Direct Prediction of Linear Free Energy Substituent Effects from 3D Structures Using Comparative Molecular Field Analysis. 1. Electronic Effects of Substituted Benzoic Acids

Ki H. Kim* and Yvonne C. Martin

Pharmaceutical Products Division, Abbott Laboratories, Abbott Park, Illinois 60064

Received July 18, 1990

We examined the ability of the comparative molecular field analysis (CoMFA) method to reproduce Hammett σ constants. The dataset includes 49 substituted benzoic acids. Molecular fields calculated with an H^+ probe and AM1 partial atomic charges produce a good fit and cross-validated estimate of σ. This estimate is more accurate than that found from a CoMFA analysis using charges based on fits to STO-3G electrostatic potential surfaces. It is also superior to that derived from regression analysis of the charges on the atoms. The relationships predicted the σ of 21 of 23 additional compounds to within 0.27 kcal/mol. We conclude that the CoMFA treatment of electrostatic effects is suitable for the examination of 3D quantitative structure–bioactivity relationships.

Introduction

Although organic molecules are three-dimensional, it is only recently that workers have attempted quantitative predictions of biological properties of molecules from their three-dimensional shape and electrostatic properties.[1-6] However, although traditional linear free energy relationships (LFER)/quantitative structure–activity relationships (QSAR) do not include consideration of 3D properties, they have been used for decades to predict chemical and biological properties of molecules.[7] Considering 3D structure promises to extend QSAR to more diverse datasets and to responses more sensitive to steric control. Therefore, we explored the most general method, CoMFA. Specifically, we asked if CoMFA descriptors include the information contained in the physicochemical descriptors typically used in QSAR. We were especially interested to explore the ability of CoMFA to predict the classic QSAR descriptor of electronic effects, the Hammett σ constant.

A CoMFA calculation uses different physicochemical descriptors than does QSAR. In a CoMFA analysis, the molecules are first superimposed in their proposed bioactive conformation. Then the potential energy field of each is calculated at various points on a lattice surrounding the molecule. The molecular field at any point in the lattice is the potential energy of interaction of some probe with the molecule. In this investigation we used the H^+ ion as the probe since the substituent effect on σ should be electrostatic in nature. For contrast we examined another probe, a methyl group, that should not predict σ.

Since the interaction energy is calculated at hundreds of points, one analyzes a CoMFA data matrix with the statistical technique of partial least squares, PLS.[8] CoMFA equations can be very complicated; hence, contour plots of regions of favorable and unfavorable potential energy values are often displayed.

For the major part of this investigation we considered a total of 49 analogues—the parent, 24 meta-, and 24 para-substituted benzoic acids. This includes all meta- and para-substituted benzoic acids for which a σ is available and for which the conformation to use is unambiguous.[9,10] We used a modified PLS method for the statistical evaluations. All equations were chosen by cross-validation. To further support the equations the σ values of other substituents were predicted from them.

Results

Correlation of Hammett σ Constants with Partial Atomic Charges. Recently, Sotomatsu et al.[11] studied a series of 27 benzoic acids. They showed that σ is linearly correlated with the partial atomic charges of the oxygens plus that of the hydrogen atom of the carboxylic acid. The charges were calculated by AM1 in the conformation in which the substituent is coplanar with the benzene ring. We extended these observations to include 49 compounds. Table I lists the partial atomic charges of the low-energy conformations and Table II the σ constants of the analogues used in this study. From eqs 1–3, Table III, we see that the literature correlation can be generalized to all 49 analogues.

Sotomatsu et al. excluded the *p*-CN derivative since it had the largest deviation. Although it had a large deviation in our studies also (0.26 from both eqs 2 and 3), we did not delete it nor any other analogue. However, the calculated σ values of the strong electron-withdrawing groups, such as SO_2CF_3, SO_2F, and SO_2CH_3, are more positive than the observed values.

The substituent in the conformation we used for eight meta and eight para analogues is not coplanar with the benzene ring. Instead the substituent is almost perpendicular to the ring. Table IV shows that the conformational energy of the coplanar conformation of these 16 analogues is higher than that of the noncoplanar conformation. Since the biological properties of a molecule may not result from its minimum energy conformation, we studied the effect of basing the calculation on these higher energy coplanar conformations. Equations 4–6, Table III, are the result. There is no relevant difference between these and the corresponding equations based on the low-energy conformations.

We next examined how the source of the partial atomic charges affects the quality of the correlation of σ. Partial atomic charges calculated by the methods of Gasteiger et al.[12,13] and Mullay[14] as implemented in TOPMOST[15] varied

(1) Simon, Z.; Badileuscu, I.; Plauchithiu, M. G.; Holban, S.; Glatt, H.; Kerek, F. *Eur. J. Med. Chem.* **1980**, *5*, 521.
(2) Hopfinger, A. J. *J. Am. Chem. Soc.* **1980**, *102*, 7196.
(3) Ghose, A. K.; Crippen, G. M. *J. Med. Chem.* **1985**, *28*, 333.
(4) Cramer, R. D. III; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
(5) Doweyko, A. M. *J. Med. Chem.* **1988**, *31*, 1396.
(6) Kato, Y.; Itai, A.; Iitaka, Y. *Tetrahedron* **1987**, *43*, 5229.
(7) Hansch, C.; Maloney, P. P.; Fujita, T.; Muir, R. *Nature* **1962**, *194*, 178.
(8) Lindberg, W.; Persson, J.-A.; Wold, S. *Anal. Chem.* **1983**, *55*, 643.
(9) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; John Wiley & Sons: New York, 1979.
(10) Charton, M. *Prog. Phys. Org. Chem.* **1981**, *13*, 119.
(11) Sotomatsu, T.; Murata, Y.; Fujita, T. *J. Comput. Chem.* **1989**, *10*, 94.
(12) Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219.
(13) Gasteiger, J.; Saller, H. *Angew. Chem., Int. Ed. Engl.* **1985**, *24*, 687.

