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

A B S T R A C T

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.

© 2009 Elsevier B.V. All rights reserved.

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).

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, G*n* and W*n* (*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 over-lap (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 pair-wise 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 anti-bonding 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

have extended our method to include a more diverse set of simple organic molecules.

The basis for using bondingness originated from a qualitative investigation of the antibonding effect and its usefulness as a qualitative explanation for the barrier to rotation in ethane [47,49]. The antibonding effect is that the destabilization of the antibonding MO (Ψ_2) is always greater than the stabilization of the bonding MO (Ψ_1) relative to the energies of the AOs (ϕ_1 and ϕ_2) combined to form the MOs.

If MOs for ethane are constructed from MOs from methyl radicals with geometry combined to form ethane, three MOs of a_1 symmetry and two pairs of degenerate e symmetry MOs are obtained. The barrier to rotation is attributed to overlap repulsion between C–H bonds and the C–H bonding predominantly occurs in the e MOs. The antibonding effect in the occupied $1e'$ and $1e''$ MOs of eclipsed ethane is greater than in the $1e_u$ and $1e_g$ MOs of staggered ethane. This is a significant π antibonding effect in the highest occupied molecular orbitals (HOMOs) e_g and e'' in the staggered and eclipsed conformations, respectively. The idea that overlap repulsion between vicinal C–H bonds should be accounted for in an additivity scheme was suggested by Smith [48] in 1999 and subsequent schemes were developed in 2001 [50] and 2005 [46].

A $B(\pi)$ value can be calculated from the AO coefficients in a MO between two atoms by adding the $B(\sigma)$ term to the dot product of the p AO coefficients on each atom. The $B(\sigma)$ is an orthogonal transformation of the AO coefficients. The transformation being the projection of the p AOs along the abstract vector between the atoms. The projection formula can be employed to ascertain the σ component of the p orbitals on each atom where the sign of the abstract vector between the atoms (**r**) is reversed for each atom, where **c** is a vector of p AO coefficients for the ith and jth atoms as shown in Eq. (1). A positive value by Eq. (1) corresponds to a bonding σ interaction, while a negative value is an antibonding σ interaction. From a $B(\sigma)$ value a $B(\pi)$ value is then calculated by Eq. (2).

$$B(\sigma) = \frac{\mathbf{r}}{|\mathbf{r}|} \cdot \mathbf{c}_i \times \frac{-\mathbf{r}}{|\mathbf{r}|} \cdot \mathbf{c}_j \tag{1}$$

$$B(\pi) = \mathbf{c}_i \cdot \mathbf{c}_j + B(\sigma) \tag{2}$$

The final $B(\pi)$ value for the molecule was ascertained by first summing $B(\pi)$ values in occupied MOs only between bonded carbon atoms with the AO coefficients of a particular MO, to get a $B(\pi)$ value for each MO for carbon-carbon bonding. The $B(\pi)$ in each MO is ordered according to its corresponding MO from highest to lowest. Contiguous negative values are added to give $B(\pi)$ for the molecule, unless a single negative value having zero or positive $B(\pi)$ for the MOs about it, is less (more negative) than a sum of contiguously negative values. In this case, the more negative value is used.

1.7. Notation for algorithms in $B(\pi)$

Bondingness (B), in particular π-bondingness [$B(\pi)$] is calculated over occupied MOs only.

Previously [46] $B(\pi)$ was summed across all C–C bonds in a MO to give a value for that MO. Contiguously negative MOs were then added together to give $B(\pi)$ for the molecule. This idea is modified in the A1.0 algorithm according to the chemical environments determined by a bond labelling algorithm which gathers information from adjacent bonds. The bonds of similar chemical environments are grouped and $B(\pi)$ is determined for each type of chemical environment. Consider 1-butene: double bonds are separated from single bonds so that $B(\pi)$ is summed between C2 and C3, and C3 and C4 for each occupied MO and contiguously negative values are added to give a specific $B(\pi)$ for C–C bonds. This is done again between C1 and C2 to give a specific $B(\pi)$ for the double bond, which is negligibly negative or zero. The molecule can

also be grouped into double bonds, single bonds and single bonds next to double bonds. The A1.1 algorithm is exactly the same as the A1.0 algorithm between atoms with bond orders less than some predefined value (normally 1.5, 2 or 3). If the bond order matches or exceeds the predefined value, then the algorithm continues in a similar manner to A1.0 with the single difference contiguously positive $B(\pi)$ values ascertained for the MOs are added together.

Parameters are ascertained by a least squares estimate (LSE) k for each specific $B(\pi)$, the product of k and $B(\pi)$ giving an adjusted specific $B(\pi)$ [$B(\pi)_{adj}$].

The A2.0 and A2.1 algorithms differ by not grouping the bonds. Instead a $B(\pi)$ value is determined for each bond by the A1.0 algorithm for A2.0 and by the A1.1 algorithm for A2.1. Then the bonds are grouped according to chemical environment and the values from the bonds of common chemical environments are added to give the specific $B(\pi)$ for each chemical environment.

