

Quantitative Structure Activity Relationship study (QSARs) of Hydroxy-2-(hydroxy phenyl) indole derivatives for Estrogen Binding to the Estrogen Receptor

Shailja Sachan*, Madhu Mishra**, B. K. Gupta**, R. S. Nigam***

*Department of Chemistry, Govt. New Science College, Rewa-486001, INDIA.

**Department of Chemistry, govt. Auto. P. G. College, Satna-485001, INDIA.

***Department of Chemistry, Rajiv Gandhi College, Sherganj, Satna-485001, INDIA.

ABSTRACT: The present study is based on modelling toxicity of environmental endocrine disruptors as estrogen binding by using distance based topological and connectivity indices. The multiple linear regression analysis reveals that ${}^0\chi^V$, ${}^1\chi^V$, ${}^2\chi^V$, ${}^3\chi^V$ are suitable parameters for modelling the toxicity as relative binding affinity for estrogen binding (RBA) to the estrogen receptor. The quality of proposed models has been tested by using cross validation procedure also.

Keywords: Estrogen Binding Affinity (RBA), Estrogen receptor, QSAR, topological and connectivity indices, Multiple linear regression analysis.

I. Introduction

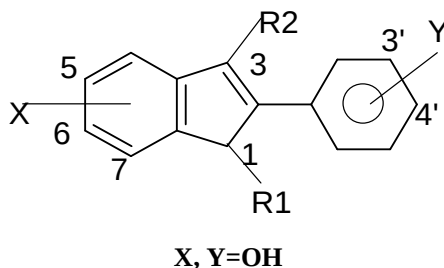
Estrogen receptor (ER) binds to estrogen response elements in target genes acts as a growth factor in normal development of the mammary gland, as well as in breast cancer, and may contribute also in the molecular events. Estrogen action on mammary epithelial cells (MECs) is mediated by estrogen receptors (ER- α and ER- β), members of the nuclear receptor superfamily and transcription factors that regulate expression of estrogen-responsive genes. It is hypothesized that ER- α contributes to the carcinogenic process by increasing cell proliferation through ER-mediated signal transduction. The rat estrogen receptor (ER) exists as two subtypes, ER α and ER β , which differ in the C-terminal ligand binding domain and in the N-terminal transactivation domain. Saturation ligand binding analysis of *in vitro* synthesized human ER α and rat ER β protein revealed a single binding component.

The estrogen database could be transformed into a knowledge base for regulatory and research purposes in part by understanding the basis of the relationship between chemical structure and estrogenic activity. Quantitative structure-activity relationship (QSAR) analysis correlates chemical structure with a specific biological activity, and the derived models are used to predict activities of untested chemicals. Ultimately, it is reasonable to expect that a number of such models may cover a number of possible mechanisms of effects of chemicals on estrogenicity (e.g., ER- β , aromatase inhibition, etc.).

Here the goal of QSAR methodology is to develop several models to predict activity/property/toxicity using correlation analysis employing statistical techniques. In the present study we have used simple as well as multiple regression analysis using maximum R^2 methods for modelling of estrogen receptors. Many estrogen responses are thought to be mediated via estrogen binding to the classical estrogen receptor (ER- α) and the ability of a chemical to compete for estradiol binding with ER and to induce estrogenic effects *in vivo*. Therefore, the determination of the pharmacophoric elements of ligands that bind to the ER is crucial to understanding the biological effects of estrogens.¹

Since the latter pioneer work had been done by Hansch and colleagues² in the 1960s, QSAR models have been applied extensively in various areas in the Chemistry and Biology.³ For environmental toxicology,⁴ QSAR is seen as a scientifically credible tool for predicting and classifying biological activities of chemicals when little or no actual data are available.⁵

Table1: Structure and activity of Hydroxy-2-(hydroxyphenyl)indole derivatives used in this study



