

BINDING AFFINITY OF TRIPHENYL ACRYLONITRILES TO ESTROGEN RECEPTORS: QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIPS

Sorana D. Bolboacă¹, Monica M. Marta¹, Lorentz Jäntschi^{1,2}

¹"Iuliu Hațieganu" University of Medicine and Pharmacy Cluj-Napoca, Romania; ²Technical University of Cluj-Napoca, Romania

ABSTRACT

AIM: The quantitative structure-activity relationship approach was applied to understand the relative binding affinity of triphenyl acrylonitriles to estrogen receptors.

MATERIAL AND METHODS: A sample of previously studied triphenyl acrylonitriles was divided into training (18 compounds) and test sets (7 compounds) using a stratified random approach. The molecular descriptor family on vertices cutting (MDFV) approach was used in order to translate the structural information into descriptors. The relationship between binding activity and structural descriptors was identified using the multiple linear regression procedure.

RESULTS: An optimal three-parameter equation with a determination coefficient of 0.9580 and a cross-validation leave-one-out parameter of 0.9408 was identified. The optimal model was assessed on a test set and a determination coefficient of 0.9004 was obtained. The MDFV model proved not to be significantly different from the previously reported model in terms of goodness-of-fit. In terms of information criteria (Akaike's, Bayesian, Amemiya, and Hannan-Quinn) and Kubinyi function, the MDFV model proved to perform better than the previously reported model.

CONCLUSION: The optimal MDFV model was able to explain ~96% of the total variance in the estrogenic binding relative affinity of triphenyl acrylonitriles and to have estimation and prediction abilities. Although there were no significant differences in terms of goodness-of-fit, the MDFV model proved to exhibit better information parameters compared to the previously reported model using the same number of molecular descriptors.

Keywords: estrogen receptors, relative binding affinity (RBA), triphenyl acrylonitriles (TPT); structure-activity relationship (SAR), molecular descriptors family on vertices (MDFV)

INTRODUCTION

The interaction of estrogens with tissues is accomplished through estrogen receptors (ER), a group of receptors activated by the hormone 17 β -estradiol.¹ Two receptors are known to date: ER α (expressed mainly in uterus, stroma of prostate, theca cells of ovary, Leydig cells of testes, epididymus, bone, breast, brain, liver and white adipose²) and ER β (expressed mainly in colon, epithelium of prostate, testis, granulose cells of ovary, bone marrow, salivary gland, vascular endothelium, brain^{3,4}).

A series of hydroxylated and non-hydroxylated triphenyl acrylonitriles (TPEs) proved to have estrogenic activities.⁵⁻⁸ The structure-activity approach has been applied to link topological features of estrogenic drugs and pharmacophoric properties^{9,10}, structural requirements of ER ligand¹¹, or reproductive toxicology.¹²

A sample of hydroxylated and non-hydroxylated triphenyl acrylonitriles for binding to calf uterus estrogen receptors has been investigated by using quantitative structure-activity relationship approach.¹ The best performing model after removing one outlier (the residual value exceeded twice the standard error of estimate) is presented in Eq (1).

$$\begin{aligned} \log RBA = & 2.114(\pm 0.243)I_{12-OH} - 0.223(\pm 0.059)S_6 \\ & - 0.147(\pm 0.050)S_{18} - 0.193(\pm 0.052)N_t \\ n = & 24, r = 0.919, r^2 = 0.845, EV = 81.415\%, \\ F = & 27.284, s = 0.595, \\ AVRES = & 0.450, PRESS = 10.021, SDEP = 0.646, \\ Pres_{av} = & 0.515, Q^2 = 0.751 \end{aligned} \quad (1)$$

where logRBA = competition for [³H] 17- β -estradiol binding, S₆ and S₁₈ are E-states of C₆ and C₁₈ respectively, I_{12-OH} = presence or absence of -OH substitution in C₁₂, N_t = number of free terminal

atoms (excluding H) in C_{12} , n = sample size; r = correlation coefficient, r^2 = coefficient of determination, F = F-values (significance level at 1%), EV = explained variance, s = standard error of estimate, $AVRES$ = average of absolute value of residuals, $PRESS$ = predictive residual sum of squares, $SDEP$ = standard deviation of error of predictions, $Pres_{av}$ = average of absolute value of predicted residuals, Q^2 = cross-validated variance.

