Modelling of Aldose Reductase Inhibitory Activity of Pyrrol-1-yl-acetic Acid Derivatives by Means of Multivariate Statistics

A. Tsantili-Kakoulidou[*,1], I. Nicolaou[2], D. Vrakas[1] and V.J. Demopoulos[2]

[1]*Department of Pharmaceutical Chemistry, School of Pharmacy, University of Athens, Panepistimiopolis, Zografou, Athens 15771, Greece;* [2]*Department of Pharmaceutical Chemistry, School of Pharmacy, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece*

Abstract: The inhibition of the aldose reductase enzyme (AR) is considered to be a promising approach to control chronic diabetes complications as well as a number of other pathological conditions. Thus considerable efforts are devoted to the development of aldose reductase inhibitors (ARIs) as possible pharmacotherapeutic agents. The establishment of adequate QSAR models would serve to this purpose. In the present study multivariate statistics was applied in order to analyse the AR inhibitory activity data of twenty three pyrrol-1-yl-acetic acid derivatives on the basis of essential molecular descriptors. The compounds contain one or two carbonyl keto groups, which serve as a bridge to link the pyrrole moiety to aromatic nuclei with or without further substitution. An adequate one component model with satisfactory statistics was obtained and validated for its robustness and predictive ability. The influence of the different descriptors in ARI activity is discussed. The derived model was further used to predict the activity of four independent compounds and the contribution of their specific structural characteristics in ARI activity was evaluated.

Key Words: Aldose reductase inhibitory activity, diabetes mellitus, principal component analysis, partial least squares analysis, pyrrol-1-yl-acetic acids, pyrrol-1-yl-3,5-difluoro-4-hydroxyphenyl derivatives.

INTRODUCTION

Aldose reductase (AR, ALR2, E.C. 1.1.1.21, alditol/ $NADP^+$ oxidoreductase) is the first enzyme of the so-called polyol pathway, and in the presence of NADPH it catalyses the reduction of the open-chain, aldehyde form of glucose to sorbitol. The enzyme is widely distributed in mammalian tissues and is a member of an oxido-reductase superfamily [1]. Activation of the aldose reductase enzyme is considered to be implicated in chronic diabetes complications [2] as well as in a number in other pathological conditions, like ischemic myocardial injury [3], abnormal proliferation of vascular smooth muscle cells [4] (which is an important feature of atherosclerosis, restenosis, and hypertension), and for bipolar and unipolar mood disorders [5]. Furthermore, liver cancers [6] as well as HeLa cervical carcinoma cells [7] overexpress AR which might contribute to their resistance to chemotherapy. Up to date, only two chemical classes of ARIs, hydantoins and carboxylic acids, have been evaluated in critical phase III trials [8,9], while initiatives aiming to the discovery of low molecular weight non-hydantoin, non-carboxylic acid structures as new chemotypes have been reported in literature [8-11]. According to crystallographic studies and molecular modelling of Aldose Reductase-Inhibitor Complexes, the binding site of AR could be classified into two regions: a polar region including the positively charged $NADP^+$, which accommodates the anionic moieties of carboxylic acids or spirohydantoin inhibitors and a nonpolar one, which accommodates the nonpolar segments of the inhibitors [12,13]. Beyond these general considerations on the binding requirements there is only limited successful implementation on using further (Quantitative) Structure Activity Relationships in the development of novel ARIs [14]. In this respect the establishment of adequate models for activity prediction would serve to elucidate the contribution of the essential molecular properties and structural characteristics and to design more effective ARIs.

In the present study multivariate data statistics was applied in order to analyse the AR inhibitory activity data of twenty three pyrrol-1-yl-acetic acid derivatives, previously synthesised and tested, "Fig. (**1**)" [15]. The compounds contain one or two carbonyl keto groups, which serve as a bridge to link the pyrrole moiety to aromatic nuclei with or without further substitution. In compound **18** the keto group has been replaced by a methylene bridge. The chemical and physicochemical properties of the compounds were parameterized by a variety of descriptors, which were used to establish a QSAR model applying multivariate data analysis (MVDA). In order to further evaluate the contribution of other structural characteristics in ARI activity the derived model was used to predict the activity of two compounds containing a di-fluorophenol group as acidic function and of two compounds containing a benzothiazole moiety directly attached at the pyrrole nucleus, "Fig. (**2**)", reported in references [11,16] respectively.

