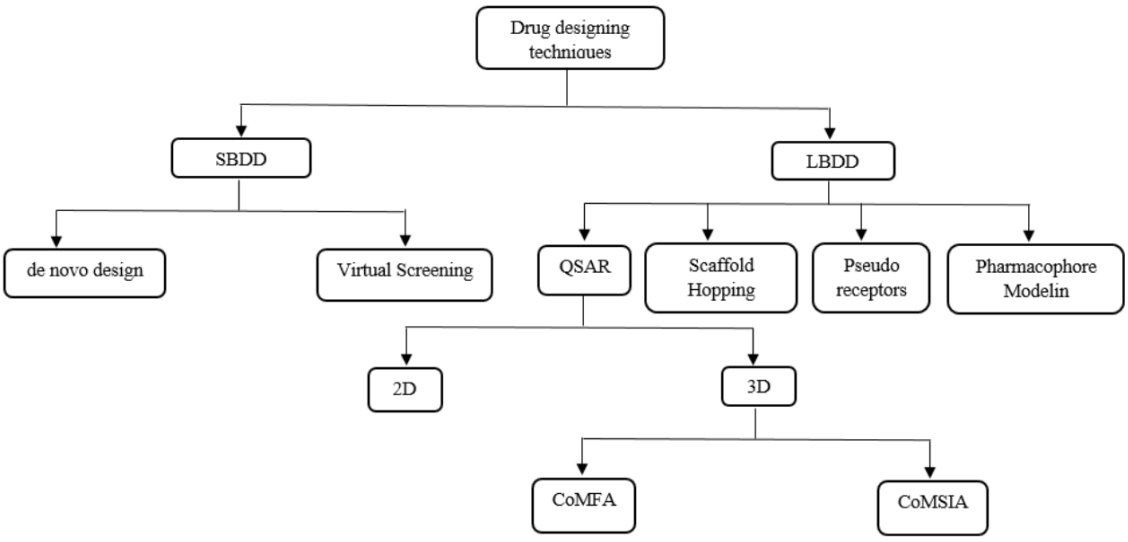

Fig I. Different groups and types of drug designing technique.

Quantitative structure-activity relationships (QSAR) have an essential role in drug design process these days, because they are cheaper alternative than the medium throughput in vitro and low throughput in vivo assays which. Also, in drug discovery and environmental

toxicology, QSAR models are now regarded as a scientifically credible tool for predicting and classifying the biological activities of untested compounds, drug resistance, toxicity prediction and physicochemical properties prediction. The QSAR methodology is based on the concept that the differences observed in the biological activity of a series of compounds can be quantitatively correlated with differences in their molecular structure. As a result, al biological activities and functions of molecules relate to specific molecular descriptors and specific regression techniques can be used to estimate the relative roles of those descriptors contributing to the biological effect.

METHODS

A. QSAR Definition and Development Quantitative structure activity relationship (QSAR) is one of the widely used approaches in ligand-based drug designing processes. In QSAR/QSPR studies quantitatively correlate and recapitulate the relationships between trends in chemical structure alterations and respective changes in biological endpoint for comprehending which chemical properties are most likely determinants for their biological activities or physicochemical properties. Quantitative Structure Activity Relationships (QSARs) mean computerized statistical method which helps to explain the observed variance in the structure changes caused by the substitution. In this concept it is assumed that the biological activity exhibited by a series of congeneric compounds is a function of various physio-chemical analysis is performed it shows that certain physio-chemical properties are favourable to the concern activity, the latter can be optimized by choosing such substituent's which would enhance such physiochemical properties. A major goal of Quantitative Structure Activity Relationship (QSAR)/ Quantitative Structure Property Relationship (QSPR) studies is to find a mathematical relationship between the activity or property under investigation, and one or more descriptive parameters or descriptors related to the structure of the molecule In QSAR, the structure of a molecule must contain the features and properties responsible for its physical, chemical, and biological activities. Figure II below describes different stages the development QSAR model process. There are a lot of softwares available for QSAR development and they are either commercial or free. We shall discuss them in detail at a later stage. These include specialized software for drawing chemical structures, interconverting chemical file formats, generating 3D structures, calculating chemical descriptors, developing QSAR models, and general-purpose software that have all the necessary components for QSAR development. For Structure Drawing or File Conversion the most common programs are ChemDraw, ACD/ChemSketch and Open Babel software. Soft wares for 3D Structure Generation are CORINA, Concord, Frog, smi23d. Descriptor Calculation can be made by using Dragon, Molconn-Z, PaDEL-Descriptor software. The first major step in a QSPR/QSAR study is the entry of the molecular structures and generation of the 3-D models. The 3-D molecular models are needed for geometric descriptor calculations. The second major step in a QSPR/QSAR study is the generation of the molecular structure descriptors. Selection of the most important descriptors is the third step and it can be achieved by using feature selection methods. The fourth major step in a QSPR/QSAR study is the generation of the QSPR/QSAR models using the descriptor sets. The fifth and last step is to validate the model by predicting the activity of compounds in the external prediction set. The results