Comp.	R ₁	R ₂	X	Y	Exp. logRBA
1	H	H	6	4'	-2.00
2	H	CH ₃	6	4'	-1.22
3	H	C ₂ H ₅	6	4'	-0.89
4	H	H	5	4'	-2.00
5	H	CH ₃	5	4'	-1.22
6	CH ₃	H	6	4'	0.58
7	C ₂ H ₅	H	6	4'	1.20
8	C ₃ H ₇	H	6	4'	0.93
9	C ₄ H ₉	H	6	4'	0.63
10	CH ₃	CH ₃	6	4'	1.00
11	C ₂ H ₅	CH ₃	6	4'	1.52
12	C ₃ H ₇	CH ₃	6	4'	1.11
13	i-C ₃ H ₇	CH ₃	6	4'	1.11
14	CH ₃	C ₂ H ₅	6	4'	0.77
15	C ₂ H ₅	C ₂ H ₅	6	4'	1.32
16	C ₃ H ₇	C ₂ H ₅	6	4'	1.28
17	CH ₃	H	5	4'	-0.10
18	C ₂ H ₅	H	5	4'	0.76
19	C ₃ H ₇	H	5	4'	1.25
20	CH ₃	CH ₃	5	4'	0.66
21	C ₂ H ₅	CH ₃	5	4'	0.98
22	C ₃ H ₇	CH ₃	5	4'	1.20
23	i-C ₃ H ₇	CH ₃	5	4'	0.54
24	C ₄ H ₉	CH ₃	5	4'	0.66
25	C ₅ H ₁₁	CH ₃	5	4'	0.36
26	C ₂ H ₅	C ₂ H ₅	5	4'	1.36
27	C ₃ H ₇	C ₃ H ₇	5	4'	0.23
28	C ₂ H ₅	CH ₃	7	4'	-1.70
29	C ₂ H ₅	H	6	3'	0.23
30	CH ₃	CH ₃	6	3'	-0.26
31	C ₂ H ₅	CH ₃	6	3'	0.48
32	C ₃ H ₇	CH ₃	6	3'	0.54
33	C ₂ H ₅	H	5	3'	0.23
34	CH ₃	CH ₃	5	3'	-0.22
35	C ₂ H ₅	CH ₃	5	3'	0.34
36	C ₃ H ₇	CH ₃	5	3'	0.87
37	(CH ₂) ₆ N(CH ₂ CH ₂) ₂	CH ₃	5	4	1.32

II. Materials and Methods

Data Sets for Analysis: W. Tong and coworkers⁶ have sought to develop potent antiestrogens based on nonsteroidal structures. Among their prototype structures are the 2-phenylindoles, for which extensive structure-activity relationship (SAR) studies have been done⁷. We have selected 37 estrogenic compounds from W Tong's data base as the training set to determine the steric and electrostatic requirements for recognition at the ER binding site. In the present study, several naturally occurring and synthetic estrogens have added to the 2-phenylindole to increase the structural diversity and to span a large range of RBA values within the training set. The chemical structures of the 37 pollutants organized by chemical class are shown in Table 1. The RBAs were calculated from a calf uterine estrogen receptor (calf ER) competitive binding assay with [³H] 17 β -estradiol (E₂). The RBA is the ratio of the molar concentration of E₂ and the competing chemical required to decrease the receptor-bound radioactivity by 50%, multiplied by 100; thus, E₂ has an RBA of 100.

Presentation of Data: In the present study Table 1 represents the structure and toxicity which is represented in terms of log RBA while, Table 2 contains various calculated topological and indicator parameters, Table 3 shows correlation matrix between toxicity and different descriptors. Table 4 shows cross-validation statistical analysis and Table 5 represents the actual, predicted and residual values for best model. Figure 1 is the graph plotted between observed and predicted toxicity.

Parameter Used: In the series of 37 Hydroxy-2-[hydroxyphenyl] indole derivatives we have used two types of descriptors:

(i) Topological Parameters: The first one is the topological descriptor which has calculated by E-Dragon software. Various topological parameters have been used such as Balaban and various Balaban type topological descriptors (J, Jhetv, Jhetp), different Randic indices (⁰ χ , ¹ χ , ² χ , ³ χ), and various connectivity parameters (⁰ χ^v , ¹ χ^v , ² χ^v , ³ χ^v). For calculation of these indices we have used hydrogen suppressed graphs. These indices are good enough to model the biological toxicities of the pollutants used in pharmaceutical industry. Wiener and Balaban type indices have been found to be the most used parameters for such studies. Also Randic and Kier and Hall type indices⁸⁻¹³ are very effective in modeling the biological activity of drug molecules.

(ii) Indicator Parameters: It is a dummy parameter used to indicate Indicator variables (parameters) some times called dummy variables or de-novo constants are used in linear multiple regression analysis to account for certain features which cannot be describe by continuous variables. In QSAR equation they normally describe a certain structural element, be it a substituent or another molecular fragment. Thus, Free Wilson analysis may be interpreted as a regression analysis approach using only indicator variables.

