Print ISSN: 0974-2115
www.jchps.com Journal of Chemical and Pharmaceutical Sciences

A DFT-based Quantitative structure activity relationship Study of organometallic estradiol derivatives

Mohamed Abd Esselem Dems[1, 3], Souhila Laib[3], Nadjia Latelli[2, 3]*, and Nadia Ouddai[3]
[1]Biotechnology Research Center, Ali Mandjli, Nouvelle Ville UV 03 BP E73, Constantine, Algeria.
[2]Laboratoire de Chimie des Matériaux et des Vivants: Activite, Reactivite, Univ. Hadj-Lakhdar, Batna, Algeria
[3]Faculte des sciences, département de chimie, Universite de Msila, BP 166 Ichbilia, 28000 M'sila, Algeria
***Corresponding author: E-Mail: latellinadjia@yahoo.fr**

ABSTRACT

In this work, we have developed QSAR model for relative binding affinity (RBA) of a diverse organometallic estradiol derivatives. The DFT method was used to calculate quantum chemical descriptors. All the results have indicated that the QSAR model that was built was robust and satisfactory (R^2 = 90.12%, Q^2_{LOO} = 86.61%, RMSE = 0.272, F = 60.6473, Q^2_{ext} =86.07%). We have, therefore, applied this model to predict the RBA for two isomers β and α wherein Mn $(CO)_3$ complex with the aromatic ring of estradiol, and the two isomers show little appreciation for the estrogenic receptor (RBA_β = 1.8% and RBA_α = 1.7%).

KEY WORDS: QSAR, DFT, Estradiol, Relative Binding Affinity, Organometallic.

1. INTRODUCTION

The bi-organometallic chemistry, built on the specificities of organometallic chemistry, is characterized by complex direct metal-carbon bond to address problems of biological interest (Jaouen, 2006; 2007; Hartinger, 2009). In 1985, the term bio-organometallic chemistry was first applied for the synthesis and the study of organometallic species in biological and medical interests. It covered complexes form using classical organometallic ligands (e.g. CO, alkyls and π-bonded species) and biomolecules (steroids, amino acids, sugars, peptides, DNA, vitamins, enzymes, antibodies).

The presence of steroidal hormone receptors in mammary tumors is one of the major necessary conditions to promote response to hormonal stimuli. The estradiol receptor assays are performed on a large scale in specially equipped hospitals throughout the world. Scientists explored several metal carbonyl complexes as markers of estradiol receptors such as $Cr(CO)_3$ which could bind to the aromatic ring of estradiol by producing two isomers: The β isomer in which $Cr(CO)_3$ is in the same face as the 17β-OH groups, and the 13β-CH_3 which had a low affinity for the estrogenic receptor (relative binding affinity "RBA" = 1.8%). However, the affinity was strong for the isomer α (RBA = 28%) (Fig.1) (Vessieres, 1988; Jaouen, 1989). The Density Functional Theory (DFT) approach was used to study this deference affinity.

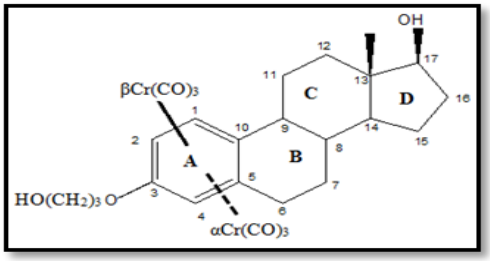


Figure.1. Optimized structures of estradiol isomer complexes with $Cr(CO)_3$

The results published in Dems (2015), shows that the relative binding affinity (RBA) is not reliant to intrinsic proprieties of estradiol. To overcome this obstacle, we opted for the QSAR model and evaluating the estrogenic activity for a range of organometallic compounds of estradiol derivatives by quantitative structure activity relationship (QSAR) models.

This type of analysis (QSAR) correlates the chemical structural characteristics with biological activity. The model can serve as a screening tool to predict the biological activity of the untested compounds (Ankley, 1997).

2. METHODS & MATERIALS

Computational Methods: All calculations for the optimization of the geometries of molecules were performed at the density functional theory DFT using the ADF (Amsterdam density function) software (Baerends, 1973), which also were carried out by generalized gradient approximations (GGA) via the exchange functional of "Perdew, Burke, and Ernzerhof's" (PBE) (John Perdew, 1997) through the use of the TZP (Triple Polarized Zeta) basis set. Frequency calculations were performed in all cases to ensure that the calculated geometries would correspond to true minima.

