PLS-Based Quantitative Structure–Activity Relationship for Substituted Benzamides of Clebopride Type. Application of Experimental Design in Drug Design

Ulf Norinder[a] and Thomas Högberg[b]

Astra Arcus AB, S-151 85 Södertälje, Sweden
[a]Karo Bio AB, P.O. Box 4032, S-141 04 Huddinge, Sweden and [b]Astra Draco, Preclinical Research and Development, P.O. Box 34, S-221 00 Lund, Sweden

Norinder, U. and Högberg, T., 1992. PLS-Based Quantitative Structure–Activity Relationship for Substituted Benzamides of Clebopride Type. Application of Experimental Design in Drug Design. – Acta Chem. Scand. 46: 363–366.

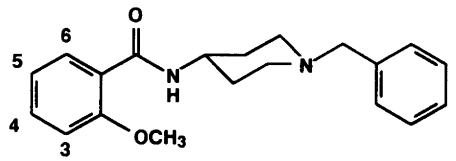
The advantageous approach of using an experimentally designed training set as the basis for establishing a quantitative structure–activity relationship with good predictive capability is described. The training set was selected from a fractional factorial design scheme based on a principal component description of physico-chemical parameters of aromatic substituents. The derived model successfully predicts the activities of additional substituted benzamides of 6-methoxy-*N*-(4-piperidyl)salicylamide type. The major influence on activity of the 3-substituent is demonstrated.

Quantitative structure–activity relationships (QSARs) have provided valuable aid in optimizing the properties of lead structures and rationalizing the structural features essential for activity of therapeutic agents.[1] A recent review article on this subject summarizes studies on CNS drugs.[2]

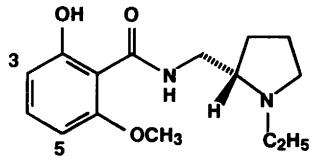
In order to develop a useful relationship with a broad predictive capability, the training set, i.e. the compounds included in the model, should span the available structural space as efficiently as possible. Today there are several methods available for the design of test series.[3] Experimental factorial design based selection methods provide such a tool for obtaining much information with a minimum number of observations (experiments).[4] They are widely used owing to their ease of performance and interpretation.

In this paper we have used a series of potential antipsychotic benzamides **1**, with affinity for the dopamine D_2 receptor, developed in our laboratories[5] to exemplify the approach whereby the training set compounds were selected based on an experimental design scheme followed by a QSAR analysis using the PLS method.[6] Previously, similar approaches have been applied to chlorinated aliphatic hydrocarbons,[7] halogenated ethers[8] and peptides.[9]

The benzamides (**1**) examined in this paper have *N*-benzyl-4-piperidinyl side chains and they display different requirements on nitrogen and aromatic substituents as well as side chain conformations compared with benzamides having *N*-ethyl-2-pyrrolidinylmethyl side chains (**2**).[10–12] The latter type of benzamide has been subject to extensive QSAR studies[12–15] and the present piperidinyl benzamides **1** have been investigated in a classical multiple regression QSAR involving substituents in the 3-, 4-, 5- and 6-positions.[5]



1



2

Fig. 1. Numbering convention for the benzamide series **1** and **2**.

Method of calculation

Experimental design. In a two-level factorial or fractional (reduced) factorial design scheme each variable is assigned a high level (designated +) and a low level (−). The number of experiments to be performed is 2^{N-r}, where N is the number of variables and r a reduction factor chosen so that $2^{N-r} > N$. It is important to include as many relevant physicochemical variables as possible when describing the substituents. However, to use the original variables in a design, even a highly reduced one, would involve too many compounds and be impractical. One method of circumventing this problem without losing too much information

and, at the same time, lowering the dimensionality, is to use a principal component analysis (PCA) of multiproperty characterization.[16] The number of significant components is determined by cross-validation.[17] The resulting principal components (latent variables) can be considered as principal properties (PP) of the original multiproperty description.[18] The signs of the PPs can then be used as a guide in a factorial or fractional factorial design.[16]

