Available on line www.jocpr.com

Journal of Chemical and Pharmaceutical Research

J. Chem. Pharm. Res., 2010, 2(4):500-507

ISSN No: 0975-7384
CODEN(USA): JCPRC5

Quantitative structure-activity relationship study of *para*-substituted *trans*- and *cis*-tamoxifen derivatives

José C. Barros, Magaly G. Albuquerque, Joaquim F. M. da Silva* and Octavio A. C. Antunes[†]

Instituto de Química, Universidade Federal do Rio de Janeiro, Av. Athos da Silveira Ramos 149, Centro de Tecnologia, Bloco A, 21944-910, Rio de Janeiro, RJ, Brazil

[†] **This paper is dedicated to the memory of Professor Octavio Antunes and his family**

ABSTRACT

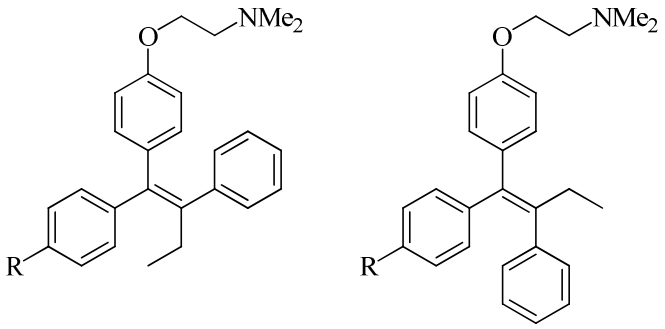
A novel quantitative structure-activity relationship model for para-substituted trans- and cis-tamoxifen derivatives is proposed, showing correlation with experimental activities. This model represents an improvement to the reported one.

Keywords: QSAR, SERM, Tamoxifen, triarylethylenes.

INTRODUCTION

Triarylethylene is the core structure of several Selective Estrogen Receptor Modulators (SERMs), such as tamoxifen and clomiphene, which are currently used in breast cancer chemotherapy [1,2] and infertility caused by polycystic ovary syndrome [3] respectively.

The industrial synthesis of SERMs containing the triarylethylene moiety employs Grignard compounds [4] or $TiCl_4$ [5], which are air-sensitive reagents. Nowadays, palladium mediated carbon-carbon coupling has emerged as a versatile approach to obtain triarylethylene derivatives from aryl halides [6-10].



trans (**1a-k**) *cis* (**2a-g**)

Figure 1. Structures of *para*-substituted *trans*- and *cis*-tamoxifen derivatives (1a-k and 2a-g).

The synthesis of a series of *para*-substituted *trans*- and *cis*-tamoxifen derivatives (**1a-k** and **2a-g**, Figure 1) and the respective pharmacological evaluation, measured as Relative Binding Affinity (RBA), were reported by McCague's group [11]. In this work, *trans* and *cis* designate the relative positions of the ethyl and phenoxyethyl-*N*,*N*-dimethylamine moieties. Our group has interest in the halogenated derivatives **1b-d** and **2b-c** from this series, since they are potential raw materials for the synthesis of new analogs to be obtained by palladium mediated C-C coupling.

More recently, Selassie's group proposed a quantitative structure-activity relationship (QSAR) model for this series of compounds based on the Hammett constants and on the I_{cis} descriptor, which is equal to 1 (one) for the *trans*-derivatives and 0 (zero) for the *cis* ones. This model, which is based only in the 2D structure of these compounds (*i.e.* classical QSAR), resulted in R^2=0.835, s=0.303, Q^2=0.779, revealing **1j** (R=CH_2OH) as outlier [12].

Since our group has been involved in both 3D-QSAR studies of SERMs [13] and in Pd catalyzed carbon-carbon coupling approaches to obtain triarylethylene derivatives [14], the present work aims to develop a new QSAR model based in both the 2D and 3D structures of this series of tamoxifen derivatives and to compare the results with the Selassie model.

RESULTS AND DISCUSSION

The 3D structures of tamoxifen derivatives were constructed in the Spartan'02 (Wavefunction Inc.) software, using as reference the X-ray crystal structure of the 4-hydroxytamoxifen (4OHT) bound to the human estrogen receptor alpha (hERα) ligand-binding domain (LBD) retrieved from the Protein Data Bank (PDB ID code 3ERT). Subsequent geometry optimization of each compound was performed in Spartan'02 at semi-empirical PM3 theory level.

From the equilibrium geometry, 13 parameters were calculated using Spartan'02 and five using Hyperchem 7.52 (Wavefunction Inc.). Also, eight parameters were obtained using ACD/LABS 9.00 (Advanced Chemistry Development Inc.) and two using literature values [12,15]. These parameters are depicted in Table 1.