The most stable structure for each compound was generated and used for calculating various quantum chemical descriptors, such as E-LUMO (energy of lowest unoccupied molecular orbital). The optimized geometries were loaded in E-DRAGON (Tetko I V, 2005; VCCLAB 2005) software which is used to calculate Molar refractivity (MR) and the Moriguchi octanol-water partition coefficient MLOGP (Moriguchi, 1992; 1994; Hansch, 1995).

QSAR Model:

Data set for analysis*:* The relative binding affinity (RBA) of estradiol derivatives was collected from literature (Vessieres, 1988; Jaouen, 1989; Wust, 1998; Luyt, 2003; Anstead, 1997) and the structures of these derivatives are shown in Fig.2 and Table.1. All biological data shown in RBA were converted to logarithmic scale (log-RBA).

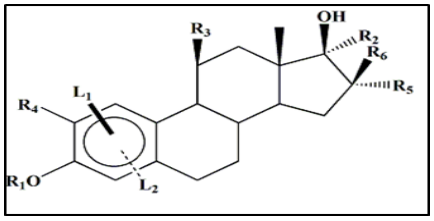
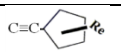
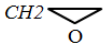


Figure.2. Substituted estradiol derivative structures

Table.1. Relative Binding Affinities of Substituted Estradiol Derivatives

Code	R1	R2	R3	R4	R5	R6	L1	L2	Log-RBA
a1	$Si(CH_3)_2tBu$	H	H	H	H	H	-	$\alpha Cr(CO)_2CSO$	-0.398
a10	H	C≡CH	H	H	H	H	-	-	1.845
a11	H	$C=CH_2$	H	H	H	H	-	-	1.800
a12	H	$(CH_2)_2$-S-Ph	H	H	H	H	-	-	1.490
a13	H	CH_3	H	H	H	H	-	-	1.643
a15	H	C=CHBr	H	H	H	H	-	-	1.851
a16	H	C=CHCl	H	H	H	H	-	-	2.061
a17	H	Ph	H	H	H	H	-	-	1.398
a18	H	H	H	H	H	H	-	-	2.000
a2	$(CH_2)_3$-OH	H	H	H	H	H	-	-	1.568
a20	H	H	$(CH_2)_3OCH_2OBr$	H	H	H	-	-	-0.027
a22	H	C≡C-$(CH_2)_2OH$	H	H	H	H	-	-	0.591
a23	H	Ph-$Cr(CO)_3$	H	H	H	H	-	-	1.041
a24	H	H		H	H	H	-	-	1.041
a25	H		H	H	H	H	-	-	1.079
a26	$Si(CH_3)_2tBu$	H	H	H	H	H	$\beta Cr(CO)_3$	-	-0.398
a28	H	$C≡C(CH_2)_2OCH_2OBr$	H	H	H	H	-	-	1.301
a29				H	H	H	-	-	1.380
a30	-$(CH_2)_3$-OH	H	H	C=CH-CH3	H	H	-	-	1.398
a31	H	C≡C-CHCl	H	H	H	H	-	-	1.290
a33	H	C=CH-CH3	H	H	OH	H	-	-	-0.155
a34	H	Ph	H	H	OH	H	-	-	-0.046
a35	H	CH3	H	H	OH	H	-	-	0.061
a36	H	H	H	H	OH	C≡CH	-	-	0.114
a37	H	H	H	H	OH	H	-	-	0.121
a4	$(CH_2)_3OH$	H	H	H	H	H	$\beta Cr(CO)_3$	-	0.243
a5	H	C≡C-$(CH_2)_6OH$	H	H	H	H	-	-	0.813
a7	$Si(CH_3)_2tBu$	H	H	H	H	H	-	-	1.041
a8	$Si(CH_3)_2tBu$	H	H	H	H	H	-	$\alpha Cr(CO)_3$	0.041
a9	H	C≡C-CH3	H	H	H	H	-	-	1.643

Method: Firstly, the multiple linear regression analysis and the variable selection were performed by the QSARINS software (Gramatica, 2014) using the Ordinary Least Square regression (OLS) method and GA-VSS (Genetic Algorithm-Variable Subset Selection) (Leardi, 1992). 20% of the molecules were selected as test set. To avoid multi collinearity, we applied the rule QUICK (Q under Influence of K) (Todeschini, 1999).

