

ARTICLE

Bond Orbital-Connection Matrix Method to Predict Refractive Indices of AlkanesChen-zhong Cao^{a*}, Shuo Gao^{a,b}

a. School of Chemistry and Chemical Engineering, Hunan University of Science and Technology, Xiangtan 411201, China; b. School of Chemistry and Chemical Engineering, Central South University, Changsha 410083, China

(Dated: Received on July 21, 2006; Accepted on November 13, 2006)

Based on the comprehension of the specific structural features affecting the refractive indices of the compounds, two descriptors $\sum X_{1CC}$ and $\sum X_{1CH}$ extracted from the bond orbital-connection matrix (BOCM) method were employed to develop a QSPR model for predicting the refractive indices of alkanes, chloroalkanes and bromoalkanes. The obtained results confirmed the usefulness of the BOCM method. $\sum X_{1CC}$ and $\sum X_{1CH}$ reflect the ability of the electronic cloud of the alkanes to be polarized; such ability is correlated with the refractive indices of substances. Therefore, the physical meaning of the obtained model can be rationally interpreted from the physical point of view. The present descriptors obtained by the BOCM method have the merit of topological indices (i.e. facility and rapid calculation of the descriptors) and the advantage of quantum descriptors (i.e. explicitly physical meaning of the parameters), which lead to an expectation of wide use in QSAR studies.

Key words: Structural descriptor, Bond orbital-connection matrix, Refractive index, Alkane, Haloalkane

I. INTRODUCTION

The molecular structure of a compound predetermines its physical properties. Quantitative structure-property/activity relationship (QSPR/QSAR) method can provide us both a better understanding of the compounds' specific behaviors from the molecule level and a direction for more efficient exploitation of new drugs and special function materials. Thus the construction of QSPR/QSAR models has received considerable attention in past years [1]. However, a chemical structure cannot be used directly in a mathematical equation to predict physicochemical properties and to overcome this difficulty, so it is necessary to derive molecular descriptors. In general, the first as well as the crucial step in QSPR/QSAR studies is to effectively extract sufficient chemical information with numerical format from molecular structures for building QSPR/QSAR models [2]. For this target, many methods have been developed [3-10]: quantum theory method, molecular topology method, and so on. Quantum theory method has very explicit physical meaning, which can provide one a physical insight in the QSPR/QSAR studies [11]. However, the application of high-level quantum calculation is still constrained by the sizes of the molecules under study. Owing to the facile calculation and high efficiency, molecular topology method has shown its vigor in the QSPR/QSAR field. Since Wiener's seminal work [12] in molecular topology, rapid development has been made and almost 400 topological in-

indices have been put forth to correlate the properties of compounds. Topological descriptors and graph theory approaches are widely used in chemistry, but they are always criticized for the lack of structural interpretation of the obtained models [2], although attempts have been made to address this issue [13-15]. It should be noted that some excellent QSAR/QSPR program packages, such as CODESSA [16] and Dragon [17], can offer researchers a very large structural descriptor pool from which one can choose parameters to build QSPR/QSAR models. These program packages are assuredly useful in QSPR/QSAR studies. However, some scholars still think that the vast amount of descriptors generated by these program packages are sometimes not enough to describe physicochemical or biological properties, and the necessity for incorporation of new descriptors is evident [13,18]. What is more, one of the drawbacks of those software packages is that they build QSAR models and select descriptors from the large automatically generated descriptor pool solely reckoning on the statistical procedures and statistical criteria, regardless of the structural features related to the considered structure-property relationship. And it is unknown which and how many descriptors will be employed in a QSAR model before the software optimizes the correlation equation, which is not helpful in understanding the structure-property relationship.

The bond orbital-connection matrix (BOCM) method was proposed recently in our laboratory [19], combining the polarizability effect index (PEI) of alkyls [20] and the adjacency matrix of graph theory [21]. The merits of the bond orbital-connection matrix method is that all those structural descriptors extracted with this approach can be calculated quickly and each descriptor

* Author to whom correspondence should be addressed. E-mail: czcao@hnust.edu.cn, Tel.: +86-732-8291336

has a transparent physical meaning, which can give us an insight to understand the relationship between the molecular structures and their properties. The bond orbital-connection matrix method has been shown to be very effective in modeling bond dissociation energy, enthalpies of formation, boiling point and density of alkanes, the enthalpies of formation of monoalkenes and monosubstituted alkanes [19,22-25]. However, the potential of this method in QSAR/QSPR modeling has still to be further verified in different systems and for different properties of compounds.

The refractive index (n) is considered to be one of the most important optical properties and it is frequently employed to characterize organic compounds [26]. It has been used as an indicator of the purity of organic compounds, but the relationship of refractive index to other optical, electrical, and magnetic properties has more significance. The relationships between refractive index and polarizability, critical temperature, surface tension, density, and boiling point, among other properties, have been reviewed [27]. The refractive index is also widely used in material science to evaluate materials [28]. However, there are only a few QSPR models for prediction of refractive indices [29,30]. In the field of molecular simulation, refractive index is employed sometimes as a molecular descriptor [31]. Therefore, it is advantageous to develop quantitative structure-refractive index models for estimation of refractive index for the untested or not synthesized compounds.

