QSARs on Antibacterial Activity of Some New not yet Synthesized, Substituted Dihydrodibenzothiepins

LASZLO TARKO[1] *, CAMELIA ELENA STECOZA[2], GEORGE MIHAI NIŢULESCU[2], CORINA ILIE[3], MARIANA CARMEN CHIFIRIUC[4]

[1] Romanian Academy, "C.D.Nenitzescu" Center of Organic Chemistry, 202B Spl.Independentei, 060023, Bucharest, Romania

[2] "Carol Davila", University of Medicine and Pharmacy, Bucharest, Faculty of Pharmacy, Department of Pharmaceutical Chemistry, 6 Traian Vuia, 020956, Bucharest, Romania

[3] National Institute for Chemical Pharmaceutical R&D – ICCF, 112 Vitan Avenue, 031299, Bucharest, Romania

[4] "Cantacuzino" National Institute of Research and Development for Microbiology and Immunology, 103 Spl. Independentei, 050096, Bucharest, Romania

This paper presents result of QSAR (Quantitative Structure Activity Relationship) study realized with the PRECLAV software. The dependent property is activity against Pseudomonas aeruginosa IC13202 strain, measured by MIC values. The calibration set includes 35 dihydrodibenzothiepins. The prediction set contains 100 others not yet synthesized dihydrodibenzothiepins, having unknown observed values of activity. There are no outliers in calibration set. The identification of "significant descriptors" and the final results are influenced by "representative sample" character of calibration set within analyzed database (calibration set + prediction set). In absence of prediction set the predictive quality of QSAR is large (r^2 = 0.9492, F =186.8, r^2_{CV} = 0.9340). Large percentage, in weight, of CH_3, C_6H_3 and SO_2 molecular fragments seems to be favorable to inhibitory activity. The logP descriptor is not predictor. In presence of prediction set the predictive quality of QSAR is lower (r^2 = 0.8708, F =48.9, r^2_{CV} = 0.8084). Within the "most active ten" prediction set dihydrodibenzothiepins nine dihydrodibenzothiepins include halogen atoms and seven dihydrodibenzothiepins include SO_2 group.

Keywords: QSAR, dibenzothiepins, Pseudomonas aeruginosa, PRECLAV

Generally speaking, the prediction set includes molecules having unknown observed values of dependent property. The QSAR studies can be made in absence or in presence of certain prediction set. In the absence of the prediction set, the purpose of QSAR studies is the identification of the molecular features with the highest impact (favorable or unfavorable) on the biochemical activity. In the presence of the prediction set the purpose is to identify the prediction set molecules having the largest computed activity.

Dihydrodibenzothiepine derivatives have been studied extensively and found to have various chemical reactivity and broad spectrum of biological activities including antidepressant, antihistaminic, antipsychotic, anti-inflammatory and antimicrobial. The search for new antimicrobial active compounds represents one of the most important directions of current medicinal chemistry. Therefore, we focused our studies on synthesis and antimicrobial activity evaluation of some new dihydrodibenzothiepine derivatives.

The structures of the prediction set molecules were selected mainly by their possibility to be synthesized in our laboratory conditions and taking into account the commercial availability of the raw materials. Also, we chose preferentially halogen substituted radicals, being known their favorable effect on the antimicrobial activity.

The calibration set and the prediction set

In our QSAR study the dependent property was the antibacterial activity against *Pseudomonas aeruginosa IC13202* bacteria strain, usually defined:

$$A = \log (k/MIC) \quad (1)$$

where:

MIC is the minimum inhibitory concentration (μg/mL)

k- real number (factor)

The analyzed chemical structures and the observed values of MIC are presented in literature [1-4]. These values are within the range [62.5, 500]. The value of *k* factor in formula (1) is *k* = 1000. Therefore, the observed values of activity are within the range [0.301, 1.204].

Figure 1 and table 1 present the chemical structure of the dibenzothiepins in calibration set and the observed value of activities.

