Quantitative Structure-Activity Relationship (QSAR) modelling for the analysis of Estrogen Receptor Beta modulators

Siddharth Singh[a], Chandrabhan Singh[a], Priya Ranjan Kumar*[a] and Sarika Srivastava[b]
[a]IMS *Engineering College, NH24, Adhyatmik Nagar, Ghaziabad, UP, India.*
[b]IMS *Ghaziabad (University Courses Campus), NH24, Adhyatmik Nagar, Ghaziabad, UP, India.*
*priyaranjan.biet07@gmail.com

Abstract

Estrogen is mediated through two estrogen receptors (ERs), ER alpha and ER beta in cell signaling. The first estrogen receptor, known as ER alpha, was cloned in 1986. At that time this was the only receptor reported that mediates estrogenic effects. Later a second ER, now known as Estrogen receptor beta, was cloned from rat prostate. ER alpha and ER beta belong to the nuclear receptors superfamily. More specifically, they belong to the family of steroid receptors that act as ligand-regulated transcription factors. ER and ER show a high sequence homology and share affinity for mostly the same ligands. ER beta have very significant role for the development of many tissues and organs, including the uterus, ovary, mammary gland, ventral prostate, brain and immune system. There are various ligands for estrogen receptors which are useful for the treatment or prevention of a variety of conditions related to estrogen functioning including: osteoporosis, bone fractures, bone loss, fibroid disease, cartilage degeneration, uterine, endometriosis, hot flashes, cardiovascular disease, increased LDL cholesterol levels, etc. The QSAR analysis of few of these ligand molecules reveals the activity modulation as the function of their structural properties.

Keywords - QSAR, Estrogen receptor beta, Estrogen, Descriptors, Regression.

Introduction

The estrogen receptor (ER) is a strong hormone-inducible transcription factor that regulates the expression of many genes (Beekman *et al.,* 1993). ERα and ERβ belong to the steroid/thyroid hormone superfamily of nuclear receptors, members of which share a common structural architecture (Alwis, 2001). They are composed of three independent but interacting functional domains: the NH2-terminal or A/B domain, the C or DNA-binding domain, and the D/E/F or ligand-binding domain (Nilsson *et al.,* 2001). The A/B domain at the NH2 terminus contains the AF-1 site where other transcription factors interact. The C/D domain contains the two-zinc finger structure that binds to DNA, and the C/F domain contains the ligand binding pocket as well as the AF-2 domain that directly contacts co-activator peptides (Nilsson *et al.,* 2001). Binding of a ligand to ER triggers conformational changes in the receptor and this leads, via a number of events, to changes in the rate of transcription of estrogen-regulated genes (Nilsson *et al.,* 2001). For the biological effects of estrogen, it has to be mediated by other receptor proteins like estradiol-17β (E2) (Jensen and Jacobsen, 1962).The discovery of ERβ has revitalized the search for improved tissue-selective estrogen receptor modulators (SERM). As, such ligands could provide the benefits of estrogens and avoid unwanted side effects of E2. In the clinical setting, these pharmaceuticals would be used for prevention or treatment of menopausal symptoms, osteoporosis, cardiovascular disease, and breast cancer in women, or other estrogen-related indications affecting both men and women (Compston, 1998; Gustafsson, 1998; Cosman and Lindsay, 1999; Negro-Vilar, 1999; Purdie and Beardsworth, 1999).

The activity of any compound is highly dependent upon its structural properties. The concept of quantitative structure-activity relationship (QSAR), the quantitative correlation of the physicochemical

properties of molecules with their biological activities was first given by Corwin Herman Hansch in 1969 (Hansch, 1969). QSAR attempt to correlate structural molecular properties (descriptors) with functions (i.e. physicochemical properties, biological activities, toxicity, etc) for a set of similar compounds, with the help of statistical methods (Saliner, 2004). As a result, a simple mathematical relationship is established:

Activity = f (structural molecular or fragment properties) ... (1)

Applications of QSAR can be extended to any molecular design purpose, including environmental sciences, prediction of different kinds of biological activity by correlation of congeneric series of compounds, lead compound optimization, classification, diagnosis and elucidation of mechanisms of drug action, and prediction of novel structural leads in drug discovery (Saliner, 2004). The goal of structure-activity modelling is to analyse and detect the determining factors for the measured activity for a particular system, in order to have an insight of the mechanism and behaviour of the studied system (Puig, 2006). QSAR model also helps us to determine which structural property influences the activity more.

This work is foccused toward establilshing the QSAR model for the estrogen receptor beta modulating compounds and to study the effect of their structural properties on the activity of these modulators.

