JOURNAL OF PHYSICAL ORGANIC CHEMISTRY
J. Phys. Org. Chem. 2002; **15:** 677–688
DOI: 10.1002/poc.535

Review Commentary

Classifications of families of homologous organic compounds based on energy by means of *ab initio* HF, MP2 and DFT optimizations[†]

Alexander Neugebauer and Günter Häfelinger*

University of Tübingen, Institute for Organic Chemistry, Auf der Morgenstelle 18, D-72076 Tübingen, Germany

Received 9 December 2001; revised 20 March 2002; accepted 24 April 2002

ABSTRACT: A method is suggested for combining the quantum chemical principle of isolated molecules and the systematics of homologous families for different functional groups in traditional organic chemistry. The classification of organic compounds is performed by statistical treatments of fundamental quantum mechanical quantities such as total and partial energies. We used the total energies (computed by DFT BLYP and B3LYP, HF and MP2 methods, which are size-extensive, applying the 6–31 G** and some other basis sets) of isolated molecules obtained by gradient optimizations to classify homologous series of organic compounds by linear least-squares regressions versus the total number of electrons. Two numerical parameters, slope (h) and intercept (s), are obtained characterizing each family of homologous compounds. Families of for example 17 different functional groups of the same CH_2-homologous unit are grouped on parallel lines with the same slope h, differing only in their s parameters. Isomeric compounds differ in their s values if they belong to the same homologous unit. Compounds of other homologous units (conjugated alkenes, cumulenes and benzenoides) are found with different slopes h on regression lines. Energy partitioning leads also to linear regressions of different statistical accuracy for the components which add fully to the observed h and s parameters. The h and s parameters allow the prediction of unknown total energies of homologous compounds from total energies of a subset of smaller compounds which belong to the same homologous family. Likewise, atomization energies can be predicted using the derived h and s parameters.

KEYWORDS: homologous families of organic compounds; energetical classifications; *ab initio* RHF, MP2 and DFT BLYP and B3LYP optimizations; estimation of total energies in homologous series; predictions of total energies of polymers; energetic differences of isomers; predictions of atomization energies; splitting of total energies

INTRODUCTION

Organic chemistry as an empirical science classifies organic compounds according to the properties of functional groups, such as double or triple CC bonds, hydroxy groups in alcohols, etc., which may be grouped in homologous series, differing by a fixed unit, e.g. C_nH_{2n+2} for saturated hydrocarbons, C_nH_{2n} for alkenes or cycloalkanes, etc.

The Beilstein system[1] of classification of organic compounds via the basic skeleton and defined functional groups and/or derivatives thereof and the systematic increase in the number of carbons in homologous series is unique because it assigns to each known or possible organic compound an appropriate place in the *Beilstein Handbook of Organic Chemistry*.[2]

Systematic behaviors of homologous series concerning experimental enthalpies of formation were investigated by Cohen and Benson.[3] They introduced additivity schemes for kinds of atoms which allow one to predict enthalpies of formation for members of homologous series.

There is no quantum chemical background for this kind of intuitional systematic classification of classical organic chemistry, because the treatment of whole classes of molecules is very difficult to reconcile with the principles of quantum mechanics, because quantum mechanical calculations themselves treat only isolated individual molecules. Therefore, in principle, no compound classes such as alcohols or aldehydes, or bond types or hybridizations, etc., are included. Šatkovskiene and Gineityte[4] suggest a method called the ‘density matrix method within the framework of the effective hamiltonian.’ This method is very formal and limited to tetrahedral systems, and its application to classifications of various organic compounds is difficult.

We present here a concept to combine the systematics of homologous families of organic chemistry with

---

*Correspondence to: G. Häfelinger, University of Tübingen, Institute for Organic Chemistry, Auf der Morgenstelle 18, D-72076 Tübingen, Germany.
E-mail: guenter.haefelinger@uni-tuebingen.de

[†] Dedicated to Prof. Dr T. M. Krygowski on the occasion of his 65th birthday.

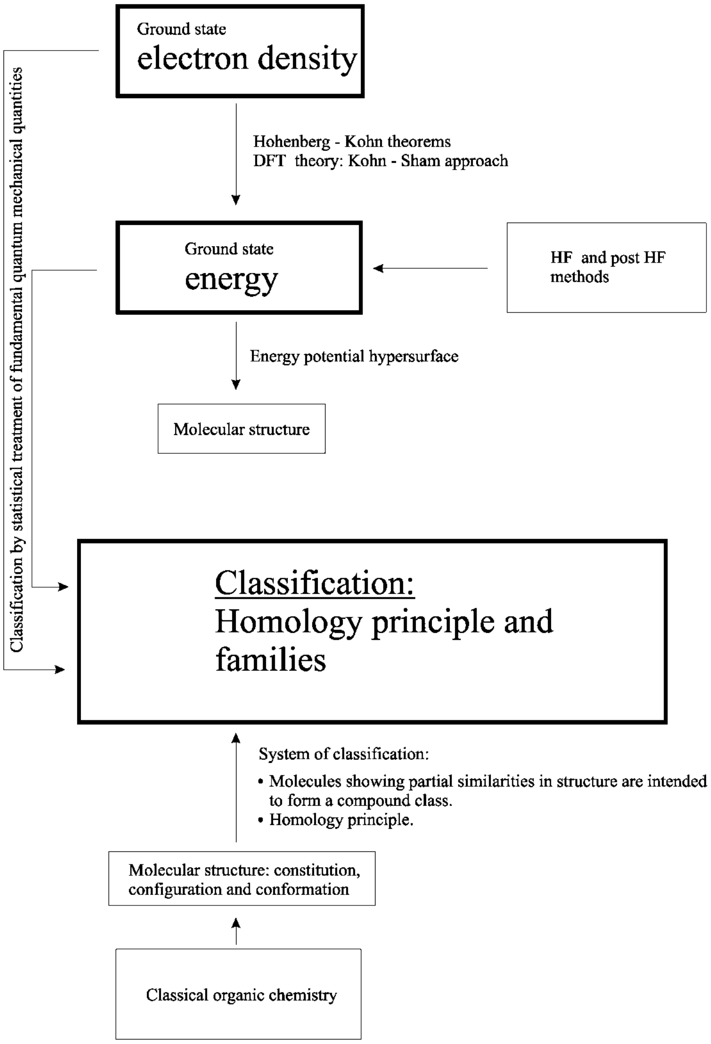

Figure 1. Survey of traditional and non-traditional classification procedures

various functionalities and the quantum chemical treatment of isolated systems. In connection with our classification scheme, the term size extensivity is of fundamental significance. It was introduced by Bartlett and Purvis[5] and is borrowed from thermodynamics. There, an extensive property is proportional to the size of a homologous system. Chapter 1 of the review of Taylor[6] deals explicitly with this and related concepts. A quantum theoretical method can be termed size extensive if the energy scales linearly with the number of electrons N. The energy of the Hartree–Fock method, the Møller–Plesset perturbation theory, the coupled cluster methods and the energy of DFT BLYP and B3LYP procedures are size extensive. Truncated configuration interaction methods (e.g. CISD) are not size extensive, whereas full CI is, however, size extensive. For size-inextensive methods our classification scheme based on the proportionality of total energies of organic compounds with the number of electrons should not work.

