Indian Journal of Chemistry
Vol. 46B, July 2007, pp. 1148-1153

Structural feature indenture for estrogen analogs as anticancer agents: *De novo* and Hansch Approach

Arun Kumar Gupta*[1], Anurekha Jain[1], Avijeet Jain[1], Anirudh Joshi[1] & S Revathi[2]

[1]Department of Pharmaceutical Chemistry, B. R. Nahata College of Pharmacy, Mhow-Neemach Road, Mandsaur (M P)

E-mail: arunkg_73@hotmail.com

[2]Smriti College of Pharmaceutical Education, Nipania, Indore (M P)

Received 7 November 2005; accepted (revised) 14 February 2007

Quantitative structure activity relationships have been performed on a series of twenty-six compounds of estrogen derivatives, for their anti-proliferative activity, in order to understand the essential structural requirement for inhibition of proliferation in estrogen dependent MCF-7 human breast cancer cells. The quantitative models derived for the study illustrates the significance of the β-hydroxy group at 17th position for the drug–enzyme interaction and allowed hydrogen-bonding interaction to estrogenic receptor in optimal manner (free rotation) as compared to oxo group (restricted orientation). The results of the study also reveal the necessity of sulfonamide moiety at 3rd position of estrogen. Additionally, presence of smaller substituents at 2nd position will be conducive for the activity.

Keywords: Estrogen analogs, Hansch approach, Fujita-Ban analysis, QSAR, antiproliferative activity, anticancer agents

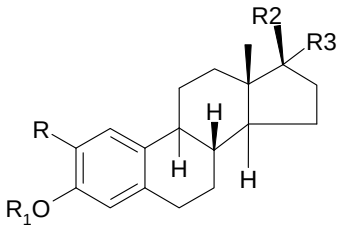
Cancer is the second leading cause of mortality in developed countries, and the discovery and development of new treatment is urgently needed due to problem with currently available treatment such as toxicities and drug resistance[1]. Many of current clinically used anticancer agents including paclitaxel, vinblastin[2-4] and compound under clinical evaluation, such as epothiolone[5] are microtublar interfering agents that bind to tubulin. Several new anticancer agents are developed recently such as Gleenec[6] which is kinase inhibitor designed as more selective with less toxicity. Similarly, discovery of the endogenous estrogen metabolite 2-methoxy estradiol (2-MeOE2) would inhibit cancer cell proliferation and angiogenesis[7] and it is 250 times more potent than estradiol. A number of potential mechanisms underlying the reversible mitotic arrest[8] and apoptosis[9-10] induced by 2-methoxy estadiol in various cancer cell lines have been proposed.2-Substituted estrone sulfamates (EMATEs) displays the similar antiproliferative activity profile to the corresponding estradiol against a range of human cancer cell lines and have 80 fold greater activity than the 2-MeOE2. 2-substituted EMATEs also inhibit the steroid sulfatase, a therapeutic target for the treatment of hormone dependent breast cancer[11]. The 2-substituted EMATEs remain to be elucidated their spectrum of biological activity encouraged us to explore the closely related structures in the search for the new anticancer compounds.

These arouse our interest for designing of novel hormone dependent breast anti-cancer agents. In the present work, we have attempted to quantify the necessary structural and physicochemical requirements for antiproliferative activity against MCF-7 human breast cancer cells.

Materials and methods

The antiproliferative activity data of estrogen analogs were taken from the reported work of Leese *et. al.*[12] (**Table I**). The antiproliferative activity data against MCF-7 human breast cancer cells (GI_{50} in μm) was converted to negative logarithmic mole dose (pGI_{50}) in order to reduce the skewness of the data set, for quantitative structure activity relationship analysis. Initially series was subjected to Fujita-Ban analysis using multiple regression technique in order to estimate the *de novo* contribution of substituents to the activity of the molecules. Further Hansch approach was carried out to establish correlations between antiproliferative activity and various substituents constants at position R, R_1 and indicator variable Iv_2 for the presence of hydroxy group at position R_2 of the molecule (**Figure 1**). The obtained model shows presence of outlier compound on the

