Comparative Molecular Field Analysis Combined with Physicochemical Parameters for Prediction of Polydimethylsiloxane Membrane Flux in Isopropanol

Rong Liu[1] and Lloyd E. Matheson[2,3]

Received March 17, 1993; accepted July 18, 1993

Comparative molecular field analysis (CoMFA) combined with various physicochemical parameters were used to develop three-dimensional quantitative structure–transportability relationships (3-D QSTR) to predict membrane flux for 108 aromatic and heteroaromatic compounds through polydimethylsiloxane (PDMS) membranes in isopropyl alcohol (IPA). Sybyl, a comprehensive computational molecular modeling package, was used to analyze the data. Optimized molecular models were selected using molecular modeling techniques. Partial least-squares (PLS) regression combined with crossvalidation or bootstrapping was used as the statistical method to establish the predictive models. Prediction was good for the steady-state flux using both steric and electrostatic field descriptors combined with a functional group classification technique. Predictive ability was substantially increased in a model using CoMFA descriptors along with log mole fraction solubility for the penetrants in isopropanol, a hydrophobic term, f_{chex}, which is used to estimate the partition coefficient between cyclohexane and water, and the addition of an intramolecular hydrogen bonding (*IHB*) term. The crossvalidated r^2 and the conventional r^2 for this model were 0.951 and 0.973, respectively. Excellent predictions are demonstrated for the membrane flux of the compounds both inside and outside the data domain.

KEY WORDS: comparative molecular field analysis; three-dimensional quantitative structure–transportability relationships; membrane diffusion prediction; molecular modeling; partition coefficient; solubility; partial least squares.

INTRODUCTION

Conventional quantitative structure–activity relationship (QSAR) techniques possess several limitations including their applicability only within narrow compound families of similar chemical structure; the lack of pertinent data needed to set up the QSAR equations; the use of only a relatively small number of predictive parameters, leading to the problem of collinearity due to the limitations of classical statistical methods; and the inability to differentiate stereoisomers, conformations, and even substitution positions (1–3). The usefulness of the QSAR technique for drug design and in other fields will increase considerably if these limitations can be overcome.

Comparative molecular field analysis (CoMFA), developed by Cramer and others (4), is one of the most promising new approaches in the three-dimensional (3-D) QSAR field. This approach is based on the fact that the most relevant calculable property values are dependent on shape and on two additional observations: (i) Molecular interactions which produce an observed biological effect are usually noncovalent; and (ii) molecular mechanics force fields, most of which treat noncovalent interactions only as steric and electrostatic forces, can account precisely for a great variety of observed molecular properties (5). It combines partial-least squares (PLS) regression, factor analysis, crossvalidation, and bootstrapping validation methods to regress hundreds of field columns against activities (2,5,6). Recent reports of CoMFA applications include the study of the effect of shape on binding affinities of 21 steroids to corticosteroid- and testosterone-binding globulins (7), quantitation of the effect of structural change on carbonyl addition for 11 carbonyl compounds (8), and correlation of binding affinities of 37 compounds to the benzodiazepine receptor inverse agonist site (9).

It has become better appreciated in recent years that the pharmacological activity of a drug molecule depends not only on its affinity for a target receptor but also on its transportability to the biophase. The study of the relatively new area of quantitative structure–transportability relationships (QSTR) is now recognized as important, especially in drug delivery research, where the prediction of penetration rate through skin and other absorptive membranes would be an extremely useful tool. While both the process and the controlling parameters of passive diffusion have been widely studied, the ability to predict the flux of a material through a membrane has not been inordinately successful to this point in time (10). It is obvious that the capability of predicting the rate of diffusion through biological membranes would be helpful not only for estimation of absorption rate and for selection of the analogue with the best absorption characteristics within a class of drugs but also for estimation of the time course of the drug in the body if its elimination kinetics are known.

The importance of the partition coefficient to the diffusion process has long been recognized (11–13). It has been shown that the relationship between the steady-state membrane diffusion rates and the partition coefficient (14,15) can be described in large part by hydrophobic parameters, such as p and f (16,17).

It has been observed that isomers with the same functional group on different ring positions not only have different diffusion rates, but also have significant differences in solubility (16). It was also reported that there is a good relationship between the steady-state flux and the log mole fraction solubility of the penetrants (18,19).

This article introduces the use of the standard Tripos force field (5.2) to build up a 3-D database and the use of CoMFA to analyze these flux data in order to establish a 3-D CoMFA QSTR model for prediction of flux through polydimethylsiloxane (PDMS) membrane material for a series of aromatic and heteroaromatic organic compounds in an isopropanol solvent system (17,18,20,21). There are two additional aspects which need to be considered further to in-

[1] Present address: American Cyanamid Company, Lederle Laboratories, Pearl River, New York 10965.
[2] Division of Pharmaceutics, College of Pharmacy, University of Iowa, Iowa City, Iowa 52242.
[3] To whom correspondence should be addressed.

0724-8741/94/0200-0257$07.00/0

crease the predictive ability and to improve the goodness of fit for the 3-D CoMFA QSTR models. The first is the consideration of the actual concentration of a molecule in contact with the membrane during the flux experiment. To do this a descriptor for solubility was added to the models. The second aspect is the addition of other important physicochemical descriptors such as the partition coefficient and hydrogen bonding ability into the model.

MATERIALS AND METHODS

Experimental

Solubility Determinations

The solubility of each solid diffusant was determined in triplicate using a procedure described by Hu and Matheson (18).

Diffusion Studies

The determination of the steady-state flux of aromatic and heteroaromatic compounds was described earlier by Hu and Matheson (18).

Data Analysis

Molecular Modeling

Compounds from the flux data set including 32 benzene, 35 pyridine, 35 quinoline, and isoquinoline derivatives plus 6 other nitrogen-containing hetroaromatic compounds were selected (18) to create a 3-D molecular structure database to be utilized to overcome the inability of conventional QSAR approaches to differentiate the effects of stereoisomers, conformations, and substituent positions on the prediction of flux. All molecular models were constructed using the SKETCH method in SYBYL Version 5.3 (2) on a Silicon Graphics 4D120GTX Graphics Workstation. The standard Tripos force field, Version 5.2 (2,22), using default tailor settings was applied for all optimization procedures, unless otherwise stated. After construction of the geometry of a molecule, the molecular energy of the structure was minimized using the CLEAN UP procedure with MAXIMIN2 followed by a DYNAMICS run without electrostatics and one more energy minimization with MAXIMIN2. The GAST-HUCK molecular charge calculation method was used for the entire data set. To be certain that an optimal conformer was obtained, the systematic conformational SEARCH was performed for each molecule with side chains. The torsional space of a side chain was searched at 10° increments. The final structure was minimized once more using MAXIMIN2.

Molecular Alignment

The following alignment rules are illustrated by structures. A through E in Fig. 1. Benzene was chosen as the reference molecule and placed at the origin and in the x–y plane. The atoms at positions 1, 3, and 5 of the benzene ring were selected as the reference atoms. These are fitted by the remaining molecules in the data set using the following criteria: (i) If there was no functional group attached to the heterocyclic ring system, the heteroatom was designated the prime or position 1 atom and was fitted to the reference carbon atom at position 1; (ii) if there was only one functional group, the aromatic carbon atom connected to the functional group was selected as the prime atom and was fitted to the reference carbon atom at position 1; (iii) if there were two or more functional groups, the aromatic carbon atom connected to the functional group with the greatest effect on flux, based on its flux fragmental value as shown in Table I (18), was selected as the prime atom; and (iv) since all the molecules in the data set had at least one six-membered aromatic ring, the six-membered ring was aligned with the reference benzene ring as just described, but with the added proviso that the arrangement also attempted to align any bulky portion of the aligned molecule in the direction of carbon atoms 4 and 5 on the reference molecule. If possible, any functional group which was not in the x–y plane was arranged in the minus z direction. After these considerations, all the conformers were individually least-squares fit to the reference molecule using the FIT option in the COMPARISON menu of SYBYL Version 5.3.