Our classification of simple organic compounds is performed by statistical treatments of fundamental quantum mechanical quantities. First the total energies of isolated systems offer an opportunity to generate a classification of families of organic compounds in the framework of only two numerical parameters based on the slope (h) and intercept (s) of regression equations related to the total number of electrons. Wiberg[7] determined a series of group equivalents (for CH_3, CH_2, CH, and C groups in various environments) to convert 6–31G* calculated molecular total energies into energies of formation at 0 K. Dewar and Storch[8] designed for this purpose atom equivalents for the atoms C, H, O and N. Analogously, Ibrahim and Schleyer[9] proposed sets of atom equivalents with about 50 parameters for 3–21G and 6–31G* basis sets to calculate heats of formation with average errors of the order of $0.8\,\text{kcal mol}^{-1}$ (1 kcal = 4.184 kJ). Later, Yala[10] used 6–31G* values and linear regression to derive 11 atomic parameters for the calculation of heats of formation which consider only the coordination number of the atoms in the molecule. These approaches were aimed at predicting of calculated heats of formation starting from total energies, not to classify homologous organic compounds using total energies.

The observed linearity of total energies with the number of electrons is similar to the linearity of experimental or calculated thermochemical heats of formation within homologous series. Therefore, our classification scheme is also able to predict energies of atomization.

Figure 1 shows a diagram of all the classification procedures mentioned above. The upper part of Fig. 1 shows the quantum chemical aspects of the classification of organic compounds based on energy and the lower section describes the traditional means of classifications by organic chemistry. For traditional organic chemistry only the structure of the molecules is important, which is given in terms of constitution, configuration and conformation. The quantum chemical approach is consistent from this point of view because of the relationship between energy and structure due to the energy potential hypersurface in the limit of the Born–Oppenheimer approximation.[11]

CALCULATION PROCEDURE

Ground-state molecular energies and ground-state densities were determined by gradient optimizations using density functional theory (DFT) methods[12] and the restricted Hartree–Fock[13] (RHF) and the Møller–Plesset perturbation method[13] of second order (MP2). DFT calculations were performed using two functionals with different levels of complexity: A gradient corrected pure functional including exchange and correlation (= BLYP)[14,15] and the three hybrid functional B3LYP[16] with part of exact exchange and approximated exchange as well as local and non-local correlation terms. For each of the calculational methods the following four basis sets were applied: Pople's 6–31G** basis set,[17,18] which is a split-valence basis set of double-zeta quality with polarization d-functions on heavy atoms and p-polarization functions on hydrogen and three flexible polarized correlation consistent (cc-p) valence basis sets of Dunning:[19] cc-pVDZ, AUG-cc-pVDZ and cc-pVTZ, which present systematic improvements in size from double zeta (DZ) to triple-zeta (TZ) quality including polarization functions. AUG denotes there additional augmention by diffuse s-type functions of higher quantum numbers. Geometry optimizations were performed using the Gaussian 98 (Revision A.7) program system.[20] Starting geometries of molecules for geometry optimizations were chosen using values of bond lengths and bond angles estimated from general hybridization rules. The structures were refined to tight cutoffs (forces and displacements) of the Gaussian 98 program.

Partial energy contributions to the total energy were obtained by a version of Gaussian 98 modified by us. For all DFT calculations we used the 'fine' integration grid of the Gaussian 98 program for enhanced accuracy of calculations.

RESULTS AND DISCUSSION

Classification based on total energy

Our classification scheme of organic compounds based on the total energy of single molecules can be described as follows: every molecule (isolated in terms of quantum chemistry) is an element of a family (in terms of traditional organic chemistry) represented by a point on a regression line of the general form

$$E(N) = hN + s \tag{1}$$

where N is the total number of electrons, h is the slope and s is the axis intercept of the regression line. The family mentioned above is again an element of a set of homologous series (same homologous unity) formed by regression equations of identical slope h, which means that regression equations of homologous families containing the same homologous unity are all strongly parallel but differing by the intercept s. Examples of considered compounds and families are collected in Table 1. The regression equations were derived by the linear least-squares method. For evaluation of the linearity of Eqn. (1) the linear correlation coefficient (R) and the standard error of estimate (*esd*) were considered. For all correlations determined by us is R close to -1.0 ($R < -0.9999999$) and *esd* is low (*esd* <0.007 hartree = 0.4 kcal mol^{-1}), which means a very good linear correlation between calculated total energies and the number of electrons. Therefore two parameters (h and s) were derived statistically which are sufficient to characterize homologous series [h, the slope of the regression Eqn. (1)] and compound classes via s, the axis intercept of the regression Eqn. (1).

The total energies for the ground state of 98 organic molecules which are listed in Table 1 were computed by four different theoretical methods, DFT (BLYP and B3LYP), and partly by HF and post HF MP2, as mentioned before. The influence of methods and basis sets on the regression parameters h and s are small and the classification scheme works satisfactorily for all examined four levels of theory. Therefore it is advantageous to use the less demanding DFT/6–31G** or DFT/cc-pVDZ method/basis set combinations, because the treatment of electron correlation is included and the procedures are computationally rather inexpensive. For a self-contained classification it is important to compute all energies of considered molecules at the same level of theory. Table 2 shows the regression parameters *h, s* and *esd* based on DFT BLYP/6–31G** and B3LYP/6–31G** calculations. The homology h parameters for 17 families with the same CH_2-homologous units are highly constant for the CH_2

Table 1. Collection of organic molecules considered for the classification scheme based on total energies