The final algorithms tested were the A3.0 and A3.1 algorithms which sum all negative $B(\pi)$ across all bonds and in each occupied MO of common chemical environment. Where the A3.1 algorithm differs by summing the positive $B(\pi)$ for bonds with bond order above or matching the defined value.

1.8. Notation in chemical environment

A further syntax was developed for computer programs to distinguish different chemical environments. This describes the chemical environment of a bond. The algorithm employed creates a label beginning with the two atoms of the bond followed by an underscore and the bond order of the bond,[1] the adjacent bonds are then similarly labelled and placed in brackets followed by the number of that kind of bond adjacent to the bond in the first part of the label. C–H bonds are usually ignored. Thus the sp^2sp^3 bond of 3-methylbut-1-ene, ignoring C–H bonds, would be labelled CC_1(CC_1)2(CC_2)1. Ignoring C–C bonds this becomes CC_1(CC_2)1. Ignoring C–C and C–H bonds prevents us from distinguishing terminal methyl, secondary, tertiary and quaternary bonding. This is intentional as we anticipate the variation in specific $B(\pi)$ to accommodate these differences and reduces clutter in the chemical environment notation.

In a few instances when there are chemical environments with small $B(\pi)$ or environments are unique to one molecule only or an environment is collinear, we combine the environment notation with another similar chemical environment. For instance CC_1(CN_1)2 is an environment unique in the test set to 1,1-dinitropropane. This was combined with CC_1(CN_1)1 environments to give CC_1(CN_1)1-2.

By the different combinations of the notation for a specific $B(\pi)$ it is evident there are many ways to quantify $B(\pi)$. However in alkanes there is little distinction between the different chemical environments of each C–C bond. There is perhaps more π-antibonding in terminal C–C bonds, but the different ways of ascertaining $B(\pi)$ give similar results. However for molecules with heteroatoms or with bond orders greater than one, specific $B(\pi)$ values can vary a great deal according to the method implied by the variations in notation. In this work a trial and error approach is adopted to ascertain the significance of an algorithm used to calculate $B(\pi)$ values.

The group methods are perhaps a little passé on the frontiers of theoretical chemistry. However they, as well as the group equiva-

---

[1] The number is actually a bond descriptor of type integer assigned in a quantum chemical input file where for the most part it represents the approximate bond order of the bond, except in the case where the bond order is 1.5, as is the case for benzenoid or nitro compounds. In this instance the integer 5 is used and we keep this notation in this work. Therefore the carbon-carbon bond in a benzene ring has the notation CC_5(CC_5)2.

lents used to correct *ab initio* or DFT total energies, do exhibit the parameters used in their methodology in a more accessible way than semiempirical and composite methods. The results of Pedley's [51] method and experimental values are organized according to the functional groups of the molecules and we have chosen his experimental values (with a few exceptions) and consequently his method to compare ours with. In extending our method we have kept in mind the question: what minimum number of parameters for specific $B(\pi)$ and steric terms are required for a model that calculates $\Delta_f H^\circ$ values, with errors comparable with current group methods, *viz.* Pedley's method?

2. Method

With any model's calculated $\Delta_f H^\circ$ values, results are most accurate when parameters are ascertained over a homologous series of closely related molecules for which experimental values are known, a confident theoretical value can then be calculated for other members of the homologous series for which there is no experimental value. One would have very little confidence in a value calculated for a non-member of the homologous series. To make a more robust model, that might be able to determine more accurate calculated values for a greater diversity of chemical structures, a test group of 345 molecules (see supporting information) has been investigated to parameterise against $\Delta_f H^\circ$ values. To compare with Pedley's scheme we used the experimental $\Delta_f H^\circ$ (g) values of his work [51], we also test against the $\Delta_f H^\circ$ (g) available from the NIST chemical webbook [52]. Recently more data have been ascertained for alkenoates by Emel'yanenko *et al.* [53].

Pedley's $\Delta_f H^\circ$ values for alkenes are ascertained from enthalpies of vaporization and combustion and an enthalpy of reaction for 1-hexene from Wiberg and Wasserman [54]. A subsequent paper by Wiberg *et al.* [55] not used by Pedley suggests improved accuracies in $\Delta_f H^\circ$ values particularly for *cis*-pent-2-ene and *cis*-hex-3-ene. Wiberg has reviewed these methods [56]. Enthalpies of reaction ascertained by Rogers *et al.* [57–61] are also not used by Pedley.[2]