In the present study, to further verify the usefulness of the bond orbital-connection matrix method in QSPR/QSAR studies, 128 alkanes whose carbon atoms ranged in 4 to 20 were chosen to construct the structure-refractive index models using the descriptors extracted by the bond orbital-connection matrix.

II. METHODOLOGY

A. Construction of estimated models

In the more recent work [24], it was found that the densities of alkanes could be estimated by Eq.(1)

$$D_{RH} = \frac{N_C}{a + g \sum X_{1CC} + k \sum X_{1CH}} \quad (1)$$

where a , g and k are coefficients; D_{RH} is the density of alkane RH, N_C is the number of carbon atoms in the alkane, and X_{1CC} and X_{1CH} are the eigenvalues of the C-C and C-H bond orbital-connection matrices [19], respectively. According to the physical principle that the refractive index of compound is proportional to its density, thus it was expected that the refractive index of the alkane n_{RH} could be estimated by Eq.(2)

$$n_{RH} = \frac{N_C}{a' + g' \sum X_{1CC} + k' \sum X_{1CH}} \quad (2)$$

where, a' , g' and k' are coefficients, and they can be obtained by regression analysis.

B. Calculation of molecular descriptors

The PEI value of an alkyl R can be easily calculated by the expression $PEI_R = \sum \Delta PEI_l$, in which ΔPEI_l is the contribution of the l th carbon atom in the alkyl R to its polarizability effect index PEI_R (for more details, see Refs.[20,32]). For the convenience of readers, the calculation of descriptors $\sum X_{1CC}$ and $\sum X_{1CH}$ [19,33] is briefly restated here.

Firstly, the alkyl's polarizability effect index PEI_{R_i} and PEI_{R_j} of the alkyl R_i and R_j which formed the alkane R_i-R_j were used as the main-diagonal element, and 1 was used as the off-diagonal element to construct the bond orbital-connection matrix CM_{ij} for C_i-C_j bond. Similarly, the PEI_{R_i} and PEI_H were used as the main-diagonal element, and 1 was used as the off-diagonal element to construct the bond orbital-connection matrix CM_{H_i} for C_i-H bond.

Then, the above matrices CM_{ij} and CM_{H_i} were determined. For each matrix CM_{ij} or CM_{H_i} , one can get two eigenvalues, and the smaller one of these two eigenvalues was assigned as X_{1CC} or X_{1CH} .

Finally, summing up all X_{1CC} (or X_{1CH}) values of a molecule, one got the parameters $\sum X_{1CC}$ (or $\sum X_{1CH}$) for the alkane molecule RH.

III. RESULTS AND DISCUSSION

In order to get the coefficients of Eq.(2), the format of Eq.(2) was transformed and the following formulation was obtained,

$$\frac{N_C}{n_{RH}} = a' + g' \sum X_{1CC} + k' \sum X_{1CH} \quad (3)$$

Taking some experimental refractive indices of alkanes [34] as the training data set (Table I), using Eq.(3) as the model, we carried out linear regression analysis and the following correlation expression was obtained:

$$\frac{N_C}{n_{RH}} = -0.87671 - 0.09638 \sum X_{1CC} - 0.70636 \sum X_{1CH} \quad (4)$$

Transforming the format of Eq.(4), yields Eq.(5):

$$n_{RH} = \frac{N_C}{-0.87671 - 0.09638 \sum X_{1CC} - 0.70636 \sum X_{1CH}} \quad (5)$$

$$r = 0.9999, RMSE = 0.0050, N = 128, F = 209652$$

where the statistics r , $RMSE$, F and N represent the correlation coefficient, the estimated root-mean-square error from the experimental refractive index values, the Fischer criterion and the sample numbers, respectively. In general, the quality of a model can be conveniently measured by the correlation coefficient r and $RMSE$ value. From the r and $RMSE$ value of the regression equation, one can say that the correlation of Eq.(5) is excellent. The average signed error, average unsigned error and $RMSE$ were -8.6×10^{-6} , 0.0035 and 0.0050

TABLE I Experimental and estimated refractive indices (at 101 kPa and 20 °C) for 128 alkanes