The study aimed to model the relationship between topological structures of a sample of triphenyl acrylonitriles with binding estrogen receptor using the molecular descriptors family obtained through vertex cutting (MDFV). The best performing model was compared with previously reported model in order to identify the method with highest performances.

MATERIAL AND METHODS

The sample of triphenyl acrylonitriles previously studied by Mukherjee et al.¹³ was included in analysis. The compound abbreviation, 2D structure and estrogenic activity expressed as relative binding affinity to the ER vis-à-vis E2 in logarithmic scale (logRBA) are given in Fig. 1. The 2D structures of the compounds were drawn using HyperChem software (version 8.04). The molecular geometry was constructed using HyperChem, and the molecule was saved as *.mol file. The molecular geometry was optimized using Molecular Modeling Pro Plus: *conformational analysis – moly minimizer – makes moderate changes – refine* (applied twice). The molecule was then saved as *.hin file (to be introduced in construction of molecular family on vertex cutting), the compounds were validated and the partial changes were computed whenever needed using HyperChem.

The sample was split in two sets: training (used to obtain the multiple-linear regression model (MLR)) and test (used to test the MLR equation obtained on training set). A number of 7 molecules were assigned to the test set, using the following approach:

- Construct of strata of the sample (25 compounds) based on experimental logRBA values. Seven strata were constructed and the frequency table was obtained.
- Extraction randomly a predefined number of molecules from each strata whenever the frequency was higher than 1 (there was one stratum with a single compound; this was assigned to training set). One stratum with observed frequency higher than 4 was randomly to provide 2 com-

pound for inclusion in test set. The following compounds were included in test set: triph024, triph016, triph004, triph017, triph006, triph019, and triph009.

The logRBA variance of the compounds included in training set proved not to be statistically significant different compared to the compounds included in test set (F-test for variances = 1.28, p-value = 0.4018); the same result was obtained when the means were compared (t-value = -0.0229, p-value = 0.9819).

- Normality test on observed logRBA for compounds included in training set (EasyFit v. 5.1.): Kolmogorov-Smirnov statistic¹⁴ equal to 0.1243 ($p = 0.9118$); Anderson-Darling statistic¹⁵ equal to 0.5128 (null hypothesis of normality accepted at significance level of 1%); and Chi-Squared statistic¹⁶ equal to 1.0122 ($p = 0.6028$).

The molecular descriptors (MDFV)¹ were calculated for each compound included in the analysis by applying the following steps (using a series of home-made programs):

- Define the investigated set of compounds: logRBA.
- Create the following tables: `triph_mdfv` table (contains the named of molecular descriptors), `triph_data` table (contains the *.hin file of molecules prepared for modelling as describe above), and `triph_prop` (contains the experimental logRBA values).
- Compute the values of molecular descriptors: candidate fragments obtained by cutting of pairs of vertices after matrix representation of the molecular graph. The MDFV descriptors have a name of 8 letters indicating how they were generated (for details of MDFV see the paper of Bolboacă and Jäntschi.¹⁷)
- Validate the MDFV descriptors: 2387280 descriptors were calculated for each molecule. 6059 descriptors proved to be valid and were used in MLR analysis after applying three validation criteria (Jarque-Bera value higher than critical value for the observed activity, identity analysis and inter-correlation higher than 0.99).

The MLR approach was applied in order to identify the best link between MDFV descriptors and logRBA. Statistica 8.0 software was used in identification of best performing model.

The identification of the best performing MDFV - MLR model on training set was performed applying the following criteria:

- Highest explanation of the observed logRBA (highest values of significant correlation coef-