In the present study we used two indicator parameters, description of these parameters are as follow; ***IR₁**: 1 if R₁ is C₂H₅, otherwise 0, ***IR₂**: 1 if R₂ is CH₃, otherwise 0

Table: 2 Calculated topological and indicator descriptors

Compd. No.	J	Jhetv	Jhetp	$^0\chi$	$^0\chi^V$	IR ₁	IR ₂
1	1.561	1.859	1.788	11.828	9.013	0	0
2	1.637	1.966	1.897	12.698	9.936	0	1
3	1.676	2.014	1.948	13.405	10.643	0	0
4	1.561	1.904	1.838	11.828	9.013	0	0
5	1.639	2.011	1.946	12.698	9.936	0	1
6	1.639	1.895	1.814	12.698	9.960	0	0
7	1.680	1.889	1.803	13.405	10.668	1	0
8	1.692	1.858	1.771	14.113	11.375	0	0
9	1.683	1.812	1.727	14.820	12.082	0	0
10	1.715	2.000	1.920	13.569	10.883	0	1
11	1.755	1.991	1.905	14.276	11.590	1	1
12	1.766	1.956	1.869	14.983	12.297	0	1
13	1.802	1.997	1.906	15.146	12.460	0	1
14	1.753	2.048	1.972	14.276	11.590	0	0
15	1.793	2.040	1.959	14.983	12.297	1	0
16	1.806	2.008	1.924	15.690	13.004	0	0
17	1.637	1.934	1.857	12.698	9.960	0	0
18	1.676	1.921	1.839	13.405	10.668	1	0
19	1.687	1.885	1.801	14.113	11.375	0	0
20	1.715	2.039	1.963	13.569	10.883	0	1
21	1.753	2.023	1.941	14.276	11.590	1	1
22	1.763	1.983	1.899	14.983	12.297	0	1
23	1.799	2.025	1.937	15.146	12.460	0	1
24	1.752	1.928	1.845	15.690	13.004	0	1
25	1.727	1.865	1.785	16.397	13.712	0	1
26	1.793	2.072	1.993	14.983	12.297	1	0
27	1.819	2.049	1.973	16.397	13.712	0	0
28	1.786	2.014	1.926	14.276	11.590	1	1
29	1.708	1.912	1.824	13.405	10.668	1	0
30	1.745	2.026	1.944	13.569	10.883	0	1
31	1.784	2.015	1.927	14.276	11.590	0	1
32	1.793	1.978	1.889	14.983	12.297	0	1
33	1.704	1.945	1.860	13.405	10.668	1	0
34	1.745	2.065	1.987	13.569	10.883	0	1
35	1.782	2.048	1.964	14.276	11.590	1	1
36	1.790	2.005	1.919	14.983	12.297	0	1
37	1.529	1.506	1.438	20.803	17.987	0	1

III. Results and Discussion

A regression analysis was carried out on desktop computer using the statistical analysis system NCSS, version 2007. The quality of generated correlation equations is evaluated by the correlation coefficient (R), standard error of estimate, a test of null hypothesis (F-test), explained variance (R²) and adjusted variance (AR²). Further the predictability of model was also measured by the cross validated R² and R² from Ridge regression.

A perusal of Table 2 indicates that high degeneracy is observed in J Jhetv and Jhetp, This is obvious as these topological indices belong to first generation topological indices proposed by Balban, However no degeneracy is observed in ${}^0\chi^v$ and thus it belongs to second generation topological indices. This indicates that ${}^0\chi^v$ will possibly prove better topological index than Balban indices in modelling the toxicity.

The data presents in Table 2 also show that presence of OH group is common in all the estrogen receptor derivatives and thus probability this play a dominant role in exhibits the toxicity.

A perusal of correlation matrix was that there was a significant correlation between toxicity and Balban or Balban type index and connectivity index. Similarly it was also pointed out that there was a strong auto correlation among connectivity indices also, so all connectivity indices gave same effect upon toxicity to avoid the problem of colinearity these indices would be used separately.

The correlation matrix in table 3 indicates that name of the descriptor used correlation sigly with the actively, therefore no un correlations are possible. Based on above results are have attempted to discover and excellent multivariate correlation for modelling the toxicity of Hydroxy-2-(hydroxyphenyl) indole derivatives using various descriptors.