The benzamides **1** were varied in the 3-, 5- and 6-positions (see Fig. 1 for the numbering convention), since only three 4-substituted derivatives were available. The 3- and 5-substituents (H, Cl, Br, OMe, Et, Pr) were characterized by nine physicochemical variables (see Table 1 for a description) and subjected to a PCA (see Table 2 for the design levels). A PP-based two-level fractional factorial design consisting of eight compounds (arbitrary choice of a small number; $N = 5$, $r = 2$) was then constructed (see Table 3 for the design matrix). Position 3 was associated with columns 1 (a) and 2 (b) of the design protocol, position 5 with columns 3 (c) and 4 (abc) while the 6-substituents (H, OH) were treated with an indicator variable (H = −, OH = +) in column 5 (ab).

QSAR analysis. The relationship between the affinity for the dopamine D_2 receptor *in vitro*, measured as inhibition of [^{3}H]spiperone binding,[19] and chemical structure was analyzed using PLS.[6] Cross-validation was used to determine the number of significant components.[17] The descriptor matrix consisted of nine physicochemical variables (Table 1) for positions 3 and 5, respectively, the sum of positions 3 and 5 as well as the squared values and an indicator variable for position 6 (H/OH). The total number of used descriptors was 55. The resulting PLS loadings were thereafter transformed back into regression coefficients for comparison purposes.[20]

The experimental and calculated affinities of **1–20** are listed in Table 4.

Table 1. A description of the physicochemical variables used in the PCA and PLS analysis.

Variable	Description
Hammett constant[21]	σ_m
Hammett constant	σ_p
Swain–Lupton[22]	*F*
Swain–Lupton	*R*
Aromatic fragment constant[21]	π
Molecular refractivity[21]	*MR*
Verloop's sterimol parameter[23]	*L*
Verloop's sterimol parameter	*B1*
Verloop's sterimol parameter	*B5*

Table 2. Principal components of the available substituents.

Subst.	PC1[a]	PC2[a]	FDL[b]
Br	1.07	2.26	+ +
Cl	1.59	1.61	
H	2.27	−2.71	+ −
Pr	−2.90	0.05	− +
OMe	−0.37	−0.55	− −
Et	−1.65	−0.66	

[a]First (PC1) and second (PC2) principal component from PCA. [b]Factorial design levels; high (+) and low (−) level.

Table 3. The fractional factorial design matrix and final substituents.

No.	Pos. 3		Pos. 5		Pos. 6	Subst.[a]		
	a	b	c	abc	ab	3	5	6
1	+	+	+	+	+	Cl	Cl	OH
2	−	+	+	−	−	(OMe)	H	H
3	+	−	+	−	−	H	H	H
4	−	−	+	+	+	Et	Cl	OH
5	+	+	−	−	+	Cl	Et	OH
6	−	+	−	+	−	(Br)	(Br)	H
7	+	−	−	+	−	H	Pr	H
8	−	−	−	−	+	Et	Et	OH

[a]Substituents in parentheses indicate a deviation from the protocol.

Table 4. Substituents and experimental and calculated affinities for [^{3}H]spiperone displacement from rat striatal tissue of compounds **1–20**.