0022-3263/91/1956-2723$02.50/0

Table I. AM1 Partial Atomic Charges[a] of Meta- and Para-Substituted Benzoic Acids

		low-energy conformation				conformation with the substituent coplanar with the benzene ring			
no.	substituent	$q_{(-O-)}$	$q_{(=O)}$	$q_{(H)}$	Σq	$q_{(-O-)}$	$q_{(=O)}$	$q_{(H)}$	Σq
1	H	−0.3170	0.3650	0.2456	−0.4364	−0.3170	−0.3650	0.2456	−0.4364
2	m-Br	−0.3143	−0.3602	0.2483	−0.4262	−0.3143	−0.3602	0.2483	−0.4262
3	m-CF_3	−0.3153	−0.3528	0.2505	−0.4176	−0.3153	−0.3528	0.2505	−0.4176
4	m-CH_3	−0.3169	−0.3658	0.2452	−0.4375	−0.3169	−0.3658	0.2452	−0.4375
5	m-Cl	−0.3144	−0.3600	0.2480	−0.4264	−0.3144	−0.3600	0.2480	−0.4264
6	m-CN	−0.3153	−0.3554	0.2498	−0.4209	−0.3153	−0.3554	0.2498	−0.4209
7	m-F	−0.3146	−0.3569	0.2483	−0.4232	−0.3146	−0.3569	0.2483	−0.4232
8	m-I	−0.3162	−0.3589	0.2480	−0.4271	−0.3162	−0.3589	0.2480	−0.4271
9	m-NH_2	−0.3146	−0.3665	0.2450	−0.4361	−0.3146	−0.3665	0.2450	−0.4361
10	m-NO_2	−0.3155	−0.3453	0.2528	−0.4080	−0.3155	−0.3453	0.2528	−0.4080
11	m-OCF_3	−0.3147	−0.3534	0.2491	−0.4190	−0.3147	−0.3534	0.2491	−0.4190
12	m-OH	−0.3156	−0.3583	0.2463	−0.4276	−0.3156	−0.3583	0.2463	−0.4276
13	m-OCH_3	−0.3159	−0.3601	0.2458	−0.4302	−0.3159	−0.3601	0.2458	−0.4302
14	m-SH	−0.3156	−0.3595	0.2472	−0.4279	−0.3156	−0.3595	0.2472	−0.4279
15	m-SCH_3	−0.3160	−0.3604	0.2464	−0.4300	−0.3160	−0.3604	0.2464	−0.4300
16	m-SCH_3	−0.3138	−0.3560	0.2497	−0.4201	−0.3138	−0.3560	0.2497	−0.4201
17	m-t-Bu	−0.3167	−0.3675	0.2449	−0.4393	−0.3167	−0.3675	0.2449	−0.4393
18	m-C_2F_5	−0.3157	−0.3528	0.2513	−0.4172	−0.3157	−0.3518	0.2509	−0.4166
19	m-CH_2Br	−0.3161	−0.3619	0.2471	−0.4309	−0.3152	−0.3639	0.2468	−0.4323
20	m-CH_2Cl	−0.3161	−0.3622	0.2472	−0.4311	−0.3151	−0.3642	0.2468	−0.4325
21	m-CH_2I	−0.3153	−0.3630	0.2472	−0.4311	−0.3154	−0.3638	0.2468	−0.4324
22	m-C_2H_5	−0.3166	−0.3664	0.2453	−0.4377	−0.3173	−0.3659	0.2451	−0.4381
23	m-SO_2CF_3	−0.3143	−0.3387	0.2557	−0.3973	−0.3146	−0.3367	0.2560	−0.3953
24	m-SO_2F	−0.3133	−0.3402	0.2557	−0.3978	−0.3131	−0.3398	0.2558	−0.3971
25	m-SO_2CH_3	−0.3157	−0.3459	0.2525	−0.4091	−0.3174	−0.3390	0.2528	−0.4036
26	p-Br	−0.3151	−0.3590	0.2480	−0.4261	−0.3151	−0.3590	0.2480	−0.4261
27	p-CF_3	−0.3133	−0.3522	0.2505	−0.4150	−0.3133	−0.3522	0.2505	−0.4150
28	p-CH_3	−0.3177	−0.3669	0.2449	−0.4397	−0.3177	−0.3669	0.2449	−0.4397
29	p-Cl	−0.3157	−0.3610	0.2476	−0.4291	−0.3157	−0.3610	0.2476	−0.4291
30	p-CN	−0.3139	−0.3553	0.2498	−0.4194	−0.3139	−0.3553	0.2498	−0.4194
31	p-F	−0.3167	−0.3625	0.2477	−0.4315	−0.3167	−0.3625	0.2477	−0.4315
32	p-I	−0.3145	−0.3587	0.2480	−0.4252	−0.3145	−0.3587	0.2480	−0.4252
33	p-NH_2	−0.3210	−0.3762	0.2429	−0.4543	−0.3210	−0.3762	0.2429	−0.4543
34	p-NO_2	−0.3113	−0.3460	0.2529	−0.4044	−0.3113	−0.3460	0.2529	−0.4044
35	p-OCF_3	−0.3163	−0.3610	0.2485	−0.4288	−0.3163	−0.3610	0.2485	−0.4288
36	p-OH	−0.3179	−0.3698	0.2455	−0.4422	−0.3179	−0.3698	0.2455	−0.4422
37	p-OCH_3	−0.3180	−0.3715	0.2449	−0.4446	−0.3180	−0.3715	0.2449	−0.4446
38	p-SH	−0.3157	−0.3655	0.2468	−0.4344	−0.3157	−0.3655	0.2468	−0.4344
39	p-SCH_3	−0.3164	−0.3664	0.2460	−0.4368	−0.3164	−0.3664	0.2460	−0.4368
40	p-SCF_3	−0.3154	−0.3565	0.2493	−0.4226	−0.3154	−0.3565	0.2493	−0.4226
41	p-t-Bu	−0.3182	−0.3667	0.2447	−0.4402	−0.3182	−0.3667	0.2447	−0.4402
42	p-C_2F_5	−0.3139	−0.3516	0.2509	−0.4146	−0.3131	−0.3512	0.2509	−0.4134
43	p-CH_2Br	−0.3162	−0.3618	0.2469	−0.4311	−0.3168	−0.3621	0.2466	−0.4323
44	p-CH_2Cl	−0.3158	−0.3617	0.2469	−0.4306	−0.3162	−0.3628	0.2468	−0.4322
45	p-CH_2I	−0.3163	−0.3625	0.2466	−0.4322	−0.3162	−0.3634	0.2468	−0.4328
46	p-C_2H_5	−0.3175	−0.3667	0.2450	−0.4392	−0.3154	−0.3691	0.2452	−0.4393
47	p-SO_2CF_3	−0.3092	−0.3390	0.2559	−0.3923	−0.3081	−0.3421	0.2564	−0.3938
48	p-SO_2F	−0.3095	−0.3391	0.2556	−0.3930	−0.3091	−0.3408	0.2557	−0.3942
49	p-SO_2CH_3	−0.3112	−0.3460	0.2528	−0.4044	−0.3082	−0.3500	0.2535	−0.4047

