Prediction of Solvation Free Energies of Small Organic Molecules: Additive-Constitutive Models Based on Molecular Fingerprints and Atomic Constants

Vellarkad N. Viswanadhan,*,† Arup K. Ghose,*,† U. Chandra Singh,‡ and John J. Wendoloski†

Department of Molecular Structure and Design, One Amgen Center Drive, Mail Stop 14-2-B, Amgen, Inc., Thousand Oaks, California 91320, and AM Technologies Inc., 130 West Rhapsody, San Antonio, Texas 78216

Received October 10, 1998

Solvation free energy is an important molecular characteristic useful in drug discovery because it represents the desolvation cost of a ligand binding to a receptor. Most of the recent developments in the estimation of solvation free energy require the use of molecular mechanics and dynamics calculations. Group contribution methods have been rarely used in the past for calculating solvation free energy because automated prediction methods have not been developed in this regard. As an aid to combinatorial library design, we explored rapid and accurate means of computing solvation free energies from the covalent structures of organic molecules and compared the results on a test set with the GB/SA solvation model. Two independent additive-constitutive QSPR methods have been developed for the computation of solvation free energy. The first is a QSPR model (HLOGS) derived using a technique that uses the counts of distinct/similar fragments and substructures for each molecule as variables in a PLS regression. The second method (ALOGS) uses an extensive atom classification scheme developed earlier for the calculation of Log *P*. A database of 265 molecules with experimentally determined solvation free energies is used to derive the HLOGS ($r = 0.97$; rms = 0.58) and ALOGS ($r = 0.98$; rms = 0.38) models, which were then tested on 27 molecules not present in the training set. A detailed comparison of the HLOGS, ALOGS, GB/SA (with AMBER* and OPLSA* potentials) on the test set showed that the HLOGS and ALOGS models give better results than the GB/SA model. Among the three methods tested, the ALOGS method gives the best result on the test set ($r = 0.96$; rms = 0.86), though the parametrization for this method is incomplete as many atom types are undetermined due to their absence in the current training set. The HLOGS method appears to handle intramolecular interactions better than the ALOGS method.

INTRODUCTION

Prediction of solvation free energies for small organic molecules is of considerable interest in drug design,[1] in the analysis of protein folding and binding,[2] and in the development of force fields[3−5] by computer simulation. Tremendous progress has been made in the development and validation of both implicit (continuum)[6−8] and explicit solvent models[9,10] for the treatment of solvation free energies of organic molecules.[3] Though these models are highly useful for accurately predicting the solvation properties of organic and biological molecules, they cannot be easily adapted for the assessment of large chemical libraries as they are based on molecular mechanics/dynamics calculations. The advent of high throughput synthesis[11,12] as a means of generating large collections of compounds of medicinal interest has entailed the necessity of computing relevant physicochemical properties such as solvation free energy and lipophilicity of organic compounds before they are synthesized, so that these calculated properties may be used in library design.[13]

Previous work by Hine and Mookerjee[14] demonstrated that solvation free energies can indeed be predicted using additive models. They developed bond contribution and group contribution models and discussed how these estimates can be used to assess the intrinsic hydrophilic character of organic compounds. However, they eliminated certain types of compounds from their training set in order to avoid large deviations for compounds with potential intramolecular hydrogen bonding. This clearly limits the applicability of their models to simple organic compounds with no internal hydrogen bonding. The present work is aimed at developing general additive-constitutive models for the prediction of solvation free energies. In the recent past, similar models have been developed for the related problem of predicting aqueous solubility of organic compounds.[15−17] While such models have the limitation that they are empirical and require considerable experimental data pertaining to different classes of organic compounds of medicinal interest, they have the advantage of simplicity and can be used for rapid, computerized estimation of physicochemical properties. Two additive-constitutive models are developed in this work. The first of these, the HLOGS model, is based on a QSPR approach utilizing a special form of molecular fingerprints called molecular holograms, which encode the structural features of a molecule as a string of numbers representing the substructures and fragments within the molecule. The second, the ALOGS model, is an atomic constant approach based on a successful "Log *P*" prediction paradigm (ALOGP[18−20]).

† Amgen, Inc.
‡ AM Technologies Inc.

10.1021/ci980154m CCC: $18.00
Published on Web 02/17/1999

METHODS

Computational methods described here are directed toward development of a "Hologram" QSAR approach (HLOGS) and an atomic constant approach (ALOGS) for prediction of solvation free energies and a comparison of the results of these approaches to those of the Generalized Born (GB/SA) approach developed by Still and co-workers.[6,21]

1. Database Preparation. The solvation free energy data for 292 simple organic molecules was taken from the examples directory of Galaxy[22] software package. These chemical structures obtained from the literature were created and stored as 3D structures in a form suitable for graphical display, using the Galaxy[22] program. The starting 3D model for each compound was built by Galaxy 2D-3D converter module. In this module a 3D structure is generated by the following steps: (i) The initial approximate structure is generated by distance geometry[23] using the bond distance and "angle distance" constraints. For the nonbonded atoms the minimum distance is set to 90% of the sum of their van der Waals radii. (ii) The resulting structure is then subjected to 200 steps of conjugate gradient molecular mechanics energy minimization. (iii) This structure is then subjected to a modified pattern search algorithm as described earlier[19,24] for torsional minimization. In this method each torsion angle is rotated within its nonrepeating domain with a specified (20°) angle increment and is set at the minimum energy angle before proceeding to the next torsion angle. The process is continued sequentially for all rotatable torsion angles and iteratively until it converges. (iv) The resulting structure was again subjected to 200 steps conjugate gradient minimization. This dataset is divided into two sets: *training* and *test.* The test set consisted of 27 molecules which included every tenth molecule of the full dataset. The training set of 265 molecules was the rest of the dataset. These training and test sets are listed in Table 1(a and b). Corresponding SD files (in the MDL format)[25] containing the 3D structural information and experimental solvation free energies for the training and test sets were used to generate molecular spread sheets containing the structures and solvation free energy data, using Sybyl[26] software.

2. Molecular Holograms and the HLOGS Model. Molecular fingerprint is a binary string representation of chemical structures designed to enhance the efficiency of chemical database searching and analysis. They can encode 2D or 3D features of molecules (or both). Two different types of 2D-fingerprints are used commonly: structural keys,[27,28] which define the presence or absence of certain groups, and hash keys, which are obtained by hashing unique structural paths. The UNITY[29] fingerprints offer a combination of keyed and hashed fingerprints, with the hashed fingerprints encoding all unique linear, branched, and cyclic fragments.