obtained by the predictions should be compared to those achieved for the training set and cross validation set to easily understand model's fitness level

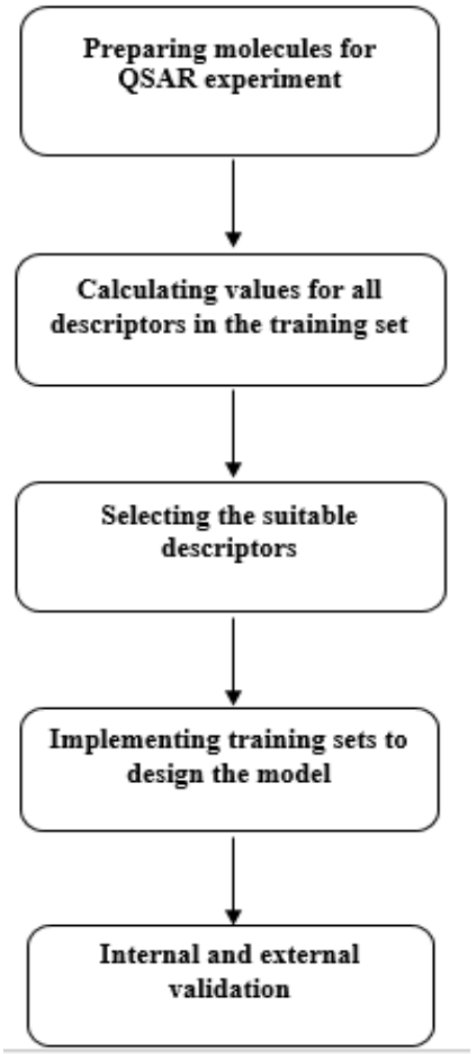

Fig II. QSAR Development process.

According to FDA (Food and Drug Administration), a drug is any substance (other than a food or device), which is used in the diagnosis, cure, relief, treatment or prevention of disease, or intended to affect the structure or function of the body. This definition is used for legal purposes, but in a lay man's term 'drug' is a pharmaceutical biomolecule or a combination of molecules that affect the body and its processes. Drug discovery starts by studying the biochemistry of the disease and the possible ways to develop a therapeutic molecule for curing the disease. So the initial outcome from the study would be the identification and analysis of specific receptors (targets) in the specialized area.

Then the identified targets must be modulated to alter their activity by performing protein receptor/target activity. Finally, drug scientist identifies the therapeutic compound in order to interact with the receptor, and the therapeutic compound could be either synthetic or naturally available.

The journey of a drug molecule from the research lab to the consumer's hand is depicted in Figure 1. Here the first compound is subjected under different assay to make them into pharmaceutically effective drugs.

Factors affecting drug discovery

Factors that affect drug discovery and development process are listed below:

1. Medicinal objective: pharmaceutical industries and researchers are keen on fixing the medicinal objective, because more precise the medicinal objective, the less likely it is to build/identify a new drug. For instance, developing a painkiller is comparatively easy than a working proton-pump inhibitor. Accordingly, the medicinal requirements affect the probability of success or failure in the drug discovery process.

2. The ability of Medicinal chemist: The role of a medicinal chemist in the drug discovery process is very crucial, involving identification/preparation of therapeutic compound and evaluating the Structure-Activity Relationships (SARs) for safety and efficacy.

3. Screening facilities: Thorough knowledge of biomolecule screening and rapid mass screening techniques are potentially important to evaluate and detect the potentially active therapeutic molecule in a short span of time.

4. Drug development facility: State of the art facility with interdisciplinary efforts in biology, chemistry, pharmaceutical researchers are essential for drug development process.