[a] Mulliken charges.

so little that we did not try the correlation. We next calculated the charges by an empirical electronegativity neutralization method.[16] Equations 7–9 in Table III show that, for predicting σ constants, charges calculated with this method are decidedly inferior to those calculated by AM1. We also calculated partial atomic charges by our implementation of the method of Weiner et al.,[17] ESPFIT. Specifically we calculated the wavefunction with the ab initio basis set STO 3G, calculated the location of points at a density of 6/Å^2 on the surface that encloses the molecule and is 1.4 Å from the van der Waals surface, calculated the electrostatic potential at these points from the wave function, and then did a least-squares fit these electrostatic potentials back to partial atomic charges centered at the atomic nuclei. The charges are listed in Table V and the correlations in eqs 10–12, Table III. Note that we did not include the four compounds that contain iodine since a basis set for it was not available. Although these ESPFIT charges are better predictors of Hammett σ for meta analogues, the correlation deteriorates for para analogues and the whole dataset. Thus, partial atomic charges calculated by AM1 best predict σ and hence, the observable property pK_a. Equations 1–3 form the reference points for the evaluation of CoMFA.

Correlation of Hammett σ Constants with CoMFA Descriptors. Are the CoMFA-calculated descriptors superior to partial atomic charges for the prediction of σ? To answer this question we used the same compounds, conformations, and charges as were used to deduce eqs 1–3 unless noted otherwise.

Clearly the interaction energy of a positive charge close to an atomic nucleus will vary dramatically with slight changes in position of the probe. This is especially a

(14) Mullay, J. *J. Am. Chem. Soc.* **1986**, *108*, 1770.
(15) Enslein, K. Health Designs, Inc., Rochester, NY 14604, 1989. TOPMOST V1.20 was used in this study.
(16) Goodford, P. J. *J. Med. Chem.* **1985**, *28*, 849. GRID V4.03 was used in this study.
(17) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.

Table II. Physicochemical Parameter Values Used in This Study

no.	substituent	σ_m[a]	σ_p[a]	$\mathcal{F}$[a]	$\mathcal{R}$[a]	σ_I[b]	σ_R[b]
1	H	0.00	0.00	0.00	0.00	0.00	0.00
2	Br	0.39	0.23	0.44	−0.17	0.47	−0.25
3	CF_3	0.43	0.54	0.38	0.19	0.40	0.11
4	CH_3	−0.07	−0.17	−0.04	−0.13	−0.01	−0.16
5	Cl	0.37	0.23	0.41	−0.15	0.47	−0.25
6	CN	0.56	0.66	0.51	0.19	0.63	0.08
7	F	0.34	0.06	0.43	−0.34	0.54	−0.48
8	I	0.35	0.18	0.40	−0.19	0.40	−0.16
9	NH_2	−0.16	−0.66	0.02	−0.68	0.17	−0.80
10	NO_2	0.71	0.78	0.67	0.16	0.67	0.10
11	OCF_3	0.38	0.35	0.38	0.00	0.39	−0.04
12	OH	0.12	−0.37	0.29	−0.64	0.24	−0.62
13	OCH_3	0.12	−0.27	0.26	−0.51	0.30	−0.58
14	SH	0.25	0.15	0.28	−0.11	0.27	0.12
15	SCH_3	0.15	0.00	0.20	−0.18	0.30	−0.38
16	SCF_3	0.40	0.50	0.35	0.18	0.42	0.08
17	t-Bu	−0.10	−0.20	−0.07	−0.13	−0.01	−0.18
18	C_2F_5	0.47	0.52	0.44	0.11	0.40	0.10
19	CH_2Br	0.12	0.14	0.10	0.05	0.20	−0.10
20	CH_2Cl	0.11	0.12	0.10	0.03	0.17	−0.08
21	CH_2I	0.10	0.11	0.09	0.03	0.17	−0.09
22	C_2H_5	−0.07	−0.15	−0.05	−0.10	−0.01	−0.14
23	SO_2CF_3	0.79	0.93	0.73	0.26	0.71	0.21
24	SO_2F	0.80	0.91	0.75	0.22	0.75	0.16
25	SO_2CH_3	0.60	0.72	0.54	0.22	0.59	0.11

[a] Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; John Wiley & Sons: New York, 1979. [b] Charton, M. *Prog. Phys. Org. Chem.* **1981**, *13*, 119.

