Journal of Computer-Aided Molecular Design, 7 (1993) 671–682
ESCOM

J-CAMD 218

A PLS QSAR analysis using 3D generated aromatic descriptors of principal property type: Application to some dopamine D2 benzamide antagonists

Ulf Norinder

Karo Bio AB, P.O. Box 4032, S-141 04 Huddinge, Sweden

Received 23 November 1992
Accepted 30 April 1993

Key words: Principal component analysis; QSAR; QSPR; PLS; 3D structures; 3PPs

SUMMARY

A simple and computationally nonintensive technique based on principal component analysis of 3D fields to derive theoretical descriptors is presented. The descriptors are then applied to a quantitative structure–activity relationship study on some dopamine D2 antagonists of benzamide type.

INTRODUCTION

Quantitative structure–activity relationships (QSARs) have provided valuable aid to elucidate underlying mechanisms of action as well as enable the development of new therapeutic agents [1]. However, the proper representation of a structure in the relationship may, in some cases, prove to be a problem. The necessary descriptors, such as substituent parameters and other physicochemical variables, may not be available for certain substituents or parts of a structure. One way to circumvent this problem is to obtain molecular descriptors through theoretical calculations.

In this article the use of aromatic descriptors generated from a 3D representation of the substituents is described and applied to a QSAR analysis of some dopamine D2 antagonists of the benzamide type. These compounds have previously been successfully analysed based on a more conventional description by using tabulated physicochemical parameters [2].

A technique similar to the one presented in this article has been used to develop theoretical descriptors for the 20 coded amino acids [3] and for some nucleic acid bases [4]. Other approaches to derive 3D-based substituent descriptors have recently been presented by Van de Waterbeemd et al. [5] and by Hemken and Lehmann [6].

The purpose here is to show that descriptors generated in the manner described in detail below based on a 3D representation may find use as variables in quantitative structure–property relationships (QSPRs).

