Modeling of Thienyl Triazoles CDK5/p25 Inhibitors by Accessible Computational Methods

著者	SUZUKI Takahiro, RAD-CURPAN Ramona, OSTOPOVICI-HALI Liliana, GOODARZI Mohammad, SAEYS Wouter, FUNAR-TIMOFEI Simona
journal or publication title	Journal of Toyo University. Natural science
volume	57
page range	123-138
year	2013-03
URL	http://id.nii.ac.jp/1060/00005998/

Modeling of Thienyl Triazoles CDK5/p25 Inhibitors by Accessible Computational Methods

Takahiro SUZUKI,[1] Ramona RAD-CURPĂN,[2] Liliana OSTOPOVICI-HALIP,[2] Mohammad GOODARZI,[3] Wouter SAEYS,[3] Simona FUNAR-TIMOFEI[2]

Abstract

Cyclin-dependent kinase 5 plays an essential role in the development of the central nervous system during mammalian embryogenesis, being required for the maintenance of human neuronal architecture. Its deregulation has profound cytotoxic effects and has been implicated in the development of neurodegenerative diseases such as Alzheimer's disease and amyotrophic lateral sclerosis. 2-Aminothienyl triazolyl derivatives were reported as potential inhibitors of cyclin-dependent kinase 5/p25 (CDK5/p25) for the treatment of Alzheimer's disease and other neurodegenerative disorders. A series of 48 triazolyl thienyl derivatives active against CDK5/p25 was previously studied by conformational analysis performed in vacuum by the OPLS_2005 force field. In this study the obtained conformers were used to calculate structural descriptors. A Multiple Linear Regression (MLR) was applied to relate the calculated descriptors to the CDK5/p25 inhibiting activity in order to reveal important descriptors for this biological activity.

Keywords: Thienyl triazoles derivatives, MLR, CDK5/p25, MarvinSketch, Dragon

1. Introduction

Alzheimer disease (AD) is multi-factorial and heterogeneous (Iqbal & Grundke-Iqbal, 2008). Independent of the aetiology, this disease is characterized clinically by chronic and progressive dementia and histopathologically by neurofibrillary degeneration of abnormally hyperphosphorylated tau seen as intraneuronal neurofibrillary tangles, neuropil threads and dystrophic neurites, and by neuritic

[1] Natural Science Laboratory, Toyo University, 5-28-20 Hakusan, Bunkyo-ku, Tokyo 112-8606, JAPAN
[2] Institute of Chemistry Timisoara of the Romanian Academy, B-dul Mihai Viteazul 24, RO-300223 Timisoara, Romania
[3] BIOSYST–MeBioS, Faculty of Bioscience Engineering, K.U. Leuven, Kasteelpark Arenberg 30, B-3001 Leuven, Belgium

(senile) plaques of β-amyloid. The neurofibrillary degeneration is apparently required for the clinical expression of AD and in related tauopathies it leads to dementia in the absence of amyloid plaques. While normal tau promotes assembly and stabilizes microtubules, the abnormally hyperphosphorylated tau sequesters normal tau, MAP1 and MAP2 (Microtubule associated protein) and disrupts microtubules. The abnormal hyperphosphorylation of tau also promotes its self-assembly into tangles of paired helical and or straight filaments.

Tau is phosphorylated by several protein kinases (Wang et al., 2007). Cyclic AMP-dependent protein kinase (PKA), calcium, calmodulindependent protein kinas II (CaMKII), glycogen synthhase kinase-3β (GSK-3β), cyclin-dependent protein kinase 5 and its activator p25 (CDK5/p25) have been shown to be associated with tangles in AD brain. Both CDK5 and GSK-3β are associated with microtubules in the brain and phosphorylation of tau by CDK5 promotes its subsequent phosphorylation by GSK-3β.

Among the phosphatases which regulate the phosphorylation of tau, protein phosphatase-2A, the activity of which is downregulated in an AD brain, is by far the major enzyme (Iqbal & Grundke-Iqbal, 2008). The inhibition of abnormal hyperphosphorylation of tau is one of the most promising therapeutic targets for the development of disease modifying drugs.

Cyclin-dependent kinase 5 (CDK5) plays an essential role in the development of the central nervous system during mammalian embryogenesis (Mapelli & Musacchio, 2003). In adults, CDK5 is required for the maintenance of neuronal architecture. Its deregulation has profound cytotoxic effects and has been implicated in the development of neurodegenerative diseases such as AD and amyotrophic lateral sclerosis.

Dhavan and Tsai, (2001) emphasize 2-aminothienyl derivatives as potential inhibitors of CDK5/p25 for the treatment of AD and other neurodegenerative disorders (Kim et al., 2002; Misra et al., 2004a; Misra et al., 2004b)

A conformational analysis study (Rad-Curpan et al., 2011) has been previously reported for a series of triazolyl thienyl derivatives active against cyclin-dependent kinase 5/p25 (CDK5/p25) (Shiradkar et al., 2007). Its purpose was to identify the local and global minima on the potential energy surface and to predict the bioactive conformation of the most active compounds of the title series. The obtained conformers have been minimized in vacuum and aqueous environment by the OPLS_2005 force field included in the Macromodel module from Schrödinger suite (Schrödinger, LLC, New York, NY, 2008).