- I. **CH_2-homologous classes of compounds (families) with units of 8 electrons**
  - Hydrocarbons:
    - **alkanes (staggered)**, C_nH_{2n+2} ($n = 1$–8, 15, 25)
      methane, ethane, propane, butane, pentane, hexane, heptane, octane, pentadecane, pentacosane
    - **alkanes (all CH bonds eclipsed)**, C_nH_{2n+2} ($n = 2$–4)
      ethane, propane, butane
    - **alkenes (*trans*)**, C_nH_{2n} ($n = 2$–6)
      ethene, propene, 2-butene, 2-pentene, 2-hexene
    - **alkynes**, C_nH_{2n-2} ($n = 2$–6)
      ethyne, propyne, 2-butyne, 2-pentyne, 2-hexyne
    - **cycloalkanes**, C_nH_{2n} ($n = 3$–6)
      cyclopropane, cyclobutane, cyclopentane, cyclohexane
  - Chlorides:
    - **alkyl chlorides**, $C_nH_{2n+1}Cl$ ($n = 1$–4)
      methyl chloride, ethyl chloride, 1-propyl chloride, 1-butyl chloride
  - Molecules containing CO bonds:
    - **alcohols (primary)**, $C_nH_{2n+2}O$ ($n = 1$–4)
      methanol, ethanol, 1-propanol, 1-butanol
    - **alcohols (secondary)**, $C_nH_{2n+1}O$ ($n = 3$–5)
      isopropanol, 2-butanol, 2-pentanol
    - **alcohols (tertiary)**, $C_nH_{2n+2}O$ ($n = 3$–5)
      tert.-butanol, 1,1-dimethylpropanol, 1,1-dimethylbutanol
    - **aldehydes**, $C_nH_{2n}O$ ($n = 1$–4)
      formaldehyde, acetaldehyde, propanal, butanal
    - **ketones**, $C_nH_{2n}O$ ($n = 3$–6)
      acetone, 2-butanone, 2-pentanone, 2-hexanone
    - **carboxylic acids**, $C_nH_{2n}O_2$ ($n = 1$–4)
      formic acid, acetic acid, propionic acid, butyric acid
    - **methyl esters**, $C_nH_{2n+2}O_2$, ($n = 2$–5)
      methyl formiate, methyl acetate, methyl propionate, methyl butyrate
  - Molecules containing CS bonds:
    - **thiols**, $C_nH_{2n+2}S$ ($n = 0$–3)
      hydrogen sulfide, methanethiol, ethanethiol, propanethiol
  - Molecules containing CN bonds:
    - **amines (primary)**, $C_nH_{2n+3}N$ ($n = 0$–3)
      ammonia, methylamine, ethylamine, propylamine
    - **amides (primary)**, $C_nH_{2n+1}NO$ ($n = 1$–3)
      formamide, acetamide, propionamide
    - **nitroalkanes (primary)**, $C_nH_{2n+1}NO_2$ ($n = 1$–4)
      nitromethane, nitroethane, 1-nitropropane, 1-nitrobutane
- II. **Conjugated alkenes (homologous units of 14 electrons)**
  - **conjugated alkenes (all-*trans*)**, $C_{2n+2}H_{2n+4}$ ($n = 0$–2)
    ethene, 1,3-butadiene, 1,3,5-hexatriene
  - **conjugated alkenes (all-*trans*)**, $C_{2n+2}H_{2n+4}$ ($n = 1$–4)
    (*E*)-2-butene, (*E,E*)-2,4-hexadiene, (*E,E,E*)-2,4,6-octatriene, (*E,E,E,E*)-2,4,6,8-decatetraene
  - **conjugated alkenes (all-cis, s-trans)**, $C_{2n+2}H_{2n+4}$ ($n = 1$–4)
    (*Z*)-2-butene, (*Z,Z*)-2,4-hexadiene, (*Z,Z,Z*)-2,4,6-octatriene, (*Z,Z,Z,Z*)-2,4,6,8-decatetraene
- III. **Cumulenes (homologous units of 6 electrons)**
  - **cumulenes**, C_nH_4 ($n = 2$–6)
    ethene, allene, butatriene, pentatetraene, hexapentaene
- IV. **Benzenoides (homologous units of 26 electrons)**
  - **linear acenes**, $C_{4n-2}H_{2n+2}$ ($n = 2$–5)
    benzene, naphthalene, anthracene, tetracene
  - **α-phenols**, $C_{4n-2}H_{2n+2}O$ ($n = 2$–5)
    phenol, 1-naphthol, 1-hydroxyanthracene, 1-hydroxytetracene
- V. **Oligovinyl chlorides (homologous units of 32 electrons)**
  - **oligovinyl chlorides**
    chlorethane, 1,3-dichlorbutane, 1,3,5-trichlorhexane, 1,3,5,7-tetrachloroctane

series with units of eight electrons with average values of −4.9113 for BLYP and −4.9148 for B3LYP calculations (all values in hartree units). The largest deviations are shown by the alkane family with unnatural all-eclipsed CH bonds (−4.9100 or −4.9135) and the series of partly strained cycloalkanes (−4.9130 or −4.9166). For con-

Table 2. Parameters h and s in hartree (= 627.5095 kcal mol^{-1}) for the classification scheme [Eqn. (1)] based on numbers of electrons and on total energies obtained by BLYP and B3LYP calculations using the 6–31G** basis set (the CH_2 homologous series are ordered by their s parameters; the linear correlation coefficients R are >0.9999999 in all cases examined)

	BLYP/6–31G**			B3LYP/6–31G**		
Family	*h*	*s*	*esd*	*h*	*s*	*esd*
Homology for 8 electrons:						
Alkanes (staggered)[a]	−4.9111	8.6273	0.0005	−4.9146	8.6229	0.0005
Alkanes (all eclipsed)	−4.9100	8.6126	0.0020	−4.9135	8.6077	0.0018
Cycloalkanes	−4.9130	0.0922	0.0073	−4.9166	0.0973	0.0077
trans-Alkenes	−4.9113	0.0345	0.0038	−4.9148	0.0397	0.0037
Amines (prim.)	−4.9109	−7.4157	0.0045	−4.9144	−7.4105	0.0047
Alkynes	−4.9119	−8.5346	0.0062	−4.9153	−8.5208	0.0060
Alcohols (prim.)	−4.9113	−27.2762	0.0022	−4.9148	−27.2598	0.0023
Alcohols (sec.)	−4.9111	−27.2892	0.0002	−4.9146	−27.2730	0.0001
Alcohols (tert.)	−4.9110	−27.2977	0.0006	−4.9145	−27.2816	0.0005
Aldehydes	−4.9117	−35.8916	0.0061	−4.9152	−35.8641	0.0061
Ketones	−4.9111	−35.9247	0.0004	−4.9146	−35.8973	0.0005
Amides	−4.9117	−51.9723	0.0033	−4.9152	−51.9330	0.0036
Methyl esters	−4.9115	−71.8411	0.0042	−4.9151	−71.7899	0.0045
Carboxylic acids	−4.9116	−71.8528	0.0045	−4.9151	−71.8036	0.0048
Nitroalkanes	−4.9111	−87.8165	0.0057	−4.9147	−87.7431	0.0020
Thiols	−4.9110	−310.9632	0.0008	−4.9145	−310.9297	0.0011
Alkyl chlorides	−4.9112	−372.3713	0.0016	−4.9147	−372.3311	0.0015
Average	−4.9113	—	0.0032	−4.9148	—	0.0030
Homology for 14 electrons:						
Conj. alkenes (all-*trans*)[b]	−5.5265	9.8813	0.0008	−5.5292	9.8736	0.0007
Conj. alkenes (all-*trans*)[c]	−5.5266	19.7188	0.0009	−5.5293	19.6982	0.0008
Conj. alkenes (all-*cis*)[c]	−5.5264	19.7134	0.0006	−5.5290	19.6930	0.0005
Average	−5.5265	—	0.0008	−5.5292	—	0.0007
Homology for 26 electrons:						
Linear acenes	−5.9070	15.9430	0.0027	−5.9093	15.9304	0.0031
Phenols	−5.9068	−12.0161	0.0019	−5.9093	−12.0160	0.0022
Average	−5.9069	—	0.0023	−5.9093	—	0.0027
Homology for 6 electrons:						
Cumulenes	−6.3440	22.9640	0.0024	−6.3455	22.9358	0.0020
Homology for 32 electrons:						
Oligovinyl chlorides	−16.8174	32.4388	0.0019	−16.8194	32.4274	0.0006

[a] C_{15} and C_{25} included.
[b] No methyl groups at each end of the molecules forming the homologous series.
[c] Methyl groups at each end of the molecules forming the homologous series.

jugated alkenes as other homologous series, differing by C_2H_2 units corresponding to 14 electrons, the average h values are −5.5265 and −5.5292, respectively. The acene series, differing by C_4H_2 units with 26 electrons, are characterized by average h values of −5.9069 and −5.9093. The largest h values are found for the cumulene series differing by one carbon atom, equivalent to six electrons, with −6.3440 and −6.3455. As already mentioned, the *esd* values are less than 0.0007 hartree. The homologous compound families with different functional groups are characterized by their specific s values. These vary in the range 23 to −372 hartree and cluster for chemically related compounds in narrow ranges, i.e. staggered and eclipsed alkanes, primary, secondary and tertiary alcohols (for the relation to isomers, see later), aldehydes and ketones or carboxylic acids and methyl esters. Figure 2 shows a graphical plot of the correlation for total energies of a selection of different homologous series with regard to the B3LYP method and the 6–31G** basis set. Figure 3 shows the graphics of the correlation for a selection of only CH_2-homologous families. This represents the parallelism of regression equations of homologous families for different functional groups containing the same CH_2-homologous unity.