Table 1. Abbreviation and source of the evaluated parameters as QSAR descriptors

Abbreviation	Parameter (unit)	Source
S^o	Standard entropy (kcal/mol °C)	Spartan'02
G^o	Standard Gibbs free energy (kcal/mol)	
H^o	Standard formation enthalpy (kcal/mol)	
MW	Molecular weight (a.m.u)	
E	Energy (kcal/mol)	
E_{aq}	Solvation energy (kcal/mol)	
μ	Dipole moment (debye)	
q_{C4}	Electrostatic charge at C_4 of aromatic ring	
E_{HOMO}	HOMO energy (kcal/mol)	
E_{LUMO}	LUMO energy (kcal/mol)	
VolCPK	CPK volume ($Å^3$)	
AreaCPK	CPK area ($Å^2$)	
Biophore[a]	Biophore length (Å)	
MR	Molar refractivity ($Å^3$/mol)	Hyperchem 7.52
RMS	Root mean square deviation of the distance from substituted derivative to 4OHT	
Polariz	Polarizability ($Å^3$)	
E.Hidr	Hydration energy (kcal/mol)	
CLogP	Calculated Log P	ACD90
PSA	Polar surface area	
FRB	Free rotatable bonds	
Parachor	Parachor (cm^3)	
D	Density (g/cm^3)	
TS	Superficial tension (dyne/cm)	
HBD	Number of hydrogen bond donors	
HBA	Number of hydrogen bond acceptors	
σ_p	Hammett constant	[15]
I_{cis}	I_{cis} = 1 for *trans* and 0 for *cis* analogs	[12]

[a] *This parameter is the distance between the para-substituent and the first ethylene carbon atom. In the 17-β-estradiol series, this 2D distance descriptor between the para-hydroxyl and the ethylenic carbon was supposed to be associated with the estrogenicity when it reaches 6Å [16].*

The QSAR equations were obtained by the stepwise Multiple Linear Regression (MLR) analysis using the SPSS 16.0 software (SPSS Inc.). The HBD (number of hydrogen bond donors) parameter was excluded from the analysis, since only one compound (**1h**) possess this value different from zero.

Analysis of the plot of experimental Log RBA x experimental Log RBA shows an acceptable dispersion of experimental data in Figure 2.

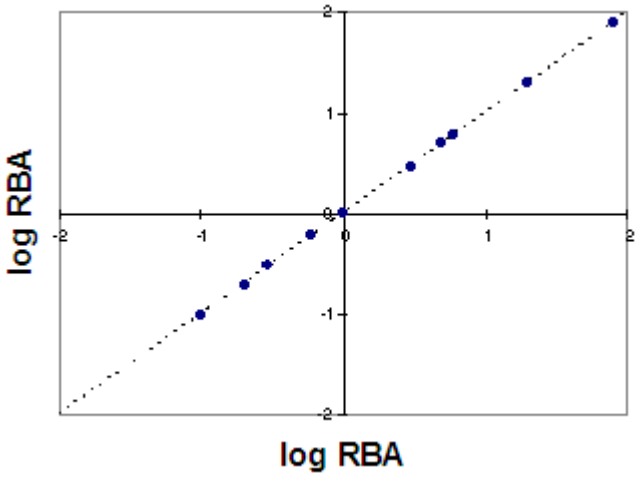


Figure 2. Log RBA x Log RBA.

Table 2 shows the calculated parameters S^0, MR, EnHidr and FRB, which were considered relevant by MLR analysis in the five best equations (Eqs. **1-5**).

Table 2. Experimental Relative Binding Affinities (RBA), Log RBA and calculated parameters S^0 (kcal/mol°C), EnHidr (kcal/mol), MR (Å^3/mol) and FRB for compounds 1a-k and 2a-g