The data reported in Table-4 indicates that J, Jhetv, Jhetp, ${}^0\chi^v$ and ${}^0\chi$ gave significant result. Jhetv and jhetp exerted same effect on toxicity, The magnitude of correlation coefficient indicates that similar results will be obtained using any one of these two descriptors, similarly ${}^0\chi^v$ and ${}^0\chi$ also exerted the same effect upon toxicity

Table: 3 Correlation matrix between toxicity, topological and indicator variables

	Exp log RBA	J	Jhetv	Jhetp	${}^0\chi$	${}^0\chi^v$	IR ₁	IR ₂
Exp log RBA	1.0000							
J	0.4390	1.0000						
Jhetv	-0.0596	0.6986	1.0000					
Jhetp	-0.0983	0.6596	0.9976	1.0000				
${}^0\chi$	0.5301	0.2274	-0.4152	-0.4293	1.0000			
${}^0\chi^v$	0.5401	0.2503	-0.3955	-0.4105	0.9996	1.0000		
IR ₁	-0.0138	-0.0871	0.2013	0.2140	-0.2794	-0.2734	1.0000	
IR ₂	0.0333	0.2788	0.1398	0.1310	0.2770	0.2860	0.02993	1.0000

Data reported in table 4 indicates that with all 37 compounds no significant result is obtained.

- LogRBA= 0.2326 (±0.2835) IR₁-0.4117(±0.3154) IR₂10.920 (±3.6716)J -5.7560(±2.7181) Jhetp

0.0481(±0.1382) ${}^0\chi^v$

N=37 R²=0.4972 MSE=0.5411 F=6.131

..... (1)

It is also seen from the structure represented of these compounds in Table 1 that compound no. 28 is only a single compound which has only X in 7th position and compd. No. 37 is a single compound which contains (CH₂)₆N(CH₂CH₂)₂ in place of R₁, both compounds are not appropriate for statistical analysis. So we have considered compd. No. 28 & 37 as outlier. After deleting the compound no 28 and 37 as outlier a dramatic change in statistics is arised, magnitude of variance is changed from 0.4927 to 0.8057 and 0.4968 to 0.8116. Based on these finding we have proposed following regressions for modelling and activating the toxicity of these indole derivatives used in this study.

After Deletion of two Outlier Compound No. 28 & 37

LogRBA= 26.2524(±14.0821) J -9.8300(±1.7838) Jhetp -0.6900 (±0.1993) ${}^0\chi$ + 0.2798 (±0.1807) IR₁ - 1.1072 (±0.2203) IR₂

N=35 R²=0.8116 MSE=0.1827 F=24.989

..... (2)

LogRBA= 26.1224 (±4.2556) J -9.6642 (±1.8121)Jhetp -0.6655 (±0.2031) ${}^0\chi^v$ + 0.2856(±0.1839) IR₁ - 1.0940(±0.2242) IR₂

N=35 R²=0.8057 MSE=0.1884 F=24.056

..... (3)

In the above equations eqn. 2 is more significant with the variance 0.8116, higher F-ratio value as 24.989 and lower mean square error (MSE) as 0.1827; there is a small difference in the value of variance, F-ratio & MSE in eqn. 3. It supports the concept that zero order randic and connectivity index are similarly effective upon toxicity in this QSAR study.