	Position			Affinities[a]			
No.	3	5	6	Exp.[b]	Calc.[c]	Calc.[d]	Calc.[e]
1	Cl	Cl	OH	−0.73	−0.71	−0.75	−0.91
2	OMe	H	H	−0.79	−0.97	−0.82	−0.78
3	H	H	H	−2.92	−2.63	−2.91	−2.52
4	Et	Cl	OH	−1.00	−0.94	−1.05	−1.09
5	Cl	Et	OH	−0.97	−1.21	−1.04	−1.32
6	Br	Br	H	−1.30	−1.08	−1.24	−1.50
7	H	Pr	H	−2.84	−3.07	−2.85	−3.00
8	Et	Et	OH	−1.41	−1.34	−1.34	−1.44
9	OMe	Br	OH	0.44	−0.12	0.48	0.11
10	H	Br	H	−2.35	−2.96	−3.32	−2.68
11	H	Et	H	−2.70	−2.86	−2.77	−2.76
12	OMe	Br	H	−0.69	−0.69	−0.21	−0.57
13	H	H	OH	−1.91	−2.06	−2.20	−1.83
14	H	Cl	OH	−1.89	−1.86	−1.80	−1.73
15	H	Br	OH	−1.61	−1.89	−1.75	−1.76
16	H	Et	OH	−2.12	−2.28	−2.07	−2.08
17	Et	H	OH	−1.30	−1.13	−1.46	−1.18
18	Br	Br	OH	−0.88	−0.51	−0.54	−0.81
19	Br	Et	OH	−1.54	−1.00	−0.89	−1.20
20	Et	Br	OH	−1.68	−0.96	−1.00	−1.12

[a]The affinities are shown in pIC_{50} [nM] values. [b]See Ref. 5 for the experimental values. [c]The values refer to the model with compounds **1–8** included. [d]The values refer to the model with compounds **1–9** included. [e]The values refer to the model with compounds **1–20** included.

Table 5. The 20 most important variables from the PLS analysis of compounds **1–20**.

Variable	Position	Coefficient
I^a	6	0.280
R	3	−0.108
σ_p^2	3	0.108
R^2	3	0.106
R^2	3+5	0.104
R	3+5	−0.101
F	3	0.073
π^2	3+5	−0.067
σ_m	3	0.054
π^2	3	−0.050
F	3	0.049
$B5$	3	0.048
L	3	0.046
MR	5	−0.046
F^2	3	0.044
$B5^2$	3	0.043
MR^2	3+5	−0.043
$B1$	3+5	0.043
π^2	5	−0.042
L^2	3	0.040

[a]Indicator variable (H = 0, OH = 1).

Results and discussion

Experimental design. The first two components, both significant, were calculated. They described 79.8 % of the variation in the multiproperty data set with individual contributions of 44.7 and 35.1 %, respectively (see Table 2 for the scores). The PCA resulted in the selection of compounds **1–8** for the training set (see Table 3 for the final substituents). However, the following considerations were made. (i) Compounds **1**, **3–5**, **7** and **8** were initially chosen since they corresponded to correct choices according to the design protocol in Table 3. (ii) None of the 20 available compounds (Table 4) matched exactly the substituent choices of #2 and #6. We therefore, arbitrarily, included two compounds which introduced new substituents into the training set in positions 3 and 5. This is a decision one is often forced to make when a limited number of compounds are available and no further synthesis is planned, or when the design targets are not synthetically accessible.

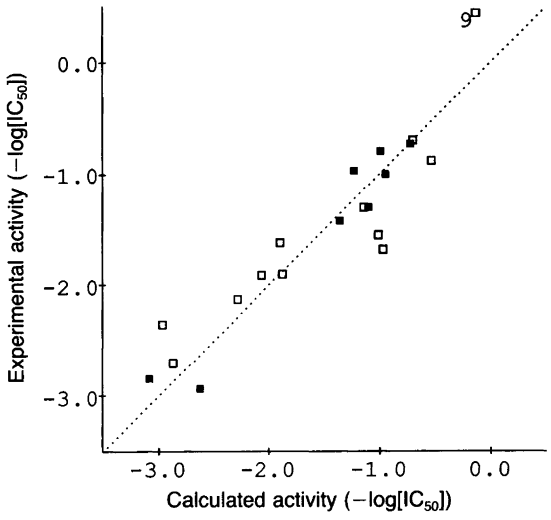


Fig. 2. Plot of calculated vs. experimental affinity of inhibition of [^{3}H]spiperone binding (■, compounds **1–8** as training set; □, predicted compounds).

