QSAR Studies with E-State Index: Predicting Pharmacophore Signals for Estrogen Receptor Binding Affinity of Triphenylacrylonitriles

Subhendu MUKHERJEE, Arup MUKHERJEE, and Achintya SAHA*

Department of Chemical Technology, University of Calcutta; 92 A.P.C. Road, Kolkata-700009, India.
Received September 21, 2004; accepted October 28, 2004

In connection to developing non-steroidal estrogen analogs, the present study explores the pharmacophore of triphenylacrylonitriles (Fig. 1) for binding affinity to estrogen receptor using Electrotopological State (E-State) indices of constituting atoms. The analysis shows the efficacy of E-State index in developing statistically acceptable model, which defines the electronic environment and topological states of diverse atoms in a molecule. The investigation concluded that electrophilic substitutions at C_6 and C_{18} of the phenyl rings (A and C rings respectively) attached to C_2 and C_1 of ethylenic moiety, along with presence of hydroxyl substitution at C_{12} (ring B) and no. of non-hydrogen free terminal atoms of the molecule have influence on the binding affinity to the estrogen receptor.

Key words QSAR; triphenylacrylonitrile; estrogen receptor; electrotopological state

Estrogens are endocrine regulators of both the male and female reproductive systems like the mammary gland, uterus, ovary, testis and prostate. They also play essential roles in non-target tissues, *e.g.*, bone, liver, or in the cardiovascular system, where estrogens have protective actions.[1—3] Such effects on different organs are the result of the interaction with the estrogen receptor (ER), of which two subtypes (ER_α and ER_β) are presently known.[4,5] The interaction with ER is also involved for the treatment in a range of diseases such as breast cancer, osteoporosis, endometrial cancer, and prostate hypertrophy.[6] ER is a member of the nuclear hormone receptor super family.[7] The principal endogenous ligand for ER in most species is 17β estradiol (E_2). Over the years, many synthetic ligands have been investigated as potential analogues of E_2,[8,9] in the search for agents that would be useful in regulating fertility, in preventing and treating breast cancer and other therapeutic areas. Selective estrogen receptor modulators (SERMs) that show tissue-dependent agonistic or antagonistic behavior, are used as first line treatment for estrogen-responsive breast cancer and for therapy against osteoporosis.[10,11] This pure estrogen antagonists (anti-estrogens) are currently in clinical development for breast cancer treatment.[12]

Several estrogenic triphenylethylenes having an NO_2, Cl or ethyl fragment as fourth substituent on the ethylene have been investigated as therapeutic agents.[13] We had earlier established some basic pharmacophore features of different triphenylethylenes with trifluoromethyl as the fourth substituent.[14] Consequently, the present work was taken up as continuation of the previous attempt in defining pharmacophore signals of different triphenylethylenes and is based on a series of triphenylethylenes with CN as the fourth substituent. The estrogenic activities exhibited by hydroxy, methylated substituents and bulky functional groups on different triphenylacrylonitrile derivatives have been reported.[15,16] The present study explores selectivity requirements of such diverse sequence of hydroxylated and non-hydroxylated triphenylacrylonitriles, including compounds with bulky hydrophobic substituents for binding to calf uterus estrogen receptor (ER).[17]

Structure–activity relationships has been drawn, investigating topological features of constituting atoms of the triphenylacrylonitrile molecular architecture for characterization of unique pharmacophore features.

Topological models directly give structural information to guide design of new molecules[18] and may predict three-dimensional structural parameters, as well.[19] The electrotopological state (E-State) of atoms has proved to be a significant tool in elucidating major drug-receptor interactions.[20—22] An atom in a molecule is part of a field of information with regard to electronic influences and topological surrounding.[20,23] Quantification of influence of this field on any atom, can correlate to the biological performance of a molecule. The contribution of an atom can be expressed as the Electrotopological State (E-State),[24] mathematically defined as

$$S_i = I_i + \Delta I_i \text{ where, } I = [(2/N)^2 \delta^v + 1]/\delta$$

and

$$\Delta I_i = \Sigma (I_i - I_j)/r^2$$

I is derived from Molecular connectivity[25] calculations and called the intrinsic state of an atom, ΔI_i is the molecular environmental effect, N is the principal quantum number, δ is the number of sigma electrons on the atom (excluding those bonding to hydrogen), δ^v is the number of valence electrons, i and j are serial numbers of atoms and r is the shortest graph distance between two atoms.

