CODEN [USA]: IAJPBB ISSN: 2349-7750

INDO AMERICAN JOURNAL OF
PHARMACEUTICAL SCIENCES

http://doi.org/10.5281/zenodo.3473306

Available online at: http://www.iajps.com ***Review Article***

PHARMACOPHORE AND QSAR STUDY: A REVIEW

Reena S R[1]*, [2]Dr .Prasobh G.R., [3]Dr .Sandhya S.M, [4]Sheeja Rekha A G , [5]Athira S

[1]Sreekrishna College of Pharmacy and Research Centre, Parassala, Thiruvananthapuram Dist, Kerala.

Article Received: August 2019 **Accepted:** September 2019 **Published:** October 2019

Abstract:

A pharmacophore is an abstract description of molecular features which are necessary for molecular recognition of a ligand by a biological macromolecule. The IUPAC defines a pharmacophore to be "an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response".A pharmacophore model explains how structurally diverse ligands can bind to a common receptor site. Furthermore pharmacophore models can be used to identify through denovo design or virtual screening novel ligands that will bind to the same receptor. The process for developing a pharmacophore model generally involves many steps.Pharmacophore modeling is a successful yet very diverse subfield of computer-aided drug design. The concept of the pharmacophore has been widely applied to the rational design of novel drugs.

*Quantitative structure–activity relationship models (**QSAR** models) are regression or classification models used in the chemical and biological sciences and engineering.*

Keywords: *Computer-aided drug design, Pharmacophore fingerprint, Protein design, Virtual screening, QSAR*

Corresponding author:

Reena S R,

Assistant Professor, Sree Krishna College of Pharmacy and Research Centre, Parassala, Trivandram Kerala, India 695502.

Email: reenachinchu88@gmail.com

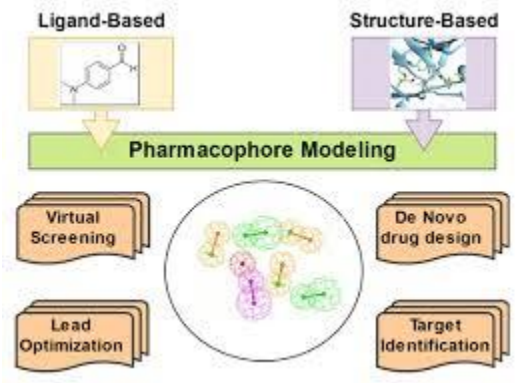
QR code

Please cite this article in press Reena S R et al., ***Pharmacophore And QSAR Study: A Review.,*** *Indo Am. J. P. Sci, 2019; 06(10).*

INTRODUCTION:

Pharmacophore approaches are successful subfields of computer-aided drug design (CADD) which have become one of the major tools in hit identification, lead optimization, and rational design of novel drugs. A pharmacophore model is the ensemble of common steric and electronic features that are necessary to ensure the optimal molecular interactions with a specific biological target and to trigger (or block) its biological response. It can be used to represent and characterize molecules on schematic 2D or 3D level by identifying the essential properties of molecular recognition. Every type of atom or group in a compound can be reduced to a pharmacophore feature (or pharmacophore fingerprints). These molecular patterns would be labeled by several chemical properties, such as hydrogen bond donors or acceptors, aromatic, cationic, *etc*, which can be used to analyze the similarity among a library of small molecules and identify the key contributing features to the biological function [1,2].

TYPE OF PHARMACOMODELING:



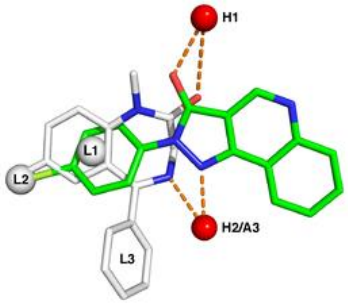
- Ligand-based pharmacophore modeling:
  In the absence of the macromolecular target structure, ligand-based pharmacophore modeling is an essential strategy for drug discovery. In this method, the common chemical characteristics from 3D structures of multiple known ligands are extracted through ligand alignment, which would represent the essential interactions between ligand and potential macromolecular target [4].