Molecular holograms are a special form of 2D fingerprints which encode more information than a traditional 2D fingerprint, by retaining a count of the number of times each unique fragment occurs, though dissimilar fragments may fall into the same bin because of the restriction on hologram size and its definition. These holograms were used to develop the HLOGS model. The HQSAR module of Sybyl[26] was used to generate molecular holograms of various lengths. The hologram lengths ranged from 71 to 401. Each hologram of a certain length consists of the number of occurrences of one or more "fingerprint" variables. Though these fingerprint variables used here are "two-dimensional", they can be extended or modified to include three-dimensional motifs. Partial Least Squares (PLS) regression is used to obtain a linear model (HLOGS) relating the hologram variables to solvation free energies. This model is then tested on the test set of 27 compounds. Additional description of the hologram variables and procedures used to obtain the holograms may be found in ref 29.

3. The ALOGS Method and Parameters for the ALOGS Method. The atom types used in the ALOGP method of predicting molecular lipophilicity has been extensively discussed in the literature. The same atom types are used for the prediction of solvation free energy. These atom types are listed in Table 2. The present analysis uses the same atom types.

The atomic log *S* parameters were determined from the general eq 1

$$\log S = \Sigma_i n_i s_i \qquad (1)$$

where n_i is the number of atoms of type i and s_i is the atomic log *S* contribution. The tabulation of data for this regression was done via the 2D-QSAR module of Galaxy.[22] A generalized inversion method[31] was used to evaluate the least squares solutions of the atomic parameters, since it guarantees the solution even for ill-conditioned matrixes. The statistical tests were performed as outlined in Zelen and Severo.[32]

4. Assessment of the Free Energy Models. To assess the calculated Solvation Free Energies from the two models, we used the Pearson Correlation Coefficient (R), RMS deviation, maximum and minimum deviations for each set, and "predictive r^2". These parameters were calculated for the overall training and test sets as well as for each subclass of molecules described earlier. The results obtained were compared with the corresponding results obtained using the GB/SA[6] model in conjunction with AMBER*[33] and OPLSA*[4] force fields available in MacroModel.[30] The application of the GB/SA model requires accurate geometries and the use of CHELPG[34] charge sets for accurate estimation of solvation free energies. These were computed using Gaussian 94 program,[35] following a HF/6-31G* optimization of the 3D structures in the test set, obtained as described under Database Preparation. The absolute solvation free energies of the 27 molecules in the test set were computed according to the procedure described earlier.[36]

Predictive r^2 [37] measures the quality of predictions relative to a simple "no model" guess, the average of all experimental log *S* values for a given set of molecules. This is given by

$$\text{predictive } r^2 = (\text{SD} - \text{"press"})/\text{SD} \qquad (2)$$

where SD is the sum of squared deviations of each measured log *S* value from their mean and "press" is the predictive sum of squared differences (the sum of squared differences between the actual and predicted log *S* values). Negative values for predictive r^2 indicate that log *S* is better estimated by "mean of values" rather than by the model under consideration.

RESULTS AND DISCUSSION

The current database of compounds is a diverse organic compound library ranging in size from 4 to 34 atoms

Table 1

(a). Experimental and Calculated Solvation Free Energies (in kcal/mol for the Training Set of 265 Molecules