In Table 3, a few corresponding statistical CH_2-homologous values for the HF/6–31G** and MP2/6–31G** methods are given. Linearity and high statistical relevance are also observed. The numerical values for the h and s parameters are of the same order of magnitude as in Table 2 and the *esd* values are less than 0.005 hartree. Average CH_2-homology h values are −4.8800 for HF and −4.8979 hartree for the MP2 calculations. The *esd* values are slightly better than those from DFT calculations.

Values for the Dunning basis set cc-pVTZ for DFT BLYP and B3LYP calculations are presented in Table 4. Three classes of homology h parameters determined by DFT calculations in Table 2 are also found for the extended cc-pVTZ basis set calculations with average values of

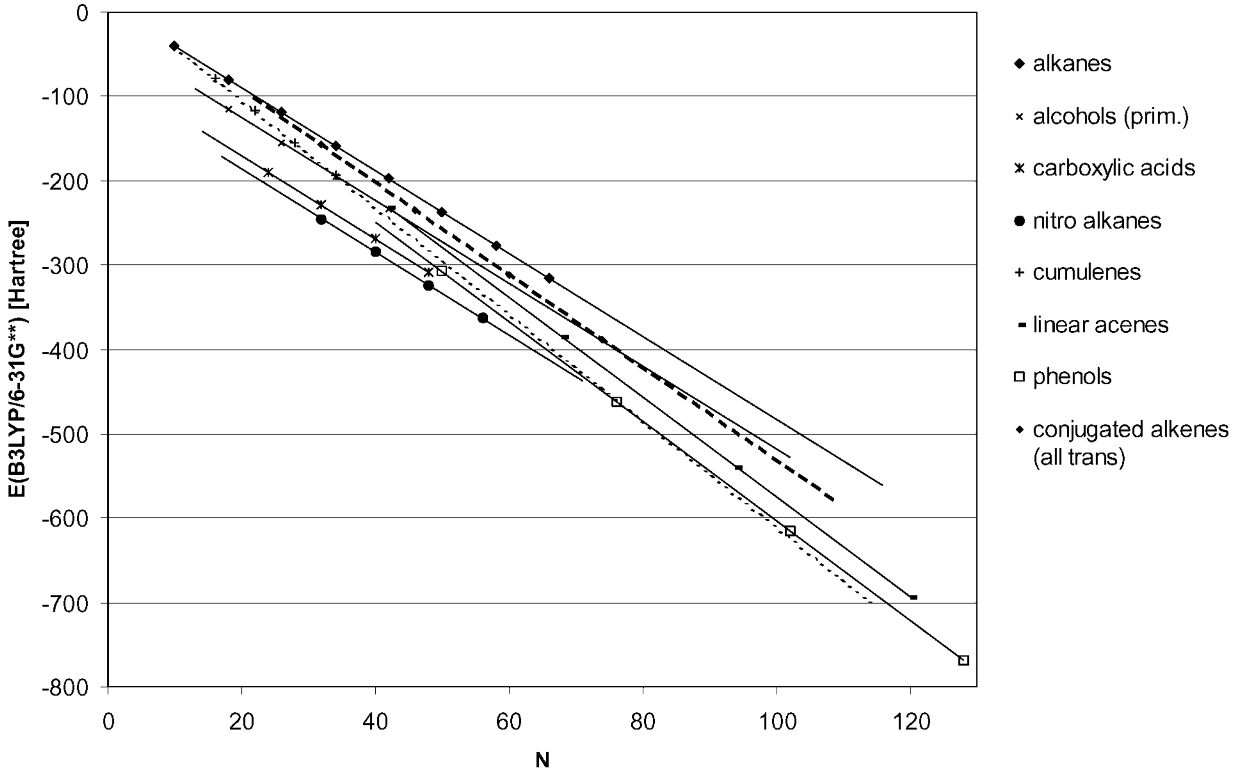

Figure 2. Plot of linear regressions for total energies of selected homologous series determined by the B3LYP/6–31G** level of theory

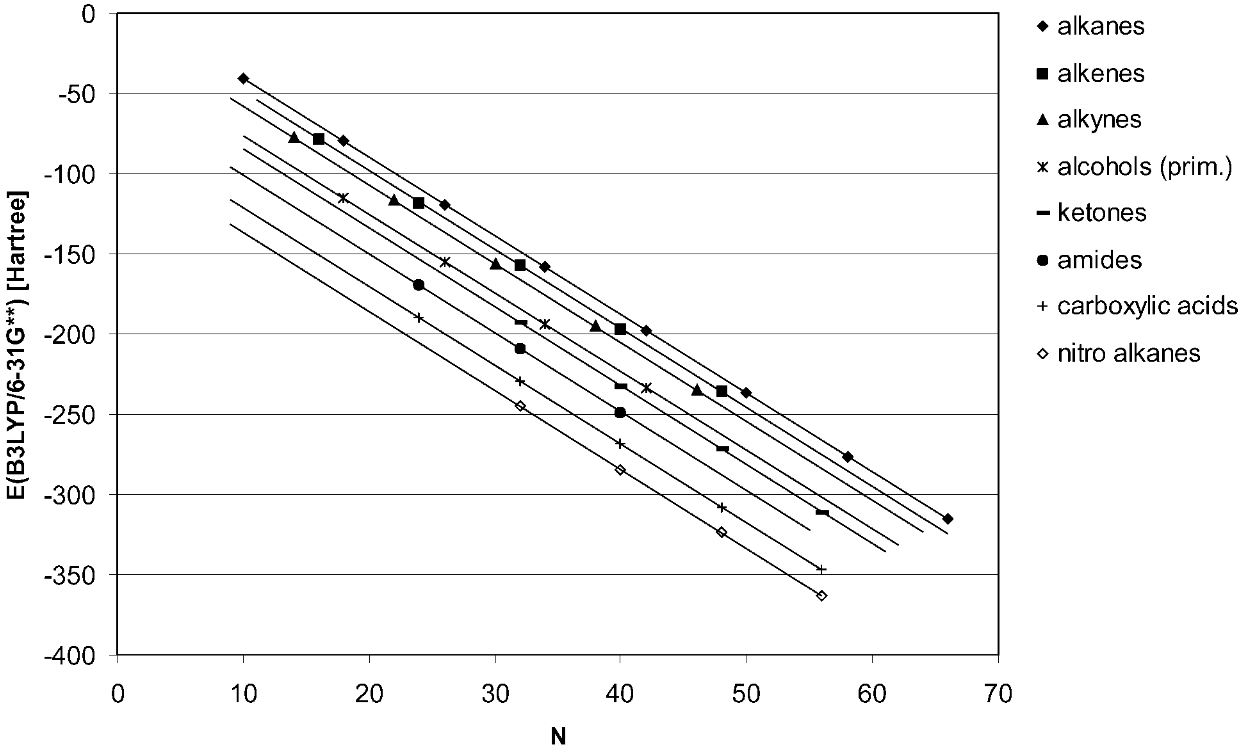

Figure 3. Plot of linear regressions for total energies of selected methylene homologous series determined by the B3LYP/6–31G** level of theory