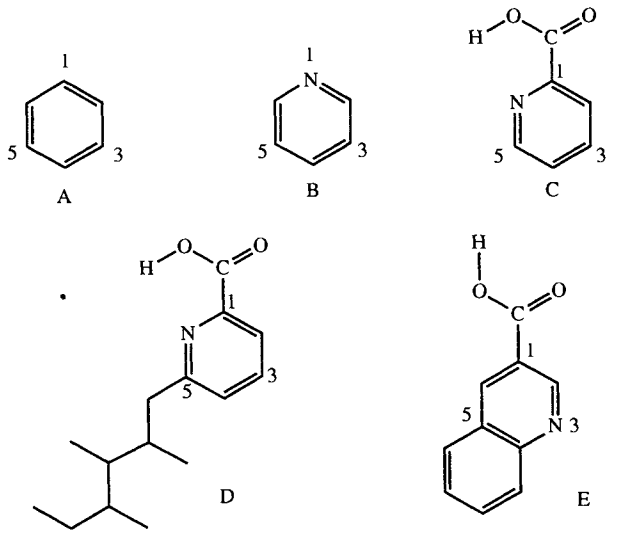

Fig. 1. Illustration of alignment rule 1. A is the reference benzene molecule with its three reference atom positions. B–E are examples of fitted molecules with their fitted atom positions. All these fitted molecules were superimposed on the reference molecule using the least-squares method.

CoMFA

CoMFA was performed using the QSAR mode of SYBYL Version 5.32. A field was calculated for benzene, the penetrant with the fastest flux in the data set, in a region where the dimensions ranged from -14.0 to $+14.0$ along the x and the y axes and from -10 to $+10$ along the z axis. Such a large region is used so that larger molecules can be fit in later work. The spacings were 2 Å along the axes and the probe atom possessed the van der Waals properties of an sp^3 carbon and a probe charge of $+1$. The region was also used for the field calculations of all the compounds of interest. The maximum energy for both the steric and the electrostatic fields was set at the default value of 30 kcal/mol. The remaining parameters were also set at the default values. Both

Table I. Classification of Functional Groups by Fragmental Values and Ring System Indicators

Functional group classification			Ring system classification	
Subgroup	Functional groups[a]	Fragmental value	Ring system	Compounds
1	Fluoro, alkyl, chloro	−0.15 to 0	1	Benzene derivatives
2	Ether, ester, ketone, aldehyde	−1.00 tp −0.28	2	Pyridine derivatives
3	Nitro	−0.95 (benzene nucleus) −0.99 (pyridine nucleus)	3	Quinoline, isoquinoline, pyrimidine, and indole derivatives and 2-quinoxalinol and acridine
4	Hydroxy, amino, carboxylic acid	−2.07 to −1.06		

[a] Arranged in order of their magnitudes.

the seric and the electrostatic fields for each conformer in the database were calculated and stored automatically into a QSAR table, which was organized so that each row corresponded to a molecular model in a Sybyl data base and a column corresponded to a steric or an electrostatic interaction energy.

Statistics

PLS and the crossvalidation method were used to analyze the CoMFA results. The minimum sigma value was set to 1 and the crossvalidation group was set to 10 for any PLS run combined with crossvalidation. For any final PLS run the minimum sigma value was set to zero. In order to improve the CoMFA results, the overall data set was divided into subgroups. A compound was placed in a subgroup based on the contribution of the substituent fragmental coefficient for calculating the steady state flux (18). The data set was classified into four subgroups as shown in Table I. The third subgroup was composed of compounds containing one or more nitro groups. These were classified into a separate subgroup because of apparent anomalous transport behaviors of these compounds. If a compound had multiple functionalities falling into more than one subgroup, it was placed into the subgroup with the highest absolute fragmental value. For example, 2-amino-4,6-dimethylpyridine was placed in subgroup number 4. Each of the subgroup indicators was multiplied by a factor of 100 in order to weight the field columns correctly.

In any PLS run combined with crossvalidation, the number of components was added until the best predictive model with the highest crossvalidated r was obtained. Then, a PLS run without crossvalidation was carried out, using the optimum number of components obtained from the best predictive model, in order to obtain the final model. This was used to generate fitted values, residual values for all compounds in the PLS data set, relative contributions of the predictors in the model, the conventional r^2 value, and the F value for the model. This final model was used to predict the target properties of the compounds which were not included in the PLS data set.

In order to test the predictive ability of each 3-D model, two sets of 8 compounds inside the data domain were randomly drawn from among the 102 benzene, pyridine, quinoline, and isoquinoline derivatives in the original data set. In each case, the test compounds were removed from the model and the model was recalculated prior to predicting the logarithm of steady-state flux, $\log J_{ssi}$, of the test compounds. Another set of 6 compounds considered to be outside of the data domain, because they were different nitrogen-containing ring systems than the 102 compounds above, was utilized to test the predictive ability of the models for compounds outside the data set.

Based on the CoMFA model, a $\log MFS_i$ term, representing the logarithm of solubility on the mole fraction scale for the diffusant in pure IPA, along with the steric and electrostatic fields was used to correlate the log steady-state flux. Each of the log MFS_i values was multiplied by a factor of 100 in order to weight the field columns correctly. The log MFS_i values for the 108 compounds are listed in Table II.

A f_{chex} term, representing the partition coefficient (23), was added to the above model and is also listed in Table II. The values of f_{chex} for all compounds in the data set were calculated based only on their substituted fragments. The ring systems were not included in the calculation in the original approach (15). Therefore, the benzene, pyridine, and quinoline derivatives will have the same values for the same functional groups regardless of the ring system. It is obvious that the benzene, pyridine, and quinoline ring systems are different in terms of partitioning ability and have different values of partition coefficient or f_{chex}. The same situation applies to their derivatives. In order to make sense of the f_{chex} values for any derivative in the data set, an indicator was assigned to the compounds based on their ring system (Table I). The ring systems outside the data domain were placed in the quinoline system because they have either a large ring size or two aromatic nitrogens and probably have lower f_{chex} values than either the benzene or the pyridine systems. Each of the ring system indicators and f_{chex} values was multiplied by a factor of 100 in order to weight the field columns correctly. The functional group classification technique was not applied in this analysis because of the use of the ring system indicators.