Alkane ^a	n_{exp}^b	$\sum X_{1\text{CC}}$	$\sum X_{1\text{CH}}$	n_{calc}^c	$n_{\text{exp}}-n_{\text{calc}}$	Alkane ^a	n_{exp}^b	$\sum X_{1\text{CC}}$	$\sum X_{1\text{CH}}$	n_{calc}^c	$n_{\text{exp}}-n_{\text{calc}}$
C5	1.3575	0.5296	-6.5340	1.3559	0.0016	25mmC8	1.4112	1.9952	-11.5282	1.4136	-0.0024
2mC4	1.3537	0.6219	-6.5407	1.3574	-0.0037	26mmC8	1.4097	1.9780	-11.5330	1.4126	-0.0029
C6	1.3749	0.7533	-7.5404	1.3708	0.0041	27mmC8	1.4086	1.9391	-11.5382	1.4111	-0.0025
3mC5	1.3765	0.8753	-7.5440	1.3737	0.0028	33mmC8	1.4165	2.1024	-11.5263	1.4160	0.0005
22mmC4	1.3688	1.0264	-7.5706	1.3724	-0.0036	34mmC8	1.4182	2.0581	-11.5196	1.4160	0.0022
23mmC4	1.3750	0.9654	-7.5496	1.3752	-0.0002	35mmC8	1.4139	2.0393	-11.5221	1.4153	-0.0014
C7	1.3878	0.9866	-8.5424	1.3828	0.0050	36mmC8	1.4139	2.0185	-11.5266	1.4143	-0.0004
2mC6	1.3848	1.0914	-8.5469	1.3847	0.0001	44mmC8	1.4144	2.1275	-11.5196	1.4174	-0.0030
3mC6	1.3887	1.1214	-8.5434	1.3862	0.0025	3eC8	1.4156	1.9220	-11.5227	1.4130	0.0026
22mmC5	1.3822	1.2794	-8.5515	1.3888	-0.0066	223mmmmC7	1.4168	2.2363	-11.5382	1.4169	-0.0001
24mmC5	1.3815	1.1998	-8.5486	1.3872	-0.0057	225mmmmC7	1.4101	2.1848	-11.5308	1.4169	-0.0068
33mmC5	1.3909	1.3200	-8.5480	1.3905	0.0004	226mmmmC7	1.4078	2.1431	-11.5376	1.4151	-0.0073
3eC5	1.3934	1.1517	-8.5400	1.3876	0.0058	234mmmmC7	1.4195	2.2022	-11.5152	1.4195	0.0000
2mC7	1.3949	1.3335	-9.5454	1.3944	0.0005	235mmmmC7	1.4169	2.1672	-11.5210	1.4180	-0.0011
22mmC6	1.3935	1.5296	-9.5480	1.3986	-0.0051	236mmmmC7	1.4131	2.1216	-11.5287	1.4160	-0.0029
23mmC6	1.4011	1.4990	-9.5407	1.3991	0.0020	245mmmmC7	1.4160	2.1754	-11.5197	1.4183	-0.0023
25mmC6	1.3925	1.4418	-9.5482	1.3965	-0.0040	246mmmmC7	1.4071	2.1189	-11.5273	1.4161	-0.0090
33mmC6	1.4001	1.5865	-9.5405	1.4012	-0.0011	255mmmmC7	1.4149	2.2158	-11.5259	1.4182	-0.0033
34mmC6	1.4041	1.5284	-9.5370	1.4004	0.0037	334mmmmC7	1.4236	2.2176	-11.5110	1.4204	0.0032
3eC6	1.4018	1.4112	-9.5350	1.3980	0.0038	335mmmmC7	1.4170	2.2648	-11.5181	1.4203	-0.0033
223mmmmC5	1.4030	1.7023	-9.5556	1.4013	0.0017	344mmmmC7	1.4235	2.3170	-11.5096	1.4225	0.0010
224mmmmC5	1.3915	1.6445	-9.5487	1.4011	-0.0096	3e2mC7	1.4174	2.0741	-11.5162	1.4168	0.0006
233mmmmC5	1.4075	1.7215	-9.5399	1.4045	0.0030	4e2mC7	1.4137	2.0602	-11.5178	1.4163	-0.0026
234mmmmC5	1.4042	1.6247	-9.5409	1.4020	0.0022	5e2mC7	1.4134	2.0346	-11.5237	1.4150	-0.0016
