Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 13 (2005) 5996–6001
Bioorganic & Medicinal Chemistry

QSAR studies of paeonol analogues for inhibition of platelet aggregation

Mukesh Doble,[a,*] S. Karthikeyan,[b] P. A. Padmawar[c] and K. G. Akamanchi[c]

[a]*Department of Biotechnology, Indian Institute of Technology Madras, Chennai 600036, India*
[b]*Center for Biotechnology, Anna University, Chennai 600025, India*
[c]*Pharmaceutical Division, Department of Chemical Technology, University of Mumbai, Matunga 400019, India*

Received 31 May 2005; revised 5 July 2005; accepted 6 July 2005
Available online 2 September 2005

Abstract—Various paeonol analogues were synthesized and tested in vitro as inhibitors of platelet aggregation. Structural properties (or descriptors) of paeonol analogues were calculated and the structure–activity relationships were determined. Several multiple linear and nonlinear regression models and back-propagation neural network model were tested and the latter using relative positive charge, hydration energy, and hydrophilic factor as inputs gave the best data fitting with $R^2 = 0.89$ and $q^2_{pre} = 0.66$. The correlation coefficient between antiplatelet inhibition activity with an interaction energy between the paeonol compounds with COX-1 enzyme is only 0.39.

1. Introduction

Blood platelets are involved in the early stages of many cardiovascular diseases namely myocardial infarction, diabetic vascular disorders, and arteriosclerosis. Aspirin and dipyridamole had been used as antiplatelet drugs, but had been identified with side effects. In addition to sodium salicylate and 3,4-dihydroxybenzoic acid, paeonol, an analgesic isolated from the root of *Paeonia suffritcosa* Andr and *Cynanchm paniculatum* (Bunge) kitaq, coumarin, chromone, and bicyclic and tricyclic derivatives of 4(3*H*)pyrimidnone[1] and 1-phenylbenzimidazoles[2] have also been found to have antiplatelet and antiischamic activity. A number (19 analogues) of analogues of this 1-(2-hydroxy phenyl)ethanone were synthetically prepared by substituting at the C-4 and the C-5 positions of the aromatic ring and were tested in vitro for antiaggregation activity triggered by ADP in blood plasma. Akamanchi et al.[3] reported the synthesis, and activity studies of these paeonol compounds.

A systematic understanding of the various substituents on the activity of the analogues helps one to rationalize the findings and also design compounds with enhanced activity. A quantitative structure–activity relationship (QSAR) study was carried out on these compounds and reported in this paper. To develop QSAR models with good predictive capability, appropriate molecular descriptors have to be chosen. Several softwares are available to calculate almost 2000 descriptors of a molecule. The problem that is faced frequently by a researcher is that of a small number of observations and a large number of molecular parameters in the descriptor pool, and one has to select the best descriptors that represent the system from this large set. At times, selecting the wrong set of descriptors could result in chance correlations or missed understanding. According to researchers, quality of QSAR results mainly depends on two factors, namely, the kind of molecular descriptor and the method used to extract the useful molecular information. These problems are addressed by applying several statistical techniques, such as cluster (dissimilarity) analysis and principal component analysis. The technique of cluster analysis involves grouping the data into subsets or clusters, such that those within each cluster are more closely related to one another than the data assigned to different clusters. The central theme of cluster analysis is calculating the degree of dissimilarity between the individual objects that are being clustered. Principal component analysis involves a mathematical procedure that transforms a number of correlated variables into a smaller number of uncorrelated variables called principal components.

