

用分子子图法计算烷烃热力学函数*

刘剑洪¹, 田德余^{2*}, 陈刚³, 刘树深³, 李志良³

(1. 深圳大学师范学院化学及生物学系, 广东深圳, 518060;

2. 国防科技大学五系, 湖南长沙 410073; 3. 湖南大学化学化工学院, 湖南长沙 410082)

摘要: 在一种新的烷烃分子子图编码方法基础上, 对烷烃的系列热力学函数进行了拟合和预测, 取得了满意结果。对吉布斯自由能 ΔG_f , 原子生成热 ΔH_{atm} , 热容 C_p , 蒸发热 ΔH_v , 液态生成热 $\Delta H_{f(l)}$ 和熵 S 的拟合方程的回归系数分别为 0.955 9, 1.000 0, 0.990 5, 0.996 9, 0.997 3 和 0.992 2。

关键词: 分子子图; QSPR; 烷烃; 热力学函数

中图分类号: O621.2 文献标识码: A

Estimation and Prediction on Thermodynamical Functions of Alkanes with Novel Molecular Subgraph

LIU Jian-hong¹, TIAN De-yu², CHEN Gang³, LIU Shu-shen³, LI Zhi-liang³

(1. Department of Chemistry and Biology, Shenzhen University, Shenzhen 518060,

2. The 5th Department, National University of Defense Technology, Changsha 410073,

3. Institute of Chemistry and Chemical Engineering, ICP&CGC, Hunan University, Changsha 410082)

Abstract: The molecular structure of alkanes is described by a novel coding method, recently developed in this laboratory, on the basis of molecular subgraph. It has been shown that there exists very good correlation between the coding and thermodynamical functions of alkanes. The correlation coefficients (R) of MLR equation for quantitative structure property relation (QSPR) on Gibbs energy (ΔG_f), atomization heat (ΔH_{atm}), heat capacity (C_p), evaporating enthalpy (ΔH_v), enthalpy ($\Delta H_{f(l)}$) and entropy (S) are respectively 0.9559, 1.0000, 0.9905, 0.9969, 0.9973 and 0.9922. The responding derivation (S) are 4.710, 4.307, 3.793, 0.484, 0.931, 0.585.

Key words: molecular subgraph; QSPR alkanes; thermodynamical function

热力学数据是化学计算与工程设计中经常用到的基础数据。随着有机化工的日益发展, 在化工设计中常常遇到热力学数据不够用的问题, 因此人们不得不利用某些理论或经验公式来估算所缺少的热力学数据。近年来许多研究人员在这一领域进行了大量开拓性工作^[1-11]。他们主要从以下四个方面来研究: (1) 利用微观性质与宏观性质之间的关系。(2) 利用分子的整体性质与构成分子的结构因素之间的联系。(3) 利用不同物质同种性能的对对应关系。(4) 利用同一物质不同物性之间的联系^[5]。

从拓扑学角度研究定量构性关系 (QSPR) 日趋频繁, 已经延伸至热力学数据领域^[1-4, 6-11]。我们根据已提出的烷烃分子子图编码方法, 对烷烃的多种热力学数据进行了拟合建模。

1 原理与方法

1.1 烷烃分子子图分解原理

烷烃的拓扑分子图为其隐氢结构式。我们对拓扑分子图进行分解时遵循以下原则: (1) 不变性: 每种分子的拓扑图分解后只有唯一的子图表示方式; (2) 唯一性: 不同分子的子图表示方式不重复;

* 收稿日期: 1999-08-05

作者简介: 刘剑洪 (1969-), 男, 副教授, 博士。

(3) 简易性: 拓扑图的分解方法应简单易行。本文采用下列方法分解烷烃的拓扑图: (1) 按命名原则选取烷烃的主链为一个基本子图; (2) 以主链上每个分支为一个亚子图; (3) 子图分类的优先原则: 依季叔仲伯碳为序, 即 $4^{\circ}C > 3^{\circ}C [> 2^{\circ}C > 1^{\circ}C > 0^{\circ}C$ (甲烷)]; (4) 用命名时主链的编码确定子图的位置; (5) 在不重复情况下, 为减少子图项的数目, 当取代基离主链的起始端较远时, 从碳链另一端开始找出此取代基的位置, 代替原位置编码。例如在 $C_1 \sim C_{10}$ 共 150 个烷烃中没有出现 6M 和 7M 子图项。

