Pergamon

Bioorganic & Medicinal Chemistry, Vol. 3, No. 10, pp. 1389–1396, 1995

0968-0896/95 $9.50 + .00

0968-0896(95)00124-7

Description of Hydrophobicity Parameters of a Mixed Set from their Three-dimensional Structures

Ki H. Kim* and Daniel H. Kim

Pharmaceutical Products Division, Abbott Laboratories, 100 Abbott Park Road, Abbott Park, IL 60064, U.S.A.

Abstract—The logarithm of capacity factors (log k') previously measured from the reversed-phase high-performance liquid chromatography (RP-HPLC) and the octanol–water partition coefficients (log P) of a mixed set of substituted benzene, furan, benzofuran, pyrrole, 1-Me-pyrrole, indole, and *N*-methyl indole derivatives are correlated with the descriptors obtained from their three-dimensional structures using the comparative molecular field analysis (CoMFA) approach. The results provide an example where log k' and log P values are calculated directly from the three-dimensional structures for a mixed set of compounds.

Introduction

In quantitative structure–activity relationship (QSAR) studies, the potency of biological compounds are often correlated with the steric, electronic, and hydrophobic parameters of substituents or the whole molecules.[1] Various physicochemical or structural descriptors have been used for such parameters. For the hydrophobic parameter, the logarithm of the *n*-octanol–water partition coefficient (log P) along with the hydrophobic constant p has often been used. Beside log P, the chromatographic capacity factor (k') measured from a reversed-phase high-performance liquid chromatography (RP-HPLC)[2–11] has sometimes been used as a hydrophobic index of bioactive compounds. The capacity factor k' is defined by eqn 1, where t_r is the retention time of the compound and t_0 is the column dead time.

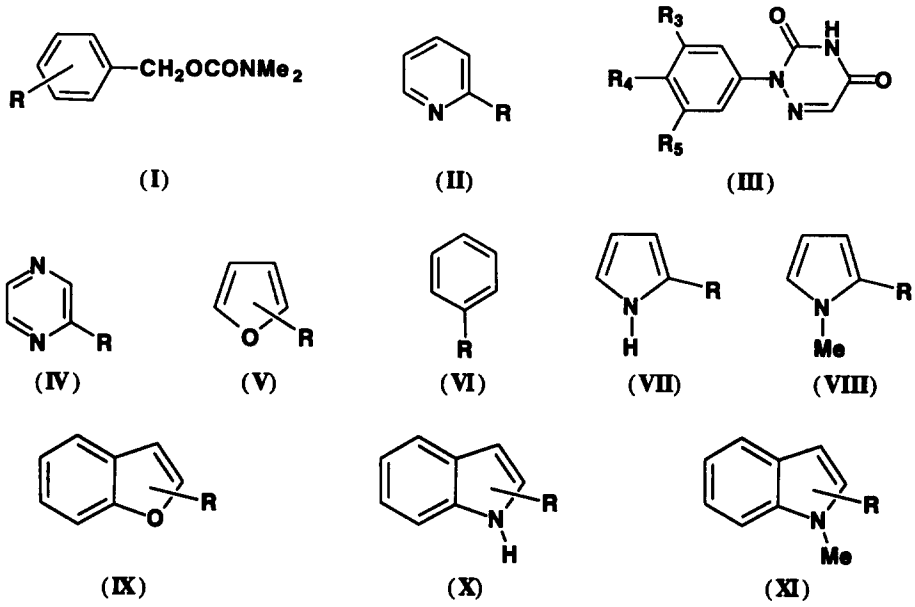
$$k' = (t_r - t_0)/t_0 \quad (1)$$
$$\log P = a \log k' + b \quad (2)$$
$$\log P = a \log k' + b\,HA + c\,HD + d \quad (3)$$
$$\log P = a \log k' + b\,HA + c\,HD + d\,\sigma + e. \quad (4)$$

Generally, there exists a linear relationship between the logarithm of the capacity factor (log k') and the corresponding log P values (eqn 2). Therefore, log k' has also been used to estimate log P. However, the relationship is often affected by the hydrogen-bonding ability of the compounds, as well as the type of stationary phase, and the composition of the mobile phase of the RP-HPLC employed. Thus, a more general relationship between log P and log k' has been proposed as shown in eqn 3. *HA* and *HD* in eqn 3 are the indicator variables for hydrogen-bonding acceptors and hydrogen-bonding donors, respectively, and σ is the Hammett electronic substituent constant or variation of it; a, b, c, d, and e in eqns 2–4 are the coefficients or error term of regression equation. This relationship is further modified in heteroaromatic systems due to the possible electronic interactions between the ring hetero-atom(s) and the substituent that might cause a modification in the hydrogen-bonding behavior of the substituent as well as the ring hetero-atom(s). In such a system, inclusion of an electronic interaction term is sometimes necessary as indicated in eqn 4.[3,4]

Since log k' remains as one of the viable options to describe the hydrophobic effects of the molecules in QSAR or to calculate the log P values, we investigated the possibility of calculating the log k' values directly from three-dimensional structures of the compounds. In our previous studies,[12] we examined the possibility with various sets of compounds using the H_2O probe and the hydrogen-bonding potentials in the comparative molecular field analysis (CoMFA) approach: substituted *N*,*N*-dimethylcarbamates (**I**), pyridine (**II**), triazine (**III**), pyrazine (**IV**), and furan (**V**) derivatives. So far, all of the previous studies were done with the series of close analogs separately. Although the results were very encouraging, the question of whether the approach can be extended to mixed structural types remained to be answered. In this study, we embarked on our study to answer this question with a mixed set of compounds. Since the measured log k' values can be affected by the experimental conditions, we used those log k' values determined under an identical condition. This set includes the derivatives of furan (**V**), benzene (**VI**), pyrrole (**VII**), 1-methylpyrrole (**VIII**), benzofuran (**IX**), indole (**X**), and 1-methylindole (**XI**). All of the log k' values were measured in Dr Yamagami's laboratory.[13]

Results and Discussion

In a preliminary investigation, the log k' values of three subsets were correlated: CO_2Me series, CONHEt series, and CONHMe series, each including six compounds. Each of these subsets contains the same aliphatic side



chain R, CO_2Me, CONHEt, or CONHMe, but includes six different ring structures (V–X). These compounds were aligned over the side chain R.