TABLE 1
AROMATIC SUBSTITUENTS AND 3PP DESCRIPTORS

Substituent	Nonbonded descriptors										Electronic descriptors				
	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5
1 Br	−71.18	−51.98	−1 49	−0 35	5.83	−3.16	5.88	0.38	−4.86	−12.27	0.54	5.02	−6.85	−4.19	1 34
2 CF_2CF_3	115.10	11.36	59 00	22 19	−20.74	14.13	−58.99	−11.35	22.69	−39 90	−135.23	36.20	12.21	−18.41	−5.07
3 CF_3	−20 23	8.38	40.65	19.53	24.87	−46.00	16 64	14.00	16.93	−1.23	−142.92	4.22	−2.65	−31.37	−2.37
4 CH_2Br	26.78	−37.48	37 95	8 03	−28.95	1 02	35.87	−30.32	4.88	21 27	20.40	52.47	22 31	−1 63	6.12
5 CH_2Cl	13.30	−39.54	37 75	9 36	−20 37	−4.58	38.17	−26.11	2 07	19.78	12.08	79 90	37.76	4.23	6.89
6 CH_2I	47 29	−39.20	40 25	5 68	−35.71	11.85	27.08	−32.44	2.96	12 80	39.21	3.79	−7 18	−2.29	6.12
7 CH_3	−60 15	−40.07	7.51	4 95	12.75	−20 61	11.00	6.88	−1.71	−16 68	64.74	7.44	−11 69	10.66	−0.44
8 Cl	−77.01	−56.44	−4 81	−2 51	5.78	5.85	0 33	0.23	−3.99	−8 83	−23.66	4.71	−6 28	−11.23	0 01
9 CN	−57.88	−40.96	−0.19	5 40	−0.25	−22.36	15.94	9 46	1.40	−26.01	−99.88	4.10	−2 70	−31.17	3.08
10 CH_2CH_3	35 64	−31.81	37.93	12 08	−22.89	9.82	19.62	−21.82	2.40	6 35	63.96	0.57	−11.63	5.79	2.56
11 F	−93.41	−73.84	−14.05	−7 85	8.24	48.26	−26.00	17.88	19.40	17.94	−55.51	2.92	−2 27	−24.99	0.41
12 H	−97.45	−79.50	−16 24	−9 13	9.61	64 11	−35.61	27.05	31.76	31.30	41 05	5 42	−10.89	5.53	−4 24
13 I	−56 49	−38.11	5 46	5.00	5.54	−22 98	17 05	3.02	−3.99	−21.22	50 74	4 81	−5.81	10.05	3 66
14 NH_2	−74.15	−49.74	−3.11	−4.00	13.72	4.36	−7.24	−2.14	−4.18	−7.75	109.21	−20.19	−32.95	23.87	−17.64
15 NO_2	−40.41	0.06	19 18	7 00	35.52	−42 80	−16 19	−13 08	−10 78	3.53	−258 35	3 45	−4.36	−56 92	5 14
16 OCF_3	−38.48	90.31	−52.47	−8.66	−45.21	10.25	11.37	0 61	3.05	−5.90	−69.81	12.45	−16.95	−15 74	1.30
17 OCH_3	−63.60	44.03	−26.86	−16 75	−10.92	26.25	−15 61	−21.56	−18 14	0.83	56 87	−36.27	54 55	12.03	1 54
18 OH	−87.53	−64.57	−11 86	−6 85	7.52	32.22	−18.15	7 53	7.25	6.56	16.88	−17.89	26 86	3 74	−1.02
19 SCF_3	−43.78	103.87	−78 35	−16 67	−52.67	2.96	−2.04	−22 76	23.61	13.92	1.08	81 65	−123 85	−0.58	3 12
20 SCH_3	−49.68	69 76	−47.32	−17 12	−25.69	1.09	−8 67	−34.59	−11.63	12.89	114 67	8.93	−17.26	22.81	−3.30
21 SH	−69.49	−38.20	−4 33	−1 76	8.80	−11.69	−1.13	−5 16	−10.58	−21.34	81.85	12.65	−20.74	15.12	−1 24
22 SO_2CF_3	125.46	6.20	62 97	5.65	1 88	19.43	−70 53	−22.10	8.13	−31.64	−235.63	159.31	72.15	50.96	−9.07
23 SO_2CH_3	74.02	15.87	73 41	24 33	8.51	7.77	−19 64	−25.43	5.56	1 95	−126.08	−95 36	−85.69	34.56	3 06
24 SO_2F	1.36	14 46	50.80	19.36	35.08	−35.27	8.33	−7.46	−3.80	22.26	−175.80	−32.48	−55.33	38.59	2.91
25 $C(CH_3)_3$	55.30	111.54	56.93	18.03	49.01	49.37	27.81	56.75	−8.21	12.15	65.58	1.04	−5.18	1.74	−0.94
26 $(CH_2)_2CH_3$	110.31	−28.20	−25.23	29.42	−56.66	−4.13	−20.12	24 02	−27.50	8.79	61.01	−0.95	−10 01	2.85	1 61
27 $(CH_2)_3CH_3$	147.85	−44 78	−68.19	−22 41	−37.37	−14.48	−8.86	55.18	−52.36	−8.54	58.05	0.51	−8 94	0.11	2 88
28 $CH_2C_6H_5$	139 02	−15.28	−10.55	−201 67	34.43	−17.66	8 93	2.04	28.58	−0.42	53.60	31.53	25.37	−61.91	20 56
29 CCH	−50.15	−28.78	−21.51	17.29	−15.86	−26 36	12.72	37.02	−0.09	−28.67	26.12	−6 81	−3.01	17.40	14.72
30 CH_2CCH	28 36	−61.33	34.56	−18 55	−19 46	−5.84	34 77	−8.43	−9.08	11.71	41 86	32.82	19 97	−27.06	58.06
31 CH_2CHCH_2	95.54	−51 98	11.77	11.46	−12.37	17 42	4.82	12.82	−40.07	32.25	53.96	10.97	−1.12	−11.76	17.30
32 $CHCH_2$	−67.07	26.75	−17.53	−8 73	10.15	−6 13	−18.11	−11.74	−23.75	−26 46	57.26	−0.58	2.20	2 55	−4.34
33 $COCH_3$	−26.40	77.95	−11.10	−7 97	16 86	−19.75	−18.28	−22 48	−20.24	21.77	−58 21	−67.96	99 77	6.02	7 11
34 $CONH_2$	−20.83	16 38	16.17	10 45	14.59	−39.98	−1.11	8 92	24 88	8 34	−75.82	−167 51	36.70	1.47	−8.48
35 $CONHCH_3$	44 25	34.82	−40 79	45 09	−40 64	−60 98	−24.47	39.32	66 90	24.19	−49 83	−168.87	21.77	8.15	−0.38
36 $CH(CH_2)_2$	−25.93	92.43	−0.08	−8.24	−4.82	23.24	32 71	9.97	3.18	−25.42	65.37	−4.13	0.94	4.60	1.93
37 $CH(CH_3)_2$	19.50	91.96	41 20	10.22	−3 16	49.14	56.84	27.06	17 29	−24.47	64.88	−0.50	−3.57	2.44	−0.59
38 $NHCH_3$	−56.83	54.56	−22.29	−11 30	−3.47	4.49	−8.28	−19.99	−23.88	−9 22	126.18	11 01	23.26	19.65	−4.26
39 $N(CH_3)_2$	−11.24	83.67	13.08	5 45	61 23	−18.90	−36.23	16.29	−32.37	34.75	143.64	17 03	9.06	10.83	−7.01
40 $CH_2CH_2C_6H_5$	180.29	−42.57	−168.20	74 56	87.29	20 65	29.37	−37.48	18.26	−9.35	15 95	8.27	3.21	−35.02	−100.89