no.	molecule name	expt (kcal/mol)	HLOGS (kcal/mol)	ALOGS (kcal/mol)	no.	molecule name	expt (kcal/mol)	HLOGS (kcal/mol)	ALOGS (kcal/mol)
1	methane	1.98	−1.42	1.98	71	chlorodifluoromethane	0.11	0.51	0.10
2	ethane	1.81	0.62	1.56	72	chlorotrifluoromethane	2.56	1.19	2.24
3	propane	2.02	2.03	1.73	73	bromotrifluoromethane	1.81	0.84	1.78
4	*n*-butane	2.18	2.17	1.91	74	1,1-difluoroethane	−0.11	−0.14	0.10
5	2-methylpropane	2.32	2.78	1.89	75	chloroethane	−0.64	−0.47	−0.99
6	*n*-pentane	2.36	2.67	2.07	76	bromoethane	−0.70	−0.83	−1.11
7	2,2-dimethylpropane	2.69	2.45	2.27	77	iodoethane	−0.73	−1.35	−0.91
8	cyclopentane	1.22	0.83	0.84	78	1,1-dichloroethane	−0.86	−1.06	−1.08
9	*n*-hexane	2.58	2.96	2.24	79	1,2-dichloroethane	−1.75	−1.28	−1.60
10	3-methylpentane	2.54	2.20	2.24	80	1,2-dibromoethane	−2.13	−1.26	−1.82
11	cyclohexane	1.24	1.38	1.01	81	1-chloro-2-bromoethane	−1.98	−1.15	−1.71
12	methylcyclopentane	1.62	1.20	1.01	82	1,1,1-trichloroethane	−0.25	−0.35	−0.25
13	2,2-dimethylbutane	2.63	2.49	2.43	83	1,1,2-trichloroethane	−1.98	−1.71	−1.96
14	*n*-heptane	2.65	3.01	2.40	84	1,1,2,2-tetrachloroethane	−2.39	−1.56	−2.33
15	2,4-dimethylpentane	2.92	2.96	2.40	85	pentachloroethane	−1.38	−1.22	−1.41
16	methylcyclohexane	1.73	1.02	1.17	86	hexachloroethane	−1.42	−1.41	−0.47
17	*n*-octane	2.93	3.07	2.57	87	2-chloro-1,1,1-trifluoroethane	0.06	0.61	0.23
18	2,2,4-trimethylpentane	2.89	3.05	2.76	88	1,1,2,2-tetrachloro-difluoroethane	0.83	0.26	0.37
19	ethylene	1.30	−0.41	1.20	89	1,1,2-trichloro-trifluoroethane	1.80	1.58	0.80
20	propylene	1.28	−0.17	1.12	90	1,1-dichloro-tetrafluoroethane	2.54	3.01	1.23
21	1-butene	1.40	1.52	1.28	91	1,2-dichloro-tetrafluoroethane	2.35	2.68	1.23
22	2-methylpropene	1.30	1.12	1.37	92	1-chloropropane	−0.36	−0.59	−0.50
23	1-pentene	1.69	1.77	1.45	93	2-chloropropane	−0.25	0.30	−0.79
24	*trans*-2-pentene	1.35	1.37	1.20	94	1-bromopropane	−0.57	0.41	−0.61
25	cyclopentene	0.57	1.19	−0.03	95	2-bromopropane	−0.48	−1.40	−0.90
26	2-methyl-2-butene	1.33	1.42	1.28	96	1-iodopropane	−0.62	−0.91	−0.41
27	3-methyl-1-butene	1.85	1.87	1.45	97	2-iodopropane	−0.47	−0.44	−0.70
28	cyclohexene	0.37	−0.36	0.14	98	1,2-dichloropropane	−1.27	−0.58	−1.40
29	4-methyl-1-pentene	1.93	1.51	1.63	99	1,3-dichloropropane	−1.92	−2.57	−1.89
30	*trans*-2-heptene	1.69	1.44	1.53	100	1,2-dibromopropane	−1.96	−2.06	−1.62
31	1-methylcyclohexene	0.68	0.99	0.40	101	1-chlorobutane	−0.14	−0.55	−0.33
32	1-octene	2.20	2.47	1.96	102	1-bromobutane	−0.41	−0.73	−0.44
33	1,3-butadiene	0.57	1.06	0.68	103	1-bromo-2-methylpropane	−0.03	0.68	−0.12
34	1,4-pentadiene	0.95	1.98	0.84	104	1-iodobutane	−0.26	−1.20	−0.25
35	2-methyl-1,3-butadiene	0.69	0.62	0.93	105	1,1-dichlorobutane	−0.70	−1.17	−0.51
36	1,5-hexadiene	1.02	0.62	1.01	106	1-chloropentane	−0.07	0.01	−0.17
37	acetylene	−0.01	−0.14	−0.01	107	2-chloropentane	0.07	0.50	−0.12
38	propyne	−0.48	−0.90	−0.22	108	3-chloropentane	0.04	−0.10	0.19
39	butyne	−0.17	0.11	−0.04	109	1-bromo-3-methylbutane	0.21	−0.46	−0.28
40	1-pentyne	0.01	−0.93	0.12	110	*cis*-1,2-dichloroethylene	−1.19	−1.04	−0.98
41	1-hexyne	0.29	−0.15	0.29	111	*trans*-1,2-dichloroethylene	−0.77	−1.04	−0.98
42	1-heptyne	0.61	0.68	0.46	112	1,2,3-trichloroethylene	−0.44	−0.95	−0.46
43	1-octyne	0.72	0.87	0.62	113	tetrachloroethylene	0.06	−0.66	0.06
44	1-nonyne	1.06	0.93	0.80	114	3-chloropropene	−0.58	−0.53	−0.46
45	benzene	−0.90	−1.08	−0.65	115	chlorobenzene	−1.02	−1.52	−0.94
46	ethylbenzene	−0.62	−0.12	−0.53	116	bromobenzene	−1.48	−1.92	−1.44
47	*o*-xylene	−0.91	−1.05	−0.75	117	1,2-dichlorobenzene	−1.38	−1.91	−1.22
48	*m*-xylene	−0.82	−0.64	−0.75	118	1,3-dichlorobenzene	−0.99	−0.88	−1.22
49	*p*-xylene	−0.82	−0.25	−0.75	119	1,4-dibromobenzene	−2.33	−2.32	−2.21
50	propylbenzene	−0.54	−0.97	−0.36	120	*p*-bromotoluene	−1.41	−1.78	−1.48
51	2-propylbenzene	−0.30	0.03	−0.36	121	1-bromo-2-ethylbenzene	−1.20	−0.65	−1.31
52	1,2,4-trimethylbenzene	−0.87	−0.93	−0.79	122	*o*-bromocumene	−0.86	−1.12	−1.15
53	butylbenzene	−0.40	−0.19	−0.19	123	dimethyl ether	−1.92	−1.27	−1.99
54	2-butylbenzene	−0.46	−1.06	−0.19	124	dimethyl sulfide	−1.56	−1.98	−1.88
55	*tert*-amylbenzene	−0.18	0.23	0.17	125	dimethoxymethane	−2.97	−3.54	−3.48
56	naphthalene	−2.45	−2.79	−2.75	126	1,3-dioxolane	−4.14	−3.47	−3.50
57	acenaphthene	−3.44	−4.31	−4.06	127	diethyl ether	−1.77	−2.27	−2.42
58	anthracene	−4.34	−4.41	−4.84	128	methylpropyl ether	−1.69	−1.16	−1.71
59	phenanthrene	−4.12	−3.36	−4.84	129	methyl isopropyl ether	−2.03	−1.66	−2.00
60	fluoromethane	−0.22	−0.55	−0.22	130	tetrahydrofuran	−3.51	−3.74	−2.98
61	trifluoromethane	0.82	1.01	0.53	131	dioxane	−5.11	−4.05	−4.05
62	tetrafluoromethane	3.21	1.69	2.67	132	ethylpropyl ether	−1.84	−1.60	−1.92
63	chloromethane	−0.54	−0.88	−0.79	133	methyl *tert*-butyl ether	−2.24	−2.46	−2.25
64	trichloromethane	−1.04	−0.36	−0.75	134	2-methyltetrahydrofuran	−3.34	−2.92	−2.78
65	tetrachloromethane	0.10	−0.14	0.97	135	tetrahydropyran	−3.16	−4.02	−2.82
66	bromomethane	−0.80	−1.51	−0.90	136	dipropyl ether	−1.17	−0.73	−1.42
67	dibromomethane	−1.99	−1.85	−1.99	137	1,2-diethoxyethane	−3.30	−4.12	−4.43
68	tribromomethane	−2.16	−1.82	−2.14	138	1,1-diethoxyethane	−3.32	−3.15	−3.45
69	iodomethane	−0.90	−1.24	−0.70	139	di-*n*-butyl ether	−0.84	−1.63	−1.09
70	chlorofluoromethane	−0.79	−0.28	−0.93	140	anisole	−1.05	−0.94	−2.06

Table 1 (Continued)