From the six compounds in the CO_2Me series, eqn 5 was obtained. The experimentally determined log k' values of these and other compounds included in this study are listed in Table 1. In eqn 5 and other correlation equations, L is the optimum number of components in the CoMFA model, n is the number of compounds used in the correlation, R^2 and R^2_{cv} are the squared correlation coefficients of the fitted model and the leave-one-out cross-validated test, respectively, and s and s_{cv} are the standard errors of estimation of the fitted model and the cross-validated test, respectively. F and p are the F-statistics and significance probability of the model. The percentages in the brackets following each equation are the amount of variance explained by each latent variable of the model.

CO_2Me series

$$\log k' = 0.048(\pm0.005)Z1_{H_2O} + 0.037(\pm0.007)Z2_{H_2O} + 0.513(0.032) \quad (5)$$
$$L = 2, n = 6, r^2 = 0.975, s = 0.079, r^2_{cv} = 0.805, s_{cv} = 0.187, F = 58.8, p = 0.004, [75, 23\%].$$

The two component model (eqn 5) explains 98% of the variance in log k' values of the CO_2Me derivatives with the R^2 of 0.98 and s of 0.08. The correlation is excellent, and its R^2_{cv} and s_{cv} values are good. The first component accounts for 75% of the variance in log k' values of these compounds. Similar but slightly better results were obtained from two other series, CONHEt and CONHMe subsets (eqns 6 and 7). Each of the two component models (eqns 6 and 7) explains above 99% of the variance in log k' values with cross-validated R^2 of 0.88 and 0.92, respectively. The results along with those of the structural type I–V obtained individually (Table 2) indicate that this approach can be extended to a mixed set of non-homologue ring structures as long as a suitable superposition can be attainable.

CONHEt series

$$\log k' = 0.079(\pm0.003)Z1_{H_2O} + 0.053(\pm0.004)Z2_{H_2O} + 1.387(0.022) \quad (6)$$
$$L = 2, n = 6, r^2 = 0.997, s = 0.053, r^2_{cv} = 0.875, s_{cv} = 0.280, F = 456.0, p = 0.0002, [83, 17\%]$$

CONHMe series

$$\log k' = 0.082(\pm0.004)Z1_{H_2O} + 0.037(\pm0.006)Z2_{H_2O} + 1.002(0.031) \quad (7)$$
$$L = 2, n = 6, r^2 = 0.993, s = 0.076, r^2_{cv} = 0.916, s_{cv} = 0.123, F = 216.9, p = 0.0006, [92, 7\%].$$

Encouraged by the results, all 59 compounds were combined in the next analysis. All of these compounds were mono-substituted analogs; one disubstituted compound (2-Br-5-CO_2Me) and one 5-substituted compound (5-methylindole) reported in the original paper were not included.[14] The result is shown in eqn 8. The eight component model explains 93% of the variance in log k' values of structurally diverse ring compounds. Since a larger number of structurally diverse compounds are included in this case, the optimum number of components in the CoMFA model is also larger than eqns 5–7. Its R^2_{cv} and s_{cv} values are excellent, considering the diverse structures included. Although the eight component model is statistically valid, the corresponding five component model accounts for most of the variance in log k' values, as indicated by the percentages of the variance accounted for by each component; the first five components explain 88% of the total variance in the data.

Table 1. Summary of the correlations between the log *k'* values and CoMFA descriptors for the combined set of furan (V), benzene (VI), pyrrole (VII), 1-methylpyrrole (VIII), benzofuran (IX), indole (X), and 1-methylindole (XI) analogs (the calculated log *k'* values were from eqn 8)