METHOD OF CALCULATION

Substituent geometries

The substituents studied (see Table 1) were attached to the para position of benzoic acid. The compounds were built and initially minimized by using the MM2-85 equations and force field incorporated in the MODEL program [7]. Conformational analysis was performed on the substituents which could adopt several conformations. However, long alkyl chains were used in their all-trans conformations. All of the generated (and MM2-85 minimized) geometries were then subjected to further complete minimization by using the AM1 Hamiltonian implemented in the MOPAC program [8]. The PRECISE keyword was used in all the AM1/MOPAC calculations. The resulting lowest conformation of each substituent was selected for further use and the benzoic moiety was superimposed on benzoic acid itself (representing the H substituent). Some of the substituents, such as Et, Pr and benzyl, adopted a nonplanar conformation with respect to the benzoic moiety.

Descriptor generation

A 3D grid was spanned around the structures (1920 points with an internal separation of 1.5 Å) and two values related to nonbonded and charge–charge interactions were computed for each grid point. A methyl probe (charge = +1) was used with the equations of Del Re et al. [9] and Coulombic point charge interactions at the AM1 level, respectively. An upper cut-off value of 50 kcal/mol was used for the nonbonded interactions. A 'missing value' was assigned to a grid point in the charge–charge calculations if the distance between the closest atom and the probe was less than the sum of the van der Waals radii. The two sets (fields) of interactions were mean centered before any statistical operations were undertaken. The number of significant principal components was determined by cross-validation [10].

The molecular descriptors, i.e. variables describing the system (compounds) under investigation, were then generated for each substituent by a separate principal component analysis (PCA) on each set where the resulting score values are the descriptors.

PLS analysis

Three datasets were analysed: Hammett's sigma-para and sigma-meta for the substituents covered (Table 2) and the inhibition of [^{3}H]spiperone binding by some 70 dopamine D2 antagonists of the benzamide type (Table 3) [2]. The first two datasets were investigated by using only the charge (electronic) set of the 3D-based descriptors (3PPs) as well as the original grid representation. The third dataset was described with the 3PP variables (positions 3 and 5, see Scheme)

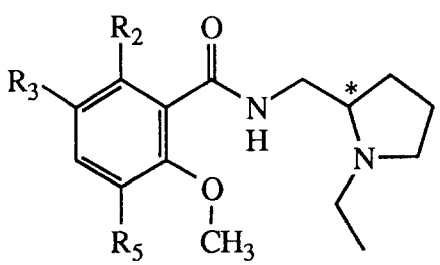

TABLE 2
TABULATED AND CALCULATED VALUES FOR HAMMETT'S SIGMA-PARA AND SIGMA-META

Substituent	Sigma-para values			Sigma-meta values		
	Tab[a]	Calculated		Tab[a]	Calculated	
		FG[b]	3PP[c]		FG[b]	3PP[c]
1 Br	0.230	0.228	0.113	0.390	0.366	0.195
2 CF_2F_3	0.530	0.531	0.574	0.470	0.533	0.495
3 CF_3	0.540	0.514	0.622	0.430	0.512	0.511
4 CH_2Br	0.140	0.094	0.002	0.120	0.112	0.125
5 CH_2Cl	0.120	0.086	0.017	0.110	0.106	0.143
6 CH_2I	0.110	0.062	−0.021	0.100	0.149	0.105
7 CH_3	−0.170	−0.147	−0.117	−0.070	−0.005	0.055
8 Cl	0.230	0.159	0.194	0.370	0.268	0.243
9 CN	0.660	0.539	0.470	0.560	0.493	0.408
10 CH_2CH_3	−0.250	−0.162	−0.111	−0.070	−0.002	0.053
11 F	0.060	0.084	0.296	0.340	0.330	0.298
12 H	0.000	−0.194	−0.043	0.000	−0.014	0.101
13 I	0.180	0.177	−0.061	0.350	0.325	0.092
14 NH_2	−0.660	−0.665	−0.288	−0.160	−0.207	−0.045
15 NO_2	0.780	0.896	1.067	0.710	0.696	0.783
16 OCF_3	0.380	0.345	0.391	0.380	0.426	0.369
17 OCH_3	−0.270	−0.196	−0.197	0.120	0.037	0.011
18 OH	−0.370	−0.227	−0.009	0.120	0.069	0.126
19 SCF_3	0.500	0.467	0.329	0.400	0.407	0.327
20 SCH_3	0.000	−0.013	−0.296	0.150	0.138	−0.053
21 SH	0.150	0.125	−0.164	0.250	0.184	0.026
22 SO_2CF_3	0.930	0.790	0.933	0.790	0.775	0.815
23 SO_2CH_3	0.720	0.700	0.796	0.600	0.646	0.684
24 SO_2F	0.910	0.883	0.942	0.800	0.835	0.790
25 $C(CH_3)_3$	−0.100	−0.170	−0.144	−0.100	−0.039	0.029
26 $(CH_2)_2CH_3$	−0.130	−0.129	−0.109	−0.070	−0.007	0.052
27 $(CH_2)_3CH_3$	−0.160	−0.085	−0.098	−0.080	−0.026	0.056
28 $CH_2C_6H_5$	−0.090	−0.157	−0.163	−0.080	−0.075	−0.051
29 CCH	0.230	0.158	0.071	0.210	0.209	0.186
30 CH_2CCH	ND[d]			0.070	0.052	0.114
31 CH_2CHCH_2	−0.190	−0.176	−0.068	0.110	0.102	0.062
32 $CHCH_2$	−0.020	−0.089	−0.133	0.050	0.056	0.038
33 $COCH_3$	0.500	0.451	0.181	0.380	0.364	0.264
34 $CONH_2$	0.360	0.331	0.308	0.280	0.290	0.341
35 $CONHCH_3$	0.360	0.422	0.265	0.350	0.367	0.315
36 $CH(CH_2)_2$	−0.210	−0.184	−0.142	−0.070	0.006	0.033
37 $CH(CH_3)_2$	−0.150	−0.147	−0.142	−0.070	−0.073	0.031
38 $NHCH_3$	−0.840	−0.594	−0.419	−0.300	−0.218	−0.134
39 $N(CH_3)_2$	−0.830	−0.713	−0.480	−0.150	−0.161	−0.185
40 $CH_2CH_2C_6H_5$	−0.120	−0.129	−0.304	−0.070	−0.071	−0.089