#[a]	R	RBA	Log RBA	S^0	EnHidr	MR	FRB
1a	H	1	0.0000	0.1818	-2.21	0.1291	8
1b	I	5	0.6990	0.1913	-1.94	0.1207	8
1c	Br	5	0.6990	0.1922	-1.88	0.1227	8
1d	Cl	1	0.0000	0.1888	-1.88	0.1292	8
1e	SMe	3	0.4771	0.1990	-2.29	0.1450	9
1f	SOMe	3	0.4771	0.2028	-3.69	0.1440	9
1g	SO_2Me	6	0.7782	0.2070	-4.8	0.1388	9
1h	SH	1	0.0000	0.1916	-3.88	0.1381	9
1i	CHO	20	1.3010	0.1936	-4.17	0.1247	9
1j	CH_2OH	80	1.9031	0.1949	-7.69	0.1353	10
1k		3	0.4771	0.1965	-3.69	0.1373	9
2a	H	0.1	-1.0000	0.1397	-2.40	0.1291	8
2b	I	0.2	-0.6990	0.1462	-1.94	0.1207	8
2c	Br	0.2	-0.6990	0.1522	-2.13	0.1227	8
2e	SMe	0.1	-1.0000	0.1563	-2.53	0.1450	9
2f	SOMe	0.3	-0.5229	0.1623	-3.86	0.1440	9
2g	SO_2Me	0.6	-0.2218	0.1628	-4.83	0.1388	9

[a] *Compound **2d**, the cis-substituted analog of **1d**, was not reported in original reference [12].*

Log RBA = -5.406 (± 0.982) + 30.919 (± 5.422) S^0 **(Eq. 1)**

Log RBA = -5.262 (± 0.829) + 26.833 (± 4.815) S^0 – 0.180 (± 0.067) EnHidr **(Eq. 2)**

Log RBA = -0.981 (± 1.056) + 28.296 (± 3.069) S^0 – 0.256 (± 0.046) EnHidr – 35.972 (± 7.699) MR **(Eq. 3)**

Log RBA = -3.467 (± 1.295) + 26.567 (± 2.744) S^0 – 0.072 (± 0.087) EnHidr – 55.551 (± 10.646) MR + 0.695 (± 0.296) FRB **(Eq. 4)**

Log RBA = -4.362 (± 0.868) + 26.279 (± 2.688) S^0 – 60.822 (± 8.419) MR + 0.913 (± 0.131) FRB **(Eq. 5)**

The statistical parameters for equations **1-5** can be found in Table 3. In this case, one can see that values are statistically significant as Fisher test associated *p*-values are less than 0.0001.

Table 3. Statistical parameters of equations 1-5

Eq.	Terms	R^2	R^2_{adj}	s	F	p <	Q^2	Q^2_{adj}	S_{PRESS}
1	1	0.684	0.663	0.4692	31.521	0.0001	0.6843	0.6843	0.1211
2	2	0.791	0.761	0.3952	26.495	0.0001	0.7910	0.7771	0.1056
3	3	0.922	0.904	0.2505	51.219	0.0001	0.9220	0.9108	0.0695
4	4	0.947	0.929	0.2158	53.170	0.0001	0.9466	0.9343	0.0623
5	3	0.944	0.931	0.2131	72.430	0.0001	0.9435	0.9355	0.0591

Equations **4** and **5** present the higher adjusted R^2 values. These equations also show comparable standard deviations and adjusted Q^2 values. However, equation **5** shows higher F test value, lower S_{PRESS} and is composed by three variables (terms or parameters), while equation **4** has four variables, thus suggesting by parsimony principle [17] that equation **5** is superior to equation **4**.

In addition, the analysis of the cross-correlation matrix of the parameters of equations **1-5** (Table 4) suggest a high correlation degree (>0.7) between FRB, EnHidr and MR. This fact corroborates the choice of equation **5** as the best model.

Table 4. Cross-correlation matrix of parameters of equations 1-5

Eq.	Parameters	Sº	EnHidr	MR	FRB
1	Sº	1			
2	Sº	1			
	EnHidr	0.32	1		
3	Sº	1			
	EnHidr	0.26	1		
	MR	-0.10	0.36	1	
4	Sº	1			
	EnHidr	-0.13	1		
	MR	0.15	-0.60	1	
	FRB	-0.27	**0.89**	**-0.78**	1
5	Sº	1			
	MR	0.09		1	
	FRB	-0.35		-0.69	1

Using model **5**, the plot of the experimental *versus* predicted Log RBA (Figure 3) and residues (Table 5) were evaluated. Both analyses show evidences of **1h** as outlier, whose residue is larger than twice the standard deviation of the residues (SD_{res})