No		Name	Log *k'* obs	cal	dev	$Z1_{H_2O}$	$Z2_{H_2O}$	$Z3_{H_2O}$	$Z4_{H_2O}$	$Z5_{H_2O}$	$Z6_{H_2O}$	$Z7_{H_2O}$	$Z8_{H_2O}$
1	IX	2-CO_2Et	1.198	1.141	0.057	10.466	1.149	8.407	2.716	−2.699	0.965	1.844	1.650
2	IX	2-CO_2Me	0.912	0.743	0.169	9.035	−1.397	4.591	1.532	0.156	0.031	1.162	−0.433
3	IX	2-CONHEt	0.494	0.582	−0.088	9.949	−0.692	6.833	0.667	−3.812	0.465	−0.456	−0.263
4	IX	2-$CONH_2$	0.135	−0.006	0.141	6.896	−5.503	−3.266	−1.235	1.015	−4.367	3.093	0.229
5	IX	2-CONHMe	0.293	0.399	−0.106	8.450	−3.066	2.825	0.430	−0.115	−0.527	−0.040	−0.512
6	IX	2-CONHPr	0.733	0.666	0.067	9.874	−1.095	6.615	−0.067	−2.030	3.974	0.327	−1.105
7	IX	H	0.992	0.739	0.253	8.800	1.361	−3.743	1.343	−0.073	−4.631	−0.329	1.376
8	V	2-CO_2Et	0.351	0.413	−0.062	−3.249	0.581	9.680	4.290	−4.145	0.565	2.004	0.934
9	V	2-CO_2Me	0.043	0.088	−0.045	−4.696	−1.912	5.901	3.250	−0.891	0.058	1.676	−0.717
10	V	2-CONHEt	−0.359	−0.081	−0.278	−3.825	−1.171	8.105	2.213	−4.826	0.522	−0.062	−0.476
11	V	2-$CONH_2$	−0.847	−0.706	−0.141	−6.860	−6.157	−2.087	0.461	−0.218	−4.533	3.493	−0.257
12	V	2-CONHMe	−0.609	−0.673	0.064	−6.221	−4.713	1.711	−0.760	−1.252	−0.377	−0.892	−1.172
13	V	2-CONHPr	−0.072	−0.018	−0.054	−3.879	−1.580	8.024	1.613	−3.248	3.856	0.695	−1.697
14	V	2-Et	0.926	0.974	−0.048	−5.087	1.646	6.029	9.611	5.038	−1.141	0.360	−2.488
15	V	2-Me	0.599	0.340	0.259	−5.300	−0.340	2.042	5.000	2.055	−2.618	0.969	−0.377
16	V	3-CO_2Et	0.528	0.604	−0.076	−5.596	2.101	−9.349	5.271	1.022	6.338	2.311	1.611
17	V	3-CO_2Me	0.223	0.433	−0.210	−5.647	0.906	−10.127	2.862	0.653	5.474	4.281	3.503
18	V	3-CO_2Pr	0.858	0.748	0.110	−5.623	2.773	−8.374	6.884	1.209	6.631	1.446	1.144
19	V	3-CONHEt	−0.335	−0.418	0.083	−6.581	1.296	−11.996	2.931	−2.561	1.322	−4.113	−3.829
20	V	3-$CONH_2$	−0.756	−0.798	0.042	−7.020	−1.451	−12.141	−3.131	−4.571	−3.349	1.434	3.704
21	V	3-CONHMe	−0.561	−0.368	−0.193	−7.023	−0.776	−12.707	−0.374	−1.811	2.360	2.003	1.734
22	V	3-CONHPr	−0.066	−0.154	0.088	−6.644	1.908	−10.776	4.338	−1.929	2.166	−3.424	−2.984
23	V	H	0.259	0.077	0.182	−4.829	0.975	−2.275	3.013	−1.296	−4.748	0.160	0.868
24	X	2-CO_2Et	1.126	0.933	0.193	9.998	0.947	7.619	1.094	−1.941	1.022	−0.103	0.357
25	X	2-CO_2Me	0.849	0.734	0.115	8.500	−1.621	3.941	0.814	1.869	0.871	−0.108	−0.248
26	X	2-CONHEt	0.393	0.428	−0.035	8.849	−2.443	4.498	−0.043	−1.342	0.841	−1.986	0.668
27	X	2-$CONH_2$	0.085	0.014	0.071	6.482	−5.906	−3.237	−0.581	2.237	−3.955	1.518	0.369
28	X	2-CONHMe	0.212	0.532	−0.320	8.091	−3.104	2.978	0.611	1.765	0.976	−1.286	0.360
29	X	2-CONHPr	0.630	0.647	−0.017	8.742	−3.081	4.404	−0.571	1.752	6.403	−0.401	−0.148
30	X	2-Me	0.708	0.994	−0.286	8.070	0.173	0.324	2.992	4.122	−2.467	−0.346	−0.131
31	X	3-CO_2Et	0.865	0.720	0.145	8.222	1.872	−12.725	1.649	0.996	4.043	−1.690	−1.087
32	X	3-CO_2Me	0.589	0.537	0.052	8.126	0.533	−13.832	−1.206	0.685	3.347	0.779	1.142
33	X	3-CONHEt	0.053	0.027	0.026	7.753	1.175	−14.666	−0.224	−1.354	0.783	−5.860	−4.530
34	X	3-CONHMe	−0.163	−0.141	−0.022	7.307	−1.203	−16.464	−4.695	−1.367	0.884	−0.416	0.549
35	X	3-Me	0.866	1.020	−0.154	8.599	2.058	−6.136	1.472	1.500	−0.917	1.291	2.498
36	X	H	0.474	0.647	−0.173	8.674	1.271	−4.078	0.604	0.410	−4.807	−1.039	0.527
37	XI	H	0.967	1.050	−0.083	9.906	7.019	−0.796	−4.236	−0.188	−3.861	2.764	−1.533
38	VIII	2-CO_2Me	0.521	0.241	0.280	−3.663	2.368	4.203	−6.873	2.050	3.146	2.845	−0.060
39	VIII	2-CONHMe	−0.303	−0.261	−0.042	−4.176	1.423	2.437	−8.541	0.994	1.321	1.126	−2.896
40	VIII	2-CONHEt	−0.086	−0.169	0.083	−3.277	2.113	4.094	−9.373	−0.538	3.025	−0.405	−0.432
41	VIII	2-CONHPr	−0.169	0.057	−0.226	−3.287	1.275	4.133	−9.584	2.451	8.463	1.787	−1.447
42	VIII	2-$CONH_2$	−0.458	−0.510	0.052	−4.595	−0.074	−2.298	−8.044	−0.210	−4.613	5.241	−1.621
43	VIII	H	0.184	0.260	−0.076	−2.281	6.167	−0.852	−4.858	−1.406	−3.779	3.535	−2.353
44	VI	CO_2Et	1.006	0.998	0.008	−1.251	5.650	11.255	1.692	−1.912	0.417	0.944	1.731
45	VI	CO_2Me	0.706	0.801	−0.095	−2.431	3.774	8.176	1.042	1.241	0.173	1.204	0.246
46	VI	CONHEt	−0.047	−0.027	−0.020	−4.280	1.829	4.079	−3.870	0.684	−2.435	−7.197	3.449
47	VI	$CONH_2$	−0.377	−0.360	−0.017	−5.232	−1.100	−0.874	−2.192	1.595	−6.963	−1.483	−0.026
48	VI	CONHMe	−0.240	−0.124	−0.116	−4.137	1.531	3.871	−2.364	0.472	−4.265	−6.944	0.593
49	VI	CONHPr	0.200	0.272	−0.072	−3.908	2.421	5.512	−3.200	1.892	0.547	−6.786	3.819
50	VI	H	0.766	0.670	0.096	−2.762	5.763	0.035	0.399	0.676	−4.368	0.265	1.376
51	VI	Me	1.109	0.885	0.224	−3.322	4.353	3.467	2.019	3.683	−2.462	0.995	0.882
52	VII	2-CO_2Et	0.331	0.029	0.302	−3.991	−0.447	7.263	0.778	−2.880	1.038	−0.598	0.431
53	VII	2-CO_2Me	0.048	−0.017	0.065	−5.308	−2.672	4.141	1.489	1.161	0.898	−0.312	0.451
54	VII	2-CONHEt	−0.362	−0.456	0.094	−5.908	−3.810	2.857	−1.909	−0.393	2.087	−3.154	1.929
55	VII	2-$CONH_2$	−0.804	−0.866	0.062	−7.226	−6.677	−3.673	−0.645	1.058	−4.859	1.512	−0.571
56	VII	2-CONHMe	−0.600	−0.559	−0.041	−6.596	−4.306	0.758	−1.265	1.526	0.304	−2.543	−0.918
57	VII	2-CONHPr	−0.086	−0.202	0.116	−5.936	−4.327	3.020	−1.999	2.687	7.551	−1.023	0.548
58	VII	2-Et	0.236	0.366	−0.130	−4.434	1.288	5.266	4.524	1.238	−1.745	0.592	−4.482
59	VII	H	−0.279	−0.008	−0.271	−4.708	0.925	−2.622	2.232	−0.889	−5.080	−0.658	0.118