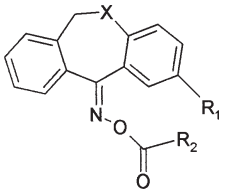

Fig. 1. Structure of the analyzed dibenzothiepins

Figure 1 and table 2 present the chemical structure of the dibenzothiepins in the prediction set.

Methods and formulas

The minimum energy geometry, for each molecule in the calibration and prediction set, was obtained by the PCModel v. 9.0 software [5], using MMX force field [6]. Then the geometry was optimized more rigorously using the quantum mechanics program MOPAC v. 9.103W [7] and included PM6 method [8]. In MOPAC analysis we used the keyword string "pm6 pulay gnorm=0.2 geo-ok bonds vectors".

In the next step, the programs MOPAC, DRAGON v. 5.4 [9] and PRECLAV v. 0907 [10-13, 21, 22] computed, for each molecule, the value for almost 2000 molecular descriptors. The molar refractivities of X, R_1 and R_2 substituents are computed using ChemSketch software [15].

**email: ltarko@cco.ro*

Table 1
SUBSTITUENTS X, R_1 AND R_2 IN CALIBRATION SET MOLECULES AND OBSERVED VALUE OF ACTIVITIES

X	R_1	R_2	Activity	X	R_1	R_2	Activity
SO_2	H	$-CH_2-CH_3$	1.204	SO_2	H	$-C_6H_3$-3,4-di-OCH_3	1.204
SO_2	H	$-C_6H_4$-*p*-C_2H_5	0.301	SO_2	H	thiophen-2-yl	1.204
SO_2	H	$-C_6H_4$-*o*-F	0.301	SO_2	CH_3	$-CH_2-CH_3$	0.903
SO_2	H	$-C_6H_4$-*m*-F	0.301	SO_2	CH_3	$-C_6H_3$-2,3-di-OCH_3	1.204
SO_2	H	$-C_6H_4$-*p*-F	0.301	SO_2	CH_3	$-C_6H_3$-2,6-di-OCH_3	1.204
SO_2	H	$-C_6H_4$-*o*-Cl	0.602	S	H	$-C_6H_4$-*m*-CH_3	0.301
SO_2	H	$-C_6H_4$-*m*-Cl	0.301	S	H	$-C_6H_4$-*p*-C_2H_5	0.301
SO_2	H	$-C_6H_4$-*p*-Cl	0.301	S	H	$-C_6H_4$-*o*-F	0.301
SO_2	H	$-C_6H_4$-*o*-Br	0.301	S	H	$-C_6H_4$-*p*-F	0.301
SO_2	H	$-C_6H_4$-*m*-Br	0.301	S	H	$-C_6H_4$-*o*-Cl	0.301
SO_2	H	$-C_6H_4$-*p*-Br	1.204	S	H	$-C_6H_4$-*p*-Cl	0.301
SO_2	H	$-C_6H_4$-*o*-I	0.301	S	H	$-C_6H_4$-*o*-Br	0.301
SO_2	H	$-C_6H_4$-*m*-I	0.301	S	H	$-C_6H_4$-*p*-I	0.301
SO_2	H	$-C_6H_4$-*p*-I	0.301	S	H	$-C_6H_4$-*o*-OCH_3	0.301
SO_2	H	$-C_6H_4$-*o*-NO_2	0.301	S	H	$-C_6H_4$-*m*-OCH_3	0.301
SO_2	H	$-C_6H_4$-*m*-NO_2	0.301	S	H	$-C_6H_4$-*o*-NO_2	0.301
SO_2	H	$-C_6H_3$-2,3-di-OCH_3	1.204	S	H	$-C_6H_4$-*m*-NO_2	0.301
SO_2	H	$-C_6H_3$-2,4-di-OCH_3	1.204				

Specific procedure [14] identified the "significant" molecular fragments. If the bond order value of chemical bond between two heavy (different from hydrogen) atoms is high, these atoms are included within the same molecular fragment. Consequently, the conjugation of certain molecular fragment with neighbor fragment(s) is low. The percents, in weight, of "significant" fragments are well correlated with the activity values.

PRECLAV program computed QSARs.