We anticipated developing a model that might eventually be suitable for calculating $\Delta_f H^\circ$ (g) for amino acids and peptides, and so molecules that exhibited functional groups common in amino acids were chosen, where there were at least a few $\Delta_f H^\circ$ (g) values in a group according to the grouping in Pedley's work. If the 345 molecules are grouped in the same way as Pedley, then 18 sub-groups of the 345 molecules test can be distinguished. The 18 groups that the molecules of the test group belong to are: alkanes, alkenes, alkynes, alcohols, ethers, aldehydes, ketones, carboxylic acids, esters, alkenoates, amines, amides, diazenes, nitriles, nitroalkanes, nitrates, thiols and benzenoids. Of the 345 molecules eight extra molecules not available in Pedley's compilation were also included. The following $\Delta_f H^\circ$ values and alternative reference sources were used: the JANAF tables [62], 3-tert-butyl-2,2,4,4-tetramethyl-3-pentanol, hydrogen cyanide, ethenetricarbonitrile and ethenetetracarbonitrile; Benson and Garland [63], pent-1-yne, pent-2-yne, 3-methylbut-1-yne and 3,3-dimethylbut-1-yne. Over the test set of 345 molecules, excepting the latter list, using around 107 parameters, Pedley's method has a standard deviation of 7 kJ mol^{-1}. This was our benchmark. A program was written to ascertain all the molecular fragments required for a group method like Pedley's. The number of molecular fragments based on a bond and its next nearest neighbours for the test set of 345 molecules was 236, for which Pedley only supplies values for about 107.

Parameters are ascertained for an additivity scheme in the following way. Some bond energies can be fixed from the atomization

enthalpies of methane, diamond [7], water, ammonia and hydrogen sulphide obtained from the JANAF tables [62] to give the values shown in Table 1.

Results are slightly improved however if LSEs are ascertained for E(C–C), E(O–H) and E(N–H) which we have done in the current model. The remaining bond energies are ascertained by the least squares method, along with other LSEs for $B(\pi)$ and a steric function. Bond energies are ascertained for bonds according to the two atoms involved in the bond only. For these parameters we are not concerned about the neighbour atoms which are used to distinguish a group in a group method. However $B(\pi)$ parameters are ascertained for bonds in different chemical environments.

Some functional groups have a non-spherical arrangement of hydrogen atoms about a heavy atom, for instance amino and hydroxyl groups. These groups have different interactions with neighbouring groups or atoms, depending on which direction the hydrogens are oriented relative to the other group or atom. The phenomenon where the spatial direction of something has a bearing on an outcome is anisotropy. Anisotropy in a functional group is usually accounted for by creating a pseudo atom to represent the lone pair of electrons usually present in such a functional group. The anisotropic problem may to some degree be accounted for by bondingness between nonbonded atoms in a molecule. Urey–Bradley forces [64] are between nearest neighbour atoms i.e. atoms not bonded to each other, but bonded to a common atom. $B(\pi)$ is directional, as is $B(\sigma)$ when p or higher l quantum number atomic orbitals are involved. Nonbonded $B(\pi)$ [nb-$B(\pi)$] is only distinguished by the two atoms involved, and no further chemical environment information. It therefore does not accrue parameters as quickly as the algorithm for bonded $B(\pi)$ values. Over the test group of 345 molecules the algorithm distinguishes the following nb-$B(\pi)$: CC, CN, CO, CS, NN, NO and OO. An nb-$B(\pi)$ value is arrived at by summing all $B(\pi)$ over all MOs for common interaction types. This much simpler algorithm works just as well if not better than an algorithm that sums contiguous negative $B(\pi)$ for each MO.

An electrostatic energy term ($V_{stat.}$), which was beneficial for molecule groups like alcohols, was also parameterised. In Eq. (3) $V_{stat.}$ is the electrostatic energy, C is a LSE, r_{ij} is the interatomic distance between atoms i and j and q_i and q_j are the said atoms respective charges.

$$V_{stat.} = C \sum \frac{q_i q_j}{r_{ij}} \tag{3}$$

To account for steric repulsion in a molecule (S_{ij}) two simple functions were tested. A purely destabilizing steric parameter would neglect the dispersive term of a formula modelling the van der Waals potential. Thus in Eq. (4) n was varied to find the best fit, r_{ij} was the distance between nonbonded H atoms and A and n were ascertained by a nonlinear least squares grid search for $n = 6$ or 12. We call this simple steric repulsion (SSR).

$$S_{ij} = A \sum \frac{1}{r_{ij}^n} \tag{4}$$

Table 1
Bond energies obtained directly from CH$_4$, diamond, H$_2$O, NH$_3$ and H$_2$S.

	Bond energy (kJ mol^{-1})
E(C–H)	415.87
E(C–C)	357.40
E(O–H)	463.50
E(N–H)	390.86
E(S–H)	366.74

---

[2] J.F. Liebman, Examiner's report, 2008.