An intramolecular hydrogen bonding indicator was added to the final 3-D model. The compounds in the data set with the ability to intramolecularly hydrogen bond are listed in Table II. The intramolecular hydrogen bonding indicator was multiplied by a factor of 100 in order to correctly weight

Table II. Experimental Flux Data and Fitted Flux Data and Their Residuals for the 3-D Predictive Models

Obs.#	Compounds				CoMFA, Subgroup Indicators, $100*log\ MFS_i$		CoMFA, $100*log\ MFS_i$, Ring Indicators, $100*f_{chex}$		CoMFA, $100*log\ MFS_i$, Ring Indicators, $100*f_{chex}$, $100*IHB$	
		$log\ MFS_i$	f_{chex}	Expt. $logJ_{ssi}$	Cal.$logJ_{ssi}$	Residual	Cal.$logJ_{ssi}$	Residual	Cal .$logJ_{ssi}$	Residual
1	Benzene	0.0000	0.050	-0.2560	-1.000	0.744	-0.833	0.577	-0.575	0.319
2	Aniline	0.0000	-2.237	-1.7500	-1.883	0.133	-1.664	-0.086	-1.572	-0.178
3	Phenol	-0.0250	-3.119	-1.5700	-1.905	0.335	-1.824	0.254	-1.782	0.212
4	Benzoic acid	-0.6979	-3.030	-2.3155	-2.491	0.175	-2.520	0.205	-2.485	0.170
5	Nitrobenzene	0.0000	-0.687	-1.7200	-1.188	-0.532	-1.465	-0.255	-1.331	-0.389
6	Acetophenone	0.0000	-0.637	-1.6400	-1.431	-0.209	-1.575	-0.065	-1.429	-0.211
7	Benzaldehyde	0.0000	-1.129	-1.4800	-1.424	-0.056	-1.389	-0.091	-1.292	-0.188
8	Ethylbenzene	0.0000	1.069	-0.5550	-1.006	-0.451	-0.880	-0.325	-0.821	0.266
9	Chlorobenzene	0.0000	0.783	-0.5400	-1.007	0.467	-0.743	0.203	-0.532	-0.008
10	Toluene	0.0000	0.542	-0.3880	-1.016	0.628	-0.919	0.531	-0.760	0.372
11	Fluorobenzene	0.0000	0.252	-0.2560	-0.999	0.743	-0.789	0.533	-0.539	0.283
12	tert-Butylbenzene	0.0000	2.096	-0.7530	-1.000	0.247	-0.929	0.176	-0.852	0.099
13	3-Hydroxybenzoic acid	-0.8804	-5.989	-3.3086	-2.674	-0.634	-3.509	0.201	-3.608	0.300
14	3-Chlorotoluene	0.0000	1.485	-0.8370	-1.069	0.232	-0.679	-0.158	-0.537	-0.300
15	3-tert-Butylphenol	-0.1322	-0.863	-1.9000	-2.138	0.238	-1.755	-0.145	-1.874	-0.026
16	4-Hydroxybenzoic acid	-0.8948	-5.989	-3.5302	-2.679	-0.852	-3.535	0.004	-3.636	0.106
17	1-Fluoro-4-nitrobenzene	0.0000	-0.275	-1.6000	-1.199	-0.401	-1.369	-0.231	-1.220	-0.380
18	2-Hydroxy-5-nitropyridine	-2.2700	-3.810	-3.7471	-3.670	-0.077	-3.775	0.028	-3.757	0.010
19	Butylbenzene	0.0000	2.096	-0.8590	-0.992	0.133	-1.005	-0.146	-0.895	0.036
20	m-Nitrobenzaldehyde	-1.4763	-1.656	-2.4900	-2.487	-0.003	-2.777	0.287	-2.633	0.143
21	4-Aminoacetophenone	-1.2757	-2.714	-3.0400	-3.070	0.030	-2.982	-0.058	-2.910	-0.130
22	4-tert-Butylbenzoic acid	-0.9893	-0.774	-2.7588	-2.915	0.156	-2.735	-0.024	-2.738	-0.021
23	3-Hydroxypyridine	-0.8477	-3.100	-2.6849	-2.594	-0.091	-2.523	-0.162	-2.620	-0.065
24	3,5-Dichloropyridine	-0.9726	1.500	-1.4823	-1.849	0.367	-1.348	-0.134	-1.293	-0.190
25	4-tert-Butylpyridine	0.0000	2.090	-1.2272	-1.009	-0.218	-1.134	-0.094	-1.273	0.045
26	3-Aminopyridine	-0.2845	-2.237	-2.6822	-2.127	-0.555	-2.012	-0.670	-2.123	-0.559
27	Pyridine	0.0000	0.020	-0.6951	-0.996	0.301	-0.977	0.282	-0.906	0.211
28	2-Aminopyridine	-0.3099	-2.190	-1.8947	-2.136	0.241	-1.949	0.054	-2.016	0.121
29	2-Chloro-6-methoxypyridine	0.0000	0.740	-1.2112	-1.419	0.208	-1.072	-0.139	-1.305	0.094
30	2-Ethylpyridine	0.0000	1.070	-0.7179	-1.017	0.299	-1.112	0.394	-1.145	0.427
31	2-Chloropyridine	0.0000	0.760	-1.0808	-0.998	-0.083	-0.854	-0.226	-0.831	-0.250
32	2-Butoxypyridine	0.0000	1.550	-1.1554	-1.374	0.218	-1.263	0.107	-1.346	0.191
33	2-Fluoropyridine	0.0000	0.240	-0.8776	-0.990	0.113	-0.899	0.021	-0.838	-0.040
34	2-Methoxypyridine	0.0000	0.000	-0.8091	-1.405	0.595	-1.367	0.558	-1.336	0.527
35	2-Methoxy-5-nitropyridine	-2.1175	-0.710	-2.6525	-3.008	0.355	-3.060	0.408	-3.091	0.439
36	2-Methoxy-5-aminopyridine	0.0000	-2.210	-2.2300	-1.930	-0.300	-1.939	-0.291	-2.194	-0.036
37	2-Methyl-5-ethylpyridine	0.0000	1.590	-0.8684	-1.039	0.171	-0.997	0.128	-1.154	0.285
38	Phenetole	0.0000	-0.110	-1.1100	-1.402	0.292	-1.379	0.269	-1.177	0.067
39	2-Hydroxypyridine	-0.8165	-3.100	-2.4986	-2.556	0.057	-2.419	-0.080	-2.487	-0.012
40	2,4-Dihydroxypyridine	-2.4535	-6.220	-4.2887	-3.951	-0.338	-4.290	0.001	-4.450	0.161
41	2-Amino-4-methylpyridine	-0.7206	-1.670	-2.2276	-2.542	0.315	-2.158	-0.070	-2.210	-0.018
42	2-Amino-5-chloropyridine	-1.4023	-1.450	-2.6247	-3.076	0.451	-2.426	0.199	-2.402	-0.223
43	2-Amino-5-nitropyridine	-2.3820	-2.900	-3.7704	-3.766	-0.004	-3.651	-0.119	-3.623	-0.148
44	2,5-Pyridinedicarboxylic acid[a]	-3.2840	-6.260	-5.2045	-4.748	-0.457	-5.309	0.104	-5.079	-0.125
45	Quinoline	0.0000	0.020	-1.4903	-1.079	-0.412	-1.276	-0.214	-1.470	-0.020
46	6-Methoxyquinoline	0.0000	0.000	-2.0969	-1.603	-0.494	-2.074	-0.023	-1.984	-0.113
47	3-Quinolinecarboxylic acid	-2.8508	-3.120	-4.4101	-4.481	0.071	-4.529	0.119	-4.587	0.177
48	4,7-Dichloroquinoline	-1.5670	1.500	-2.3913	-2.494	0.103	-2.274	-0.117	-2.144	-0.247
49	6-Methylquinoline	0.0000	0.540	-1.7474	-1.208	-0.540	-1.702	-0.045	-1.663	-0.084
50	8-Nitroquinoline	-2.2366	-0.690	-3.3947	-3.141	-0.254	-3.196	-0.199	-3.467	0.072
51	8-Hydroxyquinoline[a]	-1.5171	-2.401	-2.3583	-3.234	0.876	-2.928	0.569	-2.598	0.240
52	8-Aminoquinoline[a]	-1.0701	-2.190	-2.2781	-2.869	0.591	-2.718	0.440	-2.451	0.173
53	5-Chloro-8-hydroxyquinoline[a]	-2.2660	-2.360	-3.1655	-3.887	0.722	-3.414	0.248	-3.042	-0.124
54	5-Nitro-8-hydroxyquinoline[a]	-3.1713	-3.810	-4.2195	-4.525	0.305	-4.625	0.406	-4.258	0.038
55	4-Methoxy-2-quinolinecarboxylic acid[a]	-3.0809	-3.140	-4.6171	-4.721	0.104	-4.840	0.223	-4.391	-0.226
56	6-Quinolinecarboxylic acid	-3.0223	-3.120	-4.6724	-4.625	-0.047	-4.612	-0.060	-4.771	0.099
57	2-Hydroxy-4-methylquinoline	-2.4547	-2.580	-3.8755	-4.175	0.299	-4.118	0.242	-4.121	0.246
58	6-Aminoquinoline	-1.0013	-2.190	-3.0606	-2.918	-0.143	-3.010	-0.051	-2.960	-0.101
59	3-Aminoquinoline	-0.6925	-2.190	-2.9338	-2.663	-0.271	-2.788	-0.145	-2.837	-0.097
60	2-Hydroxyquinoline	-2.2111	-3.100	-3.8125	-3.931	0.118	-3.906	0.094	-3.800	-0.013
61	4-Hydroxyquinoline	-1.5317	-3.100	-3.6878	-3.258	-0.430	-3.188	-0.500	-3.541	-0.147
62	6-Nitroquinoline	-2.5935	-0.690	-3.6146	-3.546	-0.069	-3.728	0.113	-3.693	0.078
63	8-Quinolinecarboxylic acid[a]	-2.9393	-3.120	-4.2129	-4.444	0.231	-4.240	0.027	-4.001	-0.211
64	4-Quinolinecarboxylic acid	-2.8268	-3.120	-4.5178	-4.325	-0.193	-4.236	-0.282	-4.555	0.037
65	Butylphenylether	0.0000	1.559	-1.2500	-0.973	-0.277	-1.059	-0.191	-0.914	-0.336
66	Anisole	0.0000	-0.022	-1.0300	-1.402	0.372	-1.295	0.265	-1.067	0.037
67	m-Xylene	0.0000	1.244	-0.5800	-1.077	0.497	-0.894	0.314	-0.776	0.196
68	Methylbenzoate	0.0000	0.112	-1.4600	-1.427	-0.033	-1.352	-0.108	-1.435	-0.025
69	6-Methoxy-8-nitroquinoline	-3.1675	-0.710	-4.3323	-3.976	-0.356	-3.965	-0.367	-4.436	0.103