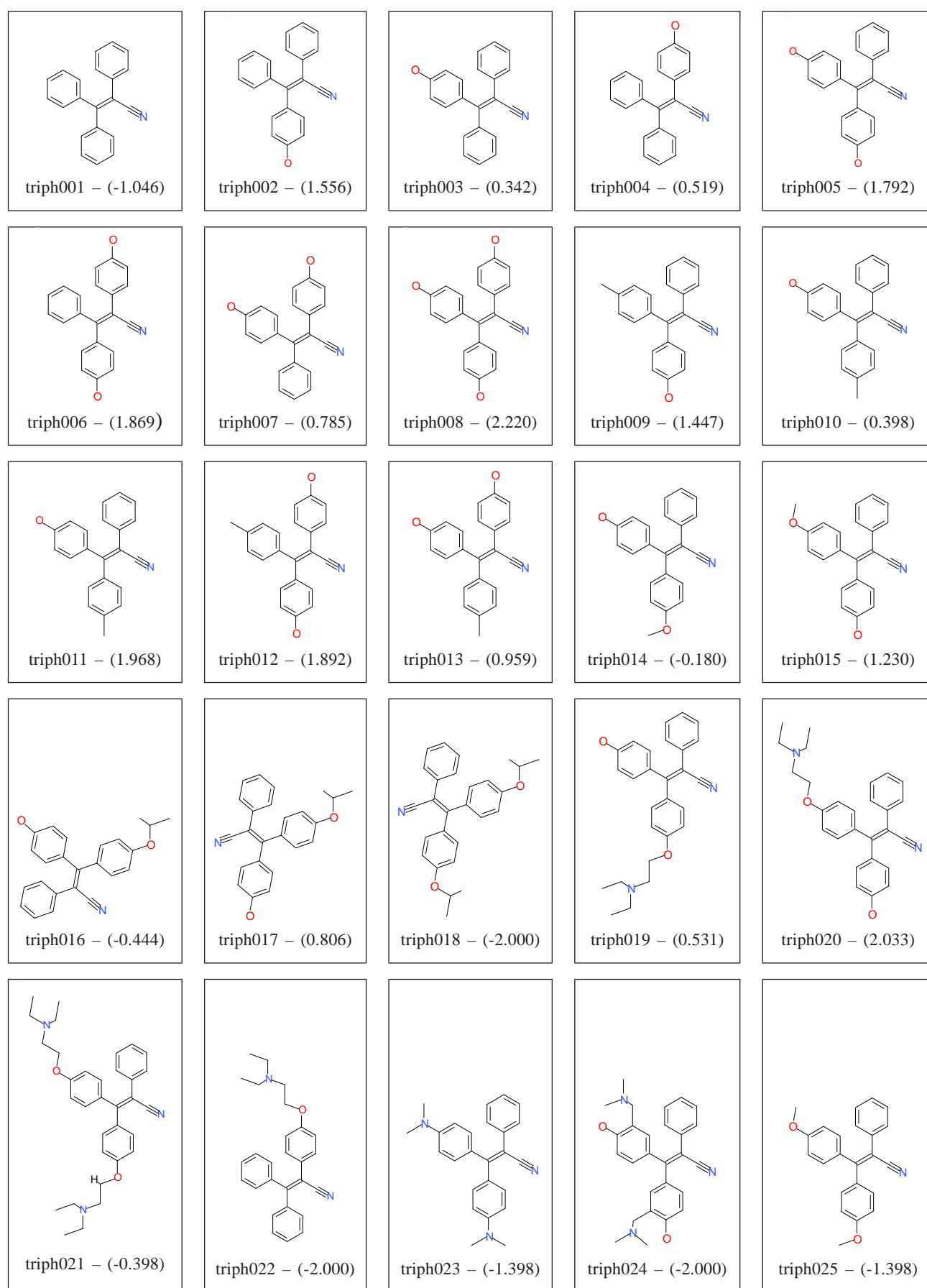


Figure 1. General structure, abbreviation and logLBA (brackets) of triphenyl acrylonitriles.

ficients between the observed and estimated activity).

- Smallest number of descriptors in the model.
- Lowest standard error of estimate.
- Highest F-value and lowest p-value associated to F-value (significance of MLR model).
- Absence of collinearity of descriptors in the model.
- Highest value of correlation coefficient in leave-one-out cross-validation.

The ability of the best performing MLR model identified in training set was test on test set.

A comparison analysis was carried on to compare the best performing MDFV-MLR model with previously reported model with higher correlation coefficient using the following statistics: Akaike's and related information criteria¹⁸ (the smallest the value the better the model was); Kubinyi function (FIT)^{19,20} (the highest the value the better the model is) and Z test for comparing two correlation coefficients²¹.

RESULTS

The best performing MDFV model in terms of goodness-of-fit is presented in Eq (3).

$$\hat{Y}_{\text{MDFV}} = 65.11(\pm 10.35) - \text{TASAAF DL} * 9.18(\pm 1.4) + \text{GLCACPD L} * (6.69 \cdot 10^{-1})(\pm 1.62 \cdot 10^{-1}) + \text{GMHAAID R} * (7.84 \cdot 10^{-5})(\pm 1.86 \cdot 10^{-5}) \quad (3)$$

where \hat{Y}_{MDFV} = estrogenic activity estimated by the model, the numbers in round brackets represent the parameter needed to compute the confidence intervals for the slope parameters; TASAAF DL, GLCACPD L, and GMHAAID R = the MDFV members.