RESULTS AND DISCUSSION

From the regression analysis it was observed that the best

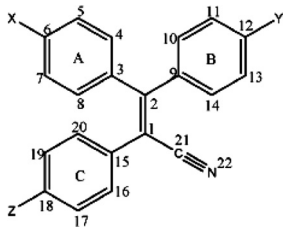
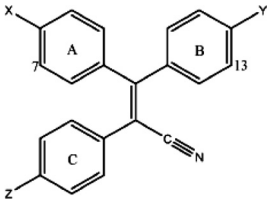


Fig. 1. General Structure of Triphenylacrylonitriles: The common atoms have been numbered 1 through 22.

* To whom correspondence should be addressed. e-mail: achintya_Saha@yahoo.com

Table 1. Observed, Calculated and Predicted Biological Activities (Relative Binding Affinity to Estrogen Receptor) of Triphenylacrylonitriles



Comp. No.	Substitutions			Log RBA (competition for [^{3}H] E_2 binding)			
	X	Y	Z	Obs.	Cal.[a)]	Cal.[b)]	Pred.[b)]
1	H	H	H	−1.046	−0.107	−0.928	−0.916
2	H	OH	H	1.556	1.256	1.008	0.873
3	OH	H	H	0.342	0.695	0.148	0.123
4	H	H	OH	0.519	0.299	−0.280	−0.488
5	OH	OH	H	1.792	2.060	2.085	2.160
6	H	OH	OH	1.869	1.663	1.657	1.579
7	OH	H	OH	0.785	1.102	0.798	0.801
8	OH	OH	OH	2.220	2.468	2.735	2.929
9	CH_3	OH	H	1.447	1.095	1.186	1.143
10	OH	CH_3	H	0.398	0.135	−0.046	−0.104
11	OH	OH	CH_3	1.968	1.781	2.136	2.168
12	CH_3	OH	OH	1.892	1.503	1.836	1.819
13	OH	CH_3	OH	0.959	0.543	0.604	0.480
14	OH	OCH_3	H	−0.180	−0.822	−0.938	−1.017
15	OCH_3	OH	H	1.230	1.458	1.525	1.568
16	OH	OiPr	H	−0.444	−0.417	−0.233	−0.200
17	OiPr	OH	H	0.806	0.894	1.327	1.401
18	OiPr	OiPr	H	−2.000	−1.582	−0.992	−0.799
19	OH	DEAE	H	0.531	−0.418	−0.235	−0.354
20	DEAE	OH	H	2.033	0.887	1.320	1.218
21	DEAE	DEAE	H	−0.398	−1.591	−1.000	−1.115
22	H	H	DEAE	−2.000	−0.698	−0.855	−0.616
23	DMA	DMA	H	−1.398	−1.906	−1.293	−1.266
24	OH	OH	H	−2.000	−0.250	—	—
	7 DMAM	13 DMAM					
25	OCH_3	OCH_3	H	−1.398	−0.563	−0.697	−0.644

Obs.=observed value; Cal.=calculated value; Pred.=predicted value; *a*) From Eq. (1); *b*) From Eq. (2). OiPr=$OCH(CH_3)_2$; DEAE=$O(CH_2)_2N(C_2H_5)_2$; DMA=$N(CH_3)_2$; DMAM=$CH_2N(CH_3)_2$.