Table II. Continued

Obs.#	Compounds				CoMFA, Subgroup Indicators, $100*log\ MFS_i$		CoMFA, $100*log\ MFS_i$, Ring Indicators, $100*f_{chex}$		CoMFA, $100*log\ MFS_i$, Ring Indicators, $100*f_{chex}$, $100*IHB$	
		$log\ MFS_i$	f_{chex}	Expt. $logJ_{ssi}$	Cal.$logJ_{ssi}$	Residual	Cal.$logJ_{ssi}$	Residual	Cal .$logJ_{ssi}$	Residual
70	Propiophenone	0.0000	-0.110	-1.6300	-1.427	-0.203	-1.522	-0.108	-1.561	-0.069
71	m-Anisaldehyde	0.0000	-0.991	-2.0900	-1.469	-0.621	-1.564	-0.526	-1.849	-0.241
72	Methyl-3-methylbenzoate	0.0000	0.814	-1.4300	-1.491	0.061	-1.294	-0.136	-1.417	-0.013
73	Ethylparaben	-0.6983	-2.320	-2.6900	-2.651	-0.039	-2.571	-0.119	-2.471	-0.219
74	3-Pyridinecaboxaldehyde	0.0000	-0.850	-1.8232	-1.425	-0.398	-1.527	-0.297	-1.772	-0.051
75	3,5-Lutidine	0.0000	1.060	-0.9487	-1.073	0.125	-1.083	0.134	-1.145	0.196
76	5-Chloro-3-pyridinol	-1.0605	-2.360	-2.6211	-2.808	0.187	-2.498	-0.123	-2.577	-0.045
77	Nicotinic acid	-2.3830	-3.120	-3.7595	-3.897	0.138	-3.640	-0.120	-3.905	0.145
78	4-Picoline	0.0000	0.540	-0.8447	-1.019	0.174	-1.139	0.294	-1.167	0.323
79	3-Acetylpyridine	0.0000	-0.610	-1.9918	-1.416	-0.576	-1.555	-0.436	-1.909	-0.083
80	6-Hydroxynicotinic acid	-3.2757	-6.240	-5.1057	-4.686	-0.420	-5.173	0.067	-5.451	0.345
81	Picolinic acid[a]	-1.7595	-3.120	-3.2816	-3.368	0.087	-3.224	-0.058	-2.903	-0.379
82	6-Choronicotinic acid	-1.5935	-2.380	-3.0980	-3.263	0.165	-3.023	-0.075	-3.290	0.192
83	Ethylnicotinate	0.0000	0.330	-1.5296	-1.387	0.143	-1.390	-0.139	-1.414	-0.116
84	Lepidine	0.0000	0.540	-1.8529	-1.106	-0.747	-1.423	-0.429	-1.733	-0.120
85	8-Hydroxyquinaldine[a]	-1.4486	-2.417	-2.3752	-3.170	0.795	-2.848	0.473	-2.576	0.201
86	2-Chlorolepidine	-0.9666	1.280	-2.2996	-2.055	-0.245	-2.127	-0.173	-2.055	-0.244
87	8-Nitroquinaldine	-2.6840	-0.170	-3.8269	-3.519	-0.308	-3.257	-0.570	-3.644	-0.183
88	4-Aminoquinaldine	-1.1561	-1.670	-3.4809	-2.977	-0.504	-2.822	-0.658	-3.259	-0.222
89	6-Methoxyquinaldine	-0.5146	0.520	-2.2467	-2.042	-0.205	-2.294	0.047	-2.184	-0.062
90	Quinaldine	0.0000	0.540	-1.6215	-1.211	-0.410	-1.687	0.065	-1.598	-0.023
91	2,4-Quinolinediol	-3.3566	-6.220	-5.4693	-4.908	-0.562	-5.511	0.042	-5.570	0.101
92	Isoquinoline	0.0000	0.050	-1.6773	-1.191	-0.487	-1.578	-0.099	-1.459	-0.219
93	2-Methyl-5-butylpyridine	0.0000	1.950	-1.1127	-1.017	-0.096	-1.252	0.139	-1.075	-0.038
94	4-Methylpyrimidine	0.0000	0.542	-1.0218	-1.010	-0.012	-1.250	0.228	-1.240	0.218
95	2-Amino-4,6-dimethylpyridine	-0.3534	-1.150	-2.2527	-2.272	0.020	-2.011	-0.242	-2.188	-0.064
96	1-Isoquinolinecarboxylic acid[a]	-2.6861	-3.030	-4.1319	-4.331	0.199	-4.166	0.034	-3.739	-0.393
97	2,4-Dimethyl-6-hydroxypyrimidine	-1.6716	-1.715	-3.3002	-3.380	0.080	-3.069	-0.231	-3.209	-0.091
98	7-Nitroindole[a]	-2.5768	-0.687	-2.6590	-3.401	0.742	-3.290	0.631	-2.961	0.302
99	Acridine	-1.3091	0.050	-2.6381	-2.219	-0.419	-2.296	-0.342	-2.635	-0.003
100	2-Quinoxalinol	-2.8665	-3.119	-4.1639	-4.478	0.314	-4.247	0.083	-4.146	-0.018
101	Indole	-0.2138	0.050	-1.8463	-1.275	-0.571	-1.491	-0.355	-1.674	-0.173
102	4-Chlorotoluene	0.0000	1.485	-0.6940	-1.053	0.359	-0.679	-0.015	-0.447	-0.247
103	2-Quinolinecarboxylic acid[a]	-2.0809	-4.220	-3.5523	-3.837	0.285	-4.318	0.766	-3.891	0.339
104	6-Isopropylquinoline	0.0000	1.480	-1.8972	-1.209	-0.688	-1.870	-0.027	-1.914	0.017
105	5-Aminoquinoline	-1.0867	-2.190	-3.1130	-2.852	-0.261	-2.637	-0.476	-2.945	-0.168
106	2,6-Dimethoxypyridine	0.0000	0.090	-1.1287	-1.457	0.328	-1.246	0.118	-1.519	0.390
107	5-Nitroquinoline	-1.4647	-0.690	-2.8620	-2.458	-0.404	-2.694	-0.168	-2.971	0.109
108	Methylparaben	-0.7376	-2.687	-2.7400	-2.636	-0.104	-2.579	-0.161	-2.498	-0.242