Keywords: Platelet aggregation; Heat of formation; Hydrophilic factor; Highest occupied molecular orbital; Neural network models.
* Corresponding author. Tel.: +91 44 2257 4107; e-mail: mukeshd@iitm.ac.in

0968-0896/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2005.07.027

2. Results and discussion

Table 1 gives the platelet aggregation inhibitory activity of the 20 compounds. Compound **3** has the highest inhibitory activity (65.36% vs 36.97% for aspirin and 36.31% for paeonol (compound **4**) at 300 μM against 5 μM ADP). Figure 1 shows the cluster analysis of the inhibition data. It appears that the data set could be grouped into three sets based on their activity, the low (0–15%), medium (15–35%), and high activity (35–100%) sets. The high activity set has only one compound (**3**) that has R^1 = H and R^2 = OH. The low and medium sets have 9 and 10 compounds, respectively. There does not seem to be any pattern between the substituent and activity.

Table 2 lists only five descriptors, namely, relative positive charge (RPCG), hydration energy (HE), heat of formation (HF), HOMO (highest occupied molecular orbital), and hydrophilic factor (Hyf), that have a correlation coefficient value greater than 0.5 with antiaggregation activity. The remaining 414 descriptors have a correlation coefficient less than 0.5 and hence are considered not significant. A negative correlation coefficient indicates increasing the descriptor decreases the antiplatelet aggregation activity. Cross-correlation studies between these five descriptors indicate that RPCG is highly correlated with HF (~0.9) and HE with Hyf (~0.77). Thus, one could select either RPCG or HF and HE or Hyf and perform regression analysis and modeling studies with only three descriptors. Studies carried out by Tavet et al.[5] with 1,4 bis(poly and mono methoxybenzoyl)-piperazines indicated that compounds bearing hydrophilic or slightly hydrophobic substituents exhibited good antiaggregation activity. Voskressensky et al.[6] have observed that log *P* value of 2.5 was optimal lipophilicity for analogues of pyrrolopyridines to exhibit antiplatelet activity. Lipophilicity was also observed as an important structural parameter by De Marco et al.[7] with derivatives of nipecotic acid anilides. Antiplatelet activity of 2-substituted phenyl and benzimidazolyl-5-methyl-4-(3-pyridyl) imidazoles was explained with micelle-water partition coefficient, while CLOGP (a descriptor used to estimate hydrophobicity of benzene derivatives) failed to correlate with the activity.[8]

Cyclic peptides related to L-aspartic acid, *N*-[3-(aminomethyl)benzoyl]-D-valyl-N2-methyl-L-arginylglycyl-, cyclic (41 → 1)-peptide (DMP 757) containing heterocyclic and other modified linking moieties exhibited in vitro antiplatelet aggregation activity.[9] The authors postulated that the lipophilic binding pocket favored a similar linking moiety for high activity. Our studies also indicate higher the hydrophilic factor the higher the antiaggregation activity. Also, for our system log *P* has a correlation coefficient of only −0.36, and hence has not been included in model development. Studies with anilides and phenyl esters of piperidine-3-carboxylic acid have indicated that high lipophilicity and increased electron density of the phenyl ring increased the activity.[10] Similar observations were made with bis-nipecotamides with varying aromatic bridges connecting the two nipecota-

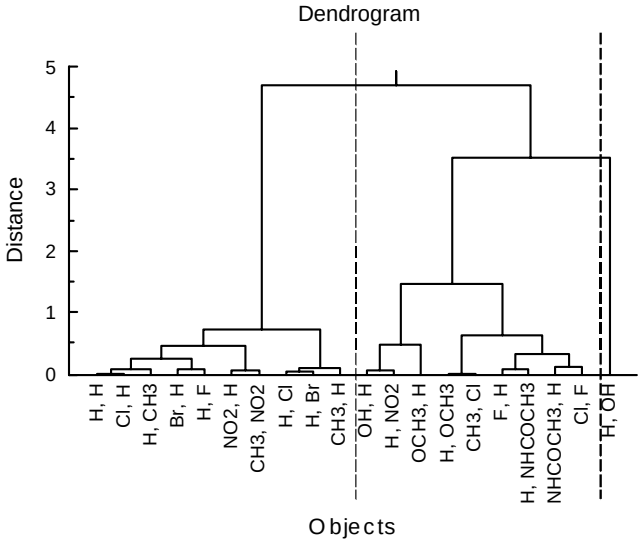
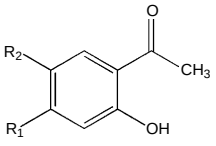


Figure 1. Results of cluster analysis.