significant relation developed with single variate was S_{21} (r=0.738) that explained 52.56% variance. This supports that presence of CN group as fourth substituent of triphenylethylene has significant influence on binding affinity to the receptor. But the overall best relation for binding affinity to ER involving 25 compounds using Multiple Regression Method was found to be

$$\text{Log RBA} = 1.261\,(\pm 0.491) + 1.906\,(\pm 0.317)\,I_{12\text{-OH}} - 0.239\,(\pm 0.077)\,S_6 - 0.169\,(\pm 0.065)\,S_{18} - 0.559\,(\pm 0.143)\,N_t \quad (1)$$

n=25, R=0.860, R^2=0.740, EV=68.811%, F=14.238 (df 4, 20), s=0.774, AVRES=0.542, PRESS=21.751, SDEP=0.933, Pres_{av}=0.697, Q^2=0.529.

Compound **24** behaved as an outlier and removed, the resultant equation was

$$\text{Log RBA} = 2.114(\pm 0.243)I_{12\text{-OH}} - 0.223(\pm 0.059)S_6 - 0.147(\pm 0.050)S_{18} - 0.193(\pm 0.052)N_t \quad (2)$$

n=24, R=0.919, R^2=0.845, EV=81.415%, F=27.284 (df 4, 20), s=0.595, AVRES=0.450, PRESS=10.021, SDEP=0.646, Pres_{av}=0.515, Q^2=0.751.

where S_6 and S_{18} are E-States of C_6 and C_{18} respectively. $I_{12\text{-OH}}$ and N_t are indicator variables signifying correspondingly presence or absence of –OH substitution in C_{12} and no. of free terminal atoms (excluding H) that indicates the degree of substitutions in the phenyl rings. For *e.g.*, in compound **16** (Table 1), N_t=4 (1 oxygen atom of OH, 2 carbons of $-CH_3$ in $OCH(CH_3)_2$ linkage and 1 nitrogen in C≡N). The 95% confidence intervals are shown in parentheses and the F-values are significant at 99% confidence level. The independent variables used are not intercorrelated ($|r| \leq 0.5$). Removal of outlier showed the intercept to be statistically insignificant and omission of the same did not affect the quality of the equation. The calculated and predicted activities obtained from the Eqs. (1) and (2) are delineated in Table 1 and Fig. 2.

The best relation developed with S_6, S_{18}, $I_{12\text{-OH}}$ and N_t (Eq. 1) that explained 68.81% variance in binding affinity. Removing an outlier, the relation (Eq. 2) explained 81.42% variance with 75.10% cross-validated variance. The model shows the importance of atoms C_6 and C_{18}, presence of hydroxyl substitution at C_{12} along with the no. of free terminal atoms of the molecules contribute towards major role in binding to the estrogen receptor. In the models, S_6 has nega-

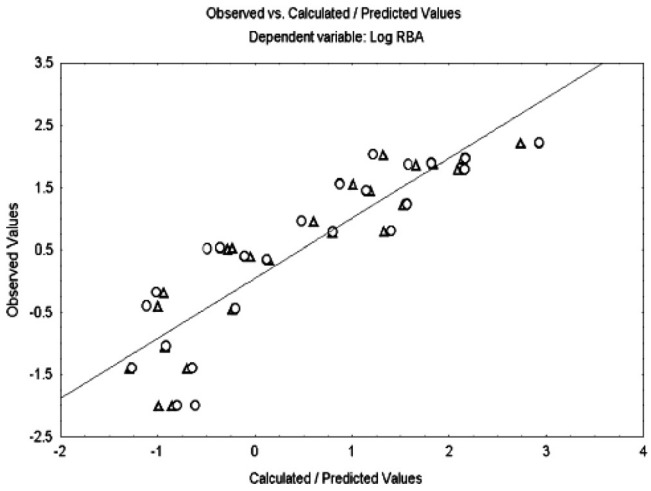


Fig. 2 Plot of Calculated and Predicted Values as per Eq. (2)
Δ, calculated value; ○, predicted value.