Table 3. Parameters *h* and *s* for the classification scheme [Eqn. (1)] for some CH_2-homologous families based on total energies obtained by HF and MP2 calculations using the 6–31G** basis set (the linear regression coefficient is >0.9999999 for all examined cases)

	HF/6–31G**			MP2/6–31G**		
Family	*h*	*s*	*esd*	*h*	*s*	*esd*
Alkanes, staggered C_nH_{2n+2} (n = 1–8, 15, 25)	−4.8797	8.5962	0.0003	−4.8975	8.6156	0.0011
trans-**Alkenes** C_nH_{2n} (n = 2–6) double bond in 2-position	−4.8798	0.0363	0.0031	−4.8979	0.0467	0.0030
Alkynes C_nH_{2n-2} (n = 2–6) triple bond in 2-position	−4.8804	−8.5010	0.0052	−4.8983	−8.5090	0.0042
Alcohols, primary $C_nH_{2n+2}O$ (n = 1–4)	−4.8799	−27.2097	0.0022	−4.8979	−27.2211	0.0020

−4.9128 for BLYP and −4.9162 for B3LYP in the CH_2-homologous series. Average *h* values for the conjugated alkene homology are −5.5283 and −5.5309 and for the cumulene family −6.3464 and −6.3477 hartree. The *s* values for each functional group are very similar to those presented in Tables 2 and 3 and the *esd* values are in the same range. Therefore, for applications no large basis set calculations such as cc-pVTZ are necessary.

In principle, one needs only one computation for a member of a new family with known homologous unity (if the average $\overline{h}$ is already known) to obtain the compound class parameter *s*. However, this method is less exact than the generation of *h* and *s* by linear regression using as many homologous compounds as possible. Our regression parameters ($\overline{h}$ as above or *h* and *s* as collected in Tables 2–4) may be used to estimate total energies of larger molecules without quantum chemical calculations.

Behavior of *E/N* versus *N*

We now look at the behavior of a normalized function, the energy per electron (*E/N*) of a homologous series. This is not a constant value but a function with a particular limiting behavior. For example, in the alkane series for the B3LYP/6–31G** calculations the (*E/N*) values decrease on going to larger alkanes. The function (*E/N*) versus *N* could be fitted analytically to the function

$$\frac{E}{N} = \left(\frac{E}{N}\right)_{\text{limit}} + bN^{-1} \qquad (2)$$

with $(E/N)_{\text{limit}} = -4.91453$ and $b = 8.62174$. The limit value $(E/N)_{\text{limit}}$ is a lower bound for $N \to \infty$ and is similar to the *h* parameter (−4.9146 hartree). For the alkyne series the (*E/N*) value increases on going to larger alkynes [$(E/N)_{\text{limit}} = -5.06985$]. Here, the $(E/N)_{\text{limit}}$ value is an

Table 4. Parameters *h* and *s* for the classification scheme [Eqn. (1)] based on total energies obtained by BLYP and B3LYP DFT calculations using the cc-pVTZ Dunning basis set (the CH_2 homologous series are ordered by their s parameters; the linear correlation coefficients *R* are >0.9999999 in all cases examined)

	BLYP/cc-pVTZ			B3LYP/cc-pVTZ		
Family	*h*	*s*	*esd*	*h*	*s*	*esd*
Alkanes C_nH_{2n+2} (n = 1–5)	−4.9126	8.6255	0.0006	−4.9159	8.6219	0.0006
trans-**Alkenes** C_nH_{2n} (n = 2–6) double bond in 2-position	−4.9128	0.0269	0.0035	−4.9162	0.0326	0.0034
Amines, primary $C_nH_{2n+3}N$ (n = 0–3)	−4.9124	−7.4295	0.0052	−4.9157	−7.4236	0.0055
Alkynes C_nH_{2n-2} (n = 2–6) triple bond in 2-position	−4.9133	−8.5501	0.0056	−4.9166	−8.5372	0.0066
Alcohols, primary $C_nH_{2n+2}O$ (n = 1–4)	−4.9128	−27.2997	0.0024	−4.9162	−27.2826	0.0023
Methyl esters $C_nH_{2n+2}O_2$ (n = 1–4)	−4.9130	−71.8845	0.0043	−4.9164	−71.8334	0.0045
Conjugated alkenes (all-*trans*) $C_{n+2}H_{2n+4}$ (n = 1–4)	−5.5284	19.7229	0.0009	−5.5310	19.7020	0.0012
Conjugated alkenes (all-*cis*) $C_{n+2}H_{2n+4}$ (n = 1–4)	−5.5282	19.7176	0.0006	−5.5308	19.6986	0.0005
Cumulenes C_nH_4 (n = 2–6)	−6.3464	22.9685	0.0025	−6.3477	22.9414	0.0021

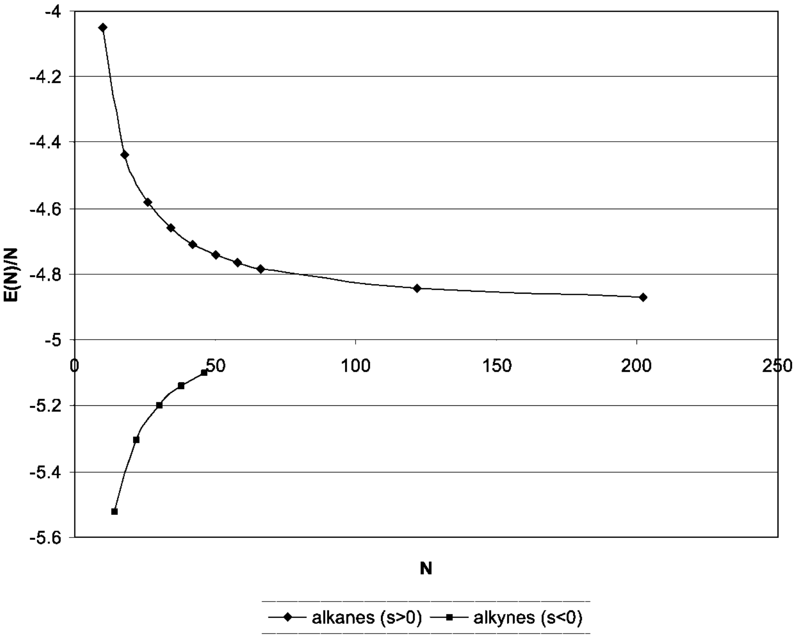

Figure 4. Graphical representation of energy per electron (*E/N*) values for homologous series (alkanes and alkynes). The criterion for decreasing (*E/N*) values is $s > 0$ (alkane series) and for increasing (*E/N*) values $s < 0$ (alkyne series)

upper bound for $N \to \infty$. Figure 4 shows graphically the behavior of (*E/N*) versus the number of electrons of homologous alkane and alkyne series, and Fig. 5 shows the same relation for benzenoids (linear acenes and α-phenols).

We find that generally *E/N* varies with the sign of *s*. It decreases for homologous series with $s > 0$ [$(E/N)_{\text{limit}}$ is a lower bound] and increases for homologous series with s <0 [$(E/N)_{\text{limit}}$ is an upper bound].