(a). Experimental and Calculated Solvation Free Energies (in kcal/mol for the Training Set of 265 Molecules (Continued)

no.	molecule name	expt (kcal/mol)	HLOGS (kcal/mol)	ALOGS (kcal/mol)	no.	molecule name	expt (kcal/mol)	HLOGS (kcal/mol)	ALOGS (kcal/mol)
141	thioanisole	−2.76	−2.36	−2.92	204	isopropylformate	−2.04	−1.91	−2.57
142	2,2′-dichlorodiethyl sulfide	−3.97	−3.52	−3.51	205	ethylacetate	−3.12	−3.03	−3.21
143	methanol	−5.14	−2.72	−4.46	206	methylpropionate	−3.01	−2.92	−2.69
144	methane thiol	−1.26	−1.58	−1.41	207	isobutylformate	−2.25	−1.92	−2.07
145	ethanol	−4.96	−5.35	−4.67	208	propylacetate	−2.89	−3.05	−2.71
146	2,2,2-trifluoroethanol	−4.35	−3.32	−3.44	209	isopropylacetate	−2.68	−2.85	−3.00
147	ethylene glycol	−7.75	−8.26	−8.95	210	methylbutyrate	−2.87	−3.16	−2.53
148	1-propanol	−4.92	−4.90	−4.17	211	isoamylformate	−2.16	−1.85	−1.95
149	2-propanol	−4.81	−4.81	−4.46	212	butylacetate	−2.58	−2.92	−2.54
150	allyl alcohol	−5.10	−2.80	−4.14	213	isobutylacetate	−2.39	−2.82	−2.21
151	1,1,1-trifluoro-2-propanol	−4.21	−4.17	−3.23	214	propylpropionate	−2.49	−2.82	−2.40
152	2,2,3,3-tetrafluoropropanol	−4.96	−5.82	−5.10	215	isopropylpropionate	−2.25	−2.36	−2.69
153	2,2,3,3,3-pentafluoropropanol	−4.20	−4.38	−4.06	216	ethylbutyrate	−2.53	−2.97	−2.74
154	hexafluoro-2-propanol	−3.81	−3.41	−1.99	217	methylpentanoate	−2.57	−2.56	−2.36
155	glycerol	−8.50	−9.09	−13.03	218	amylacetate	−2.49	−2.75	−2.38
156	1-butanol	−4.78	−4.46	−4.01	219	propylbutyrate	−2.31	−3.04	−2.24
157	2-butanol	−4.67	−3.99	−3.98	220	ethylpentanoate	−2.56	−2.36	−2.57
158	*tert*-butyl alcohol	−4.57	−4.28	−4.73	221	methylhexanoate	−2.51	−2.33	−2.18
159	2-methyl-1-propanol	−4.57	−4.59	−3.69	222	hexylacetate	−2.29	−2.49	−2.21
160	1-pentanol	−4.55	−4.12	−3.84	223	amylpropionate	−2.02	−2.51	−2.07
161	2-pentanol	−4.45	−4.24	−3.80	224	methyloctanoate	−2.07	−1.95	−1.85
162	2-methyl-1-butanol	−4.48	−4.16	−3.51	225	ethylheptanoate	−2.33	−1.80	−2.22
163	2-methyl-2-butanol	−4.49	−4.61	−4.23	226	methylbenzoate	−4.34	−4.49	−4.86
164	1-hexanol	−4.42	−3.84	−3.68	227	ethylamine	−4.67	−4.43	−5.04
165	cyclohexanol	−5.02	−5.17	−4.53	228	butylamine	−4.43	−4.74	−4.38
166	2,3-dimethylbutanol	−3.97	−3.14	−3.34	229	pentylamine	−4.14	−3.51	−4.21
167	2-methyl-3-pentanol	−3.94	−4.24	−2.98	230	hexylamine	−4.09	−3.36	−4.05
168	4-methyl-2-pentanol	−3.79	−4.10	−3.63	231	dimethylamine	−4.34	−3.85	−4.26
169	2-methyl-2-pentanol	−3.98	−3.59	−4.06	232	diethylamine	−4.12	−4.23	−4.67
170	1-heptanol	−4.31	−3.79	−3.51	233	pyrrolidine	−5.54	−5.83	−5.25
171	1-octanol	−4.16	−3.73	−3.33	234	piperidine	−5.17	−6.08	−5.07
172	phenol	−6.62	−6.67	−6.24	235	dipropylamine	−3.70	−3.62	−3.69
173	4-bromophenol	−7.20	−6.95	−7.02	236	hexamethylenimine	−4.97	−5.86	−4.90
174	thiophenol	−2.58	−1.80	−2.43	237	trimethylamine	−3.27	−2.60	−2.97
175	2-cresol	−5.94	−6.09	−6.29	238	triethylamine	−3.07	−3.30	−3.59
176	4-nitrophenol	−6.20	−5.90	−6.29	239	*n*-methylpyrrolidine	−4.02	−3.77	−3.95
177	4-hydroxybenzaldehyde	−10.61	−9.75	−10.14	240	*n*-methylpiperidine	−3.94	−3.09	−3.79
178	4-*tert*-butylphenol	−6.00	−5.95	−5.60	241	ethylenediamine	−9.88	−10.54	−9.70
179	acetaldehyde	−3.55	−2.13	−3.77	242	acetonitrile	−3.94	−3.99	−4.10
180	propanal	−3.48	−3.19	−3.47	243	propionitrile	−3.90	−4.24	−3.80
181	butanal	−3.22	−3.40	−3.30	244	butyronitrile	−3.69	−3.92	−3.63
182	pentanal	−3.07	−3.21	−3.12	245	nitroethane	−3.76	−3.04	−3.76
183	heptanal	−2.71	−2.56	−2.79	246	2-nitropropane	−3.18	−3.16	−3.18
184	octanal	−2.32	−2.50	−2.63	247	nitrobenzene	−4.17	−4.24	−3.74
185	nonanal	−2.10	−2.45	−2.46	248	2-nitrotoluene	−3.63	−3.77	−3.79
186	trans-2-butenal	−4.28	−4.82	−3.90	249	3-nitrotoluene	−3.50	−3.98	−3.79
187	*trans*-2-hexenal	−3.73	−3.47	−3.56	250	pyridine	−4.75	−5.17	−4.16
188	*trans*-2-octenal	−3.48	−3.03	−3.22	251	2-methylpyridine	−4.68	−4.49	−4.17
189	*trans-trans*-2,4-hexadienal	−4.70	−5.29	−4.43	252	3-methylpyridine	−4.84	−4.82	−4.20
190	benzaldehyde	−4.08	−4.59	−4.55	253	4-methylpyridine	−4.99	−5.42	−4.20
191	acetone	−3.85	−3.50	−3.84	254	2-ethylpyridine	−4.38	−4.79	−3.87
192	2-pentanone	−3.56	−3.26	−3.37	255	4-ethylpyridine	−4.78	−4.27	−4.03
193	2-heptanone	−3.11	−2.89	−3.03	256	2,3-dimethylpyridine	−4.88	−5.62	−4.21
194	2-octanone	−2.92	−2.56	−2.86	257	2,4-dimethylpyridine	−4.92	−4.46	−4.21
195	2-nonanone	−2.51	−2.50	−2.69	258	2,5-dimethylpyridine	−4.77	−4.41	−4.21
196	2-undecanone	−2.18	−2.39	−2.35	259	2,6-dimethylpyridine	−4.66	−3.88	−4.17
197	acetophenone	−4.64	−4.46	−4.64	260	3,4-dimethylpyridine	−5.28	−4.96	−4.26
198	acetic acid	−6.78	−4.97	−7.18	261	3,5-dimethylpyridine	−4.90	−5.13	−4.26
199	propionic acid	−6.55	−5.24	−6.88	262	2-methylpyrazine	−5.58	−5.69	−7.67
200	butyric acid	−6.44	−6.30	−6.71	263	2-ethylpyrazine	−5.53	−5.79	−7.38
201	ethylformate	−2.68	−2.38	−2.78	264	2-ethyl-3-methoxypyrazine	−4.45	−4.88	−4.26
202	methylacetate	−3.36	−2.93	−3.00	265	2-isobutyl-3-methoxypyrazine	−3.73	−3.91	−3.92
203	propylformate	−2.51	−2.47	−2.28					

