JOURNAL OF PHYSICAL ORGANIC CHEMISTRY
J. Phys. Org. Chem. 2000: **13;** 80–86

Prediction of liquid viscosity for organic compounds by a quantitative structure–property relationship

Alan R. Katritzky,[1]* Ke Chen,[1] Yilin Wang,[1] Mati Karelson,[2] Bono Lucic,[3] Nenad Trinajstic,[3] Takahiro Suzuki[4] and Gerrit Schüürmann[5]

[1]Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, P.O. Box 117200, Gainesville, Florida 32611-7200, USA
[2]Department of Chemistry, University of Tartu, 2 Jakobi Str., Tartu, EE51014, Estonia
[3]The Rugjer Boškoviæ Institute, P.O. Box 1016, HR-10001 Zagreb, Croatia
[4]Research Laboratory of Resources Utilization, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 226, Japan
[5]Department of Chemical Ecotoxicology, UFZ Centre for Environmental Research, Permoserstrasse 15, D-04318 Leipzig, Germany

Received 22 March 1999; revised 26 April 1999; accepted 27 April 1999

ABSTRACT: The liquid viscosity of 361 organic compounds containing C, H, N, O, S and/or halogens was investigated using a quantitative structure–property relationship (QSPR) approach. A five-descriptor equation was obtained with a squared correlation coefficient (R^2) of 0.854 and a standard error (S) of 0.22 log units. The data set was divided into three subsets and each of these was cross-validated by the model obtained from the complementary two subsets; the result indicates a high statistical robustness of the overall model. The average absolute error S^2 of 0.0494 log units is small compared with the overall range of eta of 1490 mPas. Copyright © 2000 John Wiley & Sons, Ltd.

KEYWORDS: liquid viscosity; organic compounds; quantitative structure–activity relationship

INTRODUCTION

Viscosity is a highly important transport property of liquids for many studies related to the transfer or movement of bulk quantities of liquids in the petroleum industries[1] and in chemical engineering in general. It would be of obvious utility to develop an effective method to predict the viscosity of compounds for which a measured value is unavailable.

Considerable work along these lines has been reported. Most existing approaches to predicting viscosity involve experimentally determined physical properties such as pressure, volume, surface tension, density, refractive index, vapor pressure, latent heat of evaporation, molecular weight and critical temperature. Viscosity models ranging from highly theoretical to entirely empirical have been reviewed by Reid *et al.*[1] and Monnery *et al.*[2] These models can be categorized as theoretical, semi-theoretical or empirical and further distinguished as predictive and correlative. For example, Monnery *et al.*[3] developed a theoretically based predictive model for calculating both liquid- and gas-phase viscosities by modifying a statistical mechanics viscosity model. Gas and liquid viscosities for a variety of 174 non-polar, polar and hydrogen-bonding compounds are correlated with average deviations of 0.5% and 1.8% and predicted with average deviations of 2.2% and 6.1%, respectively. Their work made extensive comparisons with the best recommended models determined from their review.[2] However, no clear understanding of the structure–property relationship for liquid viscosity has been attained.

An alternative approach for predicting viscosities of compounds is that of quantitative structure–property relationships (QSPR).[4,5] QSPR models are obtained on the basis of the correlation between the experimental values of the property and descriptors reflecting the molecular structure of the respective compounds. Rigorous testing of the predictive power of the equations obtained is possible. Hence the QSPR approach is a general and reliable method to predict various physicochemical properties. Suzuki *et al.*[6] used a combination of partial least-squares (PLS) and QSPR techniques to search for correlations between viscosity and 18 parameters, mostly experimental. They found five significant PLS components and proposed a nine-descriptor model for a set of 237 compounds which gave a squared correlation coefficient (R^2) of 0.92. Suzuki *et al.*[7] used the same set to develop further this nine-descriptor QSPR model for liquid viscosity using multiple linear regression (MLR) and a two-layer neural network (NN) technique. Their non-linear model had $R^2 = 0.93$ and a root-mean-square error of 0.16 log units for a prediction

**Correspondence to:* A. R. Katritzky, Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, P.O. Box 117200, Gainesville, Florida 32611-7200, USA.