Application of the classification scheme

Prediction of total energies. Total energies were predicted, as already mentioned, by means of *ab initio* MO calculations by Ibrahim and Schleyer[9] and Wiberg.[7] Ibrahim and Schleyer used the conception of additive atomic equivalents whereas Wiberg used the concept of additive group equivalents. These equivalents also lead to the prediction of thermochemical data. The prediction of total energies is equivalent to predicting heats of formation for members of homologous series.

Predictions in the alkane family. The linearity of the regression lines of homologous compounds can be used to predict estimated total energies of homologous compounds from the *h* and *s* parameters determined for a set of known total energies of smaller compounds which belong to the same homologous series. From these values, extrapolated total energies even of large molecules are accessible within chemical accuracy. This will be demonstrated using the alkane series: the classification

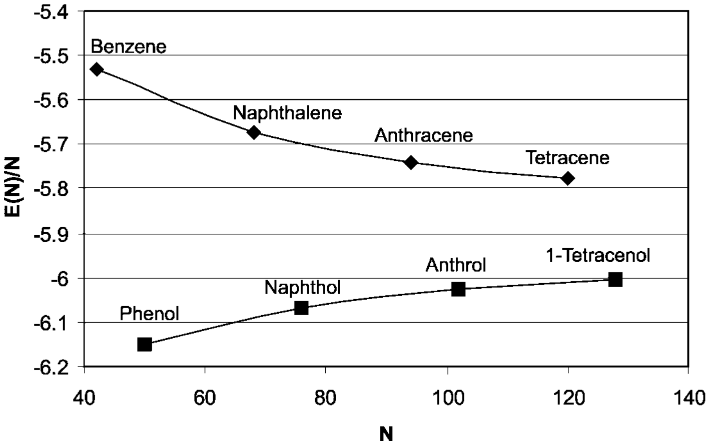

Figure 5. Graphical representation of energy per electron (*E/N*) values for benzenoids (linear acenes, $s > 0$; phenols, $s < 0$)

Table 5. Approximated (E_{approx}) and calculated (E_{opt}) total energies (hartree) of $C_{15}H_{32}$ and $C_{25}H_{52}$ for BLYP, B3LYP, HF and MP2 methods and the 6–31G** basis set [approximated values were obtained by Eqn. (1) and molecules methane to octane; optimized values were determined by fully gradient optimizations of $C_{15}H_{32}$ and $C_{25}H_{52}$ at the respective level of theory]

Compound	Method	E_{approx}[a]	E_{opt}	ΔE[b]	ΔE[b] (kcal mol^{-1})
$C_{15}H_{32}$	BLYP	−590.5220790	−590.5231090	0.0010	0.646
	B3LYP	−590.9525499	−590.9536479	0.0011	0.684
	HF	−586.7283310	−586.7288696	0.0005	0.339
	MP2	−588.9076580	−588.9102660	0.0026	1.638
$C_{25}H_{52}$	BLYP	−983.4066505	−983.4087830	0.0021	1.336
	B3LYP	−984.1165137	−984.1187988	0.0023	1.431
	HF	−977.1049520	−977.1060714	0.0011	0.703
	MP2	−980.7254990	−980.7310138	0.0055	3.464

[a] For approximated values the following h and s values were used in Eqn. (1):
BLYP: $h = -4.91077$; $s = 8.63261$
B3LYP: $h = -4.91431$; $s = 8.62761$
HF: $h = -4.87971$; $s = 8.59601$
MP2: $h = -4.89772$; $s = 8.61455$.

[b] $\Delta E = E_{approx} - E_{opt}$.

parameters h and s presented in Tables 2 and 3 for the molecules methane to octane were used to predict the total energies of pentadecane ($C_{15}H_{32}$) and pentacosane ($C_{25}H_{52}$). These approximated total energies were directly compared with calculated values using the respective level of theory. Both kinds of values for the BLYP, B3LYP, HF and MP2 methods and 6–31G** basis set are shown in Table 5. In both cases the deviations are largest for the MP2 optimization, underestimated by 1.64 and 3.46 kcal mol^{-1}, and smallest for the HF method with 0.34 and 0.70 kcal mol^{-1}. The DFT results are between these values. Keeping in mind the enormous extrapolation from C_8 to C_{15} and even to C_{25}, the observed deviations are tolerable and of experimental accuracy.

Estimation of total energies of polymers. Another example of predictions of total energies is the estimation of total energies of polymers. The h parameter of alkyl chlorides can be used to predict the total energy of polyvinylchloride (PVC). The total energy of PVC containing 100 000 monomer units (assuming an unbranched chain of monomers) is -5.3822×10^7 hartree.

A further sample for predictions of total energies of polymers is low-density polyethylene (LDPE) which consists of mainly unbranched alkane chains with molecular masses of 50 000–100 000 g mol^{-1}. Using h and s values obtained from B3LYP/6–31G** calculations, the total energy of a PE molecule consisting of 7130 monomer units (CH_2 groups) ($\sim$ 100 000 g mol^{-1}) is −280 330 hartree.

Isomerism of organic molecules. *Alkanes.* In this section, we apply our classification scheme to isomeric molecules. First, we consider the three pentane isomers *n*-pentane, isopentane and neopentane. Any of these isomers is a member of a CH_2 homologous series with the same h parameter. However, the s parameters are slightly different for all isomers. From B3LYP/6–31G** calculations ($\overline{h} = -4.91494$) we obtained absolute differences [s(pentane) $-s$(isomer)] of 0.00061 hartree (0.382 kcal mol^{-1}) for isopentane and 0.00138 hartree (0.866 kcal mol^{-1}) for neopentane, which denote the energy differences between these isomers.

Because of the high linearity of regression equations, energy differences between other isomeric systems which are members of the homologous series mentioned above are also known. An example for this is *n*-hexane–2-methylpentane predicted 0.00061 hartree = 0.382 kcal mol^{-1} and *n*-hexane–2,2-dimethylbutane predicted 0.00138 hartree = 0.866 kcal mol^{-1}.

Alcohols. Another example of isomeric series of compounds is primary, secondary and tertiary alcohols. Let us assume that all types of alcohols mentioned above are members of a CH_2-homologous series, so all h parameters are identical ($\overline{h} = -4.91494$). Using the particular s parameters, the energy differences between isomeric compounds can be easily predicted. In the following, B3LYP/6–31G** calculations are used again for estimations of h and s values. The energy differences between primary and secondary alcohols are 0.0065 hartree (4.08 kcal mol^{-1}) and between primary and tertiary alcohols 0.0101 hartree (6.34 kcal mol^{-1}).

Prediction of atomization energies (bond energies). The atomization energy is defined as the difference between the sum of atomic energies and the total energy of the molecule. A suggestion for the estimation of experimental heats of formation via atomization energies or bond energies was presented by Cohnen and Benson[3] in 1993. They introduced an additivity scheme, which subdivides an arbitrary chemical compound into a set of

Table 6. Collection of atomic energies (hartree) for C, H and O atoms

Energy	BLYP/ 6–31G**	B3LYP/ 6–31G**	BLYP/ cc-pVTZ	B3LYP/ cc-pVTZ	HF/ 6–31G**	MP2/ 6–31G**
E_C	−37.826119	−37.846279	−37.844469	−37.858575	−37.680860	−37.733829
E_H	−0.495446	−0.500273	−0.497555	−0.502156	−0.498233	−0.498233
E_O	−75.046647	−75.060611	−75.079907	−75.091857	−74.783932	−74.880037

smaller structural units in such a way that the thermochemical properties of that compound can be calculated from constants associated with the smaller units.