3e2mC5	1.4040	1.5347	-9.5347	1.4010	0.0030	3e3mC7	1.4208	2.1826	-11.5132	1.4194	0.0014
3e3mC5	1.4078	1.6351	-9.5358	1.4032	0.0046	4e3mC7	1.4206	2.1204	-11.5095	1.4187	0.0019
C9	1.4054	1.4718	-10.5392	1.4006	0.0048	5e3mC7	1.4164	2.0804	-11.5952	1.4058	0.0106
22mmC7	1.4016	1.7802	-10.5425	1.4066	-0.0050	3e4mC7	1.4207	2.1159	-11.5106	1.4184	0.0023
23mmC7	1.4088	1.7312	-10.5341	1.4068	0.0020	4e4mC7	1.4210	2.2037	-11.5087	1.4204	0.0006
24mmC7	1.4034	1.7439	-10.5347	1.4070	-0.0036	4PrC7	1.4135	1.9584	-11.5144	1.4149	-0.0014
25mmC7	1.4033	1.8454	-10.5449	1.4076	-0.0043	4i-Pr C7	1.4153	2.0915	-11.7131	1.3898	0.0255
26mmC7	1.4011	1.8132	-10.5436	1.4071	-0.0060	2233mmmmmmC6	1.4282	2.5074	-11.5088	1.4263	0.0019
33mmC7	1.4087	1.8453	-10.5337	1.4093	-0.0006	2234mmmmmmC6	1.4216	2.4178	-11.5119	1.4241	-0.0025
34mmC7	1.4108	1.7963	-10.5288	1.4091	0.0017	2235mmmmmmC6	1.4142	2.3552	-11.5217	1.4215	-0.0073
35mmC7	1.4083	1.7692	-10.5330	1.4078	0.0005	2244mmmmmmC6	1.4208	2.4382	-11.5202	1.4233	-0.0025
44mmC7	1.4076	1.8620	-10.5300	1.4103	-0.0027	2245mmmmmmC6	1.4132	2.3378	-11.5244	1.4208	-0.0076
3eC7	1.4093	1.6668	-10.5296	1.4062	0.0031	2255mmmmmmC6	1.4055	2.3523	-11.5340	1.4197	-0.0142
4eC7	1.4096	1.6803	-10.5265	1.4069	0.0027	2334mmmmmmC6	1.4298	2.4615	-11.5143	1.4246	0.0052
223mmmmC6	1.4106	1.9732	-10.5326	1.4122	-0.0016	2335mmmmmmC6	1.4196	2.3895	-11.5158	1.4230	-0.0034
224mmmmC6	1.4033	1.9369	-10.5370	1.4108	-0.0075	2344mmmmmmC6	1.4267	2.4442	-11.5083	1.4252	0.0015
225mmmmC6	1.3997	1.8917	-10.5434	1.4088	-0.0091	2345mmmmmmC6	1.4204	2.3348	-11.5124	1.4224	-0.0020
233mmmmC6	1.4141	2.0003	-10.5284	1.4135	0.0006	3344mmmmmmC6	1.4368	2.5621	-11.5018	1.4284	0.0084
234mmmmC6	1.4144	1.9256	-10.5280	1.4119	0.0025	22mm3eC6	1.4197	2.2189	-11.5105	1.4205	-0.0008
235mmmmC6	1.4051	1.8696	-10.5346	1.4097	-0.0046	23mm3eC6	1.4270	2.3488	-11.5047	1.4238	0.0032
244mmmmC6	1.4074	1.9640	-10.5333	1.4119	-0.0045	23mm4eC6	1.4226	2.2483	-11.5085	1.4213	0.0013
334mmmmC6	1.4178	2.0291	-10.5253	1.4146	0.0032	25mm3eC6	1.4157	2.1932	-11.5159	1.4192	-0.0035
3e2mC6	1.4106	1.8099	-10.5265	1.4097	0.0009	33mm4eC6	1.4269	2.3590	-11.5038	1.4241	0.0028
4e2mC6	1.4063	1.7826	-10.5409	1.4069	-0.0006	34mm3eC6	1.4267	2.3772	-11.5016	1.4248	0.0019
3e3mC6	1.4140	1.9151	-10.5239	1.4123	0.0017	33eeC6	1.4258	2.2645	-11.5010	1.4227	0.0031
4e3mC6	1.4134	1.8390	-10.5231	1.4108	0.0026	34eeC6	1.4190	2.1626	-11.3856	1.4374	-0.0184