Copyright © 2000 John Wiley & Sons, Ltd. CCC 0894–3230/2000/010080–07 $17.50

Table 1. Comparison of the experimental and predicted viscosities for the set of 361 compounds[a]

Compound	Calc. log η	Exp. log η	Compound	Calc. log η	Exp. log η
Pentane	−0.514	−0.620	1-Methyl-4-ethylbenzene	0.056	−0.160
Cyclopentane	−0.432	−0.358	Isopropylbenzene	0.055	−0.102
2-Methylpentane	−0.433	−0.509	Butylbenzene	0.136	0.015
2,2-Dimethylbutane	−0.411	−0.438	1-Isopropyl-4-methylbenzene	0.161	0.532
2,3-Dimethylbutane	−0.410	−0.425	Tetralin	0.231	0.343
Cyclohexane	−0.324	−0.009	Amylbenzene	0.240	0.124
Heptane	−0.346	−0.379	1-Phenyloctane	0.478	0.408
3-Methylhexane	−0.302	−0.430	1-Phenylnonane	0.611	0.496
2,4-Dimethylpentane	−0.314	−0.443	1-Phenylundecane	0.747	0.662
2,2,3-Trimethylbutane	−0.302	−0.237	1-Phenyldodecane	0.759	0.736
2,3-Dimethylpentane	−0.292	−0.392	1-Phenyltetradecane	0.990	0.884
Methylcyclohexane	−0.232	−0.134	1-Phenylpentadecane	1.066	0.954
2,2,3-Trimethylpentane	−0.191	−0.223	1-Methylnaphthalene	0.299	0.560
2,2,4-Trimethylpentane	−0.207	−0.299	Iodomethane	−0.580	−0.301
2-Methylheptane	−0.232	−0.283	Dichloromethane	−0.561	−0.371
3-Ethylhexane	−0.211	−0.343	Bromochlorobutane	−0.387	−0.174
Propylcyclopentane	−0.171	−0.168	Chloroform	−0.507	−0.237
Nonane	−0.174	−0.148	Bromoethane	−0.456	−0.401
Propylcyclohexane	−0.054	0.001	1,1-Dichloroethane	−0.473	−0.446
trans-1,3,5-Trimethylcyclohexane	−0.055	−0.146	1,2-Dibromoethane	−0.009	0.236
Decane	−0.088	−0.033	1,1,2-Trichloroethane	−0.294	0.044
cis-Decalin	0.143	0.529	1,1,2,2-Tetrabromoethane	1.425	0.991
Butylcyclohexane	0.029	0.117	Pentachloroethane	0.135	0.389
Undecane	0.000	0.068	1-Chloropropane	−0.493	−0.454
n-Hexylcyclopentane	0.075	0.274	1-Bromopropane	−0.349	−0.281
n-Heptylcyclopentane	0.187	0.373	1,2-Dichloropropane	−0.304	−0.063
n-Hexylcyclohexane	0.215	0.344	1,2,3-Trichloropropane	−0.126	0.406
n-Octylcyclopentane	0.236	0.464	1-Bromobutane	−0.258	−0.199
n-Heptylcyclohexane	0.304	0.447	2-Bromobutane	−0.226	0.157
n-Nonylcyclopentane	0.345	0.550	1-Chloro-2-methylpropane	−0.398	−0.335
n-Octylcyclohexane	0.395	0.544	1-Bromo-2-methylpropane	−0.221	−0.192
n-Decylcyclopentane	0.468	0.550	1-Chloropentane	−0.333	−0.237
Hexadecane	0.387	0.524	1,1-Dichloroethylene	−0.557	−0.446
n-Undecylcyclopentane	0.494	0.631	*trans*-1,2-Dichloroethylene	−0.414	−0.394
n-Decylcyclohexane	0.568	0.719	Trichloroethylene	−0.364	−0.247
n-Dodecylcyclopentane	0.602	0.708	Allyl chloride	−0.522	−0.492
n-Undecylcyclohexane	0.653	0.801	Fluorobenzene	−0.098	−0.223
n-Dodecylcyclohexane	0.730	0.876	Bromobenzene	0.034	0.054
n-Tetradecylcyclopentane	0.778	0.852	Iodobenzene	0.148	0.223
n-Pentadecylcyclopentane	0.883	0.919	1,2-Dichlorobenzene	0.109	0.155
n-Hexadecylcyclopentane	0.936	0.982	2-Fluorotoluene	0.023	−0.168
l-Pentene	−0.531	−0.706	4-Fluorotoluene	−0.011	−0.206
trans-2-Pentene	−0.569	−0.785	Methanol	0.118	−0.224
2-Methyl-1-butene	−0.547	−0.650	Ethanol	0.137	0.079
Cyclopentene	−0.549	−0.461	Ethylene glycol	1.244	1.299
cis-3-Hexene	−0.471	−0.564	1-Propanol	0.238	0.349
trans-3-Hexene	−0.452	−0.564	Glycerol	2.360	3.173
cis-2-Hexene	−0.431	−0.564	1-Butanol	0.348	0.470
Cyclohexene	−0.342	−0.187	2-Methyl-1-propanol	0.368	0.606
1-Octene	−0.261	−0.328	1,3-Butanediol	1.460	2.115
trans-2-Octene	−0.258	−0.320	2-Methyl-2-butanol	0.486	0.666
1-Nonene	−0.160	−0.208	2-Methyl-1-butanol	0.458	0.740
1-Undecene	−0.017	0.013	1,5-Pentanediol	1.369	2.107
1-Dodecene	0.090	0.114	Cyclohexanol	0.672	1.833
1-Tetradecene	0.258	0.301	2-Methyl-2,4-pentanediol	1.534	1.537
1-Pentadecene	0.347	0.391	3-Ethyl-3-pentanol	0.689	0.829
1-Heptadecene	0.523	0.556	2-Ethyl-1-hexanol	0.723	0.991
1-Octadecene	0.598	0.635	1-Nonanol	0.795	1.155
2-Methyl-1,3-butadiene	−0.540	−0.652	Allyl alcohol	0.262	0.135
Toluene	−0.143	−0.229	2-Propyn-1-ol	0.284	0.225
Ethylbenzene	−0.043	−0.169	Benzyl alcohol	0.818	0.747
m-Xylene	−0.024	−0.208	Acetaldehyde	−0.399	−0.613
p-Xylene	−0.048	−0.188	Propionaldehyde	−0.192	−0.394
1,3,5-Trimethylbenzene	0.104	0.062	Isobutyraldehyde	−0.128	−0.246