(b). Experimental and Calculated Solvation Free Energies (in kcal/mol) for the Test Set of 27 Molecules

	compound	expt	HLogS	ALogS	GB/SA AMBER*	GB/SA OPLSA*
1	2-methylpentane	2.56	3.09	2.23	2.40	2.40
2	*cis*-1,2-dimethylcyclohexane	1.60	0.76	1.34	2.12	2.12
3	1-hexene	1.73	1.96	1.62	1.43	1.42
4	2,3-dimethyl-1,3-butadiene	0.40	0.08	1.18	1.71	1.71

Table 1 (Continued)

(b). Experimental and Calculated Solvation Free Energies (in kcal/mol) for the Test Set of 27 Molecules (Continued)

	compound	expt	HLogS	ALogS	GB/SA AMBER*	GB/SA OPLSA*
5	toluene	−0.77	−0.59	−0.71	0.58	0.59
6	*tert*-butylbenzene	−0.44	−0.26	−0.01	1.48	1.49
7	dichloromethane	−1.42	−1.26	−1.51	−2.10	−2.10
8	dichlorodifluoromethane	1.71	0.64	1.82	1.98	1.98
9	chloropentafluoroethane	2.90	3.36	1.66	2.80	2.80
10	1,3-dibromopropane	−1.99	−0.94	−2.12	−3.84	−3.30
11	chloroethylene	0.50	−1.19	0.11	−1.00	−1.00
12	1,4-dichlorobenzene	−1.02	−0.90	−1.22	−1.00	−1.01
13	diethyl sulfide	−1.45	−2.02	−2.29	−0.28	−0.28
14	diisopropyl ether	−0.54	−0.14	−2.00	−0.69	−0.77
15	ethane thiol	−4.08	−1.65	−1.61	−0.83	−0.83
16	3-hexanol	−3.73	−3.33	−3.31	−3.05	−3.09
17	4-nitrophenol	−10.74	−8.12	−9.33	−12.63	−12.63
18	hexanal	−2.85	−2.89	−2.96	−4.36	−4.36
19	2-butanone	−3.76	−3.35	−3.54	−4.46	−4.46
20	methylformate	−2.82	−2.86	−2.57	−8.27	−8.31
21	ethylpropionate	−2.83	−2.75	−2.90	−4.49	−4.52
22	isoamylacetate	−2.24	−2.66	−2.38	−5.05	−5.09
23	propylamine	−4.56	−3.82	−4.55	−1.97	−1.97
24	dibutylamine	−3.38	−4.91	−3.35	0.38	0.34
25	1-nitropropane	−3.38	−1.61	−3.26	−8.25	−8.25
26	4-ethylpyridine	−4.66	−4.82	−4.04	−2.39	−2.40
27	2-isobutylpyrazine	−5.11	−7.25	−7.04	−2.15	−2.15

containing mostly mono- and bifunctional organic compounds. It contains alkanes, cycloalkanes, alkenes, aromatic and aliphatic heterocycles, halogenated aromatics and aliphatics, and alcohols along with very few polyfunctional compounds such as glycerol. The training set contains most of the functional groups of interest in medicinal chemistry except charged groups.

Molecular holograms ranging in length from 71 to 401 were computed for the training set of 265 molecules using SYBYL/HQSAR module. Each hologram of a given length (the number of bins storing the fingerprints) for a compound represents the molecule in a manner comparable to the holograms of other molecules. Each bin of the hologram contains the number of occurrences of one or more specific fragment fingerprints. The application of partial least squares (PLS) procedure for obtaining a regression model requires the specification of the maximum number of PLS components which is set at 20. Table 3 summarizes the results obtained in the training set for holograms of different length. It is seen from Table 3 that the best results are obtained for the dataset for a hologram size of 257 with 18 PLS components.

Table 1(a) shows the experimental and calculated solvation free energies obtained for the training set with these parameters. Table 1(b) shows the experimental and calculated results obtained for the test set. Comparison of the experimental and calculated results obtained using HLOGS and ALOGS methods shows that the results of both methods agree well with experimental results, but there are several exceptions. Eleven out of 265 compounds in the training set have deviations (from corresponding experimental values) greater than 1 for the ALOGS method, whereas 14 compounds have such deviations for the HLOGS method. There is no overlap between the two sets of compounds with the exception of dioxane (compound 131), indicating that these methods do not share the same limitations. Though the ALOGS method shows fewer deviations, its greatest deviation (4.53) from the experimental value occurs in the case of glycerol (compound **155**), which is reasonably well-predicted by the HLOGS method. The deviations shown by the ALOGS method for some of these compounds can be explained on the basis of intramolecular interactions. Glycerol (compound **155**) and ethylene glycol (compound **147**) contain internal hydrogen bonding groups and hence they are expected to show less hydrophilic character. Such compounds are not included in the training set of the group contribution method of Hine and Mookerjee.[14] For glycerol, this method gives a predicted value of −12.9 kcal·mol, twice as large as the experimental value (−8.4 kcal/mol). In these cases, the HLOGS method performs well. The fingerprints reflect the presence of structural fragments (of length 1−7 atoms) and thus can encode potential intramolecular hydrogen bonding information. On the other hand, this method shows large deviations (>2.0) in the cases of methane, methanol, and allyl alcohol in the training set. Six compounds in the training set showing deviations greater than 1.0 by the HLOGS model have only one or two heavy atoms (other than halogens). These are methane, ethane, ethylene, tetrafluoromethane, chlorotrifluoromethane, and methanol (compounds **1**, **2**, **19**, **62**, **72**, and **143** in Table 1(a)). For these molecules, very few bins would be occupied in the holograms; hence, the HLOGS method is not expected to perform well in such cases. Among nitrogen containing heterocycles, pyrazine derivatives (compounds **262**−**265**) are expected to show less hydrophilic character than expected by the ALOGS method because the hydrogen bonding ability of nitrogens decreases as more nitrogens are present in the ring as in pyrazine. Hence, on the average pyrazines are calculated to be more hydrophilic by the atomic constant (ALOGS) method. Particularly, 2-methylpyrazine (compound 262) shows a large deviation (of 3.0) with the ALOGS method. The hologram method (HLOGS) shows lesser deviation in the training set, though, in the test set, 2-isobutylpyrazine (compound **27**) shows large deviation with either method.