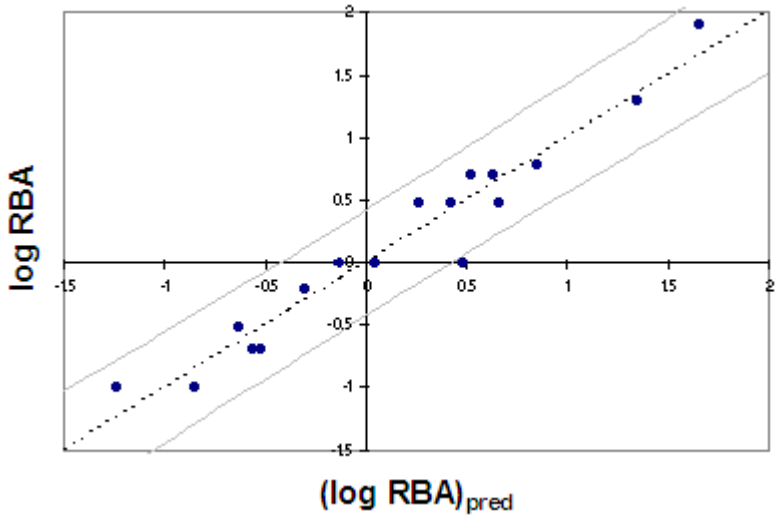


Figure 3. Experimental Log RBA *versus* calculated (Log RBA)$_{pred}$ by equation 5

Table 5. Compounds residues analysis according to equation 5

#	R	Log RBA	(LogRBA)$_{pred}$	Residue
1a	H	0.00	-0.13	-0.13
1b	I	0.70	0.63	-0.07
1c	Br	0.70	0.53	-0.17
1d	Cl	0.00	0.05	0.05
1e	SMe	0.48	0.27	-0.21
1f	SOMe	0.48	0.43	-0.05
1g	SO_2Me	0.78	0.85	0.08
1h	SH	0.00	0.49	**0.49**
1i	CHO	1.30	1.36	0.05
1j	CH_2OH	1.90	1.66	-0.24
1k	$CH(O)CH_2$	0.48	0.67	0.19
2a	H	-1.00	-1.24	-0.24
2b	I	-0.70	-0.56	0.14
2c	Br	-0.70	-0.52	0.18
2e	SMe	-1.00	-0.86	0.14
2f	SOMe	-0.52	-0.64	-0.11
2g	SO_2Me	-0.22	-0.31	-0.09
			SD$_{res}$	**0.19**

It is suggested that as **1h** is the only compound possessing a hydrogen bond donor moiety, the adopted model would not be able to detect this characteristic due to the absence of similar substructures in the sample. In fact, as already pointed, **1h** was the only compound in the sample to have HBD parameter different from zero.

Excluding the outlier **1h**, coefficients were recalculated for equation **5** furnishing the equation **6** (equation **6** and Table 6), in which analysis of residues (Table 7) and plot of Log RBA x (Log RBA)$_{pred}$ (Figure 4) did not show any outlier. In addition, the comparison of the statistical parameters between Selassie's model (excluding outlier **1j**) [12] and model **6** (Table 6) indicate that the present model is superior to the former one.

Log RBA = -4.698 (± 0.684) + 26.879 (± 2.099) S^0 - 59.612 (± 6.558) MR + 0.924 (± 0.102) FRB **(Eq. 6)**

Table 6. Statistical parameters of model 6 and comparison with Selassie model [12]

Model	Terms	R^2	R^2_{adj}	s	F	p <	Q^2	Q^2_{adj}	S_{PRESS}
Selassie[a]	2	0.835	0.812	0.3011	33.316	0.0001	0.779[c]	0.7395	0.1031
6[b]	3	0.968	0.961	0.1657	122.658	0.0001	0.9684	0.9639	0.0442

[a] *Excluding outlier **1j***
[b] *Excluding outlier **1h***
[c] *The published value of Q^2 was 0.779,[12] but in our calculations Q^2 = 0.7569, probably due to different adopted values of Hammett constants.*

Table 7. Compounds residues analysis by equation 6

#	R	log RBA	$(logRBA)_{pred}$	Residue
1a	H	0.00	-0.11	-0.11
1b	I	0.70	0.64	-0.06
1c	Br	0.70	0.55	-0.15
1d	Cl	0.00	0.07	0.07
1e	SMe	0.48	0.32	-0.15
1f	SOMe	0.48	0.49	0.01
1g	SO_2Me	0.78	0.91	0.13
1i	CHO	1.30	1.39	0.08
1j	CH_2OH	1.90	1.71	-0.19
1k		0.48	0.71	0.24
2a	H	-1.00	-1.24	-0.24
2b	I	-0.70	-0.57	0.13
2c	Br	-0.70	-0.53	0.17
2e	SMe	-1.00	-0.82	0.18
2f	SOMe	-0.52	-0.60	-0.08
2g	SO_2Me	-0.22	-0.28	-0.06
			SD_{res}	0.15