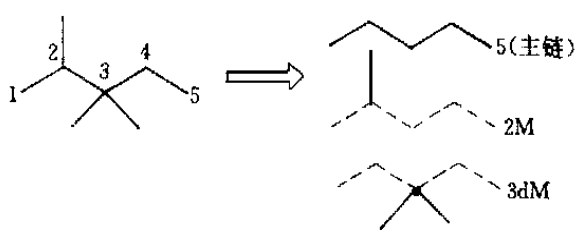


图1 烷烃分子图编码示例

Fig. 1 Illustration for coding alkanes

今以 2, 3, 3- 三甲基戊烷为例说明, 具体见图 1。(1) 基本子图即主链长为 5; (2) 两个亚子图即 2M 和 3dM。

1.2 烷烃分子的结构表征和热力学函数的相关计算

虽然可用分子图表达分子的拓扑性质, 但是从本质上来看, 分子是个非数值的数学对象。分子的各种可以测量的性质, 通常又都是用数值表达的, 因此为了把分子的拓扑性质与分子的各种可测量的性质联系起来, 必须把在分子图中获得的信息转变为一种能用数值表达的^[6]。在含 1~10 个碳原子的所有 150 种烷烃中, 根据上述原则共生成了 26 个子图项。我们将每个分子所包含的各子图项的数目作为分子的结构信息, 直接编码。显而易见, 每个分子的编码由 26 个元素组成。

图 1 的例子中, 子图项 5 (表示主链的长度)、2M (2 位的甲基) 和 3dM (3 位的二甲基) 的数目各是 1, 其余为 0。 $C_1 \sim C_{10}$ 中部分烷烃的子图编码见表 1。表中的烷烃的名称是按系统命名法简写, 方法如下: (1) 名称最后的数字表示主链长度。例如 2M5 中 5 表示主链长度为 5, 即戊烷; (2) 名称中大写的英文字母为取代基名称, 如 M 为甲基, E 为乙基, P 为丙基, iP 为异丙基; (3) 取代基前的数字表示此取代基在主链上的位置。因此, 2, 3, 3- 三甲戊烷在表 1 中的名称为 2M3M3M5。

表1 $C_1 \sim C_{10}$ 中部分烷烃的子图编码

Tab. 1 Coding of 150 alkanes

No	Alkane	1	2	3	4	5	6	7	8	9	10	2M	3M	4M	5M	2dM	3dM	4dM	3E	4E	5E	3ME	4ME	4P	3iP	4iP	3dE	
001	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
015	2M6	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
036	2M3M3M5	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
040	2M2M3M3M4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
041	9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
084	2M5M8	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
122	2M2M3M4M6	0	0	0	0	0	1	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
142	3E4E6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
149	2M3E3E5	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
150	2M4M3iP5	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0

1.3 数据与计算

表 3 中的热力学数据来自文献 [10~12]。采用自编的多元线性回归程序在 PII-233 微机上运行。因各种热力学数据的数目不同, 且缺实际数据的化合物 (样本) 不同, 故在具体计算一种数据时, 需把无实测数据的化合物从表 1 中删去, 留下有实测数据者作为校正集。还必须把校正集中所有化合物都没有的子图项删去, 否则不能进行多元线性回归运算。所以, 由此建立起来的拟合模型不能预测含有校正集中未包括的子图项的化合物 (样本)。

本文将以上经相应修正的校正集的子图编码分别与吉布斯自由能 ΔG_f , 原子生成 (化) 热 ΔH_{atm} , 液态生成热 $\Delta H_{f(l)}$, 热容 C_p , 蒸发热 ΔH_v 和熵 S 的实测值拟合和预测。

表 2 烷烃热力学数据计算结果

Tab. 2 The results on thermodynamical functions of alkane by MLR method.