Table 2. Summary of the correlations between the log k' or log P and CoMFA descriptors for the individual set of carbamate (I), pyridine (II), triazine (III), pyrazine (IV), and furan (V) analogs

Set	$Z1_{H_2O}$	$Z2_{H_2O}$	$Z3_{H_2O}$	$Z4_{H_2O}$	$Z5_{H_2O}$	Intercept	L	n	s	R^2	s_{cv}	R^2_{cv}
					log k'							
carbamate	0.083	0.042	0.072	0.072		0.830	4	19	0.125	0.939	0.368	0.355
	(±0.008)	(±0.006)	(±0.011)	(±0.011)		(±0.029)						
furan	0.103	0.027	0.058	0.068	0.040	–0.030	5	17	0.093	0.978	0.434	0.541
	(±0.006)	(±0.003)	(±0.007)	(±0.010)	(±0.011)	(±0.023)						
pyrazine	0.070	0.058	0.029	0.028		–0.371	4	19	0.154	0.910	0.322	0.518
	(±0.007)	(±0.009)	(±0.010)	(±0.014)		(±0.035)						
pyridine	0.066	0.032	0.044	0.013	0.013	0.002	5	19	0.072	0.978	0.250	0.642
	(±0.003)	(±0.004)	(±0.005)	(±0.005)	(±0.006)	(±0.016)						
triazine	0.036	0.043	0.038	0.019	0.024	0.464	5	54	0.100	0.958	0.171	0.865
	(±0.001)	(±0.003)	(±0.003)	(±0.003)	(±0.005)	(±0.014)						
					log P							
carbamate	0.145	0.062	0.117	0.051		2.321	4	19	0.201	0.941	0.590	0.384
	(±0.013)	(±0.010)	(±0.018)	(±0.019)		(±0.046)						
furan	0.173	0.026	0.065	0.088	0.047	0.922	5	17	0.116	0.982	0.500	0.548
	(±0.009)	(±0.004)	(±0.007)	(±0.012)	(±0.013)	(±0.028)						
pyrazine	0.121	0.068	0.045	0.038		–0.396	4	20	0.186	0.935	0.385	0.668
	(±0.010)	(±0.010)	(±0.012)	(±0.016)		(±0.042)						
pyridine	0.141	0.067	0.035	0.044	0.019	1.009	5	20	0.124	0.974	0.406	0.631
	(±0.007)	(±0.008)	(±0.006)	(±0.011)	(±0.009)	(±0.028)						

$$\log k' = 0.049(\pm0.003)Z1_{H_2O} + 0.093(\pm0.007)Z2_{H_2O} + 0.019(\pm0.003)Z3_{H_2O} + 0.056(\pm0.005)Z4_{H_2O} + 0.059(\pm0.010)Z5_{H_2O} + 0.021(\pm0.006)Z6_{H_2O} + 0.029(\pm0.008)Z7_{H_2O} + 0.039(\pm0.012)Z8_{H_2O} + 0.235(0.021) \quad (8)$$

$L = 8$, $n = 59$, $r^2 = 0.928$, $s = 0.158$, $r^2_{cv} = 0.775$, $s_{cv} = 0.261$, $F = 80.8$, $p = 0.0001$, [35, 27, 5, 15, 5, 2, 2, 1%].

Since the electronic interactions between the ring hetero-atom(s) and the ring substituent might affect the chromatographic behavior of these compounds, the possible additional electronic effects were investigated using the H^+ probe in CoMFA. In the previous CoMFA studies,[15–18] the H^+ probe was shown to describe the electronic effects of substituents, such as on the pK_a values of various systems. Equation 9 shows the result obtained from both the H_2O and H^+ probes. A comparison of eqns 8 and 9 shows that the inclusion of electronic descriptors does not noticeably improve the quality of eqn 8. The negligible improvement of eqn 9 over eqn 8 indicates that the electronic influences of the substituents that have not been accounted for by the H_2O probe are negligible. The excellent correlation, without any separate hydrogen-bonding indicator variables as in eqn 3, shows that the H_2O probe with the GRID hydrogen-bonding potential function sufficiently describes the hydrogen-bonding effects of various substituents influencing the log k' values.

$$\log k' = 0.058(\pm0.003)Z1_{H_2O,H^+} + 0.055(\pm0.004)Z2_{H_2O,H^+} + 0.014(\pm0.002)Z3_{H_2O,H^+} + 0.029(\pm0.004)Z4_{H_2O,H^+} + 0.040(\pm0.006)Z5_{H_2O,H^+} + 0.044(\pm0.007)Z6_{H_2O,H^+} + 0.032(\pm0.007)Z7_{H_2O,H^+} + 0.021(\pm0.006)Z8_{H_2O,H^+} + 0.235(0.018) \quad (9)$$

$L = 8$, $n = 59$, $r^2 = 0.943$, $s = 0.140$, $r^2_{cv} = 0.803$, $s_{cv} = 0.244$, $F = 104.2$, $p = 0.0001$, [51, 20, 5, 5, 5, 4, 3, 1%].

Table 2 lists the calculated log k' values using eqn 8 along with the latent variables used and the corresponding experimentally determined log k' values. Figure 1 is a plot between the observed and calculated log k' values from eqn 8.