Table 2. Descriptors that are highly correlated with antiplatelet inhibition activity

Descriptors	Correlation coefficient
Relative positive charge	−0.557
Hydration energy (kcal/mol)	−0.545
Heat of formation	−0.542
HOMO	0.512
Hydrophilic factors	0.645

Table 1. Inhibition activity of paeonol analogues



Compound No.	R_1	R_2	% Inhibition
1	H	H	4.10
2	OH	H	29.44
3	H	OH	65.36
4	OCH_3	H	36.31
5	H	OCH_3	24.47
6	C	H	3.99
7	H	Cl	8.20
8	Br	H	6.87
9	H	Br	8.78
10	F	H	1.49
11	H	F	5.44
12	CH_3	H	10.00
13	H	CH_3	3.00
14	NO_2	H	0.20
15	H	NO_2	30.40
16	$NHCOCH_3$	H	15.84
17	H	$NHCOCH_3$	20.86
18	CH_3	Cl	24.43
19	Cl	F	17.60
20	CH_3	NO_2	1.00

mide moieties. Shen et al.[2] have successfully used an electronic descriptor to describe the activities of 1-phenylbenzimidazoles. Our studies also show a positive correlation between the HOMO and activity. In addition, our studies show a negative correlation between heat of formation and activity, that is, more negative the heat of formation value (more thermodynamically stable is the molecule) higher the antiplatelet aggregation activity.

2.1. Modeling

A generic QSAR model has been suggested by Wang et al.[11] as,

$$\log[\text{Activity}] = A\log[\text{penetration}] + B\log[\text{interaction}] + C,$$

where penetration could be considered as a hydrophobicity controlled process. Interaction of the molecule and target could be due to electronic or steric[12] or electrophilic in nature or due to the molecular orbital content of the molecule.

Principal component analysis is an unsupervised learning algorithm and it is estimated from only the five descriptors short-listed and not using all the 419 descriptors. Three and four principal components are needed to explain 95.35% and 99.2% of the observed variance, respectively. Figure 2 plots the first three principal components for all the data. It is clearly seen that there are two distinct groups; the lower group relates to the medium and high activity compounds, and the upper group relates to the low activity data set. Of course, there are one or two compounds from the other group present in the wrong cluster.

Table 3 lists the various linear and nonlinear multiple regression models attempted to fit the data with the corresponding statistics. The table also lists the results of modeling studies carried out with an artificial neural network with three inputs, one hidden layer with three processing elements using TanH transfer function and Delta learning rule. Excellent model fitting is obtained with artificial neural network (ANN) models (best results are R^2 of 0.89, R^2_{adj} of 0.74, q^2_{pre} of 0.68, and F of 145.5 for a 3-3-1 model with relative positive charge, hydration energy, and hydrophilic factor as input variables). Figure 3 compares the ANN model predictions with the data. The cross-validated q^2 is good, indicating that the model has good predictive capabilities.

To test further the predictive capability of the model and also determine if the ANN was overtrained, the data were divided into two sets, namely, a learning set containing 14 data and a testing/validation set containing 6 data. The learning of the ANN was carried out using the former and the prediction with the latter data set. A 3-3-1 back propagation model using TanH transfer function and Delta learning rule gave a R^2 of 0.87 and

