Journal of Computer-Aided Molecular Design, **12:** 81–93, 1998.
KLUWER/ESCOM

A multiway 3D QSAR analysis of a series of (*S*)-*N*-[(1-ethyl-2-pyrrolidinyl)methyl]-6-methoxybenzamides

Jonas Nilsson[a,*], Evert J. Homan[b], Age K. Smilde[c], Cor J. Grol[b] & Håkan Wikström[b]
[a]*Department of Structural Chemistry, Pharmacia & Upjohn, S-75281 Uppsala, Sweden;* [b]*Department of Medicinal Chemistry, University Centre for Pharmacy, Antonius Deusinglaan 1, 9713 AV Groningen, The Netherlands;* [c]*Laboratory for Analytical Chemistry, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands*

Received 11 April 1997; Accepted 22 August 1997

Key words: alignment, molecular modeling, multilinear PLS, multiway calibration, PARAFAC, 3D QSAR

Summary

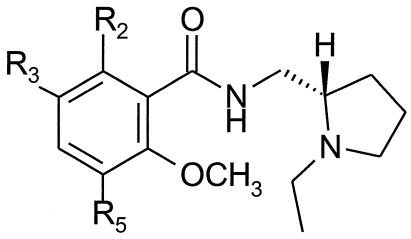
Recently, the multilinear PLS algorithm was presented by Bro and later implemented as a regression method in 3D QSAR by Nilsson et al. In the present article a well-known set of (*S*)-*N*-[(1-ethyl-2-pyrrolidinyl)methyl]-6-methoxybenzamides, with affinity towards the dopamine D_2 receptor subtype, was utilised for the validation of the multilinear PLS method. After exhaustive conformational analyses on the ligands, the active analogue approach was employed to align them in their presumed pharmacologically active conformations, using (–)-piquindone as a template. Descriptors were then generated in the GRID program, and 40 calibration compounds and 18 test compounds were selected by means of a principal component analysis in the descriptor space. The final model was validated with different types of cross-validation experiments, e.g. leave-one-out, leave-three-out and leave-five-out. The cross-validated Q^2 was 62% for all experiments, confirming the stability of the model. The prediction of the test set with a predicted Q^2 of 62% also established the predictive ability. Finally, the conformations and the alignment of the ligands in combination with multilinear PLS, obviously, played an important role for the success of our model.

Introduction

Ever since Cramer III et al. [1] introduced the CoMFA methodology, it has been available as implemented in the SYBYL molecular modeling package [2]. During the last years, new methods for the alignment of the ligands [3,4] have evolved, and other methods to generate 3D descriptors [5] have become available. Additionally, several attempts have been made to reduce the number of variables [6,7] but less efforts have been undertaken to improve the multivariate analyses. Nilsson et al. [8] introduced multilinear PLS [9] (N-PLS) as a regression method in 3D QSAR and demonstrated several advantages, as compared to bilinear PLS, i.e. traditional PLS. It was shown that multilinear PLS is more stable, increases the predictive ability, and improves the interpretation of the results. However, we still need to evaluate whether multilinear PLS really is suitable for 3D QSAR data, or if it just happened to be successful for the series of compounds used by Nilsson et al. [8,10]. Therefore, a well-known series of substituted *N*-[(1-ethyl-2-pyrrolidinyl)methyl]-derived benzamides [11,12] (**1**–**58**, Scheme 1) was chosen for a re-evaluation of this method.

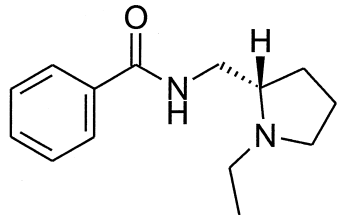
Compounds of this chemical class have been shown to possess high affinity and selectivity towards dopamine D_2 receptors, and, as a consequence, several compounds of this series have become valuable tools for in vivo and in vitro receptor binding studies (e.g. ^{3}H-raclopride (**1**) and ^{125}I-NCQ298 (**9**) [13], as well as for in vivo visualisation techniques like PET (e.g. ^{11}C-raclopride (**1**) and ^{11}C-eticlopride (**36**) and SPECT (e.g. ^{123}I-NCQ298 (**9**) [13] (see Table 1).

*To whom correspondence should be addressed.

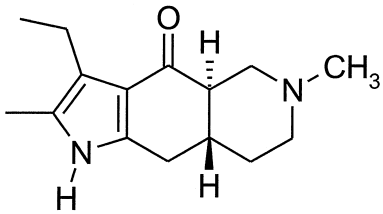


1–58

Scheme 1. Structural formulas of compounds **1**–**58**. For the definition of substituents R_2, R_3 and R_5, see Table 1.



59



60

Scheme 2. Structural formulas of (*S*)-*N*-[(1-ethyl-2-pyrrolidinyl)-methyl]benzamide (**59**) and (–)-piquindone (**60**).

Theory and methods

Prior to the multilinear PLS analysis, the conformational space accessible to the ligands was sampled by performing an exhaustive conformational analysis (see below) on the basic skeleton common to all ligands (Scheme 2, **59**). The active analogue approach was then employed to derive the presumed pharmacologically active conformation, using the rigid pyrrolo-isoquinoline (–)-piquindone (Scheme 2, **60**) as a template (see below). The best matching conformation of **59** then served as a starting point for the construction of compounds **1**–**58**, which were subsequently optimised with respect to the orientations of the aromatic substituents. Finally, all ligands were superimposed in their presumed pharmacologically active conformations (see below).

Throughout this paper lower-case characters represent scalars, boldface lower-case characters represent vectors, boldface upper-case characters represent matrices, and underlined boldface upper-case characters represent multiway matrices.