This can be rearranged to take the form of the repulsive term of a Lennard–Jones potential. If $r* = A^{1/n}$ then Eq. (4) has the form:

$$S_{ij} = \sum \left(\frac{r_*}{r_{ij}} \right)^n \tag{5}$$

If the Lennard–Jones potential is written:

$$V_{vdW} = \frac{n\varepsilon}{n-m} \left[\frac{m}{n} \left(\frac{r_o}{r} \right)^n - \left(\frac{r_o}{r} \right)^m \right] \tag{6}$$

where ε is the well depth, r_o is interaction distance minimizing V_{vdW}, m and n are integers where $m < n$ and are the hardness parameters for the attractive and repulsive terms, respectively and r is the interatomic distance. We parameterise V_{vdW} with the form shown in Eq. (7).

$$S_{ij} = \sum_r Ar^{-n} - Br^{-m} \tag{7}$$

The A and B variables of Eq. (7) can be converted to ε and r_o of Eq. (6) by use of Eq. (8)

$$r_o = \left(\frac{nA}{mB} \right)^{1/n-m} \tag{8}$$

Substituting r_o for r into Eq. (7) gives ε.

We also tested Halgren's [65,66] buffered 14 7 potential, which was not competitive with SSR or the Lennard–Jones potentials unless modified so that the nonbonded interactions with carbon were ignored.

Nonbonded interactions are ascertained between H, N and O, where we distinguish, O attached to one or two atoms as type O7 and O6, respectively and N attached to one, two or three atoms as types N61, N9 and N8, respectively. Hydrogen atoms were either type 6 attached to carbon or type 21 attached to something else. This gave 15 steric interaction types over the test set of 345 molecules. The numbers distinguishing the atom types are roughly modelled on the types of atom environments Halgren distinguishes for the buffered 14 7 potential.

All the nonbonded distances are grouped according to the notation of the respective steric interaction and used to sum the terms in Eqs. (4) or (7) and LSEs are obtained for each term. By this method common steric interaction types are more heavily weighted at longer distances. With regard to the Lennard–Jones potential there is a paucity of data at closer distances near the r_o region or closer. Thus the sign of the LSE does not in all cases conform with the sign of the respective term in the function so that some functions are a negative of the normal function and have a maximum at r_o or the two terms combine so there is no maximum or minimum and are asymptotic like. It is important in a molecular mechanics model that the potential have a minimum at r_o because the nonbonded interactions in the force field are a major component in the construction of the PES and improper functions would result in molecules flying to pieces or collapsing in on themselves. In this work this outcome

is regrettable but the author thinks we can live with the outcome as long as we realize when calculating $\Delta_f H^\circ$ values for molecules outside the test set we must be aware of some functions approaching minus infinity as inter atomic distance approaches zero. For these types of nonbonded interactions the model is limited in these instances to nonbonded distances no shorter than those encountered in the test set. The function is otherwise optimized at longer distances for the currently available data. The shortest distances encountered in the test set are given in Table 2, along with an indication of the gradient of those functions approaching minus infinity as interatomic distance approaches zero. These functions are indicated with an S, SS or G. Where an S indicates the function is steep or is about to become steep and shorter nonbonded distances will likely be erroneous. If the function is already steep for a few data in the test set then the gradient is indicated with an SS. A G indicates a more gradual potential change and the function may still be useful at slightly shorter distances. Data for all the algorithms are given in supporting information.

PM3 is the most approximate quantum chemical model we used. A model can be developed from its output in the following way. The test set of 345 molecules was geometry optimized and MO Eigen vectors and zero point energy (ZPE) determined at the PM3 level using the Spartan© [67] software package. With these calculations the following terms to be parameterised were determined: $B(\pi)$ over bonds and over nonbonded 1,3 interactions [nb-$B(\pi)$], electrostatic energy, steric energy using Eq. (7), a low lying torsional vibrational energy term (TOR) and bond energy terms.

The ZPEs were scaled by a factor of 0.9761 for PM3 and 0.9207 for HF/3-21G(*) [68,69]. A factor of 0.8217 for the STO-3G basis set was ascertained by fitting STO-3G ZPEs to 3-21G ZPEs. The LSEs were determined using a weighted least squares linear regression. The function for the weighting was the inverse square of the experimental error given in the literature for a molecule's $\Delta_f H^\circ$. These parameters were ascertained for each algorithm A1.0, A1.1, A2.0, A2.1, A3.0 and A3.1 and the standard deviations in kJ mol^{-1} over the 345 test set (S345) for the respective algorithms were 7.8, 8.2, 8.6, 8.4, 7.1 and 7.1. The t-statistics were then examined and a composite of algorithms (A1-2-3) was constructed so that for each $B(\pi)$ parameter, the algorithm with the highest t-observed value was assigned to calculate $B(\pi)$ in that chemical environment. The A1-2-3 algorithm then gave a standard deviation of 7.9 kJ mol^{-1}. With this procedure a total of 79 parameters are obtained by linear least squares regression analysis.

Changing only the quantum chemical model and ascertaining ZPEs and geometry optimizing the test set of molecules at the respective level of model and basis set, parameters were sought at the HF/STO-3G and HF/3-21G(*) levels.

The test set for HF/3-21G(*) is for 344 molecules and excludes 1-hexadecanol. Calculation of fundamental frequencies and ZPE at this level on this molecule exceeds the memory limitations in a typical 32 bit computer.