[a] Taken from Ref. 19.
[b] Full Grid (FG) representation of the substituents.
[c] 3PP representation of the substituents.
[d] Not determined.

TABLE 3
EXPERIMENTAL AND CALCULATED ACTIVITY VALUES FOR BENZAMIDES 1–70

No.[a]	Compound			Activity[b]		
	R_3	R_5	R_2	Exp[c]	Calculated	
					3PP[d]	prev[c]
1	Cl	Cl	OH	7.49	7.62	7.95
2	Br	OH	H	8.00	7.92	8.27
3	OMe	Cl	OH	7.15	7.05	7.01
4	Pr	Cl	OH	8.49	8.05	8.87
5*	NO_2	Br	OH	5.20	5.23	4.97
6	Br	Br	H	8.10	7.76	8.01
7	Me	Br	OH	8.26	8.43	8.09
8	Et	Br	H	7.96	8.32	8.23
9R	Br	OMe	H	6.84	6.95	6.96
10	I	OMe	H	9.17	9.07	9.06
11	Me	OMe	H	8.28	8.92	8.34
12	Pr	Me	OH	8.30	8.14	8.58
13	Cl	Pr	OH	6.96	6.52	7.10
14	Br	Et	OH	7.77	7.89	7.83
15	H	Et	OH	6.91	7.05	6.96
16	Et	Et	OH	8.75	8.46	8.05
17	I	OMe	OH	9.54	9.31	9.25
18	SMe	OMe	H	8.96	8.74	8.60
19	Et	OMe	H	8.89	9.06	8.68
20	Et	OMe	OH	8.89	9.30	8.87
21	Bu	OMe	H	8.57	8.68	8.63
22	Pr	H	H	7.17	7.63	8.18
23R	Cl	Cl	OH	5.53	6.09	6.46
24*	H	H	H	5.96	5.90	5.90
25	H	Ome	H	7.28	7.64	7.58
26	Cl	H	H	6.59	7.21	7.26
27	Cl	Cl	H	7.70	7.38	7.76
28	Cl	Br	H	8.25	7.45	7.75
29*	Br	H	H	7.37	6.74	6.78
30	Br	H	H	7.34	7.51	7.52
31	Br	OMe	H	8.92	8.49	8.45
32	Et	H	H	7.40	8.08	7.75
33	Et	Cl	H	8.38	8.25	8.24
34	H	H	OH	6.50	6.91	6.84
35	H	Cl	OH	7.19	7.08	7.34
36*	H	Br	OH	6.91	6.39	6.58
37	H	Br	OH	7.25	7.15	7.33
38	H	OMe	OH	8.06	7.89	7.77
39	F	H	OH	6.44	6.52	6.57
40	Cl	H	OH	7.41	7.45	7.46
41	Cl	Br	OH	7.24	7.70	7.94
42	Cl	Me	OH	7.96	7.72	7.67
43	Cl	Et	OH	7.92	7.59	7.57

TABLE 3 (continued)