The prediction of *ab initio* calculated total energies described in the last section also offers direct access to estimations of atomization energies. These can also be determined experimentally, so direct comparisons of predicted and experimental values are possible. Experimental atomization energies at 0 K were taken from the compilation of Cox and Pilcher.[21] For each homologous series one can formulate an equation for the prediction of atomization energies dependent on the number of electrons of the members of the homologous series. Hence all atomization energies of a homologous series can be computed using this equation. Equations (3)–(8) are used for the determination of atomization energies using the homology parameter *h* and the compound class parameter *s* for a selection of homologous series:

$$\Delta H_a = \left(\frac{N-2}{8}\right)(E_C + 2E_H) + 2E_H - hN - s \qquad \text{alkanes} \quad (3)$$

$$\Delta H_a = (E_C + 2E_H)\frac{N}{8} - hN - s \qquad \text{alkenes} \quad (4)$$

$$\Delta H_a = \left[(E_C + 2E_H)\left(\frac{N+2}{8}\right) - 2E_H\right] - hN - s \qquad \text{alkynes} \quad (5)$$

$$\Delta H_a = (E_C + 2E_H)\left(\frac{N-10}{8}\right) + 2E_H + 2E_O - hN - s \qquad \text{primary alcohols} \quad (6)$$

$$\Delta H_a = E_C\left(\frac{N-4}{6}\right) + 4E_H - hN - s \qquad \text{cumulenes} \quad (7)$$

$$\Delta H_a = \left[4\left(\frac{N+10}{26}\right) - 2\right]E_C + \left[2\left(\frac{N+10}{26}\right) + 2\right] \times E_H - hN - s \qquad \text{linear acenes} \quad (8)$$

where ΔH_a denotes the atomization energy to be predicted, *N* is the number of electrons of the considered homologous compound, E_C, E_H and E_O are the basis set- and method-dependent total energies of C, H and O atoms in their atomic ground state as shown in Table 6 and *h* and *s* are the classification parameters derived above. $hN + s$ represents the estimated or calculated total energy [see Eqn. (1)]. Table 7 shows examples of predicted and experimental atomization energies in the alkane series. The classification parameters *h* and *s* and E_C, E_H and E_O were obtained by B3LYP/6–31G** calculations.

Table 7 shows also direct calculated ΔH_a values obtained using the B3LYP/6–31G** level of theory. Therefore, the sum of the atomic energy of a considered molecule was subtracted from the B3LYP/6–31G** optimized value. The deviations from ΔH_a predictions using the *h* and *s* parameters are small (0.93 kcal mol^{-1} on average).

Dissection into partial energies

To gain a deeper insight into the mechanisms of the classification scheme presented in the previous section, this part consideres the role of partial energies forming the total energy of a chemical system. The total energy can be divided into the nuclear repulsion contribution (E_{NR}), the one-electron energy of α ($E_{1\alpha}$) and β ($E_{1\beta}$) spin electrons, the correlation (E_C), the exchange (E_X) and the coulomb (E_J) contributions:

$$E = E_{NR} + E_{1\alpha} + E_{1\beta} + E_C + E_X + E_J \qquad (9)$$

In the following we used the BLYP/6–31G** level of theory, so the correlation energy E_C was computed using the LYP correlation functional and the exchange energy E_X was computed using the Becke88 exchange functional. For investigations of partial energies we used a

Table 7. Prediction of atomization energies using the parameters *h* and *s* of the classification scheme (all values in kcal mol^{-1})

Compound	*N*	ΔH_a[a]	ΔH_a, predicted[b]	ΔH_a, exp[c]	$\Delta\Delta H_a$[d]
Methane	10	424.60	424.03	418.95	5.08
Ethane	18	718.21	718.94	710.75	8.19
Propane	26	1013.02	1013.85	1004.23	9.62
Butane	34	1307.80	1308.77	1299.17	9.60
Pentane	42	1602.54	1603.68	1593.19	10.49
Hexane	50	1897.28	1898.60	1887.47	11.13

[a] $\Delta H_a = \sum E_{atom} - E_{molecule}$.

[b] ΔH_a, predicted by Eqn. (2) with $h = -4.9146$ and $s = 8.6229$ derived from B3LYP/6–31G** calculations.

[c] Exp. values from Ref. 21 based on 0 K.

[d] $\Delta\Delta H_a = \Delta H_{a,predicted} - \Delta H_{a,exp}$.

Table 8. Energy partition from BLYP/6–31G** optimizations of selected molecules from the set of 98 molecules listed in Table 1 (all values in hartree)

Molecule	E_{NR}	E_1[a]	E_C	E_X	E_J	E_{total}
Methane	13.33193	−79.65961	−0.294286	−6.564951	32.702245	−40.484674
Ethane	41.85471	−188.58700	−0.551097	−12.454972	79.966865	−79.771489
Propane	81.718128	−320.19732	−0.808788	−18.343407	138.571229	−119.060171
Butane	129.249312	−467.14392	−1.066548	−24.231802	204.844132	−158.348828
Pentane	182.854491	−626.24292	−1.324231	−30.119508	277.194750	−197.637417
Hexane	241.455673	−795.32816	−1.582113	−36.009701	354.538322	−236.925987
Heptane	304.151136	−972.60634	−1.839856	−41.898183	435.978674	−276.214569
Ethene	33.124392	−169.91372	−0.498102	−11.735445	70.479475	−78.543397
Propene	70.141014	−295.81808	−0.755986	−17.627238	126.221827	−117.838473
(*E*)-2-Butene	114.797305	−437.04172	−1.013883	−23.518494	189.644029	−157.132763
(*E*)-2-Pentene	166.320379	−591.97972	−1.271274	−29.407629	259.920475	−196.417772
(*E*)-2-Hexene	223.303837	−757.82248	−1.529116	−35.296770	335.637981	−235.706560
Methanol	39.835046	−236.12178	−0.595102	−14.850320	96.054814	−115.677344
Ethanol	80.945114	−370.24238	−0.853027	−20.740505	155.919200	−154.971605
Propanol	129.041663	−518.29800	−1.110791	−26.630021	222.737090	−194.260067
Butanol	183.018705	−678.12786	−1.368535	−32.519104	295.448311	−233.548493
Formic acid	69.214789	−396.48194	−0.844525	−22.412822	160.797473	−189.727027
Acetic acid	119.582353	−549.17354	−1.102965	−28.313038	229.97963	−229.027562
Propionic acid	176.158692	−714.18820	−1.360995	−34.202075	305.27544	−268.317136
Butyric acid	235.566387	−884.84369	−1.618743	−40.090892	383.381218	−307.605718
Allene	58.744268	−271.69280	−0.703225	−16.907290	113.954821	−116.604224
Butatriene	88.229088	−381.17521	−0.907688	−22.083289	161.267937	−154.669163
Pentatetraene	121.073975	−497.32456	−1.112699	−27.260649	211.887074	−192.736863

[a] $E_1 = E_{1\alpha} + E_{1\beta}$.