[a] Compounds able to intramolecularly hydrogen bond

the factor. In order to validate the final model further, a bootstrapping validation method was used in addition to the crossvalidation. Ten runs were chosen for the bootstrapping. In general, the greater the number of runs the better. However, little is lost by drawing 10 bootstrap samples instead of 100 (2).

RESULTS AND DISCUSSION

CoMFA Studies

The alignment rule used in the CoMFA analysis was based on the strategy that the first consideration for the alignment was to emphasize that functional group which had the greatest effect on flux. The second consideration was to try to arrange the volume increment in the same direction. In this way the functional group having the greatest effect on flux for each compound in the data set can be placed in the same position for comparisons of both Coulombic and van der Waals interactions.

The crossvalidated r^2 is defined analogously to the definition of the conventional r^2 except that it is indicative not of the goodness of fit but of the actual predictive performance of the model. Table III lists the results of the CoMFA studies. The crossvalidated r^2 value for the straight CoMFA model is 0.484, which is about halfway between no model and a perfect model for prediction. Usually CoMFA is designed to analyze one class of compounds with similar structure. A crossvalidated r^2 of about 0.5 from a CoMFA analysis for a data set consisting of several classes of compounds was encouraging enough to attempt further improvements.

In order to increase the comparability of the molecular fields for all the conformers of interest in the CoMFA studies, a classification by functional group was used to subdivide the data set. The results from the CoMFA studies combined with the technique of functional group classification method are listed in Table III. The crossvalidated r^2 was increased to 0.772. The great increase in the predictive ability of combining CoMFA with the functional group classification implies linkage between this 3-D QSTR approach and conventional QSAR approaches such as Hansch's fragmen-

Table III. Comparison of the 3-D Predictive Models

	Independent values				
	CoMFA	CoMFA, subgroup indicators	CoMFA, subgroup indicators, 100 * logMFS_i	CoMFA, ring indicators, 100 * logMFS_i, 100 * f_{chex}	CoMFA, ring indicators, 100 * logMFS_i, 100 * f_{chex}, 100 * *IHB*
PLS with crossvalid. (group = 10)					
Crossvalidated r^2	0.484	0.772	0.896	0.926	0.951
Optimum No. of components	6	8	2	4	6
SE of prediction	0.945	0.628	0.419	0.353	0.290
Bootstrap (runs = 10)					
Bootstrapped r^2	—	—	—	—	0.980
SD for bootstrap r^2	—	—	—	—	0.004
Mean SE of estimate	—	—	—	—	0.183
SD for mean SE of estimate	—	—	—	—	0.066
Conventional r^2	0.740	0.914	0.907	0.950	0.973
SE of estimate	0.662	0.385	0.389	0.287	0.213
F value	47.913	130.734	510.143	488.907	605.879
n_1	6	8	2	4	6
n_2	101	99	105	103	101

tal approach and various linear free energy relationship approaches. The classification according to functional groups implies that each classified subgroup has some relationship to their various physicochemical properties such as hydrophobicity and, also on that basis, some relationship to the magnitude of the fragmental coefficient for prediction of flux. It is certainly possible that even better classification methods may exist for the data set.

Outliers from the analysis are 2,4-quinolinediol, 2,4-dihydroxypyridine, 6-hydroxynicotinic acid, 5-aminoquinoline, phenol, and 2-quinolinecarboxylic acid. The first four have negative residual values, indicating that the calculated flux is faster than the experimental flux, while the others have positive residuals, indicating that the calculated flux is slower than the experimental flux. The reasons for these are not easily understood but may be caused by selection of the conformers, alignment rules, or anomalies in the transport behavior of these compounds. The relative contributions from the normalized coefficients of each independent variable in the CoMFA model combined with the functional group classification technique are listed in Table IV.

Solubility Considerations

Thermodynamically, both a neat liquid and a saturated solution of a solid have an activity of one, but the actual number of molecules of interest in contact with the membrane is vastly different in the two situations. Solubility has been shown to have a great effect on membrane diffusion

Table IV. Relative Contributions of Normalized Coefficients of Independent Variables for the 3-D Predictive Models

	Predictive model							
	CoMFA, subgroup indicators		CoMFA, Subgroup indicators, 100 * logMFS_i		CoMFA, ring indicators, 100 * logMFS_i, 100 * f_{chex}		CoMFA, ring indicators, 100 * MFS_i, 100 * f_{chex} 100 * *IHB*	
	Norm. coeff.	Fraction	Norm. coeff.	Fraction	Norm. coeff.	Fraction	Norm. coeff.	Fraction
CoMFA (864 variable) (steric)	2.704	0.586	0.113	0.090	0.524	0.299	0.769	0.331
CoMFA (864 variable) (electrostatic)	1.113	0.241	0.059	0.047	0.118	0.067	0.187	0.081
100 * Subgroup 1	0.245	0.053	0.147	0.117				
100 * Subgroup 2	0.091	0.020	0.000	0.000				
100 * Subgroup 3	0.024	0.005	0.016	0.013				
100 * Subgroup 4	0.433	0.094	0.188	0.150				
100 * logMFS_i			0.729	0.582	0.509	0.291	0.463	0.199
100 * f_{chex}					0.469	0.268	0.543	0.234
100 * Ring System 1					0.061	0.035	0.102	0.044
100 * Ring System 2					0.004	0.002	0.023	0.010
100 * Ring System 3					0.068	0.039	0.085	0.036
IHB							0.152	0.065

(12,18). For example, the solubility of 2-quinolinecarboxylic acid in isopropanol is 18.44 mg/mL ($\log J_{ss} = -3.5523$), while that of 4-quinolinecarboxylic acid is 3.32 mg/mL ($\log J_{ss} = -4.5178$); the solubility of 5-nitroquinoline is 78.65 mg/mL ($\log J_{ss} = -2.8620$), while that of 6-nitroquinoline is 5.69 mg/mL ($\log J_{ss} = -3.6146$). The solubilities for the latter two isomers differ by a factor of 14. However, CoMFA analysis assumes that all the compounds have the same concentration. Therefore the predictor variable, $\log MFS_i$, was added to the 3-D CoMFA QSTR model.

