International Journal of Scientific Research and Management (IJSRM)
||Volume||10||Issue||04||Pages||MP-2022-624-628||2022||
Website: www.ijsrm.in ISSN (e): 2321-3418
DOI: 10.18535/ijsrm/v10i04.mp04

A Review on Quantitative Structure-Activity and Relationships (QSAR) Methods

Sakha Ram Sirvi*, Deshbandhu Joshi[2], Raghvendra Singh Bhadauria[3]

[1.] Research Scholar, Department of Pharmaceutical chemistry, Shrinathji Institute of pharmacy, upali oden, Nathdwara, dist-Rajsamand (Raj.)-313301 INDIA
[2.] Professor, Department of Pharmaceutical chemistry, Shrinathji Institute of pharmacy, upali oden, Nathdwara, dist-Rajsamand (Raj.)-313301 INDIA
[3.] Principal, Department of Pharmaceutical chemistry, Shrinathji Institute of pharmacy, upali oden, Nathdwara, dist-Rajsamand (Raj.)-313301 INDIA

Abstract:
QSAR is an analytical application that can be used to interpret the quantitative relationship between the biological activities of a particular molecule and its structure. The product of QSAR will then produce useful equations, images or models in either 2D or 3D form that would relate their biological responses or physical properties to their molecular structure. Hologram QSAR (HQSAR) uses molecular holograms and PLS to generate fragment-based structure-activity relationships. Unlike other 3D-QSAR methods, HQSAR does not require alignment of molecules, allowing automated analysis of very large data sets. CoMFA can be applied, as it often is, when the 3D structure of the receptor is unknown. To apply CoMFA, all that is needed are the activities and the 3D structures of the molecules. Comparative Molecular Similarity Indices Analysis (CoMSIA) is known as one of the newer 3D QSAR methodology. This technique is most commonly used in drug discovery to find the common features that are important in binding to the relevant biological receptor. The partial least squares (PLS) method was used to explore a linear correlation between the CoMFA and CoMSIA fields and the biological activity values.

Key words: QSAR, HQSAR, CoMFA, CoMSIA and Partial least squares.

Introduction:

Quantitative structure-activity and relationships, often simply known as QSAR, is an analytical application that can be used to interpret the quantitative relationship between the biological activities of a particular molecule and its structure [1]. It is considered a major method of chemical researching all over the world today and is frequently used in agricultural, biological, environmental, medicinal, and physical organic studies.

QSAR:

The main objective of QSAR is to observe the biological responses of a set of molecules, measure it, and statistically relate the measured activity to some molecular structure on their surface. The product of QSAR will then produce useful equations, images or models in either 2D or 3D [2, 3] form that would relate their biological responses or physical properties to their molecular structure.

A General representation of the QSAR equation

Biological Activity= co + cd1 + $(c2d1)^2$ + c3d2 + $c4d2^2$ + ….

*di = the value of the descriptor for each molecule in the series

*ci= represents a coefficient calculated by fitting variations in the data by regression analysis.

HQSAR

Hologram QSAR (HQSAR) uses molecular holograms and PLS to generate fragment-based structure-activity relationships. Unlike other 3D-QSAR methods [4], HQSAR does not require alignment of molecules, allowing automated analysis of very large data sets. Validation studies have shown that HQSAR has predictive capabilities comparable to those of much more complicated 3D-QSAR techniques.
QSAR analysis techniques are used to speed new compound discovery research by nearly every pharmaceutical, agrochemical, and biotechnology company in the world [5, 6]. HQSAR extends the applicability of this powerful technique to the people who need it most — medicinal chemists. HQSAR is fast, easy to use, and can provide accurate prediction of activity for guiding future synthesis efforts. Trials on a range of data sets have shown that HQSAR can give results comparable with sophisticated 3D-QSAR techniques, but is much easier to use.

Advantages
Automatically builds quantitative models that relate biological activity or property to chemical structure. Minimizes the need for arbitrary user input — no 3D structure [7], conformational or alignment decisions required. Applicable to large data sets numbering 1000's of structures, as well as traditional-size sets. Rapidly identifies the SAR profile of a data set. Generates models that can be readily interpreted in chemical terms, by color-coding atoms in molecular fragments that make a positive or negative contribution to the property of interest. Searches databases to make predictions for collections of structures

QSAR Performance

HQSAR has been tested on a variety of published QSAR data sets, consisting of both homogenous and heterogenous series of molecules [8]. Data sets ranging in size from 30 or 40 molecules to several thousand structures have been successfully analyzed using HQSAR. In general, HQSAR outperforms traditional descriptors such as ClogP/CMR [9] or connectivity indices, and in many cases, gives results comparable with Tripos' industry standard 3D QSAR technique, Comparative Molecular Field Analysis (CoMFA).