Statistical characteristics of the model from Eq(3) are give in Eq(4).

$$n_{\text{training}} = 18, r_{\text{training}} = 0.9788 \text{ (95\%CI [0.9427 - 0.9922])}, \\ r^2_{\text{training}} = 0.9580, r^2_{\text{adj-training}} = 0.9489, \text{CV}_{\text{training}} = 0.8660$$

$$s_{\text{est}} = 0.33, \text{F-value (p-value)} = 106 (7.16 \cdot 10^{-10}) \\ t_{\text{int}} \text{ (p-value)} = 13.49 (2.06 \cdot 10^{-9}), t_{\text{TASAAF DL}} \text{ (p-value)} = -13.72 (1.65 \cdot 10^{-9}),$$

$$t_{\text{GLCACPD L}} \text{ (p-value)} = 8.84 (4.18 \cdot 10^{-7}), t_{\text{GMHAAID R}} \text{ (p-value)} = 9.02 (3.30 \cdot 10^{-7})$$

$$\text{TASAAF DL: } T = 0.941, \text{VIP} = 1.063$$

$$\text{GLCACPD L: } T = 0.926, \text{VIP} = 1.080$$

$$\text{GMHAAID R: } T = 0.903, \text{VIP} = 1.108$$

$$r^2_{\text{loo}} = 0.9408, s_{\text{loo}} = 0.39, F_{\text{loo}} = 74 (7.88 \cdot 10^{-9}) \quad (4)$$

where training = training set, n = sample size; r = correlation coefficient; r^2 = determination coefficient; r^2_{adj} = adjusted determination coefficient;

CV = coefficient of variation; s_{est} = standard error of estimate; F-value = Fisher statistic of the MLR model; T = tolerance (collinearity diagnostic); VIF = variance inflation factor (collinearity diagnostic); r^2_{loo} = determination coefficient obtained in leave-one-out analysis; s_{pred} = standard error of predicted; F_{loo} = F-value obtained in leave-one-out analysis.

The model presented in Eq (3) was validated through its application on test set. Statistical characteristic are presented in Eq (5).

$$n_{\text{test}} = 7, r_{\text{test}} = 0.9489 \text{ (95\%CI [0.6860 - 0.9926])}, \\ r^2_{\text{test}} = 0.9004, s_{\text{test}} = 0.53, F_{\text{test}}(p_{\text{test}}) = 43 (1.21 \cdot 10^{-3}) \quad (5)$$

The values of MDFV descriptors used by model (Eq (3)), estimated and predicted logRBA, and residuals are presented in Table 1.

The goodness-of-fit of the model obtained on training set and its application on test set is presented in Fig. 2.

The best performing MDFV model (Eq (3)) was compared to the model previously reported (Eq (1), in terms of information criteria (Table 2) and of goodness-of-fit (Z test). The goodness-of-fit of the model from Eq(3) was compared with the goodness-of-fit of the previously reported model (Eq (1)) and a value of 0.795 (p-value = 0.2133) was obtained.

DISCUSSION

The linear relation between topological structures of triphenyl acrylonitriles and binding estrogens receptors was successfully modelled. The MDFV approach that implements the fragmentation of vertices in the molecular graph proved able to estimate activities.^{17,22} Seven information criteria and three weights were computed in order to compare the MDFV model with the model previously reported by Mukherjee et al.¹⁷

The best performing model was selected according to Hawkins principles²³: highest correlation coefficient, highest Fisher parameter, lowest standard error of the estimate, and smallest possible number of significant parameters ($n = 5 \cdot v$, where n = sample size, v = number of variables in the model). The model with the highest correlation coefficient, the highest Fisher parameter, the lowest standard error of estimate, and the smallest possible number of significant parameters was considered to be the best performing model (Eq (3)). This model is able to explain ~96% of the total variance in the estrogenic binding relative affinity of triphenyl acrylonitriles. The Fisher-value and its associated significance sustain the significance of the model

Table 1. Values of descriptors, observed, estimated / predicted logRBA and residuals in training and test sets