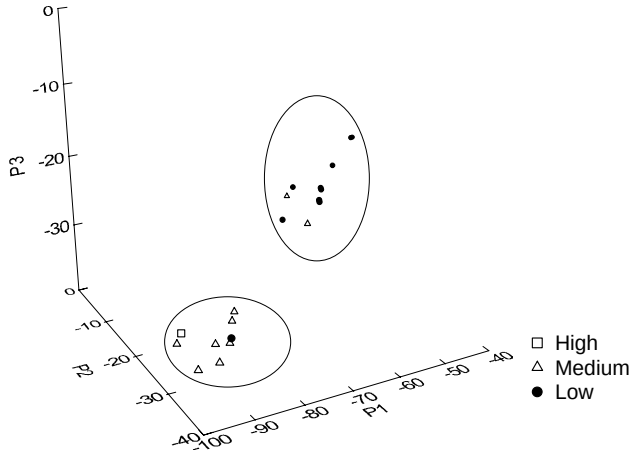


Figure 2. Plot of first three principal components of all the paeonol analogues.

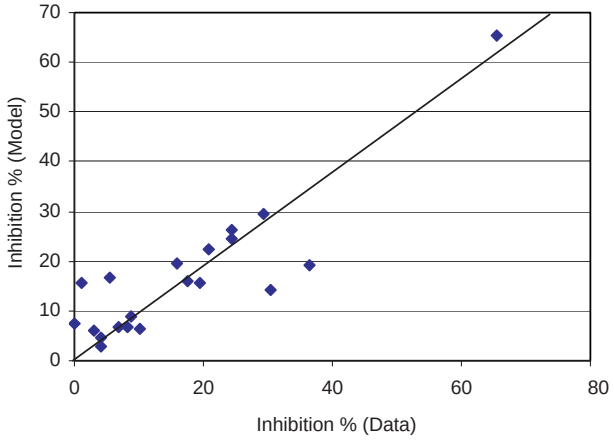


Figure 3. Comparison of back-propagation neural network model (3-3-1, inputs: HF, HOMO, Hyf, Delta learning rule, and TanH transfer function) prediction with data.

Table 3. Regression equations relating the descriptors with antiplatelet inhibition activity

Regression equation	R^2	R^2_{adj}	q^2_{pre}	RMSSE	F
18.529 + 40.610 * Hyf	0.416	0.380	0.270	11.600	12.820
−17.255 − 0.414 * HF	0.290	0.260	0.200	12.900	7.350
172.700 + 16.592 * HOMO	0.260	0.240	0.180	13.200	6.320
3.962 + 0.007 * HF − 1.710 * HOMO + 47.240 * Hyf	0.577	0.498	0.267	9.950	24.550
4.209 + 1.492 * HF − 2.456 * HOMO − 18.170 * Hyf + 0.009 * HF^2 + 0.425 * $HOMO^2$ + 126. 580 * Hyf^2	0.577	0.498	0.267	9.950	24.550
Back-propagation ANN 3-3-1, inputs: HF, HOMO, Hyf, TanH, Delta	0.780	0.700	0.640	7.160	63.800
Back-propagation ANN 3-3-1, inputs: RPCG, HE, Hyf, TanH, Delta	0.890	0.740	0.680	4.980	145.500

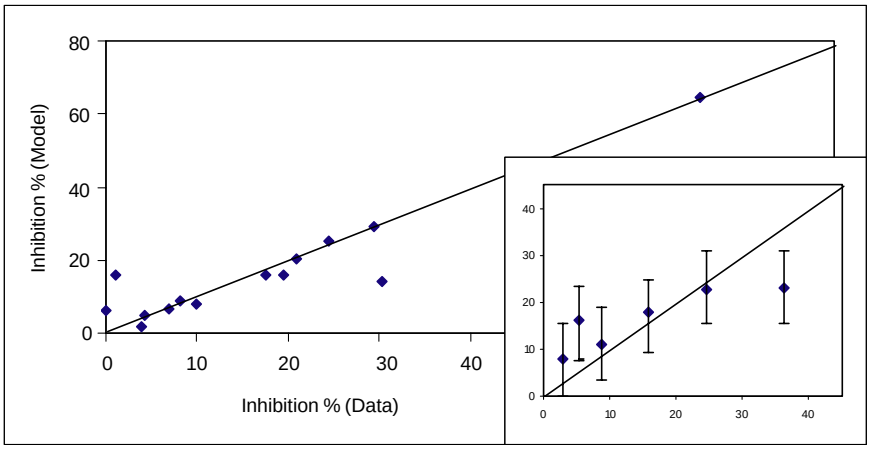


Figure 4. Comparison of back-propagation neural network model (3-2-1, inputs: RPCG, HE, Hyf, Delta learning rule, and TanH transfer function) prediction with learning data set ($n = 14$). Inset prediction with testing/validation data set ($n = 6$). Bars represent ±1 RMSSE.