Conformational analyses

Conformational analyses and pharmacophore identification (see below) were performed essentially as described by Jansen et al. [14]. Thus, conformational analyses were performed within MacroModel v. 4.5 [15], using the MM2* force field and the Monte Carlo (MC) search protocol. (*S*)-*N*-[(1-Ethyl-2-pyrrolidinyl)methyl]benzamide (**59**) was built from standard fragments and subsequently energy-minimised with default options. During all minimisations, the benzamide torsional angle was kept fixed at 0 °. For (–)-piquindone (**60**), its X-ray crystal structure (BANKIK) [16] served as the starting conformation. The conformational space accessible to both compounds was sampled by submitting the input structures to 1000 MC steps. Starting geometries were generated systematically (SUMM option) and the number of torsional angles to be adjusted in each MC step was randomly varied between 2 and n–1, n being the total number of variable torsional angles. Ring closure bonds were defined in five- and six-membered non-aromatic rings in order to allow torsional angles in these rings to be varied during the MC search procedure. Ring closure distances were limited between 0.5 and 2.0 Å. Default values were used for testing high-energy non-bonded contacts. Duplicate minimum-energy conformations (all pairs of equivalent atoms separated by less than 0.25 Å) were determined by least-squares superposition of all non-hydrogen atoms and rejected. Chiral centres were checked for conservation of the original stereochemistry before saving conformations. An energy cut-off of 21 kJ mol^{-1} was applied. Minimisations were performed using the truncated Newton conjugate gradient (TNCG) minimiser, allowing for 250 iterations per structure. A gradient of 0.01 kJ $Å^{-1}$ mol^{-1} was set as the initial convergence criterion. The local minimum-energy conformations thus obtained were submitted to a final minimisation, using the full matrix Newton-raphson (FMNR) minimiser, allowing for line searching and 1000 iterations per structure. A gradient of 0.002 kJ $Å^{-1}$ mol^{-1} was set as the final convergence criterion.

Table 1. Aromatic substitution patterns and dopamine D_2 receptor binding properties of compounds **1**–**58**

Compound	R_2	R_3	R_5	Activity[a]			Compound	R_2	R_3	R_5	Activity[a]		
				Exp[b]	Fitted[c]	Pred[d]					Exp[b]	Fitted[c]	Pred[d]
1	OH	Cl	Cl	7.49	7.75	–	30	OH	Br	NO_2	6.73	6.70	–
2	OH	OMe	Cl	7.15	7.44	–	31	OH	I	H	8.52	8.12	–
3	H	Br	Br	8.10	7.94	–	32	OH	Me	Cl	7.59	7.96	–
4	H	Et	Br	7.96	8.25	–	33	OH	Me	Me	8.11	7.98	–
5	H	I	OMe	9.17	9.14	–	34	OH	Et	H	8.54	8.32	–
6	OH	*n*-Pr	Me	8.30	8.27	–	35	OH	Et	F	8.82	8.88	–
7	OH	Cl	*n*-Pr	6.96	7.16	–	36	OH	Et	Cl	9.04	8.62	–
8	OH	H	Et	6.91	6.82	–	37	OH	Et	Br	8.64	8.32	–
9	OH	I	OMe	9.54	9.43	–	38	OH	*n*-Pr	H	8.30	7.91	–
10	H	SMe	OMe	8.96	9.09	–	39	OH	OMe	H	6.69	7.09	–
11	OH	Et	OMe	8.89	9.61	–	40	OH	OMe	Br	7.17	7.12	–
12	H	*n*-Bu	OMe	8.57	8.57	–	41	H	Br	OH	8.00	–	8.22
13	H	*n*-Pr	H	7.17	7.45	–	42	OH	*n*-Pr	Cl	8.49	–	8.20
14	H	Cl	H	6.59	6.98	–	43	OH	Me	Br	8.26	–	7.65
15	H	Cl	Cl	7.70	7.72	–	44	H	Me	OMe	8.28	–	8.34
16	H	Cl	Br	8.25	7.53	–	45	OH	Br	Et	7.77	–	8.01
17	H	Br	H	7.34	7.38	–	46	OH	Et	Et	8.75	–	8.58
18	H	Br	OMe	8.92	8.64	–	47	H	Et	OMe	8.89	–	9.12
19	H	Et	Cl	8.38	8.41	–	48	H	H	OMe	7.28	–	7.68
20	OH	H	Cl	7.19	6.96	–	49	H	Et	H	7.40	–	7.82
21	OH	H	OMe	8.06	7.83	–	50	OH	H	H	6.50	–	6.52
22	OH	F	H	6.44	6.70	–	51	OH	H	Br	7.25	–	6.67
23	OH	Cl	H	7.41	7.26	–	52	OH	Cl	Me	7.96	–	7.76
24	OH	Cl	Br	7.24	7.45	–	53	OH	Cl	OMe	8.77	–	8.45
25	OH	Cl	Et	7.92	7.63	–	54	OH	Br	F	8.15	–	8.33
26	OH	Br	H	8.08	7.70	–	55	OH	Br	Me	7.96	–	8.09
27	OH	Br	Cl	7.77	8.11	–	56	OH	Me	H	7.72	–	7.52
28	OH	Br	Br	7.59	7.80	–	57	OH	Me	*n*-Pr	6.85	–	7.41
29	OH	Br	OMe	8.85	8.90	–	58	OH	NO_2	H	5.52	–	7.30

[a] Activity values are expressed in pIC_{50} molar units.
[b] Experimental values are obtained from Ref. 11.
[c] Fitted values (training set).
[d] Predicted values (test set).

Pharmacophore identification

The pharmacophore-identifying program APOLLO [3] was used to align **59** and **60** in their presumed pharmacologically active conformations. The output files of the MC searches, containing multiple minimum-energy conformations of each compound, served as input for the VECADD module of APOLLO. In each conformation of **59** and **60**, extension vectors from the carbonyl O atom and the basic nitrogen atom pointing towards putative receptor points, as well as a centroid and a normal through the aromatic ring were defined. In all cases, minimum densities of vectors were specified, representing ideal hydrogen bonding positions. The RMSFIT module was used to determine the conformation of each ligand which gave the best overall fit with respect to the specified fitting points. When receptor points emanating from the carbonyl oxygens were fitted, the two possible points were defined as choices. When fitting aromatic rings, both extremes of the normal and the centroid were defined as choices. All points were weighed equally. Conformational energies were taken into account for determining the root mean square (rms) deviations. The rms cut-off was set to 0.5 Å. The MMDFIT module was used to extract the conformations which gave the best matches.

Substituent geometries and ligand alignment

The best fitting conformation of **59** identified by APOLLO was used as a starting point for building the compounds **1**–**58** in Table 1. The appropriate substituents were attached to the aromatic ring and the conformational space of substituents with conformational freedom was probed using the MC procedure in MacroModel as described above. For each newly introduced torsional angle, 100 MC steps were performed. All other torsional angles were fixed during these MC searches, in order to maintain the overall geometry of the pharmacophoric pattern for all compounds. The lowest energy conformations resulting from these MC searches were used for the final alignment, which was performed within SYBYL v. 6.3 [2]. The basic nitrogen atoms were protonated and centroids were defined in the aromatic rings. Then all compounds were superimposed with respect to the centroids, the carbonyl O atoms, the amide H atoms, and the protons on the protonated tertiary nitrogen atoms, using the *Fit Atoms* procedure as implemented in SYBYL.