In this study the conformers generated previously in vacuum were used to determine the structural features which influence the CDK5/p25 inhibiting activity of these compounds. Several structural descriptors were calculated and were related to this

biological activity by Multiple Linear Regression (MLR). The stability and predictive power of the MLR models were checked by several statistical criteria.

2. Materials and Methods

Forty-eight triazolyl-thienyl derivatives (Figure 1) with the CDK5/p25 inhibiting activity expressed as the logarithm of the IC_{50} values, in nM, were used in this QSAR study. The compound structures and experimental activity were taken from reference (Shiradkar et al., 2007). The training/test sets were built taking randomly 15% of the entire series, as a test set (compounds: 2a, 4a, 4b, 8a, 10b, 12a, 18a), while the remaining 85% were used as a training set (Table 1).

Structural parameters

The molecular structures of the triazolyl thienyl derivatives (Figure 1 and Table 1) were previously minimized in vacuum by the OPLS_2005 force field included in the Macromodel module from Schrödinger suite (Rad-Curpan et al., 2011).

Twenty-two types of descriptors were calculated by the Dragon software (Dragon Professional 5.5/2007, Talete S.R.L., Milano, Italy), such as constitutional, functional groups counts, topological descriptors (MAXDP - maximal electrotopological positive variation), Burden eigenvalues (BEHe1 - highest eigenvalue n. 1 of Burden matrix / weighted by atomic Sanderson electronegativities), eigenvalue-based indices, Galvez descriptors (topological charge indices), Randic descriptors (Randic molecular profiles), RDF descriptors (radial distribution function descriptors), MWC (Molecular walk counts path counts – atomic and molecular descriptors), 3D-MoRSE, atom-centred fragments (H-050 - H attached to heteroatom, S-108 – R=S group; R represents any group linked through carbon), information indices, edge adjacency indices, topological charge indices, connectivity indices, 2D-autocorrelations, molecular properties, 2D binary fingerprints, and 2D frequency fingerprints.

By MarvinSketch software (MarvinSketch v. 5.11.3, Chemaxon Ltd., Budapest, Hungary) additional structural parameters were calculated, e.g. strongest acidic pKa, ASAPlus - solvent accessible surface area of all atoms with positive partial charge, chiral center count – the number of tetrahedral stereogenic centers, fused aromatic ring count – number of aromatic rings having common bonds with other rings, stereoisomer count –the number of *R/S* and *E/Z* isomers stereo centers.