Specific procedure identified the "significant" descriptors. In presence of the prediction set PRECLAV program verifies if the calibration set (without "high outliers") is "representative sample" in calibration set + prediction set group, from the point of view of analyzed descriptor [11]. For the same calibration set and different prediction sets PRECLAV computes different QSAR equations. Consequently, in presence and absence of prediction set PRECLAV computes different QSAR equations. From this point of view, PRECLAV is fundamentally different from other QSAR computation programs, which *a priori* assume the calibration set to be a representative sample for the analyzed database.

We obtained tens thousands multilinear type (2) QSAR equations.

$$A = C_0 + \sum_{i=1}^{p} C_i \cdot D_i \quad (2)$$

where:

A is activity

C_0 - intercept

C_i - weighting factors

D_i - (values of) "significant" descriptors

p - number of QSAR descriptors

Specific formula computed the quality of equations [10]. Any descriptor included in the maximum predictive power QSAR is named "predictor". Specific procedure identified the "high outlier" molecules in the calibration set [13] and specific formula computed the relative "utility" U of predictors [10]. The predictors which present a high value (> 500) of "utility" can be considered very useful in estimating the activity, because they correlate very well with activity and do not correlate with other predictors. Each "useful" predictor offers plenty of information about why activity varies from molecule to molecule. Moreover, each "useful" predictor offers a different kind of information from the other predictors.

After computing the A_{calc} values of the activity for the prediction set molecules, PRECLAV sorts these molecules according to the estimated (computed) values. It is computed an average value A_{calc}^{m} for the estimated values and a standard deviation s of the estimated values. The program considers "high values" the activity values fulfilling the criterion (3) and "low values" the activity values fulfilling the criterion (4).

$$A_{calc} > A_{calc}^{m} + 0.5 \cdot \sigma \quad (3)$$

$$A_{calc} < A_{calc}^{m} - 0.5 \cdot \sigma \quad (4)$$

Therefore, certain prediction set molecule is "high", "moderate" or "low" active relating only to the other molecules in prediction set. In addition, the computed activity of prediction set molecules can be compared with the computed activity of calibration set molecules.

To obtain reliable enough conclusions from computed QSAR we used:

Table 2
SUBSTITUENTS X, R_1 AND R_2 IN PREDICTION SET MOLECULES

MolID	X	R_1	R_2	MolID	X	R_1	R_2
TH20	S	H	CH_3	SH20	SO_2	H	CH_3
TH40	S	H	*n*-C_3H_7	SH40	SO_2	H	*n*-C_3H_7
TH41	S	H	iso-C_3H_7	SH41	SO_2	H	iso-C_3H_7
TH50	S	H	*n*-C_4H_9	SH50	SO_2	H	*n*-C_4H_9
TH51	S	H	*sec*-C_4H_9	SH51	SO_2	H	*sec*-C_4H_9
TH52	S	H	iso-C_4H_9	SH52	SO_2	H	iso-C_4H_9
TH53	S	H	*tert*-C_4H_9	SH53	SO_2	H	*tert*-C_4H_9
TH3N	S	H	$CH{=}CH_2$	SH3N	SO_2	H	$CH{=}CH_2$
TH2M	S	H	CH_2OCH_3	SH2M	SO_2	H	CH_2OCH_3
TH2C	S	H	CH_2Cl	SH2C	SO_2	H	CH_2Cl
TH2C3	S	H	CCl_3	SH2C3	SO_2	H	CCl_3
TH3F1	S	H	$CH(F)CH_3$	SH3F1	SO_2	H	$CH(F)CH_3$
TH3F2	S	H	$(CH_2)_2F$	SH3F2	SO_2	H	$(CH_2)_2F$
TH3C1	S	H	$CH(Cl)CH_3$	SH3C1	SO_2	H	$CH(Cl)CH_3$
TH3C2	S	H	$(CH_2)_2Cl$	SH3C2	SO_2	H	$(CH_2)_2Cl$
TH3B1	S	H	$CH(Br)CH_3$	SH3B1	SO_2	H	$CH(Br)CH_3$
TH3B2	S	H	$(CH_2)_2Br$	SH3B2	SO_2	H	$(CH_2)_2Br$
TH3M	S	H	$(CH_2)_2OCH_3$	SH3M	SO_2	H	$(CH_2)_2OCH_3$
TH5R	S	H	cyclopentyl	SH5R	SO_2	H	cyclopentyl
TH5MP	S	H	CH_2- cyclopentyl	SH5MP	SO_2	H	CH_2-cyclopentyl
TH6R	S	H	cyclohexyl	SH6R	SO_2	H	cyclohexyl
TH70	S	H	CH_2-C_6H_5	SH70	SO_2	H	CH_2-C_6H_5
THH1	S	H	2-furanyl	SHH1	SO_2	H	2-furanyl
THH2	S	H	3-furanyl	SHH2	SO_2	H	3-furanyl
THH3	S	H	3-thiophenyl	SHH3	SO_2	H	3-thiophenyl
TM20	S	CH_3	CH_3	SM20	SO_2	CH_3	CH_3
TM40	S	CH_3	*n*-C_3H_7	SM40	SO_2	CH_3	*n*-C_3H_7
TM41	S	CH_3	iso-C_3H_7	SM41	SO_2	CH_3	iso-C_3H_7
TM50	S	CH_3	*n*-C_4H_9	SM50	SO_2	CH_3	*n*-C_4H_9
TM51	S	CH_3	*sec*-C_4H_9	SM51	SO_2	CH_3	*sec*-C_4H_9
TM52	S	CH_3	iso-C_4H_9	SM52	SO_2	CH_3	iso-C_4H_9
TM53	S	CH_3	*tert*-C_4H_9	SM53	SO_2	CH_3	*tert*-C_4H_9
TM3N	S	CH_3	$CH{=}CH_2$	SM3N	SO_2	CH_3	$CH{=}CH_2$
TM2M	S	CH_3	CH_2OCH_3	SM2M	SO_2	CH_3	CH_2OCH_3
TM2C	S	CH_3	CH_2Cl	SM2C	SO_2	CH_3	CH_2Cl
TM2C3	S	CH_3	CCl_3	SM2C3	SO_2	CH_3	CCl_3
TM3F1	S	CH_3	$CH(F)CH_3$	SM3F1	SO_2	CH_3	$CH(F)CH_3$
TM3F2	S	CH_3	$(CH_2)_2F$	SM3F2	SO_2	CH_3	$(CH_2)_2F$
TM3C1	S	CH_3	$CH(Cl)CH_3$	SM3C1	SO_2	CH_3	$CH(Cl)CH_3$
TM3C2	S	CH_3	$(CH_2)_2Cl$	SM3C2	SO_2	CH_3	$(CH_2)_2Cl$
TM3B1	S	CH_3	$CH(Br)CH_3$	SM3B1	SO_2	CH_3	$CH(Br)CH_3$
TM3B2	S	CH_3	$(CH_2)_2Br$	SM3B2	SO_2	CH_3	$(CH_2)_2Br$
TM3M	S	CH_3	$(CH_2)_2OCH_3$	SM3M	SO_2	CH_3	$(CH_2)_2OCH_3$
TM5R	S	CH_3	cyclopentyl	SM5R	SO_2	CH_3	cyclopentyl
TM5MP	S	CH_3	CH_2-cyclopentyl	SM5MP	SO_2	CH_3	CH_2-cyclopentyl
TM6R	S	CH_3	cyclohexyl	SM6R	SO_2	CH_3	cyclohexyl
TM70	S	CH_3	CH_2-C_6H_5	SM70	SO_2	CH_3	CH_2-C_6H_5
TMH1	S	CH_3	2-furanyl	SMH1	SO_2	CH_3	2-furanyl
TMH2	S	CH_3	3-furanyl	SMH2	SO_2	CH_3	3-furanyl
TMH3	S	CH_3	3-thiophenyl	SMH3	SO_2	CH_3	3-thiophenyl

- the physical meaning of predictors, according to MOPAC / PRECLAV / DRAGON documentation;
- the algebraic sign of predictors in QSAR;
- the computed value of utility;
- the result of virtual fragmentation.