- Structure-based pharmacophore modeling:
  The structure-based pharmacophore modeling generates chemical features of the active site and the sterical relationships from 3D structure of macromolecular target or macromolecule-ligand complex. It probes the possible interaction sites between the macromolecular target and the ligands.pharmacophore modeling services from a basic concept to a well-established CADD pipeline with applications including virtual screening, ADME-tox prediction, lead optimization, drug target identification and so on [3,5,6].

FEATURES:

Typical Pharacophore features include hydrophobic centroids , aromatic rings , hydrogen bond acceptors or donors, cations, and anions. These pharmacophoric points may be located on the ligand itself or may be projected points presumed to be located in the receptor.

The features need to match different chemical groups with similar properties, in order to identify novel ligands. Ligand-receptor interactions are typically “polar positive”, “polar negative” or “hydrophobic”. A well-defined pharmacophore model includes both hydrophobic volumes and hydrogen bond vectors [7,8,9]



(An example of a pharmacophore model of the benzodiazepine binding site on the $GABA_A$ receptor. White sticks represent the carbon atoms of the benzodiazepine diazepam, while green represents carbon atoms of the non benzodiazepine CGS-9896. Red and blue sticks are oxygen and nitrogen atoms that are present in both structures. The red spheres labeled H1 and H2/A3 are, respectively, hydrogen bond donating and accepting sites in the receptor, while L1, L2, and L3 denote lipophilic binding site)

MODEL AND DEVELOPMENT OF PHARMACOPHORE:

The process for developing a pharmacophore model generally involves the following steps:

1. **Select a training set of ligands** – Choose a structurally diverse set of molecules that will be used for developing the pharmacophore model. As a pharmacophore model should be able to discriminate between molecules with and without bioactivity, the set of molecules should include both active and inactive compounds.
2. **Conformational analysis** – Generate a set of low energy conformations that is likely to contain the bioactive conformation for each of the selected molecules.
3. **Molecular superimposition** – Superimpose ("fit") all combinations of the low-energy conformations of the molecules. Similar (bioisosteric) functional groups common to all molecules in the set might be fitted (*e.g.*, phenyl rings or carboxylic acid groups). The set of conformations (one conformation from each active molecule) that results in the best fit is presumed to be the active conformation.
4. **Abstraction** – Transform the superimposed molecules into an abstract representation. For example, superimposed phenyl rings might be referred to more conceptually as an 'aromatic ring' pharmacophore element. Likewise, hydroxy groups could be designated as a 'hydrogen-bond donor/acceptor' pharmacophore element.
5. **Validation** – A pharmacophore model is a *hypothesis* accounting for the observed biological activities of a set of molecules that bind to a common biological target. The model is only valid insofar as it is able to account for differences in biological activity of a range of molecules[10,11].
6. As the biological activities of new molecules become available, the pharmacophore model can be updated to further refine it.

STEPS IN IDENTIFYING A PHARMACOPHORE:

In general, all the algorithms for pharmacophore identification utilize the following six steps:

1) Input
2) Structure Representation
3) Pattern Identification
4) Scoring
5) Feature extraction

QSAR STUDY:

In QSAR modeling, the predictors consist of physico-chemical properties or theoretical molecular descriptors of chemicals; the QSAR response-variable could be a biological activity of the chemicals. QSAR models first summarize a supposed relationship between chemical structures and biological activity in a data-set of chemicals. Second, QSAR models predict the activities of new chemicals.

Related terms include *quantitative structure–property relationships* (*QSPR*) when a chemical property is modeled as the response variable."Different properties or behaviors of chemical molecules have been investigated in the field of QSPR. Some examples are quantitative structure–reactivity relationships (QSRRs), quantitative structure–chromatography relationships (QSCRs) and, quantitative structure–toxicity relationships (QSTRs), quantitative structure–electrochemistry relationships (QSERs), and quantitative structure–biodegradability relationships (QSBRs)."