As one might expect from the correlation indicated in eqns 2–4, there is a high correlation between the log k' and log P values of these compounds as described in eqn 10.[14] This relationship is further improved by the inclusion of an hydrogen-bonding indicator variable as shown in eqn 11. In eqn 11, HB_D takes the value of 1 for compounds with -CONH- or ring-NH, 2 for those with both -CONH- and ring-NH, and 0 for all others. The number of compounds used in this equation is 61 because one disubstituted compound (2-Br-5-CO_2Me) and one 5-substituted compound (5-methylindole) were also included. The experimentally determined log P values of these compounds except the disubstituted compound and 5-methylindole used in eqns 10 and 11 are listed in Table 3.[14] Since there is a high correlation between the log k' and log P values, and the log k' values of these compounds were successfully described

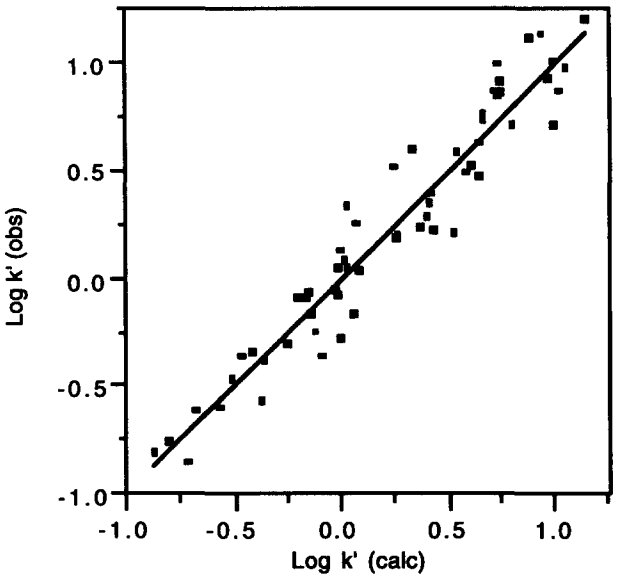


Figure 1. A plot between the observed and calculated log k' values using eqn 8.

above using the CoMFA approach, the same approach was applied for calculating these log P values. From the same 59 mono-substituted compounds used in eqn 8, eqn 12 was obtained.

$$\log k' = 0.605 \log P - 0.752 \quad (10)$$
$$n = 61,\ r^2 = 0.925,\ s = 0.149$$

$$\log k' = 0.561 \log P - 0.198\, HB_D - 0.503 \quad (11)$$
$$n = 61,\ r^2 = 0.995,\ s = 0.055$$

$$\begin{aligned}\log P = {} & 0.088(\pm 0.004)Z1_{H_2O} \\ & + 0.085(\pm 0.007)\ Z2_{H_2O} \\ & + 0.035(\pm 0.004)Z3_{H_2O} \\ & + 0.114(\pm 0.010)Z4_{H_2O} + 0.035(\pm 0.007)Z5_{H_2O} \\ & + 0.079(\pm 0.013)Z6_{H_2O} + 0.062(\pm 0.016)Z7_{H_2O} \\ & + 0.044(\pm 0.015)Z8_{H_2O} + 1.632(0.028)\end{aligned} \quad (12)$$
$$L = 8,\ n = 59,\ r^2 = 0.948,\ s = 0.213,$$
$$r^2_{cv} = 0.841,\ s_{cv} = 0.348,$$
$$F = 113.4,\ p = 0.0001,\ [49, 14, 9, 14, 3, 3, 2, 1\%].$$

The eight component model explains 95% of the variance in log P values with R^2 of 0.95 and s of 0.21. The correlation is excellent, and R^2_{cv} and s_{cv} are good, indicating the model's high predictability at least within the compounds included. The first component accounts for 49% of the variance in log P values, and the second through the eighth components explain an additional 14, 9, 14, 3, 3, 2, and 1%, respectively. As for log k', the eight component model is statistically valid. However, the corresponding four component model accounts for most of the variance in log P values as indicated by the percentages of the variance accounted for by each component; the first four components explain 86% of the total variance.

Unlike eqn 11, no separate parameter is needed in eqn 12 for the hydrogen-bonding effects of the substituents. The excellent correlation without the hydrogen-bonding indicator variable shows that the H_2O probe with the GRID hydrogen-bond potential function effectively describes the intra- or intermolecular hydrogen-bonding interactions involved with the various ring substituents in log P values.

Although eqn 12 indicates that there is no significant additional electronic influences on the log P values of these compounds, any possible electronic effects on the partitioning behavior were further examined using the H^+ probe. The result from the combination of hydrophobic and electronic effects provided in eqn 12 confirms that electronic properties do not play a noticeable role in log P values of these analogs. The negligible improvement of eqn 13 over eqn 12 indicates that the electronic influences of the substituents that have not been described by the H_2O probe and the hydrogen-bonding potential are again almost negligible.

$$\begin{aligned}\log P = {} & 0.095(\pm 0.004)Z1_{H2O,H+} \\ & + 0.101(\pm 0.007)\ Z2_{H2O,H+} \\ & + 0.072(\pm 0.009)Z3_{H2O,H+} \\ & + 0.029(\pm 0.005)Z4_{H2O,H+} \\ & + 0.015(\pm 0.004)\ Z5_{H2O,H+} \\ & + 0.059(\pm 0.011)Z6_{H2O,H+} \\ & + 0.055(\pm 0.012)Z7_{H2O,H+} \\ & + 0.020(\pm 0.008)\ Z8_{H2O,H+} + 1.632(0.027)\end{aligned} \quad (13)$$
$$L = 8,\ n = 59,\ r^2 = 0.952,\ s = 0.205,$$
$$r^2_{cv} = 0.843,\ s_{cv} = 0.347,$$
$$F = 123.6,\ p = 0.0001,\ [59, 20, 6, 3, 3, 2, 2, 1\%].$$

Table 3 lists the calculated log P values using eqn 12 along with the latent variables used and the corresponding experimentally measured log P values. Figure 2 is a plot between the observed and calculated log P values from eqn 12.

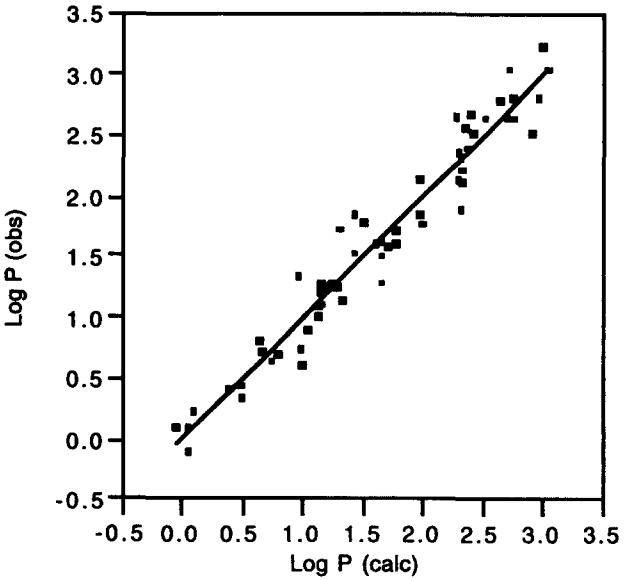


Figure 2. A plot between the observed and calculated log k' values using eqn 12.