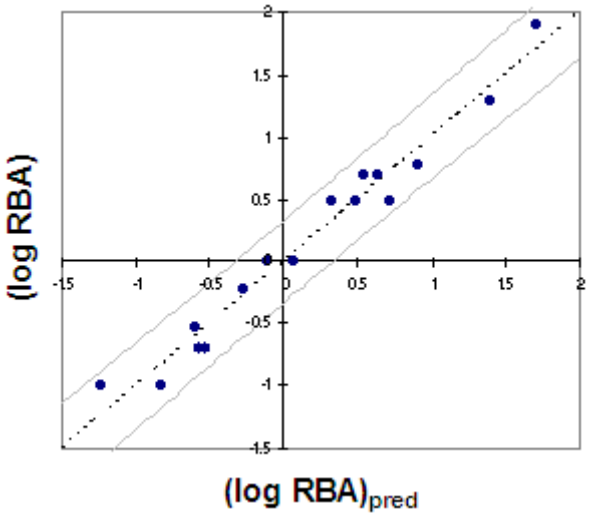


Figure 4. Experimental Log RBA *versus* predicted $(Log\ RBA)_{pred}$ by equation 6

In model **6**, the physical significance of the negative term of molar refractivity (MR) could indicate that lipophilic substituents decrease the activity while hydrophilic groups increase it, a proposition supported by experimental results that 4-hydroxytamoxifen (4OHT) is a metabolite more active in terms of RBA than tamoxifen itself [18].

The descriptors standard entropy (S^0) and free rotatable bonds (FRB) could indicate that compounds possessing some structural flexibility can be better accommodated to the receptor ER_α, thus increasing activity of these compounds.

As a conclusion, it was possible to develop a new QSAR model for a series of *para*-substituted *trans*- and *cis*-tamoxifen derivatives which has proven to be more robust than the reported one. The new model can be employed to drive the synthesis of new tamoxifen derivatives from **1b-d** using palladium mediated C-C coupling.

Acknowledgements
Authors acknowledge Brazilian agencies CAPES, CNPq and FAPERJ for financial support.

REFERENCES

[1] M Dowsett, A Howell. *Nat. Med.*, **2002**, 8(12), 1341-1344.
[2] J Peng; S Sengupta; VC Jordan. *Anticancer Agents Med. Chem.*, **2009**, 9(5), 481-499.
[3] H Al-Inany; N Johnson. *Br. Med. J.* **2006**, 332 (7556), 1461-1462.
[4] R Vardanyan; V Hruby. Synthesis of Essential Drugs, 1st Edition, Elsevier, Amsterdam, **2006**; 371–373.
[5] PL Coe; CE Scriven. *J. Chem. Soc., Perkin Trans. 1*, **1986**, 475-477.
[6] CM Nunes; J Limberger; S Poersch; M Seferin; AL Monteiro. *Synthesis*, **2009**, (16), 2761-2765.
[7] S Cacchi; G Fabrizi; A Goggiamani; D Persiani. *Org. Lett.*, **2008**, 10(8), 1597-1600.
[8] C Zhou; RC Larock. *J. Org. Chem.* **2006**, *71(8)*, 3184-3191.
[9] RA Pilli; LG Robello. *J. Braz. Chem. Soc.*, **2004**, 115(6), 938-944.
[10] CH Oh; HH Jung; KS Kim; N Kim. *Angew. Chem. Int. Ed.*, **2003**, 42(7), 805-808.
[11] R McCague; G Leclercq; N Legros; J Goodman; GM Blackburn; M Jarman; AB Foster; *J. Med. Chem.*, **1989**, 32(12), 2527-2533.
[12] CD Selassie; R Garg; S Mekapati. *Pure Appl. Chem.*, **2003**, 75(11-12), 2363-2373.
[13] EFF Cunha; RCA Martins; MG Albuquerque; RB Alencastro. *J. Mol. Model.*, **2004**, 10(4), 297-304.
[14] JC Barros; PG Lima; ALF Souza; JFM Silva; OAC Antunes. *Appl. Organometal. Chem.*, **2008**, 22(5), 249-252.
[15] C Hansch; A Leo; D Hoekman. Exploring QSAR: Volume 2: Hydrophobic, Electronic, and Steric Constants, 1st Edition, American Chemical Society, Washington, **1995**.
[16] G Klopman; SK Chakravarti. *Chemosphere*, **2003**, 51(6), 445-459.
[17] SH Unger; C Hansch. *J. Med. Chem.*, **1973**, 16(7), 745-749.
[18] GGJM Kuiper; PJ Shughrue; I Merchenthaler; J-Å Gustafsson. *Front. Neuroendocrinol.*, **1998**, 19(4), 253-286.