0.43 for the training and prediction sets, respectively. A 3-2-1 model gave a R^2 of 0.85 (R^2_{adj} of 0.73) for the training set and a R^2 of 0.57 for the prediction set. A 3-1-1 model gave a R^2 of 0.74 and 0.28 for the training and prediction sets, respectively. This study indicated that a 3-2-1 ANN model was the best one with respect to the training, as well as prediction R^2. Figure 4 compares the ANN model predictions with the learning, as well as the test data sets for the 3-2-1 network.

2.2. Binding of paeonol analogues to COX-1

Aspirin exerts its anti-inflammatory effects through selective acetylation of serine 530 on prostaglandin H2 synthase (PGHs) (COX-1).[13] The mechanism of aspirin's antiplatelet activity is thought to work predominantly through its inhibition of cyclo-oxygenase.[14] The twelve amino acids surrounding the active sites have been identified and they are 116 VAL, 120 ARG, 349 VAL, 352 LEU, 353 SER, 355 TYR, 359 LEU, 518 PHE, 523 ILE, 527 ALA, 530 SER, and 531 LEU.[15] Molecular mechanics (MM) interaction energy between these amino acids and various paeonol analogues was determined individually. Figure 5 shows a plot of % inhibition activity as a function of the binding energy for the paeonol analogues, where the binding energy or interaction energy is the difference between the total energy of the complex and the energy of the corresponding paeonol. As seen from the figure, as the interaction energy increases the antiplatelet inhibition activity possibly also increases, although such a conclusion cannot be made with certainty since the correlation coefficient is only 0.39.

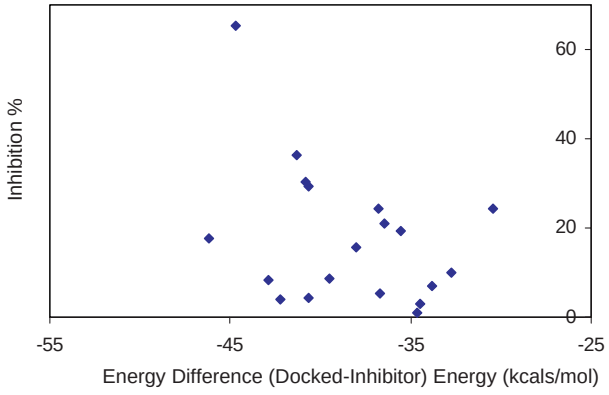


Figure 5. Relation between % inhibition activity versus binding energy for paeonol analogues with COX-1.

3. Conclusions

Twenty paeonol analogues were synthesized and tested for antiplatelet aggregation activity in vitro. QSAR studies have indicated that five best descriptors, namely, relative positive charge, hydration energy, heat of formation, HOMO, and hydrophilic factor can explain the observed anti-aggregation activity. Principal component analysis clearly shows two data clusters, representing the lower and the higher activity sets. Several multiple linear and nonlinear regression models and back propagation neural network model were developed and the latter using relative positive charge, hydration energy, and hydrophilic factor as inputs gave the best data fitting, as well as predictive R^2. The descriptors shortlisted in our study are the same as the descriptors mentioned by other researchers for several other series of anti-aggregation compounds, which indicates, irrespective of the compound, the mechanism of action is almost same. This QSAR approach developed using structural descriptors can help in understanding the properties that contribute to the action of a drug, thereby designing drugs with higher activity. Overfitting of the data will lead to a model with poor predictive capability. Hence a balance needs to be struck between overfitting the data and good predicative capability. This point was clearly seen when the data were divided into two groups, one data set used for learning/training the network and the other used for only validation/testing. The number of processing elements in the hidden layer was changed until the R^2 for the learning and the validation/testing sets were highest. Such an exercise gave a 3-2-1 back-propagation network, which had good learning as well as predictive capability.