5. The cost of a new drug: Following two factors affects the cost of drug:

(i) Cost involved in synthesis: Of about 5000-10,000 biomolecules of attention, only one therapeutic molecule may enter the pharmaceutical market.

(ii) Nature of the active molecule: Cost involved in design and fabrication will be high if the active molecule is prepared by a consuming route.

Once the drug molecule is developed, it has to reach the consumer's hand in the dosage form and it should be formulated by considering the following parameters:

- Dissolution characteristics of the drug from the dosage form;
- Solubility of the drug as a function of its physicochemical characteristics;
- Intestinal effective permeability; and
- Drug's pre-systemic metabolism.

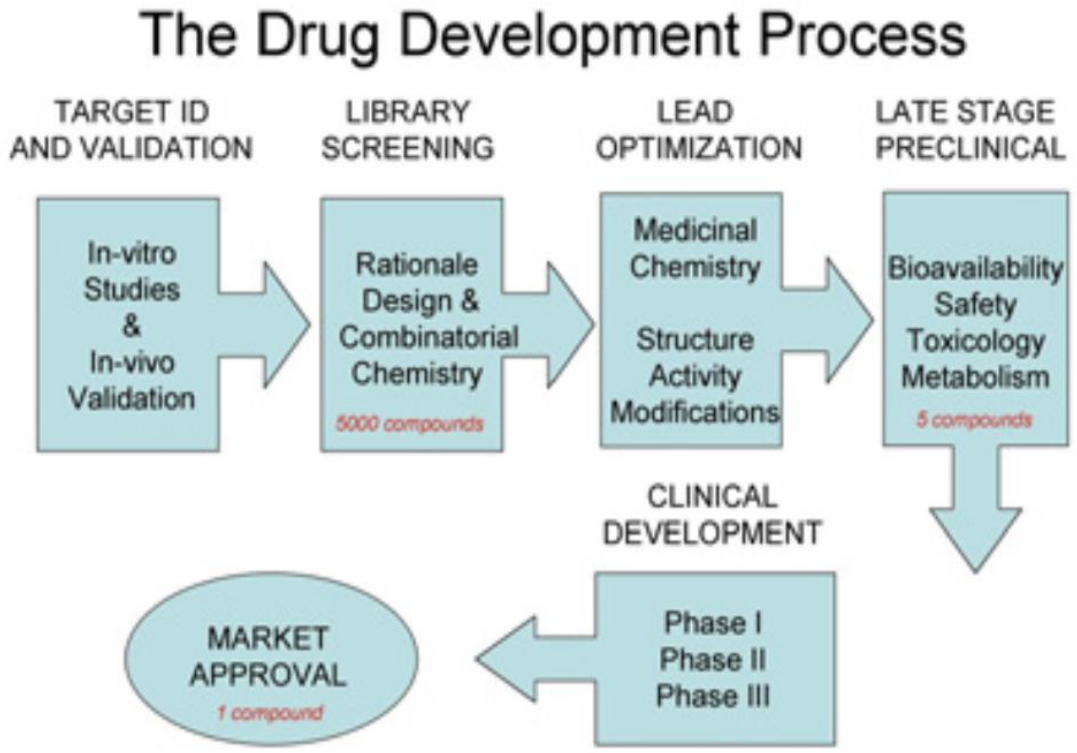

Fig. 1: New Drug development processes - From lab to consumers' hand

REFERENCES

1. Todeschini, Roberto; Consonni, Viviana. Molecular Descriptors for Chemoinformatics. Wiley. ISBN 978-3-527-31852-0.

2. Mauri, Andrea; Consonni, Viviana; Todeschini, Roberto. "Molecular Descriptors". Handbook of Computational Chemistry. Springer International Publishing. pp. 2065–2093. ISBN 978-3-319-27282-5.

3. Roy K, Kar S, Das RN (2015). "Chapter 1.2: What is QSAR? Definitions and Formulism". A primer on QSAR/QSPR modeling: Fundamental Concepts. New York: Springer-Verlag Inc. pp. 2–6. ISBN 978-3-319-17281-1.

4. Qspr/Qsar Studies by Molecular Descriptors UK ed. Edition by Mircea V Diudea

5. Quantitative Structure-Activity Relationship (QSAR) Models of Mutagens and Carcinogens 1st Edition by Romualdo Benigni

6. A Practical Guide to Rational Drug Design 1st Edition, Kindle Edition by Sun Hongmao