Mol	MDFV descriptor			logRBA	\hat{Y} -logRBA	Residuals
	TASAAF DL	GLCACPD L	GMHAAID R			
Training set						
triph018	7.440	1.693	1148.2	-2.000	-1.9825	-0.0175
triph022	7.296	-2.002	14537.0	-2.000	-2.0805	0.0805
triph023	7.408	-1.023	23340.0	-1.398	-1.7643	0.3663
triph001	7.194	-1.679	13358.0	-1.046	-1.0206	-0.0254
triph021	7.543	0.862	41710.0	-0.398	-0.3040	-0.0940
triph014	7.304	0.783	22956.0	-0.180	0.3677	-0.5477
triph003	7.270	0.772	19946.0	0.342	0.4362	-0.0942
triph025	7.350	-0.836	24907.0	-1.398	-0.9838	-0.4142
triph007	7.286	0.632	22890.0	0.785	0.4266	0.3584
triph013	7.304	0.670	29383.0	0.959	0.7962	0.1628
triph015	7.130	-0.685	24643.0	1.230	1.1164	0.1136
triph002	7.130	0.660	22774.0	1.556	1.8691	-0.3131
triph005	7.130	0.728	24238.0	1.792	2.0290	-0.2370
triph012	7.304	0.688	38360.0	1.892	1.5120	0.3800
triph011	7.130	0.752	20622.0	1.968	1.7616	0.2064
triph020	7.130	0.777	17342.0	2.033	1.5209	0.5121
triph008	7.304	1.804	39350.0	2.220	2.3352	-0.1152
triph010	7.130	-0.852	21011.0	0.398	0.7198	-0.3218
			<i>m</i>	0.375[†]	0.3753[†]	0.2422
			<i>stdev</i>	1.458	1.4269	0.2989
					<i>KS_{res} (p-value)</i>	0.1070 (0.9721)
					<i>AD_{res} (AD_{crit} 5%)</i>	0.1742 (2.5018)
Test set						
triph024	7.479	-2.367	33110.0	-2.000	-2.5490	0.5490
triph016	7.332	-0.649	25257.0	-0.444	-0.6663	0.2223
triph004	7.211	-0.716	23290.0	0.519	0.2458	0.2732
triph017	7.130	-0.694	30176.0	0.806	1.5440	-0.7380
triph006	7.231	-0.858	39450.0	1.869	1.2339	0.6351
triph019	7.373	0.865	30626.0	0.531	0.3903	0.1407
triph009	7.130	0.734	23111.0	1.447	1.9446	-0.4976
			<i>m</i>	0.390[‡]	0.3062[‡]	0.4366
			<i>stdev</i>	1.286	1.5376	0.5151
					<i>KS_{res} (p-value)</i>	0.2585 (0.6489)
					<i>AD_{res} (AD_{crit} 5%)</i>	0.4025 (2.5018)

\hat{Y} -logRBA = logRBA estimated by Eq(3) for training set and predicted logRBA for test set[†]

m = arithmetic mean; *stdev* = standard deviation;

Comparing means (significance level of 5%): [†] *p* = 0.9999; [‡] *p* = 0.6828;

KS_{res} = Kolgomorov-Smirnow statistic applied to residuals (for testing normality);

AD_{res} = Anderson-Darling statistics applied to residuals for testing normality;

AD_{crit} 5% = critical value for Anderson-Darling statistics for a significant level of 5%.