MATERIALS AND METHODS

The pyrrole derivatives **1**-**23**, "Fig. (**1**)", were synthesised and tested *in vitro* for ARI activity as reported in reference

*Address correspondence to this author at the Department of Pharmaceutical Chemistry, School of Pharmacy, University of Athens, Panepistimiopolis, Zografou, Athens 15771, Greece; Tel: + 30 210 7274530; Fax: +30 210 7274747, E-mail: tsantili@pharm.uoa.gr

1573-4064/05 $50.00+.00

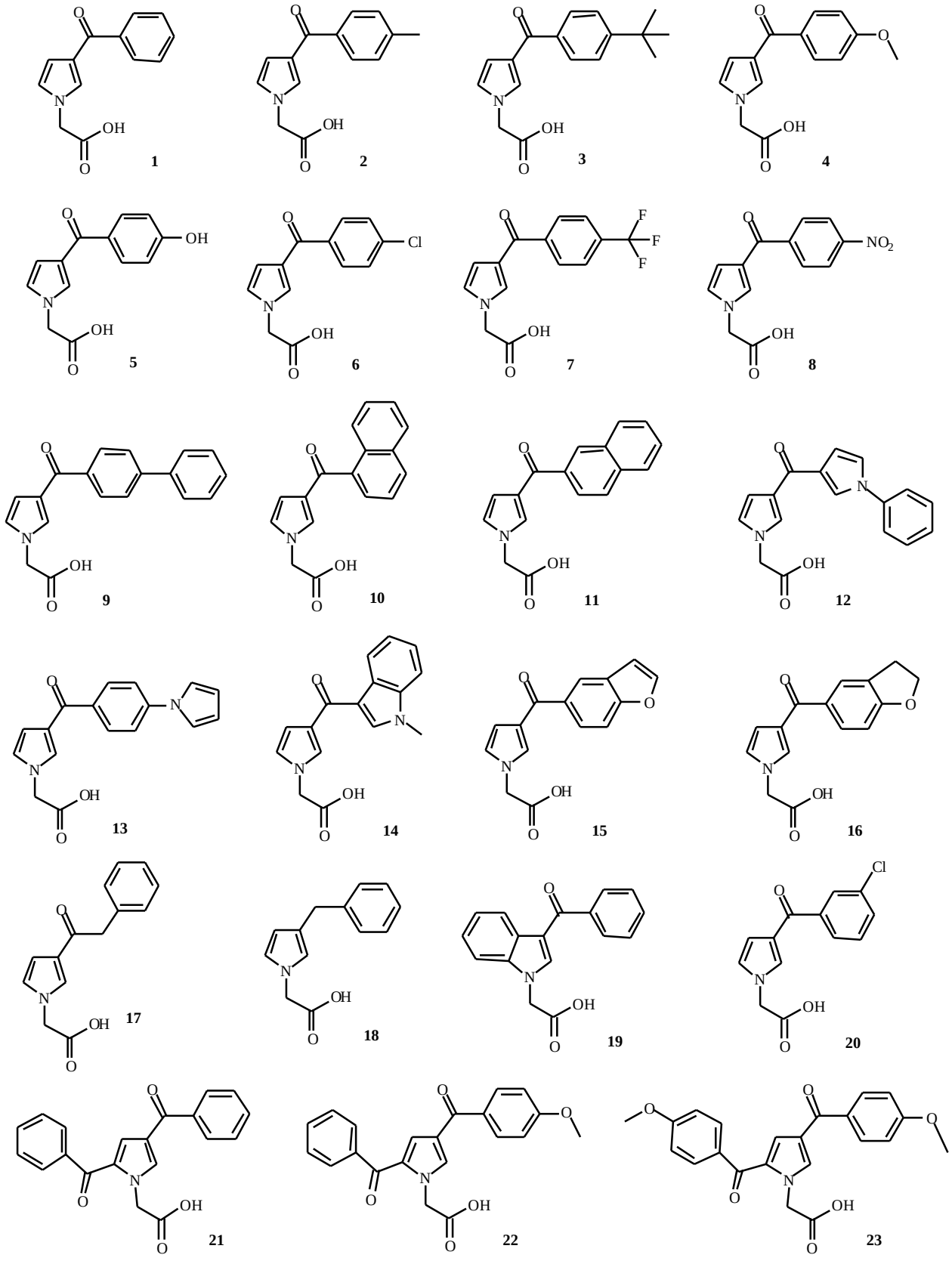
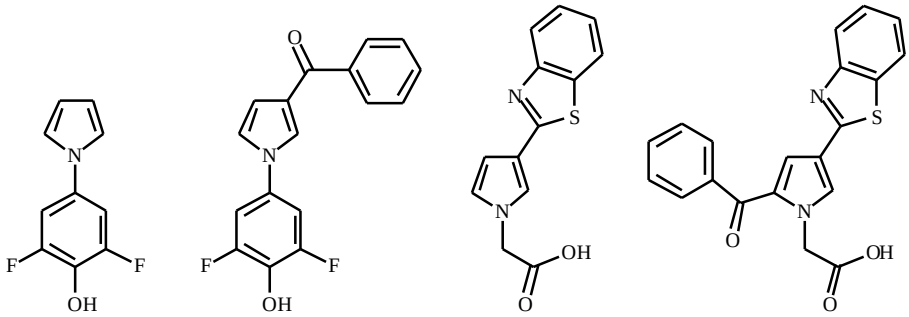