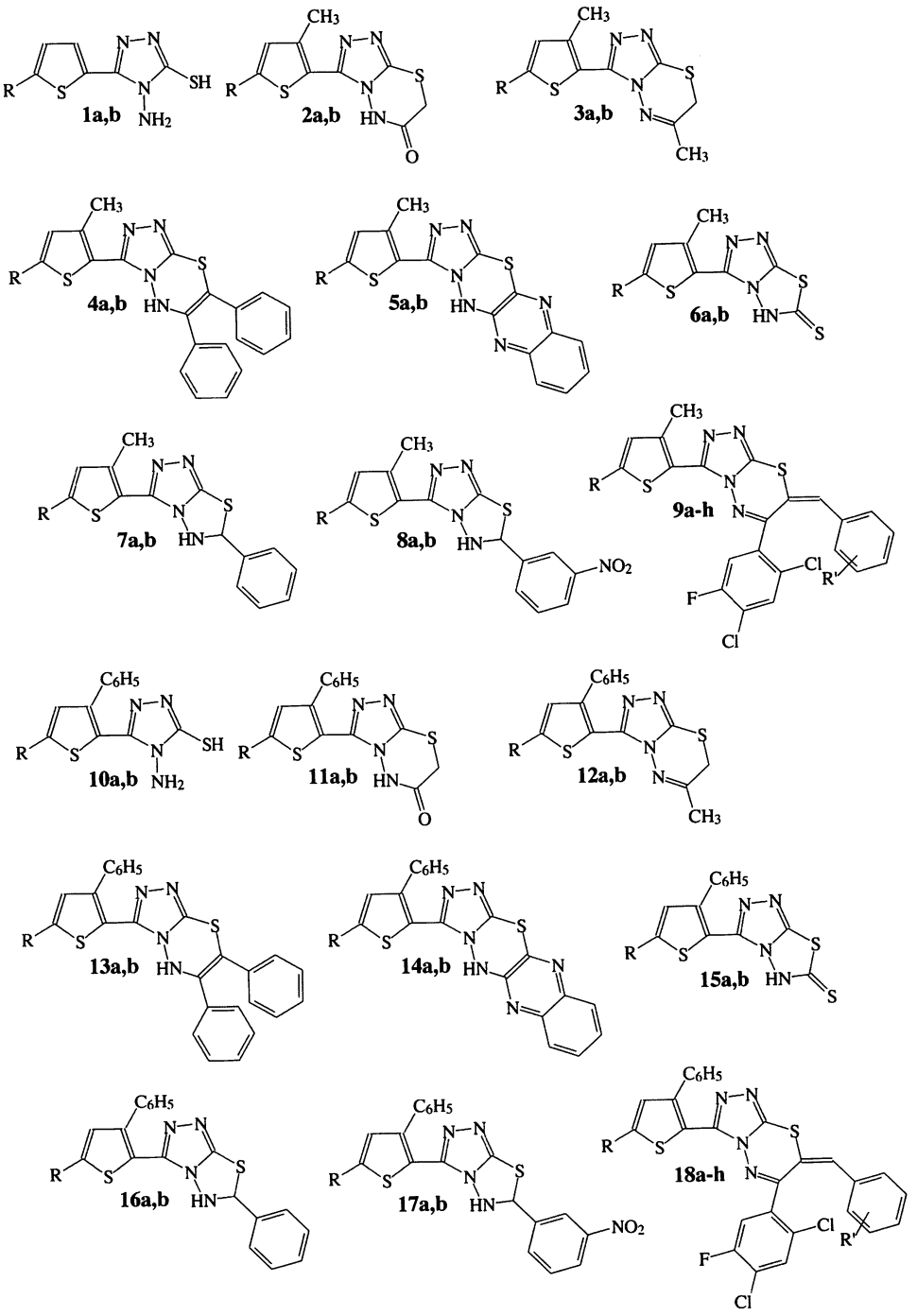

Fig 1. Triazolyl thienyl derivatives structure

Table 1. The experimental and predicted pIC_{50} values*

No	Experimental pIC_{50}	R	R'	No	Experimental pIC_{50}	R	R'
1a	7.24	$NHCOCH_3$	-	10a	7.27	$NHCOCH_3$	-
1b	7.34	$NHCOCH_2Cl$	-	***10b***	7.38	$NHCOCH_2Cl$	-
2a	6.2	$NHCOCH_3$	-	11a	6.38	$NHCOCH_3$	-
2b	6.09	$NHCOCH_2Cl$	-	11b	6.2	$NHCOCH_2Cl$	-
3a	6.19	$NHCOCH_3$	-	***12a***	6.35	$NHCOCH_3$	-
3b	6.34	$NHCOCH_2Cl$	-	12b	6.43	$NHCOCH_2Cl$	-
4a	7.36	$NHCOCH_3$	-	13a	7.4	$NHCOCH_3$	-
4b	7.14	$NHCOCH_2Cl$	-	13b	7.21	$NHCOCH_2Cl$	-
5a	7.47	$NHCOCH_3$	-	14a	7.27	$NHCOCH_3$	-
5b	7.19	$NHCOCH_2Cl$	-	14b	7.52	$NHCOCH_2Cl$	-
6a	5.49	$NHCOCH_3$	-	15a	5.69	$NHCOCH_3$	-
6b	5.13	$NHCOCH_2Cl$	-	15b	5.23	$NHCOCH_2Cl$	-
7a	7.38	$NHCOCH_3$	-	16a	7.42	$NHCOCH_3$	-
7b	7.52	$NHCOCH_2Cl$	-	16b	7.55	$NHCOCH_2Cl$	-
8a	7.19	$NHCOCH_3$	-	17a	7.22	$NHCOCH_3$	-
8b	7.28	$NHCOCH_2Cl$	-	17b	7.32	$NHCOCH_2Cl$	-
9a	5.63	$NHCOCH_3$	H	***18a***	5.84	$NHCOCH_3$	H
9b	5.47	$NHCOCH_2Cl$	H	18b	5.6	$NHCOCH_2Cl$	H
9c	5.48	$NHCOCH_3$	2-Cl	18c	5.61	$NHCOCH_3$	2-Cl
9d	5.48	$NHCOCH_2Cl$	2-Cl	18d	5.61	$NHCOCH_2Cl$	2-Cl
9e	5.47	$NHCOCH_3$	4-Cl	18e	5.61	$NHCOCH_3$	4-Cl
9f	5.53	$NHCOCH_2Cl$	4-Cl	18f	5.61	$NHCOCH_2Cl$	4-Cl
9g	5.52	$NHCOCH_3$	3-NO_2	18g	5.61	$NHCOCH_3$	3-NO_2
9h	5.51	$NHCOCH_2Cl$	3-NO_2	18h	5.57	$NHCOCH_2Cl$	3-NO_2

* test compounds are bold highlighted

Multiple linear regression (MLR)

Several MLR models (Wold & Dunn Ⅲ , 1983) were built after variable selection carried out by the Genetic Algorithm included in the MobyDigs program (Todeschini et al., 2004a) based on the RQK fitness function (Todeschini et al., 2004b). Genetic algorithm (GA) (Rogers & Hopfinger, 1994) was applied to search the feature space and select descriptors relevant to the CDK5/p25 inhibiting activity. The first step of GA is to generate a set of solutions (chromosomes) randomly which is called an initial population. Then a fitness function is used to evaluate the fitness of these individuals, and a new population is formed consisting of the fittest chromosomes as well as offspring of these chromosomes based on the notion of survival of the fittest. The leave-one-out crossvalidation fitness function was used in our study as constrained function to be optimized, a crossover/mutation trade-off parameter $T = 0.5$ and a model population size $P = 50$. Then crossover and mutation operations are performed to generate new individuals. In the subsequent selection stage, the fittest individuals evolve to the next generation. These steps of evolution continue until the stopping

conditions are satisfied. In the current work, the models were built using the simple MLR method with the selected variables from GA.

Model validity

All the statistical tests were performed at a significance level of 5 %. In MLR models, outliers can be detected by the value of residual greater than three times the value of standard error in calculation (Todeschini & Consonni, 2000), as implemented in the MobyDigs program. The Kubinyi fitness function (FIT) (Todeschini et al., 2004b) was used to check the goodness of fit of the obtained MLR models, together with other statistical criteria included in Table 2. The leave-one-out cross-validation procedure (Wold, 1978) was employed for internal validation.