Table I—Analogs of estrogen and their antiproliferative activity data



Compd	R	R_1	R_2	R_3	GI_{50}[a]	pGI_{50}[b]
1	CH_3O	H	O	O	21.3	4.672
2	CH_3O	H	OH	H	2.35	5.629
3	CH_3O	SO_2NH_2	O	O	0.30	6.523
4	CH_3O	SO_2NH_2	OH	H	0.36	6.444
5	CH_3O	$SO_2N(CH_3)_2$	O	O	>10	< 5.000
6	CH_3O	$SO_2NHCOCH_3$	O	O	168	3.775
7	C_2H_5O	H	O	O	9.04	5.044
8	C_2H_5O	SO_2NH_2	O	O	4.84	5.315
9	C_2H_5O	SO_2NH_2	OH	H	0.61	6.215
10	CH_3S	H	O	O	33.4	4.476
11	CH_3S	H	OH	H	3.96	5.402
12	CH_3S	SO_2NH_2	O	O	0.40	6.398
13	CH_3S	SO_2NH_2	OH	H	0.43	6.367
14	CH_3SO	H	O	O	>10	< 5.000
15	CH_3SO_2	H	O	O	66	4.180
16	C_2H_5S	H	O	O	31.9	4.496
17	C_2H_5S	H	OH	H	23.2	4.635
18	C_2H_5S	SO_2NH_2	O	O	35.2	4.453
19	C_2H_5S	SO_2NH_2	OH	H	23.4	4.631
20	C_2H_5	H	O	O	57.6	4.240
21	C_2H_5	H	OH	H	10.5	4.979
22	C_2H_5	SO_2NH_2	O	O	0.34	6.469
23	C_2H_5	SO_2NH_2	OH	H	0.07	7.155
24	CH_3	SO_2NH_2	O	O	0.26	6.585
25	Br	H	O	O	33	4.481
26	Br	SO_2NH_2	O	O	6.95	5.158
27	I	H	O	O	38	4.420
28	I	SO_2NH_2	O	O	1.59	5.799

[a] Concentration of 50 percent antiproliferative activity data against MCF-7 human breast cancer cells in mole.

[b] Negative logarithm of GI_{50}

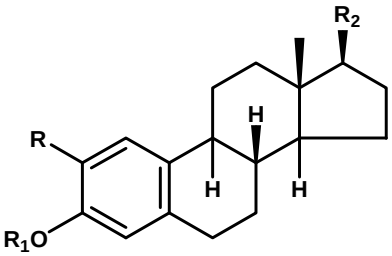


Figure 1—Lead structure of estrogen analogs used in Fujita-Ban analysis

basis of Z-score value. For better understanding of drug receptor interaction removing outlier compound and incorporated indicator variable Iv_1 for the presence of sulfonamide moiety at position R_1. Values of the substituents constants like hydrophobic (π), steric (molar refractivity or *MR*), hydrogen acceptor (*HA*), hydrogen donor (*HD*) and electronic (field effect or $\mathcal{F}$, resonance effect or $\mathcal{R}$ and Hammett's constant or σ), taken from the reported data[13,14] and Verloop parameters (value of shape of each substituent) like L, B_1, B_2, B_3, B_4 were taken from reported work of Skagerberg *et al*[15]. The data was transferred to statistical program VALSTAT[16] in order to establish the correlation between substituent constants as independent variable and antiproliferative activity as dependent variable using sequential multiple linear regression analysis method (in sequential multiple regression the program searches all permutation and combination sequentially for the data set). The ± data within the parentheses are the standard deviation, associated with coefficient of descriptors in regression equations. The best model was selected from the various statistically significant equations on the basis of observed squared correlation coefficient (r^2), standard error of estimation (SE), sequential Fischer test (F), bootstrapping squared correlation coefficient (r^2_{bs}), bootstrapping standard deviation (S_{bs}), cross validated squared correlation coefficient using leave one out procedure (q^2), chance statistics (evaluated as the ratio of the equivalent regression equations to the total number of randomized sets; a chance value of 0.001 corresponds to 0.1% chance of fortuitous correlation) and outliers (on the basis of Z-score value).