The data set

The descriptors used in this paper were generated in the GRID program [17] from three different probe atoms, the C3 probe, the CA+2 probe and the OH2 probe, reflecting the steric field, the electrostatic field and the hydrogen bonding field, respectively.

$$E_{tot} = E_{ste} + E_{ele} + E_{hb} \quad (1)$$

In contrast to SYBYL/CoMFA [2], the same interactions (of Equation 1) are calculated in each grid point independent of the type of probe atom. Thus, there are no separate steric or electrostatic fields generated, but instead the type of probe will reflect different types of fields due to the fact that the relative size of the terms in Equation 1 differs from one probe to another. For example, a charged probe, e.g. CA+2, reflects predominantly the electrostatics since E_{ele} will be the dominating term, but still the other two terms are not excluded.

Multilinear PLS analysis

Prior to traditional bilinear PLS [2,18] modelling, the data set in Figure 1 is unfolded to form a two-way matrix where each row contains the variables (the grid points) describing each molecule. The **X**-matrix is subsequently decomposed into a score vector $\mathbf{t}_1$ (I × 1) and a weight vector $\mathbf{w}_1$ (JKLM × 1) where the score vector ($\mathbf{t}_1$) is determined to have the property of maximum covariance with the dependent variable **y**, as described for the first PLS component in Figure 2a. Subsequently, the score vectors, i.e., the columns in **T** ($\mathbf{t}_1$,…,$\mathbf{t}_A$; A is the number of components) replace the original variables as regressors.

The unfolding step is not necessary in multilinear PLS but instead the decomposition is performed directly on the data set in Figure 1. The one-component decomposition (Figure 2b) comprises a score vector **t** (I × 1) and four weight vectors $\mathbf{w}^J$ (J × 1), $\mathbf{w}^K$ (K × 1), $\mathbf{w}^L$ (L × 1) and $\mathbf{w}^M$ (M × 1) corresponding to the molecular direction, the grid x-direction, the grid y-direction, the grid z-direction and the probe direction, respectively, as described in Figure 1. The decomposition procedure resembles a five-way PARAFAC decomposition [19,20]. In multiway analysis it is not always obvious which direction corresponds to the objects but in this data set it is the molecular direction. Consequently, the data were mean-centred in the molecular direction prior to analysis. Analogously to bilinear PLS, for given $\mathbf{W}^J$ ($\mathbf{w}_1^J$, …, $\mathbf{w}_A^J$), $\mathbf{W}^K$ ($\mathbf{w}_1^K$, …, $\mathbf{w}_A^K$), $\mathbf{W}^L$ ($\mathbf{w}_1^L$, …, $\mathbf{w}_A^L$) and $\mathbf{W}^M$ ($\mathbf{w}_1^M$, …, $\mathbf{w}_A^M$) the score vectors **T** ($\mathbf{t}_1$, …,$\mathbf{t}_A$) are determined to have the property of maximum covariance with **y**. Hence, the multilinear PLS model is a *partial* least-squares approximation of $\underline{\mathbf{X}}$:

$$x_{ijklm} = t_i W_j^J W_k^K W_l^L W_m^M + e_{ijklm} \quad (2)$$

The multilinear PLS algorithm [9] was recently implemented as a regression method in 3D QSAR and compared with bilinear PLS by Nilsson et al. [8].

Validation

The cross-validation experiments and the external predictions were quantified with Q^2 (CV) and Q^2 (Pred), and the quality of the calibrations with R^2, respectively. The predicted y in Equation 3 is denoted by $\hat{y}_{(i)}$, i.e., in the case of cross-validation an estimation of y_i using a model with the ith object excluded. In the case of external predictions, y_i is the response of the ith test object estimated with the complete calibration model.

$$Q^2 = \left[1 - \left(\sum_{i=1}^{I}(y_i - \hat{y}_{(i)})^2 / \sum_{i=1}^{I}(y_i - \overline{y})^2\right)\right] \times 100. \quad (3)$$

The fitted y from the calibration, in Equation 4, is denoted by $\hat{y}_i$

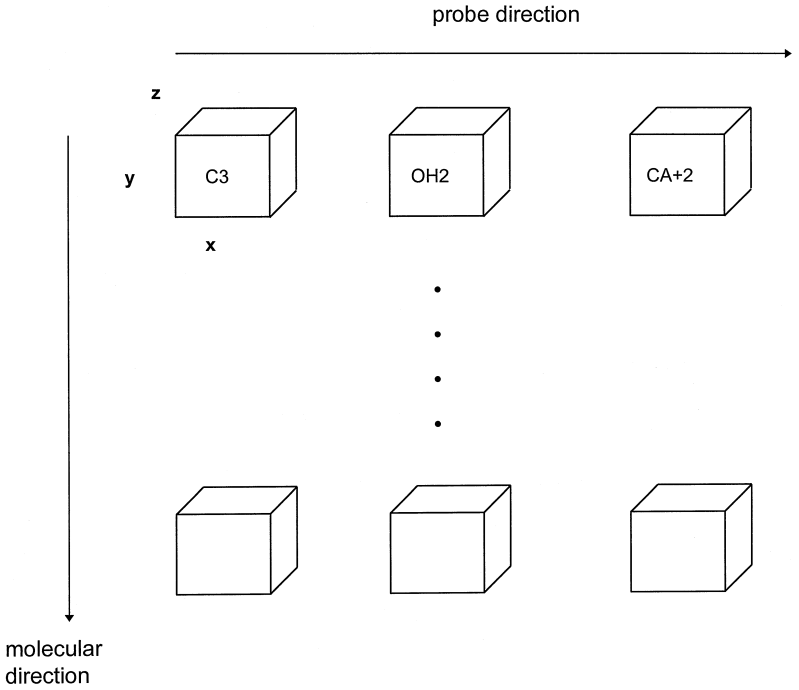


Figure 1. The complete data set defining five directions. The object direction is of 40 steps, the x-direction is of 20 steps, the y-direction is of 22 steps, the z-direction is of 20 steps and the probe direction is of 3 steps.