4. Experimental

4.1. Synthesis

More details of the synthesis of these compounds are mentioned in Akamanchi et al.[3] Although many synthetic methods can be used for the preparation of these compounds, the Fries rearrangement[4] was preferred since the reactions could be performed easily and the regioisomers formed could be readily separated with good yield. Compound **2** was prepared by the Friedel–Crafts reaction of resorcinol with glacial acetic acid in the presence of anhydrous zinc chloride. Selective methylation of compound **2** using excess methyl iodide and lithium carbonate in dimethyl formamide gave paeonol (**4**). The Fries rearrangement of 4-methoxyphenyl acetate gave predominantly compound **3**. Selective methylation of compound **3** with methyl iodide and potassium carbonate in acetone gave **5**. Compounds **6**–**20** were synthesized by the Fries rearrangement by heating a mixture of anhydrous aluminum chloride and corresponding substituted phenyl acetate. The synthesized compounds were characterized by their melting point (Campbell melting point apparatus), IR spectra (Shimadzu IR 408 using KBr disk) and ^{1}H NMR spectra (Varian EM 360, 60 MHz).

Platelet aggregation inhibitory activity of these compounds was studied in vitro at King Edward VII Memorial Hospital, Mumbai, India, from blood samples taken from donors. Platelet aggregation inhibitory activity was measured using a four-channel aggrecorder[3] (more details are given in Akamanchi et al., 1999). The concentration of adenosine diphosphate eliciting the full biphasic aggregation response (~5 μM) was established using platelet-rich plasma from each donor before the determination of inhibitory potency. Test drugs were prepared by dissolving them in methanol to prepare 2% solution. Maximum percent aggregation (MPA) obtained from the aggregation trace of control (methanol) and treated (test compound) was used to calculate % inhibition with the following formula,

$$\%\ \text{Inhibition} = 100 * (1 - \text{MPA of test compound}/\text{MPA of control}).$$

4.2. Molecular modeling studies

The molecule structures of the compounds were built using HYPERCHEM® (HYPERCUBE Inc., USA) and their minimum energy conformation was determined first by minimizing the structures using molecular mechanics MM+ force field, followed by minimizing the structures with semi-empirical quantum mechanics (AM1 method with restricted Hartree–Fock (RHF)) and finally with an ab initio method (STO-3G minimal basis set). The constitutional, topological, geometrical, charge, and quantum mechanical descriptors (419 in total) were estimated using the software DRAGON® (Milano Chemometerics, Italy), and HYPERCHEM® and the correlation coefficients for all of them with inhibition activity were determined.

4.3. Statistical methods

Statistical techniques, such as Cluster Analysis, Principal Component Analysis, and Nonlinear regression analysis, were performed on the data set using SYSTAT® software (SPSS Inc., USA). The same software was used for estimating dissimilarity distance and correlation coefficients between various molecular parameters. Neural Network simulations were carried out using NEURALWARE® software (NeuralWare, Inc., PA, USA). Back-propagation neural network with TanH transfer function was used in all cases with Delta-learning rule and 16 epochs. Estimation of dissimilarity distance and correlation coefficient was carried out to identify the best set of molecular descriptors to be used in the regression and neural network models.