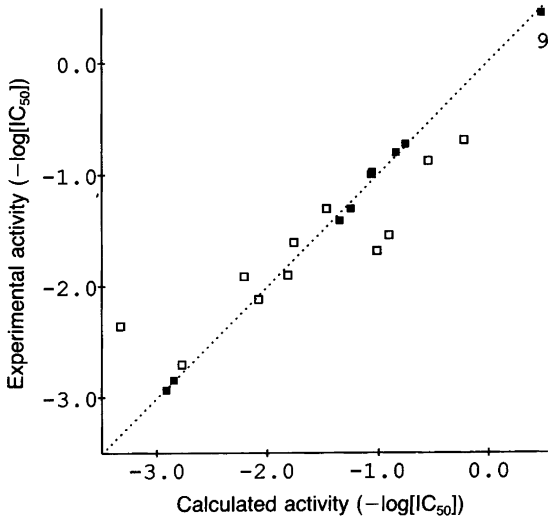


Fig. 3. Plot of calculated vs. experimental affinity of inhibition of [^{3}H]spiperone binding (■, compounds **1–9** as training set; □, predicted compounds).

QSAR analysis. The PLS analysis of the 55 descriptor variables of the training set gave two significant components with a 'predictive r^2' value of 0.95. The created model was then used to predict the affinities of the remaining 12 compounds, which resulted in a 'predictive r^2' value of 0.77. A plot of calculated vs. experimental affinities is depicted in Fig. 2.

The next choice of compound to be examined (synthesized, tested etc.), if high affinity were desired, would be compound **9**. Including **9** in the model gave an r^2 value of 0.93 and 0.59 for the new training set and the remaining 11 compounds, respectively (Fig. 3).

Incorporating all 20 available compounds in the model gave an r^2 value of 0.92 (Fig. 4).

Analysis of the regression coefficients of the last model revealed the dominating influence of the 3-substituent. The single most important variable was the indicator variable for position 6, which pointed out the favourable influence of a hydroxy-group in that position for obtaining high affinity with the [^{3}H]spiperone binding site.

The following 12 most important variables were all associated with position 3 or the corresponding variable of the sum of positions 3 and 5 (see Table 5). These results were also noted in a multiple regression QSAR analysis.[5] The study supports the distinctly different requirements on the aromatic substituents in the two benzamides series **1** and **2**.

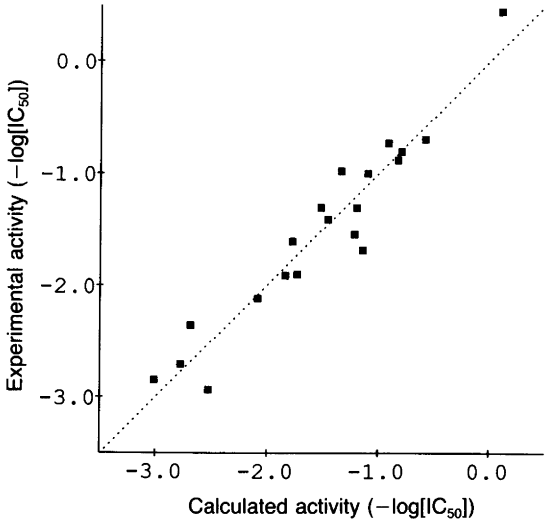


Fig. 4. Plot of calculated vs. experimental affinity of inhibition of [^{3}H]spiperone binding (■, compounds **1–20** as training set).

In the latter, the lipophilicity and steric bulk of the substituent *para* to the methoxy group (3-position in **2**) mainly determines the dopamine D_2 affinity[13,14] while the position *ortho* to the methoxy group is dominant in the former series.

Conclusions

The combined technique of a widely spanned selection of the training set compounds (initial compounds) using a PCA-based fractional factorial design followed by a PLS-supported QSAR analysis demonstrates the applicability and predictability of this approach in rational drug design.

