Thermochimica Acta 495 (2009) 14–21

Contents lists available at ScienceDirect

Thermochimica Acta

journal homepage: www.elsevier.com/locate/tca

Correlations between MO eigenvectors and enthalpies of formation for simple organic substances

Matthew C.J. Lee*, Derek W. Smith

Department of Chemistry, University of Waikato, Private Bag 3105, Hamilton, New Zealand

ARTICLE INFO

Article history:
Received 24 February 2009
Received in revised form 21 April 2009
Accepted 20 May 2009
Available online 3 June 2009

Keywords:
Enthalpy of formation
Additivity
Semiempirical
Ab initio
MOs

ABSTRACT

Bondingness, originally used in a qualitative analysis of the barrier to rotation in ethane, has been used to model $\Delta_f H^\circ$ (g) for simple organic substances. The model is parameterised with a set of 345 molecules including alkanes, alkenes, alkynes, alcohols, ethers, aldehydes, ketones, carboxylic acids, esters, alkenoates, amines, amides, diazenes, nitriles, nitroalkanes, nitrates, thiols and benzenoids. The model is compared with a current empirical scheme as well as a comparison of variations of the model using different simple steric potentials. Using bondingness and the most approximate quantum chemical models a model can be formulated that is comparable with empirical group methods but requiring less parameters.

1. Introduction

The enthalpy of formation $\Delta_f H^\circ$ can be calculated with the following methods: empirically by an additivity scheme, where this is most successfully done with a group method; semiempirically by a molecular mechanics method; by composite *ab initio* methods like G*n* or W*n* ($n = 1, 2, 3$ or 4) or by *ab initio* and density functional theory (DFT) methods corrected with group equivalents.

1.1. Empirical additivity schemes

In additivity schemes to a first approximation, the atomization energy of a molecule is approximated as the sum of constant transferable bond energy terms between any two atoms A and B [E°(A–B)]. The value of the bond energy is determined by the two atoms of the bond, and the order of the bond as specified in a valence bond structure. This approximation works for the higher members of a homologous series like the *n*-alkanes (C6 and above) but fails for structural isomers and the lower members of the series. This indicates that the C–C bond energy [E(C–C)] varies according to the groups attached. Put another way, the concept of a constant transferable bond energy holds so long as the nearest neighbours remain the same.

To account for the thermochemical differences between structural isomers, one needs to take into account the chemical environment of the bond, and this is what the group methods do. On the basis that the energy of a bond is constant as long as the nearest neighbours are the same, a molecular fragment about a bond or atom that includes all nearest neighbour atoms may be assigned a group parameter. Whether parameters are designated to a molecular fragment or bonds distinguished by the chemical environment of the neighbour atoms, the results are equivalent when the same number of parameters are used. The three common group methods by Laidler [1], Benson and Buss [2] and Allen [3] have been shown to be equivalent by Cox and Pilcher [4]. Also some account must be made for steric strain. This is usually done by multiplying the number of *gauche* 1,4 interactions in the molecule by an appropriate coefficient. The performance of a group method is as good as if not superior to other methods, if parameters are derived from accurate experimental data. A common difficulty encountered however, is that a molecule may contain a molecular fragment for which a group parameter cannot be derived from the available experimental values.

Wodrich and Schleyer [5] have recently published a group method with no parameters ascertained in a least squares way. Instead the parameters are fixed by the values of representative molecules. We note that a geminal H–H term is common to all the Gronert [6], Wodrich and Schleyer [5] and Smith [7] schemes with the corollary the latter two are equivalent. Smith determined the geminal H–H term by the method of squares. Subsequently Rogers and co-workers have proposed another scheme that minimises the number of group parameters [8].

---

* Corresponding author. Tel.: +64 7 838 4027.
E-mail address: mcjl1@waikato.ac.nz (M.C.J. Lee).

0040-6031/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.tca.2009.05.011

1.2. *Ab initio methods*

If no data exist for any species similar to a molecule under examination, a high level correlated *ab initio* calculation can be performed at a high cost in computer time. However this is only possible for very small molecules, with no more than 10 first row atoms. To reduce computer time for larger molecules, approximations of severity commensurate with molecular size must be made and parameters sought to compensate for the approximations made by the chosen model's reduced ability to calculate the full electron correlation energy and the reduced basis set size (basis set truncation error). Composite methods automatically extrapolate correction terms by varying the basis set and choosing higher and lower levels of theory to approximate corrections for such things as basis set truncation error. Composite methods minimize computational cost by using the lowest level adequate at each step in the procedure, to reproduce experimental data within or close to the limits of the experimental error of a test set of molecules. These methods are usually tested over a chemically diverse test set of small compounds. DFT methods calculate electron correlation at a significantly reduced cost. DFT local density models are known to overestimate bond dissociation energies, while HF is known to underestimate these. DFT methods are believed to perform quite well [9–11], but when tested on large test sets with larger molecules seem to perform less adequately than is commonly believed [12,13]. Bond [13] has tested the composite methods over a test set of larger compounds that more typically represent molecules in organic chemistry as well as testing the popular B3LYP density functional model which was found to perform unacceptably with the limits defined in his work.