Model Development

A QSAR is a statistical model that relates a set of structural descriptors of a chemical compound to its biological activity. To develop the QSAR model 20 ligand molecules which bind to Estrogen Receptor Beta of whom the activities are known were selected. Among these ligand molecules 80% were used as training dataset and remaining as test dataset for 2D QSAR modelling.

Table 1: List of ER beta modulating ligand molecule and their corresponding activity.

Training set compound	**Activity**	**Test set compound**	**Activity**
(9aS)-9a-butyl-6-ethyl-1-fluoro-3,8,9,10-tetrahydroindeno[2,1-e]indazol-7-one	0.3	(1R,2S)-1-[4-[2-[(3R)-3-methylpyrrolidin-1-yl]ethoxy]phenyl]-2-phenyl-tetralin-6-ol	0.6
(9aS)-9a-ethyl-1-fluoro-6-(trifluoromethyl)-3,8,9,10-tetrahydroindeno[2,1-e]indazol-7-one	0.3	1-[2-[4-(dihydroxyBLAHyl) phenoxy] ethyl] pyrrolidine-2, 5-dione	0.64
7-bromo-2-(3-fluoro-4-hydroxy-phenyl)benzofuran-5-ol	0.35	(9aS)-9a-butyl-6-(trifluoromethyl)-3,8,9,10-tetrahydroindeno[2,1-e]indazol-7-one	0.8
4-bromo-2-(4-hydroxyphenyl)-7-methoxy-benzofuran-5-ol	0.5	Estradiol	1.1
3-(2-fluoro-4-hydroxy-phenyl)-7-hydroxy-naphthalene-1-carbonitrile	0.5		
4-[8-fluoro-5-[4-(2-pyrrolidin-1-ylethoxy)phenyl]-2,3-dihydro-1-benzoxepin-4- yl]phenol	0.7		

Copyright©2017 IMSEC

(9aR)-9a-butyl-6-(trifluoromethyl)-8,9-dihydro-3H-indeno[2,3-e]indazole-7,10-dione	0.9	
hydroxy-methyl-propyl-BLAHone	1.0	
3-(6-hydrOxy-naphthalen-2-yl)-benzo[d] isooxazol-6-ol	1.0	
(9aS)-9a-butyl-1-fluoro-6-(trifluoromethyl)-3,8,9,10-tetrahydroindeno[2,1-e]indazol-7-one	1.0	
(2R,3S)-3-(2-hydroxyphenyl)-2-[4-[2-(1-piperidyl)ethoxy]phenyl]-2,3-dihydro-1,4- benzoxathiin-6-ol	1.1	
2-(3-fluoro-4-hydroxy-phenyl)-5-hydroxy-benzofuran-7-carbonitrile	1.1	
7-(3-fluoro-4-hydroxy-phenyl)-3-hydroxy-naphthalene-1-carbaldehyde	1.1	
1-chloro-8-fluoro-6-(4-hydroxyphenyl) naphthalen-2-ol	1.1	

To do QSAR modelling, nine different descriptors were selected and there values for all above listed compounds were taken from online zinc database (http://zinc.docking.org/). The list of selected descriptors and their corresponding values are given in table below.

Table 2: List of nine different descriptors with their values for selected compounds.

Sl No.	Ligand Name	xlogP	Apolar desolvation (kcal/mol)	Polar desolvation (kcal/mol)	H-bond donors	H-bond acceptors	Net charge	tPSA (Å²)	Molecular weight (g/mol)	Rotat-able bonds
1	**(9aS)-9a-butyl-6-ethyl-1-fluoro-3,8,9,10-tetrahYdroindeno[2,1-e]indazol-7-one**	4.49	9.99	-8.8	1	3	0	46	326.415	4
2	**(9aS)-9a-ethyl-1-fluoro-6-(trifluoromethyl)-3,8,9,10-tetrahydroindeno[2,1-e]indazol-7-one**	3.47	8.02	-11.03	1	3	0	46	338.304	2
3	**7-bRomo-2-(3-Fluoro-4-hydroxy-phenyl)benzofuran-5-ol**	4.29	3.57	-11.9	2	3	0	54	323.117	1
4	**4-bromo-2-(4-hydroxyphenyl)-7-methoxy-benzofuran-5-ol**	4.18	2.92	-11.43	2	4	0	63	335.153	2
5	**3-(2-fluoro-4-hydroxy-phenyl)-7-hydroxy-naphthalene-1-carbonitrile**	3.71	4.95	-12.12	2	3	0	64	279.27	1
6	**(1R,2S)-1-[4-[2-[(3R)-3-methylpyrrolidin-1-yl]ethoxy]phenyl]-2-phenyl-tetralin-6-ol**	6.54	13.92	-40.84	2	3	1	34	428.596	6