No.[a]	Compound			Activity[b]		
	R_3	R_5	R_2	Exp[c]	Calculated	
					3PP[d]	prev[c]
44	Cl	OMe	OH	8.77	8.43	8.39
45*	Br	H	OH	7.57	6.98	6.97
46	Br	H	OH	8.08	7.75	7.71
47R	Br	H	OH	6.34	6.22	6.22
48	Br	F	OH	8.15	7.70	8.42
49	Br	Cl	OH	7.77	7.92	8.21
50*	Br	Br	OH	7.80	7.23	7.46
51	Br	Br	OH	7.59	8.00	8.20
52	Br	Me	OH	7.96	8.02	7.92
53	Br	OMe	OH	8.85	8.73	8.64
54R	Br	OMe	OH	7.10	7.20	7.15
55	Br	NH_2	OH	7.48	7.95	7.66
56	Br	NO_2	OH	6.73	7.36	6.90
57	I	H	OH	8.52	8.33	8.32
58	Me	H	OH	7.72	8.18	7.60
59	Me	Cl	OH	7.59	8.36	8.10
60	Me	Me	OH	8.11	8.45	7.81
61	Me	Pr	OH	6.85	7.26	7.25
62	Et	H	OH	8.54	8.32	7.94
63	Et	F	OH	8.82	8.27	8.65
64	Et	Cl	OH	9.04	8.49	8.44
65	Et	Br	OH	8.64	8.57	8.43
66	Pr	H	OH	8.30	7.87	8.37
67	OMe	H	OH	6.69	6.87	6.51
68	OMe	Br	OH	7.17	7.12	7.00
69R	OMe	Br	OH	5.21	5.59	5.51
70	NO_2	H	OH	5.52	5.75	5.23

[a] Stereochemistry of the side chain if not (*S*)-configuration: * = racemate, R = (*R*).
[b] Activity values in pIC_{50} molar units.
[c] Model from Ref. 2 based on physicochemical descriptors.
[d] Model based on the 3PP descriptors.

related to both nonbonded and electronic interactions only since the purpose of the article is to explore the possible applicability of 3PPs as variables in QSARs.

The two alternatives that existed for position 2 (H and OH) were treated with an indicator variable I_2 (H = 0, OH = 1) as well as the chirality of the side chain, I_s (S = 1, R = −1, racemate = 0).

Each of the dependent variables was related to the descriptor matrix by using the PLS (partial least-squares projections to latent structures) method [11]. The matrices were mean-centered and the 3PP-based ones autoscaled. The number of significant components was determined by cross-validation (four validation groups for the 3PP-based matrices and a leave-one-out procedure for

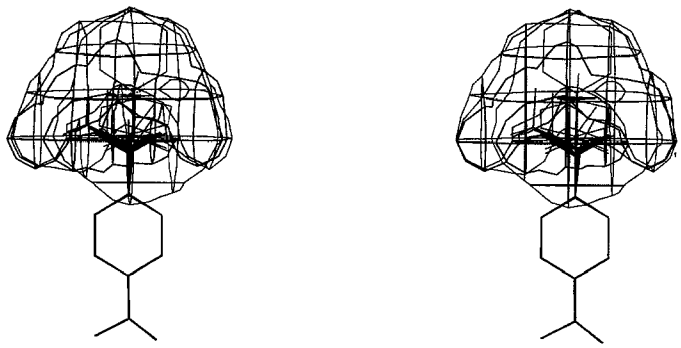

Fig. 1. Stereopicture of the nonbonded variance map at a contour level of 25. The 40 substituted benzoic acids are depicted with hydrogens omitted for clarity.

the grid-based matrices) [10]. After PLS analysis of the third dataset, the loadings were transformed back into PLS-type regression coefficients and these were further transformed into coefficients related to the points of the original 3D grid space of the benzoic acids comprising the database.

Contour Maps

The 3D contour maps were displayed by using CHEM-X [12]. The coefficients were scaled so that the largest absolute value was set to +100 or −100 depending upon the sign; the rest of the values were scaled accordingly.

RESULTS

3D-based descriptors (3PPs)

Ten and five components were extracted from the PCAs on the nonbonded and electronic interactions, respectively. They described 88.6% of the variance in the nonbonded matrix and 92.9% of the variance in the electronic matrix. The score values from the two PCAs are listed as the final descriptors in Table 1. The descriptors are available from the author on request. Contour maps of the variances of the 3D grids of both fields are depicted in Figs. 1 and 2.