Table III. Correlation of Hammett σ (pK_a Value) with the Sum of the Partial Atomic Charges at =O, O, and H(O) Atoms Calculated by Various Methods

eq	source of charges	equation	F	r	s	*press s*
		Set 1: σ_m, n = 25				
1	AM1	22.76 (±1.54)$\sum q$ + 9.94 (±0.65)	220	0.951	0.087	0.091
4	AM1 planar conformation	21.20 (±1.55)$\sum q$ + 9.28 (±0.66)	187	0.944	0.093	0.098
7	GRID	2.02 (±0.34)$\sum q$ + 1.21 (±0.16)	35	0.779	0.177	0.183
10	STO-3G ESPFIT	38.16 (±2.27)$\sum q$ + 14.35 (±0.84)	283	0.965	0.077	0.079
		Set 2: σ_p, n = 25				
2	AM1	25.16 (±1.50)$\sum q$ + 10.95 (±0.64)	281	0.961	0.115	0.123
5	AM1 planar conformation	25.49 (±1.49)$\sum q$ + 11.09 (±0.64)	294	0.963	0.112	0.119
8	GRID	2.84 (±0.54)$\sum q$ + 1.51 (±0.25)	28	0.740	0.281	0.288
11	STO-3G ESPFIT	33.07 (±3.03)$\sum q$ + 12.53 (±1.13)	119	0.922	0.169	0.175
		Set 3: $\sigma_{m,p}$, n = 49				
3	AM1	24.38 (±1.08)$\sum q$ + 10.62 (±0.46)	513	0.957	0.102	0.105
6	AM1 planar conformation	23.86 (±1.10)$\sum q$ + 10.40 (±0.47)	470	0.953	0.106	0.110
9	GRID	2.42 (±0.33)$\sum q$ + 1.36 (±0.15)	55	0.734	0.239	0.242
12	STO-3G ESPFIT	33.87 (±2.02)$\sum q$ + 12.80 (±0.75)	280	0.931	0.134	0.137

problem for lattice points inside the van der Waals surface of a compound. Hence, a major problem with the calculation of CoMFA electrostatic fields is how to treat points outside some but inside other molecules in the dataset. Cramer et al.[4] suggested that if a point is inside a compound, one should substitute the average of the electrostatic energy of those compounds for which this point is outside. Since this substitution has no obvious physical basis, we calculated the electrostatic energy only at lattice points outside every molecule. Specifically, we used only points outside the union surface of the compounds in the data set. For correlations of biological properties of molecules, these locations are sensible since we assume some of them are locations of atoms of the target macromolecule. However, for the correlation of Hammett σ constants, which are measured in aqueous solution, this choice may limit the precision of the predictions.

For each CoMFA analysis we first extracted 10 latent PLS variables. The variables were added to the equation in the order of their correlation with the dependent variable, not in the order of extraction. We chose the "best" equation by jackknifed cross-validation. This is the equation that produces the lowest or near the lowest sum of squares of (predicted − observed) values, *press s*. This means that the equation chosen is that which best predicts molecules not used in the analysis.

We first calculated the molecular fields using a probe with a charge of +1.0, a dielectric constant of the medium of 5.0, and a lattice spacing of 2 Å. Equations 13–15, Table VI, show that the fit and the *press s* are better for the CoMFA descriptors than for the partial atomic charges. Not shown in the table is that using AM1 charges calculated for the planar conformation or ESPFIT charges does not improve the quality of the fit or *press s*. The studies show that charges derived from AM1 calculations on the low-energy conformation are prefered for predicting σ with either correlation analysis or CoMFA.

We also explored other variations of the CoMFA calculations. For example, since σ is measured in the hydrogen-bonding solvent water, we examined the effect of adding a hydrogen-bond donor property to the probe. As seen from eqs 16–18, Table VI, this does not improve the results. Equations 19–21 show that changing the dielectric constant of the medium to 80 worsens the quality of the predictions presumably by reducing the differences between the compounds. Equations 22–24 show that using

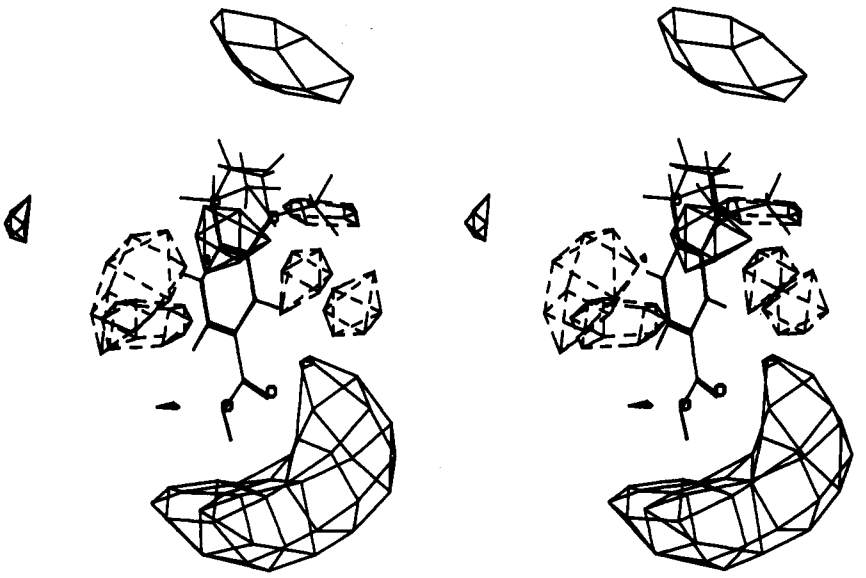

Figure 1. Stereoscopic coefficient contour map of the correlation described in eq II-1 from 25 meta-substituted benzoic acid analogues. The positive contour is in solid and the negative contour is in dash (contour shown at 0.02 level).