A comparison of the results of the CoMFA model and those of the CoMFA model plus the solubility term is listed in Table III. After the addition of the solubility term to the CoMFA model, the crossvalidated r^2 increased from 0.772 to 0.896, indicating a substantial improvement in predictive ability. The standard error of prediction or PRESS decreased from 0.628 to 0.419, indicating improvement of the predictive performance. In general, the more components used in the PLS, the lower the robustness is for the predictive model. The optimum number of components dropped from eight to two after adding the solubility term, indicating an increase in the robustness of model. The F value increased from 130.743 to 510.143 and also demonstrates a more sound regression model. The value of a conventional r^2 resulting from a PLS run varies according to the number of components used. Adding more components to a PLS run increases the conventional r^2 in the same manner as adding more variables to an ordinary regression. The conventional r^2 here corresponds to the optimum number of components resulting from the PLS run combined with the crossvalidation. It represents the goodness of fit only for the PLS model corresponding to the number of components chosen. Conventional r^2 values from two PLS models have little compatibility if they result from a different number of components. The same situation is applicable for the standard error of estimate or standard deviation.

The experimental and calculated $\log J_{ssi}$ and their residuals using the CoMFA model combined with the functional group classification technique (subgroup indicators) and the solubility term are listed in Table II. Three possible outliers emerge from the use of this model including 4-hydroxybenzoic acid, 8-hydroxyquinoline, and 8-hydroxyquinaldine. Among these three compounds, the first has a negative residual value, while the others have positive residuals. Two phenomena which may be responsible for this need to be taken into account. The first is that intramolecular hydrogen bonding increases the hydrophobicity of the compound, thereby increasing its ability to partition into the PDMS membrane and increasing its flux. 8-Hydroxyquinoline and 8-hydroxyquinaldine can both intramolecularly hydrogen bond and the underestimation of their flux can be explained on this basis. The second phenomenon is intermolecular hydrogen bonding or dimerization of the diffusant. This behavior increases the molecular volume providing a decrease in flux, but also increases the lipophilicity of the diffusant resulting in an increase in flux. The final result is dependent on which mechanism dominates the process. The behavior of 4-hydroxybenzoic acid may be explained by this process. The fact that these compounds are outliers may also be caused by selection of the conformers, alignment rules, or anomalies in the transport behavior of these compounds.

The results of the predictive tests for the random data sets both inside and outside the data domain from the model of CoMFA combined with the functional group classification technique and the solubility term are given in Table V. Only 3 of the 16 tested compounds inside the data domain have a residual higher than 0.4 log unit. Two of these three compounds, namely, 8-aminoquinoline and 5-chloro-8-hydroxyquinoline, with the highest residuals have the ability to intramolecularly hydrogen bond. Eleven of the sixteen have a residual lower than 0.3 log unit. The predictions are better for the compounds inside the data domain than for those outside the data domain.

The relative contributions from the normalized coefficients of each independent variable in the CoMFA model combined with the functional group classification technique and the solubility term are listed in Table IV.

It may appear that the reduction of optimal number of components from eight to two after addition of the solubility term indicates that the information contained in the subgroup indicators may be redundant and more uniquely described only by the solubility term. To determine if this was the case, a PLS run combined with crossvalidation was performed using only the CoMFA descriptors and the solubility term, without the subgroup indicators. The results show that the crossvalidated r^2 decreased from 0.896 to 0.835, the standard error of predictions increased from 0.419 to 0.518, and the optimal number of components remained 2. The comparison of these results demonstrates that the inclusion of the functional group classification is still significant.

Partition Coefficient Considerations

Membrane diffusion theory indicates that the partition coefficient is an important parameter in determining the magnitude of the steady-state flux (12,13). Instead of using the partition coefficient between the PDMS membrane and IPA itself, various hydrophobic parameters which estimate partition coefficient were used in the correlation analysis. This was done both because of the additivity of these parameters and because experimental determination was not required.

There are many additive, hydrophobic parameters available, but the π and f parameters are used most frequently in LFER studies (14,15). Both π and f values are based on an octanol/water solvent system. The major difference between these hydrophobic parameters is the fragmental constant method assigns an f value for hydrogen, while the π system ignores the contribution of the hydrogen atom. Both parameters were used for pyridine and quinoline derivatives in the previous LFER studies (16). Correlations of the steady-state flux for both the f and the π parameters were poor predictors, although the fragmental parameter f was slightly better. This result suggested that the octanol/water solvent system did not properly describe the PDMS/isopropanol partitioning system. The major reason may be that PDMS possesses little hydrogen bonding ability, while octanol can form hydrogen bonds with many penetrants. Hence, the f and π parameters were transformed into their corresponding cyclohexane/water parameters (f_{chex} and π_{chex}) using Seiler's I_h variable and conversion equation (23). Both f_{chex} π_{chex} correlated much better with flux, with the f_{chex} being somewhat better.

Table V. Prediction of PDMS Membrane Diffusion Rates for Compounds Inside and Outside the Data Domain Using the 3-D Predictive Models

Compounds		Predictive Models					
		CoMFA, Subgroup Indicators, 100*log MFS_i		CoMFA, Ring Indicators, 100*log MFS_i, 100*f_{chex}		CoMFA, Ring Indicators, 100*log MFS_i, 100*f_{chex}, 100*IHB	
Inside Data Domain:	Expt. log J_{ssi}	Pred. logJ_{ssi}	Residual	Pred. logJ_{ssi}	Residual	Pred. logJ_{ssi}	Residual
1st Random Data Set:							
Benzaldehyde	-1.4800	-1.3902	-0.0898	-1.3572	-0.1228	-1.2624	-0.2176
4-tert-Butylbenzoic acid	-2.7588	-2.9070	0.1482	-2.5674	-0.1914	-2.6464	-0.1124
Fluoropyridine	-0.8776	-0.9998	0.1222	-0.9068	0.0292	-0.8071	-0.0705
6-Methoxyquinoline	-2.0969	-1.5557	-0.5412	-2.1244	0.0275	-1.9753	-0.1216
6-Nitroquinoline	-3.6146	-3.5578	-0.0568	-3.7193	0.1047	-3.7289	0.1143
8-Quinolinecarboxylic acid	-4.2129	-4.4556	0.2427	-4.2369	0.0240	-3.9718	-0.2411
Propiophenone	-1.6300	-1.3823	-0.2500	-1.4715	-0.1585	-1.5282	-0.1018
Nicotinic acid	-3.7597	-3.9069	0.1472	-3.6140	-0.1457	-3.9170	0.1573
2nd Random Data Set:							
Benzoic Acid	-2.3155	-2.5011	0.1856	-2.5303	0.2148	-2.5689	0.2534
2-Hydroxy-5-nitropyridine	-3.7471	-3.6649	-0.0822	-3.7297	-0.0174	-3.7105	-0.0366
Phenetole	-1.1100	-1.3835	-0.2735	-1.3756	0.2656	-1.2591	0.1491
2-Amino-4-methylpyridine	-2.2276	-2.5664	0.3388	-2.1341	-0.0935	-2.1575	-0.0701
Quinoline	-1.4903	-1.1050	-0.3853	-1.2850	-0.2053	-1.5987	0.1084
8-Aminoquinoline	-2.2781	-2.9028	0.6247	-2.7771	0.4990	-2.5524	0.2743
5-Chloro-8-hydroxyquinoline	-3.1655	-3.9372	0.7717	-3.4510	0.2855	-3.0757	0.0898
6-Methoxyqunaldine	-2.2467	-2.0059	-0.2408	-2.2461	-0.0004	-2.2665	0.0198
Outside Data Domain:							
4-Methyl-pyrimidine	-1.0218	-0.9639	-0.0579	-1.3003	0.2785	-1.4109	0.3891
2,4-Dimethyl-6-hydroxypyrimidine	-3.3002	-3.3805	0.0803	-3.0479	-0.2523	-3.2372	-0.0630
7-Nitroindole	-2.6590	-3.5405	0.8815	-3.6225	0.9635	-3.2476	0.5886
Acridine	-2.6381	-2.1711	-0.4670	-2.2071	-0.4310	-2.5409	-0.0972
2-Quinoxalinol	-4.1639	-4.4901	0.3262	-4.2039	0.0400	-4.1862	0.0223
Indole	-1.8463	-1.2149	-0.6314	-1.4524	-0.3939	-1.7147	-0.1316

Therefore f_{chex} was used as a predictor and added to the QSAR model.