Results and Discussion

Fujita-Ban analysis was carried out in order to find out *de novo* contribution of the substituents to the activity of the molecules (**Table II**). The multivarient

Table II—Fujita-Ban matrix, calculated pGI_{50}, residual and Z-score data of estrogens analogs

Compd	R_1									R_2	R_3	Fujita-Ban Approach (eqn 2)		
	$2C_2H_5O$	$2CH_3S$	$2CH_3SO2$	$2C_2H_5S$	$2C_2H_5$	$2CH_3$	2Br	2I	$3SO_2NH_2$	$3SO_2NHAc$	17O	Cal.[a]	Cal_{res}[b]	Z[c]
1	0	0	0	0	0	0	0	0	0	0	1	4.610	0.062	0.124
2	0	0	0	0	0	0	0	0	0	0	0	5.237	0.392	0.779
3	0	0	0	0	0	0	0	0	1	0	1	5.875	0.648	1.288
4	0	0	0	0	0	0	0	0	1	0	0	6.503	-0.059	-0.117
6	0	0	0	0	0	0	0	0	0	1	1	4.610	-0.835	-1.659
7	1	0	0	0	0	0	0	0	0	0	1	4.610	0.434	0.864
8	1	0	0	0	0	0	0	0	1	0	1	5.875	-0.560	-1.114
9	1	0	0	0	0	0	0	0	1	0	0	6.503	-0.288	-0.572
10	0	1	0	0	0	0	0	0	0	0	1	4.610	-0.134	-0.266
11	0	1	0	0	0	0	0	0	0	0	0	5.237	0.165	0.328
12	0	1	0	0	0	0	0	0	1	0	1	5.875	0.523	1.039
13	0	1	0	0	0	0	0	0	1	0	0	6.503	-0.136	-0.270
15	0	0	1	0	0	0	0	0	0	0	1	4.610	-0.430	-0.854
16	0	0	0	1	0	0	0	0	0	0	1	3.607	0.889	1.767
17	0	0	0	1	0	0	0	0	0	0	0	4.235	0.400	0.796
18	0	0	0	1	0	0	0	0	1	0	1	4.873	-0.420	-0.835
19	0	0	0	1	0	0	0	0	1	0	0	5.500	-0.869	-1.729
20	0	0	0	0	1	0	0	0	0	0	1	4.610	-0.370	-0.735
21	0	0	0	0	1	0	0	0	0	0	0	5.237	-0.258	-0.513
22	0	0	0	0	1	0	0	0	1	0	1	5.875	0.594	1.180
23	0	0	0	0	1	0	0	0	1	0	0	6.503	0.652	1.297
24	0	0	0	0	0	1	0	0	1	0	1	5.875	0.710	1.411
25	0	0	0	0	0	0	1	0	0	0	1	4.610	-0.129	-0.256
26	0	0	0	0	0	0	1	0	1	0	1	5.875	-0.717	-1.426
27	0	0	0	0	0	0	0	1	0	0	1	4.610	-0.190	-0.377
28	0	0	0	0	0	0	0	1	1	0	1	5.875	-0.076	-0.152

[a] calculated negative logarithm of concentration of 50 percent antiproliferative activity data against MCF-7 human breast cancer cells in mole. [b] calculated residual value.[c] Z-score value

regression expression (Eqn 1) indicated that the many substituent having poor contribution to the antiproliferative activity, which is further supported by high standard error of the substituent coefficient.