The intrinsic hydrophilicity and hydrophobicity of halogenated compounds is of interest for several reasons.

Table 2. Classification of Atoms and Their Contributions to Solvation Free Energy (Atomic Value) (in kcal/mol)[i]

atom type no.	description[a]	atomic value	error margin	atom type no.	description[a]	atomic value	error margin
	C in				O in		
1	:CH_3R, CH4	−2.28	0.03	56	:alcohol	−1.19	0.13
2	:CH_2R_2	−2.22	0.02	57	:phenol, enol, carboxyl OH	−1.93	0.26
3	:CHR_3	−1.64	0.15	58	:═O	−3.63	0.11
4	:CR_4	−0.87	0.35	59	:Al−O−Al	2.49	0.14
5	:CH_3X	−6.55	0.11	60	:Al−O−Ar, Ar_2O	3.46	0.15
6	:CH_2RX	−5.12	0.06		:R···O···R, R−O−C═X		
7	:CH_2X_2	−0.37	0.39	61[e]	:- -O	−0.23	0.18
8	:CHR_2X	−3.29	0.16	62	:O^- (negatively charged)		
9	:$CHRX_2$	−1.82	0.24	63	:R−O−O−R		
10	:CHX_3	−1.44	0.39		Se in		
11	:CR_3X	−1.92	0.39	64	:Any−Se−Any		
12	:CR_2X_2	−4.26	0.56	65	:═Se		
13	:CRX_3	−0.12	0.14		N in		
14	:CX_4	1.13	0.39	66	:Al−NH2	−0.53	0.28
15	:═CH_2	−2.27	0.15	67	:Al_2NH	1.26	0.32
16	:═CHR	−1.70	0.10	68	:Al_3N	3.76	0.39
17	:═CR_2	−0.79	0.39	69	:Ar−NH_2, X−NH_2		
18	:═CHX	−1.03	0.26	70	:Ar−NH−Al		
19	:═CRX			71	:Ar−NAl_2		
20	:═CX_2	0.03	0.35	72	:RCO−N<, >N−X═X		
21	:≡CH	1.79	0.24	73	:Ar_2NH, Ar_3N		
22	:≡CR, R═C═R	−0.99	0.30		:Ar_2N−Al, R···N···R[f]		
23	:≡CX			74	:R≡N, R═N−	1.18	0.45
24	:R- -CH- -R	−1.54	0.02	75	:R- -N- -R,[g] R- -N- -X	−1.44	0.15
25	:R- -CR- -R	−0.93	0.08	76	:Ar−NO_2. R- -N(- -R)- -O[h]	−0.01	0.45
26	:R- -CX- -R	−2.73	0.14		RO−NO,		
27	:R- -CH- -X	−2.00	0.11	77	:Al−NO_2	−0.10	0.56
28	:R- -CR- -X	−1.58	0.22	78	:Ar−N═X, X−N═X		
29	:R- -CX- -X	0.70	0.56	79	:N+ (positively charged)		
30	:X- -CH- -X			80	unused		
31	:X- -CR- -X				F attached to		
32	:X- -CX- -X			81	:C^1_{sp3}	2.02	0.79
33	:R- -CH···X			82	:C^2_{sp3}	1.81	0.16
34	:R- -CR···X			83	:C^3_{sp3}	0.39	0.06
35	:R- -CX···X			84	:C^1_{sp2}		
36	:Al−CH═X	−1.33	0.24	85	:C^{2-4}_{sp2}, C^1_{sp}, C^4_{sp3}, X		
37	:Ar−CH═X	−0.25	0.56		Cl attached to		
38	:Al−C(═X)−Al			86	:C^1_{sp3}	1.46	0.15
39	:Ar−C(═X)−R	0.09	0.79	87	:C^2_{sp3}	1.23	0.14
40	:R−C(═X)−X, R−C≡X, X═C═X	−0.97	0.11	88	:C^3_{sp3}	−0.04	0.07
41	:X−C(═X)−X			89	:C^1_{sp2}	2.34	0.19
42	:X- -CH···X			90	:C^{2-4}_{sp2}, C^1_{sp}, C^4_{sp3}, X		
43	:X- -CR···X				Br attached to		
44	:X- -CX···X			91	:C^1_{sp3}	1.34	0.20
45	unused			92	:C^2_{sp3}	0.99	0.39
	H attached to[c]			93	:C^3_{sp3}	−0.51	0.25
46	:C^0_{sp3} having no X attached to next C	1.19	0.01	94	:C^1_{sp2}	1.84	0.26
47	:C^1_{sp3}, C^0_{sp2}	1.43	0.01	95	:C^{2-4}_{sp2}, C^1_{sp}, C^4_{sp3}, X		
48	:C^2_{sp3}, C^1_{sp2}, C^0_{sp}	−1.80	0.11	96	iodine attached to C^1_{sp3}	1.54	0.35
49	:C^3_{sp3}, C^{2-3}_{sp2}, C^{1-3}_{sp}	0.81	0.09	97−100	other iodine types		
50	:heteroatom	−1.03	0.08	101−104	halide ions		
51	:α-C[d]	1.06	0.04	105	unused		
52	:C^0_{sp3}, having 1 X attached to next carbon	0.87	0.02		S in		
53	:C^0_{sp3}, having 2 X attached to next carbon	0.96	0.13	106	:R−SH	1.87	0.56
54	:C^0_{sp3}, having 3 X attached to next carbon	0.93	0.26	107	:R_2S, RS−SR	2.61	0.45
55	:C^0_{sp3}, having 4 or more X attached to next carbon			108−110	other sulfur types		
				111−114	Si, B and unused		
				115−120	phosphorus types		