Copyright © 2000 John Wiley & Sons, Ltd.

Table 1. *continued*

Compound	Calc. log η	Exp. log η	Compound	Calc. log η	Exp. log η
2-Propenal	−0.071	−0.463	Methyl acetoacetate	0.278	0.232
Benzaldehyde	0.389	0.161	Triethylene glycol	1.887	1.690
2-Butanone	−0.355	−0.369	2,2-Dimethyl-1,3-dioxolane-4-methanol	0.911	1.041
3-Pentanone	−0.241	−0.321			
Cyclopentanone	−0.136	0.067	Tetrahydropyran-2-methanol	0.911	1.041
2,4-Pentanedione	0.366	−0.222	Tripropylene glycol	2.189	1.749
4-Heptanone	−0.039	−0.133	Eugenol	1.335	0.965
2,6-Dimethyl-4-heptanone	0.181	0.013	2,2,2-Trifluoroethanol	0.528	0.300
Acetophenone	0.238	0.258	2,2′-Dichloroethyl ether	0.010	0.382
Acetic acid	0.374	0.091	1,2-Ethanediamine	0.600	0.188
Butyric acid	0.515	0.188	Isopropylamine	−0.154	−0.419
Pentanoic acid	0.624	0.350	Butylamine	−0.052	−0.167
Hexanoic acid	0.701	0.505	Diethylamine	−0.064	−0.450
2-Ethylbutyric acid	0.740	0.519	Piperidine	0.160	0.172
Heptanoic acid	0.760	0.640	Triethylamine	−0.056	−0.444
Octanoic acid	0.854	0.766	Cyclohexylamine	0.264	0.221
Acrylic acid	0.526	0.114	Dibutylamine	0.256	−0.022
Methacrylic acid	0.591	0.121	Aniline	0.349	0.644
Oleic acid	1.626	1.589	*m*-Toluidine	0.450	0.581
Methyl acetate	−0.285	−0.419	*o*-Toluidine	0.446	0.643
Ethyl acetate	−0.226	−0.342	*N*,*N*-Dimethylaniline	0.707	0.149
Propyl formate	−0.175	−0.241	*N*-Butylaniline	0.707	0.536
Methyl propionate	−0.244	−0.322	2-Methylpyridine	−0.040	−0.094
Propyl acetate	−0.163	−0.233	4-Methylpyridine	−0.035	−0.045
Ethyl propionate	−0.191	−0.270	4-*tert*-Butylpyridine	0.263	0.175
2-Methylpropyl formate	−0.145	−0.168	Acetonitrile	−0.635	−0.426
Methyl isobutyrate	−0.181	−0.282	Propionitrile	−0.553	−0.372
Butyl acetate	−0.125	−0.136	4-Methylpentanenitrile	−0.207	−0.009
Propyl propionate	−0.092	−0.172	Acrylonitrile	−0.562	−0.456
1,2-Ethanediol diacetate	0.208	0.496	*trans*-3-Butenenitrile	−0.429	−0.289
Methyl pentanoate	−0.109	−0.147	Ethanolamine	0.909	1.366
Diethyl malonate	0.307	0.332	*o*-Phenetidine	0.731	0.784
Propyl butyrate	−0.006	−0.080	*N*-Methylpropionamide	0.511	0.783
2-Methylbutyl acetate	0.001	−0.059	*N*,*N*-Dimethylacetamide	0.222	0.331
Isoamyl acetate	−0.026	−0.060	Methyl cyanoacetate	0.623	0.446
2-Ethylhexyl acetate	0.272	0.176	Nitroethane	−0.463	−0.169
Dibutyl sebacate	1.272	0.956	Nitrobenzene	−0.015	0.308
Vinyl formate	−0.272	−0.444	2-Nitrotoluene	0.118	0.375
Vinyl acetate	−0.270	−0.376	3-Bromoaniline	0.677	0.833
Methyl methacrylate	−0.151	−0.199	Ethanethiol	−0.327	−0.602
Dimethyl maleate	0.220	0.549	Dimethyl sulfide	−0.420	−0.539
Dibutyl maleate	0.709	0.751	Ethyl methyl sulfide	−0.285	−0.428
Ethyl benzoate	0.339	0.350	Diethyl sulfide	−0.197	−0.351
Butyl benzoate	0.506	0.493	Tetrahydrothiophene	−0.172	0.018
Diisobutyl *o*-phthalate	1.384	1.477	Thiophene	−0.232	−0.179
Butyl decyl *o*-phthalate	1.759	1.740	2-Mercaptoethanol	0.693	0.532
Ethyl cinnamate	0.471	0.940	2-Methylbutane	−0.516	−0.648
Dimethoxymethane	−0.121	−0.488	Hexane	−0.431	−0.487
1,4-Dioxane	−0.033	0.079	3-Methypentane	−0.416	−0.523
Ethyl propyl ether	−0.297	−0.491	Methylcyclopentane	−0.351	−0.295
Butyl ethyl ether	−0.230	−0.376	2-Methylhexane	−0.316	−0.423
Diisopropyl ether	−0.221	−0.467	2,2-Dimethylpentane	−0.317	−0.420
Dibutyl ether	−0.064	−0.161	Ethylcyclopentane	−0.262	−0.248
1,2-Bis(methoxyethoxy)ethane	0.548	0.575	Octane	−0.214	−0.266
Diisopentyl ether	0.109	0.005	3,3-Dimethylhexane	−0.212	−0.301
Dihexyl ether	0.238	0.232	Ethylcyclohexane	−0.135	−0.074
Ethyl vinyl ether	−0.423	−0.699	Butylcyclopentane	−0.087	−0.052
Methoxybenzene	0.061	0.121	*trans*-Decalin	0.150	0.328
Dibenzyl ether	0.763	0.727	*n*-Amylcyclopentane	0.000	0.061
1,2-Epoxybutane	−0.369	−0.387	*n*-Amylcyclohexane	0.117	0.235
Tetrahydrofuran	−0.281	−0.260	Dodecane	0.067	0.178
Maleic anhydride	−0.166	0.647	Tridecane	0.161	0.275
2-Methoxyethanol	0.661	0.236	Tetradecane	0.271	0.339
Diethylene glycol	1.565	1.553	Pentadecane	0.319	0.457