pGI_{50} = 5.482(±0.309)-0.404(±0.400)2C_2H_5O-
0.156(±0.367)2CH_3S-0.857(±0.601)
2CH_3SO_2-1.263(±0.367)2C_2H_5S-
0.106(±0.367)2C_2H_5 +0.433(±0.604)2CH_3-
0.775(± 0.465)2Br-0.485(±0.465)2I
+1.114(±0.219)3OSO_2NH_2-1.262(±0.601)
3OSO_2NHCOCH3-0.445(±0.241)17O
n=26, r=0.912, r^2=0.832, SE=0.519, F=6.306 …(1)

The Eqn 1 was further optimized by removing the substituents which are insignificant at 95% confidence level; a statistical significant trivalent expression (Eqn 2) was obtained.

pGI_{50} = 5.237(±0.224) +1.266(±0.211)3SO_2NH_2-
1.002(±0.294)2C_2H_5S-0.627(±0.224)17O
n=26, r=0.848, r^2=0.719, SE=0.536, F=18.731 …(2)

Fujita-Ban analysis of antiproliferative activity of estrogen analogs inferred that the presence of *o*-sulfamate moiety at 3rd position of the A ring of estrogen is essential for the activity as compared to non substituted estrogen or *N*-acetylated *o*-sulfamate analogs. *De novo* techniques also suggested that the presence of *β*-hydroxy group at 17th position is favorable as compared to oxo group.

Fujita-Ban expression gave insight to some important structural features i.e. the *β*-hydroxy group of 17th position is optimal for the activity and allowed hydrogen-bonding interaction to estrogenic receptor in optimal manner (free rotation) as compared to oxo group (restricted orientation).

Series was further subjected to Hansch approach, using stepwise multiple linear regression method, in order to develop 2D-QSAR between inhibition of proliferation of MCF-7 cells an estrogen dependent human breast cancer cell line as dependent variable and substituents constants as independent variables**.** A tri-variant statistically significant equation was obtained (Eqn 3)**.**

pGI_{50} =-0.766(±0.404) MR-0.481(±0.157)π_1 +
0.693(±0.306) Iv_2 + 5.502
n=26, r=0.684, r^2=0.468, SE=0.737, F=6.442, outlier=1 … (3)

Equation 3 showed moderate correlation coefficient value 0.684 with one outlier (Compd No **6**). The unexplained variance in calculated activity of compd no. 6 might be due to the steric hindrance of *N*-substitution on sulfamate analogs. The fitness of the model can be improved by removing outlier. Hence remaining 25 compounds were considered for the QSAR analysis. The 1:1 correlation study revealed that indicator variable (Iv_1) contributed positively and linearly to the activity (Eqn 4) with 48.0% variance and statistical significance level better than 99.9% as it exceeded the student t-value 4.606 against tabulated $t_{0.001(2),\ 23}$ = 3.768, while molar refractivity contributed negatively to the activity (Eqn 5) with correlation coefficient value 0.437 and internal statistical significance level better than 95.0% ($t_{0.05(2),23}$=2.328 as it exceeded the tabulated value 2.069) and suggested that bulkier substitution at the parent structure is unfavorable for the inhibitory action.

pGI_{50} = 1.241(±0.269) Iv_1 + 4.721

n=25, r=0.693, r^2=0.480, SE=0.673, F=21.212

pGI_{50} =-1.080(±0.464) *MR*+ 6.679 … (4)

n=25, r=0.437, r^2=0.191, SE=0.840, F=5.420 … (5)

Stepwise multiple linear regression analysis method was employed for improving the fitness of expression, several multivariant significant equations were obtained, and equation 6 was considered as model on the basis of statistical parameters (**Table III**).

pGI_{50} =-1.014(±0.292) MR + 1.152(±0.211) Iv_1 +
0.553(±0.220) Iv_2+ 5.801

n = 25, r = 0.843, r^2 = 0.711, SE = 0.525, F =17.245, r^2_{bs} = 0.729, S_{bs} = 0.089, Chance < 0.001, q^2 = 0.577, S_{PRESS} = 0.635, S_{DEP} = 0.582 … (6)