[a] R represents any group linked through carbon; X represents any heteroatom (O, N, S, P, Se, and halogens); Al and Ar represent aliphatic and aromatic groups, respectively; ═ represents double bond; ≡ represents triple bond; - - represents aromatic bond as in benzene or delocalized bonds such as the N−O bond in a nitro group; ··· represents aromatic single bonds as the C−N bond in pyrrole. [b] Atomic hydrophobicity in the unit of log *P*(octanol−water). [c] The subscript represents hybridization and the superscript its formal oxidation number. The formal oxidation number of a carbon atom = sum of the formal bond orders with electronegative atoms, the C- -N bond order in pyridine may be considered as 2, while we have one such bond and 1.5 when we have two such bonds, the C···X bond order in pyrrole or furan may be considered as 1. [d] An α-C may be defined as a C attached through a single bond with −C═X, −C≡X, −C- -X. [e] As in nitro, *N*-oxides, [f] Pyrrole-type structure. [g] Pyridine-type structure. [h] Pyridine *N*-oxide type. [i] Error margins are shown at 80% confidence level.

Halogens (except fluorine) have a relative hydrophobic character as the halogenated ligands are known to favorably interact with hydrophobic groups in a receptor. On the other hand, halogenated compounds are in general harder to

Table 3. Results of Experiments with Different Lengths of Holograms[a]

hologram length	q^2	standard error	no. of PLS components
71	0.45	1.37	7
83	0.44	1.40	12
97	0.52	1.30	12
151	0.57	1.23	10
199	0.59	1.21	15
257	0.77	0.91	18
307	0.73	0.96	11
353	0.73	0.98	13
401	0.75	0.95	13

[a] Cross validated $r2$ (q^2), standard error, and the optimal number of PLS components are given for each run.

desolvate (compare for e.g., the solvation free energies for ethane (compound **2**), 1,1,1-trichloroethane (compound **82**), and 1,1,2-trichloroethane (compound **83**)). Out of 70 halogenated compounds in the training set, only a few show deviations greater than one by the ALOGS (compounds **89**, **90**, and **91**) HLOGS (compounds **62**, **72**, and **146**) methods. Thus, most of the halogenated compounds are reasonably well-predicted by both methods. The importance of differentiating halogen types on the basis of the oxidation number of the attached carbon is brought out by considering the solvation free energy difference between 1,1,1-trichloroethane and 1,1,2-trichloroethane (compounds **82** and **83**). In these cases, the chlorines and carbons are assigned different types, correctly reflecting the relative difference between the two molecules. One of the limitations of both methods is in the failure to account for stereoisomers. Thus, the cis and trans isomers of 1,2-dichloroethylene (compounds **110** and **111**) are assigned identical values by these methods though the experimental values are different for these isomers (by 0.42 kcal/mol). However, their average value is accounted for rather accurately by both methods.

The atomic parameters obtained by the regression procedure are listed in Table 2 along with the error margin at 80% confidence level.[38] The training set of compounds contained only 67 atom types. Therefore atomic values for the rest of the 53 atom types could not be determined. While it is possible to decrease the total number of atom types used to make the method more generally applicable with the present training set, this would be at the expense of accuracy; hence this option was not used. With the availability of more experimental data, the list of atomic constants will be expanded to cover more atom types. As the atomic constants were derived for the ALOGP atom types, a comparison of the present atomic constants (atomic solvation parameters) may be made with the atomic lipophilicity parameters (based on log P data). In the case of alkyl carbon types (types 1−14), qualitative agreement exists between the two scales, which are assigned negative values. The negative hydrophobicity values assigned to these caused some concern earlier as alkyl carbon is considered hydrophobic. However, the attached hydrogens are assigned positive values in both ALOGP and ALOGS scales, so that the totals for alkyl groups are always positive. The only alkyl carbon type which is assigned a large positive constant in both ALOGP and ALOGS scales is type 14 (CX_4). For most hydrogen and halogen types, the signs in the two scales match. However, the halogenated alkyl groups are intrinsically hydrophilic, though, as is well-known, they also possess hydrophobic character (e.g., compare chloromethane and methane). The ALOGS parameters for halogens have a positive sign with the exception of constants for atom types 88 and 93, similar to the ALOGP constants. The corresponding attached carbon types (C^0_{sp3}, C^1_{sp3}, C^0_{sp2}, etc.) are, however, assigned high negative values. Therefore, while assessing the relative hydrophilicity of different halogenated substituents on a ligand, one should consider groups of atoms such as trichloromethyl, etc. and not focus on a single atom type or constant.

Table 4

(a) Statistical Results for the Calculated Solvation Free Energies for the Compounds in the Training Set Used for the HLOGS and ALOGS Models

method	r	rms dev	max dev	min dev	max obs	min obs	pred r^2	no. of points
HLOGS	0.97	0.58	3.40	0.00	3.21	−10.61	0.95	265
ALOGS	0.98	0.38	4.53	0.00	3.21	−10.61	0.96	265

(b) Statistical Results for the Calculated Solvation Free Energies for the Compounds in the Test Set Used for the HLOGS, ALOGS, GB/SA (AMBER*), and GB/SA (OPLSA*) Models

method	r	rms dev	max dev	min dev	max obs	min obs	pred r^2	no. of points
HLOGS	0.93	1.07	2.62	0.04	2.90	−10.74	0.86	27
ALOGS	0.96	0.82	2.47	0.01	2.90	−10.74	0.92	27
GB/SA (AMBER*)	0.79	2.21	5.45	0.02	2.90	−10.74	0.40	27
GB/SA (OPLSA*)	0.80	2.20	5.49	0.01	2.90	−10.74	0.41	27

Table 4 shows the overall statistical results obtained for these methods for the training and test sets. For the training set, both methods show high correlation and predictive r^2, though the ALOGS method has a better rms deviation. The HLOGS method used only 18 PLS components (or independent variables), whereas the ALOGS method used 67 independent variables with 265 observations. Though this ratio is far from ideal for obtaining robust parameters for the ALOGS method, a reasonable "explained variance" of 0.94 is obtained for the regression model, which shows that the parameter set is reliable, though it may need additions and improvements. For the test set of molecules listed in Table 1(b), the ALOGS method shows the best correlation ($r = 0.96$) and predictive r^2 ($= 0.92$) among all methods tested. The HLOGS method shows a good correlation of 0.93, whereas the GB/SA method shows a lower correlation ($r = 0.80$) and predictive r^2 (0.41) with the OPLSA* force field. GB/SA results are very similar for OPLSA* and AMBER* force fields.