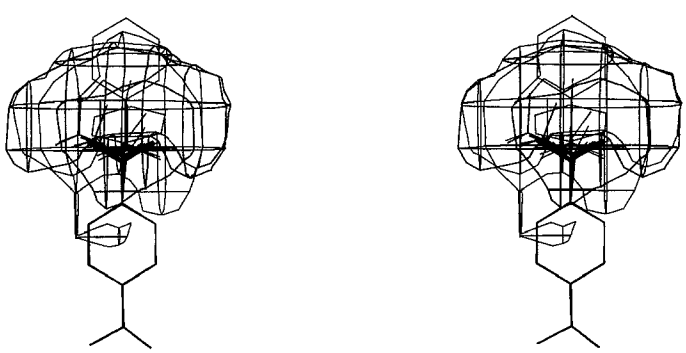

Fig. 2. Stereopicture of the charge–charge variance map at a contour level of 25. The 40 substituted benzoic acids are depicted with hydrogens omitted for clarity

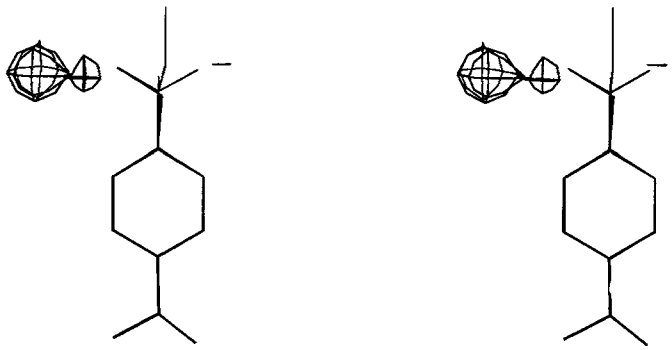

Fig. 3. Stereopicture of the positive nonbonded interactions at a contour level of 30 for position 5. The substituted benzoic acids related to the substituents occurring in position 5 (see scheme) are depicted with hydrogens omitted for clarity.

PLS analysis of Hammett's sigma-para and sigma-meta

The analysis resulted in seven and nine significant components for sigma-para and sigma-meta, respectively, based on the original grid representation of the electronic interactions. The components described 96.5% and 97.1% of the variance for sigma-para and sigma-meta, respectively. The corresponding analysis based on the 3PP electronic description gave two significant components for both datasets which explained 80.5% and 83.7% of the variance for sigma-para and sigma-meta, respectively. The tabulated and calculated (fitted) Hammett values are listed in Table 2.

PLS analysis of the benzamides

The analysis resulted in four significant components which described 85.2% of the variance in the [^{3}H]spiperone binding data. The experimental and calculated affinities are listed in Table 3. Contour maps of the PLS regression coefficients transformed into the original 3D grid space are depicted in Figs. 3–5.

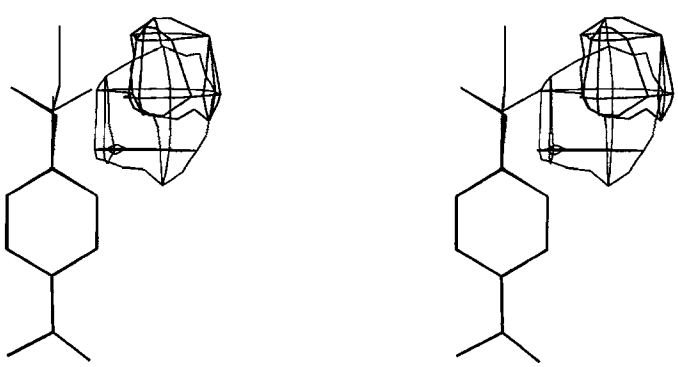

Fig. 4. Stereopicture of the negative charge–charge interactions at a contour level of −60 for position 5. The substituted benzoic acids related to the substituents occurring in position 5 (see scheme) are depicted with hydrogens omitted for clarity.

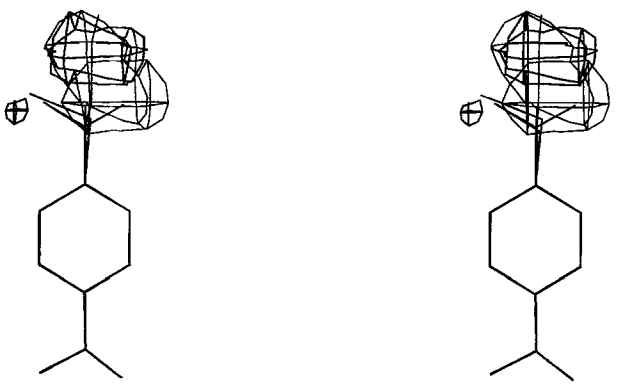

Fig. 5. Stereopicture of the positive nonbonded interactions at a contour level of 30 for position 3. The substituted benzoic acids related to the substituents occurring in position 3 (see scheme) are depicted with hydrogens omitted for clarity.

Statistics

The final statistics of the PLS models are listed in Table 4.