The prediction ability of the MLR models was checked by the Akaike Information Criterion (AIC) (Gentleman & Wilk, 1975), y-scrambling (Lindgren et al., 1996) and bootstrapping (Efron, 1987). All these statistical metrics were calculated by the MobyDigs software.

To avoid models with collinearity without prediction power, the K multivariate correlation index (Todeschini et al., 1999) was calculated. Only models with a global correlation of [XY] block (K_{XY}) greater than the global correlation of the X block (K_X) variable can be accepted, where X is the descriptor matrix and Y is the dependent variable. To each model, the K_{XY} and K_X values were calculated.

To test the external predictive ability, the following statistical measures were used (Golbraikh et al., 2003): 1) squared correlation coefficient (R^2) between the predicted and observed activities as well as squared correlation coefficient by cross-validation (q^2); 2) coefficient of determination for linear regressions with intercepts set to zero, i.e. R_0^2 (predicted versus observed activities), and $R_0'^2$ (observed versus predicted activities); 3) slopes k and k' of the above mentioned two regression lines. All these measures were applied over the test set compounds. The following conditions should be satisfied for a model with acceptable predictive ability:

$$q^2 > 0.5 \tag{1}$$

$$R^2 > 0.6 \tag{2}$$

$$\frac{(R^2 - R_0^2)}{R^2} < 0.1 \quad \text{and} \quad 0.85 \le k \le 1.15 \tag{3}$$

$$\frac{(R^2 - R_0'^2)}{R^2} < 0.1 \quad \text{and} \quad 0.85 \le k' \le 1.15 \tag{4}$$

$$\left| R_0^2 - R_0'^2 \right| < 0.3 \tag{5}$$

The predictive ability of QSAR models was also checked based on the predictive parameter R^2_{pred}(Roy et al., 2009). For a predictive QSAR model, the value of R^2_{pred} (presented in equation 6) should be higher than 0.5.

$$R^2_{pred} = 1 - \frac{\sum\left(Y_{pred(test)} - Y_{(test)}\right)^2}{\sum\left(Y_{(test)} - \bar{Y}_{training}\right)^2} \tag{6}$$

Additional statistical parameters such as Root Mean Squared Error of Prediction (*RMSEP*), Relative Standard Error of Prediction (*RSEP (%)*) and Mean Absolute Error (*MAE (%)*) were calculated to investigate the predictive ability of the models (Goodarzi et al., 2009):

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(y_{pred} - y_{obs})^2}{n}} \tag{7}$$

$$RSEP(\%) = 100\sqrt{\frac{\sum_{i=1}^{n}(y_{pred} - y_{obs})^2}{\sum_{i=1}^{n}(y_{obs})^2}} \tag{8}$$

$$MAE(\%) = \frac{100}{n}\sqrt{\sum_{i=1}^{n}\left|(y_{pred} - y_{obs})\right|} \tag{9}$$

where y_{obs} is the observed inhibiting activity of the compound in the sample, y_{pred} the predicted activity (either as internal, cross-validated or external test set prediction) and n the number of samples in the test set.

Table 2. Final MLR models

Model	Equation	r^2	q^2	q^2_{boot}	r^2_{adj}	K_x	K_{xy}	*SDEP*	*SEE*	*F*	*AIC*	$a(r^2)$	$a(q^2)$	*FIT*
MLR 1	pIC_{50} = 9.08(±0.61) – 0.48(±0.07)strongest acidic pKa + 0.01(±0.001)ASA Plus + 2.94(±0.22)chiral center count + 0.22(±0.07)fused aromatic ring count – 2.15(±0.17)stereoisomer count	0.89	0.85	0.81	0.88	37.72	39.29	0.33	0.31	56.97	0.13	0.20	-0.04	4.28
MLR 2	pIC_{50} = –42.09(±5.06) – 0.62(±0.04)MAXDP + 12.63(±1.23)BEHe1 + 0.38(±0.06)H-050 – 1.39(±1.26)S-108	0.94	0.93	0.92	0.94	39.32	48.22	0.23	0.22	146.44	0.06	0.35	0.21	10.10

* r^2 represents the squared correlation coefficient, q^2 – leave-one-out cross-validation parameter, q^2_{boot} – bootstrapping parameter, $a(r^2)$ and $a(q^2)$ – y-scrambling variables, r^2_{adj} –adjusted r^2, SDEP – Standard Deviation Error in Prediction, *F*- Fischer test, *SEE* – Standard Error of Estimate, *AIC*- Akaike Information Criterion, the multivariate *K* correlation indices (K_x -the multivariate correlation index of the matrix of X descriptors and K_{xy} - the multivariate correlation index of the matrix of X descriptors and Y response variable), *FIT*- the Kubinyi fitness function.

3. Results and Discussion

Several MLR models were built after variable selection carried out by the Genetic Algorithm. The best prediction of CDK5/p25 inhibiting activity was obtained with the following MLR model included in Table 2.