Fig. (1). Structures of the analysed pyrrol-1-yl-acetic acid derivatives **1-23**.



24 **25** **26** **27**

Fig. (2). New structures, **24-27**, evaluated by the model.

[15]. The derivatives **24**-**27**, "Fig. (**2**)", were accordingly reported in references [11,16]. The percentage of aldose reductase inhibitory activity (%Inh) was converted to logit values [17] according to the equation: $logit_{(ARI)}=log(\%Inh/(100-\%Inh)$. The %Inh refers to a concentration equal to 10^{-6} M of the compounds tested against aldose reductase isolated from rat lenses. The %Inh and the $logit_{(ARI)}$ values are presented in Table **1**.

Table 1. Percentage of AR Inhibitory Activity, Observed and Predicted $Logit_{(ARI)}$ Values of the Analysed Pyrrole Derivatives 1-23

Compound	%Inh	$logit_{(ARI)}$ observed	$logit_{(ARI)}$ predicted	Residual
1	27	-0.432	-0.327	-0.105
2	46	-0.070	-0.209	0.139
3	41	-0.158	0.045	-0.203
4	72	0.410	-0.077	0.487
5	45	-0.087	-0.144	0.057
6	41	-0.158	-0.129	-0.029
7	47	-0.052	-0.096	0.044
8	18	-0.658	0.420	-1.078
9	76	0.501	0.258	0.243
10	58	0.140	0.143	-0.003
11	79	0.575	0.187	0.388
12	67	0.308	0.227	0.081
13	80	0.602	0.234	0.368
14	38	-0.213	0.086	-0.299
15	40	-0.176	-0.001	-0.175
16	42	-0.140	-0.050	-0.090
17	29	-0.389	-0.310	-0.079
18	9	-1.005	-0.712	-0.293
19	38	-0.213	-0.034	-0.179
20	44	-0.105	-0.135	0.030
21	82	0.658	0.898	-0.240
22	93	1.123	1.200	-0.077
23	96	1.380	1.446	-0.066

Structural Geometry Optimisation and Calculation of 3D Descriptors

Geometry optimization was achieved using the program SPARTAN SGI v. 5.1.3 OpenGL (Wavefunction, Inc., 18401 Von Karman Avenue, Suite 370, Irvine, CA 92612 USA) in two steps, AM1 and Monte Carlo (MMFF94). RHF/6-31G* was used in order to calculate the following descriptors:

Surface Area and Volume of Electron Density Area (EDSA, EDVOL), the maximum and minimum Electrostatic Potentials and their difference (ESP_{max}, ESP_{min} ESP_{diff}), Energy parameters (E_{HOMO}, E_{LUMO}) and their Energy bandgap ($E_{bandgap}$) and Dipole Moment (D).

AM1-SM2 was used to calculate Hydration Energy (HE).

Optimized structures were further introduced to the program Molgen v. 4.0 (International Copyright 1996, P. Baricic, M. Mackov & J. E. Slone) to calculate the descriptors:

Water Accessible Surface Area (WASA), Water Accessible Volume (WAVOL) and Polar Molecular Surface Area (PSA, [18]).