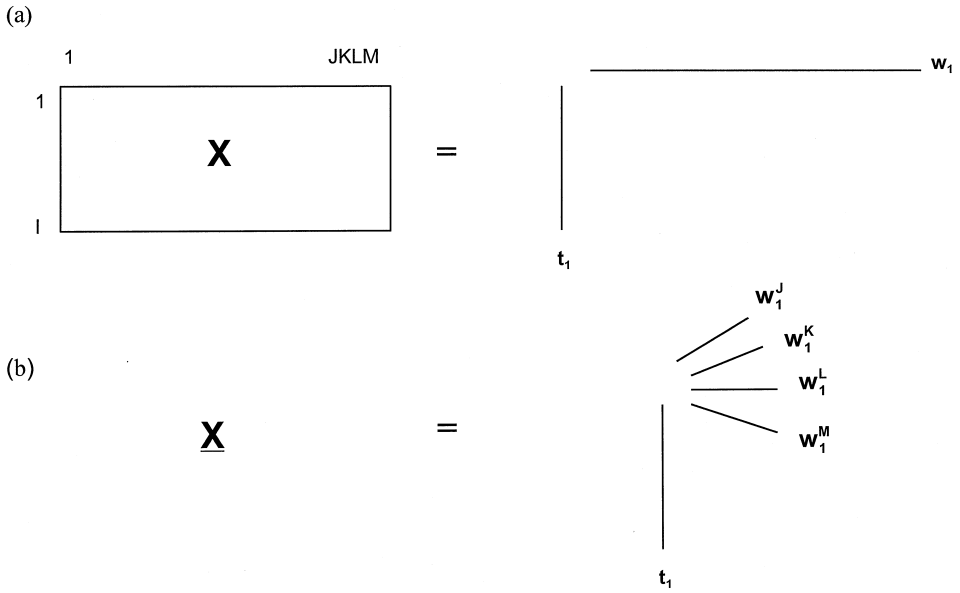


Figure 2. Graphical representation of the different regression methods: (a) bilinear PLS and (b) quintilinear PLS.

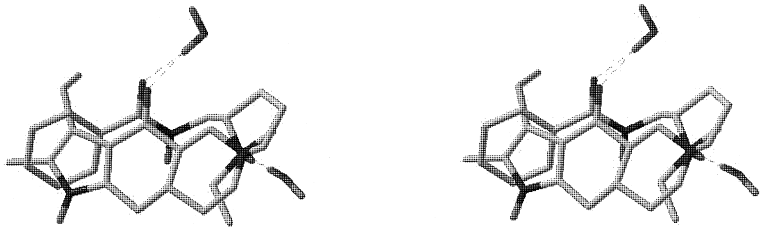


Figure 3. Stereo representation of the superposition of the best matching conformations of **59** (conformation 6) and **60** (conformation 2), as identified by APOLLO. For clarity purposes, alkyl hydrogens have been omitted. The water molecules mimic amino acid residues of the receptor, capable of forming hydrogen bonds with the ligands.

Figure 4. Stereo representation of the final alignment of compounds **1**–**58**. For clarity purposes, alkyl hydrogens have been omitted.

$$R^2 = \left[1 - \left(\sum_{i=1}^{I}(y_i - \hat{y}_i)^2 / \sum_{i=1}^{I}(y_i - \overline{y}_i)^2\right)\right] \times 100. \quad (4)$$

In addition to the traditional 'leave-one-out' (LOO) cross-validation, 'leave-three-out' (L3O) and 'leave-five-out' (L5O) cross-validations were also performed, where in each experiment the objects were left out randomly, but only once. The results were reported as the average Q^2 of 20 cross-validation experiments [21,22].

Partial PLS coefficients

The results from 3D QSAR studies are often presented as comprehensive iso-contour plots of the partial regression coefficients $\mathbf{b}_{PLS}(b_1, \ldots, b_R;\ R = JKLM)$ [2,23] in Equation 5, where x_r is the rth grid point, i.e. variable, and b_r is the corresponding coefficient.

$$\hat{y} = b_1x_1 + \cdots + b_rx_r + \cdots + b_Rx_R. \quad (5)$$

Basically, the coefficients $\mathbf{b}_{PLS}$ are needed for the predictions of the biological activity $\hat{y}$ of new molecules, but since the sizes and signs of the coefficients reveal the relative influence of each grid point on **y**, they are also suitable for the interpretation. That is, an external compound, not included in the training set, with a substituent protruding into a region with positive b_r's will produce a positive (repulsive) field x_r in this region and, consequently, have positive influence on $\hat{y}$. If the region has negative b_r's, however, the opposite is valid.

Results

Conformational analyses

The MC procedure identified 32 and 3 minimum-energy conformations for **59** and **60**, respectively. The two lowest energy conformations of **60** were equal in energy, conformation 1 being identical to the minimised X-ray crystal structure, while in conformation 2 the ethyl side chain points in the opposite direction.

Pharmacophore identification

APOLLO identified conformation 6 of **59** (ΔE = 5.80 $kJ\,mol^{-1}$) and conformation 2 (ΔE = 0.00 $kJ\,mol^{-1}$) of **60** as best matches with respect to the indicated fitting points, with an RMS deviation of 0.26 Å (Figure 3). In conformation 6 of **59**, the (1-ethyl-2-pyrrolidinyl)methyl side chain adopts a half-folded conformation.

Substituent geometries and ligand alignment

The orientation of the 6-OMe group in the lowest energy conformations depended on the 5-substituent. When R_5 was H, OMe, OH or F, the 6-OMe group was oriented coplanar with respect to the aromatic ring. In all other cases, i.e. R_5 being alkyl, Cl, Br or