An intercorrelation analysis of the selected molecular descriptors from the final MLR models was performed with the STATISTICA software (STATISTICA 7.1, Tulsa, StatSoft Inc, OK, USA) and is presented in Tables 3 and 4. The selected descriptors are not intercorrelated. To test model collinearity variance inflation factors (*VIF*) (Neter et al.,1985) were calculated by the STATISTICA software (STATISTICA 7.1, Tulsa, StatSoft Inc, OK, USA) for the descriptors included in the final MLR models, in addition to the *K* multivariate correlation index (Todeschini et al., 1999). According to Chatterjee and Price (1991), if *VIF* shows values >10, or if the tolerance remains below 0.10, then the model present multicolinearity. For *VIF* <5, no significant colinearity is present. The *VIF*, tolerance values and the *K* multivariate correlation index indicate the absence of multicollinearity in the final MLR models.

The applicability domain of the model with 41 training compounds was evaluated by leverage analysis expressed as Williams plot (see Figures 2 and 3), in which the standardized residuals and the leverage values were plotted. These plots confirm the absence of outliers and highly influential points (the leverage average value being of 0.146 and 0.122, respectively).

Experimental versus predicted inhibiting activity values are presented in Figures 4 and 5 for the selected MLR models.

Table 3. Correlation matrix, variance inflation factors (*VIF*) and toleranceof the selected descriptors included in MLR 1 model

	Strongest acidic pKa	ASAPlus	Chiral Center Count	Fused Aromatic Ring Count	Stereoisomer Count	*VIF*	Tolerance
Strongest acidic pKa	1					1.57	0.64
ASAPlus	0.43	1				3.03	0.33
Chiral Center Count	0.08	-0.11	1			1.51	0.66
Fused Aromatic Ring Count	0.14	-0.41	-0.18	1		1.82	0.55
Stereoisomer Count	0.28	0.69	0.32	-0.57	1	3.13	0.32

Table 4. Correlation matrix, and variance inflation factors (*VIF*) and tolerance of the selected descriptors included in MLR 2 model

	MAXDP	BEHe1	H-050	S-108	*VIF*	Tolerance
MAXDP	1				1.69	0.59
BEHe1	0.35	1			1.60	0.63
H-050	-0.60	-0.57	1		2.08	0.48
S-108	-0.29	-0.31	0.16	1	1.19	0.84

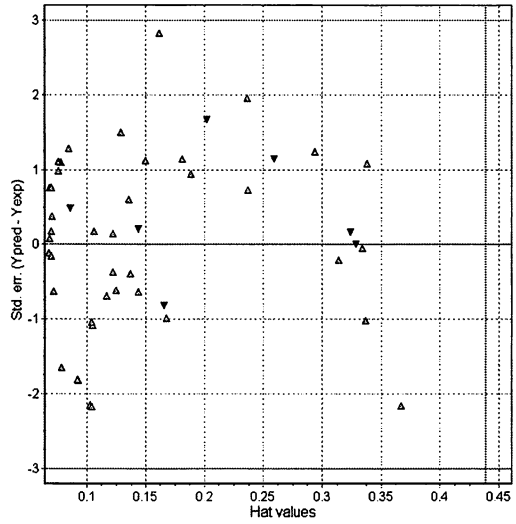

Fig. 2. Williams plot: jackknifed residuals of MLR 1 model versus leverages. Training compounds are marked by white triangles and test compounds by black triangles

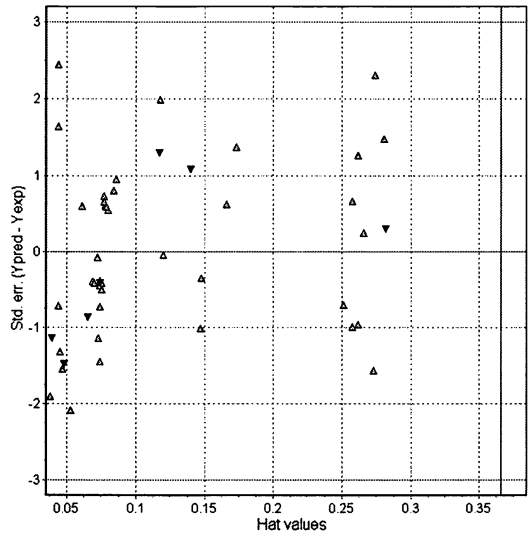

Fig. 3. Williams plot: jackknifed residuals of MLR 2 model versus leverages. Training compounds are marked by white triangles and test compounds by black triangles.

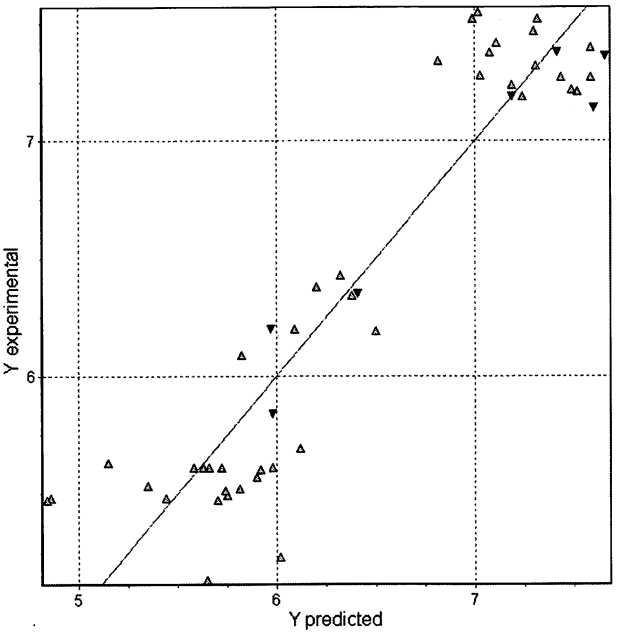

Fig. 4. Observed versus predicted inhibiting activity of MLR 1 model (training compounds are marked by white triangles and test compounds by black triangles)