The goodness of the regression fits were estimated using parameters, such as R^2, R^2_{adj}, q^2_{pre} (also known as validation R^2), RMSSE, and F ratio, which are defined below,

$$R^2 = 1 - \text{SSE}/\text{TSS}$$

$$R^2_{\text{adj}} = 1 - (n-1)(1-R^2)/(n-p-1)$$

$$q^2_{\text{pre}} = 1 - \text{PRESS}/\text{TSS}$$

$$\text{RMSSE} = \text{mean sum of square of the error} = \sqrt{\text{SSE}/n}$$

$$F = (n-2)R^2/(1-R^2)$$

where n, number of data points; p, number of parameters.

$$\text{TSS} = \sum (y_{\text{data},i} - y_{\text{avg}})^2$$

$$\text{SSE} = \sum (y_{\text{model},i} - y_{\text{data},i})^2$$

$$y_{avg} = \sum y_{\text{data},i}/n$$

$y_{\text{data},\ i}$ = data points; $y_{\text{model},\ i}$ = model predictions.

PRESS = A model is developed with $(n-1)$ data points and the nth point is predicted using this model. This exercise is carried out for all the points. The sum of squares of the difference between these predicted data (using the 'leave-one-out' scheme) and the actual values is called PRESS. Predictive capability of a model is expressed by q^2_{pre}. RMSSE is an indication of the mean deviation of the prediction from the data. A large F indicates the model fit is not a chance occurrence.

References and notes

1. Roma, G.; Braccio, M. D.; Carrieri, A.; Grossi, G.; Leoncini, G.; Signorello, M. G.; Carotti, A. *Bioorg. Med. Chem.* **2003**, *11*, 123.
2. Shen, Q.; Lu, Q. Z.; Jiang, J. H.; Shen, G. L.; Yu, R. Q. *Eur. J. Pharm. Sci.* **2003**, *20*, 63–71.
3. Akamanchi, K. G.; Padmawar, P. A.; Thatte, U. M.; Rege, N. N.; Dahanukar, S. A. *Pharm. Pharmacol. Commun.* **1999**, *5*, 323.
4. Martin, R. *Org. Prep. Proced. Int.* **1992**, *24*, 373.
5. Tavet, F.; Lamouri, A.; Heymans, F.; Dive, G.; Touboul, E.; Blavet, N.; Godfroid, J. J. *J. Lipid Mediators Cell Signalling* **1996**, *15*, 145.
6. Voskressensky, L. G.; De Candia, M.; Carotti, A.; Borisova, T. N.; Kulikova, L. N.; Varlamov, A. V.; Altomare, C. *J. Pharm. Pharmacol.* **2003**, *55*, 323.
7. De Marco, M.; De Candia, M.; Carotti, C. S.; De Candia, E.; Altomare, C. *J. Pharm. Sci.* **2004**, *22*, 153.
8. Tanaka, A.; Nakamura, K.; Nakanishi, I.; Fujiwara, H. *J. Med. Chem.* **1994**, *37*, 4563.
9. Wityak, J.; Fevig, J. M.; Jackson, S. A.; Johnson, A. L.; Mousa, S. A.; Parthasarathy, A.; Wells, G. J.; Degrado, W. F.; Wexler, R. R. *Bioorg. Med. Chem. Lett.* **1995**, *5*, 2097.
10. De Candia, M.; Summo, L.; Carrieri, A.; Altomare, C.; Nardecchia, A.; Cellamare, S.; Carotti, A. *Bioorg. Med. Chem.* **2003**, *11*, 1439.
11. Wang, X.; Yin, C.; Wang, L. *Chemosphere* **2002**, *46*, 1045.
12. Schultz, T. W.; Cronin, M. T. D.; Walker, J. D.; Aptula, A. O. *J. Mol. Struct. (THEOCHEM)* **2003**, *622*, 1.
13. Loll, P. J.; Picot, D.; Garavito, R. M. *Nat. Struct. Biol.* **1995**, *2*, 605.
14. Bradford, B.; Worrall, M. D.; Karen, C. *J. Neurosurgeons Focus* **2000**, *8*, 14.
15. http://bioportal.weizmann.ac.il/oca-bin/vcofc.cgi?num=8&pdb_id=1pth&xid=00204200001115748120.