DISCUSSION

The object, as previously mentioned, is to explore the usefulness of aromatic substituent descriptors generated from a relatively simple and computationally rather nonintensive procedure based on a 3D representation of the substituents. This type of 3D description of structures has gained considerable interest through the recently developed 3D QSAR techniques [13,14]. Kim has shown that substituent parameters of Hammett type and other properties can be predicted by using a GRID-CoMFA approach [15,16]. To verify that this is also possible with the equations and procedure used in this work, these correlations were made for Hammett's sigma-

TABLE 4
FINAL STATISTICS OF THE DERIVED PLS QSAR MODELS

Property	N[a]	PC[b]	r^{2} [c]	r^{2}(CV)[d]	s[e]	F[f]	Desc.[g]
Sigma-para	39	7	0.965	0.828	0.089	120.80	FG
Sigma-para	39	2	0.805	0.691	0.193	74.47	3PP
Sigma-meta	40	9	0.971	0.914	0.053	112.03	FG
Sigma-meta	40	2	0.845	0.769	0.111	101.11	3PP
[^{3}H]spiperone binding	70	4	0.852	0.721	0.381	93.46	3PP
[^{3}H]spiperone binding	70	4	0.854	0.751	0.378	95.31	Ref. 2

[a] Number of compounds.
[b] Number of PLS components.
[c] Ordinary correlation coefficient squared.
[d] Cross-validated correlation coefficient squared.
[e] Standard deviation.
[f] Statistical F-value, the quota between explained and unexplained variance of the dependent variable.
[g] Full grid (FG) or 3PP description of the substituents.

para and sigma-meta. Since the variance, at this level of approximation, of the 3D grid points is concentrated around the substituents and not the benzoic moiety (see Figs. 1 and 2), it should be possible to also correlate Hammett's sigma-meta to the same set of formally para substituted structures used for the sigma-para analysis. The good statistical results, especially the high cross-validated r^2 values (Table 4) obtained for the two relationships, give some credibility to the fact that grid-based electronic interactions are capable of describing electronic effects. However, the purpose here is not to create 3D equivalents of Hammett's substituent parameters but to develop alternative descriptors for use in structure–property investigations in general.

A separate PCA was performed on each of the two sets of interactions (nonbonded and charge–charge) and both PCAs described a large portion of the variance in the respective original matrices. The resulting score vectors of the two PCAs are the new descriptors (see Table 1).

The electronic descriptors were subsequently tested for their ability to predict the same Hammett parameters as previously analysed. The results of these two PLS analyses were encouraging since the computed descriptors also seemed to be able to predict Hammett's sigma-para and sigma-meta with good accuracy and high cross-validated r^2 values (Table 4), albeit somewhat lower than in the analysis based on the full grid representation. However, this is not a very surprising result in view of the different objectives of the two projection methods (PCA and PLS). The purpose of the former one is to derive components which describe as much of the variance in the descriptor matrix as possible while the latter method uses the descriptor matrix to extract components which explain as much of the variance in the dependent variable as possible. The PCA-extracted components are therefore not as 'dedicated' as the PLS ones to correlate to a certain dependent variable (property).

Why then construct and use a database of 3D-based substituent descriptors instead of establishing a 3D QSAR model based on the actual compounds? This question is especially valid in view of the advantage of the PLS projection scheme compared with that of PCA as previously mentioned. Certainly one is more likely to be able to establish a better relationship with the former approach. But what about the predictability of new compounds? How well-balanced with respect to variance (variation of structures and/or substituents) are these 3D QSAR models? The grid-based 3D QSAR models give reasonable predictions as long as the new structures are within the 3D space already covered by the previously investigated structures with reasonable variance. New structures with elements perturbing 'unknown' (not covered) space will be poorly predicted, unless the part of the perturbing substituent/substructure has no effect on the property studied, because all the grid coefficients of this space will be close to zero. This is especially critical for the nonbonded interactions because of their rapid decline with increasing distance from the closest atom.

The construction of a well-balanced 3D database with respect to variance of the grid points around the database structures provides a good basis for deriving descriptors capable not only of structural interpolations (new compounds having substituents within the 3D domains already covered by the compounds comprising the QSAR model) but also of structural extrapolations, i.e., of compounds with substituents included in the database where the substituents probe 3D space not investigated/covered by the compounds of the QSAR model. Such extrapolations are virtually impossible for models based on the grid representation unless very favourable circumstances prevail, as discussed above. However, the PCA-derived descriptors certainly suffer from the same limitations as any other descriptors, e.g., when predicting compounds with substituents not covered by the compounds used to derive the relationship.

These new descriptors generated through a PCA on 3D grid interactions can be seen as 3D equivalents (3PPs) of the traditional 'principal properties' (PPs) [17], which are PCA-derived descriptors from multivariate characterizations. This latter approach has been used for characterizing aromatic substituents in toxicological evaluations [18].

How can one use the database of 3PPs?