CoMFA:
Comparative Molecular Field Analysis is a 3D QSAR technique based on data from known active molecules. CoMFA can be applied, as it often is, when the 3D structure of the receptor is unknown. To apply CoMFA, all that is needed are the activities and the 3D structures of the molecules. Of course, activities have to be measured, but 3D structures can be determined [10] either by measurement (crystal X-ray analysis) or by calculation from the 2D diagram and (optionally) subsequent optimization.
The aim of CoMFA is to derive a correlation between the biological activity of a set of molecules and their 3D shape, electrostatic and hydrogen bonding characteristics. This correlation is derived from a series of superimposed conformations, one for each molecule in the set. These conformations are presumed to be the biologically active structures, overlaid in their common binding mode. Each conformation is taken in turn, and the molecular fields around it are calculated. The fields, usually electrostatic and steric (van der Waals interactions), are measured at the lattice points of a regular Cartesian 3D grid; the lattice spacing is typically 2 Å [11]. The "measured" interaction is between the molecule and a probe atom (an sp^3-hybridized carbon with +1 charge).

Active molecules are placed in a three-dimensional grid (2-Å spacing) encompassing all of the molecules. At each grid point, steric energy (Lennard-Jones potential) and electrostatic energy are measured for each molecule by a probe atom (sp^3-hybridized carbon with +1 charge). To minimize domination by large steric and electrostatic energies, all energies that exceed a specified value (default 30 kcal/mol) are set to the cutoff value [12]. CoMFA uses a partial least-squares (PLS) analysis to predict activity from energy values at the grid points.

CoMSIA
Comparative Molecular Similarity Indices Analysis (CoMSIA) is known as one of the newer 3D QSAR methodology. This technique is most commonly used in drug discovery to find the common features that are important in binding to the relevant biological receptor. In this technique, both steric and electrostatic features, hydrogen bond donor, hydrogen bond acceptor and hydrophobic fields are considered. The fields are evaluated by a PLS analysis similar to the CoMFA formalism [13-15]. Models of comparative statistical

significance were obtained. Field contribution maps were produced for the different models. Due to cutoff settings in the CoMFA fields and the steepness of the potentials close to the molecular surface, the CoMFA maps are often rather fragmentary and not contiguously connected. This makes their interpretation difficult. The maps obtained by the CoMSIA approach are superior and easier to interpret. CoMSIA maps highlight the regions with in the area occupied by the ligand skeletons which is require for a particular physicochemical property important for activity. This is a more significant guide to trace the features that really matter especially with respect to the design of novel compounds.

CoMSIA was performed using the QSAR option of SYBYL. Five physicochemical properties (steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor) were evaluated, using a common probe atom with 1 Å radius, charge +1, hydrophobicity +1, hydrogen-bond donor and acceptor properties +1. The grid extended beyond the molecular dimensions by 2.0 Å in all directions [16]. Different resolution steps were tested: from 1.0 to 4.0 Å in steps of 0.5 Å. Different column filterings σmin (from 0.0 to 4.0 in steps of 0.5) and attenuation factors α (from 0.1 to 0.8 in steps of 0.1) were analysed as well. The predictive power of the models was assessed by the leave-oneout cross-validated (LOO-CV) coefficient q2 LOO, the standard error of prediction (SEP), and the residuals between the experimental (IC50exp) and predicted by LOO-CV (IC50pred) binding affinity. According to the residuals the peptides were classified into 3 categories: very well predicted peptides with residuals ≤ 0.5 log unit, well predicted peptides with residuals between 0.5 and 1.0 and poorly predicted peptides with residuals > 1.0. A mean residual value for the set and its standard deviation were also calculated. More robust cross-validation test was also performed, dividing the sets into 5 groups, developing a number of parallel models from the reduced data with one of the groups omitted, and then predicting the affinities of the excluded peptides. The mean of the q2 values from 20 runs is given as q2 CV5. As the affinity range for the separate alleles was slightly different, the ratio of the SEP to affinity range was used as an additional assessment of the statistical validity of the models. This ratio should be for further analyses.[3]