Table-4: Regression Analysis of derivatives

Model No.	Para. Used	Ai(1----5)	Intercept	MSe	R ²	F-Ratio
1	J Jhetv ⁰ _{χ^v}	A ₁ = 11.2702(±3.9123) A ₂ = -5.9918(±2.9859) A ₃ = 0.0446(±0.1430)	-7.7902	0.54946	0.4565	9.241
2	J JhetrP ⁰ _χ	A ₁ =10.5561(±3.3619) A ₂ = -5.6586(±2.6521) A ₃ = 0.0563(±0.1320)	-7.9608	0.54376	0.4622	9.453
3	J JhetrP ⁰ _{χ^v}	A ₁ = 10.4389(±3.4096) A ₂ = -5.5875 (±2.6487) A ₃ = 0.0608 (±0.1318)	-7.7921	0.54325	0.4627	9.472
4	IR ₁ J JhetrP ⁰ _χ	A ₁ =0.1800(±0.2840) A ₂ = 9.8970 (±3.5486) A ₃ = -5.3377 (±2.7240) A ₄ = 0.0773 (±0.1373)	-7.7843	0.55380	0.4689	7.062
5	IR ₂ J JhetrP ⁰ _χ	A ₁ =0.3794(±0.3109) A ₂ = 11.7793 (±3.4845) A ₃ = -6.1965 (±2.6693) A ₄ = 0.0195 (±0.1344)	-8.4455	0.53581	0.4861	7.567
6	IR ₁ IR ₂ J Jhetp ⁰ _{χ^v}	A ₁ = 0.2326 (±0.2835) A ₂ = -0.4117(±0.3154) A ₃ = 10.920 (±3.6716) A ₃ = -5.7560(±2.7181) A ₃ = 0.0481(±0.1382)	-8.1363	0.5411	0.4972	6.131
7	IR ₂ IR ₁ J JhetrP ⁰ _χ	A ₁ = 0.2302 (±0.2834) A ₂ =--0.4140 (±0.3154) A ₃ =-11.0478(±3.6171) A ₄ = -5.8351 (±2.7203) A ₅ =0.0430 (±0.1382)	-8.2639	0.5416	0.4968	6.122
After Deletion of two Outlier Compound No. 28 and 37						
8	J JhetrP ⁰ _χ IR ₁ IR ₂	A ₁ = 26.2524(±14.0821) A ₂ = -9.8300(±1.7838) A ₃ =-0.6900 (±0.1993) A ₄ =0.2798 (±0.1807) A ₅ =-1.1072 (±0.2203)	-16.346	0.1827	0.8116	24.989
9	J JhetrP ⁰ _{χ^v} IR ₁ IR ₂	A ₁ =26.1224 (±4.2556) A ₂ =-9.6642 (±1.8121) A ₃ =-0.6655 (±0.2031) A ₄ = 0.2856(±0.1839) A ₅ = -1.0940(±0.2242)	-18.593	0.1884	0.8057	24.056

Table-5: Cross Validation Statistical parameters

Model	n	R ² _{cv}	R ² _{adj}	R ² _{ridge}	R ²
6	37	0.1803	0.4161	0.4913	0.4972
9	35	0.7206	0.7722	0.7820	0.8057
8	35	0.7286	0.7791	0.7874	0.8116

In order to confirm above standard we have also used the cross validation process & ridge regression for study of multicollinearity. Results are reported in table 4&5 It is clear that model 8 is best in this study Further confirmation is also established by estimating the predicted activity by and by plotting graph between observed and predicted activity. A plot is also obtained between variance inflation factors with K value. The value of VIF is less than from which also support the importance of the model no 8.

Table-6: Screening Summary

C. No	Actual	Predicted	Residual
1.	-2	-1.094	-0.906
2	-1.22	-0.777	-0.443
3	-0.89	-0.721	-0.169
4	-2	-1.578	-0.422
5	-1.22	-1.198	-0.022
6	0.58	0.062	0.518
7	1.2	1.053	0.147
8	0.93	0.92	0.01
9	0.63	0.639	-0.009

C.No.	Actual	Predicted	Residual
19	1.25	0.499	0.751
20	0.66	-0.007	0.667
21	0.98	1.013	-0.033
22	1.2	0.924	0.276
23	0.54	1.389	-0.849
24	0.66	0.688	-0.028
25	0.36	0.144	0.216
26	1.36	1.085	0.275

10	1	0.408	0.592
11	1.52	1.413	0.107
12	1.11	1.292	-0.182
13	1.11	1.767	-0.657
14	0.77	0.428	0.342
15	1.32	1.413	-0.093
16	1.28	1.335	-0.055
17	-0.1	-0.406	0.306
18	0.76	0.601	0.159

27	0.23	0.73	-0.5
28	0.23	0.488	-0.258
29	-0.26	-0.134	-0.126
30	0.48	0.864	-0.384
31	0.54	0.71	-0.17
32	0.23	0.035	0.195
33	-0.22	-0.55	0.33
34	0.34	0.454	-0.114
35	0.87	0.342	0.528

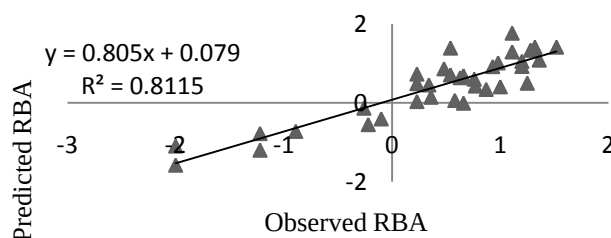


Fig-1: Graph drawn between actual and predicted RBI activity of hydroxyl-2-(hydroxyphenyl) indole derivatives.