Table 2
The minimum nonbonded distance found for each steric interaction in a molecule in the test set and the appearance of the function at that distance, where G is gradual, S is steep and SS is very steep. Minus means the function is decreasing, while a plus means the function increases as nonbonded distance decreases. Space indicates the function is a typical Lennard–Jones function and is not remarkable.

		H21-H6	H6-H6	H6-N61	H6-N8	H6-N9	H21-O7	H6-O6	H6-O7	N61-N61	N8-O7	O7-O7
Nonbonded distance (Å)		2.2	1.8	3	2.5	2.3	2.3	2.3	2.2	3.6	2.6	2.6
Number of interactions in test set		1048	21567	185	352	137	32	1256	1431	21	44	48
3-21G(*)	A1-2-3		+S	−S	−S					−G		+G
	A3.1			−S	−S					−S		+S
STO-3G	A1-2-3		−G	−S	−G					−SS	−SS	−G
	A3.1		−G	−SS		+G				−S	−SS	−S
PM3	A1-2-3	+G	−G	−S	−S		−G			−S		+G
	A3.1		−G	−G						−S		+S

Table 3
Experimental – calculated $\Delta_f H^\circ$ ($\Delta\Delta_f H^\circ$) values for the small molecule test set. Values are in kJ mol^{-1}.

	Experimental error	PM3		STO-3G		3-21G(*)	
		$\Delta\Delta_f H^\circ$ A1-2-3	$\Delta\Delta_f H^\circ$ A3.1	$\Delta\Delta_f H^\circ$ A1-2-3	$\Delta\Delta_f H^\circ$ A3.1	$\Delta\Delta_f H^\circ$ A1-2-3	$\Delta\Delta_f H^\circ$ A3.1
Methane	0.4	0.5	1.4	1.3	0.8	0.8	1.1
Ethane	0.3	3.1	4.5	2.7	2.9	3.0	3.2
Propane	0.5	3.8	5.9	4.7	4.8	3.2	4.2
Butane	0.6	3.0	1.3	0.5	0.6	3.0	0.6
Ethene	0.3	1.7	0.6	0.3	0.5	0.6	0.4
Ethyne	0.7	−0.4	0.1	0.3	0.4	0.4	0.5
Methanol	0.2	0.2	−0.3	0.6	0.4	0.8	1.1
Dimethyl ether	0.5	0.9	1.8	1.0	0.3	0.5	0.1
Formaldehyde	0.5	2.3	2.3	2.1	2.1	2.2	3.3
Propanone	0.7	1.9	−0.7	0.8	0.0	1.4	−1.4
Formic acid	0.5	0.2	0.1	0.2	−0.1	−0.3	−0.1
Methyl methanoate	0.7	−1.0	0.9	1.6	0.3	0.0	−0.6
Methyl 2-propenoate	0.8	−12.4	−4.8	−7.4	−9.8	−6.6	−7.0
Methylamine	1.0	−1.5	1.9	−0.4	1.6	0.5	1.2
Methanamide	2.0	2.1	3.8	0.0	1.4	−1.8	−2.9
Trans-butylmethyldiazene	2.1	2.0	0.9	0.2	5.2	0.0	0.1
Hydrogen cyanide	8.4	−15.9	−9.5	−27.2	−27.6	−38.4	−24.0
Nitromethane	0.6	−0.1	0.0	−0.5	−0.9	0.0	−0.6
Methyl nitrate	4.3	−1.4	2.6	−1.0	−5.5	2.0	0.9
Methanethiol	0.6	1.0	0.1	0.7	0.7	0.4	0.3
Benzene	0.7	0.1	0.1	−0.5	−0.4	0.0	−0.5
S21[a]		4.8	3.2	6.5	6.9	8.9	5.8
S345[b]		7.9	7.1	6.1	6.2	6.6	6.1

[a] Standard deviation for the 21 molecules in the small molecule test set.
[b] Standard deviation for the 345 molecules over which parameters were ascertained.

In the composite of algorithms the prescription of assigning the algorithm with the largest t-observed value, for calculating $B(\pi)$ in the various chemical environments was not strictly followed, and this was the case with CN_3 $B(\pi)$ for STO-3G and 3-21G(*) and with NN_2 for PM3. These had the largest t-observed value with A1.0 or A2.0, but all A1.0, A2.0 and A3.0 algorithms give a negligible $B(\pi)$ value and an LSE of the order of 10^4 to 10^8. In this instance the next highest sensible t-observed value was used to select either A1.1, A2.1 or A3.1.

It must be stressed that in most models methane has a specific parameter, or is left out of the test set and some test sets also exclude ethane. The 345 test set includes all the smallest molecules

in each sub-group of molecules. It is with these molecules (the small molecule test set) that we want to monitor and check the performance of each model. Of particular importance to us are the values calculated by the model for methane, ethane, propane and butane. These molecules are shown in Table 3 along with experimental $\Delta_f H^\circ$ – calculated $\Delta_f H^\circ$ ($\Delta\Delta_f H^\circ$) and the standard deviation for the small molecule test set (S21) as well as S345. Data for the remaining algorithms is supplied in supporting information. The S21 values are dominated by the largest experimental and calculated disparities. For Pedley's experimental values of Table 3, S21 values are dominated by the hydrogen cyanide and to a lesser extent methyl 2-propenoate $\Delta\Delta_f H^\circ$ values.