Continued Table I.

Alkane ^a	$n_{\text{exp}}^{\text{b}}$	$\sum X_{1\text{CC}}$	$\sum X_{1\text{CH}}$	$n_{\text{calc}}^{\text{c}}$	$n_{\text{exp}}-n_{\text{calc}}$	Alkane ^a	$n_{\text{exp}}^{\text{b}}$	$\sum X_{1\text{CC}}$	$\sum X_{1\text{CH}}$	$n_{\text{calc}}^{\text{c}}$	$n_{\text{exp}}-n_{\text{calc}}$
2234mmmmC5	1.4147	2.1052	-10.5625	1.4104	0.0043	3 <i>i</i> -Pr2mC6	1.4195	2.2279	-11.5102	1.4207	-0.0012
2244mmmmC5	1.4069	2.1074	-10.5396	1.4140	-0.0071	22344mmmmmmC5	1.4307	2.6037	-11.5120	1.4278	0.0029
2334mmmmC5	1.4222	2.1414	-10.5252	1.4170	0.0052	3e223mmmmC5	1.4420	2.5709	-11.4998	1.4289	0.0131
23mm3eC5	1.4221	2.0373	-10.5208	1.4155	0.0066	3e234mmmmC5	1.4333	2.4966	-11.4999	1.4274	0.0059
24mm3eC5	1.4131	1.9431	-10.5428	1.4100	0.0031	33ee2mC5	1.4343	2.4130	-11.4958	1.4264	0.0079
33eeC5	1.4206	1.9716	-10.5164	1.4147	0.0059	24mm3 <i>i</i> -PrC5	1.4246	2.3679	-11.5060	1.4240	0.0006
C10	1.4102	1.7204	-11.5348	1.4074	0.0028	C12	1.4216	2.2268	-13.5220	1.4184	0.0032
2mC9	1.4099	1.8295	-11.5372	1.4092	0.0007	C13	1.4256	2.4834	-14.5138	1.4230	0.0026
3mC9	1.4125	1.8670	-11.5315	1.4107	0.0018	C14	1.4290	2.7412	-15.5046	1.4270	0.0020
4mC9	1.4123	1.8833	-11.5280	1.4115	0.0008	C15	1.4315	3.0014	-16.4948	1.4306	0.0009
5mC9	1.4116	1.8878	-11.5268	1.4117	-0.0001	C16	1.4345	3.2623	-17.4838	1.4339	0.0006
22mmC8	1.4082	2.0324	-11.5370	1.4131	-0.0049	C17	1.4369	3.5242	-18.4720	1.4368	0.0001
23mmC8	1.4146	2.0092	-11.5280	1.4139	0.0007	C18	1.4390	3.7874	-19.4602	1.4395	-0.0005
24mmC8	1.4091	2.0018	-11.5368	1.4125	-0.0034	C20	1.4425	4.3159	-21.4344	1.4443	-0.0018

^a The m, e, Pr and *i*-Pr represent methyl, ethyl, propyl and isopropyl.^b The experimental refractive index was taken from Ref.[34].^c Calculated by Eq.(5).

TABLE II Refractive index prediction for the external validation set

Alkane	$n_{\text{exp}}^{\text{a}}$	$\sum X_{1\text{CC}}$	$\sum X_{1\text{CH}}$	$n_{\text{pred}}^{\text{b}}$	$n_{\text{exp}}-n_{\text{pred}}$	Alkane	$n_{\text{exp}}^{\text{a}}$	$\sum X_{1\text{CC}}$	$\sum X_{1\text{CH}}$	$n_{\text{pred}}^{\text{b}}$	$n_{\text{exp}}-n_{\text{pred}}$
C4	1.3326	0.3203	-5.5244	1.3357	-0.0031	45mmC8	1.4190	2.0709	-11.5168	1.4167	0.0023
2mC5	1.3715	0.8540	-7.5453	1.3728	-0.0013	4eC8	1.4151	1.9423	-11.5178	1.4141	0.0010
23mmC5	1.3919	1.2403	-8.5454	1.3889	0.0030	224mmmmC7	1.4092	2.2090	-11.5255	1.4181	-0.0089
223mmmmC4	1.3864	1.4076	-8.5526	1.3920	-0.0056	233mmmmC7	1.4202	2.2674	-11.5794	1.4116	0.0086
C8	1.3974	1.2266	-9.5424	1.3924	0.0050	244mmmmC7	1.4142	2.2480	-11.5188	1.4198	-0.0056
4mC7	1.3979	1.3765	-9.5392	1.3965	0.0014	345mmmmC7	1.4229	2.2352	-11.5108	1.4207	0.0022
2mC8	1.4031	1.5800	-10.5423	1.4024	0.0007	22mm4eC6	1.4131	2.2928	-11.5106	1.4219	-0.0088
2233mmmmC5	1.4236	2.2165	-10.5260	1.4185	0.0051	24mm3eC6	1.4225	2.2569	-11.5071	1.4217	0.0008
22mm3eC5	1.4123	2.0193	-10.5258	1.4143	-0.0020	24mm4eC6	1.4235	2.3046	-11.5116	1.4220	0.0015

^a The experimental refractive index was taken from Ref.[34].^b Predicted by Eq.(5).

respectively. The calculated refractive indices of alkanes RH are listed in Table I.

Descriptor $X_{1\text{CC}}$, obtained by the matrix \mathbf{CM}_{ij} , has a positive value; $X_{1\text{CC}}$ is regarded as the relative bond energy of C–C bond in the alkane molecule. Descriptor $X_{1\text{CH}}$, obtained by the matrix \mathbf{CM}_{iH} , has a negative value; $X_{1\text{CH}}$ is regarded as the relative bond energy of C–H bond in the alkane molecule. The larger the value of the descriptor $X_{1\text{CC}}$ (or $X_{1\text{CH}}$) is, the weaker the C–C (or C–H) bond is [19]. Furthermore, more recent work [35] has shown that for these alkanes with the same number of carbon atoms, the larger the value of the descriptors $\sum X_{1\text{CC}}$ and $\sum X_{1\text{CH}}$ are, the looser the valence electrons grasped by the nucleus are. From Eq.(5), one can see that the coefficients of the descriptors $\sum X_{1\text{CC}}$ and $\sum X_{1\text{CH}}$ are both negative. Because the descriptors $\sum X_{1\text{CC}}$ and $\sum X_{1\text{CH}}$ are in the denominator position of the fraction, the negative coefficients in front of the descriptors indicate that the denomi-

nator decreased with increasing values of the two descriptors. When the numerator in the fraction is constant, the fraction increases with the increasing values of the two descriptors. In other words, the refractive indices increase with increasing values of the two descriptors $\sum X_{1\text{CC}}$ and $\sum X_{1\text{CH}}$ for those alkanes containing the same number of carbon atoms. That is to say, the looser the valence electrons grasped by the nucleus in a molecule are, the larger the refractive index of the molecule is. This result agrees with the physical intuition.