G4 and W4 are the latest in the series, Gn and Wn (n = 1, 2, 3, 4) [14–18] of composite procedures where in G4 parameters have been extended for the first, second and third row compounds. Only the first and second rows were examined in the procedures prior to G4. Recently Schwabe and Grimme [19] have proposed double hybrid density functionals as well as modified post-HF methods without calculation of vibrational data to calculate reaction energies for larger molecules.

1.3. *Molecular mechanics*

Molecular mechanics (MM) use empirical functions related to molecular geometry. The functions used by a MM method are the method's force field.

A force field specifies, among other things, potential energy. When the potential energy is related to nuclear positions of the molecule, a potential energy surface (PES) can be defined. The PES is a multidimensional nonlinear function of the molecular nuclear positions. A molecular geometry ascertained at a minimum on the PES is then considered geometry optimized. The sum of the potential energy functions at the resultant molecular geometry gives a steric energy (SE). The SE is parameterised with bond increments to calculate $\Delta_f H^\circ$ values.

1.4. *Semiempirical quantum mechanics*

For large chemical systems as encountered in organic chemistry and biochemistry and pharmaceutical research, the semiempirical neglect of diatomic differential overlap methods (NDDO) provide an alternative to molecular mechanics. These methods have a minimum basis set that ignores core electrons with the common approximation that atomic orbitals on neighbouring atoms do not overlap. NDDO methods are parameterised to experimental data to reproduce equilibrium geometries, heats of formation, dipole moments and ionization potentials. The common methods are Austin model 1 (AM1 [20]), modified neglect of differential overlap (MNDO [21] and MNDO/d [22]) and parametric method 3 (PM3 [23]). PM3 has been updated by PM6 [24]. Jorgensen *et al.* have included extra terms in the core repulsion formula to apply a pairwise distance directed Gaussian function (PDDG) between bonded atoms [25–27]. AM1 has been reparameterised to a training set of 1736 molecules and is now called RM1 [28]. RM1 is easily implemented in programs that already have AM1 as no line of code needs to be changed except for the values of the parameters. Because NDDO methods are parameterised from experimental data they calculate $\Delta_f H^\circ$ directly without calculating $\Delta H(T)$. However NDDO $\Delta_f H^\circ$ values are not accurate enough to correctly order the stability of structural isomers. Jorgensen and co-workers [29] have recently compared semiempirical MO methods.

1.5. *Systematic corrections*

The semiempirical methods and *ab initio* methods can be improved with the use of atom, bond or group equivalents methods. In the simplest of these only atom equivalents are ascertained. Wiberg [30,31] and subsequently Ibrahim and Schleyer [32] independently ascertained group equivalents for HF methods.

Allinger *et al.* included two more terms TOR and POP [33–39]. TOR is a correction for low lying torsional vibrations not accounted for in the harmonic approximation of vibrational frequency calculations and is approximated by a coefficient with the number of single bonds in a molecule about which there is free rotation, excluding methyl groups. POP is a correction for excess energy in $\Delta_f H^\circ$ due to population of higher energy conformers.

Herndon [40] ascertained atom equivalents by least squares estimates for the total energy (ΔE_{Tot}) and number of carbon and hydrogen atoms over a group of 65 saturated and unsaturated, as well as, strained hydrocarbons for the HF model. Liu and Chen [41] retrained the Herndon test group for DFT and MP2 single point energy calculations with large basis sets geometry optimized and thermally corrected with a smaller basis set. This was done with similar regression analysis as Herndon, but included a regression constant.

Habibollahzadeh *et al.* [42] ascertained valency dependent atom equivalents for DFT when ΔE_{Tot} is corrected with $\Delta H(T)$ calculated and geometry optimized with the 6-31G(d,p) basis set.

Mole *et al.* [43] ascertained atom equivalents for six DFT models using a test group of 23 molecules and showed B3LYP to perform best.

Repasky *et al.* [44] used a training set of 329 molecules and a test set of 583 molecules including the training set, to ascertain 61 group equivalents including TOR for AM1, MNDO and PM3. AM1 and PM3 performed about equally well with PM3 having a slightly better mean absolute error.

Winget and Clark [12] have tested the B3LYP density functional method with atom equivalents over 845 compounds.

Delley [45] has compared 25 electronic structure models over test groups ranging from a subset of 234 molecules for MP2, to the complete set of 592 molecules and atoms for a number of DFT models.

1.6. *Bondingness*

In our previous article we used a single parameter to describe the variation in C–C bond energies [E(C–C)] to account for the variation in molecular structure among the structural isomers of alkanes [46]. We termed this π-antibondingness [$B^*(\pi)$] which Smith [47,48] considered might be a consequence of the antibonding effect within the occupied MOs of a molecule. Here we define π-bondingness [$B(\pi)$] to be the negative of $B^*(\pi)$ such that $B^*(\pi) = -B(\pi)$; likewise for σ-bondingness $B^*(\sigma) = -B(\sigma)$. By ascertaining parameters for $B(\pi)$ in different chemical environments we