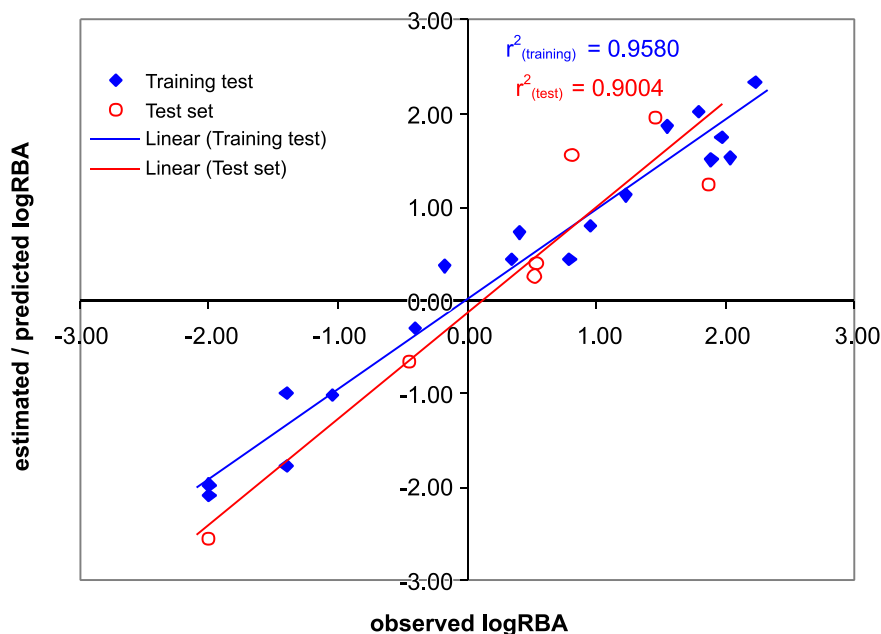


Figure 2. Goodness-of-fit of models: training versus test.

Table 2. Validation and comparison of the models

Parameter	Model	
	Eq(3)	Eq(1)
AIC _c (corrected Akaike information criterion)	-104.49	-52.16
w _i (AIC _c)	1.00	4.33·10 ⁻¹²
AIC _{R2} (AIC based on determination coefficient)	1.94	2.96
w _i (AIC _{R2})	0.62	0.38
AIC _u (McQuarrie and Tsai corrected AIC)	-4.34	-0.86
w _i (AIC _u)	0.85	0.15
BIC (Bayesian information criterion)	-99.48	-45.18
APC (Amemiya prediction criterion)	0.00	0.10
HQC (Hannan-Quinn criterion)	-107.08	-53.02
FIT (Kubinyi function)	8.72	2.59

w_i = Akaike weights for model *i*.

Parameters: Smallest the better excepting FIT and w_i (where largest the better).

from Eq (3) while the t-values and associated significance sustain the significance of the MDFV descriptors used by Eq (3). The characteristics of the descriptor's contribution to the binding estrogenic activity of triphenyl acrylonitriles revealed the followings:

- The interaction between structure and binding activity proved to fulfill via bonds (topology - *T*) and space (geometry - *G*).
- Dominant atomic properties were represented

by electronic affinity (*A*), melting point (*L*), and relative atomic mass (*M*).

- The structure on activity scale is logarithm (*L*) and reciprocal (*L*).

The binding estrogens receptor affinity of TPT proved to be of geometric and topological nature. It depended on compound electronic affinity, melting point and relative atomic mass (Eq (3)).

The validity of a linear model is sustained by the absence of collinearity within descriptors used

by the model. The analysis of collinearity was carried out for the model presented in Eq (3) by computing two parameters: tolerance and variance inflation factor. Tolerance, defined as the difference between 1 and determination coefficient, show the degree of instability in the regression coefficients. The variance inflation factor gives the degree to which the standard error of the predictor is increased due to the predictor's correlation with the other predictors in the model. Values less than 0.10 for tolerance and values greater than 10 for variance inflation factor indicates the presence of multicollinearity.²⁴

The prediction ability of the model in Eq (3) was analyzed in leave-one-out experiment (internal validation) and by applying the model on an external set of compounds (test set, external validation). Internal and external validation analyses must be conducted in SAR analysis since high internal predictivity is not necessary related to high external predictivity²⁵ (effect known as Kibinyi paradox²⁶).

The leave-one-out experiment, 17 experiments were conducted with 17 compounds in training set (for identification of MLR model) and 1 compound in test set (the application of the identified MLR model). The values of leave-one-out determination coefficient (0.9408) proved to be close to the determination coefficient of the model from Eq (3) (0.9580) and indicates a good prediction ability (~2% difference between these two determination coefficients). The external validation of the model presented in Eq (3) was carried out and the model was applied on a sample of 7 compounds. The difference between determination coefficient obtained in test set and the determination coefficient of the model in Eq (3) proved to be of 4%. Thus, the predictivity analysis (internal and external) conducted for the model presented in Eq (3), revealed that the model is reliable and could be used to predict the relative binding affinity on estrogen receptors of triphenyl acrylonitriles. Furthermore, the reliability of the model presented in Eq (3) is sustained by the absence of significant difference between the observed and estimated mean of logRBA (training set), respectively observed and predicted mean of logRBA (test set) ($p > 0.6$, significance level of 5%).