7	**1-[2-[4-(dihydroxyBLAHyl)phenoxy]ethyl]pyrrolidine-2,5-dione**	3.22	6.68	-19.31	2	8	0	106	499.519	5
8	**4-[8-fluoro-5-[4-(2-pyrrolidin-1-ylethoxy)phenyl]-2,3-dihydro-1-benzoxepin-4-yl]phenol**	5.39	12.69	-40.06	2	4	1	43	446.542	6
9	**(9aS)-9a-butyl-6-(trifluoromethyl)-3,8,9,10-tetrahydroindeno[2,1-e]indazol-7-one**	4.25	9.68	-14.64	1	3	0	46	348	4
10	**(1R,2S)-2-phenyl-1-[4-[(2S)-2-pyrrolidin-1-ylpropoxy]phenyl]tetralin-6-ol**	6.40	13.69	-35.82	2	3	1	34	428.596	6
11	**(9aR)-9a-butyl-6-(trifluoromethyl)-8,9-dihydro-3H-indeno[2,3-e]indazole-7,10-dione**	3.59	8.74	-9.5	1	4	0	63	362.351	4
12	**hydroxy-methyl-propyl-BLAHone**	3.89	8.81	-8.59	1	2	0	37	282.383	2
13	**7-bromo-2-(4-hydroxy-2-methyl-phenyl)-1,3-benzoxazol-5-ol**	3.78	0.6	-8.67	2	4	0	66	320.142	1
14	**3-(6-hydroxy-naphthalen-2-yl)-benzo[d]isooxazol-6-ol**	3.58	3.96	-11.72	2	4	0	66	277.279	1
15	**(9aS)-9a-butyl-1-fluoro-6-(trifluoromethyl)-3,8,9,10-tetrahydroindeno[2,1-e]indazol-7-one**	4.54	9.56	-11.06	1	3	0	46	366.358	4
16	**Estradiol**	3.43	4.73	-4.93	2	2	0	40	272.388	0
17	**(2R,3S)-3-(2-hydroxyphenyl)-2-[4-[2-(1- piperidyl)ethoxy] phenyl**	5.84	10.31	-41.81	3	5	1	63	464.607	6
18	**2-(3 fluoro-4-hydroxy-phenyl)-5-hydroxy-benzofuran-7-carbonitrile**	3.23	3.36	-15.17	2	4	0	77	269.231	1
19	**7-(3-flUoro-4-hydroxy-phenyl)-3-hydroxy-naphthalene-1-carbaldehyde**	4.00	4.17	-12.25	2	3	0	58	282.27	2
20	**1-chloro-8-fluoro-6-(4-hydroxyphenyl)naphthalen-2-ol**	4.89	4.94	-10.32	2	2	0	40	288.705	1

Copyright©2017 IMSEC

The structural properties play critical role for the activity of any compound. In order to establish the relation between the structural properties called descriptors and the activity of corresponding compound, the multiple regression analysis was carried out. Multiple regression analysis helps to establish the relationship between the input variables i.e. descriptors of ligand molecule and output i.e. the activity of that compounds (Singh and Kumar, 2014). The behavior of the system is approximated to the following multiple linear regression equation.

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7 + b_8X_8 + b_9X_9 \quad \ldots (2)$$

Where Y is the calculated activity of ligand molecule, X_1 is xlogP value, X_2 is Apolar desolvation value, X_3 is Polar desolvation value, X_4 is number of H-bond donors, X_5 is number of H-bond acceptor, X_6 is Net charge, X_7 is total Polar Surface Area (tPSA), X_8 is total Molecular weight and X_9 is number of Rotatable bonds. b_0, b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8 and b_9 are linear coefficients of regression equation. An online regression tool form the url http://www.xuru.org/rt/toc.asp was used for generating regression equation. The following regression equation was generated after doing multi regression analysis.

$$Y = -2.741817536 \times 10^{-1} X_1 - 2.685843365 \times 10^{-2} X_2 + 1.716219813 \times 10^{-2} X_3 + 5.510215127 \times 10^{-1} X_4 - 3.565596641 \times 10^{-2} X_5 + 2.244982402 \times 10^{-1} X_6 - 8.325690992 \times 10^{-3} X_7 - 3.811445153 \times 10^{-3} X_8 + 2.788983205 \times 10^{-1} X_9 + 2.544924944 \quad \ldots (3)$$

Result and Discussion

The table below shows the actual activity and calculated activity of testing dataset through Multiple Regression equation generated above.