Equation 6 has better correlation coefficient (r=0.843), which accounts for more than 71.0% of the variance in the activity; also the inter-correlation among the parameters is less <0.100 (**Table IV**). The P value is less than 0.001 for each physiochemical parameters involved in the multivariant model generation, suggested that each independent variable contributed linearly. The data showed overall internal statistical significance level better than 99.9% as it exceeded the tabulated $F_{(3,21\ \alpha\ 0.001)}$ = 8.99. The model was further subjected for outlier by Z-score method and no compound was found to be an outlier, which suggested that the model is able to explain the structurally diverse analogs, and is helpful in designing of more potent compounds using physiochemical

Table III—Substituent constants, calculated, calculated (loo) value with residual and Z-score data of estrogen analogs used in QSAR study

Compd	Parameters			QSAR Model				
	MR	Iv_1	Iv_2	Cal.[a]	Cal_{res}[b]	Z-Score	$Cal_{(loo)}$[c]	$Cal_{(loo)res}$[d]
1	0.787	0	0	5.003	-0.331	-0.675	5.067	-0.395
2	0.787	0	1	5.556	0.073	0.149	5.533	0.096
3	0.787	1	0	6.155	0.368	0.750	6.092	0.431
4	0.787	1	1	6.707	-0.264	-0.537	6.775	-0.331
7	1.247	0	0	4.537	0.507	1.033	4.478	0.565
8	1.247	1	0	5.688	-0.373	-0.760	5.732	-0.417
9	1.247	1	1	6.241	-0.026	-0.054	6.245	-0.031
10	1.382	0	0	4.400	0.077	0.156	4.390	0.086
11	1.382	0	1	4.952	0.450	0.916	4.865	0.537
12	1.382	1	0	5.551	0.847	1.724	5.438	0.960
13	1.382	1	1	6.104	0.262	0.534	6.058	0.309
15	1.349	0	0	4.433	-0.253	-0.515	4.464	-0.283
16	1.842	0	0	3.933	0.563	1.146	3.773	0.723
17	1.842	0	1	4.486	0.148	0.302	4.436	0.199
18	1.842	1	0	5.085	-0.632	-1.286	5.293	-0.839
19	1.842	1	1	5.638	-1.007	-2.051	5.990	-1.359
20	1.030	0	0	4.757	-0.517	-1.053	4.823	-0.584
21	1.030	0	1	5.309	-0.331	-0.673	5.383	-0.404
22	1.030	1	0	5.908	0.560	1.141	5.841	0.628
23	1.030	1	1	6.461	0.694	1.413	6.333	0.821
24	0.565	1	0	6.380	0.205	0.418	6.324	0.261
25	0.888	0	0	4.901	-0.419	-0.854	4.967	-0.486
26	0.888	1	0	6.052	-0.894	-1.821	6.181	-1.022
27	1.394	0	0	4.388	0.033	0.067	4.383	0.037
28	1.394	1	0	5.539	0.259	0.528	5.504	0.295

[a] calculated negative logarithm of concentration of 50 percent antiproliferative activity data against MCF-7 human breast cancer cells in mole.
[b] calculated residual value.
[c] calculated (leave-one-out) negative logarithm of concentration of 50 percent antiproliferative activity data against MCF-7 human breast cancer cells in mole.
[d] calculated (leave-one-out) residual value.

Table IV—Inter-correlation matrix of substituent constants used in QSAR analysis

	π	*MR*	*L*	B_1	B_2	B_3	B_4	*HA*	$\mathcal{R}$	$\mathcal{F}$	Iv_1	Iv_2
π	1.000											
MR	0.307	1.000										
L	0.176	0.866	1.000									
B_1	0.124	0.377	0.007	1.000								
B_2	0.017	0.718	0.839	0.309	1.000							
B_3	0.281	0.046	0.300	0.693	0.260	1.000						
B_4	0.590	0.110	0.045	0.634	0.188	0.719	1.000					
HA	0.759	0.352	0.039	0.521	0.067	0.266	0.177	1.000				
$\mathcal{R}$	0.082	0.313	0.042	0.674	0.027	0.647	0.538	0.628	1.000			
$\mathcal{F}$	0.395	0.192	0.151	0.575	0.183	0.353	0.569	0.276	0.042	1.000		
Iv_1	0.133	0.083	0.059	0.138	0.052	0.205	0.205	0.027	0.128	0.145	1.000	
Iv_2	0.060	0.090	0.157	0.280	0.309	0.147	0.255	0.021	0.135	0.233	0.053	1.000