A series of parameters were computed in order to compare the best performing MDFV model (Eq (3)) with the model identified by Mukherjee et al. (Eq (1)). The MDFV model (Eq (3)) systematically obtained the best expected values: the smallest

values of information criteria (AIC_c , AIC_{R^2} , AIC_u , BIC, APC and HQC), highest values of Akaike's weights ($w_i(AIC_c)$, $w_i(AIC_{R^2})$, $w_i(AIC_u)$) and of the Kubinyi function (FIT). According to the results presented in Table 2, based on AIC values it could be concluded that the model from Eq(3) is more useful compared to model from Eq (1).

As far as the goodness of fit of the models on Eq (1) and Eq (3) according to Z test was concerned, these models were not statistically different ($p = 0.2133$). Even if the models are not statistically different in terms of goodness-of-fit the MDFV model proved to be better in terms of information criteria and Kubinyi function.

The present study aimed to model the relative binding affinity to estrogen receptors of TPT by using information extracted from the matrix representation of the compounds. A valid and reliable model with three descriptors was obtained. The model proved its reliability in training and test sets. The analyzed sample size is the main limitation of the present research. Current research in our laboratory aims to characterize activities / properties of other classes of compounds by using the MDFV approach.

CONCLUSIONS

In the present study an alternative SAR model relating the relative binding affinity on estrogen receptors with the molecular structure of triphenyl acrylonitriles by means of structural descriptors with the multiple linear regression approach. The best performing model proved to be able to explain ~96% of the total variance in the estrogenic binding relative affinity of triphenyl acrylonitriles and exhibits better information parameter compared to previously reported model with the same number of molecular descriptors involved. The assessment in test set of 7 triphenyl acrylonitriles not used in identification of the best MDFV model suggests that it performs predictively.

ACKNOWLEDGEMENTS

The research was supported by CNCSIS-UEFISCSU (project PNII-IDEI458/206/2007).

REFERENCES

1. Dahlman-Wright K, Cavailles V, Fuqua SA, et al. International Union of Pharmacology. LXIV. Estrogen receptors. *Pharmacol Rev* 2006;58(4):773-81.
2. Jensen EV, Jordan VC. The estrogen receptor: a model for molecular medicine. *Clin Cancer Res* 2003;9(6):1980-9.