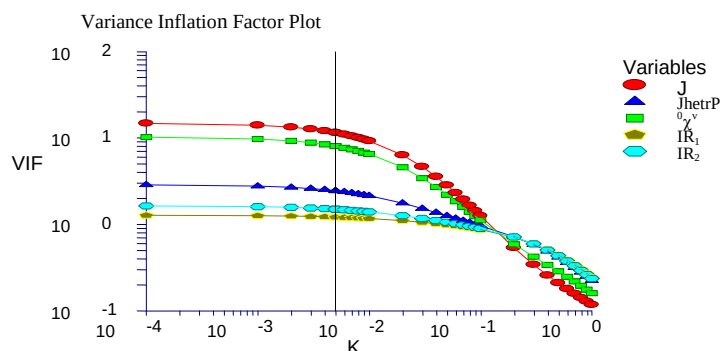


Figure 2: Plot between VIF and K.
IV. CONCLUSION:

Spurred by the need of new anti-pollutants or less toxic pollutants possess heterocyclic systems in their structure, such as indole ring. In view of the results and discussions presented above, we conclude that topological properties such as Weiner, Balaban, various connectivity and valence connectivity descriptors can successfully be used for modelling toxicity of various this types of pollutants. These results will help medical as well as environmental scientists in the designing and prediction of toxicity of pollutants and thus the synthesis of new compounds exhibiting less toxicity then those reported in this paper.

ACKNOWLEDGEMENT

Author's express deep sense of gratitude to Prof. V. K. Agrawal, Director, NITTTTR, Bhopal for giving the opportunity to work under his expert guidance in this new area and very thankful to him for his illuminating guidance, constant inspiration, useful feedback and suggestions. One of the author S. Sachan is also thankful to UGC Bhopal for providing financial support to purchase NCSS Software.

REFERENCE

- [1] G.M.Anstead ,K. E. Carlson , J. A. Katzenellenbogen , *Steroids*, Vol. 62, (1997), 268-303.
- [2] T. Fujita , J. Iwasa , C. Hansch , *J. Am. Chem. Soc.*, Vol. 86, (1964), 5175-5180.
- [3] C.Hansch ,A. Leo , *Exploring QSAR-Fundamentals and Applications in Chemistry and Biology*, Washington DC: *The American Chemical Society*, (1995).
- [4] S. P. Bradbury , *Toxicol. Lett.*, Vol. 79, (1995), 229-237.
- [5] C.Hansch, B. R. Telzer ,L. Zhang , *Toxicol.*, Vol. 25(1), (1995), 67-89.
- [6] W. Tong , R. Perkins ,R. Strelitz ,E. R. Collantes , S. Keenan ,W. J. Welsh , W. S. Branham ,D. M. Sheehan , *Environ. Health Perspect*, Vol. 105 (10), (1997), 1116-1124.
- [7] A. E. Von ,J. Prekajac ,J. Strohmeier , *J. Med. Chem.*, Vol. 27, (1984), 1439-1447.
- [8] P. V. Khadikar ,S. Karmakar, V. K. Agrawal, *J. Chem. Inf. Comp. Sci.*, 41 (2001) 934.
- [9] V. K.. Agrawal, J. Singh ,B. Shaik, S. Singh,S. Sikhima,P.V.Khadikar,C.T.Cupuran;*Bioorg.Med.chem* , 15 (2007) 6501.
- [10] V. K. Agrawal, V. K. Dubey, J. Singh, P.V. Khadikar, *Oxid. Commun.*, 1 (2008) 27.
- [11] V. K. Agrawal, J. Singh, S. Singh, P. V. Khadikar, *Oxid. Commun.*, 1 (2008) 2.
- [12] V. K. Agrawal, J. Singh, B. Shaik, S. Chaturvedi, P. V. Khadikar, *Oxid. Commun.*, 31 (2008) 264.
- [13] V. K. Agrawal, J. Singh, B. Shaik, S. Singh, O. Deeb, P. V. Khadikar, C. T. Supuran , *Chemical Biology Drug Design*. 71 (2008) 244.