To test the quality of prediction of Eq.(5), the refractive indices of 18 alkane which were not included in the training set for building the QSPR model were predicted. Table II listed the molecules of the external validation set along with their prediction and experimental values of refractive index. The estimated *RMSE* and the unsigned error from the experimental values were only 0.0048 and 0.0037, respectively, which shows the

TABLE III Experimental and estimated refractive indices (at 101 kPa and 20 °C) for 44 chloroalkanes RCl and 23 bromoalkanes RBr

Haloalkane	$\sum X_{1CC}$	$\sum X_{1CH}$	n_{exp}^a	n_{calc}^b	$n_{\text{exp}}-n_{\text{calc}}$	Haloalkane	$\sum X_{1CC}$	$\sum X_{1CH}$	n_{exp}^a	n_{calc}^b	$n_{\text{exp}}-n_{\text{calc}}$
PrCl	0.1356	-3.9385	1.3879	1.3923	-0.0044	25mm2ClC6	1.4418	-9.0523	1.4232	1.4299	-0.0067
PrCl	0.1356	-3.9604	1.3777	1.3825	-0.0048	2eClC6	1.3679	-9.0002	1.4319	1.4371	-0.0052
BuCl	0.3203	-4.9612	1.4021	1.4029	-0.0008	244mmm2ClC5	1.6445	-9.0566	1.4308	1.4349	-0.0041
s-BuCl	0.3203	-4.9881	1.3971	1.3937	0.0034	C9Cl	1.4718	-9.9849	1.4345	1.4410	-0.0065
i-BuCl	0.3921	-4.9753	1.3984	1.4020	-0.0036	2ClC9	1.4718	-10.0157	1.4420	1.4360	0.0060
t-BuCl	0.3921	-5.0166	1.3857	1.3879	-0.0022	3m3ClC8	1.6165	-10.0499	1.4351	1.4343	0.0008
ClC5	0.5296	-5.9741	1.4127	1.4131	-0.0004	4m4ClC8	1.6299	-10.0496	1.4360	1.4347	0.0013
2ClC5	0.5296	-6.0031	1.4069	1.4051	0.0018	3e3ClC7	1.6668	-10.0496	1.4400	1.4356	0.0044
3ClC5	0.5296	-6.0085	1.4082	1.4036	0.0046	223mmm3ClC6	1.9732	-10.0589	1.4465	1.4420	0.0045
3m1ClC4	0.6219	-5.9831	1.4084	1.4148	-0.0064	C10Cl	1.7204	-10.9812	1.4379	1.4458	-0.0079
2m2ClC4	0.6219	-6.0346	1.4055	1.4004	0.0051	PrBr	0.1356	-3.9385	1.4343	1.4463	-0.0120
22mm1ClC3	0.7552	-6.0071	1.4044	1.4140	-0.0096	i-PrBr	0.1356	-3.9604	1.4251	1.4360	-0.0109
ClC6	0.7533	-6.9826	1.4199	1.4216	-0.0017	BuBr	0.3203	-4.9612	1.4401	1.4454	-0.0053
3ClC6	0.7533	-7.0199	1.4163	1.4129	0.0034	s-BuBr	0.3203	-4.9881	1.4366	1.4360	0.0006
2m2ClC5	0.8540	-7.0439	1.4126	1.4110	0.0016	i-BuBr	0.3921	-4.9753	1.4348	1.4405	-0.0057
4m2ClC5	0.8540	-7.0195	1.4113	1.4168	-0.0055	t-BuBr	0.3921	-5.0166	1.4278	1.4261	0.0017
2eClC4	0.8753	-6.9976	1.4222	1.4227	-0.0005	C5Br	0.5296	-5.9741	1.4447	1.4477	-0.0030
3m3ClC5	0.8753	-7.0476	1.4210	1.4110	0.0100	2BrC5	0.5296	-6.0031	1.4413	1.4395	0.0018
33mmClC4	1.0264	-7.0130	1.4161	1.4248	-0.0087	3BrC5	0.5296	-6.0085	1.4441	1.4380	0.0061
23mm2ClC4	0.9654	-7.0532	1.4191	1.4130	0.0061	2m1BrC4	0.6219	-5.9889	1.4451	1.4435	0.0016
22mm3ClC4	1.0264	-7.0553	1.4182	1.4148	0.0034	3m1BrC4	0.6219	-5.9831	1.4420	1.4452	-0.0032
ClC7	0.9866	-7.9861	1.4256	1.4291	-0.0035	22mm1BrC3	0.7552	-6.0071	1.4370	1.4385	-0.0015
2ClC7	0.9866	-8.0165	1.4221	1.4229	-0.0008	C6Br	0.7533	-6.9826	1.4478	1.4502	-0.0024
3ClC7	0.9866	-8.0249	1.4228	1.4212	0.0016	3BrC6	0.7533	-7.0199	1.4472	1.4414	0.0058
4ClC7	0.9866	-8.0269	1.4237	1.4208	0.0029	2m1BrC5	0.8540	-6.9967	1.4495	1.4470	0.0025
2m2ClC6	1.0914	-8.0483	1.4200	1.4199	0.0001	3m1BrC5	0.8753	-6.9896	1.4496	1.4486	0.0010
3m3ClC6	1.1214	-8.0516	1.4250	1.4202	0.0048	3m3BrC5	0.8753	-7.0476	1.4525	1.4350	0.0175
23mm2ClC5	1.2403	-8.0536	1.4264	1.4236	0.0028	33mm1BrC4	1.0264	-7.0130	1.4440	1.4432	0.0008
24mm2ClC5	1.1998	-8.0519	1.4180	1.4227	-0.0047	C7Br	0.9866	-7.9861	1.4502	1.4531	-0.0029
23mm3ClC5	1.2403	-8.0583	1.4318	1.4227	0.0091	2BrC7	0.9866	-8.0165	1.4503	1.4469	0.0034
22mm4ClC5	1.2794	-8.0308	1.4180	1.4296	-0.0116	4BrC7	0.9866	-8.0269	1.4495	1.4448	0.0047
3e3ClC5	1.1517	-8.0529	1.4400	1.4209	0.0191	C9Br	1.4718	-9.9849	1.4533	1.4583	-0.0050
ClC8	1.2266	-8.9872	1.4305	1.4354	-0.0049	C10Br	1.7204	-10.9812	1.4557	1.4605	-0.0048
3m3ClC7	1.3679	-9.0522	1.4317	1.4278	0.0039						