Table 3. Summary of the correlations between the log *P* values and CoMFA descriptors for the combined set of furan (V), benzene (VI), pyrrole (VII), 1-methylpyrrole (VIII), benzofuran (IX), indole (X), and 1-methylindole (XI) analogs (the calculated log *k'* values were from eqn 12)

No		Name	Log *P*			$Z1_{H_2O}$	$Z2_{H_2O}$	$Z3_{H_2O}$	$Z4_{H_2O}$	$Z5_{H_2O}$	$Z6_{H_2O}$	$Z7_{H_2O}$	$Z8_{H_2O}$
			obs	cal	dev								
1	IX	2-CO_2Et	3.05	3.052	−0.002	10.875	−0.525	6.427	1.500	1.716	−2.167	1.219	3.297
2	IX	2-CO_2Me	2.53	2.424	0.106	9.393	−3.087	0.852	0.925	0.723	0.872	−1.257	1.684
3	IX	2-CONHEt	2.22	2.333	−0.113	10.829	−1.829	4.646	−0.607	1.807	−3.379	−0.312	0.716
4	IX	2-$CONH_2$	1.54	1.425	0.115	7.107	−6.780	−8.980	0.359	−4.880	0.498	0.635	2.436
5	IX	2-CONHMe	1.85	1.968	−0.118	9.102	−4.360	−1.515	−0.305	0.461	0.266	−1.461	1.019
6	IX	2-CONHPr	2.65	2.527	0.123	11.210	−1.247	5.857	−1.067	3.452	−1.932	0.916	−2.209
7	IX	H	2.67	2.277	0.393	8.308	1.599	−4.040	1.773	−4.653	−1.603	0.832	−1.160
8	V	2-CO_2Et	1.50	1.646	−0.146	−3.339	−1.970	6.371	1.960	3.265	−3.247	1.103	2.429
9	V	2-CO_2Me	1.00	1.118	−0.118	−4.838	−4.442	0.892	1.517	2.460	0.201	−1.017	1.138
10	V	2-CONHEt	0.61	1.014	−0.404	−3.427	−3.093	4.814	−0.150	3.464	−4.116	−0.043	0.124
11	V	2-$CONH_2$	−0.11	0.060	−0.170	−7.139	−8.282	−9.207	0.905	−3.193	−0.349	0.749	1.723
12	V	2-CONHMe	0.23	0.101	0.129	−5.780	−6.209	−3.190	−2.299	1.632	−0.470	−1.888	−0.435
13	V	2-CONHPr	1.10	1.153	−0.053	−3.056	−2.635	5.945	−0.548	5.099	−2.827	0.834	−3.071
14	V	2-Et	2.40	2.374	0.026	−5.958	0.259	3.853	9.264	1.802	2.839	−1.041	−3.829
15	V	2-Me	1.85	1.430	0.420	−5.998	−1.665	−0.794	4.807	−0.753	0.584	−0.129	−1.437
16	V	3-CO_2Et	1.78	2.003	−0.223	−6.149	5.039	−8.123	1.524	6.083	1.924	1.763	2.708
17	V	3-CO_2Me	1.28	1.649	−0.369	−6.204	3.309	−9.633	−0.147	3.980	2.252	4.146	1.370
18	V	3-CO_2Pr	2.36	2.294	0.066	−6.187	5.871	−6.878	2.796	7.234	1.755	0.844	4.183
19	V	3-CONHEt	0.72	0.656	0.064	−6.975	4.840	−10.166	−0.434	3.834	−2.889	−4.018	−0.585
20	V	3-$CONH_2$	0.09	−0.054	0.144	−7.596	0.057	−12.294	−3.766	−3.162	−3.033	3.686	−0.929
21	V	3-CONHMe	0.34	0.491	−0.151	−7.436	1.708	−12.547	−2.837	1.554	−0.391	2.271	−0.784
22	V	3-CONHPr	1.20	1.151	0.049	−7.090	5.427	−8.767	0.889	4.943	−1.974	−3.024	1.304
23	V	H	1.34	0.963	0.377	−5.846	0.308	−3.795	2.363	−3.028	−2.415	0.939	−1.787
24	X	2-CO_2Et	3.22	2.999	0.221	10.682	0.086	7.377	0.894	1.958	−1.228	0.397	1.377
25	X	2-CO_2Me	2.78	2.633	0.147	9.142	−2.407	1.752	0.970	1.497	2.548	−1.004	0.797
26	X	2-CONHEt	2.32	2.315	0.005	10.113	−2.563	2.629	−0.789	2.335	−1.268	0.094	0.437
27	X	2-$CONH_2$	1.61	1.608	0.002	6.987	−6.448	−8.273	1.165	−3.732	1.243	0.695	1.194
28	X	2-CONHMe	1.90	2.328	−0.428	9.083	−3.485	0.025	0.107	1.949	1.794	−0.779	0.372
29	X	2-CONHPr	2.80	2.757	0.043	10.666	−1.795	4.403	−0.999	5.223	1.520	2.250	−3.344
30	X	2-Me	2.53	2.913	−0.383	8.001	0.256	−0.158	4.597	−2.141	2.323	−0.196	−1.467
31	X	3-CO_2Et	3.04	2.732	0.308	8.385	6.324	−9.393	−0.208	3.308	1.171	−1.517	1.270
32	X	3-CO_2Me	2.57	2.351	0.219	8.283	4.429	−11.252	−2.233	1.009	1.782	1.433	−0.243
33	X	3-CONHEt	1.62	1.767	−0.147	8.044	6.007	−11.028	−1.841	1.627	−2.089	−5.538	−0.968
34	X	3-CONHMe	1.25	1.297	−0.047	7.608	2.696	−14.322	−5.169	−1.525	−0.329	0.326	−2.065
35	X	3-Me	2.80	2.957	−0.157	8.249	3.740	−4.346	2.097	−1.775	1.618	2.628	−0.948
36	X	H	2.14	2.295	−0.155	8.270	1.820	−3.574	1.763	−4.678	−0.874	0.354	−2.052
37	XI	H	2.64	2.760	−0.120	8.475	6.160	4.661	−0.176	−8.164	0.676	−0.572	−0.516
38	VIII	2-CO_2Me	1.80	1.501	0.299	−4.068	0.756	7.677	−4.240	−1.455	4.653	−1.061	2.956
39	VIII	2-CONHMe	0.71	0.802	−0.092	−4.522	−0.067	5.769	−5.425	−2.680	3.567	−3.259	0.177
40	VIII	2-CONHEt	1.09	1.123	−0.033	−2.948	1.860	10.017	−6.536	−1.344	0.991	−0.875	0.296
41	VIII	2-CONHPr	1.63	1.651	−0.021	−2.367	2.159	11.328	−6.404	1.773	4.327	1.464	−2.624
42	VIII	2-$CONH_2$	0.45	0.489	−0.039	−5.709	−2.195	−1.263	−2.949	−8.846	1.388	0.065	2.850
43	VIII	H	1.15	1.337	−0.187	−4.118	4.606	3.863	−1.241	−7.438	0.318	−0.657	−0.953
44	VI	CO_2Et	2.67	2.700	−0.030	−1.766	3.852	14.549	2.380	0.642	−1.585	1.365	3.086
45	VI	CO_2Me	2.12	2.330	−0.210	−3.103	1.820	9.923	2.278	−0.079	1.883	−0.480	2.176
46	VI	CONHEt	1.28	1.244	0.036	−3.873	2.475	9.061	−2.558	−2.861	−3.108	0.243	1.141
47	VI	$CONH_2$	0.64	0.739	−0.099	−5.755	−1.851	−0.550	1.228	−6.992	−1.208	−0.330	0.301
48	VI	CONHMe	0.90	1.040	−0.140	−4.096	1.408	7.054	−1.018	−3.319	−2.767	−1.928	−0.583
49	VI	CONHPr	1.72	1.773	−0.053	−3.169	3.779	11.913	−2.073	−0.827	−2.223	2.078	−0.090
50	VI	H	2.13	1.973	0.157	−4.233	5.368	4.326	2.604	−5.424	−0.574	1.136	−0.620
51	VI	Me	2.69	2.406	0.284	−4.459	3.506	6.586	4.633	−3.391	2.135	0.692	0.378
52	VII	2-CO_2Et	1.73	1.311	0.419	−3.594	−1.759	6.067	−0.410	3.073	−2.058	0.011	0.855
53	VII	2-CO_2Me	1.27	1.143	0.127	−5.017	−3.964	0.905	0.493	2.733	1.646	−0.804	0.681
54	VII	2-CONHEt	0.80	0.638	0.162	−4.732	−3.509	1.818	−3.063	3.186	−1.038	0.171	−0.669
55	VII	2-$CONH_2$	0.09	0.046	0.044	−7.233	−7.805	−9.326	0.772	−3.270	0.566	0.195	0.299
56	VII	2-CONHMe	0.42	0.395	0.025	−6.011	−4.788	−2.026	−1.824	1.878	1.684	−2.352	−1.610
57	VII	2-CONHPr	1.25	1.196	0.054	−4.211	−2.853	3.550	−2.705	6.315	2.352	2.240	−4.127
58	VII	2-Et	1.59	1.715	−0.125	−5.176	−0.477	3.861	5.150	0.819	1.646	−2.633	−3.104
59	VII	H	0.75	0.981	−0.231	−5.637	0.565	−3.326	2.306	−3.224	−1.881	0.402	−2.566