Other Molecular Descriptors

Lipophilicity (ClogP, [19]) was calculated using the program MaclogP v. 1.0.3 (Biobyte Corp. 201 West 4th St. Suite 204 Claremont, CA 91711 USA).

Molar Refractivity (MR, [20]) was calculated using the program Molgen.

Polarizability was calculated using the program ACD v. 4.0.1 (Advanced Chemistry Development inc., 90 Adelaide Street West Toronto, Ontario M5H 3V9, Canada).

The number of hydrogen bond acceptor sites (HA) was calculated based on the total number of nitrogen and oxygen atoms [21].

In total seventeen descriptors were calculated, which are available upon request.

Statistical Analysis

Multivariate Data Analysis including Principal Component (PCA) and Partial Least Squares Analysis (PLS) was performed using SIMCA-P v.8.0 (Umetrics AB Umea, Sweden, http://www.umetri.se/) software package. The data matrix included the descriptors + the response variable, $logit_{(ARI)}$. Data were centred and scaled to unit variance. The scores plot of the two first Principal Components served as a tool to explore the uniformity in the behaviour of the compounds. PLS Analysis, a regression extension of PCA, was used to relate the information between the block of the variables X (descriptor matrix) and the response variable Y corresponding to the $logit_{(ARI)}$ [22,23]. The predictive ability of the models was evaluated using several statistical tools [24]. First, internal validation (cross validation) was applied using the default option of SIMCA-P. As a second tool permutation testing, based on the randomisation of the response data was used. Finally, the parent set of compounds was split into different training and test sets. PLS Analysis was repeated for the training sets and the models derived were used to predict the activity of the test sets.

RESULTS AND DISCUSSION

PCA Overview

PCA was initially applied to an X-data matrix containing all molecular descriptors and the $logit_{(ARI)}$ values. Six Principal Components (A=6) proved significant leading to a PCA model with R^2=0.970 and Q^2=0.800. The scores plot of the first two Principal Components was used to explore the uniformity in the behaviour of the compounds, "Fig. (**3**)". Compound **18,** which lacks a carbonyl keto group, is inactive and lies apart from the other compounds. The most active compounds **21**, **22** and **23**, having two carbonyl keto groups, form their own cluster. The nitro derivative **8** was found to be a strong outlier lying outside the Hotelling T^2 ellipse.

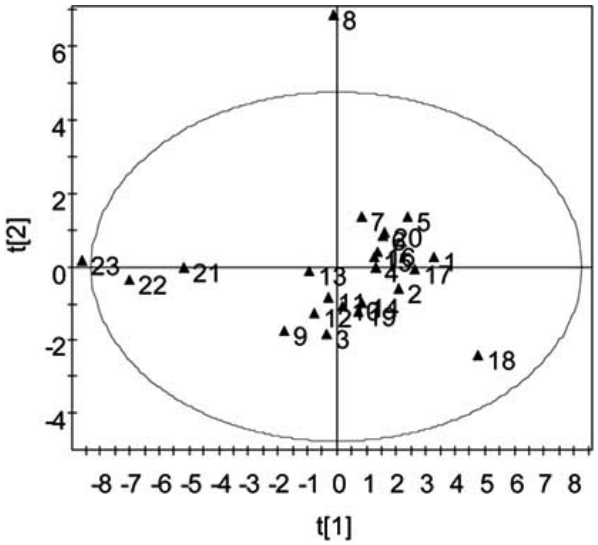


Fig.(3). PCA model: Plot of the scores of the fisrt two Principal Components t [1] versus t [2].

PLS Analysis

PLS Analysis was performed initially with the whole set of descriptors as X-matrix, followed by variable selection on the basis of the Variable Influence on Projection. The activity of compound **8** was mispredicted by any model. Taking into consideration the strongly deviating behaviour of compound **8** in the PCA scores plot, discussed above, this failure should rather be attributed to inadequate description of the chemical nature of the nitro group by the variables used. Consequently compound **8** was excluded from the statistical analysis. The most adequate PLS model was further chosen according to the following criteria: i) its simplicity concerning the number of components and the number of original variables used, ii) its validation according to the permutation test iii) its predictive ability after dividing the data set into training and test sets. These criteria were best fulfilled by a one component model (A=1) derived using eleven original variables with the following statistics:

R^2=0.838, Q^2=0.810, RMS=0.222. The predicted $logit_{(ARI)}$ values are included in Table **1**.