Copyright © 2000 John Wiley & Sons, Ltd.

Table 1. *continued*

Compound	Calc. log η	Exp. log η	Compound	Calc. log η	Exp. log η
n-Nonylcyclohexane	0.476	0.635	Cyclohexanone	−0.064	0.347
Heptadecane	0.505	0.573	Propionic acid	0.435	0.042
n-Tridecylcyclopentane	0.736	0.781	Isobutyric acid	0.488	0.118
n-Tridecylcyclohexane	1.052	0.949	2-Methylbutyric acid	0.561	0.382
cis-2-Pentene	−0.539	−0.695	2-Ethylhexanoic acid	0.847	0.887
2-Methyl-2-butene	−0.547	−0.686	Nonanoic acid	0.917	0.920
1-Hexene	−0.441	−0.585	Methyl formate	−0.319	−0.458
trans-2-Hexene	−0.431	−0.564	Ethyl formate	−0.277	−0.396
1-Heptene	−0.342	−0.456	Acetic anhydride	−0.189	−0.042
1-Decene	−0.079	−0.094	Methyl butyrate	−0.186	−0.237
1-Tridecene	0.194	0.212	Isopropyl acetate	−0.200	−0.245
1-Hexadecene	0.446	0.476	Ethyl butyrate	−0.125	−0.173
1,5-Hexadiene	−0.323	−0.561	2-Methylpropyl acetate	−0.075	−0.157
Benzene	−0.237	−0.186	Pentyl acetate	−0.014	−0.034
o-Xylene	−0.021	−0.092	Ethyl pentanoate	−0.039	−0.072
1,2,4-Trimethylbenzene	0.082	0.048	Propyl isobutyrate	−0.018	−0.080
Propylbenzene	0.055	−0.068	Ethyl acrylate	−0.149	−0.210
sec-Butylbenzene	0.152	0.019	Diethyl maleate	0.436	0.553
1-Phenylhexane	0.301	0.223	Methyl benzoate	0.249	0.311
1-Phenylheptane	0.390	0.316	Dibutyl *o*-phthalate	1.361	1.299
1-Phenyldecane	0.634	0.579	Bis(2-ethylhexyl) *o*-phthalate	2.347	1.911
1-Phenyltridecane	0.876	0.812	Diethyl ether	−0.415	−0.611
Styrene	−0.079	−0.124	Tetrahydropyran	−0.190	−0.083
α-Methylstyrene	0.036	−0.099	1,2-Diethoxyethane	0.045	−0.187
Dibromomethane	−0.165	0.008	Bis(2-methoxyethyl) ether	0.294	0.299
Carbon tetrachloride	−0.411	−0.014	Butyl vinyl ether	−0.240	−0.301
Iodoethane	−0.412	−0.228	Ethoxybenzene	0.148	0.093
1,1,2-Trichlorotrifluoroethane	−0.008	−0.148	Propylene oxide	−0.427	−0.485
2-Chloropropane	−0.502	−0.398	Furan	−0.279	−0.420
1-Chlorobutane	−0.419	−0.339	2-Ethoxyethanol	0.701	0.312
2-Chlorobutane	−0.410	−0.385	4-Hydroxy-4-methyl-2-pentanone	0.984	0.462
2-Chloro-2-methylpropane	−0.399	−0.291	2-Hydroxybenzaldehyde	1.429	0.462
1-Iodo-2-methylpropane	−0.106	−0.058	Trifluoroacetic acid	0.573	−0.033
cis-1,2-Dichloroethylene	−0.414	−0.331	Benzoyl bromide	0.238	0.291
Tetrachloroethylene	−0.167	−0.053	Pentylamine	0.025	0.008
Chlorobenzene	−0.109	−0.098	Tributylamine	0.459	0.189
1,3-Dichlorobenzene	0.073	0.036	Pyrrole	0.207	0.131
3-Fluorotoluene	0.010	−0.216	Methylphenylamine	0.441	0.373
Benzyl chloride	0.024	0.146	*N*, *N*-Diethylaniline	0.540	0.339