Secondly, the robustness of the models and their predictivity were evaluated by the coefficient of determination (R^2), coefficient of determination adjusted (R^2_{adj}), standard deviation(s), Fisher's value (F) cross-

validation techniques (Q^2_{LMO}; bootstrap), and (Q^2_{LOO},) (Allen, 1974). The real predictive capability of each model developed on the training set is verified on an external validation parameter Q^2_{EXT}.

Finally, the presence of outliers and chemicals, which were very structurally influential in determining model parameters, was verified by applicability domain (AD) (Eriksson, 2003) and was discussed according to the Williams plot.

3. RESULTS AND DISCUSSION

The correlation matrices between the logarithm of relatives binding affinity (log RBA) and all of descriptors (E-LUMO, MR and MLOGP) is given in Table.2. It shows that the descriptors are not correlated between the descriptors.

Table.2. Correlation matrix between biological activities, log-RBA, and physicochemical descriptors

	Log-RBA	MR	E-LUMO
MR	-0.410		
E-LUMO	-0.835	-0.014	
MLOGP	0.477	-0.013	-0.440

The relationship between the structural descriptors and the binding affinity (log RBA) is modelled on the equation (1).

Log RBA = 3.157 - 0.026*MR + 0.100* MLOGP - 0.011*E-LUMO (1)

The values of R^2= 90.12% and R^2_{adj} = 88.64% attest the good fitting performances of the model which is highly significant (with great value of the F-value = 60.82).

The cross-validated square correlation coefficient of the model is Q^2_{LOO} =86.65%. That showed good correlation between predicted activity and actual activity. This value reflects the accuracy of the models. The similarity SDEP and SDEC means that internal capacity prediction models are not too dissimilar from their adjustment authority.

The difference between Q^2_{LOO} and Q_{LMO}30%, is about1%; this shows best stability in internal validation while at the same time the validation by the bootstrap Q^2boot = 83.49% confirms the stability and internal prediction model.

The quality of Q^2ext=86.07% and the small S = 0.272 values confirm the high predictivity of this model (Table.3). The QUIK rule is based on the K multivariate correlation index (Table.3) that measures the total correlation of a set of variables. The descriptors are not correlated (medium K_x: 20.92%) and the difference in the correlation between the block of X variables plus the response Y (K_{xy}) and that of X (K_{xx}) is sufficiently high:

$\Delta = k_{xy} - k_x$ = 21.44% (Table.3). The analysis of the AD, in the Williams plot of this model Fig.3, confirms the absence of outliers and influential points in the selected QSAR models.

Table.3. Diagnostic statistics for the Selected Model

R^2	Q^2_{LOO}	Q^2_{boot}	Q_{LMO}30%	Q^2_{ext}	R^2_{adj}	SDEP	N_{TR}
90.120	86.65	83.49	87.30	86.07	88.64	0.29	24
$SDEP_{ext}$	SDEC	K_x	K_{xy}	F	RMSE	PRESS	N_{TEST}
0.295	0.25	20.92	42.37	60.82	0.27	1.99	6

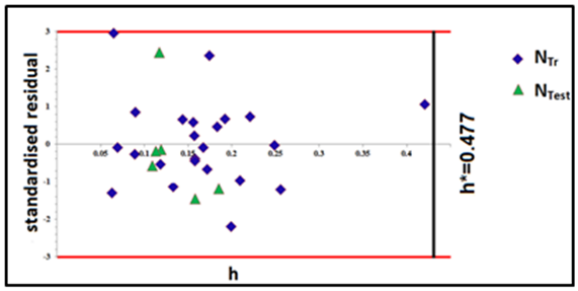


Figure.3.Plot of standardized residuals versus leverages, dash lines represent ± 3 standardized residual, dotted line represents warning leverage (h*=0.477)

The difference between R^2 and Q^2_{LOO} is small. In view of these observations, we conclude that the QSAR equation model is fairly robust.

The experimental values according to the values predicts of log RBA are presented graphically in Fig.4, which showed dispersion characteristic of a good fit.