The histogram of the variable coefficients is presented in "Fig. (**4**)". It was found that polarity and bulk properties exerted a strong positive influence in AR inhibitory activity. The negative electrostatic potential ESP_{min} related to the carboxylic group was also found to be favourable to the activity. These results are in accordance with the postulated existence of a polar and a non polar region in the binding site. The energy parameters E_{LUMO} and $E_{bandgap}$ showed a negative contribution. In combination with the low influence of E_{HOMO}, which was therefore not included in the final model, the assumption of charge transfer complex formation, previously reported [25,26] does not seem to hold for the investigated pyrrole derivatives. A possible explanation may be that in these structures the carbonyl keto group serves as a flexible bridge and is not included in a constrained planar region for the proposed charge transfer interaction [25,26]. Lipophilicity showed a moderate positive effect when considered in the PLS analysis. Since its influence on projection was rather low it was not included in the final model.

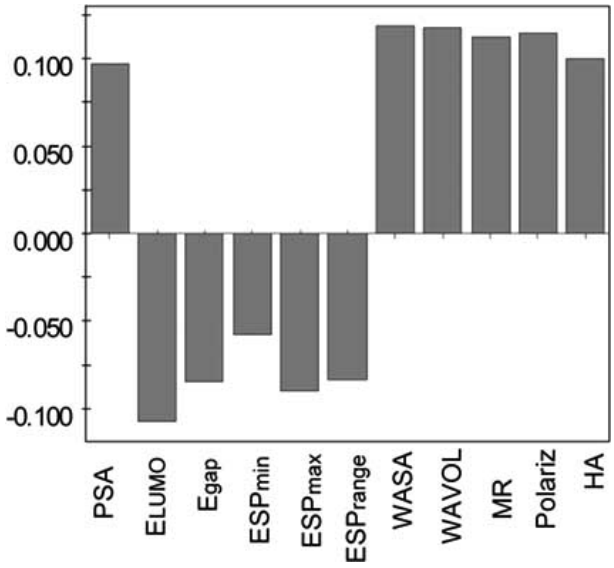


Fig. (4). Coefficients histogram of the PLS model.

Validation of the PLS Model

The overfit of the model after 10 permutations of the response variable was examined, by plotting R^2 and Q^2 versus the correlation coefficient Ry between permuted and originally data. The intercepts of both R^2 and Q^2 regression lines were below zero, indicating robustness of the model. The results of the permutation testing are presented in "Fig. (**5**)".

For further validation of the model the compound set was divided into a training set and a test set. The compounds for the test set were selected so that the whole ellipse of the PCA scores plot and the whole activity range was represented. Thus, the test set contained the compounds **2, 12, 14, 17** and **22**. Compound **8** was also not included in the training set. The training set (seventeen compounds) was subjected to

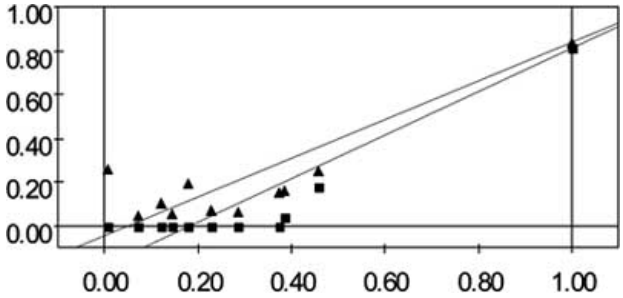


Fig.(5). Permutation test: ✿ R^2: intercept = -0.04, ✿ Q^2: intercept = -0.18.

PLS analysis using the entire X-matrix. After variable selection the same one component model with analogous statistics was obtained (A=1, R^2=0.818, Q^2=0.777, RMS=0.238). The observed versus predicted $logit_{(ARI)}$ values for both the training and the test compounds are shown in "Fig. (**6**)".