Comparison of the results of the CoMFA model; of the CoMFA model plus the solubility term, and of the CoMFA model combined with both the solubility and the f_{chex} terms is listed in Table III. After the addition of the f_{chex} term, the crossvalidated r^2 increases from 0.896 and 0.926. This has greatly improved the predictive ability compared to the previous 3-D predictive model involving only the CoMFA and solubility terms. The decrease in the PRESS or predictive error from 0.419 to 0.353 also shows an obvious improvement in the predictive performance. The optimum number of components increases from two to four and the F value decreases slightly, from 510.143 to 488.907. These, however, do not affect the robustness of the PLS model because an optimum number of components of four is not large (8) and an F value of 490 is still very large. The conventional r^2 is 0.950 and the standard error of the estimate is 0.287, indicating excellent goodness of fit for this 3-D predictive model.

The experimental and calculated logJ_{ssi} and their residuals from the 3-D model using CoMFA, logMFS_i, and f_{chex} (combined with the ring system indicators) are listed in Table II. There are five possible outliers from the use of this model including 3-aminopyridine, 4-aminoquinaldine, benzene, 7-nitroindole, and 2-quinolinecarboxylic acid. Among these five possible outliers, the amino compounds have negative residual values and the last three have positive deviations. The reasons for the first three outliers in this model are not clear yet. The deviations of 7-nitroindole and 2-quinolinecarboxylic acid may be caused by their intramolecular hydrogen bonding behavior.

The results of the predictive tests for this model are given in Table V. Only one compound has a residual higher than 0.3 log unit. Eleven of the sixteen have a residual lower than 0.2 log unit. The predictions were greatly improved for the data inside the data domain compared with the previous 3-D model. The highest residuals are still borne by the two compounds with the ability to intramolecularly hydrogen bond. The predictions for the data outside the domain were also improved overall. Some, however, became worse. The predictive residual for 7-nitroindole, for example, increased about 0.1 log unit. But the fact that five of the six compounds have a residual less than 0.5 log unit and four of them have a residual less than 0.4 log unit indicates good predictive ability for the data outside the data domain.

The relative contributions from the normalized coefficients of each independent variable in this 3-D model including CoMFA, logMFS_i, and f_{chex} (combined with the ring system indicators) are listed in Table IV.

Hydrogen Bonding Considerations

The ability of a compound to participate in intramolecular hydrogen bonding has been shown to produce a signif-

icant increase in membrane flux because it increases the lipophilicity of the molecule. Many compounds, such as 8-hydroxyquinoline, 8-hydroxyquinaldine, 2-quinolinecarboxylic acid, and 8-aminoquinoline, with the ability to intramolecularly hydrogen bond diffused significantly faster than their predicted flux (12). It was also found in the previous 3-D models that some outliers were compounds capable of intramolecular hydrogen bonding such as 7-nitroindole. Therefore, an indicator variable (*IHB*) for intramolecular hydrogen bonding was assigned to the compounds with this capability and added to the previous 3-D model.

The results for the final 3-D predictive model using CoMFA, log mole fraction solubility, f_{chex} (combined with the ring system indicators), and intramolecular hydrogen bonding terms are listed in Table III. After the addition of the *IHB* indicator, the crossvalidated r^2 increases from 0.926 to 0.951 and the PRESS or predictive error drops from 0.353 to 0.290. The conventional r^2 is 0.973 and the standard error of the estimate is 0.213, indicating an excellent goodness of fit for the 3-D predictive model. The optimum number of components as a result of the crossvalidation run is six, which does not affect the robustness of the model considering the number of predictors used. Bootstrapping is a safeguard against the promulgation of an artifactual, nonpredictive, QSAR equation. The results of the bootstrap sampling demonstrate a strong support of the final PLS model. The bootstrapped r^2 is 0.980, which is slightly higher than the conventional r^2. The standard deviation of the bootstrapped r^2 or the standard deviation of conventional r^2's in 10 random drawings is 0.004. The mean of the standard error of estimate in the 10 bootstrap samplings is 0.183, which is lower than the 0.213, the standard error of the estimate for the PLS equation. The standard deviation of the standard errors of estimate generated in 10 bootstrapping runs is 0.066, which is also very low. The difference between the statistical parameters calculated from the original data set and the average of the parameters calculated from the many bootstrap samplings is a measure of the bias of the original calculation. Therefore, all these verify a good correlation of the log steady-state flux with the CoMFA fields and other selected physicochemical terms and no artifactual PLS-overlooked correlation.

The experimental and calculated $\log J_{ssi}$ and their residuals from this 3-D model using CoMFA, $\log MFS_i, f_{chex}$ (combined with the ring system indicators), and *IHB* are listed in Table II. A plot of experimental $\log J_{ssi}$ versus fitted $\log J_{ssi}$ from this 3-D model is provided in Fig. 2. Four possible outliers emerge from the use of this model including 3-aminopyridine, 2-ethylpyridine, 2-methoxypyridine, and 2-methoxy-5-nitropyridine. Among these four possible outliers, the first has a negative residual value while the last three compounds have positive residuals. After adding the intramolecular hydrogen bonding indicator, all the outliers in the previous model are removed except for 3-aminopyridine, which does not intramolecularly hydrogen bond. The residual values for the three new outliers (2-ethylpyridine, 2-methoxypyridine, and 2-methoxy-5-nitropyridine) in the final model are relatively small, ranging from 0.43 to 0.53. The presence of an ortho-substituted lipophilic group (referring the substituted position relative to the heteroaromatic nitrogen) is a common characteristic of these compounds. The

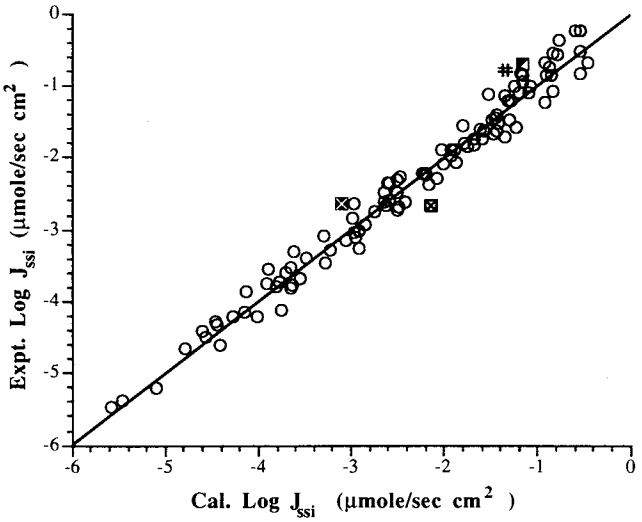

Fig. 2. Relationship between observed and calculated $\log J_{ssi}$ of the entire data set for the final 3-D predictive model involving CoMFA, $\log MFS_i, f_{chex}$ (combined with the ring system indicators), and *IHB*. Possible outliers: (⊠) 3-aminopyridine (obs. 26); (#) 2-methoxypyridine (obs. 34); (⊠) 2-methoxy-5-nitropyridine (obs. 35); (◥) 8-hydroxyquinaldine (obs. 88).

steric effect of the ortho-substituted group on the hetero nitrogen may decrease the effect of the negative partial charge of the aromatic nitrogen, which may, in turn, increase the membrane flux.