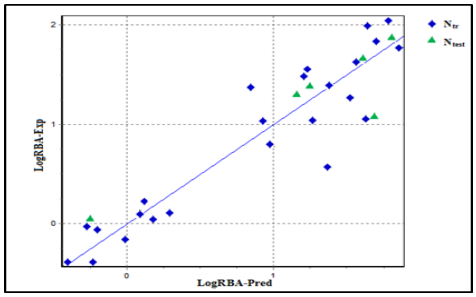


Figure.4. Regression line of the experimental and predicted values of Log-RBA for QSAR model generated by MLR

The QSAR model demonstrate a significant correlation of MR, MLOGP and E-LUMO with log RBA. The predictions' values reveal a good agreement with the experimental values. The descriptors E-LUMO, MLOGP and MR play a crucial role in enhancing the log RBA values of the estrogenic receptor.

Testing: The α isomer of estradiol complexed at $Cr(CO)_3$ moiety showed a good binding affinity (RBA 28%) but the β isomer had low RBA 1.8%. This interesting experimental result prompted us to develop a QSAR model to find the RBA of the two isomers β and α for $Mn(CO)_3$ complex with the aromatic ring of estradiol (Fig.5) .

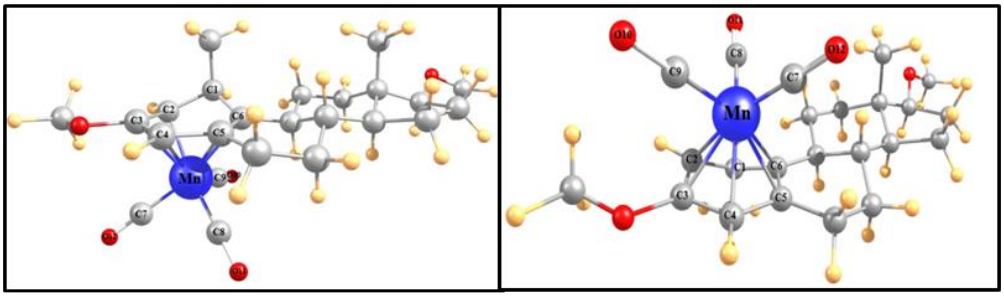


Figure.5. Structures of estradiol isomer complexes with $Mn(CO)_3$

However, no experimental measurement was performed for $Mn(CO)_3$. The two isomers showed little appreciation for the estrogenic (RBA_β = 1.8% and RBA_α = 1.7%) (Table.4).

Table.4. Predicted log RBA and Values of three selected descriptors of two estradiol isomers β and α complexes with $Mn(CO)_3$

Isomer	Code	MLOGP	MR	E-LUMO	Log-RBA Pred	RBA
α	B1	2.09	85.66	84.41	0.24	1.74
β	B2	2.17	85.93	82.97	0.26	1.81

4. CONCLUSION

The RBA was correlated with three descriptors: quantum chemical descriptor E-LUMO, physic chemical descriptors MR and octanol-water partition Moriguchico efficient MLOGP. The QSAR model is stable, and with good internal and external predictive performance and good quality of fit. This model is proposed to predict the RBA of untested organometallic derivatives of estradiol.

5. ACKNOWLEDGEMENTS

The authors are very thankful to the « University Lyon 1 and CNRS UMR 5180 Sciences Analytiques; Laboratoire de Chimie Physique Théorique, bâtiment Dirac, 43 boulevard du 11 November 1918, 69622 Villeurbanne Cedex (France), for offering the computing facilities and helpful discussion with the scientists.

REFERENCES

Allen DM, The relationship between variable selection and data augmentation and a method for prediction, Technometrics, 16, 1974, 125-127.

Ankley G, Bradbury S.P, Hermens J, Mekenyan O.G, Tollefsen K.E, Current approaches to the use of structure activity relationships (SARs) in identifying the hazards of endocrine-disurpting chemicals to wildlife, Proceedings of the EMWAT Workshop, SETAC-Europe The Netherlands, 1997, 19-40.

Anstead GM, Carlson KE, Katzenellenbogen Aj, The estradiol pharmacophore, Ligand structure-estrogen receptor binding affinity relationships and a model for the receptor binding site, Steroids, 62, 1997, 268-303.