tive coefficient, so substitutions that tend to decrease the magnitude of S_6 will be favored. Compound **3** with hydroxyl (e^- withdrawing) substitution at C_6 causing decrease in electron density around the atom resulted in marked increase in activity as compared to compound **1**, where X=H. Furthermore it has been found earlier that hydroxy substitution in X, Y or Z enhances the receptor binding affinity of non-steroidal estrogen analogs[26] and loss of this H-bonding capacity of phenolic OH group results in loss of RBA to the estrogen receptors.[27] Thus electrophilic substitution at C_6 is preferential. E-State value at C_{18} also has negative contribution, so e^- withdrawing substitutions at Z must also be favored, as this will tend to decrease the e^- density around C_{18} and correspondingly the E-State value. Compound **4** with hydroxyl (e^- withdrawing) substitution in C_{18} resulted in higher activity than compound **1** (Z=H). From the result, it was also observed that presence of hydroxyl substitution at C_{12} has influence on binding affinity to ER. This is evident from the relative binding affinity exhibited by compound **8** with hydroxyl substitution at Y as compared to compound **5**. Compound **8** exhibits the highest binding affinity in the series while compound **5** with Y=H exhibits much lesser binding affinity. From the regression, it was also found that the factor N_t has negative contribution towards binding affinity, so compounds with lesser number of terminal atoms (excluding H) should also result in increased binding affinity. In general compounds with more branched side chain substitutions resulting in more N_t (Compounds **21**, **23**, **25**) has reduced binding affinities in comparison to rest of the series.

MATERIALS AND METHODS

A diverse series of 25 molecules tested for estrogen receptor binding belonging to the triphenylacrylonitrile derivatives[17] were used in this study. Biological activity (estrogenic activity) is expressed in terms of logarithm of relative binding affinity to the ER *vis-à-vis* E_2 (Log RBA). The E-States of different atoms were calculated by means of a JAVA2 based program *ETSA-CS*,[28] which was standardized using established sets of data. Statistical analysis was performed by *Statistica version 5.0*[29] using standard and forward stepwise multiple regression methods. Statistical parameters of the regression equation considered are: r or R (correlation coefficient), EV (explained variance), F (variance ratio) with df (degree of freedom), s (standard error of estimate) and AVRES (average of absolute values of residuals). Leave-One-Out (LOO) cross-validation[30] was performed that generated PRESS (predictive residual sum of squares), SDEP (standard deviation of error of predictions), $Pres_{av}$ (average of absolute value of predicted residuals) and Q^2 (cross-validated variance). Significance of the regression coefficients was assessed by the 't' test. In case the intercept of an equation was statistically insignificant and omission of the same did not affect the quality of the equation, exclusion of the intercept was achieved to get statistically more acceptable equation. A compound was considered as outlier in the equation when the residual value exceeded twice the standard error of estimate.

CONCLUSION

In view of these observations, the present study could account for some of the structural requirements of triphenylacrylonitriles for ER binding affinity. The studies support that *para* electrophilic substitutions in the phenyl rings (A and C respectively) along with presence of *para* hydroxyl group in the phenyl ring B (attached to C_2 and C_1 of ethylenic moiety), which are capable of active hydrogen bonding with the receptor, are essential for activity. The studies also support the presence of less terminal atoms (less branched chains) should result in more active compounds.