Table 4
Comparison of SSR, the Lennard–Jones and buffered 14 7 potentials. All $B(\pi)$ values are ascertained at the HF/3-21G(*) level. Values are in kJ mol^{-1}.

	Experimental error	SSR		Lennard–Jones		Buffered 14 7	
		$\Delta\Delta_f H^\circ$ A1-2-3	$\Delta\Delta_f H^\circ$ A1.1	$\Delta\Delta_f H^\circ$ A1-2-3	$\Delta\Delta_f H^\circ$ A1.1	$\Delta\Delta_f H^\circ$ A1-2-3	$\Delta\Delta_f H^\circ$ A1.1
Methane	0.3	1.5	1.1	1.1	1.1	1.5	1.4
Ethane	0.4	3.2	3.1	2.9	2.7	2.5	0.8
Propane	0.5	4.7	3.3	3.3	3.1	3.0	1.5
Butane	0.7	0.2	3.2	3.1	2.8	2.7	1.4
Ethene	0.5	1.0	2.3	0.5	1.8	2.1	2.4
Ethyne	0.8	0.5	0.1	0.6	0.0	0.5	0.7
Methanol	10.0	4.0	2.7	2.3	2.7	12.6	9.2
Dimethyl ether	0.5	1.5	2.4	2.1	2.1	3.8	1.9
Formaldehyde	0.5	2.3	2.6	1.2	2.1	2.6	5.7
Propanone	0.6	0.9	−1.5	0.5	−2.5	3.8	8.4
Formic acid	0.6	−0.5	−1.7	−0.6	−1.1	0.0	−0.9
Methyl methanoate	12.0	26.0	15.3	22.0	11.4	34.4	15.6
Methyl 2-propenoate	0.8	−5.7	−6.9	−5.2	−6.6	−3.5	−6.6
Methylamine	1.0	2.9	0.7	1.7	2.5	2.4	2.1
Methanamide	7.0	18.7	−3.8	−4.7	−5.3	14.1	9.6
Trans-butylmethyldiazene	2.2	−0.2	0.0	0.0	0.0	−1.2	−1.0
Hydrogen cyanide	8.4	−0.9	−4.2	−18.6	−36.3	−41.0	−42.0
Nitromethane	7.0	−13.0	0.2	−6.4	0.1	−6.8	14.3
Methyl nitrate	1.0	−0.2	0.0	0.1	0.1	−0.3	0.1
Methanethiol	0.6	−0.3	−0.1	0.3	0.3	−1.1	−0.8
Benzene	0.5	0.0	−0.4	0.0	-0.7	1.3	1.9
S21[a]		8.1	4.3	7.0	8.9	12.9	11.3
S344[b]		8.3	9.2	8.2	11.0	9.8	10.9

[a] Standard deviation for the 21 molecules in the small molecule test set.
[b] Standard deviation for the 344 molecules over which parameters were ascertained.

Heretofore the test set has been trained with Pedley's experimental values. This was done primarily to gauge our model with a current empirical model and with weighted regression analysis the two models are comparable. Without weighting data in the regression analysis, standard deviations well below $6\,kJ\,mol^{-1}$ are easily attained but at the expense of the smaller molecules.

To compare SSR, the Lennard–Jones and the buffered 14 7 potentials over the 344 molecule test set, the most recent experimental data are taken from NIST if available, otherwise Pedley's values are still used. This data set is much more difficult to fit and standard deviations for the models are much higher. This is perhaps not surprising as Pedley's work presents a single recommended experimental data per molecule chosen in part for the fit in the trend of molecules with which it is grouped. And for this data those of us not expert thermochemists are pleased to have someone choose for us. Data comparing steric potentials using the A1-2-3 or A1.1 algorithms is shown in Table 4. A comparison of all the algorithms is given in supporting information.

3. Conclusion

The use of $B(\pi)$ in a quantitative model provides a means to calculate gaseous $\Delta_f H^\circ$ comparable with Pedley's empirical scheme. However because $B(\pi)$ can be used to vary the C–C bond energy without extra structural parameters, adjacent C–H and C–C bonds can be ignored and consequently the method does not require as many parameters as a method that uses group equivalents. The steric energy as well as $B(\pi)$ calculations for a molecule also provide a mechanism that will distinguish conformers and even rotamers which group methods are incapable of doing. Both methods do not take into account the population of higher energy conformers present in the gas phase. Population analysis would likely be the next step to improve the accuracy of the model. The models that give the best treatment of alkanes use the modified buffered 14 7 potential and the A1.0 or A1.1 algorithms for $B(\pi)$, but the large S344 shows the model is one of the worst, along with the Lennard–Jones potential and the A1.1 algorithm, over the remaining molecules in the test set. A reasonable compromise is the A1-2-3 algorithm and the Lennard–Jones potential which gives reasonable estimates for the small alkanes as well as having the best overall performance with the lowest S344 of $8.2\,kJ\,mol^{-1}$. We recommend a model with an A1-2-3 algorithm and Lennard–Jones potential which is most robust when unique or collinear parameters are combined so that few or no molecules have a zero $\Delta\Delta_f H^\circ$. From experience the presence of parameters unique to a molecule tends to give more negative Lennard–Jones potentials when using linear least squares regression and Eq. (7). Perl scripts to train a test set of molecules as well as calculate $\Delta_f H^\circ$ using the parameters from the training set are available upon request from the corresponding author. The scripts can assimilate output for molecules calculated with the Spartan'04 or Gamess [70] (update 11-04-2008) quantum chemical programs.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tca.2009.05.011.