Baerends EJ, Ellis DE, Ros P, Self-consistent molecular Hartree, Fock, Slater calculations, the computational procedure, J.Chem. Phys, 2 (1), 1973, 41-51.

Dems M.A, Latelli N, Ouddai N, The Relation Between Haptotropic Rearrangements and RBA In Estradiol Complexed, JCHPS, 8 (2), 2015, 233-238.

Eriksson L, Jaworska J, Worth A, Cronin M, Mc Dowell RM, Gramatica P, Methods for reliability, uncertainty assessment, and applicability evaluations of regression based and classification QSARs, Environ. Health Perspect, 111 (10), 2003, 1361-1375.

Gramatica P, Cassani S, Chirico N, QSARINS-chem, Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS, Journal of Computational Chemistry, software and new update, 35, 2014, 1036-1044.

Gramatica P, Chirico N, Papa E, Cassani S, Kovarich S, QSARINS: A New Software for the Development, Analysis, and Validation of QSAR MLR Models, Journal of Computational Chemistry, Software news and updates, 34, 2013, 2121-2132.

Hansch C, Leo A, Exploring QSAR: Fundamentals and Applications in Chemistry and Biology, American Chemical Society, Washington DC, 1995.

Hartinger C G, Dyson P J, Bio organometallic chemistry from teaching paradigms to medicinal applications, Chem. Soc. Rev, 38, 2009, 391- 401.

Jaouen G Top S, Vessieres A, Bio organometallics, Wiley-VCH, Weinheim, 2006.

Jaouen G, Dyson PJ, In Comprehensive Organometallic Chemistry III, Eds, R. H. Crabtree, D. M. P. Mingos, Elsevier, Ltd, Oxford, 12, 2007.

Jaouen G, Vessieres A, The use of high affinity binding bioligands modified by transition metal carbonyl moieties, Pure & Appl. Chem, 61 (3), 1989, 565-572.

John Perdew P, Burke K, Ernzerhof M, Generalized Gradient Approximation Made Simple, Phys, Rev. Lett, 78, 1997, 1396.

Langreth D.C, Vosko S.H, Response Functions and Nonlocal Approximations, Adv. Quantum Chem, 21, 1990,175.

Leardi R, Boggia R, Terrile M, Genetic algorithms as astrategy for feature selection, Journal of Chemometrics, 6, 1992, 267-281.

Luyt G L, Bigott H M, Welchb M J, Katzenellenbogena A j, 7α and 17α-Substituted Estrogens Containing Tridentate Tricarbonyl Rhenium/Technetium Complexes: Synthesis of Estrogen Receptor Imaging Agents and Evaluation Using MicroPET with Technetium-94m, Bioorganic & Medicinal Chemistry, 11, 2003, 4977–4989.

Moriguchi L, Hirono S, Liu Q, Nakagome I, Matsushita Y, Simple method of calculating octanol/water partition coefficient, Chem. Pharm, Bull, 40, 1992, 127–130.

Moriguchi L, Hirono S, Nakagome I, Hirono H, Comparison of reliability of logP values for drugs calculated by several methods, Chem.Pharm.Bull, 42, 1994, 976-978.

Tetko I.V, Gasteiger J, Todeschini R, Mauri A, Livingstone D, Ertl P, Palyulin V.A, Radchenko E.V, Zefirov N.S, Makarenko A.S, Tanchuk V.Y, Prokopenko V.V, Virtual computational chemistry laboratory design and description, J. Comput. Aid. Mol. Des, 19, 2005, 453-463.

Todeschini R, Maiocchi A, Consonni V, The K Correlation Index, Theory Development and its Application in Chemometrics, Chemom. Int. Lab. Syst, 46, 1999, 13-29.

Vessieres A, Jaouen G, Gruselle M, Rossignol J.L, Savignac M, Top S, Greenfield S, Synthesis And Receptor Binding Of Polynuclear Organometallic Estradiol Derivatives, J. steroid Biochem, 30 (1-6), 2009, 301-305.

Wust F, Carlson K.E, Katzenellenbogen J.A, Spies H, Johannsen B, Synthesis and binding affinities of new 17a-substituted estradiol-rhenium “n11” mixed-ligand and thio ether carbonyl complexes, Steroids, 63, 1998, 665-671.