parameters. The leave-one-out cross validation method was employed for the prediction of activity, and a q^2 value (in the biological activity data of leave one compound) of 0.3 corresponds to a confidence limit greater than 95%, which minimizes the risk of finding significant explanatory equation for the biological activity just by mere chance. The Cross-validated squared correlation co-efficient (q^2=0.577) suggested a good internal consistency as well as predictive ability of the biological activity. The r^2_{bs} is at par with the conventional squared correlation coefficient (r^2). Randomized biological activity test (Chance < 0.001) revealed that the results were not based on chance correlation. In general, the model fulfills the statistical validation criteria to a significant extent to be a useful theoretical base for proposing more active compounds. The Hansch analysis indicated that both indicator variables Iv_1 & Iv_2 contributed positively while molar refractivity contributed negatively to the expression. Analysis revealed that the presence of sulfonamide moiety at 3rd position (Iv_1) and β-hydroxy group (Iv_2) at 17th position is essential for the optimum activity. Molar refractivity (MR) which is representative of bulkiness/molar volume of the substituents play key role at 2nd position of the ring and suggested that small substituents are more favorable as compared to the larger one.

Acknowledgment

Authors are thankful to the Director, B R Nahata College of Pharmacy, Mandsaur for providing facilities to complete this work.

References

1 Haskel C M, *Cancer Treatment* (W B Saunders: Philadelphia, PA), **2001**.

2 Eckhardt S, *Curr Med Chem Anti-cancer Agents,* 2, **2002**, 419.

3 Lu M C, *Cancer Chemothearapeutic Agents,* (American Chemical society Washington, DC), **1995**.

4 Altmann K H, *Curr Opin Chem Biol,* 5, **2001**, 424.

5 Wartmann M, Altmann K H, *Curr Med Chem Anti-cancer Agents*, 2, **2002**, 123.

6 O'Dwyer M E & Druker B J, *Curr Cancer Drug Targets,* 1, **2001**, 49.

7 Fotsis T, Zhang Y M, Pepper M S, Adlercreutz H, Montesano R, Nawroth P P & Schweigerer L, *Nature*, 368, **1994**, 237.

8 Attalla H, Makela T P, Adlecreutz H & Andresson L C, *Biochem Biophys Res Commun,* 228, **1996**, 467.

9 Nakagawa-Yagi Y, Ogane N, Inoki Y & Kitoh N, *Life Sci* 58, **1996**, 1461.

10 Schmacher G, Kataoka M, Roth J A & Mukhopadhyay T, *Clin Cancer Res*, 5, **1999**, 493.

11 Purohit A, Hejaz H A M, Walden L, Mac Carthy M L, Packham G, Potter B V & Eeed M J, *Int J Cancer*, 85, **2000**, 584.

12 Leese M P, Hejaz H A M, Mahon M F, Newman S P, Purohit A, Reed M J & Potter B V L, *J Med Chem*, 48, **2005**, 5243.

13 Hansch C & Leo A, *Substituent constants for correlation analysis in chemistry and biology*, (John Wiley & Sons, New York), **1979**.

14 Wolff M E, *Burger's medicinal chemistry and drug discovery,* (John Wiley & Sons, New York), **1994**.

15 Skagerberg B, Bonelli D, Clementi S, Cruciani G & Ebert C, *Quant Struct Act Relat,* 8, **1989**, 204.

16 Gupta A K, Arockia B M & Kaskhedikar S G, *Indian J Pharm Sci,* 66, **2004**, 396.