	ΔG_f	ΔH_{atm}	C_p	ΔH_V	$\Delta H_f(t)$	S
b_0	39.537	7417.235	273.661	32.953	60.378	51.003
b_1	-68.085	-3411.418	-174.957	-17.643	-37.878	-31.103
b_2	-54.922	-2236.603	-153.449	-11.752	-30.156	-26.998
b_3	-50.752	-1064.106	-129.431	-6.824	-23.896	-22.685
b_4	-43.449	110.809	-106.971	-2.040	-18.247	-18.414
b_5	-33.680	1283.286	-85.385	2.950	-11.938	-13.827
b_6	10.042	2454.983	23.138	8.073	-6.131	2.785
b_7	14.394	3628.965	21.374	13.487	0.000	2.466
b_8	15.998	1178.347	23.757	3.701	6.073	2.679
b_9	9.790	1174.533	22.489	4.147	11.572	2.545
b_{10}	22.188	1174.730	43.398	3.678	6.812	5.808
b_{11}	29.947	1180.028	46.051	3.072	5.894	5.550
b_{12}	32.379	2358.667	48.800	6.470	5.804	5.622
b_{13}	28.739	2350.631	46.705	7.501	7.094	7.280
b_{14}	27.569	2347.404	49.609	6.162	13.871	7.766
b_{15}	23.843	2347.029	48.654	8.751	11.879	7.926
b_{16}	48.570	4696.071	66.445	8.721	11.332	9.297
b_{17}	43.843	-	70.071	17.231	10.860	9.029
b_{18}	42.111	-	71.010	-	23.719	12.262
b_{19}	50.810	-	65.913	-	-	9.570
b_{20}	41.811	-	72.510	-	-	10.512
b_{21}	62.181	-	92.456	-	-	13.646
N	134	44	134	59	52	134
R	0.9559	1.0000*	0.9905	0.9969	0.9973	0.9922
S	4.710	4.307	3.793	0.484	0.931	0.585
F	56.579	22558.790	277.353	489.386	483.154	336.445

* 实际计算机结果为 0.999998

2 结果与讨论

2.1 回归方程

用多元线性回归得到各热力学函数的回归方程, 其中各回归系数及统计参量见表 2。表 2 中 $b_0 \sim b_{21}$ 分别为各回归方程的系数, N 为样本数, R 为拟合的相关系数, S 为标准偏差, F 为统计检验值。由此可知, 它们的相关性均很好, $R > 0.99$, 仅有一函数 (ΔG_f) 例外, 其 $R = 0.956$ 。其中原子化热 (ΔH_{atm}) 结果最佳, 其相关系数极接近 1, 高达 $R = 0.999998$ 。

2.2 估计与预测

部分具体的计算值见表 3。表 3 中每个热力学函数有实测值 (Exp.), 计算值 (Calc.), 绝对误差 (Err.) 三项, 分别表示实测值、计算值及其偏差。三项都有表示此物质在校正集中, 而仅有计算值这一项表示此物质在预测集中, 计算值为预测结果。

2.3 初步结论

从计算结果看, 除对吉布斯自由能 ΔG_f 的拟合效果稍差外 ($R = 0.959$, $S = 4.710$), 其它拟合方程的回归系数 (R) 都在 0.990 以上。这说明, 用本文所述的分子子图编码方法预测烷烃的一系列热力

学函数是切实可行的,有待进一步推广应用。

表3 烷烃热力学函数的计算值与实测值

Tab.3 Comparison of calculated and experimental values of thermodynamical functions on alkanes