References

1. Martin, Y. *Quantitative Drug Design,* Marcel Dekker, New York 1978.
2. Gupta, S. P. *Chem. Rev. 89* (1989) 1765.
3. Pleiss, M. A. and Unger, S. H. In: Sammes, P. G. and Taylor, J. B., Eds., *Comprehensive Medicinal Chemistry, Vol. 4,* Pergamon, Oxford 1990, pp. 561–587.
4. Box, G. E. D., Hunter, W. G. and Hunter, J. S. *Statistics for Experimenters,* Wiley, New York 1978.
5. de Paulis, T., Hall, H., Kumar, Y., Rämsby, S., Ögren, S. O. and Högberg, T. *Eur. J. Med. Chem. 25* (1990) 507.
6. Wold, S., Albano, C., Dunn, W. J., III, Edlund, U., Esbensen, K., Geladi, P., Hellberg, S., Johansson, E., Lindberg, W. and Sjöström, M. In: Kowalski, B. R. Ed., *Chemometrics; Mathematics and Statistics in Chemistry,* NATO ASI Series C138, Reidel, Dordrecht 1984, pp. 17–96.
7. Dunn, W. J., III, Wold, S., Edlund, U., Hellberg, S. and Gasteiger, J. *Quant. Struct.-Act. Relat. 3* (1984) 131.
8. Hellberg, S., Wold, S., Dunn, W. J., III, Gasteiger, J. and Hutchings, M. G. *Quant. Struct.-Act. Relat. 4* (1985) 1.
9. Hellberg, S., Sjöström, M. and Wold, S. *Acta Chem. Scand., Ser. B 40* (1986) 135.
10. Högberg, T., Rämsby, S., de Paulis, T., Stensland, B., Csöregh, I. and Wägner, A. *Mol. Pharmacol. 30* (1986) 345.
11. Högberg, T., Norinder, U., Rämsby, S. and Stensland, B. *J. Pharm. Pharmacol. 39* (1987) 787.
12. (a) Högberg, T., Rämsby, S., Ögren, S. O. and Norinder, U. *Acta Pharm. Suec. 24* (1987) 289; (b) Högberg, T. *Drugs Fut. In press.*
13. de Paulis, T. and Hall, H. In: Kuchar, M., Ed., *QSAR in Design of Bioactive Compounds,* Prous, Barcelona 1984, 199–208.
14. Norinder, U. and Högberg, T. *Quant. Struct.-Act. Relat. 10* (1991) 1.
15. Högberg, T. and Norinder, U. In: Krogsgaard-Larsen, P. and Bundgaard, H., Eds., *A Textbook of Drug Design and Development,* Harwood, London. *In press.*
16. Skagerberg, S., Bonelli, D., Clementi, S., Cruciani, G. and Ebert, C. *Quant. Struct.-Act. Relat. 8* (1989) 32.
17. Wold, S. *Technometrics 20* (1979) 379.
18. Hellberg, S., Sjöström, M., Skagerberg, B. and Wold, S. *J. Med. Chem. 30* (1987) 1126.
19. See Ref. 5 for the experimental details.
20. Öhman, J. *Calibration and Background Correction with Partial Least Squares Regression in Liquid Chromatography,* Thesis, University of Umeå 1988.
21. Hansch, C. and Leo, A. J. *Substitution Constants for Correlation Analysis in Chemistry and Biology,* Wiley, New York 1979.
22. Swain, C. G. and Lupton, E. C. *J. Am. Chem. Soc. 90* (1968) 4328.
23. Verloop, A. and Tipker, J. In: Hadzi, D. and Jerman-Blazic, B., Eds., *Pharmaco Chemistry Library. QSAR in Drug Design and Toxicology,* Elsevier, Amsterdam 1987.

Received March 15, 1991.