References

[1] K.J. Laidler, Can. J. Chem. 34 (1956) 626–648.
[2] S.W. Benson, J.H. Buss, J. Chem. Phys. 29 (1958) 546–572.
[3] T.L. Allen, J. Chem. Phys. 31 (1959) 1039–1049.
[4] J.D. Cox, G. Pilcher, Thermochemistry of Organic and Organometallic Compounds, Academic Press Inc., London & New York, 1970.
[5] M. Wodrich, P. Schleyer, Org. Lett. 8 (2006) 2135–2138.
[6] S. Gronert, J. Org. Chem. 71 (2006) 1209–1219.
[7] D.W. Smith, J. Chem. Soc., Faraday Trans. 92 (1996) 1141–1147.
[8] A.A. Zavitsas, N. Matsunaga, D.W. Rogers, J. Phys. Chem. A 112 (2008) 5734–5741.
[9] K.E. Riley, B.T. Op't Holt, K.M. Merz, J. Chem. Theory Comput. 3 (2007) 407–433.
[10] S.F. Sousa, P.A. Fernandes, M.J. Ramos, J. Phys. Chem. A 111 (2007) 10439–10452.
[11] E.N. Brothers, G.E. Scuseria, J. Phys. Chem. A 112 (2008) 13706–13711.
[12] P. Winget, T. Clark, J. Comput. Chem. 25 (2004) 725–733.
[13] D. Bond, J. Org. Chem. 72 (2007) 5555–5566.
[14] J.A. Pople, M. Head-Gordon, D.J. Fox, K. Raghavachari, L.A. Curtiss, J. Chem. Phys. 90 (1989) 5622–5629.
[15] L.A. Curtiss, K. Raghavachari, G.W. Trucks, J.A. Pople, J. Chem. Phys. 94 (1991) 7221–7230.
[16] L.A. Curtiss, K. Raghavachari, P.C. Redfern, V. Rassolov, J.A. Pople, J. Chem. Phys. 109 (1998) 7764.
[17] L.A. Curtiss, K. Raghavachari, P.C. Redfern, J. Chem. Phys. 126 (2007) 084108.
[18] A. Karton, E. Rabinovich, J.M.L. Martin, B. Ruscic, J. Chem. Phys. 125 (2006), 144108/144101–144108/144117.
[19] T. Schwabe, S. Grimme, Acc. Chem. Res. 41 (2008) 569–579.
[20] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, J. Am. Chem. Soc. 107 (1985) 3902.
[21] M.J.S. Dewar, W. Thiel, J. Am. Chem. Soc. 99 (1977) 4899.
[22] W. Thiel, A.A. Voityuk, Int. J. Quantum Chem. 44 (1992) 807–829.
[23] J.J.P. Stewart, J. Comput. Chem. 10 (1989) 209–220.
[24] J.J.P. Stewart, J. Mol. Model 13 (2007) 1173–1213.
[25] I. Tubert-Brohman, C.R.W. Guimarães, W.L. Jorgensen, J. Chem. Theory Comput. 1 (2005) 817–823.
[26] I. Tubert-Brohman, C.R.W. Guimarães, M.P. Repasky, W.L. Jorgensen, J. Comput. Chem. 25 (2004) 138–150.
[27] M.P. Repasky, J. Chandrasekhar, W.L. Jorgensen, J. Comput. Chem. 23 (2002) 1601–1622.
[28] G.B. Rocha, R.O. Freire, A.M. Simas, J.J.P. Stewart, J. Comput. Chem. 27 (2006) 1101–1111.
[29] K.W. Sattelmeyer, J. Tirado-Rives, W.L. Jorgensen, J. Phys. Chem. A 110 (2006) 13551–13559.
[30] K.B. Wiberg, J. Comput. Chem. 5 (1984) 197–199.
[31] K.B. Wiberg, J. Org. Chem. 50 (1985) 5285–5291.
[32] M.R. Ibrahim, P.v.R. Schleyer, J. Comput. Chem. 6 (1985) 157.
[33] N.L. Allinger, L.R. Schmitz, I. Motoc, C. Bender, J.K. Labanowski, J. Phys. Org. Chem. 3 (1990) 732–736.
[34] N.L. Allinger, L.R. Schmitz, I. Motoc, C. Bender, J.K. Labanowski, J. Am. Chem. Soc. 114 (1992) 2880–2883.
[35] N.L. Allinger, L.R. Schmitz, I. Motoc, C. Bender, J.K. Labanowski, J. Comput. Chem. 13 (1992) 838–841.
[36] N.L. Allinger, L. Yang, I. Motoc, C. Bender, J.K. Labanowski, Heteroatom Chem. 3 (1992) 69–72.