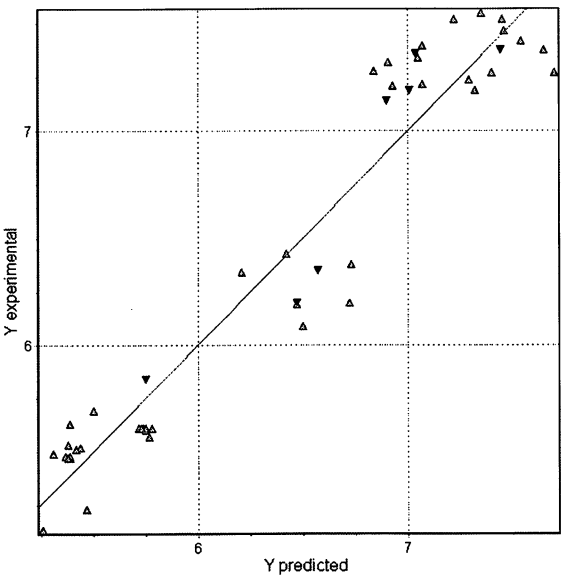

Fig. 5. Observed versus predicted inhibiting activity of MLR 2 model (training compounds are marked by white triangles and test compounds by black triangles)

Model 2 is better internally validated than model 10, according to all statistical tests (Table 2). The predictive ability of the best MLR models was tested (see Tables 5 and 6) using (*RMSEP*), Relative Standard Error of Prediction (*RSEP (%)*) and Mean Absolute Error (*MAE (%)*) (see Model Validity section). Both MLR models present a good predictive ability. In the mean time higher r^2 values must be accompanied by q^2 values as near as possible to the r^2 ones (Gramatica, 2007) (to avoid overfitting, which was, also, checked by the *RMSEP*, *RSEP* and *MAE* values).

Table 5. Predictive power results for the external test set (eqs. 1-6)

Model	R^2	$\frac{R^2 - R_0^2}{R^2}$	$\frac{R^2 - R_0'^2}{R^2}$	k	k'	$\lvert R_0^2 - R_0'^2 \rvert$	R^2_{pred}
10	0.875	0.01	0.06	1.01	0.99	0.05	0.913
11	0.927	0.04	0.01	0.98	1.02	0.03	0.898
CoMFA*	0.605	0.02	0.96	1.00	0.99	0.57	0.865
CoMSIA*	0.895	0.03	0.14	1.00	1.00	0.09	0.592

* predicted pIC_{50} values taken from [Ul Haq, et al.,2011]

Table 6. Predictive power results for the external test set (eqs. 7-9)

Model	*RMSEP*		*RSEP (%)*		*MAE (%)*	
	Training	Test	Training	Test	Training	Test
10	0.23	0.22	3.60	3.17	6.90	16.78
11	0.33	0.23	5.16	3.42	8.14	15.84
CoMFA*	0.22	0.58	3.35	0.58	6.06	23.86
CoMSIA*	0.14	0.33	2.13	5.19	5.30	18.41

* predicted pIC_{50} values taken from [Ul Haq, et al.,2011]

The maximal electrotopological positive variation, the presence of R=S groups, where R represents any group linked through carbon, high values of strongest acidic pKa and increased number of *R/S* and *E/Z* isomers stereo centers decrease the CDK5/p25 inhibiting activity. Compounds having increased values of electronegativities, higher solvent accessible surface area of all atoms with positive partial charge, H atoms attached to heteroatoms, increased number of tetrahedral stereogenic centers and of aromatic rings having common bonds with other rings are favourable for the CDK5/p25 inhibiting activity.

In a previous study of the same series of compounds (Ul Haq et al., 2011) high statistical fitting results were obtained by CoMFA (r^2 = 0.930, q^2 = 0.737, *SDEP* = 0.46, *SEE* = 0.24, F = 93.29, 5 components for the steric and electrostatic model) and CoMSIA (r^2 = 0.972, q^2 = 0.779, *SDEP* = 0.42, *SEE* = 0.15, F = 193.92, 6 components for the steric (S), electrostatic (E), hydrophobic (H), hydrogen donor (D) and acceptor (A) model) analysis. The model predictivity was checked in that study only by the predictive-r squared (correlation coefficient for test set predictions), having the value of 0.78 for the CoMFA (S,E) and 0.95 for the CoMSIA (S,E,H,D,A) models,

the authors concluded good predictive ability of these models. We calculated the predictivity criteria: Golbraikh-Tropsha tests (Golbraikh et al., 2003), R^2_{pred}, *RMSEP*, *RSE (%)* and *MAE (%)* values (Tables 5 and 6) based on the CoMFA and CoMSIA predicted CDK5/p25 pIC_{50} values (Ul Haq et al.,2011). The Golbraikh-Tropsha criteria calculated by eqs. 3 and 5 (for the CoMFA model) and the criteria calculated by eq. 5 (for the CoMSIA model) were not fulfilled. The predictive ability criteria (eqs. 7–9) do not indicate predictive CoMFA and CoMSIA models. The authors also concluded that steric requirements are playing a major role in order to optimize the biological activities of the compounds bearing large substituents at the phenyl ring substituted to the thiadiazole ring.