Table IV. Heat of Formation ($H°_f$)[a] of Meta- and Para-Substituted Benzoic Acids in Which the Substituent Has Two Accessible Conformations

		$H°_f$ (kcal/mol)	
no.	substituent	conformation with the substituent not coplanar with the aromatic ring	conformation with the substituent coplanar with the aromatic ring
18	m-C_2F_5	-314.62047	-312.99385
19	m-CH_2Br	-70.20734	-68.35465
20	m-CH_2Cl	-82.57182	-81.49122
21	m-CH_2I	-58.58937	-56.31542
22	m-C_2H_5	-81.35312	-80.90478
23	m-SO_2CF_3	-127.82020	-124.26876
24	m-SO_2F	-10.50856	-7.68907
25	m-SO_2CH_3	1.21334	4.73012
42	p-C_2F_5	-314.28257	-312.72946
43	p-CH_2Br	-70.22949	-68.30776
44	p-CH_2Cl	-82.50379	-81.42569
45	p-CH_2I	-58.69319	-56.32342
46	p-C_2H_5	-81.54678	-81.13134
47	p-SO_2CF_3	-126.45185	-123.12836
48	p-SO_2F	-10.00106	-7.42365
49	p-SO_2CH_3	1.69098	4.45143

[a] Calculated using AM1. Stewart, J. J. P. MOPAC V5.0 (QCPE No. 455).

1 Å spacing instead of 2 Å does not change the quality of the results although the calculations are at least 1 order of magnitude longer.

Equations 25–27 show that using a methyl probe does not lead to a good fit. This result shows that our results are probably not statistical artifacts and it thereby increases our confidence that the H^+ CoMFA results reflect real relationships.

In summary, CoMFA descriptors are not only able to describe the electronic effect of substituents on the σ values of substituted benzoic acids, but also are superior to partial atomic charges for this purpose.

CoMFA Prediction of Inductive/Field and Resonance Effects. The electronic effect of a substituent results from both its inductive/field and resonance effect on the property of interest. Since the correlation of σ with CoMFA descriptors is imperfect, we wondered if CoMFA includes only inductive or only resonance effects. We, therefore, studied the correlation of CoMFA descriptors with various properties of substituents, Table II.

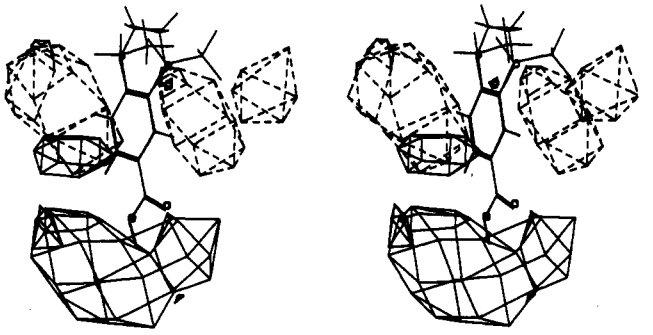

Figure 2. Stereoscopic coefficient contour map of the correlation described in eq II-2 from 25 para-substituted benzoic acid analogues. The positive contour is in solid and the negative contour is in dash. (The contour is shown at 0.07 level: contoured at a lower level to show the essential region.)

Table VII shows that the Swain–Lupton inductive/field parameter $\mathcal{F}$ (eqs 28–30) and the Charton σ_1 parameter (eqs 31–33) are significantly described by CoMFA electrostatic descriptors. However, the fit and predictive abilities of these equations are inferior to those of the composite parameter σ. To explore this further we also correlated the $\mathcal{F}$ and σ_1 values of 4-substituted bicyclo-[2.2.2]octane-1-carboxylic acids since in this system resonance effects are by definition absent. Equations 34 and 35, Table VII, show that the quality of these fits improved over the separate meta or para benzoic acid series, but not really over the total dataset. (The *press s* from eqs 34 and 35 is high because of the low number of compounds in the dataset: this means there are not enough compounds on which to base accurate predictions.) We conclude that although inductive-field effects are significantly correlated with CoMFA electrostatic calculations based on AM1 charges, these charges apparently include resonance effects as well.

In contrast, eqs 36–37 show that the two resonance parameters $\mathcal{R}$ and σ_R are not fit well by CoMFA electrostatic descriptors based on AM1 charges. We conclude that the AM1 charges include a balance of resonance and inductive/field effects of substituents and expect that CoMFA calculations based on these charges are reasonable to use for correlating biological data.

Figures 1–3 show the contours that result from eqs 13–15. Note that in each case the positive contour encloses the site of the reaction studied, the carboxyl group. The