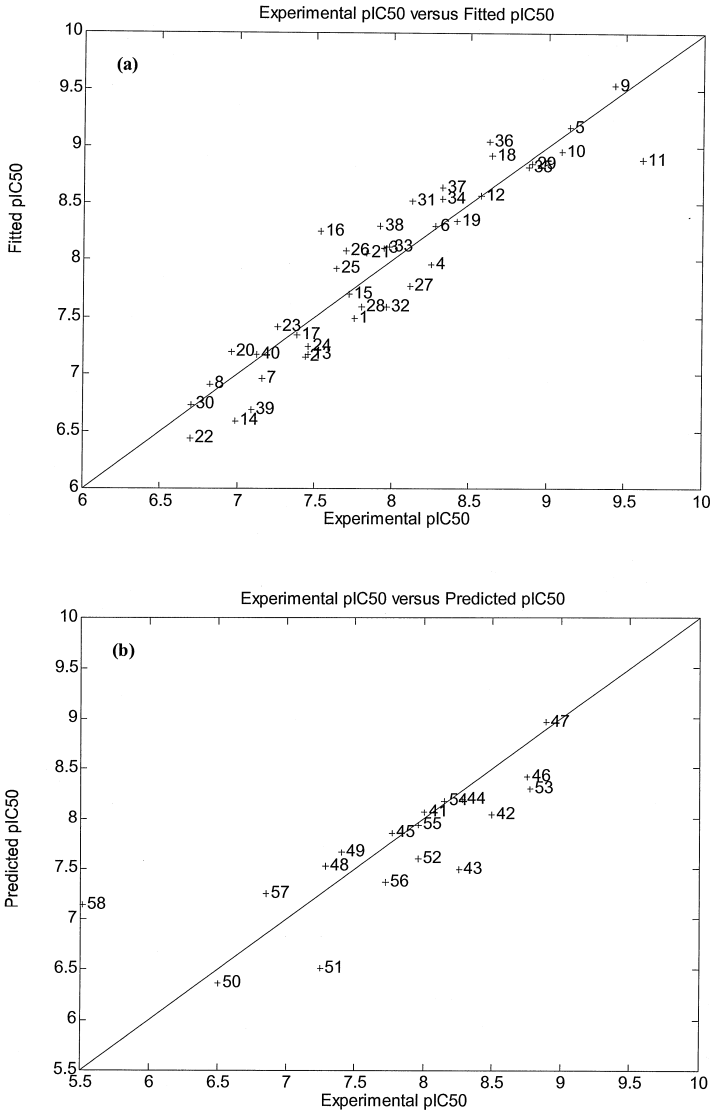


Figure 5. The performance of Model 2, showing (a) experimental pIC_{50} values versus fitted pIC_{50} values and (b) experimental pIC_{50} values versus predicted pIC_{50} values, after six N-PLS components.

NO_2, the 6-OMe group adopted a conformation perpendicular to the plane of the aromatic ring. In both orientations an intramolecular hydrogen bond between the 6-OMe O atom and the amide H atom was formed. All alkyl substituents longer than methyl at the 3- and 5-positions adopted all-trans conformations perpendicular to the plane of the aromatic ring, while OMe and OH groups at these positions had a coplanar orientation. All OH substituents at the 2-position adopted coplanar conformations, forming an intramolecular hydrogen bond between the 2-OH H atom and the carbonyl O atom. The final alignment of compounds **1**–**58** is shown in Figure 4.

The data set

The grid was large enough to enclose all the aligned molecules with at least 4 Å in all directions, where the x-, the y- and the z-directions were divided into 20 (J = 20), 22 (K = 22) and 20 (L = 20) parts with a step size of 1 Å, respectively. Three different probes (M = 3) were considered: the C3 probe, the OH2 probe and the CA+2 probe. The total data set consisted of 58 molecules and 26400 variables. As described above and

Table 2. Calibration[a] of the complete model, Model 1 ($40 \times 26\,400$)

#LV	$\mathbf{R}^2$ ($\underline{\mathbf{X}}$)	$\mathbf{R}^2$ (**y**)	$\mathbf{Q}^2$ (LOO)
1	5	47	29
2	11	59	44
3	14	69	52
4	17	75	56
5	20	79	56
6	**21**	**81**	**57**
7	22	85	56
8	24	87	62

[a] All values are expressed as percentages.

in Figure 1, we have defined five different directions: the molecular direction, the grid x-direction, the grid y-direction, the grid z-direction and finally the probe direction.

The 58 molecules were divided into a training set and a test set, selected with help from a principal component analysis (PCA) of **X** (58×26400; unfolded descriptor matrix in Figure 2a). Six main clusters were identified in the three first principal components describing 60% of the variation in **X**. The number of compounds selected from each cluster depended on the size of the cluster. This resulted in 18 compounds which were selected for the test set (Table 1, compounds **41**–**58**) and the remaining 40 compounds comprised the training set (Table 1, compounds **1**–**40**).

Multilinear PLS analysis

Different data preprocessing methods have been discussed in the literature [5], but the only data preprocessing applied on our data set was column mean-centering in the molecular direction [8]. Thus, the multiway matrix $\underline{\mathbf{X}}$ was unfolded, column mean-centred and, subsequently, 'back-folded' before the regression analysis was performed.

The number of significant N-PLS components was estimated by leave-one-out cross-validation of the complete model, Model 1, and we found six N-PLS components (Table 2; Q^2 = 57%) to be optimal. Important, in order not to lose information during the variable reduction step, we performed the variable reduction starting from a model one component more complex than optimal. Accordingly, the absolute sum of the weights from the first seven N-PLS components was calculated, each direction apart. A position in a direction was considered significant and selected only if it exceeded a lower cut-off value. An arbitrary cut-off value of 0.1 generated a reduced data set with 2940 number of variables, called Model 2. Stated differently, only variables with high weights from Model 1 were selected and included in Model 2. The probe mode was left intact; hence, variables from all three probes were included in the reduced data set. (One may argue why we did not perform the variable selection from a model with 10 N-PLS components having a Q^2 of 62%. The rationale for this was that the absolute sum of the weights barely changed going from seven to 10 N-PLS components and we did not want to risk influencing the selection with insignificant variation.)

Table 3 summarises the calibration and validation results, from the first 10 N-PLS components of Model 2. The six first N-PLS components used 21% of the variation in $\underline{\mathbf{X}}$ to explain 87% of the variation in **y**, and the Q^2's from the three cross-validation experiments LOO, L3O and L5O were calculated to be 63%, 62% and 62%, respectively. The calculated pIC_{50} values and the predicted pIC_{50} values after the sixth N-PLS component are tabulated in Table 1, and plotted against the experimental pIC_{50} values in Figures 5a and 5b for the training set and the test set, respectively.

Each subsequent component focuses on different variables, i.e., different regions in the grid which can be determined with help from Figure 6. The weights in Figures 6a and 6b correspond to positions in the x- and y-directions in Figure 7. For the first N-PLS component, the weights in Figures 6a and 6b are high in positions 7–13 and 5–9, respectively, corresponding to the substituents on the benzamide 3-position (Figure 7, area 1). Analogously, N-PLS component two focuses on substituents on the benzamide 5-position (area 2), N-PLS component three on large substituents on the benzamide 3-position (area 3), N-PLS component four on the out-of-plane 6-OMe (area 4), N-PLS component five on the amide part and substituents on the benzamide 2-position (area 5), and N-PLS component six on large substituents on the benzamide 5-position (area 6). All components account for variation from all three probes except for the fifth, which accounts for variation only from the CA+2 probe (Figure 6d). (In multilinear PLS, consecutive components are not orthogonal, which implies that some overlap of information may exist between different components. However, it has been shown [8] in multilinear PLS that each N-PLS component accounts for smaller easy-to-interpret regions, while in bilinear PLS each PLS component accounts for more variation but not that well defined regions.)