Results and discussions

In absence of prediction set the number of "significant" descriptors is 442 and we obtained type (2) QSAR equation where:

$C_0 = 0.6171$

$C_1 = - 0.3126$

D_1 is number of CH groups in benzene rings [9] (U = 1000)

$C_2 = - 0.0137$

D_2 is product of molar refractivities computed for X and R_1 substituents (U = 685)

$C_3 = 4.5435$

D_3 is lowest eigenvalue #8 of Burden matrix weighted by atomic Sanderson electronegativities [9, 16] (U = 947)

The quality of above QSAR is large enough ($r^2 = 0.9492$; $F = 186.8$; $r^2_{CV} = 0.9340$). There are no outliers in calibration set. The minimum correlation descriptor/activity is computed for D_3 ($r^2 = 0.2257$). The maximum intercorrelation between descriptors is computed for D_1/ D_2 pair ($r^2 = 0.3552$).

The percents, in weight, of some molecular fragments are well correlated (directly or inversely) with the values of activity: C_6H_4 (r = - 0.7468), CH_3 (r = 0.6730), C_6H_3 (r = 0.5416) and SO_2 (r = 0.4662).

Because of the structure of the computed QSAR and the result of virtual fragmentation we think:

-the presence of unsubstituted benzene rings is unfavorable for the antibacterial activity;

-the presence of substituted benzene rings by alkyl groups is favorable to antibacterial activity;

-there is somehow synergetic effect of X and R_1 groups

-the presence of SO_2 (instead S) groups is favorable for the antibacterial activity;

Table 3
THE "MOST ACTIVE" TEN MOLECULES IN PREDICTION SET

MolID	Estimated activity	MolID	Estimated activity
TM2C3	1.524	SM2C3	1.505
SM3C2	1.489	SM3C1	1.455
SM2M	1.454	TM3C2	1.431
TM3B2	1.427	SM2C	1.410
SM3F2	1.405	SM3B2	1.401

-the influence of molecular shape is large but difficult to interpret;

-there is no influence of lipophilicity on biochemical activity.

In the presence of the prediction set the number of "significant" descriptors is only 44. Because of much smaller number of "significant" descriptors we think the calibration set is not quite "representative sample" in calibration set + prediction set group. We obtained different enough type (2) QSAR equation where:

C_0 = - 6.6595
C_1 = 2.3319
D_1 is R autocorrelation of lag #5 weighted by atomic polarizabilities [9, 17, 18]
C_2 = 0.5023
D_2 is 3D-MoRSE signal #15 [9, 19]
C_3 = 2.0692
D_3 is information content index (neighborhood symmetry of 1-order) [9, 20]
C_4 = - 0.2698
D_4 is topological charge index of order 3 [9]

The quality of above four predictors QSAR is good enough (r^2 = 0.8708; F = 48.9; r^2_{CV} = 0.8084). Using this equation the maximum activity computed for calibration set molecules is 1.225, the average activity computed for calibration set molecules is 0.511±0.354 and the average activity computed for prediction set molecules is 1.010±0.333.

According to criterion (3), this equation identified 39 molecules in prediction set having "high values" of activity. The large number of molecules identified as "high active" in prediction set is, probably, the statistical effect of small gap between maximum and minimum observed value of activity in calibration set, only 0.9 logarithmic units. The Table 3 includes the molecules identified as "the most active 10 molecules" in prediction set and the computed values of activity for these molecules. In table 2 prediction set 32% molecules include halogen atoms and 50% molecules include SO_2 chemical group. However, in the table 3 "most active" molecules 90% molecules include halogen atoms and 70% include SO_2 group. Favorable effect of halogen atoms and SO_2 group on activity is obvious.

Conclusions

Large percentage, in weight, of CH_3, C_6H_3 and SO_2 molecular fragments seems to be favorable for the activity.