Table V. Atomic Charges Calculated by the ESPFIT[a] Method

no.	substituent	$q_{(-O-)}$	$q_{(=O)}$	$q_{(H)}$	Σq
1	H	−0.3152	−0.2784	0.2177	−0.3759
2	*m*-Br	−0.3144	−0.2730	0.2212	−0.3662
3	*m*-CF_3	−0.3142	−0.2724	0.2210	−0.3656
4	*m*-CH_3	−0.3153	−0.2794	0.2173	−0.3775
5	*m*-Cl	−0.3143	−0.2708	0.2223	−0.3628
6	*m*-CN	−0.3140	−0.2711	0.2220	−0.3631
7	*m*-F	−0.3137	−0.2742	0.2195	−0.3684
8	*m*-I	-	-	-	-
9	*m*-NH_2	−0.3140	−0.2793	0.2170	−0.3763
10	*m*-NO_2	−0.3140	−0.2680	0.2231	−0.3589
11	*m*-OCF_3	−0.3139	−0.2724	0.2198	−0.3665
12	*m*-OH	−0.3145	−0.2748	0.2180	−0.3712
13	*m*-OCH_3	−0.3146	−0.2757	0.2177	−0.3726
14	*m*-SH	−0.3140	−0.2771	0.2177	−0.3734
15	*m*-SCH_3	−0.3145	−0.2777	0.2172	−0.3750
16	*m*-SCF_3	−0.3135	−0.2747	0.2197	−0.3686
17	*m*-t-Bu	−0.3152	−0.2805	0.2168	−0.3789
18	*m*-C_2F_5	−0.3144	−0.2720	0.2213	−0.3652
19	*m*-CH_2Br	−0.3146	−0.2764	0.2190	−0.3719
20	*m*-CH_2Cl	−0.3146	−0.2755	0.2197	−0.3703
21	*m*-CH_2I	-	-	-	-
22	*m*-C_2H_5	−0.3151	−0.2796	0.2172	−0.3775
23	*m*-SO_2CF_3	−0.3131	−0.2656	0.2243	−0.3544
24	*m*-SO_2F	−0.3129	−0.2658	0.2246	−0.3541
25	*m*-SO_2CH_3	−0.3138	−0.2691	0.2222	−0.3607
26	*p*-Br	−0.3148	−0.2757	0.2199	−0.3706
27	*p*-CF_3	−0.3134	−0.2713	0.2212	−0.3635
28	*p*-CH_3	−0.3161	−0.2811	0.2170	−0.3802
29	*p*-Cl	−0.3142	−0.2732	0.2209	−0.3664
30	*p*-CN	−0.3130	−0.2694	0.2222	−0.3602
31	*p*-F	−0.3157	−0.2795	0.2185	−0.3766
32	*p*-I	-	-	-	-
33	*p*-NH_2	−0.3183	−0.2893	0.2146	−0.3929
34	*p*-NO_2	−0.3125	−0.2668	0.2237	−0.3555
35	*p*-OCF_3	−0.3154	−0.2793	0.2189	−0.3758
36	*p*-OH	−0.3164	−0.2847	0.2168	−0.3843
37	*p*-OCH_3	−0.3167	−0.2853	0.2166	−0.3854
38	*p*-SH	−0.3163	−0.2843	0.2167	−0.3839
39	*p*-SCH_3	−0.3166	−0.2849	0.2161	−0.3854
40	*p*-SCF_3	−0.3155	−0.2794	0.2183	−0.3766
41	*p*-t-Bu	−0.3160	−0.2812	0.2164	−0.3809
42	*p*-C_2F_5	−0.3139	−0.2709	0.2212	−0.3636
43	*p*-CH_2Br	−0.3148	−0.2765	0.2189	−0.3725
44	*p*-CH_2Cl	−0.3146	−0.2748	0.2198	−0.3697
45	*p*-CH_2I	-	-	-	-
46	*p*-C_2H_5	−0.3159	−0.2810	0.2169	−0.3800
47	*m*-SO_2CF_3	−0.3115	−0.2638	0.2250	−0.3503
48	*p*-SO_2F	−0.3116	−0.2638	0.2250	−0.3503
49	*p*-SO_2CH_3	−0.3127	−0.2681	0.2230	−0.3579

[a] See text.

contour is positive because substituents that withdraw electrons make this region more electropositive and hence decrease the energy needed to remove the H^+ (increase the σ values). The negative contours arise because of the requirement of electrical neutrality in the molecule as a whole. Note that they are near the substituents. These contours conform to our qualitative expectation of what the CoMFA should detect.

Prediction of σ Constants. The true test of an equation is its ability to predict the values of compounds not included in its derivation. Therefore, we predicted the σ values for 23 additional substituents.[9] The σ values were forecast using both partial atomic charges, eq 3, and CoMFA, eq 15. Table VIII shows the values. Although there is good agreement between the observed and calculated values for both equations, the advantage of CoMFA is that its predictions are closer to the observed value in 14 of the 23 examples. Considering eq 15, for 12 of the 23 compounds the deviation between observed and predicted σ value is 0.10 or less, and for 18 compounds the deviation is 0.20 or less. The corresponding values for eq 3 are 9 and 16.

The conformation of compound 8, *m*-CH_3NH-benzoic acid, chosen for consistency with the molecules in the original dataset predicts σ with a deviation of −0.27 and −0.19 for eqs 3 and 15. However, there is a slightly lower energy conformation from which the deviation from eq 15 is 0.02. This result suggests that there are conformational influences on the CoMFA predictions of σ. It raises to 13 the number of compounds predicted to within 0.10 and to 19 the number predicted to within 0.20 log units.

The three large outliers from the fit to eq 3 are the *p*-CN, *p*-NH_2, and *p*-OH analogues with deviations of 0.26, −0.21, and −0.21. The two largest outliers from eq 15 are the *p*-OH and *p*-NH_2 for which the deviations are −0.16 and −0.27. For this reason, it is not surprising that neither eq 3 nor eq 15 predict the large negative σ value of *p*-$NHCH_3$ and *p*-$N(CH_3)_2$. However, for *p*-NH_2, *p*-OH, and *p*-$NHCH_3$ the equation based on the para-substituted analogues only (eq 14) predicts the σ values with deviations of −0.03, −0.12, and −0.17. Similarly, although the deviation of the predicted σ value of *p*-SO_2NH_2 from eq 15 is 0.39, the deviation is only 0.02 from eq 14. These results bring to 15 the number of compounds predicted to within 0.10 and to 21 the number predicted to within 0.20 log units. The differences in precision of prediction suggest there are

Table VI. Correlation of Hammett σ (pK_a) with Molecular Fields Calculated in Different Ways from AM1 Point Charges

eq	variation	Z^a	F^b	*r*	*s*	*press s*
	Set 1: σ_m, n = 25					
13	standard (probe q=1; dielectric=5; lattice spacing 2Å)	5	204	0.991	0.042	0.052
16	probe has H-bond donor properties	5	128	0.985	0.052	0.072
19	dielectric constant of medium = 80	7	82	0.985	0.055	0.088
22	1-Å lattice spacing	5	174	0.989	0.045	0.064
25	methyl probe (probe q = 0)	1	6	0.449	0.252	0.297
	Set 2: σ_p, n = 25					
14	standard	8	107	0.991	0.068	0.104
17	probe has H-bond donor properties	4	70	0.966	0.115	0.154
20	dielectric constant of medium = 80	2	56	0.914	0.173	0.214
23	1-Å lattice spacing	8	112	0.993	0.063	0.106
26	methyl probe (probe q = 0)	1	8	0.499	0.362	0.409
	Set 3: $\sigma_{m,p}$, n = 49					
15	standard	7	119	0.976	0.082	0.093
18	probe has H-bond donor properties	7	153	0.981	0.072	0.102
21	dielectric constant of medium = 80	4	67	0.927	0.136	0.158
24	1-Å lattice spacing	7	130	0.978	0.072	0.093
27	methyl probe (probe q = 0)	1	12	0.445	0.316	0.371

[a] Number of latent variables included. [b] Statistical F test of the significance of the least-significant variable to enter the equation.