REFERENCES AND NOTES

1) Rosano G. M., Sarrel P. M., *Cardiologia*, **39**, 275—279 (1994).
2) Inestrosa N. C., Marzolo M. P., Bonnefont A. B., *Mol. Neurobiol.*, **17**, 73—86 (1998).
3) Lindsay R., *Endocrinol. Metab. Clin. North Am.* (US), **27**, 399—409 (1998).
4) Kuiper G. G. J. M., Enmark E., Pelto-Huikko M., Nilsson S., Gustafsson J. Å., *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 5925—5930 (1996).
5) Barkhem T., Carlsson B., Nilsson Y., Enmark E., Gustafsson J. Å., Nilsson S., *Mol. Pharmacol.*, **54**, 105—112 (1998).
6) Levenson A. S., Jordan V. C., *Eur. J. Cancer*, **35**, 1628—1639 (1999).
7) Tsai M. J., O'Malley B. W., *Ann. Rev. Biochem.*, **63**, 451—486 (1994).
8) Gao H., Katzenellenbogen J. A., Garg R., Hansch C., *Chem. Rev.*, **99**, 723—744 (1999).
9) Anstead G. M., Carlson K. E., Katzenellenbogen J. A., *Steroids*, **62**, 268—303 (1997).
10) Dutertre M., Smith C. L., *J. Pharmacol. Exp. Ther.*, **295**, 431—437 (2000).
11) Rosati R. L., Jardine P. D. S., Cameron K. O., Thompson D. D., Ke H. Z., *J. Med. Chem.*, **41**, 2928—2931 (1998).
12) Howell A., Osborne C. K., Morris C., Wakeling A. E., *Cancer*, **89**, 817—825 (2000).
13) Harper M. J. K., Walpole A. L., *Nature* (London), **212**, 87 (1966).
14) Mukherjee S., Mukherjee A., Saha A., *Bioorg. Med. Chem. Lett.*, **14**, 897—900 (2004).
15) Miquel J. F., Sekera A., Chaudron T., *C. R. Acad. Sci.* (Paris), **286**, 151— (1978).
16) Pons M., Michel A., Crastes de Paulet A., Gilbert J., Miquel J. F., Précigoux G., Hospital M., Ojasoo T., Raynaud J. P., *J. Steroid Biochem.*, **20**, 137—145 (1984).
17) Bignon E., Pons M., Crastes de Paulet A., Doré J. C., Gilbert J., Abecassis J., Miquel J. F., Ojasoo T., Raynaud J. P., *J. Med. Chem.*, **32**, 2092—2103 (1989).

18) Rose K., Hall L. H., *SAR QSAR Environ. Res.*, **14**, 113—129 (2003).
19) Estrada E., Molina E., Perdomo-Lopez I., *J. Chem. Inf. Comput. Sci.*, **41**, 1015—1021 (2001).
20) Hall L. H., Mohney B., Kier L. B., *Quant. Struct.–Act. Relat.*, **10**, 43—51 (1991).
21) Huuskonen J., *Chemosphere*, **51**, 949—953 (2003).
22) Roy K., Chakraborty S., Saha A., *Bioorg. Med. Chem. Lett.*, **13**, 3753—3757 (2003).
23) Kier L. B., Hall L. H., *Pharm. Res.*, **7**, 801—807 (1990).
24) Kier L. B., "Chemometric Methods in Molecular Design (Methods and Principles in Medicinal Chemistry)," Vol. 2, ed. by Waterbeemd H. V. de, VCH Verlagsgesellschaft mbh, Weinheim, 1995, pp. 39—44.
25) Kier L. B., Hall L. H., *J. Pharm. Sci.*, **70**, 583—589 (1981).
26) Jordan V. C., Mittal S., Gosdon B., Koch R., Lieberman M. E., *Environ. Health Perspect.*, **61**, 97—110 (1985).
27) Fang H., Tong W., Shi L. M., Blair R., Perkins R., Branham W., Hass B. S., Xie Q., Dial S. L., Moland C. L., Sheehan D. M., *Chem. Res. Toxicol.*, **14**, 280—294 (2001).
28) Developed at Cyber-Mate, Srirampur, Hooghly, West Bengal (India).
29) Brand name of StatSoft, Inc., Tulsa (USA).
30) Wold S., Eriksson L., "Chemometric Methods in Molecular Design (Methods and Principles in Medicinal Chemistry)," Vol. 2, ed. by Waterbeemd H. V. de, VCH Verlagsgesellschaft mbh, Weinheim, 1995, pp. 312—318.