PLS

The partial least squares (PLS) method was used to explore a linear correlation between the CoMFA and CoMSIA fields and the biological activity values . It was performed in two stages. First, cross-validation analysis was done to determine the number of components to be used. This was performed using the leave-one-out (LOO) method to obtain the optimum number of components and the corresponding cross-validation coefficient, q^2. The value of q^2 that resulted in a minimal number of components and the lowest cross-validated standard error of estimate (S_{cv}) was accepted. The column filtering values (σ_{min}) was set to 2.0 kcal/mol in order to speed up the analytical process and reduce noise. Second, the optimum number of components were used to derive the final PLS model, with no validation method . The CoMFA and CoMSIA results were then graphically interpreted by field contribution maps.

Molecular Docking

The three dimensional structures known may be represented to show different views of the structures. With complex molecular mechanics programs it is possible to superimpose one structure on another. The same approach is used to superimpose the three dimensional structure of a potential drug on its possible target site. This process, which is often automated, is known as docking. Molecular docking is used to predict the structure of the intermolecular complex formed between two molecules. The small molecule called Ligand usually interacts with protein's binding sites. Binding sites are areas of protein known to be active in forming of compounds [14, 16]. There are several possible mutual conformations in which binding may occur. These are commonly called binding modes.

It also predicts the strength of the binding, the energy of the complex; the types of signal produced and calculate the binding affinity between two molecules using scoring functions. The most interesting case is the type protein-ligand interaction, which has its applications in medicine.

Types of molecular docking –

The following are majorly used method for docking-

- **Lock and Key\Rigid Docking** – In rigid docking, both the internal geometry of the receptor and ligand is kept fixed and docking is performed.

- **Induced fit\Flexible Docking** - An enumeration on the rotations of one of the molecules (usually smaller one) are performed. Every rotation the surface cell occupancy and energy is calculated; later the most optimum pose is selected.

Major steps in molecular docking:

Step I – Building the Receptor

This step the 3D structure of the receptor should be considered which can be downloaded from PDB [9, 13, 16]; later the available structure should be processed. This should include removal of the water molecules from the cavity, stabilizing the charges, filling the missing residues, generation the side chains etc according to the parameters available. The receptor should be biological active and stable state.

Step II – Identification of the Active Site

After the receptor is built, the active site within the receptor should be identified. The receptor may have many active sites but the one of the interest should be selected [17]. Most of the water molecules and heteroatom if present should be removed.

Step III – Ligand Preparation

Ligands can be obtained from various databases like ZINC, PubChem or can be sketched using tools Chemsketch. While selecting the ligand, the LIPINSKY'S RULE OF 5 should be applied. The rule is important for drug development where a pharmacologically active lead structure is optimized stepwise for increased activity and selectivity, as well as drug-like properties as described.

For selection of a ligand according to the LIPINSKY'S RULE:

- Not more than 5 –H bond donors.
- Molecular Weight NOT more than 500 Da.
- Log P not over 5.
- NOT more than 10 H bond acceptors.

Step IV- Molecular docking

This is the last step, where the ligand is docked onto the receptor and the interactions are checked [18]. The scoring function generates score depending on which the best fit ligand is selected.

Conclusion:

The main objective of QSAR is to observe the biological responses of a set of molecules, measure it, and statistically relate the measured activity to some molecular structure on their surface. HQSAR does not require alignment of molecules, allowing automated analysis of very large data sets. Validation studies have shown that HQSAR has predictive capabilities comparable to those of much more complicated 3D-QSAR techniques. The aim of CoMFA is to derive a correlation between the biological activity of a set of molecules and their 3D shape, electrostatic and hydrogen bonding characteristics. This correlation is derived from a series of superimposed conformations, one for each molecule in the set. These conformations are presumed to be the biologically active structures, overlaid in their common binding mode. CoMSIA was performed using the QSAR option of SYBYL. Five physicochemical properties (steric, electrostatic, hydrophobic, and hydrogen-bond donor and acceptor) were evaluated, using a common probe atom with 1 Å radius, charge +1, hydrophobicity +1, hydrogen-bond donor and acceptor properties +1. The grid extended beyond the molecular dimensions by 2.0 Å in all directions.