Table 3. Calibration[a] and validation[b] of Model 2 (40 × 2940)

#LV	$\mathbf{R}^2$ ($\underline{\mathbf{X}}$)	$\mathbf{R}^2$ (**y**)	$\mathbf{Q}^2$ (LOO)	$\mathbf{Q}^2$ (L3O)[c]	$\mathbf{Q}^2$ (L5O)[c]	$\mathbf{Q}^2$ (Pred)[d]
1	5	44	23	20	19	36
2	8	64	41	39	38	45
3	13	73	48	47	46	64
4	15	79	57	54	53	56
5	18	81	59	59	58	62
6	**21**	**87**	**63**	**62**	**62**	**62**
7	25	88	64	63	63	57
8	26	90	64	64	63	63
9	30	91	66	64	64	62
10	30	92	68	63	63	59

[a] All values are expressed as percentages.
[b] LOO stands for leave-one-out, L3O for leave-three-out and L5O for leave-five-out.
[c] Average from 20 cross-validation experiments.
[d] Predictions of the external test set (18 × 2940).

The partial PLS coefficients $\mathbf{b}_{PLS}$ of Equation 5 are presented as stereo iso-contour plots in Figures 8a–c for the three probes after six N-PLS components.

Discussion

The key issue in 3D QSAR modeling is to find a simple and straightforward model with high predictive ability. Traditionally, bilinear PLS is utilised as the regression method but recently it has been shown that multilinear PLS [8–10] is more stable, simpler and offers improved predictability.

An important step in 3D QSAR analyses is the alignment of the ligands under investigation, i.e. the relative positioning of the ligands in the fixed lattice, prior to the generation of the 3D descriptors. Even when the ligands possess a large degree of conformational flexibility, a single conformation has to be selected for each ligand. In order to be able to extrapolate the results of a 3D QSAR analysis in terms of receptor residues surrounding the ligands, we thought it essential to align the ligands in their pharmacologically relevant, i.e., receptor binding conformations. Therefore, we employed the active analogue approach to determine the presumed pharmacologically active conformations of the ligands [24]. This approach is based on the assumption that ligands binding to the same binding site share the same pharmacophoric pattern, i.e. the three-dimensional arrangement of structural features essential for recognition by the receptor. The pharmacophoric pattern of a flexible ligand can be determined by comparing it with that of a rigid analogue (template), in which the activity is retained. We chose (–)-piquindone as the template molecule. Although belonging to different chemical classes, (–)-piquindone and *N*-[(1-ethyl-2-pyrrolidinyl)methyl]-derived benzamides share several structural and pharmacological characteristics, a prerequisite when employing the active analogue approach. Thus, both classes of compounds contain an aromatic ring capable of π–π stacking with receptor residues, a carbonyl functionality, and a basic nitrogen atom at a certain distance from the aromatic ring, capable of forming hydrogen bonds with receptor residues. In addition, the binding to dopamine D_2 receptors is highly stereoselective and sodium-dependent for both classes, suggesting that they may share the same binding site and binding mode. This makes (–)-piquindone a suitable template for *N*-[(1-ethyl-2-pyrrolidinyl)methyl]-derived benzamides, as has been shown in the past [13,25,26].

A thorough conformational analysis was performed on *N*-[(1-ethyl-2-pyrrolidinyl)methyl]benzamide (**59**), which constitutes the basic skeleton of the series. Although all compounds contain a 6-OMe substituent, this functionality was omitted from the basic skeleton, since its orientation was dependent on the adjacent substituent and thus not identical for all compounds. Conformation 6 (ΔE = 5.80 kJ mol^{-1}) of **59** was identified by APOLLO as fitting best on conformation 2 of (–)-piquindone with respect to the defined fitting points (Figure 3). In this conformation of **59**, the *N*-[(1-ethyl-2-pyrrolidinyl)methyl] side chain adopts a half-folded conformation. This finding is consistent with previous reports [13,25], although

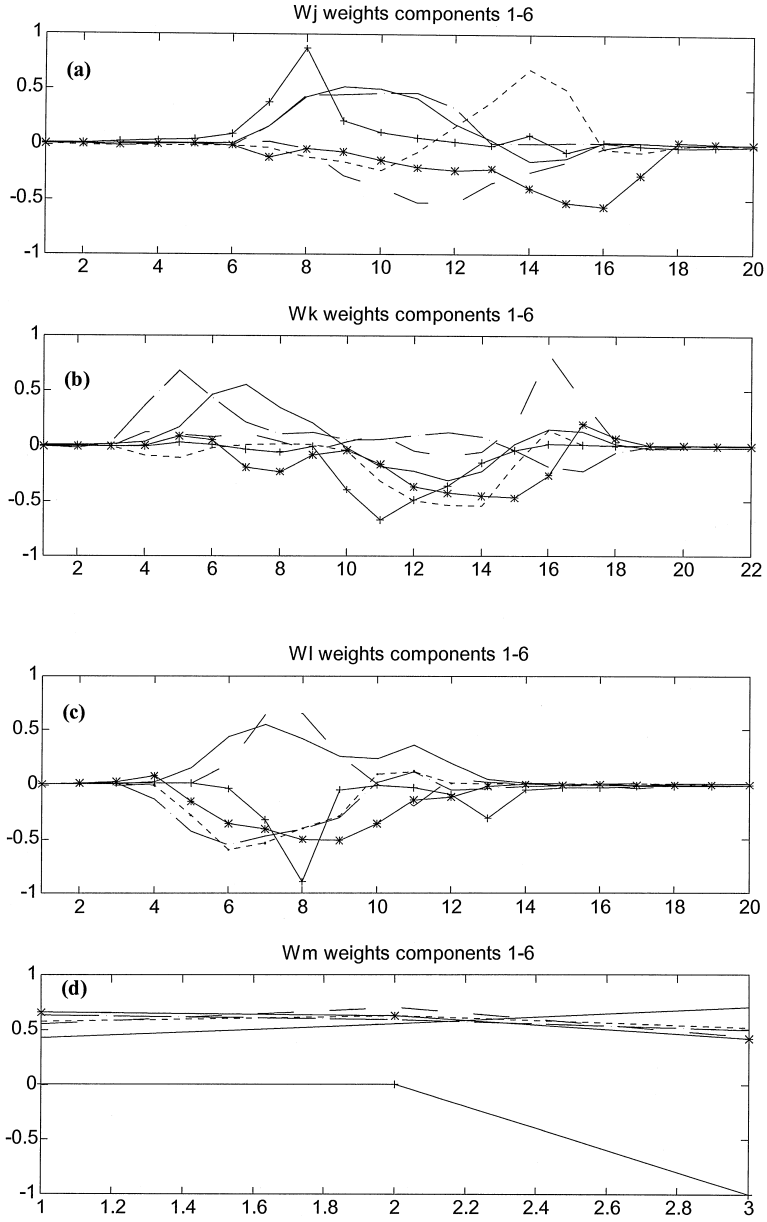


Figure 6. The weights from the first six components of the complete model from (a) the x-direction, (b) the y-direction, (c) the z-direction and (d) the probe direction. N-PLS(1) is represented by a solid line, N-PLS(2) a dotted line, N-PLS(3) a dashed-dotted line, N-PLS(4) a dashed line, N-PLS(5) a + solid line and N-PLS(6) a ∗ solid line.

the conformation we identified is not exactly identical to previously reported conformations of the side chain. This is probably the result of differences in the fitting procedures.