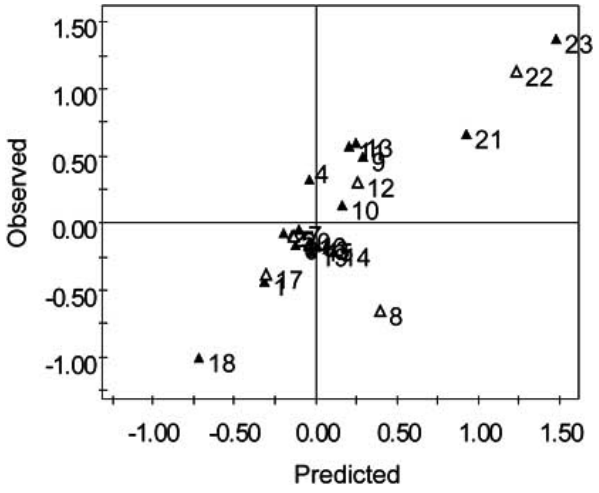


Fig.(6). Observed versus predicted logit(ARI) values: ✿ training set, Δ test set.

Correlation between observed and predicted $logit_{(ARI)}$ values of the test compounds separately was accompanied by R^2=0.920. The nitro derivative **8** remained an outlier and was not included in the correlation.

Prediction of Derivatives 24-27

The PLS model was further used to predict the ARI activity of four compounds containing structural characteristics not included in the set of the analyzed compounds. Observed and predicted $logit_{(ARI)}$ values and the corresponding residuals are listed in Table **2**.

Compounds **24** and **25** containing a di-fluorophenol group as acidic function were reasonably predicted by the PLS model. For both compounds a positive residual was obtained indicating a slightly stronger contribution of the di-fluorophenol group than predicted by the model. Concerning the benzothiazole derivatives the activity of compound **26** was considerably underestimated by the model, while that of

Table 2. Percentage of AR Inhibitory Activity, Observed and Predicted Logit$_{(ARI)}$ Values of Compounds 24-27

Compound	%Inh	logit$_{(ARI)}$ observed	logit$_{(ARI)}$ predicted	Residual
24	13	-0.826	-1.118	0.292
25	68	0.327	-0.053	0.380
26	80	0.602	-0.214	0.816
27	94	1.195	1.110	0.084

compound **27** was well predicted. It should be noted that compound **27** contains a carbonyl keto bridge, an essential feature in all active pyrrole derivatives analysed, which possibly predominates in the outcome of the prediction. This is not the case for compound **26**, which lacks a carbonyl keto group. Its higher than predicted activity should therefore be attributed to the presence of the benzothiazole ring. Actually, it has been postulated that there is a hitherto unrecognised binding site on the enzyme with strong affinity for benzothiazoles, located away from the site which binds the acidic groups [27]. This feature is not incorporated in the PLS model, since it was not present in the analysed compounds.

CONCLUSIONS

The presented PCA/PLS model is focused on a set of acidic pyrrolyl-derivatives with aldose reductase enzyme inhibitory activity. The derived model was found to be in accordance with the postulated existence of a polar and a nonpolar region at the active site of the aldose reductase enzyme. However, the developed model did not support the importance of a charge transfer complex proposed for a different chemical class of aldose reductase inhibitors (i.e. non-pyrrolyl) [25,26]. It is worth noting that the developed QSAR model of the present work, correctly revealed the strong contribution to inhibitory activity of the difluorophenol group in a pyrrolyl-derived blind test set. In contrast, the model underestimated the importance of the benzothiazole residue in this blind test set, supporting the notion for a hitherto unrecognized binding site, on the active site of the enzyme, with strong affinity for benzothiazoles.

REFERENCES

[1] Carper, D.A.; Wistow, G.; Nishimura, C.; Graham, C.; Watanabe, K.; Fujii, Y.; Hayashi, H.; Hayaishi, O.A. *Exp. Eye Res*., **1989**, *49*, 377.

[2] Miyamoto, S. *Expert Opin. Ther. Patents*, **2002**, *12*, 621.

[3] Hwang, Y.C.; Sato, S.N.; Tsai, J.Y.; Yan, S.D.; Bakr, S.; Zhang, H.P.; Oates, P.J.; Ramasamy, R. *FASEB J*., **2001**, *15*, U77.

[4] Ruef, J.; S.Q. Liu, S.Q.; Bode, C.; Tocchi, M.; Srivastava, S.; M.S. Runge, M.S.; Bhatnagar, A. *Arterioscler. Thromb. Vasc. Biol.*, **2000,** *20*, 1745.