Hine and Mookerjee[14] developed two group contribution methods for solvation free energy based on bond contributions and atom group contributions. However, they deleted 80 compounds from their 192 compound database because solvation free energies for those 80 compounds were influenced by long range polar interactions within the molecule causing the predictions to deviate considerably from experimental data. For the rest of the data, they obtained very good fits (with an RMS less than 0.2 kcal/mol), though they did not use a separate test set. Deleting the badly predicted compounds serously limits the applicability of their models because most of the compounds of medicinal interest possess multiple polar groups. The present ALOGS and HLOGS methods have been developed and tested considering all molecules in the database and thus they are more general

than previous models.[14] The HLOGS method appears suitable to account for intramolecular interactions by considering substructural fragments which can include these interactions, though this is not a validated conclusion. The ALOGS method is shown to be more predictive, but the current version of the model (training database and atom types) does not fully account for intramolecular hydrogen bonding.

CONCLUSION

The results demonstrate that the HLOGS and ALOGS models presented here are good predictors of log *s*. While the ALOGS method gives better statistical results, the HLOGS method appears to offer more generality with regard to the applicability to many types of organic compounds based on the present training set. In cases of molecules with potential intramolecular hydrogen bonding groups such as glycerol, the HLOGS method performs better than the ALOGS method, though, in several cases, even the HLOGS method does not offer good predictions. Results with the test data set show that these methods are capable of accurate predictions, comparable to or better than the GB/SA results for the test set. However, since the training set does not contain charged compounds, the models presented here will not be predictive for these types of compounds. The present QSPR study offers a general validation of "hologram QSAR" (HQSAR) procedure for a database of organic compounds containing a large variety of functional groups.

REFERENCES AND NOTES

(1) Kollman, P. A. *Acc. Chem. Res.* **1996**, *29*, 461−469.
(2) Eisenberg, D.; McLachlan, A. D. *Nature (London)* **1986**, *319*, 199−203.
(3) Jorgensen, W. L.; TiradoRives, J. *Persp. Drug. Discuss. Des.* **1995**, *3*, 123−138.
(4) Jorgenson, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.
(5) Jorgenson, W. L.; Nguyen, T. B. *J. Comput. Chem.* **1993**, *14*, 195.
(6) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(7) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R.; Sitkoff, D.; Nicholls, A.; Ringalda, M.; Goddard, W. A. I.; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 11875.
(8) Papazyan, A.; Warshel, A. *J. Phys. Chem. B.* **1997**, *101*, 11254−11264.
(9) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269−6271.
(10) Jorgenson, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R.; Klein, M. *J. Chem. Phys.* **1984**, *79*, 926−935.
(11) Wilson, S. R., Wilson, S. R., Czarnick, A. W., Eds.; John Wiley: New York, 1997; pp 1−23.
(12) Armstrong, R. W.; Combs, A. P.; Tempest, P. A.; Brown, S. D.; Keating, T. A. *Acc. Chem. Res.* **1996**, *29*, 123−131.
(13) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. In *Rational Drug Design*; Parrill, A., Reddy, M. R., Eds.; American Chemical Society: Washington, DC, 1998 (in press).
(14) Hine, J.; Mookerjee, P. K. *J. Org. Chem.* **1974**, *40*, 292−298.
(15) Bodor, N.; Harget, A.; Huang, M.-J. *J. Am. Chem. Soc.* **1991**, *113*, 9480−9483.
(16) Yalkowsky, S. H.; Banerjee, S. *Aqueous solubility, methods of estimation for organic compounds*; Marcel Dekker: New York, Basel, and Hong Kong, 1992.
(17) Klopman, G.; Wang, S.; Balthasar, D. M. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 474−482.
(18) Ghose, A. K.; Crippen, G. M. *J. Comput. Chem.* **1986**, *7*, 565−577.
(19) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163−172.
(20) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. *J. Phys. Chem.* **1998**, *102*, 3762−3772.
(21) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem.* **1997**, *101*, 3005−3014.
(22) Galaxy 2.3 Molecular Modeling Software Package; A. M. Technologies, Inc.: San Antonio, TX, 1997.
(23) Crippen, G. M.; Havel, T. F. *Distance Geometry and Molecular Conformation*; Research Studies Press: New York, 1988.
(24) Ghose, A. K.; Logan, M. E.; Treasurywala, A. M.; Wang, H.; Wahl, R. C.; Tomczuk, B.; Gowravaram, M.; Jaeger, E. P.; Wendoloski, J. J. *J. Am. Chem. Soc.* **1995**, *117*, 4671−4682.
(25) Integrated Scientific Informarion System (ISIS) v 2.0.2; MDL Information Systems, Inc.: San Leandro, CA, 1997.
(26) Sybyl 6.3 Molecular Modeling Software; Tripos Associates: St. Louis, 1998.
(27) MACCS-II and ISIS Software; Molecular Design Ltd.: San Leandro, CA, 1998.
(28) Barnard, J. M.; Downs, G. M. Fingerprint Descriptor Package, 3.1 version; Barnard Chemical Information Ltd.: Sheffield, 1995.
(29) (a) Unity Reference Manual; Tripos Associates: St. Louis, MO, 1998. (b) Lowis, D. R. *HQSAR: A New, Highly Predictive QSAR Technique*; Tripos Technical Note, Tripos Associates: St. Louis, MO, 1997.
(30) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T. *J. Comput. Chem.* **1990**, *11*, 440−467.
(31) Krishnamurthy, E. V.; Sen, S. K. *Computer Based Numerical Algorithms*; East-West Press: New Delhi, 1976.
(32) Zelen, M.; Severo, N. C.; Abramovitz, M.; Stegun, I. A. In *Handbook of Mathematical Functions*; Abramovitz, M., Stegun, I. A., Eds.; Dover: New York, 1972; pp 925−964.
(33) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1996**, *118.*, 2309−2309.
(34) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.
(35) Frisch, M. J.; Head-Gordon, M.; Schelegel, H. B.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Defrees, D. J.; Fox, D. J.; Whiteside, R. J.; Seeger, R.; Melius, C. F.; Baker, J.; Martin, R.; Kahn, L. R.; Stewart, J. J. P.; Fluder, E. M.; Topiol, S.; Pople, J. A. *GAUSSIAN 88*; Gaussian: Pittsburgh, PA, 1988.
(36) Technical Manual, Macromodel V6.0; Available from Schrodinger, Inc., Jersey City, NJ, 1999.
(37) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
(38) Atomic Solvation Constants were derived by Ghose, A. K.; Singh, U. C., at A.M. Technologies, Inc.: San Antonio, TX.

CI980154M