^a The experimental refractive index was taken from Ref.[34].

^b Calculated by Eq.(7).

good prediction accuracy of Eq.(5). Thus, Eq.(5) can be recommended for predicting the refractive indices of the untested compounds with confidence.

Compared with Kier and Hall's result ($r^2=0.998$, $S=0.043$) [36], the present paper obtained a better result. More recently, Du *et al.* has built a good QSAR model for calculating the refractive indices of 27 alkanes ($r=0.9600$, $s=0.0028$) [30]. Thought the prediction error was somewhat higher than that of Du's model, the present model (Eq.(5)) correlated a larger data set than Du's model did. What is more, each descriptor

employed in the present model has a very explicit physical meaning and the QSPR model can be interpreted from the physical point of view, which offers us an insight into the structure-property relationship.

The good result obtained in the prediction of the refractive indices for alkanes encourages further exploration of the usefulness of the present model (i.e. Eq.(2)) in other molecular systems. The present research found that Eq.(2) could be also applied to predict the refractive indices for haloalkanes.

For and the 44 chloroalkanes RCl (Table III), 23 bro-

moalkanes RBr using Eq.(2) as the model, one can get:

$$n_{\text{RCl}} = \frac{N_{\text{C}}}{-0.59271 - 0.11236 \sum X_{1\text{CC}} - 0.70143 \sum X_{1\text{CH}}} \quad (6)$$

$$r = 0.9999, RMSE = 0.0059, N = 44, F = 70863$$

$$n_{\text{RBr}} = \frac{N_{\text{C}}}{-0.59594 - 0.00152 \sum X_{1\text{CC}} - 0.67802 \sum X_{1\text{CH}}} \quad (7)$$

$$r = 0.9999, RMSE = 0.0062, N = 23, F = 53500$$

Because the number of the experimental refractive indices of iodoalkanes is too small, we just investigated the structure-refractive index relationship for chloroalkanes RCl and bromoalkanes RBr, and the results were satisfying.

IV. CONCLUSION

A general 3-descriptor QSAR model with explicit physical meaning was developed to predict the refractive indices of alkanes RH, chloroalkanes RCl and bromoalkanes RBr. The obtained results illustrated the usefulness of the bond orbital-connection matrix method. The present descriptors obtained by the BOCM method have the merit of topological indices (i.e. facility and rapid calculation of the descriptors) and the advantage of quantum descriptors (i.e. explicit physical meaning of the parameters). Furthermore, the selection of the descriptors in the present QSAR model was on the basis of the comprehension of the specific structural features that affect the refractive index of the compounds, which may be more helpful in understanding the structure-property relationship.

V. ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No.20472019), the Natural Science Foundation of Hunan Province (No.06JJ20002) and the Key Scientific Research Fund of Hunan Province Education Department (No.04A015). We would like to express our appreciation to Mr. Chuen L. Wen and En X. Liang for offering some of references.

- [1] C. Hansch, A. Leo and D. Hoekman, *Exploring QSAR. Fundamentals and Applications in Chemistry and Biology*, Washington DC: American Chemical Society, (1995).
- [2] M. Pompe and M. Randić, *J. Chem. Inf. Model.* **46**, 2 (2006).
- [3] R. Todeschini and V. Consonni, *The Handbook of Molecular Descriptors, in the Series of Methods and Principles in Medicinal Chemistry*, R. Mannhold, H. Kubinyi and H. Timmerman, Edn., New York: Wiley-VCH, Vol.11, (2000).
- [4] L. Xu and C. Y. Hu, *Application of the Chemical Graph Theory*, Beijing: Science Press, (2000).
- [5] H. W. Xin, *Topology of Molecules*, Hefei: University of Science and Technology of China Press, (1991).
- [6] C. M. Nie and M. F. Fan, *Chin. J. Chem. Phys.* **13**, 71 (2000).
- [7] L. L. Mu and C. J. Feng, *Chin. J. Chem. Phys.* **16**, 197 (2003).
- [8] R. Zeng, C. Cao and H. Yuan, *Chin. J. Chem. Phys.* **18**, 372 (2005).
- [9] C. Cao and L. Liu, *Chin. J. Chem. Phys.* **18**, 707 (2005).
- [10] H. Y. Yao, D. J. Hua, C. G. Le, X. P. Zhu and M. B. Luo, *Chin. J. Chem. Phys.* **17**, 426 (2004).
- [11] M. Karelson, V. S. Lobanov and A. R. Katritzky, *Chem. Rev.* **96**, 1027 (1996).
- [12] H. Wiener, *J. Am. Chem. Soc.* **69**, 17 (1947).
- [13] M. Randić and J. Zupan, *J. Chem. Inf. Comput. Sci.* **41**, 550 (2001).
- [14] E. Estrada and H. González, *J. Chem. Inf. Comput. Sci.* **43**, 75 (2003).
- [15] M. Randić and M. Pompe, *J. Chem. Inf. Comput. Sci.* **41**, 575 (2001).
- [16] <http://www.semichem.com/codessa.html>
- [17] <http://www.disat.unimib.it/chm/dragon.html>
- [18] E. Estrada, E. J. Delgado, J. B. Alderete and G. A. Jaña, *J. Comput. Chem.* **25**, 1787 (2004).
- [19] C. Cao and H. Yuan, *J. Chem. Inf. Comput. Sci.* **43**, 600 (2003).
- [20] C. Cao and Z. Li, *J. Chem. Inf. Comput. Sci.* **38**, 1 (1998).
- [21] S. M. Ulam, *A Collection of Mathematical Problems*, New York: John Wiley Sons, (1960).
- [22] C. Cao and S. Gao, *J. Mol. Struct. (THEOCHEM)*. **718**, 153 (2005).
- [23] C. Cao, S. Gao and R. Zeng, *J. Mol. Struct. (THEOCHEM)*. **728**, 85 (2005).
- [24] C. Cao and S. Gao, *Internet Electron. J. Mol. Des.* **4**, 671 (2005).
- [25] C. Cao, L. Jiang and H. Yuan, *Internet Electron. J. Mol. Des.* **2**, 621 (2003).
- [26] I. L. Finar, *Organic Chemistry*, 2nd Edn., London: Longman, Vol.2, (1975).
- [27] J. R. Partington, *An Advanced Treatise on Physical Chemistry*, London: Longmans, Green, Vol.4, (1953).
- [28] J. Bicerano, *Prediction of Polymer Properties*, 2nd Edn., New York: Marcel Dekker Inc., (1996).
- [29] K. Q. Wang, *Chin. J. Org. Chem.* **15**, 6599 (1995).
- [30] X. H. Du and J. G. Guang, *Chin. J. Chem. Phys.* **18**, 211 (2005).
- [31] S. Grigoros, *J. Coumpt. Chem.* **11**, 493 (1990).
- [32] C. Cao, *Substituent Effects in Organic Chemistry*, Beijing: Science Press, (2003).
- [33] C. Cao and S. Gao, *Acta Phys. Chim. Sin.* **21**, 1028 (2005).
- [34] D. R. Lide, Ed., *CRC Handbook of Chemistry and Physics*, Internet Version 2005, 85th edn., Boca Raton: CRC Press, FL, 2005, <http://www.hbcnetbase.com>.
- [35] S. Gao and C. Cao, *Acta Phys. Chim. Sin.* (Accepted, in press)
- [36] L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure-Activity Analysis*, Letchworth, Hertfordshire, England: Research Studies Press, Ltd., 26, (1986).