It is clear from Figures 6a–d that positions corresponding to grid points, predominantly in the periphery of the grid, have low weights, and, by omitting these variables, a reduced model (Model 2) with 2940 number of variables was obtained. Model 2 described slightly more of the variance in **y** (R^2 = 87%) as compared to Model 1 (R^2 = 81%), and the cross-validated Q^2 (LOO) was slightly increased from 57% to 63% after six N-PLS components in Models 1 and 2, respectively. The stability of Model 2 was confirmed since the cross-validated Q^2 (62%) did not change, i.e. it did not decrease when larger groups of com-

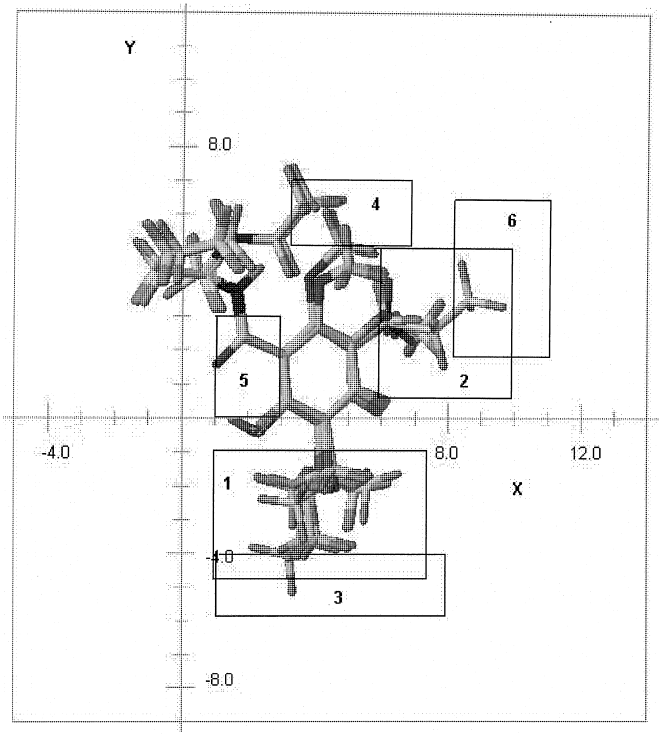


Figure 7. The 40 aligned training molecules, enclosed in the grid, viewed in the x- and y-directions.

pounds were left out each time (Table 3). Important, additional components did not significantly increase (or decrease) the cross-validated Q^2. Therefore, in order to keep the number of components at a minimum and avoid insignificant variation, we used six N-PLS components for Model 2.

The role of the cross-validated Q^2 is to act as a descriptor for the predicted Q^2, which it often does too optimistically [8]. In this model, however, the cross-validated Q^2 (62%) is in good agreement with the predicted Q^2 (62%). First now, knowing that the model is stable and possesses high predictability we may interpret the iso-contour plots in Figures 8a–c.

From this model we can conclude that the conformation of the 6-OMe group has great influence on the pIC_{50} value. In a planar conformation, the 6-OMe group protrudes in a region (I) with negative (dark grey) C3 PLS coefficients (Figure 8a) and, consequently, grid points within this region will have repulsive (positive) interactions with the 6-OMe group. Thus, the 6-OMe group in a planar conformation affects the pIC_{50} value negatively. For compounds with the 6-OMe group in a perpendicular conformation, the 6-OMe group protrudes in a positive (light grey) C3 PLS coefficient region (II), which favours a high pIC_{50} value (i.e. high affinity). At the same time, hydrogen bonding (Figure 8b) with the 6-OMe oxygen becomes favourable in region III, which promotes a high pIC_{50} value. From the CA+2 fields (Figure 8c) it is also clear that the substituents in positions 5 and 6 must possess negative electrostatic potential in order to promote a high pIC_{50} value. These are the most obvious conclusions that can be drawn from Figures 8a–c.

The substituents at the 2- and 3-positions have less influence on the pIC_{50} value as compared to substituents at the 5- and 6-positions [27,28]. However, a substituent with positive electrostatic potential in the 3-position is favourable. The size of the substituent is of minor importance. Also, important hydrogen bonding sites are, obviously, not present in the vicinity of positions 2 and 3. This is consistent with results reported by Norinder et al. [12].

The importance of the basic nitrogen on the pyrrolidinyl moiety has already been established elsewhere [26], but is not recognised in our model since the variance in the GRID descriptors in this region is very low.