1,2-Propanediol	1.345	1.660	Pyridine	−0.287	−0.011
1,3-Propanediol	1.265	1.748	2,4,6-Trimethylpyridine	0.174	0.176
2-Butanol	0.367	0.592	Butyronitrile	−0.411	−0.210
1-Pentanol	0.389	0.602	Methacrylonitrile	−0.555	−0.407
3-Methyl-1-butanol	0.439	0.708	Morpholine	0.377	0.348
1-Hexanol	0.477	0.702	*p*-Phenetidine	0.703	1.111
1-Heptanol	0.621	0.846	Nitromethane	−0.530	−0.188
1-Octanol	0.623	0.951	3-Nitrotoluene	0.100	0.367
1-Decanol	0.858	1.158	1-Butanethiol	−0.195	−0.300
m-Cresol	0.832	1.265	Thiophenol	0.228	0.093
Butyraldehyde	−0.096	−0.342	Trimethylene sulfide	−0.271	−0.195
Pentanal	0.007	−0.290	2,2′-Thiodiethanol	1.781	1.814
Acetone	−0.502	−0.498			
2-Pentanone	−0.218	−0.296			
4-Methyl-2-pentanone	−0.129	−0.233			

[a] Values of G_I and HDCA(2) for these compounds are available on request.

set of 124 compounds. However, both of these models include four experimental physical properties as molecular descriptors. Although efficient fragment methods are available to estimate molar refraction, critical temperature and cohesive energy, it is more difficult to deal with the molar magnetic susceptibility. Obviously, the models are most useful for compounds for which these experimental values are available.

In recent years, the CODESSA[8] QSPR software has been successfully applied to the prediction of a variety of

Copyright © 2000 John Wiley & Sons, Ltd.

Table 2. The best five-descriptor correlation model of viscosity (log η) for the data set of 361 diverse structures

N	Descriptor	$X + \Delta X$	t-Test	R^2	R_{cv}^2	F
0	Intercept	-10.3 ± 1.7	-6.06			
1	HDCA(2)	1.77 ± 0.08	23.0	0.339	0.330	184.3
2	G_I	0.000557 ± 0.000016	34.2	0.811	0.804	766.2
3	N_{rings}	2.78 ± 0.41	6.78	0.830	0.822	579.0
4	FPSA(3)	20.2 ± 3.1	6.58	0.843	0.836	477.7
5	E_{min}(C)	0.0897 ± 0.0165	5.44	0.854	0.844	414.1
		$R^2 = 0.854$, $F = 414.1$, $S = 0.22$, $R_{cv}^2 = 0.840$				

physical properties of compounds,[4,5] including boiling points,[9] critical temperatures,[10] vapor pressures,[11] glass transition temperatures of polymers[12] and critical micelle concentrations of surfactants.[13] CODESSA can generate a large number of quantitative descriptors of the constitutional, topological, geometrical, electrostatic, and quantum chemical characteristics of a molecule. The primary purpose of the present work was to develop a QSPR model of viscosity for diverse compounds utilizing the CODESSA descriptors derivable solely from the structures of compounds.