Table 9. Statistical parameters (h_i and s_i derived from linear regression for partial energies $E_i = h_i N + s_i$ (in hartree = 627.5095 kcal mol^{-1}) against the number of electrons N for BLYP/6–31G** calculations (R = linear correlation coefficient; *esd* = estimated standard deviation)

	Family				
Statistical parameter	Alkanes	Alkenes	Alcohols (prim.)	Carboxylic acids	Cumulenes
$R(E_{N\ R})$	0.99317	0.99656	0.99819	0.99932	0.99849
$R(E_1)$	−0.99706	−0.99855	−0.99924	−0.99969	−0.99957
$R(E_C)$	−0.99999	−0.99999	−0.99999	−0.99999	−0.99999
$R(E_X)$	−0.99999	−0.99999	−0.99999	−0.99999	−0.99999
$R(E_J)$	0.99641	0.99822	0.99907	0.99963	0.99943
$R(E_{total})$	−0.99999	−0.99999	−0.99999	−0.99999	−0.99999
$esd(E_{N\ R})$	13.6270	7.23255	4.55252	3.24047	2.55567
$esd(E_1)$	27.2556	14.4641	9.09601	6.43848	5.08330
$esd(E_C)$	0.00033	0.00016	0.00007	0.00025	0.00019
$esd(E_X)$	0.00066	0.00118	0.00039	0.00439	0.00200
$esd(E_J)$	13.6284	7.23377	4.54513	3.19793	2.52711
$esd(E_{total})$	0.00057	0.00378	0.00226	0.00451	0.00244
$h(E_{N\ R})$	6.1286	5.9567	5.9706	6.9454	4.8889
$h(E_1)$	−18.7427	−18.3997	−18.4259	−20.3762	−18.1952
$h(E_C)$	−0.0322	−0.0322	−0.0322	−0.0323	−0.0341
$h(E_X)$	−0.7361	−0.7363	−0.7362	−0.7365	−0.8625
$h(E_J)$	8.4714	8.3002	8.3125	9.2881	7.8589
$h(E_{total})$	−4.9111	−4.9113	−4.9113	−4.9116	−6.3441
$s(E_{N\ R})$	−66.2829	−69.0779	−70.9077	−99.9035	−46.9294
$s(E_1)$	144.4287	138.2765	102.0802	97.3731	124.8547
$s(E_C)$	0.0283	0.0172	−0.0151	−0.0705	0.0480
$s(E_X)$	0.7954	0.0441	−1.5990	−4.7392	2.0665
$s(E_J)$	−70.3427	−69.2255	−56.8345	−64.5127	−57.0760
$s(E_{total})$	8.6268	0.0344	−27.2762	−71.8528	22.9638

substantially smaller set of five homologous series with individual molecules presented in Table 8, which shows the computed individual partial energies. Table 9 presents the linear correlation statistics $E_i(N) = h_iN + s_i$ for each partial energy E_i, where h_i denotes the slope and s_i the axis intercept of the regression equation. Table 9 shows that the correlations for E_{NR}, E_1 and E_{J} are statistically much lower than those for E_{C}, E_{X} and E_{total}. The values $h(E_{\mathrm{C}})$ and $h(E_{\mathrm{X}})$ values are remarkably constant and relatively small for the $\mathrm{CH_2}$-homologous series but different for the cumulene series. The largest values with different signs are observed for $h(E_1) > h(E_{\mathrm{J}}) > h(E_{\mathrm{NR}})$. The contributions due to correlation (E_{C}) and exchange (E_{X}) to the total energy are substantially lower than those for E_1, E_{J} and E_{NR}. The s_i values differ strongly from family to family but with the same order of magnitude as in the h_i partition.

Another general statement is the behavior of the slopes h_i and the axis intercepts s_i of partial energies, which are strictly additive:

$$h = \sum_i h_i \tag{10}$$

$$s = \sum_i s_i \tag{11}$$

Therefore, the linear regression equation used for the classification of organic compounds [Eqn. (1)] can be written as

$$E(N) = \left(\sum_i h_i\right)N + \sum_i s_i \tag{12}$$

CONCLUSIONS

We developed a chemical concept to combine the systematics for families of homologous compounds with different functional groups of traditional organic chemistry with the quantum chemical treatment of distinct single molecules. A statistical approach of fundamental quantum chemical properties such as total energies was carried out by correlating these quantities against the number of electrons for families of homologous series. The classification of these families was derived using the slopes (h) as homology parameters and the axis intercepts (s) as functional group dependent substance parameters of the derived linear regression equations.

In summary, our approaches provide a simple opportunity to classify families of homologous organic compounds by means of quantum chemical parameters which may used in numerical extrapolations for total energies of larger systems (even polymers) and predictions of thermochemical data such as energies of atomization. Isomers in different homologous series with the same h values are characterized by differences in their s values.

Acknowledgement

We thank Dr Wolfhard Koch, Institute of Physical and Theoretical Chemistry, University of Tübingen, for valuable comments.

REFERENCES

1. *How to use the Beilstein Handbook of Organic Chemistry*. Springer: Berlin, 1995.
2. *Beilstein Handbook of Organic Chemistry*. Beilstein Institut für Literatur der Organischen Chemie, Frankfurt am Main, 1881–2001.
3. Cohen N, Benson SW. *Chem. Rev.* 1993; **93**: 2419–2438.
4. Šatkovskiene D, Gineityte V. *J. Mol. Struct. (THEOCHEM)* 1994; **311**: 137–140.
5. Bartlett RJ, Purvis GD. *Int. J. Quantum Chem.* 1978; **14**: 561.
6. Taylor PR. *Lecture Notes in Quantum Chemistry: European Summer School*. Springer: Berlin, 1994.
7. Wiberg KB. *J. Comput. Chem.* 1984; **5**: 197.
8. Dewar MJS, Storch DM. *J. Am. Chem. Soc.* 1985; **107**: 3898–3902.
9. Ibrahim MR, Schleyer PvR. *J. Comput. Chem.* 1985; **6**: 157–167.
10. Yala Z. *J. Mol. Struct.* 1990; **207**: 217–223.
11. Born M, Oppenheimer JR. *Ann. Phys.* 1927; **84**: 457.
12. Koch W, Holthausen MC. *A Chemist's Guide to Density Functional Theory*. Wiley-VCH: Weinheim, 1999.
13. Hehre WJ, Radom L, Schleyer PvR, Pople JA. *Ab Initio Molecular Orbital Theory*. Wiley: New York, 1986.
14. Becke AD. *Phys. Rev. A* 1988; **38**: 3098.
15. Lee C, Yang W, Parr RG. *Phys. Rev. B* 1988; **37**: 785.
16. Becke AD. *J. Chem. Phys.* 1993; **98**: 5648.
17. Hehre WJ, Ditchfield R, Pople JA. *J. Chem. Phys.* 1972; **56**: 2257–2261.
18. Hariharan PC, Pople JA. *Theor. Chim. Acta* 1973; **28**: 213–222.
19. Dunning TH. *J. Chem. Phys.* 1989; **90**: 1007.
20. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresmann JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill P. M. W., Johnson B, Chen W, Wong MW, Andres JL, Head-Gordon M, Replogle ES, Pople JA. Gaussian 98, Revision A.7. Gaussian: Pittsburgh, PA, 1998.
21. Cox JD, Pilcher G. *Thermochemistry of Organic and Organometallic Compounds*. Academic Press: London, 1970.