As an example, biological activity can be expressed quantitatively as the concentration of a substance required to give a certain biological response.

Additionally, when physicochemical properties or structures are expressed by numbers, one can find a mathematical relationship, or quantitative structure-activity relationship, between the two. The mathematical expression, if carefully validated can then be used to predict the modeled response of other chemical structures [12,13,14].

ESSENTIAL STEPS IN QSAR STUDIES:

Principal steps of QSAR/QSPR including
(i) Selection of Data set and extraction of structural/empirical descriptors
(ii) variable selection,
(iii) model construction and
(iv) validation evaluation.

Types:

1)Fragment based (group contribution):

Analogously, the "partition coefficient"—a measurement of differential solubility and itself a component of QSAR predictions—can be predicted either by atomic methods (known as "XLogP" or "ALogP") or by chemical fragment methods (known as "CLogP" and other variations). It has been shown that the logP of compound can be determined by the sum of its fragments; fragment-based methods are generally accepted as better predictors than atomic-based methods. Fragmentary values have been determined statistically, based on empirical data for known logP values. This method gives mixed results and is generally not trusted to have accuracy of more than ±0.1 units.

Group or Fragment based QSAR is also known as GQSAR. GQSAR allows flexibility to study various molecular fragments of interest in relation to the variation in biological response. The molecular fragments could be substituents at various substitution sites in congeneric set of molecules or could be on the basis of pre-defined chemical rules in case of non-congeneric sets. GQSAR also considers cross-terms fragment descriptors, which could be helpful in identification of key fragment interactions in determining variation of activity. Lead discovery using Fragnomics is an emerging paradigm. In this context FB-QSAR proves to be a promising strategy for fragment library design and in fragment-to-lead identification endeavours.

An advanced approach on fragment or group-based QSAR based on the concept of pharmacophore-similarity is developed. This method, pharmacophore-similarity-based QSAR (PS-QSAR) uses topological pharmacophoric descriptors to develop QSAR models. This activity prediction may assist the contribution of certain pharmacophore features encoded by respective fragments toward activity improvement and/or detrimental effects [15,16].

2)3D-QSAR

The acronym **3D-QSAR** or **3-D QSAR** refers to the application of force field calculations requiring three-dimensional structures of a given set of small molecules with known activities (training set). The training set needs to be superimposed (aligned) by either experimental data (e.g. based on ligand-protein crystallography) or molecule superimposition software. It uses computed potentials, e.g. the Lennard-Jones potential, rather than experimental constants and is concerned with the overall molecule rather than a single substituent. The first 3-D QSAR was named Comparative Molecular Field Analysis (CoMFA) by Cramer et al. It examined the steric fields (shape of the molecule) and the electrostatic fields which were correlated by means of partial least squares regression (PLS).

The created data space is then usually reduced by a following feature extraction (see also dimensionality reduction). The following learning method can be any of the already mentioned machine learning methods, e.g. support vector machines. An alternative approach uses multiple-instance learning by encoding molecules as sets of data instances, each of which represents a possible molecular conformation. A label or response is assigned to each set corresponding to the activity of the molecule, which is assumed to be determined by at least one instance in the set (i.e. some conformation of the molecule) [17,18].

On June 18, 2011 the Comparative Molecular Field Analysis (CoMFA) patent has dropped any restriction on the use of GRID and partial least-squares (PLS) technologies.