DATA AND METHODOLOGY

Data set

The experimental values of viscosities (η) were taken from the combination of the training set and prediction set of Ref. 7. A total of 361 structurally diverse organic compounds containing C, H, O, N, S and/or halogen atoms were included in the data set, with a range of experimental η values between 0.197 and 1490 mPa·s. Table 1 shows the compounds and their corresponding experimental viscosities as log η at 20 °C.

Computational methods

The quantitative structure–property relationship was developed using the Microsoft Windows version of the CODESSA program.[8] The program can calculate around 1000 descriptors, and the search for the best statistical relationship between calculated descriptors and experimental property data can be carried out using several techniques.[4,13] Structural optimization of the 361 compounds was performed using the semi-empirical quantum-chemical AM1 parameterization[14] with MOPAC 6.0 program.[15] The MOPAC output files were loaded into the CODESSA program along with the experimental log η to calculate molecular structural descriptors,[4] including constitutional, topological, geometrical, electrostatic and quantum chemical descriptors. Altogether over 600 molecular descriptors were calculated for each of the 361 compounds studied. Then the statistical analysis to find the best QSPR model was carried out using a heuristic method based on the linear regression technique. This procedure is based on the scale forward selection technique[16] and has been described in detail elsewhere.[17] In addition, the data were treated by best regression method[8] (see Table 4) by using a polynomial expansion, and by application of a neural network approach, but none of these alternative techniques improved the quality of the final regression equation.

RESULTS AND DISCUSSION

The best five-parameter correlation equation obtained for the entire data set of 361 compounds had $R^2 = 0.854$ and is presented in detail in Table 2. The cross-validated correlation coefficient $R_{cv}^2 = 0.840$, as compared with the correlation coefficient R^2, indicates the stability of the regression equation. Calculated log η values are given in Table 1. According to the t-test, the most important descriptor is the hydrogen-bonding donor charged surface area, HDCA(2), defined as

$$\mathrm{HDCA}(2) = \sum \frac{q_D \sqrt{S_D}}{S_{tot}} \tag{1}$$

where q_D is the partial charge on the hydrogen-bonding donor (H) atom (s), S_D denotes the exposed surface area of this atom which is defined as that part of the total surface area contributed by the respective atom and S_{tot} is the total molecular surface area calculated from the van der Waals' radii[18] of the atoms (overlapping spheres). The summation in the formulation of HDCA(2) [Eqn. (1)] is performed over all simultaneously possible hydrogen bonding donor and acceptor pairs per solute molecule. Also, hydrogen atoms attached to carbons connected directly to a carbonyl or cyano group were included as possible hydrogen bonding donor centers (their effectiveness is, of course, much smaller because of the smaller partial charge on them).

Copyright © 2000 John Wiley & Sons, Ltd.

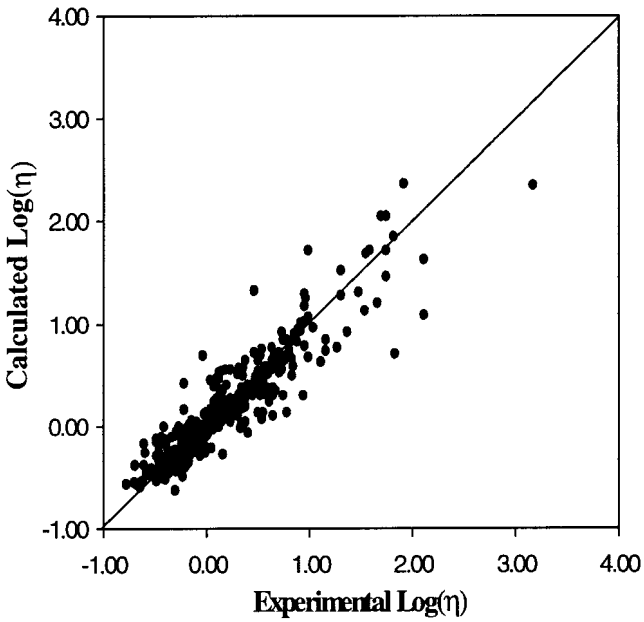

Figure 1. Scatter plot of the calculated versus experimental log η values for 361 compounds