3. Kuiper GG, Enmark E, Peltö-Huikko M, Nilsson S, Gustafsson JA. Cloning of a novel receptor expressed in rat prostate and ovary. *Proc Natl Acad Sci USA* 1996;93(12):5925-30.
4. Mosselman S, Polman J, Dijkema R. ER β : identification and characterization of a novel estrogen receptor. *FEBS Lett* 1996;392:49-53.
5. Harper MJK, Walpole AL. Contrasting endocrine activities of cis and trans isomers in a series of substituted triphenylethylenes. *Nature* 1966;212(5057):87.
6. Miquel JF, Sekera A, Chaudron T. Synthèse de polyphényléthylenes et Synthèse de polyphényléthylenes et interférences avec le récepteur oestrogène d'utérus de souris. *C R Acad Sci (Paris)* 1978;286:151-4.
7. Pons M, Michel A, Crastes de Paulet A, et al. Influence of new hydroxylated Triphenylethylene (TPE) derivatives on estradiol binding to uterine cytosol. *J Steroid Biochem* 1984;20(1):137-45.
8. Pons M, Bignon E, Chastes De Paulet A, Gilbert J, Ojasoo T, Raynaud JP. Hydroxylated triphenylacrylonitriles adopt a unique orientation within the binding site of the estrogen receptor. *J Steroid Biochem* 1990;36(5):391-7.
9. Spyrakis F, Cozzini P. How computational methods try to disclose the estrogen receptor secrecy - Modeling the flexibility. *Curr Med Chem* 2009;16(23):2987-3027.
10. Mukherjee S, Nagar S, Mullick S, Mukherjee A, Saha A. Pharmacophore mapping of selective binding affinity of estrogen modulators through classical and space modeling approaches: exploration of bridged-cyclic compounds with diarylethylene linkage. *J Chem Inf Model* 2007;47(2):475-87.
11. Islam MA, Nagar S, Das S, Mukherjee A, Saha A. Molecular design based on receptor-independent pharmacophore: Application to estrogen receptor ligands. *Biol Pharm Bull* 2008;31(7):1453-60.
12. Cronin MTD, Worth AP. (Q)SARs for predicting effects relating to reproductive toxicity. *QSAR and Combinatorial Science* 2008;27(1):91-100.
13. Mukherjee S, Mukherjee A, Saha A. QSAR studies with E-state index: predicting pharmacophore signals for estrogen receptor binding affinity of triphenylacrylonitriles. *Biol Pharm Bull* 2005;28(1):154-7.
14. Kolmogorov A. Confidence limits for an unknown distribution function. *The annals of mathematical statistics* 1941;12(4):461-3.
15. Anderson TW, Darling DA. Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Annals of Mathematical Statistics* 1952;23(2):193-212.
16. Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine* 1900;50:157-75.
17. Bolboacă SD, Jäntschi L. Comparison of quantitative structure-activity relationship model performances on carboquinone derivatives. *Scientific World Journal* 2009;9:1148-66.
18. Akaike H. Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* 1969;21:243-7.
19. Kubinyi H. Variable selection in QSAR studies. I. An evolutionary algorithm. *Quantitative structure-activity relationships* 1994;13:285-94.
20. Buckland ST, Burnham KP, Augustin NH. Model selection: An integral part of inference. *Biometrics* 1997;53(2):603-18.
21. Hinkle DE, Wiersma W, Jurs SG. *Applied statistics for the behavioral sciences*. 1988, 2nd ed. Boston: Houghton Mifflin Company .
22. Bolboacă SD, Marta MM, Stoenoiu CE, Jäntschi L. Molecular descriptors family on vertex cutting: relationships between acelazolamide structures and their inhibitory activity. *Applied Medical Informatics* 2009;25(3-4):65-74.
23. Hawkins DM. The problem of overfitting. *J Chem Inf Comput Sci* 2004;44(1):1-12.
24. Stevens J. *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates, 2002.
25. Kubinyi H, Hamprecht FA, Mietzner T. Three-dimensional quantitative similarity-activity relationships (3D QSiAR) from SEAL similarity matrices. *J Med Chem* 1998;41:2553-64.
26. van Drie JH. Pharmacophore discovery - lessons learned. *Curr Pharm Des* 2003;9:1649-64.

СХОДСТВО СВЯЗЫВАНИЯ ТРИФЕНИЛ АКРИЛОНИТРИЛА С РЕЦЕПТОРАМИ ЭСТРОГЕНА – ЗАВИСИМОСТЬ МЕЖДУ ХИМИЧЕСКОЙ СТРУКТУРОЙ И СВОЙСТВАМИ

S. Bolboacă, M. Marta, L. Jäntschi

РЕЗЮМЕ

ЦЕЛЬ: Исследовать относительное сходство связывания трифенил акрилонитрила с рецепторами эстрогена с помощью анализа зависимости между химической структурой и свойствами.

МАТЕРИАЛ И МЕТОДЫ: Определенное количество уже исследованного трифенил акрилонитрила разделено на случайном принципе на учебные и тестовые комплекты (соответственно 17 и 18 соединений). С целью привести структурную информацию в структурные дескрипторы авторы использовали анализ молекулярной дескрипторной группы (MDFV). Зависимость связывания рецептора с структурным дескриптором определена с помощью полимерного линейного регрессионного анализа.

РЕЗУЛЬТАТЫ: Выведено оптимальное трипара-

метрическое уравнение с детерминационным коэффициентом 0.9580 и скрещенным достоверным параметром 0.9408. Оптимальная модель испробована на тестовом комплекте соединений, при чем получена стоимость детерминационного коэффициента 0.9004. Не отмечена сигнификантная разница между MDFV моделью и уже изученной моделью по отношению к статистическому соотношению. В отношении информационных критериев (критерии Akaike, Bayes, Akaike, and Hannan-Quinn) и функции Kibinui модель MDFV оказалась лучше, чем изученная раньше модель.

ЗАКЛЮЧЕНИЕ: Оптимальная MDFV модель успела объяснить приблизительно 96% общей дисперсии относительного сходства связывания трифенил акрилонитрила с рецепторами эстрогена, кроме того она демонстрирует оценочные и прогностические свойства. Хотя и не отмечены сигнификантные различия в статистической адекватности, оказалось, что эта модель имеет более хорошие информационные параметры, чем уже изученная модель, которая использует то же самое число молекулярных дескрипторов.