Table VII. Correlation of Inductive/Field or Resonance Substituent Effects with Molecular Fields[a]

eq.	variation	Z	F	r	s	press s
	Set 1: Meta-Substituted and Unsubstituted Analogues, n = 25					
28	$\mathcal{F}_m$	5	47	0.962	0.074	0.100
31	σ_{I-m}	3	34	0.909	0.102	0.108
	Set 2: Para-Substituted and Unsubstituted Analogues, n = 25					
29	$\mathcal{F}_p$	4	28	0.920	0.104	0.123
32	σ_{I-p}	1	43	0.808	0.138	0.141
36	$\mathcal{R}_p$	3	12	0.796	0.168	0.194
37	σ_{R-p}	2	10	0.696	0.200	0.206
	Set 3: Meta- and Para-Substituted and Unsubstituted Analogues, n = 49					
30	$\mathcal{F}_{m,p}$	6	49	0.936	0.089	0.101
33	$\sigma_{I-m,p}$	5	40	0.906	0.100	0.111
	Set 4: 4-Substituted Bicyclo[2.2.2]octane-1-carboxylic Acids, n = 10					
34	$\mathcal{F}$	2	21	0.927	0.104	0.211
35	σ_I	2	19	0.921	0.104	0.192

[a] Model equation was chosen at the minimum *press s* value.

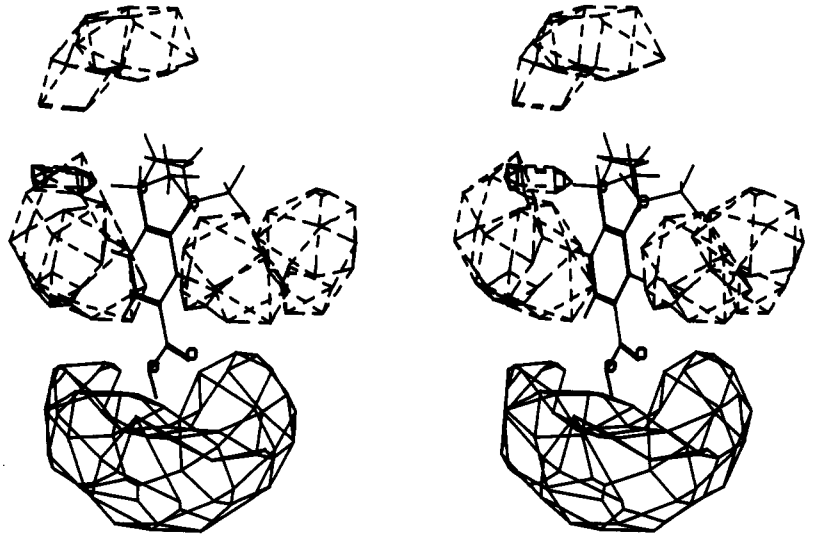

Figure 3. Stereoscopic coefficient contour map of the correlation described in eq II-3 from 49 meta- and para-substituted benzoic acid analogues. The positive contour is in solid and the negative contour is in dash (contour shown at 0.02 level).

differences between eqs 14 and 15 although statistically they appear to be identical. The differences in the regions of negative contours shown in Figures 2 and 3 reflect these differences in the equations.

The method for selecting the points for which to calculate the electrostatic contribution to σ appears to be the reason for the failure to predict the σ values of m-$OCOCH_3$, since this compound occupies regions in space not occupied by any of the analogues in the reference dataset.

Overall, these results show that eq 15 predicted an equilibrium constant to within 0.27 kcal/mol for 21 of the 23 compounds. This is quite respectable considering the simple nature of the calculations involved.

Discussion

We have oversimplified the description of the substituent effect on the pK_a of benzoic acids by basing our calculations on the unsolvated neutral form of the molecules. We thus have not explicitly included solvation or the effect of substituents on the relative stability of the benzoate anion. This might explain why we did not get a good CoMFA description of the resonance effects of substituents.

We did not include calculations of the anion because it would not be correct to calculate the partial charges on the unsolvated anion, since this species is not found in solution. Instead, one should also include solvent molecules and do the calculation on the complex. Clearly such a calculation would involve a lot more computer time and would also present the ambiguity as to where to place the solvent molecules, how many solvent molecules to use, and the relative orientation of the solute and the solvent. Additionally, in ligand binding to a macromolecule, our primary interest, the macromolecular binding site is more fixed in space since the side chains of a protein are not as free to move as are individual water molecules. Accordingly, the substituent effect on pK_a is not a perfect model for the substituent effect on the electrostatic contribution to the binding affinity of a ligand for a macromolecule. For these reasons we chose to examine the correlations based only on the unsolvated neutral molecule.

We calculated the molecular fields only at regions outside the union volume of the molecules in the dataset. This choice of region makes sense when correlating bioactivity, since, if the protein binds all the molecules similarly, we would not expect to find protein atoms there. However, in the correlation of pK_a, clearly water moves in to solvate the molecule or ion. Thus, it is possible that, in the more fixed matrix of a protein structure, CoMFA will more accurately describe the electrostatic effects of substituents.