Table 3. Verification of statistical validity

Set	R^2	R_{cv}^2	S
I	0.838	0.825	0.22
II	0.860	0.844	0.23
III	0.864	0.852	0.22

The second most important descriptor in this regression is the gravitational index (G_I) over all bonded atoms i, j in the molecule, defined as[19]

$$G_I = \sum_{ij}^{\text{all bonded atoms}} (m_i m_j / r_{ij}^2) \tag{2}$$

where m_i and m_j are the atomic masses of the bonded atoms and r_{ij} denotes the respective bond lengths. As reported in our previous study on the normal boiling point[17] and the solubility of gases and vapors in water,[19] the gravitational index reflects the effective mass distribution in the molecule and effectively describes the intermolecular dispersion forces in the bulk liquid media. The combination of the two descriptors HDCA(2) and G_I adequately represents the forces of intermolecular attraction; HDCA(2) is related to the hydrogen-bonding ability of compounds and G_I with the dispersion and cavity-formation effects in liquids. The two descriptors are almost orthogonal because the intercorrelation coefficients between them is lower than 0.1.

Descriptor N_{rings}, the relative number of rings in the molecule, is related to the size and shape of the molecule as defined in the CODESSA Manual.[8] Descriptor FPSA(3) is the fractional positive partial charged surface area and is defined as the ratio between the positively charged surface area and total surface area of the molecule.[20] FPSA(3) describes electrostatic polar interactions between molecules and is calculated by the equation

$$\text{FPSA}(3) = \frac{\sum_{A}^{N_A} S_A q_A}{S} \tag{3}$$

where S_A is the surface area of a positively charged atom, q_A is the corresponding positive atomic charge, N_A is the number of the positively charged atoms in the molecule and S is the total surface area of the molecule. The charges in HDCA(2) and FPSA(3) are from quantum chemical calculation (for definition, see Ref. 8). The descriptor E_{min} (C) is defined as the minimum atomic state energy for a C atom. The state energy characterizes the magnitude of the perturbation experienced by a C atom in the molecular environment in comparison with the isolated atom. The last three descriptors probably all reflect intermolecular interactions; however, these three descriptors have much less impact than the hydrogen-bonding related descriptor HDCA(2) and the molecular mass descriptor G_I. The plot of the calculated versus experimental log η of the complete set using the five-descriptor model is shown in Fig. 1. The standard error S is 0.22 log units.

In order to demonstrate further the absence of chance correlations, the whole data set was divided into three subsets (by using numbers 1, 4, 7, …; 2, 5, 8, …; and 3, 6,

Table 4. Correlation obtained by the multiregression approach

N	Descriptor	$X + \Delta X$
0	Intercept	-0.92965 ± 0.02865
1	G_I (all bonds)	$(120.59 \pm 3.41) \times 10^{-5}$
2	HA-dependent HDCA(2) (Zefirov's PC)	1.80371 ± 0.11448
3	Maximum electron–electron repulsion for an H—O bond	0.37779 ± 0.03487
4	Minimum electron–nucleus attraction for an H—O bond	-0.16765 ± 0.01550

$R^2 = 0.861$, $S = 0.22$ (with $N = 361$ in denominator), $F = 554.4$
Cross-validation (leave-one out): $R_{cv}^2 = 0.852$, $S_{cv} = 0.22$

Copyright © 2000 John Wiley & Sons, Ltd.

9, …) and each subset was predicted by using the other two subsets as the training set. In this procedure, the same descriptors were retained, but the coefficients were allowed to vary. The results are shown in Table 3, which indicates that the model has a high statistical stability and validity.

While the present work was in progress, Ivanciuc *et al.*[21] independently developed a five-descriptor QSPR model for liquid viscosity with 337 compounds using the CODESSA program. Their data set differs somewhat from ours, but possesses about the same diversity. The five descriptors in Ivanciuc's *et al.* model are molecular weight, Randic connectivity index of order 3, hydrogen-donor charged surface area HDCA(2), maximum electrophilic reactivity index for a C atom, and maximum atomic orbital electronic population. The squared correlation coefficient and standard error are $R^2 = 0.846$ and $S = 0.27$, respectively. Using these five descriptors to correlate our data set, the correlation coefficient and standard error are $R^2 = 0.854$ and $S = 0.22$, which are the same as those obtained by our present model. Since completion of the work described in the main body of this paper, we have applied the recently developed novel technique of descriptor selection based on multiregression.[22,23] Using this approach, the best possible multilinear regression model for liquid viscosity containing four descriptors is selected among all the descriptors from data set. This model possesses better statistical parameters [$R^2 = 0.861$, $S = 0.215$, $R_{cv}^2 = 0.852$, S_{cv} (leave-one-out) = 0.223], than our five-parameter correlation, emphasizing the importance of descriptor selection in the modeling process. The descriptors involved in the model are: gravitation index (all bonds), HA-dependent HDCA(2) (Zefirov's PC), maximum electron–electron repulsion for an H—O bond and minimum electron–nucleus attraction for an H—O bond (see Table 4). This work is still in progress.