Large percentage, in weight, of C_6H_4 molecular fragments seems to be unfavorable for the activity.

The influence of molecular shape is ambiguous and difficult to interpret.

The logP descriptor is not predictor.

The "representative sample" feature of calibration set in calibration set + prediction set group have a large influence on predictive power of computed QSAR.

The most active ten molecules in prediction set include halogen atoms and a SO_2 group.

References

1.STECOZA, C, E., CĂPROIU, M, T., DRĂGHICI, C, ILIE C., CHIRIŢĂ, I. C., Rev. Chim. (Bucureşti), **59**, no. 12, 2008, p. 1348

2.STECOZA, C. E., CĂPROIU, M. T., DRĂGHICI ,C., CHIFIRIUC, M. C., DRĂCEA, N. O., Rev. Chim. (Bucureşti), **60**, no. 2, 2009, p. 137

3.ILIE, C., STECOZA ,C. E., CĂPROIU, M. T., HĂU ,R., GUŢĂ ,R., NĂNĂU-ANDREESCU, D., Rev. Chim. (Bucureşti), **60**, no. 6, 2009, p. 588

4.CHIFIRIUC, M. C., STECOZA, C. E., DRĂCEA, O., LARION, C., ISRAIL, A. M., accepted for publication, Rom. Biotech. Lett., 2009

5.** The last version of PCModel is available from Serena Software, Box 3076, Bloomington, IN 47402-3076, USA, Internet page http://www.serenasoft.com/

6.*** Improved version of MM2, see Allinger N. L., J. Am. Chem. Soc., **99**, 1977, p. 8127

7.*** The last version of MOPAC is available from Internet page http://www.openmopac.net/

8. STEWART, J. J. P., J. Mol. Model., **13**, 2007, p. 1173

9. *** The last version of DRAGON (documentation included) is available from Talete srl., via V. Pisani, 13-20124, Milan, Italy, Internet page http://www.talete.mi.it

10.*** The last version of PRECLAV (documentation included) is available from Center of Organic Chemistry – Bucharest, email: ltarko@cco.ro or pfilip@cco.ro

11. TARKO, L., Rev. Chim. (Bucuresti), **56**, no. 6, 2005, p. 639

12. TARKO, L., LUPESCU I., GROPOSILA - CONSTANTINESCU D., ARKIVOC, 2005, X, p. 254

13. TARKO, L., STECOZA C. E., ILIE C., CHIFIRIUC M. C. Rev. Chim. (Bucuresti), **60**, no. 5, 2009, p. 476

14. TARKO, L., Rev. Chim. (Bucuresti), **55**, no. 7, 2004, p. 539

15. ** The last version of ChemSketch software is available from Advanced Chemistry Development (ACD), 110 Yonge Street, 14th floor, Toronto, Ontario, Canada M5C 1T4, see website http://www.acdlabs.com/download/chemsketch/

16.BURDEN, F.R., J. Chem. Inf. Comput. Sci. **29**, 1989, p. 225

17.CONSONNI, V., TODESCHINI, R., PAVAN, M., J. Chem. Inf. Comput. Sci., **42**, 2002, p. 682

18.CONSONNI, V., TODESCHINI ,R., PAVAN ,M., GRAMATICA, P., J. Chem. Inf. Comput. Sci., **42**, 2002, p. 693

19.SCHUUR, J.H., SELZER, P., GASTEIGER, J., J. Am. Chem. Soc., **36**, 1996, p. 334

20.MAGNUSON, V.R., HARRISS, D.K., BASAK, S.C., Physical and Theoretical Chemistry, R.B. King (Ed.), Elsevier, Amsterdam, **1983**, p. 178

21. TARKO, L., PINTILIE, L., NEGUT, C., ONISCU, C., CAPROIU, M.T., Rev. Chim. (Bucuresti), **59**, no. 2, 2008, p. 185

22. DONE, R., MANDRILA, G., TARKO, L., Rev. Chim. (Bucuresti), **60**, no. 10, 2009, p. 992

Manuscript received: 17.10.2009