The results described above demonstrate the applicability of CoMFA in calculating log k' and log P values of a mixed set of heterocyclic systems directly from their three-dimensional structures. The reason why such excellent correlations were obtained from the CoMFA descriptors derived from the interaction energies using a H_2O probe and the hydrogen-bonding potential may be understood if one considers close correlations existing between the log P values of organic compounds and their physicochemical properties such as the molecular surface area or volume describing their 'steric' or bulk properties and the hydrogen-bond donating and accepting properties.[19-27] The H_2O probe used in the GRID energy calculations is an electrically neutral group and has a van der Waals radius of 1.70. The H_2O probe can donate up to two hydrogen-bonds and accept up to two hydrogen-bonds. Therefore, the energy calculated using the probe and the hydrogen-bonding potential in GRID comes from the steric and hydrogen-bonding interactions. Thus, the steric and the hydrogen-bond donating and accep-ting properties of the H_2O probe may be responsible for the superb role of the probe mimicking log P. It is not surprising to see the similar results for log k' since there is a close relationship between log P and log k'.

Conclusion

The method described in this study calculates the log k' and log P values from the three-dimensional structures of the molecules. The results obtained show much promise in calculating hydrophobic parameters directly from their three-dimensional structures. Since an estimation of hydrophobicity in heterocyclic systems is often more complicated, the results are especially encouraging. The results also indicate that it is not necessary to include log k' along with the CoMFA descriptors from the H_2O probe in the correlation to describe hydrophobic effects.

Experimental

Partition coefficients and logarithm of the chromatographic capacity factor (k')

The *n*-octanol–water log P and log k' values were taken from the literature.[13] The measurements of log P were done at pH values where the solute existed as the neutral form. The log k' values were measured using a commercial Capcellpack C_{18} column, and the data obtained from the 50% (v:v) of MeOH in H_2O as the eluent (log k' (M50)) was used in this analysis.

Molecular modeling

The starting coordinates were generated using CONCORD[28,29] in the graphics modeling package for small molecules at Abbott. In homolog series, the compound with a largest substituent was modeled first, and smaller substituents were built from the compound. All geometric variables were optimized with AM1 of AMPAC[30] with a VAX 9000 computer. The compounds were aligned over the ring structure among the same ring analogs, and over the side chains among the different ring analogs.

CoMFA energy calculation

The electrostatic and hydrophobic potential energy fields of each molecule were calculated at various lattice points surrounding the molecule using a H^+ or H_2O probe group, respectively, with the program GRID.[31-34] A van der Waals radius of 1.70 and a charge of 0.0 were used for the H_2O probe, and a van der Waals radius of 0.00 and a charge of 1.0 were used for the H^+ probe. The dielectric constant of environment and molecules were set to 5 and 4, respectively. For each molecule, the energies at a total of 2100 grid points were calculated with 2 Å spacing in a lattice of 26 × 28 × 18 (X = −11.5–14.5; Y = −14.5–13.5; Z = −10.5–7.5). In order to investigate the 'best' position for the lattice points, the grid box was shifted by 0.5, 1.0, and 1.5 Å in each of the directions X, Y, and Z. All the energy values with a value greater than 4.0 kcal mol^{-1} were truncated to 4.0. Any lattice point for which the standard deviation of the energies is less than 0.05 was discarded. These procedures reduced the number of lattice points to 216 and 1575 for the H_2O and H^+ probes, respectively.