No	Alkane	ΔG_f (300K)			ΔH_{atm} (300K)			C_p (300K)			ΔH_v (298K)			$\Delta H_f(l)$			S (300K)		
		Exp.	Calc.	Err.	Exp.	Calc.	Err.	Exp.	Calc.	Err.	Exp.	Calc.	Err.	Exp.	Calc.	Err.	Exp.	Calc.	Err.
002	2																22.50	22.500	0.000
003	3				4000.45	4005.816	-5.366					15.310					28.79	30.222	-1.432
004	4		-28.548		5176.18	5180.632	-4.452		98.704			21.201					35.34	36.482	-1.142
005	2M3				5184.56	5184.164	0.396					19.011					36.95	37.034	-0.084
006	5		-15.385		6349.91	6353.128	-3.218		120.212			26.42	26.129	0.291			41.40	42.131	-0.731
007	2M4		-18.506		6357.95	6358.980	-1.030		121.842			24.59	24.902	-0.312			42.95	43.295	-0.345
008	2M2M3				6369.46	6364.483	4.977					21.78	21.780	0.000			45.61	44.094	1.517
009	6		-11.215		7524.10	7528.044	-3.944		144.230			31.55	30.913	0.637			47.52	48.440	-0.920
010	2M5		-4.05	-5.344	1.294	7531.22	7531.476	-0.256	143.01	143.350	-0.340	29.86	29.830	0.030			48.82	48.943	-0.124
011	3M5		-2.12	-0.991	-1.129	7528.54	7527.662	0.879	140.88	141.585	-0.705	30.27	30.276	-0.006			48.28	48.025	0.255
012	2M2M4		-7.42	-6.360	-1.060	7542.48	7539.299	3.181	142.26	142.103	0.157	27.69	27.671	0.019			51.00	50.354	0.646
013	2M3M4		-1.77	-4.112	2.342	7534.69	7533.513	1.177	142.21	143.216	-1.006	29.12	29.049	0.071			49.48	49.189	0.292
014	7		9.50	-3.911	13.412	8698.16	8700.521	-2.361	166.00	166.690	-0.690	36.55	35.903	0.647			53.63	54.248	-0.618
015	2M6		4.90	-1.174	6.074	8705.32	8706.391	-1.071	165.40	167.368	-1.968	34.80	34.614	0.186			54.93	55.253	-0.323
016	3M6		6.60	3.179	3.421	8702.68	8702.577	0.103	164.50	165.604	-1.104	35.08	35.060	0.020			54.35	54.334	0.016
017	2M2M5		2.10	6.803	-4.703	8716.54	8711.795	4.745	167.70	163.610	4.090	32.43	32.598	-0.168			57.05	56.003	1.047
018	2M3M5		7.60	9.051	-1.451	8709.63	8706.009	3.621	161.80	164.723	-2.923	34.24	33.977	0.263			54.83	54.943	-0.008
019	2M4M5		4.90	10.654	-5.754	8712.39	8706.206	6.184	171.70	167.106	4.594	32.88	33.508	-0.628			56.17	54.748	1.422
020	3M3M5		4.80	14.562	-9.762	8711.94	8703.759	8.181	166.70	166.263	0.437	33.02	33.629	-0.609			56.07	54.010	2.060
021	3E5		12.47	13.353	-0.883	8700.04	8700.157	-0.117	166.80	166.916	-0.116	35.22	34.880	0.340			53.77	52.991	0.779
022	2M2M3M4		6.30	8.034	-1.734	8715.20	8713.832	1.368	164.20	163.476	0.724	32.04	31.818	0.222			56.63	56.248	0.382
023	8		17.67	5.857	11.813	9872.22	9872.218	0.002	188.70	188.276	0.424	41.48	41.026	0.454			59.74	60.378	-0.638
024	2M7		13.37	6.130	7.240	9879.26	9878.868	0.392	188.20	189.828	-1.628	39.68	39.604	0.076			60.98	61.060	-0.080
025	3M7		13.79	10.483	3.307	9876.41	9875.054	1.356	186.82	188.064	-1.244	39.83	40.050	-0.220			60.34	60.142	0.199
026	4M7		17.40	12.086	5.314	9878.87	9875.252	3.621	188.03	190.446	-2.416	39.67	39.581	0.089			60.17	60.052	0.118
027	2M2M6		12.15	10.973	1.177	9888.51	9886.711	1.799	189.33	187.628	1.702	37.29	37.382	-0.092			62.63	62.312	0.318
028	2M3M6		17.20	13.221	3.979	9877.71	9880.924	-3.214	185.18	188.742	-3.562	38.79	38.761	0.029			60.40	61.147	-0.747
029	2M4M6		13.07	14.824	-1.754	9883.19	9881.121	2.069	193.35	191.125	2.225	37.76	38.292	-0.532			61.47	61.057	0.413
030	2M3M6		11.40	8.616	2.784	9886.42	9886.419	0.001	186.52	189.857	-3.337	37.86	37.685	0.175			62.26	62.347	-0.087
031	3M3M6		15.13	18.732	-3.602	9883.90	9878.675	5.225	191.96	190.281	1.679	37.93	38.413	-0.483			61.58	60.319	1.261
032	3M4M6		18.43	19.177	-0.747	9876.77	9877.307	-0.537	182.72	189.360	-6.640	39.02	38.738	0.282			60.23	60.139	0.092
033	3E6		18.53	17.523	1.007	9874.65	9875.073	-0.423	190.58	190.935	-0.355	39.40	39.664	-0.264			59.88	59.300	0.580
034	2M2M3M5		19.45	21.197	-1.747	9883.90	9886.328	-2.428	186.77	184.983	1.787	36.91	36.745	0.165			61.44	61.897	-0.457
035	2M2M4M5		15.70	22.801	-7.101	9887.92	9886.526	1.394	189.45	187.366	2.084	35.13	36.277	-1.147			61.97	61.807	0.163
036	2M3M3M5		20.04	24.604	-4.564	9880.22	9882.107	-1.887	188.20	189.401	-1.201	37.22	37.330	-0.110			60.63	60.822	-0.193
037	2M3M4M5		20.76	25.048	-4.289	9881.22	9880.739	0.481	192.72	188.480	4.240	37.61	37.655	-0.045			60.98	60.642	0.338
038	2M3E5		20.68	23.395	-2.715	9874.99	9878.505	-3.515	193.05	190.054	2.996	38.52	38.581	-0.061			59.69	59.803	-0.113
039	3M3E5		24.36	27.748	-3.388	9878.75	9874.690	4.060	189.07	188.290	0.780	37.99	39.027	-1.037			60.46	58.885	1.575
040	2M2M3M3M4		24.04	23.587	0.453	9889.68	9889.930	-0.250	188.28	188.154	0.126		35.172				62.40	62.233	0.167
041	9				11046.20	11046.200	0.000					46.44	46.440	0.000			65.64	66.451	-0.811