The results of the predictive tests for this model are given in Table V. The highest predictive residual for compounds inside the data domain is 0.2743. Twelve of the sixteen have a residual lower than 0.2 log unit. After adding the intramolecular hydrogen bonding indicator, the predictions for the compounds with the ability to intramolecularly hydrogen bond have been greatly improved. The predictions for the data outside the domain were also much improved, especially for 7-nitroindole, for which the predictive residual decreased from 0.9635 to 0.5886. The predictive residuals of 4-methylpyrimidine and indole are 0.381 and −0.1316, respectively. The predictive residuals for remaining compounds are less than 0.1 log unit. Thus, the final model demonstrates an excellent predictive ability for the compounds outside the data domain.

The relative contributions from the normalized coefficients of each independent variable in the final 3-D model are listed in Table IV. After adding the *IHB* indicator to the predictive model, the relative importance of the CoMFA steric and electrostatic fields still remains. Their contributions to the model total more than 45%, with the steric field contributing 33.1% and the electrostatic field 8.1%. The contribution of the intramolecular hydrogen bonding indicator contributes only 6.5%; however, it clarifies the classification of the data set and plays an important role in increasing the predictive ability of compounds which possess this ability.

CONCLUSION

CoMFA combined with the use of statistical methods such as PLS shows promise in the development of a new 3-D QSTR approach for prediction of membrane flux. The use of

a functional group classification technique can increase the predictive performance of CoMFA. This is especially important for this investigation, in which several classes of compounds were involved in the CoMFA studies.

The combination of CoMFA with the solubility and the partition coefficient descriptors has greatly improved the predictive ability over the CoMFA models without these physicochemical parameters. This implies that considerations of experimental conditions and suitable parameters describing the membrane diffusion process are necessary in 3-D QSTR development. The improvement of both predictive performance and goodness of fit after the addition of an intramolecular hydrogen bonding term indicates that some compensation of molecular structural description for CoMFA molecular fields is needed according to the results of the PLS analysis. The excellent predictions of membrane flux for the compounds both inside and outside the data domain demonstrate that the 3-D predictive model involving CoMFA steric and electrostatic fields, solubility and partition coefficient terms, and an intramolecular hydrogen bonding indicator is a promising approach for the prediction of membrane flux.

ACKNOWLEDGMENTS

The flux data of heteroaromatic compounds were generated by M. W. Hu and those of aromatic compounds were generated by P. Vayumhasuwan and D. Moeckly. Financial support from the Graduate College, University of Iowa, for the molecular modeling studies is gratefully acknowledged. Part of this paper was presented at the 1990 AAPS annual meeting and the Graduate Symposium in Drug Delivery and Pharmaceutical Technology at the 1991 AAPS annual meeting.

Coordinates of all the molecules in this study and other input data will be provided by the authors upon request.

REFERENCES

1. S. Borman. New QSAR techniques eyed for environmental assessment. *C&EN* 68:20–23 (1990).
2. *Theory Manual of SYBYL, Molecular Modeling Software, Version 5.4,* Tripos Associates, Inc., St. Louis, MO, 1991.
3. K. F. Koehler. 3D QSAR analysis of the benzodiazepine inverse agonist receptor. Reprint of presentation at 1990 Tripos Users Guide Meeting, May 1990, St. Louis, MO.
4. R. D. Cramer III and M. Miline. The lattice model: A general paradigm assisted drug design. Abstracts of American Chemical Society, J. L. Fauchère (ed.), COMP 44, 1979.
5. R. D. Cramer III, D. E. Paterson, and J. D. Bunce. Recent advances in comparative molecular field analysis (CoMFA). In J. L. Fauchère (ed.), *QSAR: Quantitative Structure-Activity Relationships in Drug Design, Proceedings of the 7th European Symposium on QSAR,* Alan R. Liss, New York, 1988, pp. 161–165.
6. R. D. Cramer III, J. D. Bunce, D. E. Paterson, and I. E. Frank. Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quant. Struct.-Act. Relat.* 7:18–25 (1988).
7. R. D. Cramer III, D. E. Paterson, and J. D. Bunce. Comparative molecular field analysis (CoMFA). I. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* 110:5959–5967 (1988).
8. M. Clark, R. D. Cramer III, D. M. Jones, D. E. Patterson, and P. E. Simeroth. Comparative molecular field analysis (CoMFA). 2. Toward its use with 3D-structural databases. *Tetrahedron Comp. Method.* 3:47–59 (1990).
9. M. S. Allen, K. F. Koeher, *et al.* Synthetic and computer-assisted analyses of the pharmacophore for the benzodiazepine receptor inverse agonist site. *J. Med. Chem.* 33:2343–2357 (1990).
10. D. W. Osborne and A. H. Amann. *Topical Drug Delivery Formulations,* Marcel Dekker, New York, 1990.
11. M. H. Jacobs. *Diffusion Processes.* Springer-Verlag, New York, 1967.
12. H. A. Daynes. The process of diffusion through a rubber membrane. *Proc. Roy. Soc. Lond.* A97:286 (1920).
13. R. M. Barrer. Permeation, diffusion and solution of gases in organic polymers. *Trans. Faraday Soc.* 35:628 (1939).
14. A. Leo, C. Hansch, and D. Elkins. Partition coefficients and their uses. *Chem. Rev.* 71:525 (1971).
15. R. F. Rekker. *The Hydrophobic Fragmental Constant,* Elsevier, New York, 1977.
16. M. Hu. PH.D. thesis, University of Iowa, Iowa City, 1990.
17. D. M. Moeckly and L. E. Matheson. The development of a predictive method for the estimation of diffusion rates through polydimethylsiloxane membranes. I. Identification of critical variables for a series of substituted benzenes. *Int. J. Pharm.* 77:151–162 (1991).
18. M. Hu and L. E. Matheson. The development of a predictive method for the estimation of diffusion rates through polydimethylsiloxane membranes. III. Application to a series of substituted pyridines. *Pharm. Res.* 10:732–736 (1993).
19. L. E. Matheson and M. Hu. The development of a predictive method for the estimation of diffusion rates through polydimethylsiloxane membranes. IV. Application to a series of substituted quinolines. *Pharm. Res.* 10:839–842 (1993).
20. L. E. Matheson, P. Vayumhasuwan, and D. M. Moeckly. The development of a predictive method for the prediction of flux through polydimethylsiloxane membranes. II. Derivation of a diffusion parameter and its application to multisubstituted benzenes. *Int. J. Pharm.* 77:163–168 (1991).
21. P. Laoratthaphong, Ph.D. thesis, University of Iowa, Iowa City, 1989.
22. M. Clark, R. D. Cramer III, and N. V. Opdenbosch. Validation of the general purpose Tripos 5.2 force field. *J. Comp. Chem.* 10(8):982–1012 (1989).
23. P. Seiler. Interconversion of lipophilicities from hydrocarbon/water systems into the octanol/water system. *Eur. J. Med. Chem.* 9:473 (1974).