The simpler MLR models employed in this study using more accessible descriptors gave worse fitting results, but a better predictive ability compared to the more elaborated CoMFA, CoMSIA and docking models. The information derived from the CoMFA and CoMSIA contour maps indicates (Ul Haq et al.,2011) less information on structural features of triazolyl thienyl derivatives for the CDK5/p25 inhibiting activity in comparison to the information derived from the MLR models.

4. Conclusion

Multiple linear regression in combination with a genetic algorithm for variable selection was used to model the CDK5/p25 inhibiting activity of a series of triazolyl thienyl derivatives. The most stable conformers of these compounds obtained previously by conformational analysis performed in vacuum by the OPLS_2005 force field were used. Structural descriptors were calculated by the Dragon and Chemaxon software and were used in MLR models to extract structural features important for CDK5/p25 inhibiting activity. Several criteria for internal and external validation were employed. It was found that compounds susceptible to high acidic properties and with increased number of *R/S* and *E/Z* isomers stereo centers decrease the CDK5/p25 inhibiting activity. Favorable for this biological activity are increased positive charges on compound solvent accessible area, increased number of tetrahedral stereogenic centers and of aromatic rings having common bonds with other rings. MLR models obtained by accessible software gave acceptable statistical results with better predictive ability than the more elaborated previously published CoMFA/CoMSIA/ docking models.

Acknowledgements

This project was financially supported by Project 1.1 of the Institute of Chemistry Timisoara of the Romanian Academy. The authors are indebted to Chemaxon Ltd. for offering access to its software.

References

Chatterjee, S.; Price, B.(1991) *Regression Analysis by Example,* 2nd ed., John Wiley & Sons, New York.

Dhavan, R.; Tsai, L.H. (2001) A decade of CDK5, *Nat. Rev. Mol. Cell Biol.,* **2**, 749–759.

Efron, B.(1987) Better bootstrap confidence intervals. *J. Am. Stat. Assoc.,* **82**, 171-200.

Gentleman,J.F; Wilk, M.B.(1975) Detecting outliers. II. Supplementing the direct analysis of residuals. *Biometrics*, **31**, 387-410.

Golbraikh, A.; Shen, M; Xiao, Z.; Xiao, Y.D.; Lee, K.H.; Tropsha,A. (2003) Rational selection of training and test sets for the development of validated QSAR models. *J. Comput. Aid. Mol. Des.,* **17**, 241-253.

Goodarzi,M.; Deshpande, S.; Murugesan, V.; Katti, S.B.;Prabhakar, Y.S. (2009) Is feature selection essential for ANN modeling? *QSAR Comb. Sci.* **28**, 1487-1499.

Gramatica, P.(2007) Principles of QSAR model validation: internal and external. *QSAR Comb., Sci.,* **26**(5), 694 – 701.

Iqbal, K.; I. Grundke-Iqbal, I.(2008) Alzheimer neurofibrillary degeneration: significance, etiopathogenesis, therapeutics and prevention, *J. Cell Mol. Med.,* **12**, 38–55.

Kim, K.S.; Kimball, S.D.; Misra, R.N.; Rawlins, D.B.; Hunt, J.T.; Xiao, H.Y.; Lu, S.; Qian, L.; Han, W.C.; Shan, W.; Mitt, T.; Cai, Z.W.; Poss, M.A.; Zhu, H.; Sack, J.S.; Tokarski, J.S.; Chang, C.Y.; Pavletich, N.; Kamath, A.; Humphreys, W.G.; Marathe, P.; Bursuker, I.; Kellar, K.A.; Roongta, U.; Batorsky, R.; Mulheron, J.G.; Bol, D.; Fairchild, C.; Lee, F.Y.; Webster, K.R.(2002) Discovery of aminothiazole inhibitors of cyclin-dependent kinase 2: synthesis, X-ray crystallographic analysis, and biological activities. *J. Med. Chem.,* **45**, 3905–3927.

Lindgren, F.; Hansen, B.; Karcher, W.; Sjöström, M.; Eriksson, L.(1996) Model validation by permutation tests: Applications to variable selection. *J. Chemometr.,* **10**, 521-532.

Mapelli, M.; Musacchio,A.(2003) The structural perspective on CDK5, *Neurosignals,* **12**, 164–172.

Misra, R.N.; Xiao, H.Y.; Kim, K.S.; Lu, S.; Han, W.C.; Barbosa, S.A.; Hunt, J.T.; Rawlins, D.B.; Shan, W.; Ahmed, S.Z.; Qian, L.; Chen, B.C.; Zhao, R.; Bednarz, M.S.; Kellar, K.A.; Mulheron, J.G.; Batorsky, R.; Roongta, U.; Kamath, A.; Marathe, P.; Ranadive, S.A.; Sack, J.S; Tokarski,J.S.; Pavletich, N.P.; Lee, F.Y..; Webster, K.R.; Kimball, S.D.(2004a) N-(cycloalkylamino)acyl-2-aminothiazole inhibitors of cyclin-dependent kinase 2. N-[5-[[[5-(1,1-dimethylethyl)-2-oxazolyl]methyl]thio]-2-thiazolyl]-4-piperidinecarboxamide (BMS-387032), a highly efficacious and selective antitumor agent. *J. Med. Chem.*, **47**, 1719–1728.