3) Chemical descriptor based

In this approach, descriptors quantifying various electronic, geometric, or steric properties of a molecule are computed and used to develop a QSAR. This approach is different from the fragment (or group contribution) approach in that the descriptors are computed for the system as whole rather than from the properties of individual fragments. This approach is different from the 3D-QSAR approach in that the descriptors are computed from scalar quantities (e.g., energies, geometric parameters) rather than from 3DAn example of this approach is the QSARs developed for olefin polymerization by half sandwich compounds

Evaluation of the quality of QSAR models:

QSAR modeling produces predictive models derived from application of statistical tools correlating biological activity (including desirable therapeutic effect and undesirable side effects) or physico-chemical properties in QSPR models of chemicals (drugs/toxicants/environmental pollutants) with descriptors representative of molecular structure or properties. QSARs are being applied in many disciplines, for example: risk assessment, toxicity prediction, and regulatory decisions in addition to drug discovery and lead optimization. Obtaining a good quality QSAR model depends on many factors, such as the quality of input data, the choice of descriptors and statistical methods for modeling and for validation. Any QSAR modeling should ultimately lead to statistically robust and predictive models capable of making accurate and reliable predictions of the modeled response of new compounds.
For validation of QSAR models, usually various strategies are adopted:

1. Internal validation or cross-validation (actually, while extracting data, cross validation is a measure of model robustness, the more a model is robust (higher q2) the less data extraction perturb the original model);
2. external validation by splitting the available data set into training set for model development and prediction set for model predictivity check;
3. blind external validation by application of model on new external data and
4. data randomization or Y-scrambling for verifying the absence of chance correlation between the response and the modeling descriptors [19].

The success of any QSAR model depends on accuracy of the input data, selection of appropriate descriptors and statistical tools, and most importantly validation of the developed model. Validation is the process by which the reliability and relevance of a procedure are established for a specific purpose; for QSAR models validation must be mainly for robustness, prediction performances and applicability domain (AD) of the models [20,21,22].

Some validation methodologies can be problematic. For example, *leave one-out* cross-validation generally leads to an overestimation of predictive capacity. Even with external validation, it is difficult to determine whether the selection of training and test sets was manipulated to maximize the predictive capacity of the model being published.
Different aspects of validation of QSAR models that need attention include methods of selection of training set compounds, setting training set size and impact of variable selection for training set models for determining the quality of prediction. Development of novel validation parameters for judging quality of QSAR models is also important [23,24,25].

APPLICATION:

Chemical:
One of the first historical QSAR applications was to predict boiling points.

It is well known for instance that within a particular family of chemical compounds, especially of organic chemistry, that there are strong correlations between structure and observed properties. A simple example is the relationship between the number of carbons in alkanes and their boiling points. There is a clear trend in the increase of boiling point with an increase in the number carbons, and this serves as a means for predicting the boiling points of higher alkanes.

A still very interesting application is the Hammett equation, Taft equation and pKa prediction methods [25,26].

Biological:
The biological activity of molecules is usually measured in assays to establish the level of inhibition of particular signal transduction or metabolic pathways. Drug discovery often involves the use of QSAR to identify chemical structures that could have good inhibitory effects on specific targets and have low toxicity (non-specific activity). Of special interest is the prediction of partition coefficient log *P*, which is an important measure used in identifying "druglikeness" according to Lipinski's Rule of Five.

While many quantitative structure activity relationship analyses involve the interactions of a family of molecules with an enzyme or receptor binding site, QSAR can also be used to study the interactions between the structural domains of proteins. Protein-protein interactions can be quantitatively analyzed for structural variations resulted from site-directed mutagenesis [27,28,29].

It is part of the machine learning method to reduce the risk for a SAR paradox, especially taking into account that only a finite amount of data is available. In general, all QSAR problems can be divided into coding and learning [30,31,35].

REFERENCES:

1. Wermuth CG, Ganellin CR, Lindberg P, Mitscher LA (1998). "Glossary of terms used in

medicinal chemistry (IUPAC Recommendations 1998)". Pure and Applied Chemistry. **70** (5): 1129–1143. doi:10.1351/pac199870051129.
2. Madsen U, Bräuner-Osborne H, Greenwood JR, Johansen TN, Krogsgaard-Larsen P, Liljefors T, Nielsen M, Frølund B (2005). "GABA and Glutamate receptor ligands and their therapeutic potential in CNS disorders". In Gad SC (ed.). Drug Discovery Handbook. Hoboken, N.J: Wiley-Interscience/J. Wiley. pp. 797–907. ISBN 0-471-21384-5.
3. Duarte, CD; et al. (2007), "Privileged structures: a useful concept for the rational design of new lead drug candidates", Mini Rev Med Chem, **7** (11): 1108–1119, doi:10.2174/138955707782331722, PMID 18045214.
4. Kier LB (September 1967). "Molecular orbital calculation of preferred conformations of acetylcholine, muscarine, and muscarone". Mol. Pharmacol. **3** (5): 487–94. PMID 6052710.
5. Kier LB (1971). Molecular orbital theory in drug research. Boston: Academic Press. pp. 164–169. ISBN 0-12-406550-3.
6. Ehrlich P (1909). "Über den jetzigen Stand der Chemotherapie". Ber. Dtsch. Chem. Ges. **42**: 17–47. doi:10.1002/cber.19090420105
7. Newman DJ, Cragg GM. Natural products as sources of new drugs over the last 25 years. J Nat Prod. 2007;70(3):461–477.
8. Lourenço AM, Ferreira LM, Branco PS. Molecules of natural origin, semi-synthesis and synthesis with anti-inflammatory and anticancer utilities. Curr Pharm Des. 2012;18(26):3979–4046.
9. Wikberg JES, Spjuth O, Eklund M, Lapins M. Chemoinformatics Taking Biology into Account: Proteochemometrics. In: Guha R, Bender A, editors. Computational Approaches in Cheminformatics and Bioinformatics. Hoboken: John Wiley & Sons; 2011:57–92.
10. Reardon S. Project ranks billions of drug interactions. Nature. 2013; 503(7477):449–450.
11. Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. Br J Pharmacol. 2011;162(6):1239–1249.
12. Krasavin M, Karapetian R, Konstantinov I, et al. Discovery and potency optimization of 2-amino-5-arylmethyl-1,3-thiazole derivatives as potential therapeutic agents for prostate cancer. Arch Pharm (Weinheim). 2009;342(7):420–427.
13. Kaul P. Drug discovery: Past, present and future. In: Jucker E, editor. Progress in Drug Research, Volume 50. Berlin: Springer Science and Business Media; 1998:9–105.
14. Veselovsky AV, Zharkova MS, Poroikov VV, Nicklaus MC. Computer-aided design and discovery of protein-protein interaction inhibitors as agents for anti-HIV therapy. SAR QSAR Environ Res. 2014;25(6):457–471.
15. Song CM, Lim SJ, Tong JC. Recent advances in computer-aided drug design. Brief Bioinform. 2009;10(5):579–591.
16. Taft CA, Da Silva VB, Da Silva CH. Current topics in computer-aided drug design. J Pharm Sci. 2008;97(3):1089–1098.
17. Bajorath J. Integration of virtual and high-throughput screening. Nat Rev Drug Discov. 2002;1(11):882–894.
18. Hopkins AL, Keserü GM, Leeson PD, Rees DC, Reynolds CH. The role of ligand efficiency metrics in drug discovery. Nat Rev Drug Discov. 2014;13(2):105–121.
19. Ballester PJ, Mangold M, Howard NI, et al. Hierarchical virtual screening for the discovery of new molecular scaffolds in antibacterial hit identification. J R Soc Interface. 2012;9(77):3196–3207.
20. Boyd MR. The position of intellectual property rights in drug discovery and development from natural products. J Ethnopharmacol. 1996;51(1–3):17–25; discussion 25–27.
21. Thiel KA. Structure-aided drug design's next generation. Nat Biotechnol. 2004;22(5):513–519.
22. J.H. van Drie (2007). "Monty Kier and the Origin of the Pharmacophore Concept" (PDF). Internet Electronic Journal of Molecular Design. **6**: 271–279.
23. 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.
24. Ghasemi, Pérez-Sánchez; Mehri, Pérez-Garrido (2018). "Neural network and deep-learning algorithms used in QSAR studies: merits and drawbacks". Drug Discovery Today. **23** (10): 1784–1790. doi:10.1016/j.drudis.2018.06.016. PMID 29936244.
25. Nantasenamat C, Isarankura-Na-Ayudhya C, Naenna T, Prachayasittikul V (2009). "A practical overview of quantitative structure-activity relationship". Excli J. **8**: 74–88. doi:10.17877/DE290R-690.
26. Nantasenamat C, Isarankura-Na-Ayudhya C, Prachayasittikul V (Jul 2010). "Advances in computational methods to predict the biological activity of compounds". Expert Opinion on Drug Discovery. **5** (7): 633–54. doi:10.1517/17460441.2010.492827. PMID 22823204.
27. Yousefinejad S, Hemmateenejad B (2015).