[37] L.R. Schmitz, I. Motoc, C. Bender, J.K. Labanowski, N.L. Allinger, J. Phys. Org. Chem. 5 (1992) 225–229.
[38] J. Labanowski, L. Schmitz, K.-H. Chen, N.L. Allinger, J. Comput. Chem. 19 (1998) 1421–1430.
[39] L.R. Schmitz, K.-H. Chen, J.K. Labanowski, N.L. Allinger, J. Phys. Org. Chem. 14 (2001) 90–96.
[40] W.C. Herndon, Chem. Phys. Lett. 234 (1995) 82–86.
[41] M.H. Liu, C. Chen, J. Comput. Chem. 27 (2006) 537–544.
[42] D. Habibollahzadeh, M.E. Grice, M.C. Concha, J.S. Murray, P. Politzer, J. Comput. Chem. 16 (1995) 654–658.
[43] S.J. Mole, X. Zhou, R. Liu, J. Phys. Chem. 100 (1996) 14665–14671.
[44] M.P. Repasky, J. Chandrasekhar, W.L. Jorgensen, J. Comput. Chem. 23 (2002) 498–510.
[45] B. Delley, J. Phys. Chem. A 110 (2006) 13632–13639.
[46] M.C.J. Lee, D.W. Smith, Thermochim. Acta 435 (2005) 43–47.
[47] D.W. Smith, J. Chem. Educ. 75 (1998) 907–909.
[48] D.W. Smith, Phys. Chem. Chem. Phys. 1 (1999) 4065–4068.
[49] D.W. Smith, J. Chem. Educ. 77 (2000) 780–784.
[50] D.W. Smith, Phys. Chem. Chem. Phys. 3 (2001) 3562–3568.
[51] J.B. Pedley, Thermochemical Data and Structures of Organic Compounds, Thermodynamics Research Center, Texas, 1994.
[52] H.Y. Afeefy, J.F. Liebman, S.E. Stein, NIST Chemistry WebBook, NIST Standard Reference Database Number 69, ed., 2005.
[53] V.N. Emel'yanenko, A.V. Toktonov, S.A. Kozlova, S.P. Verevkin, V. Andrushko, et al., J. Phys. Chem. A 112 (2008) 4036–4045.
[54] K.B. Wiberg, D.J. Wasserman, J. Am. Chem. Soc. 103 (1981) 6563–6566.
[55] K.B. Wiberg, D.J. Wasserman, E. Martin, J. Phys. Chem. 88 (1984) 3684–3688.
[56] K.B. Wiberg, in: J.F. Liebman, A. Greenberg (Eds.), Molecular Structure and Energetics, VCH Publishers Inc., New York, 1987.
[57] D.W. Rogers, E.L. Crooks, J. Chem. Thermodyn. 15 (1983) 1087–1092.
[58] D.W. Rogers, E. Crooks, K. Dejroongruang, J. Chem. Thermodyn. 19 (1987) 1209–1215.
[59] D.W. Rogers, K. Dejroongruang, J. Chem. Thermodyn. 20 (1988) 675–680.
[60] D.W. Rogers, K. Dejroongruang, J. Chem. Thermodyn. 21 (1989) 1115–1120.
[61] D.W. Rogers, K. Dejroongruang, S.D. Samuel, W. Fang, Y. Zhao, J. Chem. Thermodyn. 24 (1992) 561–565.
[62] M.W. Chase, C.A. Davies, J.R. Downey, D.J. Frurip, R.A. Mcdonald, A.N. Syverud, J. Phys. Chem. Ref. Data 14 (1985).
[63] S.W. Benson, L.J. Garland, J. Phys. Chem. 95 (1991) 4915–4918.

[64] H.C. Urey, C.A. Bradley, Phys. Rev. 38 (1931) 1969–1978.
[65] T.A. Halgren, J. Comput. Chem. 17 (1996) 520–552.
[66] T.A. Halgren, J. Am. Chem. Soc. 114 (1992) 7827–7843.
[67] 2004 Spartan. In *Spartan'04*. 18401 Von Karman Avenue, Suite 370, Irvine, CA 92612 U.S.A, Wavefunction, Inc.
[68] B.J. Lynch, Y. Zhao, D.G. Truhlar, Database of Frequency Scaling Factors for Electronic Structure Methods, 2007.
[69] A.P. Scott, L. Radom, J. Phys. Chem. 100 (1996) 16502–16513.
[70] M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, et al., J. Comput. Chem. 14 (1993) 1347–1363.