Misra, R.N.; Xiao, H.; Williams, D.K.; Kim, K.S.; Lu, S.; Keller, K.A.; Mulheron, J.G.; Batorsky, R.; Tokarski, J.S.; Sack, J.S.; Kimball, S.D.; Lee, F.Y.; Webster, K.R.(2004b) Synthesis and biological activity of N-aryl-2-aminothiazoles: potent pan inhibitors of cyclin-dependent kinases. *Bioorg. Med. Chem. Lett.*, **14**, 2973–2977.

Neter, J.; Wasserman, W.; Kutner, M. H. (1985) *Applied linear statistical models: Regression, analysis of variance, and experimental designs*. Homewood, IL: Irwin.

Rad-Curpan,R.; Ostopovici-Halip, L.; Funar-Timofei, S.; Goodarzi, M.; Suzuki,T.(2011) Conformational analysis for a series of Thienyl Triazoles derivatives that act as cyclin-dependent Kinase 5/P25 inhibitors, *Journal of Toyo University, Natural Science,* No.55, 79 – 90,

Rogers, D.; Hopfinger, A.J.J (1994) Application of Genetic Function Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships, *J. Chem. Inf. Comput. Sci.*, **34**, 854-866.

Roy, P.R.; S. Paul, S.; Mitra, I.; Roy, K.(2009) On two novel parameters for validation of predictive QSAR models, *Molecules,* **14**, 1660-1701.

Shiradkar, M.R.;Padhalingappa, M.B.; Bhetalabhotala,S.;Akula, K.C.;Tupe, D.A.; Pinninti, R.R.; Thummanagoti, S.(2007) A novel approach to cyclin-dependent kinase 5/p25 inhibitors: A potential treatment for Alzheimer's disease, *Bioorganic & Medicinal Chemistry,* **15**, 6397–6406.

Todeschini, R.; Consonni, V.; Maiocchi, A.(1999) The K correlation index theory development and its application in chemometrics. *Chemom. Intell. Lab. Syst.*, **46**, 13-29.

Todeschini, R.; Consonni, V. (2000) *Handbook of Molecular Descriptors*. Wiley, Weinheim, p. 369

Todeschini,R.; Consonni, V.; Mauri, A.; Pavan, M.(2004a) In: Leardi R, editor. *Nature-inspired Methods in Chemometrics: Genetic Algorithms and Artificial Neural Networks*. Chapter 5, Amsterdam: Elsevier; pp. 141-167.

Todeschini,R.; Consonni, V.;Mauri, A.; Pavan,M.(2004b) Detecting 'bad' regression models: multicriteria fitness functions in regression analysis. *Anal. Chim. Acta,* **515**, 199-208.

Ul Haq,Z.; ,Uddin,R.; Wai, L.K.; Wadood, A.; Lajis, N.H.(2011) Docking and 3D-QSAR modeling of cyclin-dependent kinase 5/p25 inhibitors, *J. Mol. Model.*, **17**, 1149-1161.

Wang, J.Z.; Grundke-Iqbal, I.; Iqbal, K. (2007) Kinases and phosphatases and tau sites involved in Alzheimer neurofibrillary degeneration, *Eur. J. Neurosci.*, **25**, 59–68.

Wold, S.(1978) Cross validatory estimation of the number of components in factor and principal components models. *Technometrics.*, **20**, 397-405.

Wold, S.; Dunn III, W. J.(1983) Multivariate quantitative structure-activity relationships (QSAR): conditions for their applicability. *J. Chem. Inf. Comp. Sci.*, **23**, 6-13.

要　旨

サイクリン依存性キナーゼ5/p25阻害活性を有するチエニル・トリアゾール誘導体の計算化学の手法によるモデル化

鈴木孝弘, Ramona RAD-CURPĂN, Liliana OSTOPOVICI-HALIP, Mohammad GOODARZI, Wouter SAEYS, Simona FUNAR-TIMOFEI

サイクリン依存性キナーゼ5（CDK5）は、脳で発見されたセリン・スレオニン型リン酸化酵素であり、CDK5の異常活性化がアルツハイマー病や薬物依存の病態に関与していることが指摘されている。本研究では、サイクリン依存性キナーゼ5/p25に対して阻害活性を持つ48種類のトリアゾリルーチエニル誘導体について、最も活性の高い化合物を設計するために、計算化学の手法を適用した。高性能な分子設計シミュレーション・ツールである「Schrödinger suite」にあるマクロモデルモジュールを用いて、各分子の気相および水中での安定なコンフォメーションを検討した。それにより得られたコンフォメーションのエネルギーが最小化するように、Polak-Ribiere 勾配法を用いた OPLS_2005力場によって、RMS勾配度が0.01 kcal/Å･mol以下で収束する条件で探索した。その最適構造に基づいて、分子の特性を表す種々の分子記述子をDragonと Chemaxonのソフトウェアを用いて計算した。それらの分子記述子とCDK5/p25の阻害活性との定量的な相関を重回帰分析によって検討したところ、統計的に有意なQSARモデルを構築できた。本モデルによる阻害活性の予測値は、3次元QSARの手法であるCoMFA/CoMSIA/ドッキングモデルよりも優れていることを明らかにした。