"Chemometrics tools in QSAR/QSPR studies: A historical perspective". Chemometrics and Intelligent Laboratory Systems. 149, Part B: 177–204. doi:10.1016/j.chemolab.2015.06.016.
28. Ghasemi, Pérez-Sánchez; Mehri, fassihi (2016). "The Role of Different Sampling Methods in Improving Biological Activity Prediction Using Deep Belief Network". Journal of Computational Chemistry. **38** (10): 1–8. doi:10.1002/jcc.24671. PMID 27862046.
29. Tropsha A, Gramatica P, Gombar VJ (2003). "The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models". QSAR &Comb. Sci. **22**: 69–77. doi:10.1002/qsar.200390007.
30. Gramatica P (2007). "Principles of QSAR models validation: internal and external". QSAR &Comb. Sci. **26** (5): 694–701. doi:10.1002/qsar.200610151. hdl:11383/1668881.
31. Chirico N, Gramatica P (Aug 2012). "Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection". Journal of Chemical Information and Modeling. **52** (8): 2044–58. doi:10.1021/ci300084j. PMID 22721530.
32. Ghasemi, Pérez-Sánchez; Mehri, fassihi (2018). "Deep neural network in QSAR studies using deep belief network". Applied Soft Computing. **62**: 251–259. doi:10.1016/j.asoc.2017.09.040.
33. Tropsha, Alexander (2010). "Best Practices for QSAR Model Development, Validation, and Exploitation". Molecular Informatics. **29** (6–7): 476–488. doi:10.1002/minf.201000061. ISSN 1868-1743. PMID 27463326.
34. Patani GA, LaVoie EJ (Dec 1996). "Bioisosterism: A Rational Approach in Drug Design". Chemical Reviews. **96** (8): 3147–3176. doi:10.1021/cr950066q. PMID 11848856.
35. Brown N (2012). Bioisosteres in Medicinal Chemistry. Weinheim: Wiley-VCH. ISBN 978-3-527-33015-7.
36. Thompson SJ, Hattotuwagama CK, Holliday JD, Flower DR (2006). "On the hydrophobicity of peptides: Comparing empirical predictions of peptide log P values". Bioinformation. **1** (7): 237–41. doi:10.6026/97320630001237. PMC 1891704. PMID 17597897.
37. Wildman SA, Crippen GM (1999). "Prediction of physicochemical parameters by atomic contributions". J. Chem. Inf. Comput. Sci. **39** (5): 868–873. doi:10.1021/ci990307l.
38. Ajmani S, Jadhav K, Kulkarni SA. "Group-Based QSAR (G-QSAR)".
39. Manoharan P, Vijayan RS, Ghoshal N (Oct 2010). "Rationalizing fragment based drug discovery for BACE1: insights from FB-QSAR, FB-QSSR, multi objective (MO-QSPR) and MIF studies". Journal of Computer-Aided Molecular Design. **24** (10): 843–64. Bibcode:2010JCAMD..24..843M. doi:10.1007/s10822-010-9378-9. PMID 20740315.
40. Prasanth Kumar S, Jasrai YT, Pandya HA, Rawal RM (November 2013). "Pharmacophore-similarity-based QSAR (PS-QSAR) for group-specific biological activity predictions". Journal of Biomolecular Structure & Dynamics. (1): 56–69. doi:10.1080/07391102.2013.849618. PMID 24266725.