[5] Regenold, W.T.; Kling, M.A.; Hauser, P. *Psychoneuroendocrinology*, **2000**, *25*, 593.

[6] K.W.Y. Lee, K.W.Y.; Ko, B.C.B.; Jiang, A.R.; D.L. Cao, D.L.; S.S.M. Chung, S.S.M. *Anti-Cancer Drugs*, **2001,** *12*, 129.

[7] Lee, E.K.; Regenold, W.T.; Shapiro, P. *Anti-Cancer Drugs,* **2002**, *13*, 859-868.

[8] Mylari, B.L.; Armento, S.J.; Beebe, D.A.; Conn, E.L.; Coutcher, J.B.; Dina, M.S.; O'Gorman, M.T.; Linhares, M.C.; Martin, W.H.; Oates, P.J.; Tess, D.A.; Withbroe, G.J.; Zembrowski, W.J. *J. Med. Chem.*, **2003**, *46*, 2283.

[9] Costantino, L.; Ferrari, A. M.; Gamberini, M. C.; Rastelli, G. *Bioorgan. Med. Chem.*, **2002**, *10*, 3923.

[10] de la Fuenta, J.A.; Manzanaro, S.; Martin, M.J.; de Quesada, T.G.; Reymundo, I.; Luengo, S.M.; Gago, F. *J. Med. Chem.*, **2003**, *46*, 5208.

[11] Nicolaou, I.; Zika, C.; Demopoulos, V.J. *J. Med. Chem*., **2004**, *47*, 2706.

[12] Singh, S.B.; Malamas, M.S.; Hohman, T.C.; Nilakantan, R.; Carper, D.A.; Kitchen, D. *J. Med. Chem*., **2000**, *43*, 1062.

[13] Urzhumtsev, A.; Tete-Favier, F.; Mitschler, A.; Barbanton, J.; Parth, P.; Urzhumtseva, L.; Biellman, J-F.; Podjarny, A.D.; Moras, D.A. *Structure*, **1997**, *5*, 601.

[14] Iwata, Y; Arisawa, M.; Hamada, R.; Kita, Y.; Mizutani, M.Y.; Tomioka, N.; Iitai, A.; Miyamoto, S. *J. Med. Chem.,* **2001,** *44*, 1718.

[15] Nicolaou, I.; Demopoulos, V.J. *J. Med. Chem.,* **2003**, *46*, 417.

[16] N. Zaher, N.; Nicolaou, I.; Demopoulos, V.J. *J. Enzym. Inhib. Med. Ch.,* **2002**, *17*, 131.

[17] Chalmers, D.L.; Scholz, G.H.; Topliss, D.J.; Kollinatis, E.; Munro, S.L.A.; Craik, D.J.; Iskander, M.N.; Stockigt, J.R. *J. Med. Chem.,* **1993**, *36*, 1272.

[18] Demopoulos, V.J.; Anagnostou, C.; Nicolaou, I. *Pharmazie,* **2002**, *57*, 652.

[19] Leo, A.J. *Chem. Rev*., **1993,** *93*, 1281.

[20] Ghose, A.K; Crippen, G.M. *J. Chem. Inf. Compo. Sci*., **1987,** *27*, 21.

[21] Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Freeney, P.J. *Adv. Drug. Deliv.Rev.,* **1997,** *23*, 3.

[22] Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Wold, S. Introduction to Multi-and Megavariate Data Analysis using Projection Methods (PCA and PLS), Umetrics, Umea, Sweden, **1999**.

[23] Eriksson, L.; Johansson, E. *Chemom. Intell. Lab. Syst.,* **1996,** *34*, 1.

[24] Wold, S.; Eriksson, L. In *Chemometric Methods in Molecular Design*, van de Waterbeemd, Ed; VCH, Weinheim, **1995**, pp.309-318.

[25] Kador, P.F.; Sharples, N.E. *Mol. Pharmacol.*, **1983,** *24*, 521.

[26] Kador, P.F.; Kinoshita, J.H.; Sharpless, N.E. *J. Med. Chem.*, **1985,** *28*, 841.

[27] Ashizawa, N.; Aotsuka, T. *Drugs of the Future,* **1998**, *23*, 521.

Received: 27 September, 2004 Accepted: 05 March, 2005