Table 3: Result of test dataset after calculating activity through regression equation

Sl No.	Ligand Name	Actual Activity	Calculated Activity	Error
6	(1R,2S)-1-[4-[2-[(3R)-3-methylpyrrolidin-1-yl]ethoxy]phenyl]-2-phenyl-tetralin-6-ol	0.6	0.653322	0.053322
7	1-[2-[4- (dihydroxyBLAHyl) phenoxy] ethyl] pyrrolidine- 2,5-dione	0.64	0.576117	0.063882
9	(9aS)-9a-butyl-6-(trifluoromethyl)-3,8,9,10-tetrahydroindeno[2,1-e]indazol-7-one	0.8	0.718690	0.081309
16	Estradiol	1.1	1.052343	0.047656
Residual Sum of Squares (r^2)				**1.140109**

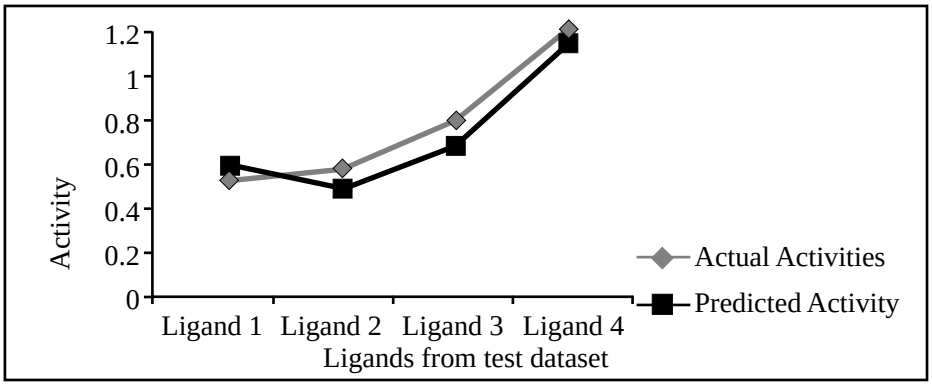

Figure 1: Plot for the actual and calculated activity of test dataset by regression equation.

The higher accuracy of the model makes it suitable for the prediction of other modulator ligand molecules as well. This model can also be used to assess the effect of any descriptor mentioned above on the activity of respective compound. After the structural analysis, it has been found that the xlogP value, number of Hydrogen bond donor, Net charge and number of rotatable bonds of any modulator highly affect the activity of that compound. Any slight variation in these descriptors shows large change in the activity. Whereas the remaining descriptors having slightly less effect on the activity. Hence it can also be concluded that the activity of any compound is highly dependent upon its structural properties.

References

Alwis, I.D. 2001. Mechanism of ovarian disruption by neonatal des exposure: a further investigation into ovarian function in the hamster cheek pouch. Masters thesis, Wichita State University, U.S.A.

Beekman, J.M., Allan, G.F., Tsai, S.Y., Tsai, M.J., O'Malley, B.W. 1993. Transcriptional activation by the estrogen receptor requires a conformational change in the ligand binding domain. *Molecular Endocrinology,* 7 (10), 1266-74.

Compston, J.E. 1998. Selective oestrogen receptor modulators: potential therapeutic applications. *Clinical Endocrinology,* 48, 389–391.

Cosman, F., Lindsay, R. 1999. Selective estrogen receptor modulators: clinical spectrum. *Endocrine Reviews,* 20, 418–434.

Gustafsson, J.A. 1998. Therapeutic potential of selective estrogen receptor modulators. *Current Opinion in Chemical Biology,* 2, 508–511.

Hansch, C. 1969. A quantitative approach to biochemical structure-activity relationships. *Accounts of Chemical Research,* 2, 232-239.

Jensen, E.V., Jacobsen H. 1962. Basic guides to the mechanism of estrogen action. *Recent Progress in Hormone Research,* 18, 387–41.

Negro-Vilar, A. 1999. Women's health in the next millenium: endocrinology is leading the way. *Endocrine Reviews,* 20, 241–242.

Nilsson,S., Mäkelä,S., Treuter, E., Tujague, M., Thomsen, J., Andersson, G., Enmark, E., Pettersson, K., Warner, M., Gustafsson, J.A. 2001. Mechanisms of Estrogen Action. *Physiological Reviews,* 81 (4), 1535-1565.

Puig, M. B. 2006. Three Dimensional Simulitary of Molecules with biological interest on the basis of molecular interaction potentials. Doctoral Thesis, *Universitat Pompeu Fabra.*

Purdie, D.W., Beardsworth, S.A. 1999. The selective oestrogen receptor modulation: evolution and clinical applications. *British Journal of Clinical Pharmacology,* 48, 785–792.

Saliner, A.G. 2004. Molecular quantum similarity in QSAR: Applications in computer Aided molecular Design. Doctoral Thesis, *University of Girona.*

Singh, R.K., Kumar, P.R. 2014. Development of artificial neural network modeling of p-cresol biodegradation. *International Journal of Advanced Biotechnology and Research,* 5 (1), 43-53.

Copyright©2017 IMSEC