Partial least squares (PLS) calculations

Unless otherwise noted, 10 orthogonal latent variables were first extracted by the standard PLS algorithm.[35] The number of latent variables in the 'best' correlation model was chosen as that which significantly minimized the sum of the squares of the difference in activity between the leave-one-out cross-validation test. After the number of latent variables was established, the 'best' correlation model was derived including all the compounds. The final model was further validated by the overall and the stepwise F-statistics. If F-statistics did not support the model, the least significant latent variable was eliminated and the model was rederived. The variables Z1–Z8 in the correlation equations are the first through eighth latent variables from the PLS analysis. The probe group used in deriving the latent variables is indicated as a subscript following each latent variable.

References

1. Martin, Y. C. *Quantitative Drug Design*; Dekker: New York, 1978; p. 425.

2. Terada, H. *Quant. Struct.–Act. Relat.* **1986**, *5*, 81.

3. Miyake, K.; Mizuno, N.; Terada, H. *J. Chromatogr.* **1988**, *439*, 227.

4. Yamagami, C.; Takami, H.; Yamamoto, K.; Miyoshi, K.; Takano, N. *Chem. Pharm. Bull.* **1984**, *32*, 4994.

5. Minick, D. J.; Frenz, J. H.; Patrick, M. A.; Brent, D. A. *J. Med. Chem.* **1988**, *31*, 1923.

6. Yamagami, C.; Ogura, T.; Takao, N. *Chem. Pharm. Bull.* **1991**, *39*, 1217.

7. Thus, J. L. G.; Kraak, J. C. *J. Chromatogr.* **1985**, *320*, 271.

8. Tayar, N. E.; Tsantili-Kakoulidou, A.; Roethlisberger, T.; Testa, B.; Gal, J. *J. Chromatogr.* **1988**, 237.

9. Taylor, P. J. In *Quantitative Drug Design*; C. A. Ramsden, Ed.; Pergamon: Oxford, 1990; Vol. 4, pp. 241–294.

10. Hammers, W. E.; Meurs, G. J.; De Ligny, C. L. *J. Chromatogr.* **1982**, *247*, 1.

11. Braumann, T. *J. Chromatogr.* **1986**, *373*, 191.

12. (a) Kim, K. H. *Quant. Struct.-Act. Relat.* **1995**, *14*, 8; (b) Kim, K. H. *J. Comput.-Aid. Mol. Des.*, in press.

13. Yamagami, C.; Yokota, M.; Takao, N. *J. Chromatogr. A* **1994**, *662*, 49.

14. Yamagami, C.; Takao, N. *Chem. Pharm. Bull.* **1992**, *40*, 925.

15. Kim, K. H.; Martin, Y. C. *J. Med. Chem.* **1991**, *34*, 2056.

16. Kim, K. H.; Martin, Y. C. *J. Org. Chem.* **1991**, *56*, 2723.

17. Kim, K. H.; Martin, Y. C. In *QSAR: Rational Approaches to the Design of Bioactive Compounds*; C. Silipo; A. Vittoria, Eds; Elsevier: Amsterdam, 1991; p. 151.

18. Kim, K. H. *Quant. Struct.–Act. Relat.* **1992**, *11*, 127.

19. Pearlman, R. S. In *Partition Coefficient*; W. J. I. Dunn; J. H. Block; R. S. Pearlman, Eds; Pergamon Press: New York, 1986; p. 3.

20. Leo, A.; Hansch, C. *J. Med. Chem.* **1976**, *19*, 611.

21. Camilleri, P.; Watts, S. A.; Boraston, J. A. *J. Chem. Soc. Perkin Trans. II* **1988**, 1699.

22. Dunn, W. J. I.; Koehler, M. G.; Grigoras, S. *J. Med. Chem* **1987**, *30*, 1121.

23. Leahy, D. *J. Pharm. Sci.* **1986**, *75*, 629.

24. Tunon, I.; Silla, E.; Pascual-Ahuir, J. L. *Protein Engng* **1992**, *5*, 715.

25. Kamlet, M. J.; Doherty, R. M.; Abraham, M. H.; Marcus, Y.; Taft, R. W. *J. Phys. Chem.* **1988**, *92*, 5244.

26. Lien, E. J.; Banerjee, S.; Khan, A.; Gao, H.; Wang, F. In *Trends in QSAR and Molecular Modeling 92, Proceedings of the 9th European Symposium on Structure–Activity Relationships: QSAR and Molecular Modeling*; C. G. Wermuth, Ed.; ESCOM: Leiden, 1993; p. 303.

27. El Tayar, N.; Testa, B. In *Trends in QSAR and Molecular Modeling 92, Proceedings of the 9th European Symposium on Structure–Activity Relationships: QSAR and Molecular Modeling*; C. G. Wermuth, Ed.; ESCOM: Leiden, 1993; p. 101.

28. Rusinko, III A.; Sheridan, R. P.; Nilakantan, R.; Haraki, K. S.; Bauman, N.; Venkataraghavan, R. *J. Chem. Info. Comp. Sci.* **1989**, *29*, 251.

29. Rusinko, III A.; Skell, J. M.; Balducci, R.; McGarity, C. M.; Pearlman, R. S. The University of Texas at Austin and Tripos Associates: St. Louis, Missouri, U.S.A.

30. Dewar, M. J. S.; Zoebish, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

31. Boobbyer, D. N.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. *J. Med. Chem.* **1989**, *32*, 1083.

32. Wade, R. C.; Clark, K. J.; Goodford, P. J. *J. Med. Chem.* **1993**, *36*, 140.

33. Wade, R. C.; Goodford, P. J. *J. Med. Chem.* **1993**, *36*, 148.

34. Goodford, P. *J. Med. Chem.* **1985**, *28*, 849.

35. Wold, S.; Johansson, E.; Cocchi, M. In *3D QSAR in Drug Design. Theory Methods and Applications.*; H. Kubinyi, Ed.; ESCOM: Leiden, 1993; p. 523.

(*Received in U.S.A.* 12 *May* 1995; *accepted* 28 *June* 1995)