In the CoMFA analysis of bioactivity, one would expect the electrostatic contours to give a map of the electrostatic features of the macromolecular binding site. In such a binding site there might be several spatially separate regions of electrostatic interaction between the ligands and the protein. Hence, substituent effects on the interaction

Table VIII. Prediction of Hammett σ Values of Substituted Benzoic Acids

no.	substituent	obs σ	calculated σ eq 3	calculated σ eq 15
1	m-CH=CH$_2$	0.05	-0.01	**0.02**
2	m-CH$_2$CN	0.16	**0.17**	0.14
3	m-CHO	0.35	0.18	**0.27**
4	m-CH$_2$OCH$_3$	0.02	**-0.03**	0.13
5	m-COCH$_3$	0.38	0.13	**0.26**
6	m-CONH$_2$	0.28	0.12	**0.28**
7	m-NCS	0.48	0.34	**0.44**
8	m-NHCH$_3$	-0.30	-0.03	**-0.11**[a]
9	m-N(CH$_3$)$_2$	-0.15	**-0.12**	-0.21
10	m-OCOCH$_3$	0.39	**0.15**	0.70
11	m-SCN	0.41	**0.35**	0.49
12	m-SO$_2$NH$_2$	0.46	**0.64**	0.76
13	p-CH=CH$_2$	-0.02	**-0.04**	-0.12
14	p-CH$_2$CN	0.01	0.18	**0.07**
15	p-CHO	0.42	0.34	**0.45**
16	p-CH$_2$OCH$_3$	0.03	**0.01**	-0.15
17	p-COCH$_3$	0.50	0.30	**0.39**
18	p-CONH$_2$	0.36	0.30	**0.34**
19	p-NCS	0.38	0.18	**0.19**
20	p-NHCH$_3$	-0.84	-0.48	**-0.49**[b]
21	p-N(CH$_3$)$_2$	-0.83	-0.57	**-0.58**
22	p-SCN	0.52	0.30	**0.59**
23	p-SO$_2$NH$_2$	0.57	**0.79**	0.89[c]

[a] For a slightly lower energy conformation the calculated σ is -0.28. [b] Equation 14 predicts a σ value of -0.69. [c] Equation 14 predicts a σ value of 0.55. A conformation that is 2 kcal/mol higher in energy has predicted σ values of 0.44 (eq 15) and 0.75 (eq 3).

of a ligand with a macromolecule may involve substituent effects on atoms at more than one position on the ligand. Because in CoMFA one does not measure substituent effects with respect to only one site, but lets the data decide the relationships, CoMFA is more attractive than traditional QSAR to study the electrostatic contributions to substituent effects on bioactivity.

Methods

Molecular Modeling. The starting coordinates were generated with CONCORD.[18] The core benzoic acid conformation was planar. All geometric variables were optimized with AM1 of MOPAC.[19,20] For meta-substituted benzoic acids, the conformation chosen has the substituent on the same side of the molecule as the carbonyl oxygen of the acid. The molecules were aligned by superimposing the unsubstituted benzoic acid moiety.

Partial atomic charges were calculated with AM1 or our modification of the method of Weiner, et al.[17] described above. (For sulfur atoms the MNDO parameters were used in AM1.) The coordinates and partial atomic charges for each molecule are in the supplemental material.

CoMFA Descriptor Calculation. The steric and electrostatic CoMFA descriptors were obtained by first calculating the interaction energies with the program GRID. A zero van der Waals radius and a charge of 1.0 was used for the H^+ probe and a radius of 1.95 Å and a charge of 0.0 was used for the methyl probe. For each molecule the energies at a total of 720 grid points were calculated with 2-Å spacing in a lattice of 14 × 16 × 18 Å.

Several considerations reduced the number of points to be considered with PLS. All steric energies with a value greater than 4.0 kcal/mol were truncated to 4.0. Any lattice point for which the standard deviation is less than 0.05 was discarded. To select only electrostatic energies calculated outside the union volume of the molecules in the dataset, we discarded any lattice point for which the steric energy for may molecule of the dataset is 4.0 kcal/mol or greater. For example, these procedures reduced the number of lattice points to 656, 654, and 637 for eqs 13, 14, and 15.

PLS Calculations. Because of earlier experience (manuscript in preparation) we did not use the standard PLS method, but instead a modification of it. We first extracted 10 orthogonal latent variables by the standard PLS algorithm. We observed that the order of extraction might not be the order of the correlation of the variables with the dependent property. Therefore, we added the variables to the equation in the order of their correlation with the dependent variable. The "best model" was chosen as that which minimizes the sum of squares of (predicted minus observed) using predictions made from leave-one-out jackknife method.

Supplementary Material Available: Coordinates and AM1 partial atomic charges for 49 benzoic acids (49 pages). Ordering information is given on any current masthead page.

(18) Rusinko, A. III; Skell, J. M.; Balducci, R.; McGarity, C. M.; Pearlman, R. S. The University of Texas at Austin and Tripos Associates, St. Louis, MO, 1988.

(19) Stewart, J. J. P. MOPAC V5.0 (QCPE No. 455). Ran with the keywords NOINTER and XYZ.

(20) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

The Perimeter Model and Magnetic Circular Dichroism of Porphyrin Analogues

Jacek Waluk[1] and Josef Michl*

Center for Structure and Reactivity, Department of Chemistry and Biochemistry, The University of Texas at Austin, Austin, Texas 78712-1167

Received July 16, 1990

The simple perimenter model is used to analyze the electronic structure of a series of conjugated macrocycles formally related to the $C_{20}H_{20}^{2+}$ perimeter, such as porphyrin, porphycene, secophyrin (parent of texaphyrin), and several that have not yet been synthesized. Particular attention is paid to consequences for UV–vis absorption and magnetic circular dichroism and to the effect of substitution and benzo annelation on these properties.

It has been known for some time that magnetic circular dichroism (MCD) of numerous cyclic approximately or exactly planar π-electron systems may be not only successfully computed at the semiempirical PPP or INDO/S

0022-3263/91/1956-2729$02.50/0