参考文献

- [1] Wiener H. Structural determination of paraffin boiling points [J]. J. Am. Chem. Soc. 1947, 69: 17-20.
- [2] Wiener H. J. Phys. Chem., 1948, 52 (14): 2636-2638.
- [3] Balaban A T. Applications of graph theory in chemistry [J]. J. Chem. Inf & Comput. Sci., 1985, 25 (2): 334-343.
- [4] Wessel M D, Jurs P C. Prediction of normal boiling points of hydrocarbons from molecular structure [J]. J Chem. Inf & Comput. Sci., 1995, 35 (1): 68-76.
- [5] Peterson K L. J Chem. Inf & Comput. Sci., 1995, 35 (5): 896-904.
- [6] Randic M. J Math Chem., 1991, 7 (1): 155-168.
- [7] 辛厚文. 分子拓扑学 [M]. 合肥: 中国科学技术大学出版社. 1992, 第1版.
- [8] 杨家安, 江元生. 饱和链线图特征和热力学性质 [J]. 化学学报. 1983, 41: 884-895.
- [9] 刘树深, 杨万平, 曹晨忠, 李志良. 改良人工神经网络方法预测烷烃物理性质 [J]. 化工学报. 1998, 49: 245-250.
- [10] 赵国良, 靳长德. 有机热力学数据的估算 [M]. 北京: 人民教育出版社. 1983, 第一版.
- [11] Gakh A A, Gakh E G, Sumpter B G, Noid D W. Neural network-graph theory approach to the prediction of the physical properties of organic compounds [J]. J. Chem Inf & Comput Sci. 1994, 34 (5): 832-839.
- [12] Weast R C (ed.), Handbook of Chemistry and Physics [M]. 70th, Boca Raton: CRC Press. Inc. 1989.
- [13] 王化云. 烷烃异构体的结构排序与热力学性质的变化规律 [J]. 化学学报. 1993, 51: 216-223.
- [14] 姚瑜元, 许禄, 袁秀顺. 一种新的拓扑指数用于饱和链烷烃类化合物的结构/性质相关性研究 [J]. 化学学报. 1993, 51: 463-469.