CONCLUSIONS

A quantitative structure–property relationship approach was used to predict the liquid viscosity of organic compounds for a diverse set of 361 compounds. A general five-descriptor model was developed with a squared correlation coefficient of 0.854 and a standard error of 0.22 log units. The most important descriptor is the hydrogen-bonding donor charged surface area, which reflects that hydrogen bonding is a key factor to control liquid viscosity.

In comparison with previous linear and non-linear models,[7] our present QSPR model and the essentially equivalent model of Ivanciuc *et al.*,[21] are more general and practical methods for predicting liquid viscosities.

Acknowledgements

We thank Dr O. Ivanciuc (Bucharest) for informing us of his parallel work in this area and for discussions.

REFERENCES

1. R. C. Reid, J. M. Prausnitz and B. E. Poling, *The Properties of Gases and Liquids*, 4th ed. McGraw-Hill, New York (1987).
2. W. D. Monnery, W. Y. Svrcek and A. K. Mehrotra, *Can. J. Chem. Eng.* **73**, 3–40 (1995).
3. W. D. Monnery, A. K. Mehrotra and W. Y. Svrcek, *Ind. Eng. Chem. Res.* **37**, 652–659 (1998).
4. A. R. Katritzky, V. S. Lobanov and M. Karelson, *Chem. Soc. Rev*. 279–287 (1995).
5. A. R. Katritzky, M. Karelson and V. S. Lobanov, *Pure Appl. Chem.* **69**, 245–248 (1997).
6. T. Suzuki, K. Ohtaguchi and K. Koide, *Comput. Chem. Eng.* **20**, 161–173 (1996).
7. T. Suzuki, R.-U. Ebert and G. Schüürmann, *J. Chem. Inf. Comput. Sci.* **37**, 1122–1128 (1997).
8. A. R. Katritzky, V. S. Lobanov and M. Karelson, *CODESSA Version 2.0 Reference Manual*. (1994).
9. A. R. Katritzky, V. S. Lobanov and M. Karelson, *J. Chem. Inf. Comput. Sci.* **38**, 28–41 (1998).
10. A. R. Katritzky, L. Mu and M. Karelson, *J. Chem. Inf. Comput. Sci.* **38**, 293–299 (1998).
11. A. R. Katritzky, Y. Wang, S. Sild, T. Tamm and M. Karelson, *J. Chem. Inf. Comput. Sci.* **38**, 720–725 (1998).
12. A. R. Katritzky, S. Sild, V. S. Lobanov and M. Karelson, *J. Chem. Inf. Comput. Sci.* **38**, 300–304 (1998).
13. P. D. T. Huibers, V. S. Lobanov, A. R. Katritzky, D. O. Shah and M. Karelson, *Langmuir* **12**, 1462–1470 (1996).
14. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *J. Am. Chem. Soc.* **107**, 3902–3909 (1985).
15. J. J. P. Stewart, MOPAC 6.0. QCPE No. 455 (1990).
16. N. R. Draper and H. Smith, *Applied Regression Analysis*. Wiley, New York (1966).
17. A. R. Katritzky, L. Mu, V. S. Lobanov and M. Karelson, *J. Phys. Chem.* **100**, 10400–10407 (1996).
18. R. S. Pearlman, In *Molecular Surface Areas and Volumes and Their Use in Structure Activity Relationships*, edited by S. H. Yalkowsky, A. A. Sinkula and S. C. Valvani. Marcel Dekker, New York (1979).
19. A. R. Katritzky, L. Mu, V. S. Lobanov and M. Karelson, *J. Chem. Inf. Comput. Sci.* **36**, 1162–1168 (1996).
20. D. T. Stanton and P. C. Jurs, *Anal. Chem.* **62**, 2323–2329 (1990).
21. O. Ivanciuc, T. Ivanciuc, P. A. Filip and D. Cabrol-Bass, *J. Chem. Inf. Comput. Sci.* submitted for publication.
22. B. Lucic and N. Trinajstic, *J. Chem. Inf. Comput. Sci.* **39** 121–132 (1999).
23. B. Lucic, N. Trinajstic, S. Sild, M. Karelson and A. R. Katritzky, *J. Chem. Inf. Comput. Sci.* **39**, 610–621 (1999).

Copyright © 2000 John Wiley & Sons, Ltd.