The first, and simplest, way is to scan the database as such for interesting entries indicated by an already established relationship. The second possibility is to transform the derived PLS regression coefficients of the established QSAR model back into coefficients related to the original 3D grid space of the database compounds and investigate the resulting contour maps that indicate important areas of interactions which, in turn, may help to identify new (additional) substituents of potential interest which should be included into the database. It is this latter way of using the database that seems to be the most interesting one from a structure–property point of view. Thus, these maps seem to give the possibility of expanding the 3PP database with new substituents of interest from the basis of a previously developed QSAR (QSPR) – something which is not as easily done with traditional substituent databases of a physicochemical type based on experimental data or other databases which do not take the 3D structures of the substituents into account in an explicit manner. From this aspect, it is also of interest to let the initially developed descriptor database be of reasonable size and variation. In this way, old descriptors will stay mathematically rather similar when new substituents are included and new PCAs performed. Thus, previously derived models do not need to be recalculated every time the database is updated, especially in view of the error limits of the property investigated.

The 3PP descriptors were then further investigated through the analysis of the ability to inhibit [^{3}H]spiperone binding of some 70 substituted *N*-ethyl pyrrolidinyl benzamides which had previously been studied by using the same statistical method (PLS) but based on a more conventional physicochemical description of the substituents (sigma-para, sigma-meta, F, R, pi, L, B_1, B_5) [2]. The result of the analysis is good and the explained variance of the dependent variable, 85%, is well on a level with the results obtained in the previous investigation (86% explained variance). Both relationships point out the favourable influence of a methoxy group in the 5 position. This is well seen in the positive nonbonded and negative electronic interaction contour maps (Figs. 3 and 4), where the contours indicate the favourable negative electronic interactions from the oxygen as well as the positive nonbonded interactions from the methyl group. The 3 position is favoured by relatively large substituents such as Et, I and Pr (in that order of decreasing affinity). This is also well described by the positive nonbonded contour map (Fig. 5). The importance of a hydroxy group in position 2 as well as an *S* stereochemistry of the pyrrolidinyl side chain is also indicated in this QSAR, as was the case in the model mentioned in Ref. 2.

CONCLUSIONS

The three examples presented in this article seem to indicate that the 3PPs may serve as useful descriptors when developing QSAR models. These descriptors and the procedure with which they were derived may therefore represent interesting alternatives to other descriptors since the former ones are relatively easy to derive for new substituents of interest.

However, there is also the question of whether the substituents themselves should be aligned in

some manner or not [20]. The present 3PPs are not based upon an alignment of the respective substituent as such. Further development of this type of approach and more applications of 3PPs in quantitative structure–property correlations will hopefully bring additional light in the question of substituent alignment.

REFERENCES

1 Martin, Y.C., Quantitative Drug Design, Marcel Dekker, New York, 1978.
2 Norinder, U. and Högberg, T., Acta. Pharm. Nordica, 4 (1992) 73.
3 Norinder, U., Peptides, 12 (1991) 1223.
4 Norinder, U. and Jonsson, J., Quant. Struct.–Act. Relatsh., submitted for publication.
5 Van de Waterbeemd, H., Carrupt, P.-A., Testa, B. and Kier, L.B., In Wermuth, C.G. (Ed.) Trends in QSAR and Molecular Modelling 92, Proceedings of the 9th European Symposium on Structure–Activity Relationships, ESCOM, Leiden, 1993, pp. 69–75.
6 Hemken, H.G. and Lehmann, F., P.A., Quant. Struct.–Act. Relatsh., 11 (1992) 332.
7 Prof. K. Steliou, Department of Chemistry, University of Montreal, Montreal, Quebec, Canada H3C 3J7. Model version 2.96.
8 Stewart, J.J.P., MOPAC Version 6.0, QCPE, program no. 455, University of Indiana, Bloomington, IN 47405, U.S.A.
9 Del Re, G., Gavuzzo, E., Giglo, E., Lelj, F., Mazza, F. and Zappia, V., Acta Crystallogr., B33 (1977) 3289.
10 Wold, S., Technometrics, 20 (1979) 379.
11 Wold, S., Albano, C., Dunn III, W.J., 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, Reidel, Dordrecht, 1984, pp. 17–95.
12 Chem-X, developed and distributed by Chemical Design Ltd., Oxford, England.
13 Cramer III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
14 Norinder, U., J. Comput.-Aided Mol. Design, 4 (1990) 381.
15 Kim, K.H. and Martin, Y.C., J. Org. Chem., 56 (1991) 2723.
16 Kim, K.H., Quant. Struct.–Act. Relatsh., 11 (1992) 127.
17 Hellberg, S., Sjöström, M., Skagerberg, B. and Wold, S., J. Med. Chem., 30 (1987) 1126.
18 Skagerberg, B., Borelli, D., Clementi, S., Cruciani, G. and Ebert, C., Quant. Struct.–Act. Relatsh., 8 (1989) 32.
19 Hansch, C. and Leo, A., Substituents Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York, NY, 1979.
20 Clementi, S., personal communication.