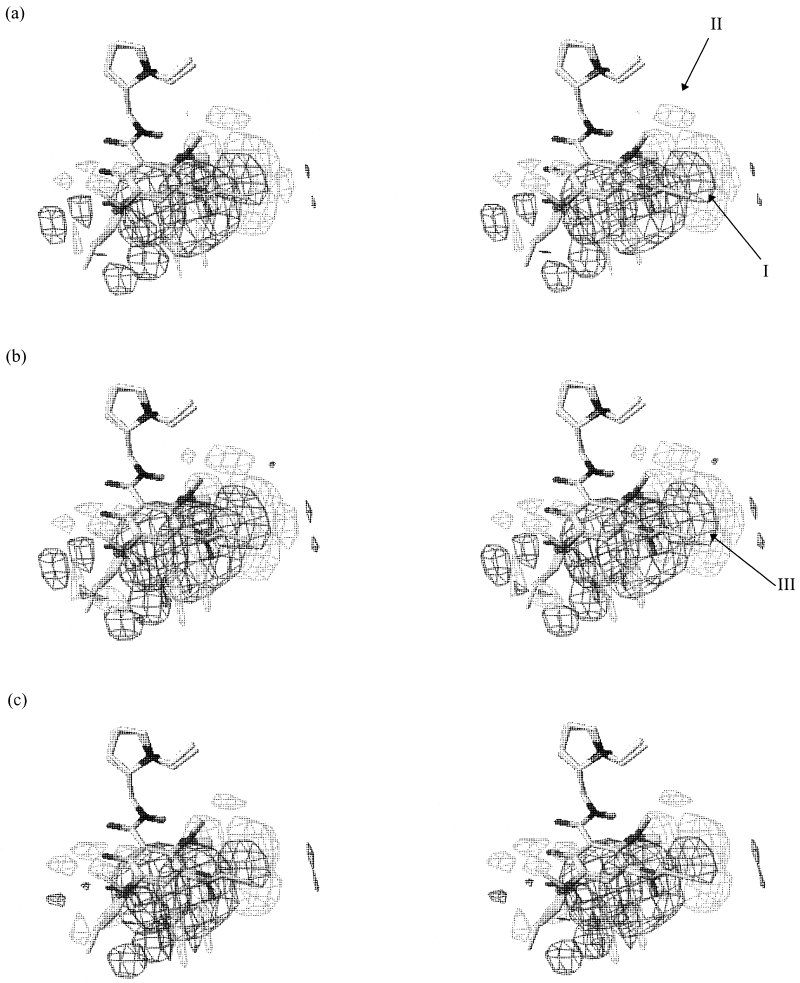


Figure 8. The $\mathbf{b}_{PLS}$ coefficients from Model 2 after six N-PLS components, showing (a) the C3 probe, (b) the OH2 probe and (c) the CA + 2 probe on the |0.001| level. Regions with negative and positive $\mathbf{b}_{\mathrm{PLS}}$ coefficients are pictured in dark and light grey, respectively.

Conclusions

We conclude that 3D QSAR modelling with multilinear PLS works very well and definitely is an alternative to the traditional bilinear PLS method.

We have demonstrated the importance of a proper strategy for the conformational analyses and the alignment procedure in order to succeed in 3D QSAR modelling. We found that the conformations of the ligands, to a great extent, explained the difference in pIC_{50} values among the ligands. Further, interpretations of our model are in line with what others have reported and confirm the reliability of our model [11,12].

We used multilinear PLS to reduce the number of variables by omitting variables with low weights. That is, only variables with variation correlated with the pIC_{50} values were considered in the final model (Model 2). The cross-validated Q^2 for Model 2 was 62%, independent of how many groups were left out each time. This result confirms that our data set is homogeneous and that our model is stable. Normally [10], leave-one-out cross-validation produces higher Q^2, as compared to when larger groups of compounds are left out each time. The cross-validated Q^2 (62%) from this model is a perfect estimation of the predicted Q^2, which was also found to be 62%. This is most certainly an effect originating from the multilinear PLS [9] method, which has been discussed by Nilsson et al. [8] and Smilde [23].

References

1. Cramer III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
2. SYBYL – Molecular Modeling Software, v. 6.3, SYBYL inclusive the CoMFA module, Tripos, St. Louis, MO, U.S.A., 1996.
3. APOLLO – Automated PharmacOphore Location through Ligand Overlap; the APOLLO program is available upon request from Konrad F. Koehler (koehler@irbm.it), 1996.
4. Jones, G., Willet, P. and Glen, R.C., J. Comput.-Aided Mol. Design, 9 (1995) 532.
5. Clementi, S., GOLPE, v. 3.0, Perugia, Italy, 1995.
6. Baroni, M., Costantino, G., Cruciani, G., Riganelli, D., Valigi, R. and Clementi, S., Quant. Struct.-Act. Relatsh., 12 (1993) 9.
7. Norinder, U., J. Chemometrics, 10 (1996) 95.
8. Nilsson, J., de Jong, S. and Smilde, A.K., J. Chemometrics, 11 (1997) 511.
9. Bro, R., J. Chemometrics, 10 (1996) 47.
10. Nilsson, J., Wikström, H., Smilde, A.K., Glase, S., Pugsley, T.A., Cruciani, G., Pastor, M. and Clementi, S., J. Med. Chem., 40 (1997) 833.
11. Norinder, U., J. Comput.-Aided Mol. Design, 7 (1993) 671.
12. Norinder, U. and Högberg, T., Acta Pharm. Nord., 4 (1992) 73.
13. Högberg, T., Drugs Future, 16 (1991) 333.
14. Jansen, J.M., Copinga, S., Gruppen, G., Molinari, E.J., Dubocovich, M.L. and Grol, C.J., Bioorg. Med. Chem., 4 (1996) 1321.
15. Mohamadi, F., Richards, N.G.J., Guida, W.C., Liskamp, R., Lipton, M., Caufield, C., Chang, G., Hendrickson, T. and Still, W.C., J. Comput. Chem., 11 (1990) 440.
16. Olson, G.L., Cheung, H.-C., Morgan, K.D., Blount, J.F., Todaro, L., Berger, L., Davidson, A.B. and Boff, B., J. Med. Chem., 24 (1981) 1026.
17. Goodford, P.J., GRID, University of Oxford, Oxford, U.K., 1995.
18. Geladi, P. and Kowalski, B.R., Anal. Chem. Acta, 185 (1986) 1.
19. Carrol, J. and Chang, J.J., Psycometrika, 35 (1970) 283.
20. Harschman, R.A., UCLA Working Papers in Phonetics, 16 (1970) 1.
21. Cruciani, G., Baroni, M., Costantino, G., Riganelli, D. and Skagerberg, B., J. Chemometrics, 6 (1992) 335.
22. Baroni, M., Clementi, S., Cruciani, G., Costantino, G., Riganelli, D. and Oberrauch, E., J. Chemometrics, 6 (1992) 347.
23. Smilde, A.K., J. Chemometrics, 11 (1997) 367.
24. Jansen, J.M., Ph.D. Thesis, Three dimensions in drug design. Probing the melatonin receptor, University of Groningen, The Netherlands, 1995.
25. Högberg, T., Rämsby, S., Ögren, S.O. and Norinder, U., Acta Pharm. Suecia, 24 (1987) 289.
26. Rognan, D., Sokoloff, P., Mann, A., Martres, M.P., Schwartz, J.C., Costentin, J. and Wermuth, C.G., Eur. J. Pharmacol, 189 (1990) 59.
27. De Paulis, T., Tayar, N.A., Carrupt, P.-A., Testa, B. and Van de Waterbeemd, H., Helv. Chim. Acta, 74 (1991) 241.
28. De Paulis, T., Kumar, Y., Johansson, L., Rämsby, S., Hall, H., Sallemark, M., Angeby Moller, K. and Ögren, S.O., J. Med. Chem., 29 (1986) 61.