References

1. Wiley MR, Weir LC and Briggs S., "Structure-Based Design of Potent, Amidine- Derived Inhibitors of Factor Xa: Evaluation of Selectivity, Anticoagulant Activity, and Antithrombotic Activity", *J. Med. Chem.*, 2000, 43, 883-899.
2. Grimmett MR, Imidazole and Benzimidazole synthesis, Academic Press Publishing Harcourt Brace & Company, 1997, pp 85.

3. Tanaka A, Terasawa T and Takasugi H, “Identification and Structure-Activity Relationships of a Novel Series of *N*-Alkyl-*N*-(heteroaryl-substituted benzyl)-*N*'-arylureas”, *J. Med. Chem.*, 1998, 41, 2390-2410.
4. Kukarni S.K., Hand Book of Experimental Pharmacology,3rd Edn, Vallabh Prakashan,2007, pp.148-150
5. Vogel Wolfgang H., Scholkens Berhward A., Drug Discovery and Evaluation; Pharmacological Assays, 2nd Edn , March 2002, pp. 869-870.
6. Ghangale G. R, Mahale Tushar and Jadhav N.D, “Evaluation of Antiulcer Activity of Ocimum Sanctum in Rats”, Veterinary World, December , 2009,2(12),pp. 465-466
7. Nitin Dubey and Nidhi Dubey, “Antiulcer Activity of a Traditional Pearl Preparation: Mukta Bhasm.” *Research J. Pharm. and Tech.*.April.-June, **2009**,*2(2)*,287-290.
8. G.Vinothaposhan and K.Sundar, “Anti-ulcer activity of Mimosa pudica leaves against gastric ulcer in rats”, Research Journal of Pharmaceutical, Biological and Chemical Sciences(RJPBCS), October – December, 2010, 1(4),pp. 606-611
9. Dinesh Kumar Patidar , “Anti-Ulcer Activity Of Aqueous Extract Of Murraya Koenigii In Albino Rats”, International Journal of Pharma and Bio Sciences , Jan-Mar 2011,2(1),pp.524-529
10. Irini, A; Doytchinova; Darren R. Flower Edward Jenner Institute for Vaccine Research, Compton, Berkshire,U.K. ; Journal of Computer-Aided Molecular Design. 2002, 16, 535–544.
11. Giraud F, Guillon R, Logé C, Pagniez F, Picot C, Borgne ML, et al. Synthesis and structure-activity relationships of 2-phenyl-1-[pyridinyl- and piperidinylmethyl) amino] -3-(1H-1,2,4-triazol-1-yl) propan-2-ols as antifungal agents. Bioorgan Med Chem Lett. 2009;19:301–304.
12. Masubuchi M, Ebiike H, Kawasaki KI, Sogabe S, Morikami K, Shiratori Y, et al. Synthesis and biological activities of benzofuran antifungal agents targeting fungal N-myristoyltransferase. Bioorgan Med Chem. 2003;11:4463–4478.
13. Andriole VT. Current and future antifungal therapy: New targets for antifungal therapy. Int J Antimicrob Ag. 2000;16:317–321.
14. De Pauw BE. New antifungal agents and preparations. Int J Antimicrob Ag. 2000;16:147–150.
15. Andriole VT. Current and future antifungal therapy: New targets for antifungal agents. J Antimicrob Chemoth. 1999;44:151–162.
16. Khalafi-Nezhad A, Soltani Rad MN, Mohabatkar H, Asrari Z, Hemmateenejad B. Design, synthesis, antibacterial and QSAR studies of benzimidazole and imidazole chloroaryloxyalkyl derivatives. Bioorgan Med Chem. 2005;13:1931–1938.
17. Arjmand F, Mohani B, Ahmad S. Synthesis, antibacterial, antifungal activity and interaction of CT-DNA with a new benzimidazole derived Cu (II) complex. Eur J Med Chem. 2005;40:1103–1110.
18. Khabnadideha S, Rezaeia Z, Khalafi-Nezhadb A, Pakshirc K, Heirana MJ, Shobeiria H. Design and synthesis of 2-methyl and 2-methyl-4-nitro imidazole derivatives as antifungal agents. Iran J Pharm Sci. 2009;5:31–36.